محمد پور تهمتن رضا علی ۱۳۳۸
مفاهیم و روش های آمار حیاتی و شاخصهای بهداشتی – تالیف دکتر رضاعلی محمد پور تهمتن- این کتاب بر اساس سر فصل مصوب دانشگاهی طراحی شده است.

- ۱۳۸۶- انتشارات آوای مسیح ساری ..... ریال
۲۲۰ ص – جدول – نمودار- ISBN .978-964-2769-05-6
فهرست نویسی بر اساس اطلاعات فیپا (فهرست نویسی پیش از انتشار)
۱- آمار زیستی ۲- شاخصهای بهداشتی ۳- کاربرد روشها در تحقیقات بهداشتی و پزشکی

عنوان پشت جلد به انگلیسی Concepts & Methods of Biostatistics and Health Indices
کتابخانه ملی ایران ۱۰۶۴۱۰۴

مفاهیم و روش های آمار حیاتی و شاخصهای بهداشتی
نویسنده: دکتر رضا علی محمد پور تهمتن
ناشر: ....– انتشارات آوای مسیح ساری
طراحی روی جلد:
حروف چینی:
تیراژ:
چاپ اول: ۱۳۸۶
شابک: ISBN
Printed in Iran
واحد پخش و فروش: ساری- جاده خزر آباد-مجتمع دانشگاه علوم پزشکی مازندران- کتابفروشی مرکزی – تلفن صندوق پستی
بابل: - صندوق پستی تلفن
کلیه حقوق این اثر برای مولف محفوظ است. هرگونه استفاده از مطالب کتاب برای آموزش و تحقیقات با ذکر نام مولف و ناشر بلا مانع است.

دانشگاه علوم پزشکی و خدمات بهداشتی درمانی
مازندران

بسم الله الرحمن الرحیم

مفاهیم و روشهای

آمارحیاتی وشاخص های بهداشتی

دکتر رضاعلی محمدپور تهمتن
عضو هیات علمی دانشگاه علوم پزشکی مازندران

فهرست مطالب **صفحه**

پیش گفتار .....

فصل اول

آشنائی با مفاهیم اساسی .....

تعریف آمار .....

متغیر ها و سنجش آنها .....

انواع متغیرها .....

مراحل یک تحقیق علمی .....

نقش آمار در پژوهش های علمی و پزشکی .....

نرم افزارهای کامپیوتری آماری .....

فصل دوم

جمع آوری داده ها و نمونه گیری .....

روش های جمع آوری داده ها .....

نمونه گیری .....

انواع نمونه گیری .....

فصل سوم

طبقه بندی و نمایش داده ها .....

توصیف داده ها با جدول .....

نمایش داده ها با نمودار .....

فصل چهارم

توصیف عددی داده ها .....

شاخص های مرکزی .....

شاخص های پراکندگی .....

تغییرات یکنواخت در داده ها .....

فصل پنجم

احتمالات وتوزیع های احتمالات .....

توزیع دوجمله ای .....
توزیع پوآسون .....
توزیع نرمال .....
فصل ششم
برآورد .....
نمونه گیری، آماره و توزیع نمونه ای .....
استنباط در باره جامعه .....
فاصله اطمینان میانگین .....
فاصله اطمینان نسبت .....
برآورد حجم نمونه .....

فصل هفتم
آزمون فرضیه .....
خطای نوع اول و دوم .....
انواع آزمون ها .....
آزمون های t .....
آزمون کای دو .....

فصل هشتم
تحلیل همبستگی و رگرسیون .....
همبستگی .....

رگرسیون .....

فصل نهم
تجزیه و تحلیل واریانس .....

فصل دهم
شاخص های بهداشتی و جمعیتی

شاخص های بهداشتی .....

انواع شاخص های بهداشتی .....

میزان ها، نسبت ها و اعداد مطلق .....

شاخص های مربوط به باروری .....

شاخص های مربوط به مرگ و میر .....

شاخص های مربوط به بیماری ها .....

شاخص های مربوط به سنجش وضعیت خدمات رسانی بهداشت و درمان .....

شاخص های اقتصادی، اجتماعی، فرهنگی مرتبط با بهداشت .....

شاخص های جمعیتی .....

فهرست منابع .....

ضمیمه .....

پیش گفتار

استفاده از مفاهیم وروشهای آماری در سالهای اخیر، افزایش چشمگیری داشته است. بسیاری ازتحقیقات مربوط به علوم پزشکی در شاخه های پزشکی، دندانپزشکی، داروسازی، بهداشت وپرستاری و مامائی ورشته های پیرا پزشکی نیز از اصول و روشهای آماری استفاده می کنند. علیرغم وجود منابع مختلف در زبان انگلیسی برای آموزش آمار به دانشجویان و علاقه مندان، در سالهای اخیر کتابهای فارسی نیز تالیف شده است. بعضی از این منابع بخشی از نیاز های آموزشی را تامین می کنند.ولی تنوع رشته های تحصیلی و موضوعات تحقیقاتی باعث شده است تا همچنان تالیف و ترجمه کتابهای آماری به زبان ساده وعمومی لازم باشد. تا در آموزش دانشجویان دوره های کاردانی وکارشناسی رشته های مختلف، مفاهیم پایه و اصول و روشهای اساسی علم آمار را آموزش دهد و برای دانشجویان کارشناسی ارشد، اساتید، پزشکان و محققین محترم که برای انجام تحقیقات مختلف، نیاز به فهم مبانی و اصول آماری دارند نیز مورد استفاده قرار گیرد. این کتاب برای دو واحد درس آمار حیاتی و یک واحد درس شاخصهای بهداشتی در دانشگاه های علوم پزشکی می تواند مورد استفاده قرار گیرد. با تغییر مثالها و استفاده از تمرینهای موضوعی در در سهای آمار عمومی و کاربردی در سایر رشته های دانشگاهی، مورد استفاده قرار گیرد. در آموزشهای کارگاهی و دوره های باز آموزی در زمینه روش تحقیق نیز می تواند مورد استفاده قرار گیرد.

بر اساس اجزاء و مراحل علم آمار، در فصل اول کتاب، مفاهیم کلی، ودر فصل دوم تا چهارم، روش های جمع آوری، طبقه بندی، نمایش و توصیف داده ها توضیح داده شد. مفاهیم احتمال وتوزیع های دوجمله ای و پوآسن و نرمال در فصل پنجم ارائه گردید. در فصل ششم وهفتم از مباحث آمار تحلیلی به موضوعات برآورد و آزمون فرضیه اشاره شد. در فصل هشتم و نهم از مباحث آمار پیشرفته به موضوعات تحلیل همبستگی و رگرسیون و آنالیز واریانس به صورت مختصر اشاره شد. در فصل آخر شاخص های بهداشتی توضیح داده شد.

تجربیات چندین ساله مولف در آموزش آمار و انجام مشاوره های تحقیقاتی در دانشگاه ها باعث شده است که کتاب حاضر را تالیف نمایم که در این راه از منابع مختلف انگلیسی و فارسی استفاده کردم که از مولفین آنها تشکر می کنم ولی باید اذعان نمایم که شالوده مطالب بیشتر کتابهای منتشر شده در زبان فارسی تا کنون دو کتاب روشهای آماری وشاخصهای بهداشتی است که یکی، نوشته اساتید محترم دکتر کاظم محمد ، دکتر ملک افضلی ونهابتیان است که ازبنیانی محکم و جامعیت برخوردار است که تغییر یا افزایش مطالب آن به سادگی امکان پذیر نیست، و دیگری

نوشته واین و دانیل که ترجمه دکتر سید محمد تقی آیت اللهی است . از همه این اساتید سپاسگزارم.

روشهای آمار تحلیلی وپیشرفته و اختصاصی تر برای کاربرد های پزشکی و بهداشت انشاءالله در جلد دیگری ارائه خواهد شد. راهنمائی خوانندگان محترم در باره هر گونه اشکال وابهام احتمالی مطالب کتاب موجب امتنان خواهد بود.

دکتر رضاعلی محمدپورتهمتن
دانشگاه علوم پزشکی مازندران
۱۳۸۶

فصل اول :
آشنایی با مفاهیم اساسی

هدف های فصل اول: با مطالعه این فصل شما باید بتوانید:
۱- علم آمار را تعریف کنید و هدف ها و مراحل آنرا برشمارید.
۲- متغیر ها و انواع آنها را توضیح دهید.
۳- با جایگاه آمار در پژوهش های بهداشتی آشنا شوید.
۴- نرم افزارهای کامپیوتری که در انجام کارهای آماری مفید هستند را نام ببرید.

۱-۱- مقدمه:

آمار به عنوان یک لغت در زبان فارسی به معنی حساب، شمارش یا احصائیه است که معادل کلمه *statistics* در زبان انگلیسی به کار رفته است. Statistics از کلمه لاتین *status* مشتق شده است که به معنی تعیین وضعیت است.
آمار در اصطلاحات روزمره وعادی به معنی *اعداد وارقام* به کار می رود. آمار از نظر تاریخی به عنوان یک علم زمینه های متفاوتی دارد. از یک طرف در گذشته به دلیل کاربرد های زیاد سرشماری وجمع اوری داده های جمعیتی، اقتصادی واجتماعی جهت تامین نیاز های حکومتی، علم آمار به عنوان زیر شاخه ای از علوم انسانی واجتماعی شکل گرفت واز طرفی با توسعه علم احتمالات در ریاضیات وگسترش فرمول های آن، شاخه آمار ریاضی توسعه پیدا کرد. ودر سال های اخیر با به کارگیری این علم در جهت گسترش دامنه بسیاری از علوم دیگر آمار کاربردی رشد چشمگیری داشته است.

۱-۲- تعریف آمار :

برای آمار تعاریف متعددی بکار رفته است که در اینجا به یکی از آنها که هدف ها واجزا وروش های این علم را بهتر بیان می کند اشاره می کنیم.
آمار علمی است که در باره ی جمع آوری، طبقه بندی، خلاصه کردن و نمایش داده ها، جهت تجزیه و تحلیل آنها به منظور قضاوت وارزشیابی بحث می کند. انتقال اطلاعات و نتایج حاصل به منظور استنباط از وضع موجود و تصمیم گیری یا برنامه ریزی بر مبنای آن به کار برده می شود. تصمیم در مورد مجموعه بزرگی از داده ها با توجه به بررسی بخشی از آن بر عهده آمار است. قسمتی از آمار که به جمع آوری ، طبقه بندی و نمایش داده ها مربوط می شود *آمار توصیفی* و

قسمتی که در مورد استنتاج و تحلیل اطلاعات می پردازد تحت عنوان *استنباط آماری* یا *آمار استنباطی* معروف می باشد.

آمار حیاتی[1] یا آمار زیستی :

علم آمار کاربردهای فراوانی در رشته های گوناگون علوم مانند ،بازرگانی ،آموزش و پرورش روانشناسی ، کشاورزی ، اقتصاد ، مهندسی ، پزشکی ، بهداشت و ... دارد.

کاربرد آمار در علوم زیستی ، پزشکی و بهداشت منجر به توسعه شاخه ای از علم آمار گردیده است که به *آمار حیاتی* یا *آمار زیستی* معروف گشته است. گاهی اوقات اعداد وارقامی که نشان دهنده تعداد وقایع مهم زندگی مانند تولد، ازدواج، طلاق ومرگ افراد در جامعه می باشد تحت عنوان آمار وقایع حیاتی یا به اختصار آمار حیاتی گفته می شود که نباید با آمار حیاتی یا آمار زیستی به عنوان یک شاخه علمی یکسان تلقی گردد، اگرچه می توان گفت که این شاخه علمی شامل بحث های مربوط به این وقایع نیز است که در بخش شاخص های بهداشتی به بررسی آنها می پردازیم.

جامعه آماری، واحد و نمونه آماری :

در اصطلاحات آماری به مجموعه ای از افراد یا اشیاء یا هر پدیده دیگری که حداقل در یک صفت ثابت مشترک باشند *جامعه آماری* گفته می شود جامعه آماری ممکن است محدود یا نا محدود باشد.

هر عضو جامعه آماری به عنوان یک *واحد آماری* و هر تعدادی از آنها را که مورد بررسی قرار گرفته *نمونه آماری* از جامعه مورد نظر به حساب می آیند.

۳-۱- متغیرها:

غالباً برای بررسی جامعه آماری مطالعه صفات دیگری که از یک فرد به فرد دیگر قابلیت تغییر دارد مورد توجه قرار می گیرد. به اینگونه صفت یا صفات، متغیر[2] گفته می شود.

به عنوان مثال اگر جامعه آماری بیماران مراجعه کننده در یک روز مشخص به درمانگاه معین باشند صفت ثابت که باعث تشکیل این جامعه آماری می گردد همان مراجعه افراد بیمار می باشد ولی برای بررسی های گوناگون ممکن است سن بیماران مورد پرسش قرار گیرد.

سن متغیری است که میزان آن از یک فرد به فرد دیگر تغییر می کند.

جنس مراجعین نیز متغیر دیگری است که قابلیت تغییر از یک فرد به فرد دیگر برای آن وجود دارد. متغیرهای زیادی را می توان نام برد. قد دانش آموزان یک کلاس ، وزن نوزادان در هنگام تولد

---

[1] - Biostatistics

[2] - variable

، سن مادران در هنگام اولین زایمان ، تعداد فرزندان خانوارهای یک روستا ، شغل رئیس خانوار و ... متغیرهای دیگر هستند.

۴-۱- نتایج مشاهدات یا مقیاس سنجش متغیرها:

نتیجه مشاهده و سنجش صفات متغیر معمولاً به یکی از اشکال زیر است که ممکن است *انواع مقیاس های اندازه گیری* نیز گفته شود:

۱- اسمی

۲- رتبه ای

۳- فاصله ای

۴- نسبتی

در حالتی که بتوان بر اساس نتیجه مشاهدات تنها در مورد انتساب فرد مورد مطالعه به گروه یا طبقه خاصی اظهار نظر کرد و نام آن طبقه یا گروه را مشخص کرد صفت اسمی خوانده می شود.

مثلاً در مورد جنس می توانیم گروه مرد و زن را در نظر گرفته که هر فرد می تواند در یکی از گروهها قرار گیرد.

یا در مطالعه گروه خون افراد، هر فرد با داشتن گروه خون خاص، در یک گروه معین قرار می گیرد. متغیرهائی مانند جنس و گروه خون به صورت اسمی در گروههای مختلف سنجیده می شوند . اگر افراد جامعه را به اساس اینکه بیمار هستند یا نیستند طبقه بندی نمائیم به صورت اسمی نتیجه مشاهدات را گروه بندی نموده ایم.

علاوه بر گروه بندی افراد بر اساس یک ویژگی ممکن است در بعضی از مشاهدات نتیجه سنجش یک خاصیت را با بیان رتبه فرد یا شیء در رابطه با سایر افراد بیان کرد مثلاً برای سنجش یک بیماری ، افراد را بر حسب شدت یا ضعف بیماری در گروههای مختلف قرار داد . بدیهی است فردی که در گروه شدید قرار دارد از صفت مورد نظر نسبت به فردی که در گروه متوسط قرار دارد بیشتر برخوردار بوده است و همینطور فردی که در گروه متوسط قرار دارد نسبت به گروه خفیف از ویژگی مورد بحث، بیشتر برخوردار است.

نکته ای که در این مقیاس وجود دارد این است که نمی توان گفت که شدت بیماری در گروه متوسط دو برابر خفیف است. یا اینکه متوسط به همان اندازه از گروه خفیف بیشتر بیمارند که گروه شدید نسبت به گروه متوسط بیماری شدید تری دارند.

در این مقیاس نمی توان فاصله دو رتبه متمایز و یا نسبت اندازه بین آنها را مشخص کرد ولی خاصیت صفت اسمی در اینجا نیز وجود دارد. اینگونه صفات را *رتبه ای* می گویند در نوع دیگر مشاهدات با قرار دادن یک مبدأ نتیجه سنجش را بر حسب واحدی ثابت و معینی اندازه گیری نمود

و در نتیجه فاصله دو شیئی یا دو فرد را از نظر صفت مورد بررسی معین کرد ولی بـا توجـه باینکـه این نوع اندازه گیری فاقد صفر ذاتی می باشد نمی توان نسبت انـدازه خاصـیت مـورد مطالعـه را در افراد مشخص کرد به همین دلیل به این مقیاس *فاصله ای* گفته می شود.

به دلیل استفاده از واحد معین می توان فاصله دو اندازه را دقیقاً معلوم نمود ولی چون صفر در ایـن مقیاس ،صفر حقیقی نبوده و به معنی عدم ویژگی مورد نظر نمی باشد بلکه حالـت قـرار دادی دارد نمی توان نسبت دو اندازه مختلف را با هم تعیین نمود. مثلاً اگر درجـه حـرارت جسـمی ۴۰ درجـه سانتیگراد و دیگری ۲۰ درجه سانتیگراد باشد نمی توان گفت که جسم اول دو برابـر دیگـری گرمـا دارد چون در گذشته تصور می شد که وقتی جسمی همه گرمای خود را از دست داده اسـت دمـای آن صفر خواهد بود مثلا وقتی آب یخ زده است هیچ گرمایی ندارد و دمای آن را صفر در نظر گرفته اند ولی با ارائه تئوری حرکت مولکولی در فیزیک مشخص شده است که گرما همان حرکت مولکول ها است بنا بر این در دمای صفر درجه هم مولکول ها حرکت دارند یعنی صفر در اندازه گیـری دمـا به معنی عدم وجود گرما یا صفر واقعی نیست. این نوع مقیـاس نیـز از خاصـیت اسـمی و رتبـه ای برخوردار است.

نوع دیگر مقیاسها مقیاس نسبتی است که نه تنها می توان فرد مورد مطالعه را به گروهـی منتسـب کرد یا رتبه آنرا تعیین نمود یا فاصله آنرا با افراد دیگر سنجید بلکه می توان نسبت انـدازه خاصـیت مورد مطالعه را در دو فرد مشخص کرد . مثلاً در اندازه گیری طول اشیاء صفر به معنی نبودن طـول بوده لذا اگر جسمی ۴۰ سانتی متر طول و دیگری ۲۰ سانتی متـر طـول داشـته باشـد مـی تـوانیم بگوئیم که طول اولی دو برابر دومی است.

۵-۱- انواع متغیرها

با توجه به نتیجه سنجش متغیرها می توان متغیرها را به صورتهای *کیفی* و *کمی* تقسیم بندی کرد. متغیرهائی که نتیجه آنها به صورت اسمی ورتبه ای باشد معمولاً تحـت عنـوان کیفـی و متغیرهـای فاصله ای و نسبتی را تحت عنوان کمی تقسیم بندی می کنند . متغیرهای کمی به صورتهای *گسسته* و *پیوسته* تقسیم می گردد. به عبارت دیگر متغیری که بین هر دو مقدار آن نمی توان مقدار دیگری را برای آن در نظر گرفت. مثلا یک خانوار ۲ فرزندی و دیگری ۳ فرزنـدی باشـد نمـی تـوان خـانوار دیگری تصور کرد که بین ۲ تا ۳ فرزند مثلا ۲/۳ فرزند داشته باشد.

متغیر کمی گسسته (منفصل) :

متغیرهائی مانند تعداد افراد خانوار ، تعداد بیماران بستری شده در یـک بخـش در زمـان معـین یـا میزان مرگ و میر و ... که مقادیر آن مجموعه شمارش پذیر اعداد یا زیر مجموعه ای از آن را اختیار می کند متغیر *کمی گسسته* نامیده می شوند. به عبارت دیگر، متغیری کـه بـین هـر دو مقـدار آن

نمی توان مقدار دیگری را برای آن در نظر گرفت. به عنوان مثال، اگر یک خانوار ۲ فرزندی و خانوار دیگر ۳ فرزندی باشد نمی توان خانوار دیگری تصور کرد که بین ۲ و ۳ مثلا ۲/۳ فرزند داشته باشد.

متغیر کمی پیوسته (متصل):

متغیرهائی مانند وزن ،طول قد که بین هر دو مقدار آن می توان مقدار دیگری برای آن تصور کرد متغیر *پیوسته* نامیده می شوند . در این نوع متغیرها ممکن است نتایج مشاهدات بصورت اعدادی با تعداد اعشار دلخواه اندازه گیری شده باشند . بدیهی است این تقریب به میزان دقت وسیله اندازه گیری در عمل وابسته است.

متغیر ساده یا مرکب:

متغیرها از نظر دیگر به متغیر ساده یا مرکب تقسیم می شود.

متغیر مرکب : متغیری است که نتیجه سنجش آن با یک عدد و یا خصوصیات نشان داده می شود اما لازمه اش بررسی و اندازه گیری چندین *متغیر ساده* است .

مثال ۱- در بررسی میزانDMF مردم شهر تهران، برای تعیین اندازه این متغیر باید تعداد دندانهای پوسیده ، افتاده و پر شده مورد بررسی قرار گیرد .

مثال ۲- وضعیت اقتصادی خانوار یک متغیر مرکب است که برای سنجش آن می تواند متغیرهای ساده ای مثل وضع درآمد ، بعد خانوار ، محل سکونت ، نحوه تصرف واحد مسکونی و ...مورد سنجش قرار گیرد.

انواع متغیرها از نظر نقش وارتباط آنها با یکدیگر:

دربررسی ارتباط متغیرها با یکدیگر معمولًا آنها را به چهار گروه تقسیم می کنند که به همراه مثال توضیح داده می شود.

۱- متغیر مستقل (Independent) :

متغیری که به طور مستقل تغییر می کند ولی تغییرات آن باعث تغییرات متغیر دیگر که به آن وابسته است می شود. معمولا در تحقیقاتی که هدف آنها تعیین ارتباط بین متغیرها است یا اثر متغیری را بر دیگری می خواهیم بررسی کنیم، رفتار متغیر مستقل مورد مطالعه قرار می گیرد تا اثرات آن در تغییرمتغیر وابسته که غالبا همان بروز مشکل است، مشخص شود.

مثال ۱- در بررسی رابطه بین مصرف سیگار با سرطان ریه، وضعیت سیگار (کشیدن یا نکشیدن) متغیر مستقل است.

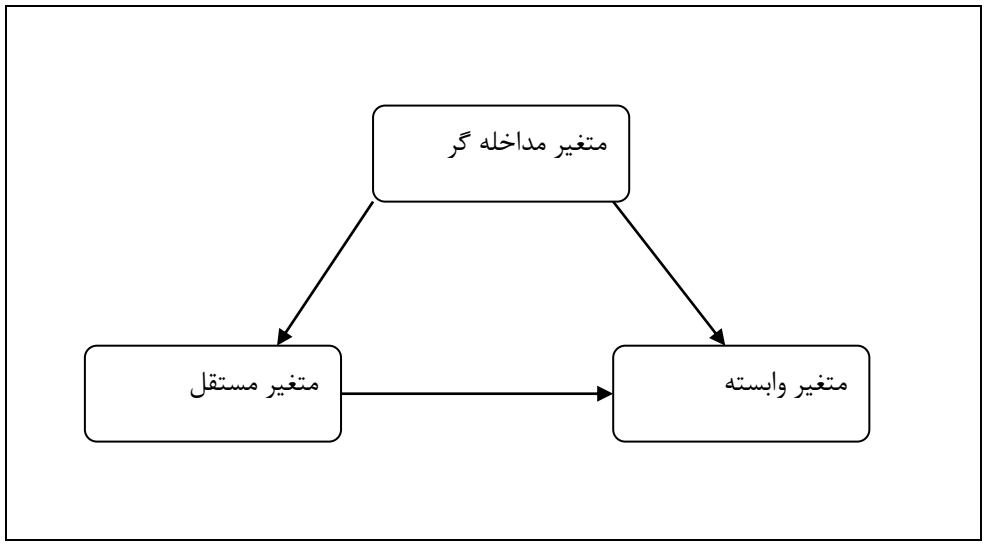
مثال ۲- در بررسی تاثیر نوع دارو در درمان هپاتیت، نوع دارو متغیر مستقل است.

۲- متغیر وابسته (Dependent) یا غیر مستقل :

متغیری که در نتیجه ی تغییر یک یا چند متغیر مستقل تغییر می کند *وابسته* نام دارد. متغیر وابسته همان مشکل یا پیامد است که ممکن است در مدل های آماری برای تعیین رابطه متغیرها، مستقل به عنوان *ورودی* ومتغیر وابسته به عنوان *خروجی* در نظر گرفته شود.
در مثالهای بالا ، سرطان ریه و نتیجه درمانی هپاتیت متغیرهای وابسته هستند.

۳- متغیر مداخله گر یا مخدوش کننده: Confounding variable or Intervening variable

به متغیری است که روی متغیر مستقل و یا روی متغیر وابسته یا بر هر دوی آنها تاثیر می گذارد و باعث می شود که نتوانیم اثر متغیر مستقل را بر متغیر وابسته ارزیابی کنیم، متغیر *مداخله گر یا مخدوش کننده* می گویند.



مثال : در بررسی اثر آهن تکمیلی بر هموگلو بین خون، آیا فقط آهن تکمیلی است که بر وضعیت هموگلوبین تاثیر می گذارد . مثلا عادت غذایی ، بیماری انگلی ، بیماری مزمن ، سن ، جنس ، میزان استرس و وضعیت متابوبیسم فرد نیز بر روی وضعیت هموگلوبین اثر دارد .

هدف بسیاری از تحقیقات علمی شناسایی متغیرهای مداخله گر وکنترل اثرات آنها بر ارتباط متغیر مستقل و وابسته است. لذا محققین باید این اثرات را حذف یا مشخص نمایند تا میزان اثر باقی مانده بر متغیر وابسته که ناشی از متغیرهای مستقل است بصورت شفاف معلوم شود. برای این کار می توانند چهار رفتار متفاوت با متغیرهای مداخله گر داشته باشند.

الف – نادیده گرفتن : این عمل موجب به بیراهه کشیده شدن تحقیق می شود زیرا نتایج تحقیق فقط به تغییر متغیر ارتباط نداشته بلکه به عوامل دیگری بستگی داشته است این نوع تحقیق اگر انجام نشود بهتر است.

ب – از مطالعه خارج کردن : در این مطالعه متغیرهای مداخله گر را شناسایی و حذف می کنیم. در این مطالعه به دو مشکل بر می خوریم . یکی اینکه پیدا کردن نمونه مشکل است و دیگر اینکه مطالعه از نظر اعتبار خارجی و تعمیم پذیری دچار مشکل می شود.

ج – بررسی و مشابه سازی : این روش و اصیل ترین روش تحقیق است . در این جا متغیر مداخله گر وارد مطالعه شده و قدرت فعالیت آنها را تحت کنترل و تنظیم در می آوریم . مشکل عمده آن این است که این روش خیلی عملی نیست . از محدودیت های دیگر آن تعدد متغیرهای مداخله گر و قابل پیش بینی نبودن اثرات آن می باشد .

د – بررسی بدون مشابه سازی : به علت تعدد متغیرهای مداخله گر و قابل پیش بینی نبودن اثرات آن این روش، در صورتی که روش ج امکان پذیر نباشد انجام می گیرد.

بهترین رفتار با متغیر مداخله گر به ترتیب بررسی مشابه سازی ، بررسی بدون مشابه سازی، حذف متغیر و نادیده گرفتن است.

در عمل و در واقعیت محقق ترکیبی از این چهار رفتار را دارد . محقق بعضی را نادیده گرفته و بعضی را حذف و بعضی را مشابه سازی و بعضی را بررسی بدون مشابه سازی می کند . یکی از دلایلی که از نظر اعتبار هیچ تحقیقی صد در صد ایده ال نیست همین رفتار متفاوت با متغیرهای مداخله گر که میزان اعتبار تحقیق را متاثر می سازد.

۴- متغیر زمینه ای (background variable) :

متغیرهایی هستند که جزء خصوصیات نمونه هستند و معمولا در همه تحقیقات وجود داشته و اغلب به طور متناسب مورد بررسی قرار می گیرند تا خصوصیات گروهها و یا جوامع روشن شوند . بر اساس پیشینه این متغیر اثر روی نتیجه تحقیق ندارد اما محقق احتمال اثر آن را بر روی نتایج تحقیق می دهد و هم اینکه محقق بابررسی این متغیرها وضعیت نمونه ها را بهتر مشخص می کند .

مثال : سن ، جنس ، وضعیت تاهل ، خصوصیات مکانی

متغیرهای زمینه ای در هر زمان ممکن است متفاوت باشد مثال : بررسی اثر انسولین برقند خون متغیرهای زمینه ای در گذشته وزن ، سن ، جنس ، بود و هم اکنون در پیشینه آمده است که جنس و سن از متغیرهای مداخله گر است بنابراین دیگر زمینه ای محسوب می شود .

۶-۱- نقش آمار در پژوهشهای علمی :

اگر موضوع علم آمار را جمع آوری ، طبقه بندی و نمایش ، تجزیه و تحلیل اطلاعات بدانیم واضح است که آمار در قلمرو تمام تحقیقات علمی به کار گرفته می شود . به عبارت دیگر در مرحله جمع آوری اطلاعات شیوه های مختلف آمار گیری که ممکن است همه افراد جامعه آماری را در بر بگیرد (سر شماری) یا بخشی از افراد جامعه را شامل شود (نمونه گیری) به محقق کمک می کند که به بهترین روش اطلاعات را گرد آوری نماید و برای تائید یا رد مدل مفروض یا اصلاح نظریه موجود و حتی طرح ریزی تحقیقات دیگر روشهای مربوط به آمار استنباطی را به کار ببرد.

در تحقیقات مربوط به پزشکی و بهداشت نیز آمار یکی از ابزارهای مهم به شمار می آید که در ارتباط نزدیک با شاخه های دیگر علمی مانند اپیدمیولوژی ، پزشکی اجتماعی و سایر علوم بهداشتی قرار دارد و لذا مطالعه آمار به منظور هدایت تحقیقات و کمک به گسترش معلومات تحقیقاتی در این بخشها، ضروری خواهد بود.

مراحل یک تحقیق علمی :

در پژوهشهای علمی، مراحل زیر که گاهی به روش علمی تحقیق موسوم است موردتوجه قرار می گیرد.

۱- بیان مسئله و مشخص کردن هدف مورد بررسی یکی از مراحل مهم یک تحقیق به شمار می رود .

۲- طراحی مطالعه و انتخاب روش و استراتژی خاصی که منجر به دستیابی هدف تحقیق گردد.

۳- جمع آوری داده ها و سازمان دهی آنها :

در هر تحقیقی تهیه اطلاعات واقعی باتوجه به هدف مورد نظر اهمیت اساسی دارد . داده های آماری جمع آوری و با استفاده از روشهای خاصی سازماندهی و خلاصه می گردند.

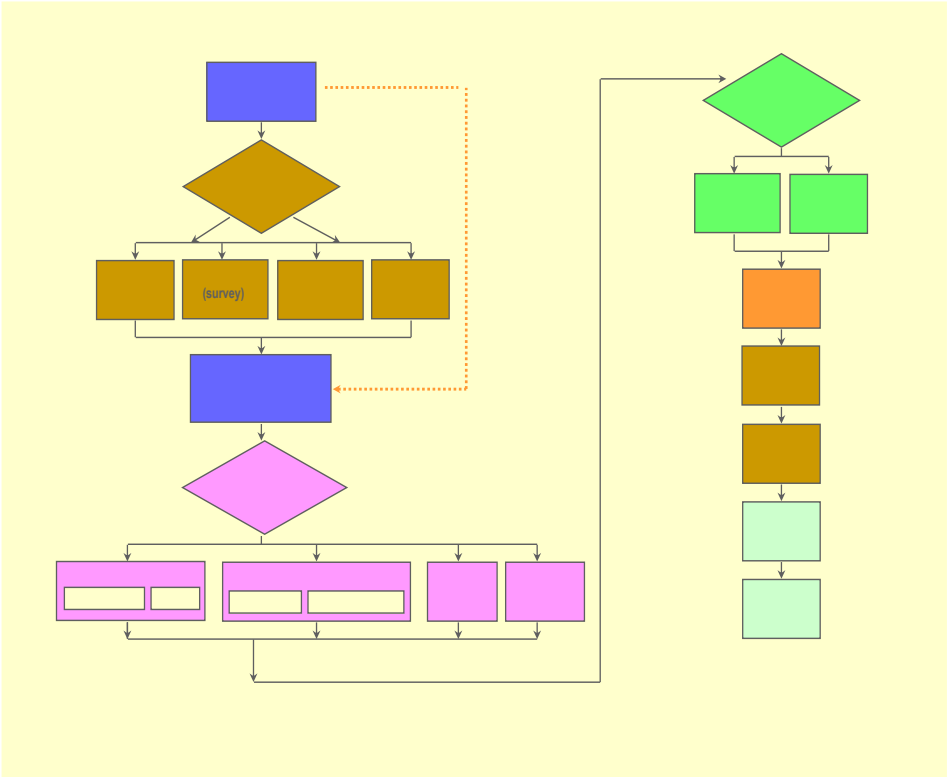
۴- تجزیه و تحلیل داده ها :

اطلاعاتی که بوسیله روشهای خاصی به دست آمده اند به عنوان منبع اساسی برای کسب معلومات جدید درباره پدیده مورد مطالعه بکار می روند و برای ارائه یا تأئید فرضیه ای مورد تجزیه و تحلیل قرار می گیرند .

۵- بیان یافته ها و گزارش :

اثبات یا رد فرضیه ها و پاسخگوئی به سئوالاتی که در بیان مسئله یا هدف مطالعه مطرح بوده است در این مرحله صورت می گرد و نتایج حاصل به صورت یک گزارش تنظیم و پیشنهادات لازم ارائه می گردد.

این مراحل در نمودار کلی زیر مشخص شده است.



نرم افزارهای کامپیوتری آماری

استفاده از نرم افزارهای کامپیوتری باعث شده است که محاسبات پیچیده مربوط به روشهای آماری به سهولت و در زمان کمتری انجام گیرد. بنابر این، توجه محققین رشته های مختلف، به نکات تشخیصی در انتخاب روشها و مدلها، مهمتر از انجام محاسبات و حفظ فرمولهای آنها است. لذا

نکات تشخیصی در انتخاب روشها و تفسیر نتایج و گزارش آنها در نگارش مقالات پژوهشی از نکات مهمی است که استقبال بیشتری را باعث شده است.

مهمترین نرم افزارهای آماری عبارتند از: **SPSS، SAS، STATA و Minitab** که کاربرد های وسیعی دارند.**SPSS (Statistical Package for the Social Science)** نرم افزار آماری است که برای کاربرد روش های آمار در علوم اجتماعی تهیه شد ولی در سایر زمینه ها در سطح وسیع مورد استفاده قرار گرفت. **Epi-Info** نیز نرم افزاری با کاربرد هایی در مطالعات اپیدمیولوژیکی و آمار پزشکی است.

نرم افزار داده پردازی و آماری **Microsoft Excel** هم برای وارد کردن داده ها و هم برای تحلیل های اولیه آماری مورد استفاده قرار می گیرد. از معروفترین نرم افزارهای گرافیکی برای رسم نمودارهای آماری **HG و Stat. Graph یا SG** است.

در بخشهای مختلف این کتاب، مواردی از خروجی های نرم افزار **SPSS** پس از اجرای روشهای آماری ارائه شده است.

پیشنهاد کار تحقیقاتی

به منظور آشنایی با پرسشنامه پیشنهاد طرح تحقیقاتی، یک نسخه از آن در اینجا ارائه می گردد. شما می توانید آن را از واحد امور پژوهشی دانشکده تهیه نمایید و قسمتهای مهم آن را که به آمار مربوط می شود در کلاس مورد بررسی قرار دهید.

جمهوری اسلامی ایران

دانشگاه علوم پزشکی و خدمات بهداشتی درمانی .....

معاونت پژوهشی

پرسشنامه طرح تحقیقاتی

قسمت اول- اطلاعات مربوط به مجری طرح و همکاران

۱-۱- نام و نام خانوادگی مجری طرح :

۱-۲- نشانی : تلفن :

محل کار : تلفن :

۱-۳- شغل و سمت فعلی مجری طرح :

۱-۴- سازمان متبوع :

۱-۵- محل اجرای طرح :

۱-۶- درجات علمی و سوابق تحصیلی طرح دهنده (به ترتیب از لیسانس به بالا)

درجه تحصیلی	رشته تحصیلی و تخصصی	دانشگاه	کشور	سال دریافت

۱-۷- مشخصات همکاران:

ردیف	نام و نام خانوادگی	شغل	درجه علمی	نوع همکاری	امضای همکار

۱-۸- فهرست آخرین طرح‌های تحقیقاتی مصوب و مقالات منتشر شده مجری طرح (حداکثر ۳ طرح پژوهشی و ۳ مقاله)

قسمت دوم- اطلاعات مربوط به طرح پژوهشی

۲-۱) عنوان طرح:

فارسی :

انگلیسی :

۲-۲- نوع طرح : بنیادی ☐ کاربردی ☐

۲-۳- معرفی طرح و بیان مسأله :

۲-۴- بررسی متون (Literature review) :

۲-۵- اهداف و فرضیات (OBJECTIVES & HYPOTHESIS)

هدف (اهداف) کلی طرح (General objectives) :

هدف (اهداف) اختصاصی طرح (Specific Objectives) :

فرضیات (یا سوالات تحقیق) :

قسمت سوم- اطلاعات مربوط به روش اجراء طرح و انتخاب نمونه ...

۳-۱- طراحی تحقیق و روش اجرا (Design and Methodology) :

(لطفاً اطلاعات زیر بطور دقیق نوشته شود):

(روش تحقیق و توجیه آن، خصوصیات دقیق نمونه (جامعه آماری, تعداد نمونه, معیارهای ورود و خروج ...)و تجزیه و تحلیل آماری، تعریف گروه شاهد، مکان اجرای تحقیق، معرفی پرسشنامه یا فرم اطلاعاتی* و روایی و پایایی آن، ابزار با ذکر جزئیات، کیت‌ها و مواد مصرفی کلیدی با ذکر مشخصات، داروها و دارونماها, شرح مداخله و روش انجام کار و ...)

۲-۳- جدول متغیرها

نام متغیر	نقش متغیر		مقیاس				واحد		ابزار سنجش	نوع رفتار با متغیر مداخله‌گر			
	مستقل	وابسته	زمینه‌ای	مداخله‌گر	کمی	کیفی				نادیده گرفتن	ثبت کردن	حذف کردن	مشابه سازی کردن

۳-۳- تعریف واژه‌ها (تعریف عملیاتی) :

۴-۳- ملاحظات اخلاقی (Ethical Consideration) :

(حتی‌الامکان بطور مستند و با ذکر منابع بی‌خطر بودن تحقیق برای موارد انسانی توجیه شود)

۵-۳- جدول زمانی مراحل اجراء طرح پس از تصویب (Guntt Chart) :

ردیف	شرح هر یک از فعالیتهای اجرائی طرح	سال اجرا	زمان اجرا									
			ماه ۱	ماه ۲	ماه ۳	ماه ۴	ما ه ۵	ما ۶ه	ماه ۷	ماه ۸	ما ه ۹	

فصل چهارم- اطلاعات مربوط به هزینه ها

* در صورتی که برای گردآوری داده‌ها از پرسشنامه یا فرم اطلاعاتی استفاده می‌کنید، یک نسخه از آن ضمیمه شود.

۱-۴- آیا برای این طرح از سازمانهای دیگر نیز درخواست اعتبار شده است؟ بلی ☐ خیر ☐
در صورت مثبت بودن جواب لطفاً نام سازمان و نتیجه حاصل را ذکر فرمائید.

۲-۴- هزینه پرسنلی با ذکر مشخصات کامل و میزان اشتغال هر یک و حق الزحمه آنها :

ردیف	نوع فعالیت	انجام دهنده*	کل ساعات کار برای طرح	حق الزحمه در ساعت	جمع (ریال)
جمع هزینه های تخمینی					

* در صورتی که تعداد افراد انجام دهندهٔ هر فعالیت بیش از یک نفر هستند در این قسمت اسامی افراد نوشته شود:

۳-۴- هزینه آزمایشات و خدمات تخصصی که توسط دیگر مؤسسات صورت می گیرد :

موضوع آزمایش یا خدمات تخصصی	مرکز سرویس دهنده	تعداد کل دفعات	هزینه برای هر دفعه	جمع (ریال)
جمع هزینه های آزمایشات و خدمات تخصصی				

۳-۴- فهرست وسایل و موادی که باید از اعتبار این طرح از داخل یا خارج کشور خریداری شود :

نام دستگاه یا مواد	شرکت سازنده و یا فروشنده	کشور	مصرفی یا غیرمصرفی	آیا در ایران موجود است؟	تعداد لازم	قیمت واحد - کل	تایید کارشناس طرح
جمع هزینه های وسایل و مواد							

۴-۴- هزینه های دیگر :

الف- هزینه های تکثیر اوراق پرسشنامه ها : .. .. ریال

ب- هزینه های تکثیر نشریات و کتب مورد لزوم : ..... ریال
ج- هزینه تایپ و صحافی (گزارش پایانی) : ..... ریال
ج- سایر هزینه ها (با ذکر مورد) : ..... ریال

۴-۷- جمع هزینه های طرح :

الف- جمع هزینه های پرسنلی	ریال
ب- جمع هزینه های آزمایشات و خدمات تخصصی	ریال
پ- جمع هزینه های وسایل	ریال
ت- جمع هزینه های مسافرت	ریال
ث- جمع هزینه های دیگر	ریال
ج- جمع کل	ریال

۴-۸- مبلغی که از منابع دیگر کمک خواهد شد و نحوه مصرف آن :
۴-۹- باقیمانده هزینه‌های طرح که تأمین آن درخواست میشود :

صحت مطالب مندرج در پرسشنامه را تأیید می نمائیم.
محل امضای مجری طرح : محل امضای معاون پژوهشی :
منابع و مأخذ (References) :

نمونه سؤالات فصل اول :

مفاهیم وتعاریف:

۱- مفاهیم زیر را تعریف کنید ؟
آمار
آمار زیستی
متغیر
نمونه

۲- کدامیک از صفهای زیر صفت متغیر کیفی است ؟
الف- تعداد خانه های بهداشتی شهرستانهای مازندران
ب- نوع بیماری

ج- مذهب

۳- کدام یک از مباحث زیر از اجزاء آمار استنباطی یا تحلیلی محسوب می شود؟

الف- محاسبه شاخص های مرکزی وپراکندگی ب- آزمون فرضیه

ج- جمع آوری داده ها د- طبقه بندی داده ها

۴- تفاوت قضاوتهای آماری و قضاوتهای ریاضی در کدام یک از موارد زیر توجیه می شود؟

الف- قضاوتهای ریاضی قطعی وقضاوتهای آماری احتمالی هستند.

ب- قضاوتهای ریاضی وقضاوتهای آماری تفاوتی ندارند.

ج- در قضاوتهای آماری استثناء کردن موارد باعث نقض حکم می شود ولی در ریاضی نمی شود.

د- قضاوتهای ریاضی احتمالی ولی قضاوتهای آماری قطعی هستند.

متغیرها ومقیاس آنها:

۵- کدام یک از متغیرهای زیر یک متغیر کمی گسسته است؟

الف- تعداد فرزندان خانوارهای یک روستا ب- وزن نوزاد

ج- قد کودکان د- میزان کلسترول

۶- نتیجه آزمایش پزشکی بر حسب مثبت یا منفی گزارش شده است. مقیاس به کار گرفته شده از کدام نوع است؟

الف- اسمی ب- ترتیبی ج- فاصله ای د- نسبتی

۷- متغیر وزن نوزادان در هنگام تولد با کدام مقیاس سنجیده شود مناسبتر است؟

الف- اسمی ب- رتبه ای ج- فاصله ای د- نسبتی

۸- کدام یک از مقیاس سنجش متغیرها کاملتر است؟

الف- اسمی ب- رتبه ای ج- فاصله ای د- نسبتی

۹- در بررسی عوامل موثر بر فشارخون هر یک از متغیرهای سن و شغل به ترتیب چه نقشی را دارند؟

الف- مستقل و وابسته ب- مستقل و مستقل

ج- وابسته ووابسته د- وابسته و مستقل

۱۰- برای متغیر های زیر کدام مقیاس سنجش مناسب تر است؟ جلوی آن بنویسید.

۱- وزن نوزاد بر حسب گرم ۲- جنس نوزاد

۳-گروه خونی ۴- سن

۵-میزان هموگلوبین ۶- طبقه اجتماعی اقتصادی

۷- درجه حرارت بدن ۸- فشار خون سیستولیک

۹- تعداد فرزند

۱۰- تعداد مراجعه مادران به مرکز در دوران بارداری

تمرین (کار عملی) :

۱- یک کارشناس بهداشتی می خواهد میزان شیوع سوء تغذیه را در کودکان زیر ۶ سال ساکن در مناطق روستایی یک شهرستان تحقیق کند. جدول متغیر های زیر را که بخشی از پرسشنامه پیشنهاد طرح تحقیقاتی دانشگاه های علوم پزشکی کشور است، برای این تحقیق تکمیل نمایید. میزان شیوع در فصل دهم تعریف شده است.

جدول متغیرها

نام متغیر	نقش متغیر					مقیاس		واحد	مقیاس سنجش	نوع رفتار با متغیر مداخله‌گر			
	وابسته	زمینه‌ای	مداخله‌گر	کمی	کیفی	کمی	کیفی			نادیده گرفتن	ثبت کردن	حذف کردن	مشابه سازی کردن

۲- برای آشنایی با قابلیتهای نرم افزار کامپیوتری Excel در زمینه روشهای آماری، با مراجعه به راهنمای آن، لیستی از قابلیتها را تهیه و در کلاس ارائه نمایید.

فصل دوم :
جمع آوری داده ها ونمونه گیری

هدف های فصل دوم: با مطالعه این فصل شما باید بتوانید:
۱- روش های جمع آوری داده ها را توضیح دهید.
۲- برای یک بررسی ساده آماری پرسشنامه تدوین کنید.
۳- هر یک از روشهای نمونه گیری را تعریف کنید.
۴- در تحقیقات بهداشتی در سطح کار آموزی روش نمونه گیری مناسب را انتخاب و به کار ببرید.

۱-۲- روشهای جمع آوری داده ها :

داده ها :

علاوه بر آشنائی با متغیرها و نحوه سنجش آنها لازم است که در هر مطالعه ای روشهای گرد آوری داده ها مورد توجه قرار گیرد. وقتی متغیری مورد مطالعه قرار می گیرد مقدار یا نتیجه سنجش آن برای هر فرد یک داده نامیده میشود که برای افراد نمونه یا جامعه همان مجموعه داده ها یاداده های خام را تشکیل می دهد . معمولاً هدفهای تحقیق تعیین چگونگی توزیع متغیرها یا ارتباط بین آنها ست . اگر داده های مختلفی از یک یا چند متغیر داشته باشیم ممکن است اطلاعات مورد نیاز در زمینه تعیین چگونگی توزیع یا ارتباط بین توزیعهای مختلف از دل داده ها بیرون کشیده شود که اصطلاحاً ممکن است گفته شود اطلاعات (Information) همان عصاره داده های خام (Data) است.

مثال : بهورزی در یک روستا مشاهده می کند که تعداد مادرانی که به دلیل اسهال فرزندان خود در دو هفته گذشته به خانه بهداشت مراجعه کردند از حد معمول بیشتر شده است . او سن و جنس کودک ،و نوع منبع آب آشامیدنی و نوع تغذیه را از مادران سئوال می کند . پاسخ این ها همان داده های خام هستند . اگر با طبقه بندی و تجزیه و تحلیل داده ها مشخص شود که بین منبع آب آشامیدنی و اسهال کودکان ارتباطی وجود دارد می گوییم این همان اطلاعاتی است که از دل این داده ها به دست آمده است ممکن است حجم داده ها بسیار زیاد باشد ولی اطلاعات مفید و مهم را در یک یا چند جمله می توان بیان کرد. برای جمع آوری داده ها بهورز می تواند روشها و ابزارهای مختلفی را مورد استفاده قرار دهد که بر حسب شرایط و نوع متغیرها می تواند متفاوت باشد . شرایط مختلفی می تواند گردآوری داده ها را تحت تأثیر قرار دهد . اگر این شرایط اثرات ناخواسته

و کنترل نشده ای بر سنجش متغیرها ایجاد کند به عنوان منابع خطا محسوب می شوند و صحت و اعتبار داده ها را مخدوش می کنند در چنین شرایطی نباید انتظار داشت که اطلاعات مفید و ناجی از دل داده های خام که تحت تأثیر منابع خطا قرار گرفته اند قابل استنتاج باشد. بنابر این اهمیت توجه به روشهای مختلف گردآوری داده ها در هر شاخه از تحقیقات علمی بسیار مهم است.

انتخاب روش مناسب برای جمع آوری داده ها، چه در سر شماری باشد یا در نمونه گیری، بستگی به نوع تحقیق، اهداف، امکانات دارد و متناسب با وضعیت جامعه یا نمونه و نوع مطالعه ممکن است به یکی یا ترکیبی از روشهای زیر صورت گیرد :

۱- مشاهده
۲- مصاحبه یا پر کردن پرسشنامه توسط مأموران آموزش دیده
۳- روش پرسشنامه کتبی
۴- آزمایش
۵- استفاده از اطلاعات سیستم ثبت (گزارشات قبلی)
۶- سایر روش ها مانند مکالمه تلفنی و پست الکترونیکی یا استفاده از اینترنت یا سیستم های ماهواره ای

الف- مشاهده:

مشاهده در اینجا به معنی درک حسی از پدیده ها و نتایج تغییرات آن است و شامل همه ادراکات حسی مانند دیدن ، شنیدن ، لمس کردن و ... است که مهمترین آن مشاهده با چشم است. مشاهده ، ثبت وضبط دقیق تمامی جوانب حادثه ، رفتار، گفتار و پدیده ها مبتنی بر الگویی است که از قبل تنظیم شده است .

مثلاً یک تکنسین آزمایشگاه با مشاهده عامل بیماری در زیر میکروسکوپ نتیجه یک تست را مثب گزارش می کند.

یک بهورز با اندازه گیری وزن کودک میزان آن را ۱۵ کیلو گرم اعلام می کند در حقیقت او برای ثبت داده وزن کودک از روش مشاهده استفاده کرده است.

مشاهده متاثر از حواس بوده ، پس می تواند منشا خطا باشد . برای اجتناب از خطا باید مشاهده را استاندارد کرد . مشاهده استاندارد شده با استفاده از ابزار معین را اندازه گیری گویند . مشاهده به عنوان مهمترین و متداولترین روش جمع آوری داده ها است که به صورت مختلف تقسیم بندی می شود:

۱- مشاهده مستقیم یا بدون واسطه

وقتی مشاهده گر به طور مستقیم و بدون واسطه صفت متغیر مورد نظر را مشاهده می کند اندازه گیری قد و وزن و یا شنیدن صدای قلب بیمار به وسیله پزشک مثال هایی از این نوع مشاهدات هستند.

۲- مشاهده غیر مستقیم یا با واسطه

گاهی اوقات نتیجه مشاهده دیگران در تعیین حالت یا مقدار یک صفت در جایی ثبت می شود و یا از آن مشاهده کننده پرسیده می شود در آنصورت به آن مشاهده غیر مستقیم یا با واسطه می گویند.

مشاهده رفتارها:

مشاهده شونده ممکن است انسانها یا حیوان یا اشیاء باشند گاهی اوقات هدف مطالعه مشاهده رفتارهای انسانی یا حیوانی است. چون رفتارهای انسانها یا حیوانها در هنگامی که حضور مشاهده گر را احساس کنند تحت تأثیر قرار می گیرد و از حالت طبیعی و واقعی خود خارج می شود در آنصورت مشاهده نمی تواند وضعیت رفتار واقعی فرد را نشان دهد و همراه با خطا خواهد بود بنابراین مشاهده را از این نظر به دو گروه تقسیم می کنند :

۱- مشاهده مشارکتی :

در این روش ،مشاهده گر خود را در بروز رفتارهای مورد مشاهده اش مشارکت می دهد تا حضورش در تغییر رفتار مشاهده شونده مؤثر نباشد ،مثلاً کودکی کن تا اعتماد کودکان را به خودت جلب نمائید باعث می شود که در حین بازی با آنها بتوان رفتارهای مسالمت جویانه یا پرخاشگرانه آنها را مطالعه کرد . این روش در مطالعات اجتماعی مفید خواهد بود.

۲- مشاهده غیر مشارکتی یا دوربین مخفی

مشاهده کننده پدیده مورد مشاهده را بطور پنهانی یا آشکار ملاحظه می نماید ولی خود در آن رفتار مشارکت ندارد. اگر محققی بخواهد رفتارهای نوجوانان بزه کار یا معتادان را مطالعه کند باید به شیوه های مخفی رفتارهای آنها را زیر نظر داشته باشد.

استفاده از روش مشاهده در جمع آوری داده ها با توجه به شرایط تحقیق ممکن است به صورتهای دیگری نیز تقسیم بندی شود: مشاهدات آزمایشگاهی، که معمولاً با ابزار آزمایشگاهی صورت می گیرد. مشاهدات میدانی که برای بررسی پدیده ها در طبیعت و مطالعات اجتماعی در میدان (field) جامعه مورد استفاده قرار می گیرد ، مشاهدات بدون ابزار و مشاهدات با ابزار از این نمونه ها هستند.

ب)مصاحبه :

جمع آوری داده ها از طریق پرسیدن از یک یا چند نفر و ثبت پاسخ آنها بـه عنـوان روش مصـاحبه شناخته می شود. مصاحبه مبتنی بر پرسش مستقیم و رودررو از پاسخگوست .
گفتگو و سؤالها در اطراف اهداف و فرضیه تحقیق متمرکز است . مصاحبه ها ممکن است بـه شـکل های مختلفی انجام گیرد. موفقیت مصاحبه به هوشیاری ، تجربه و مهارت مصـاحبه کننـده بسـتگی دارد. امکان تاثیرگذاری مصاحبه کننده بر پاسخگو وجود دارد . معمولاً بـر اسـاس میـزان آگـاهی و مهارت مصاحبه کننده و نیز با توجه به موضوع مورد مطالعه مصاحبه را به دو گروه تقسیم می کنند

۱- مصاحبه های ثابت یا غیر قابل انعطاف:

در این حالت معمولاً سئوالات مشخص تهیه و در یک پرسشنامه نوشته مـی شـود و همـه مصـاحبه کنندگان از افراد مورد مصاحبه به صورت یکنواخت می پرسند و پاسخها را ثبت مـی کننـد و هـیچ گونه دخل و تصرف در سئوالات یا تغییر آنها از طرف مصاحبه کنندگان مجاز نیست تا امکان ایجـاد یکنواختی بیشتری فراهم شود این شیوه برای جامعه ای با حجم زیاد مناسب تر است.

۲- مصاحبه انعطاف پذیر یا باز:

زمانی که مصاحبه شوندگان بر اساس موضـوعات کلـی درزمـان مصـاحبه اجـازه تغییـر یـا تفسـیر سئوالات را داشته باشند و بر اساس پاسخهای مصاحبه شوندگان جریان مصاحبه را هـدایت نماینـد می گوئیم مصاحبه حالت ثابت ندارد و به صورت باز ، آزاد یا انعطاف پذیر انجام می گیرد.
بدیهی است در جمعیت های زیاد که باید از مصاحبه کنندگان زیادی استفاده کرد به دلیـل اینکـه آنها توانائیهای مختلف دارند نمی توان آنها را در جریان مصاحبه آزاد گذاشت تا بـا طـرح سـئوالات مختلف و نحوه برقراری ارتباط با مصاحبه شوندگان تغییرات زیادی در نتایج پاسخها ایجاد کنند که این امر طبقه بندی و استخراج داده ها را مشکل خواهد ساخت.

ج) پرسشنامه (کتبی)

پرسش نامه کتبی از ابزارهای جمع آوری داده هاست که سئوالات در آن نوشته شده است و شخص مورد مطالعه پاسخ پرسش ها را در آن می نویسد.
وقتی که بعضی از متغیرها قابل مشاهده نباشد و یانتوانیم از طریق مصاحبه داده هـا را جمـع آوری کنیم . استفاده از روشهای پرسش نامه ای متـداولترین روشهاسـت . نکـات زیـادی در مـورد تهیـه پرسش نامه خوب باید مورد توجه قرار گیرد. که برخی از آنها مربوط به ساختار و تدوین و ویـرایش یک پرسشنامه است و برخی دیگر مربوط به محتوا و چگونگی سئوالات پرسشنامه است که در ایـن قسمت به تعدادی از آنها اشاره می شود.
انتخاب نوع سوال در پرسشنامه ها و محتوای آنها در پژوهش های مختلف متفاوت است.

انواع سوالات:

سوالات یک پرسشنامه ممکن است به شکل مختلف باشد که در مباحث سنجش و اندازه گیری یا منابع آموزشی به آنها اشاره می شود. ولی به طور کلی می توان آنها را به صورت زیر تقسم بندی کرد:

۱- سوالات کوتاه جواب
۲- سوالات چند گزینه ای
۳- سوالات کلی یا تشریحی

سوالات از نظر نوع پاسخ آنها به دو نوع ***باز و بسته*** تقسیم می شود که در هنگام طرح سوالات باز چارچوب پاسخهای ممکن تعیین نمی شود و پاسخ دهنده در جواب سوالات باز باید با ذهنیت خودش جواب های احتمای را مشخص نماید ولی در هنگام طرح سوالات بسته کلیه پاسخهای احتمالی این گونه سوالات در جلوی آنها نوشته می شود و پاسخ دهنده یک یا چند تا از پاسخ ها را انتخاب می نماید.

ویژگیهای سوال باز:

اگر خوب تدوین شود اطلاعات مفید در زمینه های متنوعی به دست می دهد .
در مسائلی چون شناخت عقاید و انگیزه ها مفید و احتمالا ضروری است .
طرح آن نیازمند مهارت ، تجربه و دقت زیاد است .
استخراج و طبقه بندی پاسخهای آن غالبا دشوار است .
تعیین روایی و پایایی آن اهمیت دارد.

ویژگیهای سوال بسته:

اگر خوب طرح شود مشکل امتناع از پاسخ را حل می کند .
به بررسی روابط متغیرهای پیش بینی شده کمک می کند و اطلاعات اضافی به دست نمی دهد .
داده های ساده ، مستقیم و قابل اعتمادی را به دست می دهد .
استخراج و طبقه بندی داده ها را تسهیل می کند .

نکات مهم در تهیه پرسشنامه:

محققی که پرسشنامه را طرح می نماید باید نکات زیر را در نظر داشته باشد :

۱. دقیقا چه چیزی را می خواهد بسنجد؟
۲. سوالات را چه افرادی پاسخ خواهند داد ؟
۳. سطح اطلاع و درک آنها از موضوع در چه حدی است ؟
۴. آیا بیشتر با سوادند یا بی سواد ؟

۵. اطلاعات گرد آمده چگونه تجزیه و تحلیل می شوند ؟ با دست یا با کامپیوتر طبقه بندی و تحلیل می شود؟

در تهیه پرسشنامه نکات محتوایی و نکات ظاهری و ویرایشی زیادی باید مورد توجه قرار گیرد. یک پرسشنامه خوب بایستی دارای ویژگیهای زیر باشد:

۱) دارای عنوان ، مشخصات و توضیحات ضروری برای پاسخگو باشد .
۲) ظاهری آراسته و برانگیزاننده داشته باشد .
۳) به گونه ای باشد که ثبت و کدگذاری داده ها را آسان گرداند .
۴) جمله بندی پرسشها دقیق و خالی از ابهام باشد .
۵) پرسشها موجب واکنشهای احساسی و عاطفی نگردد .

مراحل مختلف تهیه یک پرسشنامه:

مهمترین گام در *تهیه پرسشنامه تدوین وتنظیم صحیح آن* است .

۱. از متغیرها بعنوان نقطه آغاز شروع نمائید .
۲. پرسش ها را بطور واضح و کوتاه و با استفاده از کلمات روز مره بنویسید .
۳. مطمئن شوید که با هر سوال ، هر بار ، تنها یک نکته را مطرح می کنید .
۴. مطمئن شوید که پرسش شما نحوه پاسخ را القاء ننماید .
۵. مطمئن شوید که پرسشنامه با توجه به جلب همکاری نفرات پاسخگو طرح شده باشد.
۶. سئولاتی را بمنظور کنترل پاسخ هایی که به پرسش های حساس قبلی داده شده طرح نمائید .
۷. پرسشنامه را حتی الامکان کوتاه طرح نمائید .
۸. مطمئن شوید که در نهایت از تمام سئوالاتی که در پرسشنامه گنجانیده اید استفاده می شود .
۹. مطمئن شوید که تمام مصاحبه کنندگان و مصاحبه شوندگان فهم مشابهی از سئوالات طرح شده را دارند .
۱۰. برای اطمینان از درست بودن سازمان ، واژه بندی و ترتیب و توالی سئوالات پرسشنامه را قبل از شروع طرح امتحان کنید .
۱۱. آزمایش قبلی پرسشنامه برای تصمیم در این مورد که آیا ترجمه آن به یک یا چند زبان محلی ضرورت دارد یا ندارد ، لازم است .

اجزاء پرسشنامه:

هدف از تنظیم پرسشنامه طرح سوالاتی است که به وسیله آنها متغیرهایی مورد سنجش قرار گرفته تا اهداف طرح تامین گردد. پرسشنامه دارای چند قسمت است .

۱- عنوان و شناسنامه:

قسمتی است که به پرسشنامه هویت می دهد وشامل نام، محل اجراء، عنوان طرح تحقیقاتی، تاریخ و شماره پرسشنامه است. اگر لازم باشد برای جلب همکاری و اعتماد پاسخگو پیرامون اهداف طرح و محرمانه ماندن اطلاعات، توضیحاتی نوشته می شود.

۲- بدنه و قسمت اصلی :

متغیرها و مقیاس سنجش آنها در این قسمت آورده می شود. تمامی متغیرهای هدف شامل متغیر وابسته، مستقل، مداخله گر و زمینه ای را آورده می شود. تقدم و تأخر متغیرها در پرسشنامه بستگی به حساسیت کار شما دارد . در تحقیقات علوم پزشکی بهتر است اول متغیرهای زمینه ای، بعد مداخله گر، بعد متغیرهای اصلی به ترتیب اول متغیر وابسته و سپس مستقل را در پرسشنامه وارد کنیم. متغیرها را با واحد اندازه گیری شان ثبت کنید مثلا سن به سال وزن به کیلوگرم مشخص شود.

۳- ارزیابی وکنترل :

باید دقت کرد که پرسشنامه درست تکمیل شود لذا نام و نام خانوادگی بررسی کننده همراه با امضاء باید زیر هر پرسشنامه تکمیل شده درج گردد. یک نفر به عنوان کنترل کننده نیز پرسشنامه تکمیل شده را کنترل کند و امضاء کند .

درج تاریخ تنظیم بررسی و کنترل ضروری است . باید بررسی و کنترل در یک زمان صورت بگیرد یا کمترین زمان اختلاف را داشته باشد تا در صورت کم وکاستی در تنظیم، دسترسی فرد مورد پرسش امکان پذیر باشد.

د- استفاده از اطلاعات موجود و سیستم ثبت:

استفاده از اطلاعات موجود ، روشی ساده و ارزان برای جمع آوری داده ها است .

بهترین و شاید تنها راه برای دستیابی به اطلاعات مربوط به وقایع گذشته است .

در استفاده از اطلاعات موجود بایستی به روایی داده ها ، تعریف مفاهیم ، روشهای محاسبه ، تفاوتهای زمانی و مکانی ، تناسب داده ها با اهداف پژوهش و ... توجه کرد .

۲-۲- نمونه گیری:

اغلب به جای بررسی کلیه عناصر و اعضاءجامعه فقط تعدادی از اعضاءآن را به عنوان نمونه مورد بررسی قرار داده و نتایج حاصل را به کل جامعه تعمیم می دهند، به این روش *نمونه گیری* گفته می شود. در مقابل مطالعه همه اعضاء جامعه آماری را *سرشماری* می گویند.

تعداد کل اعضاءنمونه را حجم نمونه و تعداد کل اعضاءجامعه را حجم جامعه می گویند.

نمونه گیری در موارد زیر توصیه می گردد که برخی از این موارد به دلیل مشکلاتی است که در انجام یک مطالعه به روش سر شماری وجود دارد. اما در بعضی از مواقع اگر چه سرشماری همه اعضای جامعه آماری مشکل نباشد ولی جایز نیست:

۱- جامعه آماری نامحدود و یا درحجم وسیع باشد.

۲- بودجه آمارگیری کم باشد و یا هزینه جمع آوری اطلاعات از طریق سرشماری گران باشد.

۳- نیروی انسانی کافی جهت سرشماری در اختیار نباشد.

۴- نتایج آمارگیری در زمان کوتاه مورد نیاز باشد.

۵- کنترل بیشتر در نمونه گیری دقت را افزایش می دهد.

۶- سرشماری همه اعضای جامعه جایز نباشد.

ضمنا یادآوری می گردد که نمونه گیری در بعضی از موقعیت ها ضروری است در این موقعیت ها نوع مطالعه به گونه ای است که با انجام بررسی بر روی واحد های مورد مطالعه، به آنها خسارت وارد می شود یا واحد های مورد مطالعه از بین می روند. در چنین موقعیت هایی برای انجام مطالعات آماری مجبوریم از نمونه گیری به جای سر شماری استفاده کنیم. مثلا برای آزمایش خون افراد در تعیین وجود عامل بیماری زا باید به نمونه ای از خون اکتفا کرد. یا در مطالعات کنترل کیفی از محصولات تولیدی باید به نمونه ای از آنها اکتفا کرد و همه محصولات را از بسته بندی خارج نکرد.

۳-۲- انواع نمونه گیری :

نمونه گیری به طور کلی بر دو نوع است :

۱- نمونه گیری غیر تصادفی (قصدی)

۲- نمونه گیری تصادفی

نمونه گیری غیرتصادفی :

در نمونه گیری غیر تصادفی افراد نمونه با شناخت قبلی که محقق نسبت به جامعه دارد به دلخواه انتخاب می شوند یکی از مهمترین نکاتی که در مورد این نوع نمونه گیری مطرح است این است که تعمیم نتایج حاصل از نمونه گیری قصدی به جامعه صحیح نیست.

انگیزه های مختلفی در انتخاب نمونه به صورت غیر تصادفی موثر است که در بعضی از موقعیت ها راحتی کار در دسترس بودن نمونه ها و یا عدم تمایل افرادی که به تصادف انتخاب شده اند جهت شرکت در مطالعه همگی باعث می شوند که در برخی تحقیقات بهداشتی و پزشکی نمونه گیری از

افراد در دسترس یا مراجعین به مراکز یا داوطلبین یا گروه های خاص مانند معتادین زندانی به صورت غیر تصادفی انجام گیرد.

الف- نمونه گیری سهل و آسان (convenience sampling)

در مطالعاتی که برای برآورد پارامترهای جامعه دقت زیادی لازم نیست نمونه های در دسترس مانند مراجعین به یک مرکز بهداشتی و در مانی یا خانوارهای همسایه و یا افراد یک اردوگاه را می توان انتخاب کرد.

ب- نمونه گیری سهمیه ای (quota sampling)

اگر نمونه های در دسترس را در هریک از گروه های موجود در جامعه به نسبت جمعیت آنها انتخاب نمایند به نوعی سهمیه بندی رعایت می شود در این صورت احتمال این که نمونه ها معرف جامعه باشند افزایش می یابد.

مثال: انتخاب افراد مراجعه کننده به مراکز درمانی را به صورت سهمیه ای در سطح شهر به گونه ای انتخاب نمایند که درمانگاه های عمومی و شبانه روزی دولتی و خصوصی را شامل شود.

نمونه گیری تصادفی :

در این نوع نمونه گیری انتخاب افراد به طور شانسی (تصادفی)صورت می گیرد به طوریکه احتمال یا شانس انتخاب هر یک از واحد های جامعه معین می باشد . به این نوع نمونه گیری تصادفی (نمونه گیری بر حسب احتمال) نیز می گویند.نمونه گیری های تصادفی ممکن است به صورت های مختلف انجام گیرد :

۱) نمونه گیری تصادفی ساده :

با استفاده از روشهایی مانند استفاده از جدول اعداد تصادفی یا قرعه کشی می توان افراد نمونه را انتخاب کرد ممکن است در هر بار انتخاب مقدار احتمال انتخاب شدن افراد ثابت بوده و برای همه افراد جامعه یکسان باشد که در این صورت نمونه گیری تصادفی ساده با جایگذاری می باشد به عبارت دیگر اگر پس از انتخاب یک فرد از نمونه دوباره همان فرد شانس انتخاب شدن راداشته باشد و به جامعه برگردانده شود تا در انتخاب بعدی نیز شرکت داده شود نمونه گیری با جایگذاری می باشد . در حالتی که پس از انتخاب یک عضو نمونه از میان بقیه اعضاء افراد دیگر انتخاب شوند با هر بار انتخاب عضو نمونه، احتمال انتخاب برای فرد بعدی بیشتر خواهد شد که در این صورت نمونه گیری بدون جایگذاری خوانده می شود.

مثال - اگر بخواهیم از بین ۳۰ دانشجوی یک کلاس که به ترتیب از ۱ تا ۳۰ شماره دارند به روش تصادفی ساده یک نمونه تصادفی ۵ تایی انتخاب کنیم با استفاده از جدول اعداد تصادفی (ضمیمه) به شرح زیر عمل می کنیم.

ابتدا به تصادف بر روی جدول اعداد تصادفی یک نقطه را انتخاب می کنیم. اگر به سمت راست چپ بالا یا پایین به ترتیب عددهای دو رقمی (به تعداد رقم های ۳۰ یعنی حجم جامعه) را جدا می کنیم. هر یک از این اعداد در فاصله ۰۰ تا ۳۰ باشد نشان دهنده شماره دانشجویی است که در نمونه انتخاب شده است. و اگر هر یک از این اعداد دو رقمی بیشتر از ۳۰ باشد آنرا ندید گرفته و عدد بعدی را در نظر می گیریم که شماره عضو بعدی نمونه خواهد بود. بدین ترتیب ادامه می دهیم تا ۵ عضو نمونه انتخاب شود. از روش قرعه کشی نیز می توان نمونه گیری تصادفی ساده را انجام داد.

۲) نمونه گیری تصادفی به طریق طبقه بندی :

این روش زمانی بکار می رود که اطلاعات دقیقتری در مورد جامعه آماری وجود داشته باشد و بتوان جامعه را بر اساس یک ویژگی خاصی به جمعیتهای کوچکتر تقسیم بندی کرد. بطوری که این جمعیت ها طبقات یا قشرهایی از جامعه را تشکیل داده باشند در اینصورت از هر قشر یا طبقه به نسبت درصد کل جامعه تعدادی بعنوان اعضاءنمونه انتخاب می شوند.

مثلاً در نمونه گیری از جمعیت ساکن در یک استان ابتدا به شهرستانهای مختلف تقسیم بندی کرده و به نسبت درصد جمعیت استان که در شهرستانها سکونت دارند تعداد نمونه مربوط به هر شهرستان را مشخص و از آن شهرستان انتخاب می کنند.

نکته مهم در این نوع نمونه گیری این است که ملاک مرزبندی طبقات باید واضح وروشن باشد تا هیچ فردی شانس انتخاب در بیش از یک طبقه را پیدا نکند و هیچ کسی از جامعه نباشد که عضو هیچ طبقه ای به حساب نیامده باشد. ملاک طبقه بندی می تواند جنس، محل سکونت، گروه های سنی و... باشد. در عمل معمولا معیار تخصیص نمونه به طبقات مختلف نسبت حجم طبقات به حجم جامعه است ولی در بعضی موارد ممکن است هزینه سرانه نمونه گیری در طبقات متفاوت باشد لذا علاوه بر حجم طبقات، هزینه را در فرمول های تعیین حجم نمونه در نظر گرفت، که باید به کتابهای نمونه گیری مراجعه کرد.

۳) نمونه گیری به طریق خوشه ای :

در نمونه گیری خوشه ای واحد نمونه گیری فرد نیست بلکه یک گروه از افراد جامعه به عنوان یک خوشه به تصادف انتخاب و سپس واحدهای جامعه در آن خوشه ها انتخاب می گردند . مثلاً برای انتخاب تعدادی نمونه در یک نواحی روستائی یک شهرستان می توان هر یک از روستاها را به عنوان یک خوشه تلقی نموده و ابتدا تعدادی از خوشه ها را به تصادف انتخاب و سپس افراد خوشه ها انتخابی را مورد بررسی قرار داد. اگر جمعیت روستاها متفاوت است می توان خوشه های یکنواخت تری مانند ۱۰ یا ۲۰ خانواری تعریف کرد.

۴) نمونه گیری به روش منظم (سیستماتیک):

در این روش انتخاب واحد ها از یکدیگر مستقل نیست بطوریکه پس از انتخاب اولین عضو نمونه به صورت تصادفی با منظور کردن یک فاصله معین به ترتیب افراد بعدی انتخاب می گردند . در حقیقت این نوع نمونه گیری ممکن است به شبه تصادفی نیز موسوم گردد. برای این کار از روش زیر استفاده می کنیم:

فرض کنیم بخواهیم از یک جامعه N عضوی، که لیستی با شماره های از ۱ تا N دارد، نمونه ای به حجم n انتخاب کنیم. ابتدا نسبت $k=\frac{N}{n}$ را محاسبه می کنیم واز بین ۱ تا k یک عدد را به تصادف انتخاب می نمائیم که همان شماره اولین عضو نمونه در جامعه خواهد بود. حال اگر به آن شماره مقدار k را اضافه نماییم شماره نمونه دوم تعیین می گردد. با ادامه این روند شماره تمامی نمونه ها انتخاب می شود.

باید توجه شود در جمعیت هائی که تغییرات پدیده های مورد بررسی دارای تناوب است اگر K با دوره تناوب پدیده مورد بررسی یکسان باشد ممکن است همه نمونه ها از یک وضعیت نسبتا ثابتی برخوردار باشند. که این امر باعث خطا در تعمیم نتایج می شود. برای جلوگیری از این امر ممکن است K را تغییر داده تا با دوره تناوب متغیر یکسان نباشد.

۵) نمونه گیری چند مرحله ای :

در این روش در هر مرحله ممکن است با استفاده از یکی از روشهای فوق طبقات یا خوشه ای معین انتخاب و سپس در مراحل بعدی نمونه گیری، مجددا واحدهای دیگری از افراد انتخاب گردند . مثلاً در مرحله اول برای انتخاب تعدادی نمونه در یک استان ابتدا شهرستانهایی به تصادف انتخاب و سپس در هر شهرستان مناطق شهری یا روستائی به عنوان خوشه انتخاب و در مرحله بعدی در هر خوشه تعدادی از افراد خانوار بصورت تصادفی ساده انتخاب شوند.

نکات کاربردی در تحقیقات:

انواع تورش (Bias) در نمونه گیری:

۱- بیماران مراجعه کننده به مراکز یا بستری بیمارستان ها، نمونه همه بیماران جامعه نیستند.

۲- داوطلبین با غیر داوطلبین خصوصیات یکسانی ندارند.

۳- نمونه هایی که در مطالعه حاضر به همکاری نیستند با بقیه تفاوت هایی دارند.

۴- مطالعه افراد زندانی و کسانی که به اجبار در اختیار دیگران هستند، ممکن است نشان دهنده شدت یا عدم حساسیت موضوع از نظر آنها نباشد.

نکاتی در باره گزارش نتایج تحقیق:

۱- در انتشار نتایج تحقیق باید به معیار های ورود نمونه ها در مطالعه توجه داشت.
۲- دلایل اینکه بعضی افراد از مطالعه خارج شده ان چیست؟ (معیار خروج)
۳- اختلاف نمونه با بقیه جمعیت بیان شود.
۴- تعمیم پذیری نتایج به جامعه قابل تشخیص باشد.

پروژه عملی:

الف- پرسشنامه ای جهت بررسی میزان آگاهی، نگرش و عملکرد دانشجویان یا یکی از گروه های اجتماعی دیگر را در باره یک بیماری یا یک موضوع بهداشتی تهیه و در سطح محدود کاربرد آن را ارزیابی کنید.
ب- اگر بخواهید تحقیقی در باره موضوع سوال ۱ در سطح جامعه آماری مربوط اجرا کنید چه روش نمونه گیری را ترجیح می دهید؟ چرا؟
ج- روشهای جمع آوری داده ها و روش نمونه گیری را در مورد یک مقاله تحقیقی منتشر شده در کلاس بحث نمایید. نقاط قوت و ضعف آن را بررسی کنید.

نمونه سؤالات فصل دوم:

۱- کدام یک از روش های گردآوری داده ها برای یک مطالعه تجربی مناسب تر است؟
الف- مشاهده ب- مصاحبه
ج- مکالمه د- ثبت
۲- در تحقیقاتی که مبتنی بر داده های موجود یا دست دوم هستندکدام یک از روش های گردآوری داده ها به کار می رود؟
الف- مشاهده ب- مصاحبه
ج- مکالمه د- ثبت
۳- در کدام یک از روشهای نمونه گیری انتخاب نمونه ها مستقل از یکدیگر نیست؟
الف- تصادفی ساده ب- تصادفی خوشه ای
ج- تصادفی سیستماتیک یامنظم د- تصادفی طبقه بندی شده
۴- کدام یک از موارد زیر از منابع تورش در نمونه گیری به حساب می آیند؟
الف- استفاده از داوطلبین ب- بیماران با دوره بیماری کوتاه
ج- توجه به موارد ثبت شده د- همه موارد

۵- در تحقیقاتی که در سطح شهرستان انجام می شود و تعدادی از کلاس ها انتخاب وسپس بعضی از دانش آموزان آن کلاسها مطالعه می شوند چه نوع نمونه گیری انجام می شود؟

الف- تصادفی ساده با جایگذاری ب- طبقه ای

ج- خوشه ای د- ساده بدون جایگذاری

۶- در تحقیقاتی که در سطح استان انجام می شود و نمونه ای از خانوارها انتخاب وسپس بعضی از افراد آن خانوارها مطالعه می شوند چه نوع نمونه گیری انجام می شود؟

الف- تصادفی ساده با جایگذاری ب- طبقه ای

ج- خوشه ای د- طبقه ای و خوشه ای

فصل سوم :
طبقه بندی و نمایش داده ها

هدف های فصل سوم: پس از مطالعه این فصل خواننده باید بتواند:

۱- نسبت به طبقه بندی داده های جمع آوری شده اقدام نماید.

۲- ویژگی های نمودارهای مهم آماری را بیان کند.

۳- نسبت به انتخاب نمودار مناسب و رسم آن اقدام کند.

۴- با استفاده از نرم افزار های کامپیوتری نمودار مورد نظرش را رسم کند.

۱-۳- توصیف داده ها به وسیله جدول :

پس از جمع آوری داده ها به منظور فهم و درک شهودی در مورد توزیع داده ها لازم است با شرایط و ضوابطی، داده ها را در قالب جداول خاصی قرار داد این جداول برای مشاهدات مربوط به متغیرهای کیفی یا کمی پیوسته ممکن است با کمی اختلاف تهیه گردند.

الف- اگر نتیجه مشاهدات مربوط به صفات کیفی باشد، غالباً گروه بندی نتایج به سادگی انجام خواهد گرفت.

مثال ۱- در یک آزمایشگاه نوع گروه خون ۲۵ نفر به صورت زیر مشخص گردید.

A, O, A, B, A, A, A, AB, O, A, A, A, O, B, O, B, A, B, O, B, O, A, B, B, A

برای گروه بندی این اطلاعات کافی است تعداد هر یک از گروههای خونی را معین کرده بعنوان فراوانی آن گروه محسوب نمائیم که نتایج حاصل در جدول فراونی زیر می آید.

جدول ۱-۳- توزیع فراوانی گروههای خونی روی ۲۵ نفر از مراجعین به یک آزمایشگاه

گروه خونی	فراوانی (f)	فروانی نسبی
A	۱۱	۰/۴۴
B	۷	۰/۲۸
AB	۱	۰/۰۴
O	۶	۰/۲۴
جمع	n=۲۵	۱/۰۰

ب- اگر نتیجه مشاهدات مربوط به متغیرهای کمی گسسته باشد، مشخص نمودن فراوانی های مقادیر متغیر نیز کاری دشوار نیست دراین نوع متغیرها باید در تمایز متغیر از فراوانی آن دقت لازم به عمل آید.

مثال ۲: فرض کنید تعداد فرزندان یک نمونه ۵۰ خانواری از روستای الف، به صورت زیر می باشد.

۳، ... ، ۰ ، ۲، ۲، ۳، ۵، ۱ ، ۲، ۰، ۴، ۲

به منظور خلاصه کردن و سازماندهی اطلاعات جدول توزیع فراوانی این داده ها، به صورت زیر تهیه می گردد . برای تشکیل جدول توزیع فراوانی ممکن است با استفاده از خط نشان تعدا خانوارهایی که صفر ، ۱ ،۲ یا ... فرزند دارند را شمارش کرده که تعداد کل این خط نشانها برای هر یک از مقادیر متغیر یا طبقات و زیر گروههای کوچکتر از مقادیر متغیر، همان فراوانی ساده یا مطلق آن مقدار یا طبقه را تشکیل می دهد.

جدول (۲-۳)- توزیع فراوانی تعداد فرزندان خانوارهای روستای الف را نشان می دهد.

تعداد فرزند	خط نشان	(f) فراوانی ساده	فراوانی نسبی $\frac{f_i}{n}$ ( )	(F)فراوانی تجمعی	فراوانی تجمعی نسبی
۰	////////////////	۱۵	۰/۳۰	۱۵	۰/۳۰
۱	//////////	۱۰	۰/۲۰	۲۵	۰/۵۰
۲	//////////////	۱۳	۰/۲۶	۳۸	۰/۷۶
۳	//////	۶	۰/۱۲	۴۴	۰/۸۸
۴	///	۳	۰/۰۶	۴۷	۰/۹۴
۵	///	۳	۰/۰۶	۵۰	۱/۰۰
جمع	۵۰	۵۰	۱		

در جدول توزیع فراوانی معمولاً ستونهای فراوانی ساده ، فراوانی تجمعی و فراوانی تجمعی نسبی محاسبه می گردد.

فراوانی نسبی :

برای محاسبه فراوانی نسبی از فرمول زیر استفاده می شود.

$$\text{فراوانی نسبی} = \frac{\text{فراوانی طبقه}}{\text{فراوانی کل}}$$

بدیهی است که مجموع فراوانی نسبی برای همه طبقات ۱ خواهد بود که به صورت زیر نشان داه می شود . که K برابر تعداد طبقات و n برابر حجم نمونه است.

$$\sum_{i=1}^{k}\frac{f_i}{n}=1\Leftrightarrow\sum_{i}^{k}f_i=n$$

فراوانی تجمعی :

اگر فراوانی ساده هر طبقه را با فراوانی ساده طبقات قبل از آن جمع نموده و برای آن طبقه منظور نمائیم فراوانی تجمعی یا تراکمی آن طبقه نامیده می شود که به صورت زیر می توان نوشت :

$$F_i=\sum_{j=1}^{i}f_j=f_1+f_2+\ldots+f_i=\sum_{j=1}^{i-1}f_j+f_i=f_i+F_{i-1}$$

و F_i فراوانی تجمعی طبقه i- ام و f_i فراوانی ساده طبقه j- ام می باشد که j از ۱ تا i تغییر می کند.

در مثال قبل، فراوانی تجمعی طبقه اول همان فراوانی ساده طبقه اول یعنی ۱۵ خواهد بود و فراوانی تجمعی طبقه دوم برابر ۲۵=۱۰+۱۵ و فروانی تجمعی طبقه سوم برابر ۳۸+۱۳+۱۰+۱۵ و ... است.

بدیهی است فراوانی تجمعی طبقه آخر همیشه در جدول توزیع فراوانی برابر n خواهد بود.

فراوانی تجمعی نسبی :

اگر فراوانی تجمعی هر طبقه را بر فراوانی کل تقسیم نمائیم فراوانی تجمعی نسبی آن محاسبه می گردد . به عبارت دیگر می توان با تجمعی کردن فراوانی نسبی طبقات نیز فراوانی تجمعی نسبی را محاسبه کرد.

مثلاً فراوانی تجمعی نسبی طبقه اول ۰/۳۰ = $\frac{15}{50}$ و فراوانی تجمعی نسبی طبقه دوم ۰/۵۰ = $\frac{25}{50}$ و ... است.

درصد فراوانی نسبی و درصد فراوانی تجمعی نسبی :

اگر هر یک از فراوانی های نسبی یا فراوانی تجمعی نسبی را در عدد ۱۰۰ ضرب کنیم درصد فراوانی نسبی و یا درصد فراوانی تجمعی نسبی به دست خواهد آمد.

ج- گروه بندی نتایج مشاهدات در حالتی که متغیر کمی پیوسته باشد:

در این گونه مشاهدات درنظر گرفتن نکاتی برای گروه بندی ضرورت دارد.

تعیین تعداد طبقات به ابتکار و تشخیص محقق وابسته است ولی معمولاً توصیه می شود که تعداد طبقات کمتر از ۵ و بیشتر از ۲۵ نباشد و بهتر است که تعداد طبقات بین ۸ تا ۱۵ برای نمونه های زیاد و بین ۵ تا ۱۰ برای نمونه های کم، طوری انتخاب شود که بر اساس آن و بر اساس دامنه تغییرات داده ها ، طول هر طبقه به صورت عددی سر راست و گرد شده، تعیین گردد.

مرحله اول: دامنه تغییرات محاسبه می شود.

R= مقدار حداکثر (ماکزیمم) - مقدار حداقل (مینیمم)

در حالتی که اعداد با تقریب کمتر از x گرد شده باشند به اندازه x به مقدار اختلاف مقدار حداقل و حداکثر اضافه می گردد.

مرحله دوم: اگر k تعداد طبقات فرض شود از فرمول $(C=\frac{R}{K})$ مقدار (c) طول طبقه، محاسبه می گردد.گاهی اتفاق می افتد که طول طبقات یکسان در نظر گرفته نشود ولی برای سهولت کار اغلب اوقات مساوی در نظر گرفته می شود.

مرحله سوم: با توجه به شمارش مقادیر داده ها در هر یک از طبقات، فراوانی داده ها تعیین می گردد.

مثال ۳- فرض کنید نمونه ای مرکب از ۱۰۰ نفر انتخاب و قد هر یک را بر حسب سانتیمتر ثبت کنیم پس از تعیین مقدار دامنه تغییرات و تعداد طبقات طول هر طبقه محاسبه می گردد.

نتایج حاصل در جدول (۳-۱) خلاصه گردیده است.

جدول (۳-۳) توزیع فراوانی و فراوانی نسبی و تجمعی قد نمونه ای به حجم ۱۰۰ نفر

شماره طبقه	حدود طبقات	وسط طبقات	فراوانی f	فراوانی نسبی $\frac{f}{n}$	فراوانی تجمعی
۱	۱۴۵-۱۵۴	۱۴۹/۵	۳	۰۳	۳
۲	۱۵۵-۱۶۴	۱۵۹/۵	۷	۰/۰۷	۱۰
۳	۱۶۵-۱۷۴	۱۶۹/۵	۲۴	۰/۲۴	۳۴
۴	۱۷۵-۱۸۴	۱۷۹/۵	۳۲	۰/۳۲	۶۶
۵	۱۸۵-۱۹۴	۱۸۹/۵	۲۸	۰/۲۸	۹۴
۶	۱۹۵-۲۰۴	۱۹۹/۵	۶	۰/۰۶	۱۰۰
			n=۱۰۰	۱/۰۰	

در تهیه جدول ها باید به نکات زیر توجه کرد :

۱- جدول ها باید حتی الامکان ساده باشند.

۲- جدول ها باید عنوانی رسا و کامل داشته باشند و معمولاً با شماره جدول های مختلف از هم متمایز گردند.

۳- اطلاعات داخل جدول باید از منبع معتبر و مآخذ آن در زیر جدول معرفی گردد.

۴- در جدول از علائم شناخته شده یا تعریف شده استفاده گردد.

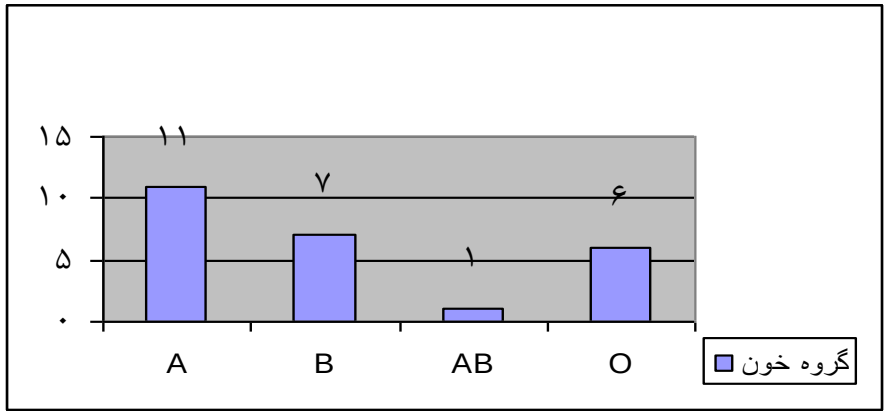
۵- در جدول ستون جمع و سطر جمع آمده باشد.

۲-۳- نمایش داده ها بوسیله نمودار:

هر چند جدول ها تا حدودی وضعیت توزیع داده ها را مشخص می نماید ولی به منظور درک آسانتر می توان اطلاعات جداول را بوسیله نمودارهای مختلف نشان داد. برای نمایش اطلاعات ممکن است از نمودار های زیر استفاده نمود :

۱- نمودار نرده ای – میله ای (Bar Diagram)

معمولاً برای نمایش اطلاعات مربوط به صفات کیفی از این نوع نمودار استفاده می گردد . به عنوان مثال : اگر توزیع فراوانی مربوط به جدول (۱-۱) را در نظر بگیریم . نمودار نرده ای زیر را می توان برای آن ترسیم کرد :



شکل ۱-۳- توزیع گروههای خونی در نمونه ای به حجم ۲۵ نفر از مراجعین به یک آزمایشگاه

گاهی اوقات ممکن است به جای استفاده از ستونهای مستطیل شکل از خطوطی بـه طـول فراوانـی هر طبقه استفاده نمود.

۲- نمودار دایره ای (Pie Diagram)

مانند نمودار نرده ای معمولاً اطلاعات مربوط به صفات کیفی که نتایج سنجش آنهـا بصـورت اسـمی یا رتبه ای طبقه بندی می شوند را بوسیله نمودار دایره ای نشان می دهند . بـرای رسـم ایـن نـوع نمودارها با رسم شعاع هائی سطح دایره را به تناسب فراوانی های هر گروه تقسیم مـی کنیـد. بـرای این کار ممکن است دایره را به صورت ۳۶۰ درجه در نظر گرفته و از فرمول زیر قسمتی از دایره که مربوط به هر گروه می شود را محاسبه کرد.

قسمتی از دایره (بر حسب درجه) که به گروه i ام مربوط است = $\frac{360}{n} \times f_i$

فراوانی نسبی طبقه i ام ×۳۶۰ = $\frac{f_i}{n}$ ×۳۶۰=

در حالت دیگر ممکن است مساحت دایره را بر حسب درصد فراوانی نسبی هر طبقه تقسیم کرد.

درصد فراوانی نسبی طبقه i-ام = فراوانی نسبی طبقه i ×۱۰۰ = $\frac{f_i}{n}$ ×۱۰۰ = $\frac{۱۰۰}{n}$ ×fi

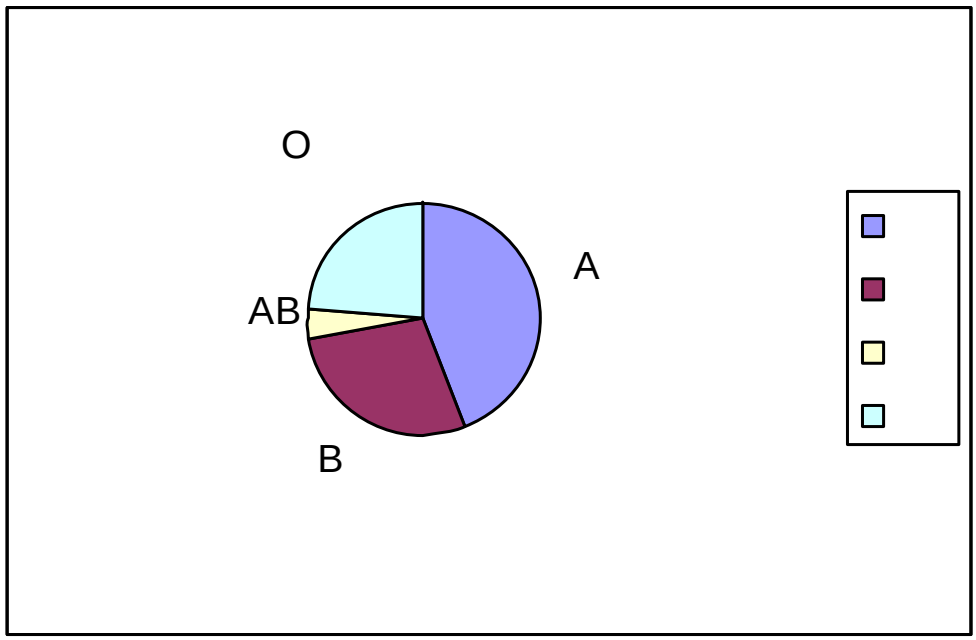
حال به عنوان مثال اطلاعات جدول (۱-۱) را به صورت نمودار دایره ای نشان می دهیم.

قسمتی که مربوط به گروه خونی A است بر حسب درجه = $\frac{360}{25}$ ×۱۱=۱۵۸/۴

قسمتی که مربوط به گروه خونی B است بر حسب درجه = $\frac{۳۶۰}{۲۵}$ ×۷=۱۰۰/۸

قسمتی که مربوط به گروه خونی AB است بر حسب درجه = $\frac{۳۶۰}{۲۵}$ ×۱=۱۴/۴

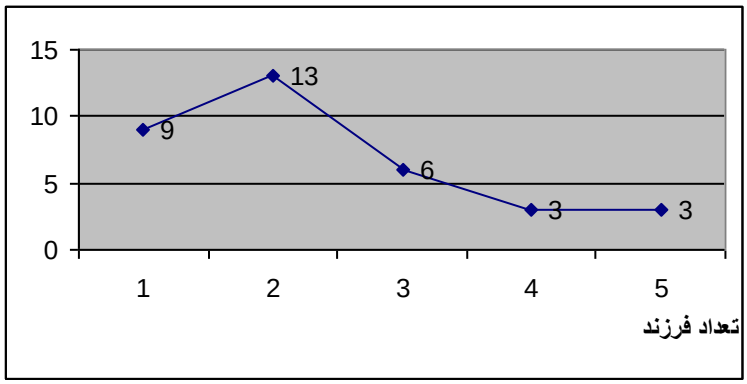
قسمتی که مربوط به گروه خونی O است بر حسب درجه = $\frac{۳۶۰}{۲۵}$ ×۶=۸۶/ ۴



شکل (۲-۳)- توزیع گروههای خونی در نمونه ای به حجم ۲۵ نفر از مراجعین به آزمایشگاه .

۳- نمودار چند گوش یا به صورت خط شکسته :

این نمودار معمولاً برای نمایش داده های کمی گسسته استفاده می شود . روی محور طولها مقـادیر متغیر ورودی محور عرضها مقادیر مربوط به فراوانی را معین می کنیم . نقاط متناظر بـرای طـول و عرض معین را بهم وصل نموده ، نمودار چند گوش یا خط شکسته بدست می آید.



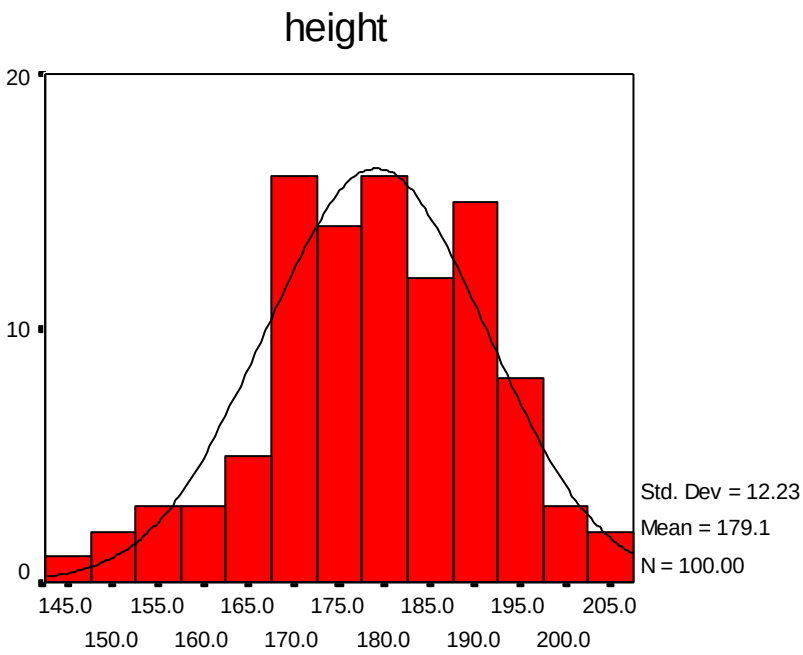
شکل (۳-۳)- توزیع فراوانی تعداد فرزندان نمونه ای مرکب از ۵۰ خانوار از روستای الف

۴- بافت نگار یا هیستوگرام (Histogram) :

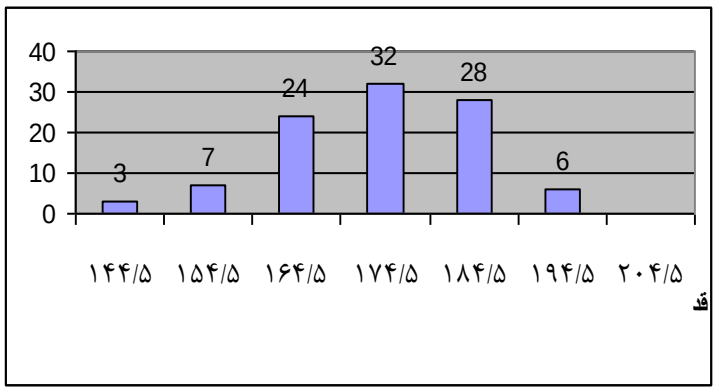
برای نمایش اطلاعات کمی پیوسته از هیستو گرام استفاده مـی شـود. ارتفـاع مسـتطیلهائی کـه در هیستو گرام رسم می شود . معمولاً برابر فراوانی ساده یا نسبی هر طبقه در نظر گرفته می شود ولی در صورتیکه به نسبت معینی عرض مستطیلی از بقیه متفاوت باشد در ارتفاع همان نسبت بصـورت عکس باید تأثیر نماید مثلاً اگر عرض یک مستطیل که دامنه آن طبقه است دو برابر طبقـات دیگـر باشد باید ارتفاع آن که فراوانی آن طبقه می باشد نصف گردد.

برای مقایسه توزیع متغیر بر اساس سطح زیر هیسـتوگرام و توزیـع نرمـال بیشـتر نـرم افـزار هـای کامپیوتری هنگام رسم هیستوگرام ها تغیرات متغیر را با منحنی توزیع نرمال بـر روی یـک نمـودار نشان می دهند تا میزان همخوانی یا عدم همخوانی توزیع متغیـر بـا توزیـع نرمـال مشـخص شـود (شکل ۴-۳).

برای این عدم همخوانی ضریب چولگی یا کشیدگی از راست یا چپ محاسبه مـی شـود کـه مقـدار کشیدگی از سمت راست یا چپ را نسبت به منحنی نرمال تعیین می کند.



شکل (۴-۳) هیستوگرام فراوانی ساده اندازه های قد ۱۰۰ نفر



اگر در هیستو گرام به جای کرانه های واقعی طبقات تنها از نقاط وسط طبقات بر روی محور طولها استفاده نمائیم بعبارت دیگر اگر در نقاط وسط اضلاع فوقانی مستطیلها رابه هم وصل نمائیم نمـودار چند گوش یا چند ضلعی فراوانی بدست خواهد آمد.

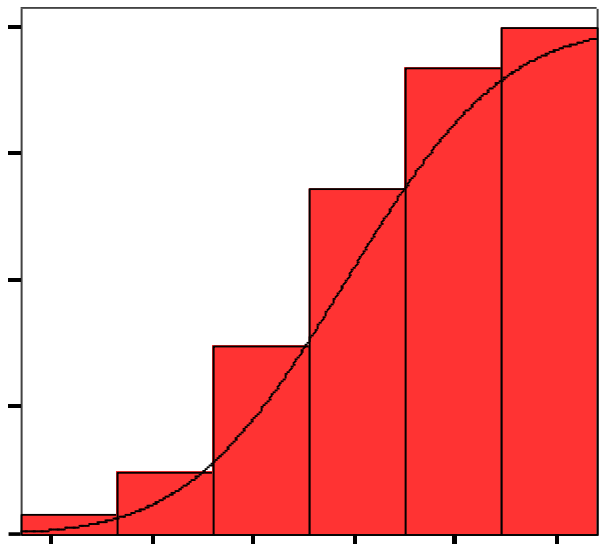
نکته:

تفسیر هیستوگرام ها برای متغیر های پیوسته بر پایه شکل و مساحت آن ها ست که در نواحی مختلف توزیع شده است. اگر هیستوگرام بر حسب فراوانی نسبی طبقات رسم شود می توان مساحت زیر هیستوگرام را برابر یک واحد مربع در نظر گرفت که با مفهوم آماری مجموع فراوانی های نسبی برابر یک است متناظر خواهد بود. پس از رسم منحنی حدی ناشی از چند ضلعی هیستوگرام، و انتگرال گیری از آن، مساحت سطح زیر منحنی در کل دامنه تغییرات متغیر مربوطه برابر یک می شود. از طرفی احتمال وقوع کلیه حالات ممکن نیز برابر یک است. بنابر این بین مفاهیم فراوانی نسبی و احتمال قرار گرفتن مقادیر متغیر در یک فاصله و مساحت سطح زیر منحنی در آن فاصله یک تناظر یک به یک برقرار است. این موضوع مبنای مباحث توزیع های آماری و احتمالاتی است که در قالب فرمول های ریاضی ارائه می شود و کاربرد های وسیعی پیدا کرده است.

۵- هیستوگرام تجمعی

اگر به جای استفاده از اندازه های فراوانی ساده یا نسبی بر روی محور عرضها از فراوانی تجمعی استفاده شود هیستو گرام فراوانی تجمعی رسم خواهد شد.

بدیهی است هیستو گرام فراوانی تجمعی همواره صعودی خواهد بود. اگر نقاط متناظر با کرانه های بالای طبقات و فراوانی های تجمعی هر طبقه را به هم وصل نمائیم نمودار چند ضلعی فراوانی (اوجایو) بدست خواهد آمد.



height

شکل (۵-۳)- هیستو گرام تجمعی و چند ضلعی اوجایو قد ۱۰۰ نفر (مربوط به جدول ۳-۳)

صدکها(percentiles):

توجه کنید که اگر در رسم هیستو گرام تجمعی از فراوانی نسبی تجمعی یا درصد آن استفاده شود به سادگی می توان مشخص کرد که مقدار متغیر برای چند درصد افراد نمونه از یک مقدار معین کمتر یا بیشتر است و یا **صدک** های مختلف را برای اندازه های متغیر محاسبه نمود. مثلاً برای تعیین صدک ۲۵ ام (یا چارک اول) یا صدک پنجاهم (میانه)و یا صدک هفتادوپنجم (چارک سوم) می توان بر روی محور عرضها در حالتیکه درصد فراوانی تجمعی نسبی را منظور نموده ایم نقاط ۲۵ درصد یا ۵۰ درصد یا ۷۵ درصد را مشخص نموده و خطوطی موازی محور طولها رسم نموده تا چند ضلعی اوجایو را قطع نماید و از نقاط تقاطع بر محور طولها عمودهائی رسم گردد. نقاط بدست آمده به ترتیب نشان دهنده مقادیر چارک اول، میانه و چارک سوم (صدک بیست و پنجم ، میانه ،صدک هفتادوپنجم) می باشد. با انتخاب k درصد می توان صدک kام را مشخص کرد.

۶- نمودار های دیگر:

نمودار نقطه ای، نمودار ساقه و برگ، نمودار جعبه ای و نمودار پراکنش (scatter diagram) از انواع دیگر نمودار ها هستند که در نرم افزار های کامپیوتری می توان آن ها را برحسب نوع و تعداد

متغیر ها مورد استفاده قرار داد. برای تعیین همبستگی بین دو متغیر کمّی از نمودار پراکنش استفاده می شود. نمونه هایی از آن در فصل هشتم در مبحث تحلیل همبستگی رسم شده است.

پیشنهاد تمرین عملی:

برای انجام طبقه بندی داده ها یا ارائه جدول توزیع فراوانی افراد بر حسب متغیرهای یک تحقیق به وسیله نرم افزار، داده های واقعی یا فرضی را وارد جدول یا همان صفحه گسترده نرم افزارهای Excel یا SPSS نمایید. برای این کار داده های هر سطر مربوط به هر فرد و ستونها نشان دهنده پاسخ هر سوال پرسشنامه یا همان متغیرها هستند که پس از کدگذاری وارد کامپیوتر می شود. پس از ذخیره کردن داده ها در فایل مشخص، نمودارها و جداول مناسب را تهیه کنید.

نمونه سؤالات فصل سوم:

۱- فراوانی تجمعی آخرین طبقه در یک جدول توزیع فراوانی کدام است؟

الف- نصف فراوانی کل ب- فراوانی کل

ج- فراوانی مطلق آخرین طبقه د- مجموع فراوانی های نسبی

۲- در یک جدول توزیع فراوانی، مجموع فراوانیهای نسبی برابر چند است؟

الف- صفر ب- یک

ج- ۱۰۰ د- n

۳- در صورتی که بخواهیم میزان مرگ ومیر در یک کشور را در ۱۰ سال متوالی گذشته با نمودار نشان دهیم از کدام یک از نمودارهای معمول بهتر است استفاده شود؟

الف- هیستوگرام ب- ستونی ج- چند بر فراوانی د- دایره ای

۴- برای متغیرهای کمی گسسته چه نوع نموداری مناسبتر است ؟

الف- دایره ای ب- ستونی یا نرده ای

ج- هیستوگرام د- چند بر یا پلی گون

۵- در رسم هیستوگرام، اگر طول طبقه ای نصف طول طبقات دیگر باشد فراوانی آن طبقه چگونه تصحیح می شود؟

الف- نصف می شود. ب- دو برابر می شود.

ج- تغییری نمی کند. د- هیچکدام

۶- برای متغیر های زیر چه نموداری مناسب تر است در جلوی آن بنویسید.

۱- جنس نوزاد ۲- سن

۳- درجه حرارت بدن ۴- تعداد و فرزند

۵- میزان کلسترول خون ۶- همبستگی وزن و قد

تمرین:

۱- در زمینه رشته تحصیلی خود مجموعه ای از داده های واقعی را جمع آوری کرده جدول توزیع فراوانی ، فراوانی نسبی و نمودار مناسبی برای آن ارائه نمائید ؟

۲- در یک مطالعه سلامت روانی ۴۰۴ بیمار که به درمانگاه پوست بیمارستان بوعلی ساری مراجعه کردند با استفاده از پرسشنامه GHQ24 ارزیابی شد. تعداد ۲۵۴ نفر (۶۰/۶ درصد) نمره ۲۳ یا کمتر را کسب کرده اند و دارای سلامت روانی بودند و ۱۵۹ نفر (۳۹/۴ درصد) نمره ۲۴ یا بیشتر داشتند که مشکوک به اختلال روانی بودند (زهره حاج حیدری و همکاران ۱۳۸۴).

نمودار مناسب برای نمایش این داده ها کدام است؟ رسم کنید.

۳- توزیع فراوانی زنان همسر دار ۱۵-۴۹ ساله برحسب تعداد فرزند به دنیا آمده در یک شهرستان در جدول زیر آمده است.

نسبت به محاسبه فراوانی نسبی و تجمعی اقدام نمایید.

تعداد فرزند	فراوانی
۰	۶۵
۱	۱۹۲
۲	۲۳۹
۳	۱۰۵
۴	۴۹
۵	۱۱
بیشتر یا مساوی ۶	۸
جمع	۶۶۰

۴- توزیع فراوانی زنان همسر دار ۱۵-۴۹ ساله برحسب تعداد فرزند مورد انتظار در یک شهرستان در جدول زیر آمده است.

تعـــــداد فرزنـــــد موردانتظار	فراوانی
۰	۳
۱	۱۲۲
۲	۴۴۳
۳	۶۵
۴	۲۶
۵	۱
جمع	۶۶۰

الف- نسبت به محاسبه فراوانی نسبی و تجمعی اقدام نمایید.

ب- توزیع فراوانی زنان را برحسب تعداد فرزندان به دنیا آمده در تمرین قبل و تعداد فرزندان مطلوب از نظر آنها را در یک دستگاه مختصات رسم نمایید و با هم مقایسه کنید.

۵- جدول توزیع فراوانی زنان همسردار برحسب نوع روش یا وسیله پیشگیری از بـارووری در زیـر آمده است:

الف- نسبت به تکمیل جدول اقدام نمایید.

ب- نمودار مناسبی برای این داده ها رسم کنید.

نوع روش	فراوانـــــی ساده	فراوانی نسبی	فراوانی تجمعی
قرص	۱۱۷		
IUD	۲۶		
آمپول	۷		
کاندوم	۱۰۸		
بستن لوله	۱۱۵		
طبیعی	۲۱۶		
هیچ روش	۷۴		
نورپلنت	۱		

فصل چهارم :
توصیف عددی داده ها

هدف های فصل چهارم: با مطالعه این فصل شما باید بتوانید:
۱- شاخص های مرکزی را نام برده و هر یک را تعریف کنید.
۲- نسبت به محاسبه شاخص های مرکزی به صورت دستی یا کامپیوتری اقدام نمایید.
۳- شاخص های پراکندگی را نام برده و هر یک را تعریف کنید.
۴- ویژگی های هر یک از شاخص ها و استفاده مناسب از آنها را بیان کنید.
۵- خروجی کامپیوتری نرم افزار spss را در زمینه شاخص های توصیفی تفسیر کند.

۱-۴- شاخص های عددی

اگر چه جداول و نمودارها برای نمایش توزیع بکار می روند ممکن است در بعضی موارد چندان گویا نباشند لذا برای مشخص نمودن چگونگی توزیع متغیر در جامعه یا نمونه استفاده از اندازه های محدودی مناسب تر خواهد بود.
این اندازه ها برای نشان دادن مرکزیت توزیع یا چگونگی پراکندگی آن بکار می روند در این قسمت بعضی از مهمترین آنها مورد بحث قرار خواهند گرفت . اندازه هایی که وضعیت مرکزیت توزیع را مشخص می نمایند شاخصهای مرکزی و اندازه هائی که میزان پراکندگی توزیع را معین نمایند شاخصهای پراکندگی گویند.

۱-۴- الف : شاخصهای مرکزی

۱- میانگین (Mean)

میانگین برای تعدادی از داده ها عبارت است از خارج قسمت مجموع داده ها بر تعداد آنها ۰ همچنان این داده ها مربوط به نمونه باشد فرض نمائیم بصورت x_1وx_2و....x_n نشان داده شده باشند ، میانگین این داده ها را با $\bar{x}$ (x-bar)نشان داده و بصورت زیر محاسبه می گردد:

$$\bar{x}=\frac{1}{n}(x_1+x_2+...+x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i$$

که در آن $\sum_{i=1}^{n}$ (سیگما) علامت مجموع می باشد.

اگر بخواهیم کلیه اندازه های یک جامعه را با هم جمع نمائیم و بر تعداد اعضاء جامعه تقسیم نمائیم می توان آنرا با حروف یونانی μ نشان داده و میانگین جامعه بنامیم . غالباً در عمل میانگین نمونه ای از داده ها محاسبه و میانگین جامعه را بر اساس آن برآورد می نمایند.

مثال: وزن ۵ نفر از نوزادان درهنگام تولد بر حسب کیلو گرم بصورت ۴/۵ و ۲/۵ و ۳/۶ و ۴و ۲/۹ ثبت شده است میانگین وزن این نوزادان را محاسبه نمائید:

حل: در حالیکه داده ها در جدول توزیع فراوانی طبقه بندی شده باشند برای محاسبه میانگین از روش زیر استفاده می شود :

$$\bar{X} = \frac{4.5+2.5+3.6+4+2.9}{5} = \frac{17.5}{5} = ۳/۵ \text{ کیلو گرم}$$

بعنوان مثال اگر در یک مطالعه ای از یک متغیر کمی گسسته نتیجه مشاهدات بصورت طبقه بندی شده در جدول زیر آمده باشد :

f_i	x_i
f_1	x_1
f_2	x_2
.	.
.	.
f_k	X_K

X_i مقدار صفت برای گروه i (ام) و f_i فراوانی مربوط به این گروه می باشد : در آن صورت خواهیم داشت

$$\bar{x}=\frac{x_1 f_1+x_2 f_2+\ldots+x_k f_k}{f_1+f_2+\ldots+f_k}=\frac{\sum_1^k f_i x_i}{n}$$

مثال) میانگین تعداد فرزندان را در مورد داده های جدول (۲-۱) محاسبه نمائید :

$$\bar{x}=\frac{۰\times۱۵+۱\times۱۰+۲\times۱۳+۳\times۶+۴\times۳+۵\times۳}{۵۰}=\frac{۰+۱۰+۲۶+۱۸+۱۲+۱۵}{۵۰}=\frac{۸۱}{۵۰}$$
$$=۱/۶۲$$

در حالتیکه داده های طبقه بندی شده مربوط به یک متغیر کمی پیوسته باشد در آنصورت میانگین با تقریب امکان پذیر است بطوریکه برای هر طبقه وسط آن طبقـه را بعنـوان نماینـده آن طبقـه در نظر گرفته و در فراوانی آن ضرب می نمائیم . و مجموع همه آنها را برای کلیه طبقـات بدسـت مـی آوریم در آن صورت میانگین خارج قسمت مجموع محاسبه شده بر تعداد کل نمونه است.

مثال) میانگین مربوط به داه های طبقه بندی شده متغیر کمی پیوسـته در جـدول زیـر محاسـبه شده است.

طبقات	حدود واقعی	نماینـده طبقـات (x_i)	فراوانی (f_i)	$x_i f_i$
۱۰-۱۴	۹/۵-۱۴/۵	۱۲	۵	۶۰
۱۵-۱۹	۱۴/۵-۱۹/۵	۱۷	۱۰	۱۷۰
۲۰-۲۴	۱۹/۵-۲۴/۲	۲۲	۱۲	۲۶۴
۲۵-۲۹	۲۴/۵ -۲۹/۵	۲۷	۸	۲۱۶
۳۰-۳۴	۲۹/۵-۳۴/۵	۳۲	۵	۱۶۰
۳۵-۳۹	۳۴/۵-۳۹/۵	۳۷	۴	۱۴۸
جمع			۴۴	۱۰۱۸

$$\bar{x}=\frac{\sum x_i f_i}{n}=\frac{۱۰۱۸}{۴۴}=۲۳/۱۳$$

۲) میانه (median)

میانه شاخص مرکزی دیگری است که عبارت است از مقداری که برای نصف افراد مقدار متغیر از آن بزرگتر یا مساوی و برای نصف دیگر از آن کوچکتر یا مساوی است.

اگر داده ها را به ترتیب صعودی یا نزولی مرتب نمائیم آنگاه اندازه صفت را برای فرد وسط ($\frac{n+۱}{۲}$) بعنوان میانه انتخاب می کنیم چنانچه n فرد باشد عدد وسط مقدار مربوط به شماره $\frac{n+۱}{۲}$ خواهد بود و اگر n زوج باشد میانگین مقدار وسط که شماره های $\frac{n}{۲}$ و $\frac{n}{۲}+۱$ خواهند داشت میانه داده ها بحساب می آید.

مثال) میانه وزن نوزادان در مثال صفحه قبل را بدست آورید:

حل) اول وزنها را بصورت صعودی مرتب می کنیم :

۲/۵و۲/۹و۳/۶و۴و۴/۵

چون در اینجا $n = ۵$ پس $\frac{۵+۱}{۲}=۳$ لذا سومین مقدار یعنی ۳/۶ میانه خواهد بود و می نویسیم :

میانه (median) = ۳/۶

مثال) در سری ۵و۸و۴و۱۰و۶و۷و۱۲و۹ میانه را تعیین کنید ؟

پس از مرتب کردن داده ها : ۴،۵،۶،۷،۸،۹،۱۰،۱۲

چون تعداد نمونه زوج است و $n=۸$ لذا $\frac{n}{۲}=۴$ و $\frac{n}{۲}+۱=۵$ شماره های دو مقدار وسط را در داده ها معین می نماید لذا میانه برابر $\frac{۷+۸}{۲}=۷/۵$ خواهد بود.

محاسبه میانه در جدول توزیع فراوانی :

برای محاسبه میانه از جدول توزیع فراوانی بصورت زیر عمل می کنیم :

مرحله اول : ستون فراوانی تجمعی را تشکیل می دهیم.

مرحله دوم : مقدار $\frac{n}{۲}$ یعنی نصف حجم نمونه را بدست آورده و آنرا با مقادیر ستون فراوانی تجمعی مقایسه می کنیم اولین طبقه ای که فراوانی تجمعی آن از $\frac{n}{۲}$ بیشتر است را طبقه ای که میانه در آن است می نامیم .

مرحله سوم : در این مرحله میانه را از رابطه زیر محاسبه می نمائیم .

$$Median = L_i + \frac{\frac{n+1}{2} - F_{i-1}}{f_i} \times c$$

که در آن L_i حد پائینی طبقه میانه و F_{i-1} فراوانی تجمعی طبقه قبل از میانه و f_i فراوانی ساده طبقه میانه و c طول آن طبقه است.

مثال) میانه داده های جدول قبل را محاسبه نمائید ؟

طبقات	f_i	F_i
۹/۵-۱۴/۵	۵	۵
۱۴/۵-۱۹/۵	۱۰	۱۵
۱۹/۵-۲۴/۵	۱۲	۲۷
۲۴/۵-۲۹/۵	۸	۳۵
۲۹/۵-۳۴/۵	۵	۴۰
۳۴/۵-۳۹/۵	۴	۴۴
	۴۴	

$$\frac{n+۱}{۲} = \frac{۴۴+۱}{۲} = ۲۲/۵$$

$$Med = ۱۹/۵ + \frac{(\frac{۴۴+۱}{۲} - ۱۵)}{۱۲} \times ۵ = ۲۲/۲$$

۳- نما (Mode)

یکی دیگر از شاخصهای مرکزی نما (مد) است که مقداری از متغیر که بیشترین فراوانی را دارا می باشد نما نام دارد مثلاً در جدول فروانی زیر نما عدد ۳ می باشد :

تعداد فرزندان	تعداد خانوار
۰	۵
۱	۱۰
۲	۱۲
۳	۱۵
۴	۱۱

مثال : اگر مقادیر یک متغیر بصورت ۲، ۵ ، ۶ ، ۷، ۴، ۵، ۳، ۵ باشد مقدار نما عدد ۵ خواهد بود.
ممکن است یک سری از داده ها رو یا چند نما داشته باشد که همه آنها بیشتر از بقیه تکرار شده اند. نما ساده ترین نوع شاخص مرکزی است و بدلیل اعتبار کمتری که دارد معمولاً کمتر مورد استفاده قرار می گیرد.

ویژگیهای میانگین ، میانه، نما:

الف) ویژگیهای میانگین :

۱- یکتائی : برای هر مجموعه از داده ها تنها یک میانگین وجود دارد.

۲- در محاسبه میانگین همه داده ها نقش دارند به عبارت دیگر میانگین تحت تاثیر همه مقادیر داده هاست. در حالتهائی که بعضی از اندازه ها خیلی کوچک باشند تأثیر زیاد آنها بر میانگین ممکن است گمراه کننده باشد.

۳- میانگین شاخص مرکزی ساده با کاربردهای بیشتری برای متغیر های کمی با مقیاس نسبتی است .

ب) ویژگیهای میانه :

۱- یکتائی : میانه برای یک مجموعه داده ها یکتا است.

۲- محاسبه آن ساده است.

۳- میانه تحت تاثیر رتبه داده هاست و نشان دهنده موقعیت مکانی است که فرد وسط از نظر مقادیر آن متغیر قرار می گیرد.

۴- بطور جدی از اعداد خیلی بزرگ یا خیلی کوچک تأثیر نمی پذیرد.

۵- میانه برای متغیر های با مقیاس رتبه ای مناسبترین شاخص مرکزی است.

ج- ویژگیهای نما :

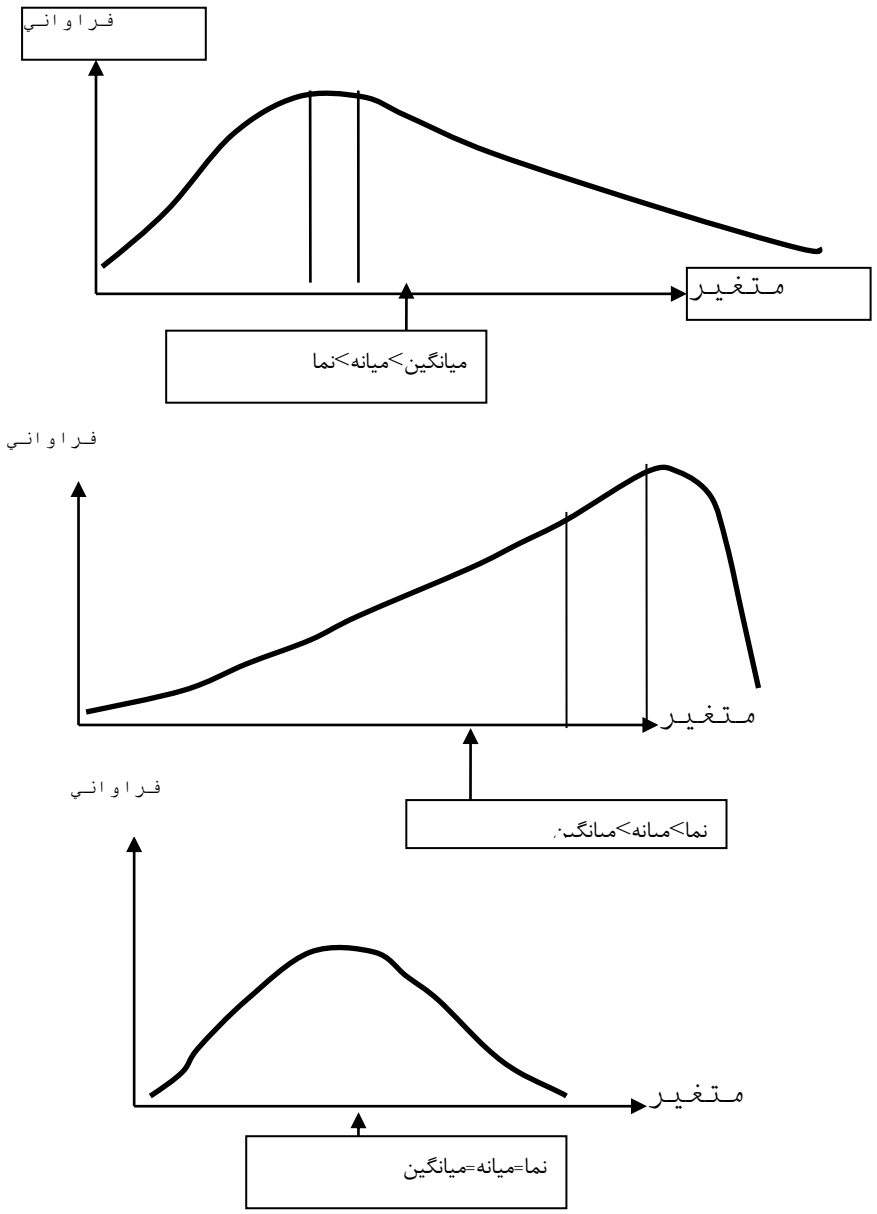
۱- نما بعلت سادگی و اهمیت کم آن مورد توجه زیاد نیست.

۲- نما در یک مجموعه از داده ها منحصر به فرد نیست. ممکن است داده ها هیچ مقدار نمایی نداشته باشند یا بیشتر از یک مقدار نمای آنها باشد.

۳- در نمایش داده های کیفی نما کاربرد بیشتری دارد.

مقایسه شاخصهای مرکزی :

اگر توزیع متقارن باشد و دارای یک قله در آنصورت هر ۳ ملاک گرایش به مرکز بر هم منطبق خواهد بود . و اگر توزیع نا متقارن باشد مثلاً بسمت راست یا چپ چاوله باشد این ملاکها با هم فرق دارند حالتهای فوق را در نمودارهای زیر می توان نشان داد.



شکل مقایسه میانگین، میانه و نما در توزیع های نامتقارن و متقارن

۱-۴- ب: شاخص های پراکندگی :

در بخش قبل شاخصهای مرکزی مورد بحث قرار گرفت هر چند توزیع های مختلف با شاخصهای مرکزی تا حدودی شناخته می شوند ولی تنها شاخصهای مرکزی نمی تواند برای مقایسه توزیعها یا تعیین چگونگی پراکندگی داده ها کافی باشد ممکن است دو جامعه از نظر شاخصهای مرکزی مانند میانگین یا میانه یکسانی داشته باشند ولی مقادیر دیگر متغیر در دو جامعه وضعیت هماهنگ یا مشابهی نداشته باشد.

بعنوان مثال ممکن است دو دانش آموز دارای میانگین یکسان در یکسال تحصیلی بوده ولی یکی از آنها دارای نمرات پراکنده ای شامل نمرات کمتر از ۱۰ و نیز نمرات بسیار بالائی باشد ولی دانش آموز دیگر دارای هیچ نمره کمتر از ۱۰ نبوده و همه نمرات او نزدیک به میانگین بوده باشد. برای نشان دادن پراکندگی نمرات از شاخصهای پراکندگی استفاده می شود.

۱- طول دامنه تغییرات :

ساده ترین شاخص پراکندگی طول میدان تغییرات است که با حرف R مشخص می گردد برای بدست آوردن آن اختلاف کمترین مقدار صفت را از بیشترین مقدار آن محاسبه نمود :

$R= X_{max} - X_{min}$

چون در محاسبه R فقط از دو مقدار حداکثر و حداقل داده ها استفاده می شود این شاخص تنها بیان کننده طول دامنه تغییرات داده ها و راجع به مقادیر داخل این فاصله گویائی لازم را ندارد.

۲- انحراف چارکی:

چارک اول، دوم و سوم یعنی صدک های ۲۵، ۵۰ و ۷۵ داده ها را به ۴ قسمت تقسیم می کند. تغییرات داده ها را می توان بر اساس انحراف بین چارک اول و سوم تا حدودی مشخص کرد. فاصله بین اولین و سومین چارک را دامنه انحراف چارکی می گویند.

۳- میانگین قدر مطلق انحرافات (Mean of Deviation (M.D.) :

اگر میانگین را از همه مقادیر صفت کم نموده اعداد حاصل بدلیل بزرگتر یا کوچکتر بودن از میانگین بصورت مثبت یا منفی خواهد بود و جمع جبری اختلاف اعدام از میانگین برابر صفر خواهد بود . اگر مجموع قدر مطلق اختلاف از میانگین را محاسبه و بر تعداد مشاهدات تقسیم نمائیم عبارت حاصل میانگین انحرافات نامیده می شود و بصورت زیر نشان داده می شود.

$$M.D. = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

مثال : میانگین انحرافات ۳ عدد ۱۵ ، ۱۰ ، ۱۱ که میانگین آنها برابر ۱۲ است برابر است با:

$$M.D. = \frac{|۱۵-۱۲| + |۱۰-۱۲| + |۱۱-۱۲|}{۳} = \frac{۳+۲+۱}{۳} = ۲$$

۴- واریانس (پراش) Variance:

اگر به جای استفاده از میانگین قدر مطلق انحرافات از میانگین مجذور انحرافات استفاده شود بعبارت دیگر هر یک از مقادیر رامنای میانگین نموده و اعداد حاصل را به توان دوم رسانده (نتایج بدست آمده مجذور انحرافات از میانگین گفته می شود) سپس مجموع مجذورات را بر تعداد اعضاء جامعه تقسیم نمائیم شاخص حاصل واریانس نام دارد.

این شاخص اشکال اعمال جبری بر روی قدر مطلقها را نداشته و پراکندگی داده هار با بخوبی نشان می دهد.

واریانس را برای N عدد بصورت زیر محاسبه و علامت δ^2 (سیگمادو) نشان می دهند.

$$\delta^2 = \frac{\sum_{1}^{N}(x_i - \mu)^2}{N}$$

با استفاده از روشهای معمول در ریاضیات می توان عبارت $\sum(x_i - \mu)^2$ را بصورت زیر نوشت :

$$\begin{aligned}\sum(x_i-\mu)^2 &= (\sum(x_i^2 - 2\mu x_i + \mu^2) = \sum x_i^2 - 2\mu\sum x_i + N\mu^2 \\ &= \sum(x_i^2 - 2\mu(N\mu) + N\mu^2 = \sum x_i^2 - 2N\mu^2 + N\mu^2 \\ &= \sum x_i^2 - N\mu^2\end{aligned}$$

و بدین ترتیب می توان واریانس را بصورت ساده تر از فرمول زیر حساب نمود:

$$\delta^2 = \frac{\sum x_i^2 - N\mu^2}{N}$$

اگر نتیجه مشاهدات بصورت گروه بندی شده باشد واریانس از فرمول زیر محاسبه می شود :

(K تعداد طبقات است)

$$\delta^2 = \sum_{1}^{k} \frac{f_i (x_i - \mu)^2}{N}$$

با استفاده از روش بالا عبارت حاضر بصورت زیر در می آید :

$$\delta^2 = \frac{\sum f_i x_i^2 - N\mu^2}{N}$$

چون در محاسبه واریانس اختلاف از میانگین بصورت مجذور آمده است لذا اختلافات بزرگتر از حد واقعی نشان داده شده است برای داشتن معیار پراکندگی می توان جذر واریانس را بکار برد که با آن انحراف معیار (استاندارد) می گویند.

و بصورت زیر نشان داده می شود :

$$\delta = \sqrt{\delta^2} = \sqrt{\frac{\sum f_i x_i^2 - N\mu^2}{N}} \quad \text{یا} \quad \sqrt{\frac{\sum f_i (x_i - \mu)^2}{N}}$$

تبصره :

چون در عمل بیشتر با نمونه انتخاب شده از جامعه مواجه خواهیم بود معمولاً بجای δ^2 و δ از برآوردهای آنها یعنی از واریانس و انحراف معیار نمونه که با S^2 و S نشان داده می شود استفاده می گردد:

در بخش استنباط آماری ثابت می شود که اگر در فرمول واریانس نمونه از n-۱ بجای n در مخرج کسر استفاده شود واریانس نمونه برآورد خوبی از واریانس جامعه خواهد بود لذا داریم :

$$s^2 = \frac{\sum_{1}^{n} (x_i - \bar{x})^2}{n-1}$$

نکته:

اگر یک نمونه n تایی از یک جمعیت انتخاب شود دارای یک مقدار ثابت میانگین و واریانس می باشد ولی از آن جایی که هر نمونه n تایی می تواند به صورتهای مختلفی از یک جامعه انتخاب شود به صورت نظری می توان گفت که واریانس نمونه متغیری است با توزیع t که W Gusset آن را معرفی نمود. چون در تغییرات واریانس نمونه ای به جای n در مخرج کسر از n-1 استفاده شود برآورد بهتری از واریانس جامعه ارائه می شود در حالتی که تعداد نمونه کم است از n-1 استفاده می شود که به آن درجه آزادی می گویند. برای توجیه بهتر درجه آزادی به این مثال توجه نمایید.

فرض کنید به شما بگویند ۵ عدد بیان کنید که مجموع آن ها ۱۰۰ باشد در آنصورت شما در انتخاب اعداد ۴ درجه آزادی دارید و عدد پنجم وابسته به ۴ عدد دیگر است و برابر تفاضل مجموع ۴ عدد از ۱۰۰ می باشد. به عبارت دیگر در انتخاب ۵ عدد شما ۴ درجه آزادی دارید. چون تغییرات n-1 نمونه در واریانس نمونه ای موثر است می توان گفت که S^2 دارای n-1 درجه آزادی است.

واریانس نمونه در حالتی که اطلاعات طبقه بندی شده است : یا

$$\begin{cases} s^2 = \dfrac{\sum_{1}^{k}(x_i - \bar{x})^2 f_i}{n-1} \\ s^2 = \dfrac{\sum_{1}^{k} f_i x_i^2 - n\bar{x}^2}{n-1} \end{cases}$$

انحراف استاندارد(معیار) نمونه :

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad \text{یا} \quad s = \sqrt{\frac{\sum x_i^2 f_i - n\bar{x}^2}{n-1}}$$

مثال- واریانس و انحراف استاندارد اعداد ۲،۵،۴،۱ را بدست آورید :

$$\sigma^2 = \sum \frac{(x_i - \mu)^2}{N}$$

$$\sigma^2 = \frac{(2-3)^2 + (5-3)^2 + (4-3)^2 + (1-3)^2}{4} = \frac{10}{4} = 2.5$$

$$\Rightarrow \sigma = \sqrt{2.5} = 1.58$$

مثال) واریانس و انحراف استاندارد جدول توزیع فراوانی زیر را محاسبه نمائید :

حدود طبقات	f_i	x_i وسط طبقات	$x_i f_i$	$X_i^2 f_i$
۵/۹-۱۴/۵	۵	۱۲	۶۰	۷۲۰
۱۴/۵-۱۹/۵	۱۰	۱۷	۱۷۰	۲۸۹۰
۱۹/۵-۲۴/۵	۱۲	۲۲	۲۶۴	۵۸۰۸
۲۴/۵-۲۹/۵	۸	۲۷	۲۱۶	۵۸۳۲
۲۹/۵-۳۴/۵	۵	۳۲	۱۶۰	۵۱۲۰
۳۴/۵-۳۹/۵	۴	۳۷	۱۴۸	۵۴۷۶
جمع	۴۴		۱۰۱۸	۲۵۸۴۶

$$\mu = \frac{\sum x_i f_i}{n} = \frac{1018}{44} = 23.1$$

$$\delta^2 = \frac{\sum x_i^2 f_i - N\mu^2}{N} = \frac{25846 - 44(23.14)^2}{44}$$

$$= \frac{25846 - 23560}{44} = 51.9$$

$$\delta = \sqrt{51.9} = 7.2$$

۲-۴- تأثیر تغییرات یکنواخت در مشاهدات:

الف) اگر x متغیری با مقادیر x_1 و x_2 و ... x_n به صورت متغیر جدیدی با افزایش مقدار ثابت c به هر یک از مقادیر بصورت متغیر 'x با مقادیر زیر تبدیل یابد :

$X_1+C, X_2+C, \ldots, X_n+C$

در آن صورت میانگین و واریانس متغیر حاصل را مورد بحث قرار می دهیم :

$$\mu_{x'} = \frac{(x_1+c)+(x_2+c)+\ldots(x_n+c)}{n} = \frac{\sum x_i + nc}{n} = \mu_x + c$$

که در آن μ_x میانگین متغیر x و $\mu_{x'}$ میانگین متغیر 'xاست

$$\sigma_{x'}^2 = \frac{\sum [(x_i+c)-(\mu+c)]^2}{N} = \frac{\sum (x_i-\mu)^2}{N} = \sigma_x^2$$

که در آن $\delta_{x'}^2$ واریانس 'x و δ_x^2 واریانس x است بنابر این :

نتیجه ۱ : هر گاه داده ها را با عدد ثابتی جمع (یا کم) کنیم میانگین بهمان اندازه زیاد(یا کم)می شود. ولی درواریانس تغییری حاصل نمیگردد.

ب) اگر کلیه مقادیر x در عدد ثابتی مانند k ضرب گردد متغیر جدید'x دارای مقادیر زیر خواهد بود.

x':kx ۱ و kx۲ و... و Kx$_n$

لذا میانگین و واریانس آن که با $\mu_{x'}$ و $\delta^2_{x'}$ نشان داده می شود بصورت زیر در می آید :

$$\mu_{x'} = \frac{kx_1 + kx_2 + ..kx_n}{N} = \frac{k(x_1 + x_2 + ... + x_n)}{N} = \frac{k\sum x_i}{N} = k\mu_x$$

$$\delta^2_{x'} = \frac{\sum (kx_1 - k\mu)^2}{N} = \frac{\sum k^2 (x_i - \mu)^2}{N} = k^2 \delta^2_x$$

نتیجه۲ : هر گاه داده ها را در عدد ثابتی ضرب (یا تقسیم) کنیم میانگین و انحراف معیار به همان نسبت بزرگ (یا کوچک) می شود ولی واریانس به نسبت مجذور عدد ثابت بزرگ (یا کوچک) می شود.

محاسبه میانگین و واریانس از روش کوتاه :

با استفاده از تغییرات فوق می توانیم روش کوتاه تری در محاسبه میانگین و واریانس ارائه نمائیم اگر مقداری از x_i که معمولاً بیشترین فراوانی رادارد x_0 بنامیم و از کلیه x_i ها کم نموده و حاصل را بر فاصله طبقه (h) تقسیم نمائیم متغیر جدید 'x بدست خواهد آمد :

$$x'_i = \frac{x_i - x_0}{h}$$

پس از اینکه میانگین و واریانس متغیر جدید محاسبه گردید اگر میانگین 'x را در فاصله گروه (h) ضرب و با x_0 جمع نمائیم میانگین متغیر اصلی (x) محاسبه می شود:

اگر واریانس 'x را در مربع فاصله گروه یعنی h^2 ضرب نمائیم واریانس x محاسبه می شود چون:

$$x_i = x'_i.h + x_0 \Rightarrow \begin{cases} \mu_x = \mu_x'.h + x_0 \\ \delta^2_x = \delta^2_{x'}.h^2 \end{cases}$$

مثال) با استفاده از روش کوتاه میانگین و واریانس جدول فراوانی مثال قبل را محسابه نمائید :

حدود طبقات	f_i	وسط طبقات	$x'i=\frac{x_i-22}{5}$	$X'_i f_i$	$X'^2_i f_i$
۵/۹-۱۴/۵	۵	۱۲	-۲	-۱۰	۲۰
۱۴/۵-۱۹/۵	۱۰	۱۷	-۱	-۱۰	۱۰
۱۹/۵-۲۴/۵	۱۲	۲۲	۰	۰	۰
۲۴/۵-۲۹/۵	۸	۲۷	۱	۸	۸
۲۹/۵-۳۴/۵	۵	۳۲	۲	۱۰	۲۰
۳۴/۵-۳۹/۵	۴	۳۷	۳	۱۲	۳۶
جمع	۴۴			۱۰	۹۴

$$\mu x'=\frac{\sum x'_i \cdot f_i}{N}=\frac{10}{44}=0.227 \Rightarrow \mu x=0.227\times 5+22=23.1$$

$$\sigma^2_{x'}=\frac{94-44(0.227)^2}{44}=\frac{94-2.267}{44}=2.08$$

$$\Rightarrow \sigma^2_x=2.08\times 5^2=52.1$$

$$\sigma=\sqrt{52.1}=7.21$$

۳-۴- ضریب تغییرات (coefficient of variation) :

شاخص های میانگین و انحراف معیار از نوع خود صفت بوده و واحد اندازه گیری صفت را دارا هستند مثلاً اگر متغیر مورد بررسی وزن و واحد اندازه گیری آن کیلو گرم باشد نسبت انحراف معیار به میانگین را برای متغیری خاص بدست آورده و آنرا بر حسب درصد بیان می نمائیم اثر واحد اندازه گیری در نسبت حاصل حذف شده و کمیت حاصل یک مشخص کننده نسبی است که به آن *ضریب تغییرات* گفته و با C.V. نشان می دهند.

$$C.V.=\frac{\delta}{\mu}\times ۱۰۰$$

این شاخص برای مقایسه تغییرات دو صفت یا یک صفت با دو واحد مختلف بکار می رود. در مطالعات پزشکی برای بررسی و مقایسه تغییرات آنزیمها، پارامترهای فیزیولوژیک بدن، آزمایشات بیوشیمیایی که واحد های اندازه گیری متفاوتی دارند مورد استفاده قرار می گیرد.

مثال- اگر میانگین و انحراف معیار وزن بدن دانش آموزان یک کلاس به ترتیب ۴۰ و ۲ کیلو گرم باشد و نیز میانگین و انحراف معیار نمرات آنها در درس ریاضی برابر ۱۵ و ۱/۵ واحد باشد تغییرات این دو صفت را با هم مقایسه نمائید :

$$CV\ (\text{وزن}) = \frac{۲}{۴۰} \times ۱۰۰ = ۵$$

$$CV(\text{نمرات}) = \frac{1.5}{15} \times ۱۰۰ = ۱۰$$

بنابر این تغییرات نمرات ریاضی دانش آموزان ۲ برابر تغییرات وزن آنهاست اگر چه بدون محاسبه ضریب تغییرات هر دو شاخص میانگین و انحراف معیار وزن بیشتر از نمرات بوده است.

نمونه سؤالات فصل چهارم:

۱- کدام یک از موارد زیر از ویژگیهای میانگین نیست؟

الف- میانگین یکی از شاخصهای مرکزی است.

ب- میانگین در یک مجموعه از داده ها منحصر به فرد (یکتا) است.

ج- میانگین تحت تاثیر همهٔ مقادیر داده هاست.

د- میانگین همواره بیشتر یا مساوی ۵۰ درصد داده هاست.

۲- اگر همه مقادیر x برابر ۱۰ باشد میانگین وانحراف معیار به ترتیب چقدر است؟

الف- یک وصفر ب- ۱۰ وصفر

ج- ۱۰ ویک د- ۱۰ وبینهایت

۳- اگر میانگین وانحراف معیار x به ترتیب برابر ۵ و۱ باشد، میانگین وانحراف معیار $2x-1$ برابر است با ...؟

الف- ۹ و۱ ب- ۹ و۲

ج- ۵ و۱ د- ۵ و ۲

۴- در یک توزیع نامتقارن استفاده از کدام شاخص مرکزی موجه تر است؟

الف- نما ب- میانگین

ج- میانه د- هیچ تفاوتی ندارد.

۵- برای متغییر های کمی پیوسته کدام نوع نمودار مناسبتر است؟

الف- هیستوگرام ب- چند بر

ج- ستونی د- دایره ای

۶- برای مقایسه تغییرات یک متغیر با دو واحد اندازه گیری متفاوت از .....استفاده می شود.

الف- دامنه تغییرات ب- واریانس

ج- میانگین د- ضریب تغییرات

۷- واریانس چیست؟

الف- مجذور متوسط انحرافات از میانگین

ب- متوسط مجذور انحرافات از میانگین

ج- متوسط جذر انحرافات از میانگین

د- مجموع مجذور انحرافات از میانگین

۸- انحراف معیار یک صفت همواره چگونه است؟

الف- بزرگتر از واریانس ب- کو چکتر از واریانس

ج- مثبت د- نامنفی

۹- در یک بررسی روی DMF نتایج زیر بدست آمده است.

DMF	۰	۱	۲	۳	۴
تعداد	۱	۱۱	۱۳	۷	۳

میانه، نما و میانگین به ترتیب برابرند با :

الف- ۱۳،۱۳،۱۳ ب- ۲ ، ۲، ۲

ج- ۱۲، ۱۳، ۲ د- ۲،۱۳،۲

۱۰- اگر از کلیه داده ها، میانگین آنها را کم کنیم، انحراف معیار داده ها همیشه:

الف- کوچکتر می شود ب- بزرگتر می شود

ج- برابر صفر می شود د- تغییری نمی کند

۱۱- میانگین صفتی برابر ۲۵ و ضریب تغییرات برابر ۲۰ درصد است. واریانس صفت برابر است با:

الف- ۲۵ ب- ۵

ج- ۱/۵ د- ۱/۲۵

۱۲- میانگین تعداد فرزندان ۳۰ خانوار ۲ و میانگین تعداد فرزندان ۲۰ خانوار ۳ فرزند است میانگین تعداد فرزندان همه این خانوارها چقدر است؟

الف- ۲/۵ ب- ۲

ج- ۳ د- ۲/۴

۱۳- هرگاه داده ها را با عدد ثابتی جمع یا کم کنیم، میانگین آنها :

الف- به همان اندازه زیاد یا کم می شود

ب- به همان اندازه کم یا زیاد می شود

ج- به اندازه مربع آن عدد زیاد یا کم می شود

د- تغییری نمی کند

۱۴- هرگاه داده ها را در عدد ثابتی ضرب کنیم، میانگین آنها :

الف- به نسبت مربع آن اندازه زیاد می شود

ب- به نسبت مربع آن اندازه کم می شود

ج- به نسبت اندازه آن عدد زیاد می شود

د- تغییری نمی کند

مسائل:

۱- اطلاعات مربوط به تعداد فرزندان خانوار های یک روستا در جدول زیر آمده است:

الف- میانگین ومیانه و نما را تعیین کنید.

ب- واریانس و انحراف معیار را محاسبه کنید.

ج- دامنه تغییرات و ضریب تغییرات را محاسبه کنید.

تعداد فرزند	فراوانی
۰	۴
۱	۶
۲	۷
۳	۵
۴	۳
جمع	۲۵

۲- اگر میانگین و انحراف معیار X به ترتیب برابر μ و σ باشد میانگین و انحراف معیار $Z=\frac{X-\mu}{\sigma}$ چقدر است؟

۳- توزیع فراوانی زنان بر حسب تحصیلات و میانگین تعداد فرزندان و انحراف معیار هر گروه در جدول زیر آمده است. میانگین و انحراف معیار کل زنان را محاسبه نمایید.

تحصیلات زن	فراوانی	میانگین تعداد فرزند	انحراف معیار
بی سواد	**۲۵**	**۳/۶**	**۱/۸**
ابتدایی	**۱۴۳**	**۲/۵**	**۱/۳**
راهنمایی	**۱۴۱**	**۲**	**۱**
دیپلم	**۲۳۱**	**۱/۵**	**۱**
دانشگاهی	**۱۳۰**	**۲/۶**	**۱/۶**

فصل پنجم :
احتمال و توزیع های آن

هدفهای فصل پنجم: پس از مطالعه این فصل شما باید قادر باشید:

۱- احتمال را تعریف کنید.

۲- دو تا از قوانین مهم احتمال را نام ببرید.

۳- توزیع احتمال دوجمله ای را تعریف و محاسبات لازم را انجام دهید.

۴- توزیع نرمال را تعریف و خواص آن را بیان کنید.

۵- نمونه هایی از کاربرد های توزیع ها را در علوم بهداشتی نام ببرید.

۵-۱- احتمال:

پدیده های زیادی در طبیعت وجود دارند که به پدیده های تصادفی موسومند احتمالات در مورد شانس وقوع این پدیده ها بحث می کند بطوریکه اندازه این شانس وقوع را احتمال وقوع آن پیش آمد می گویند.

فضای نمونه ای : مجموعه کلیه نتایج ممکن یک آزمایش تصادفی را فضای نمونه ای آن آزمایش گویند.

پیشامد: هر زیر مجموعه از یک فضای نمونه ای را یک پیشامد گویند.

مثال : اگر سکه ای ۲ بار پرتاب گردد فضای نمونه ای بصورت زیر خواهد بود ، که معمولاً با s نشان داده می شود :

s={ (خط ،خط) ،(شیر ، خط) ،خط ،شیر)، (شیر ،شیر) }

مجموعه A= {(شیر،شیر)} یک پیشامد است که وقوع هردو بار شیر آمدن رویه باللائی سکه را نشان می دهد.

تعریف احتمال :

احتمال یک حادثه که اندازه شانس وقوع آن است غالباً با p نشان داده می شود . اگر A یک حادثه (یا پیشامد) باشد همواره داریم :

$$0 \le P(A) \quad , \quad P(A) \le 1$$

اگر نتیجه آزمایش بتواند به N صورت همتراز و ناسازگار رخ دهد بعبارت دیگر شانس وقوع همه صورتها مساوی و وقوع توام آنها امکان پذیر نباشد و از این N صورت M صورت آن برای وقوع حادثه معین A مساعد باشد ،گوئیم احتمال حادثه A یک کسر متعارفی بصورت $\frac{M}{N}$ است.

یعنی : $P(A) = \frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات ممکن}} = \frac{M}{N}$

مثال : در پرتاب دو سکه در مثال قبل احتمال اینکه هر دو بار شیر ظاهر شود برابر است با:

$$P(\text{شیر ،شیر}) = P(A) = \frac{۱}{۴}$$

مثال : در کیسه ای ۳ مهره آبی و ۵ مهره سیاه وجود دارد اگر به تصادف ۲ مهره ازکیسه خارج شود احتمال اینکه :

الف) هر دو مهره آبی باشد چقدر است ؟ (در صورتیکه مهره دوم بدون جایگذاری مهره اول خارج شده باشد ؟ اگر حادثه مورد نظر ظاهر شدن هر دو مهره آبی باشد، آنگاه :

$$P(A) = \frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات ممکن}} = \frac{۳\times۲}{۸\times۷} = \frac{۶}{۵۶} = \frac{۳}{۲۸}$$

بیان دیگر عبارت فوق استفاده از ترکیب $\binom{8}{2}$ بعنوان تعداد حالات ممکن و ترکیب $\binom{3}{2}$ بعنوان تعداد حالات مساعد می باشد و لذا :

$$P(A) = \frac{\binom{3}{2}}{\binom{8}{2}} = \frac{\frac{۳!}{۲!۱!}}{\frac{8!}{2!6!}} = \frac{۳\times۲}{۵۶} = \frac{۶}{۵۶} = \frac{۳}{۲۸}$$

ب) هر دو مهره آبی باشد چقدر است ؟ در صورتیکه پس از خارج نمودن مهره اول آنرا به کسیه برگردانیم و مهره دوم راخارج نمائیم (با جایگذاری).

$$P(A) = \frac{۳\times۳}{۸\times۸} = \frac{۹}{۶۴}$$

قوانین احتمال:

۲-۵- احتمال حاصل جمع:

در تعبیر مجموعه ایاگر Aو B دو پیش آمد باشند بطوریکه $A\cap B=\phi$ در آنصورت A و B را دو *پیشامد ناسازگار* گویند.

احتمال اجتماع دو پیشامد Aو B وقتی که A و B ناسازگار هستند بصورت زیر خواهد بود:

$P(A\cup B)=P(A)+P(B)$ که $A\cap B$ به معنی آن است که A یا B رخ دهد.

در حالت کلی برای هر دو پیشامد A و B خواهیم داشت :

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

که در آن $A\cap B$ پیشامد توام A و B می باشد.

مثال : اگر یک تاس را دوبار بریزیم احتمال اینکه حداقل یک بار روی ۵ ظاهر شود را حساب نمائید :

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=\frac{۱}{۶}+\frac{۱}{۶}-\frac{۱}{۳۶}=\frac{۱۱}{۳۶}$$

تذکر : روابط زیر برای هر پیشامد A و پیش آمد عدم وقوع آن یعنی $\bar{A}$ برقرار است :

$A\cup\bar{A}=S$ (فضای نمونه ای) $A\cap\bar{A}=\phi$

$P(A)+P(\bar{A})=۱$ $P(\bar{A})=1-P(A)$

۳-۵- احتمال حاصل ضرب :

دو حادثه A و B را مستقل می گویند اگر وقوع یا عدم وقوع یکی در وقوع یا عدم وقوع دیگری تأثیر نداشته باشند : اگر A و B مستقل باشند در آنصورت :

$$P(A\cap B)=P(A).P(B)$$

در حالت کلی برای هر دو پیش آمد A و B :

$$P(A\cap B)=P(A).P(B|A)$$

که در آن $P(B|A)$ احتمال وقوع B به شرطی که A اتفاق یافته باشد بوده و به احتمال مشروط B به شرط A موسوم است بعبارت دیگر :

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

مثال : بفرض اینکه پسر یا دختر بودن یک فرزند در احتمال پسر یا دختر بودن فرزند بعدی تأثیر نداشته باشد و در هر بار احتمال پسر بودن $\frac{۱}{۲}$ باشد می خواهیم این احتمال را که ۴ فرزند متوالی یک خانواده پسر باشد محاسبه کنیم:

اگر A۱و A۲وA۳و A۴ به ترتیب حادثه پسر بودن فرزند اول و دوم و سوم و چهارم باشد در آنصورت :

$P(A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

۴-۵- توزیع احتمال :

توزیع احتمال عبارت است از یک مدل ریاضی برای توزیع فراوانی واقعی این مدلها ممکن است برای یک متغیر گسسته بکار روند یا برای یک متغیر پیوسته ۰ اگر تابعی با مقادیر عددی از پیشامدهای فلضای نمونه ای را در نظر بگیریم در حقیقت با متغیر تصادفی سروکار خواهیم داشت. دراینجا هدف بررسی توزیعهای مختلف احتمال نیست و تنها صورت مختصر توزیع احتمال دو جمله ای و نرمال مورد بحث قرار می گیرند.

۵-۵- توزیع دو جمله ای :

آزمایشی را در نظر بگیرید که نتایج ممکن آن را بتوان بدو صورت و قوع یا عدم وقوع A دسته بندی کرد. اگر نتیجه به وقوع A ختم شود گوئیم یک پیروزی (موفقیت) و در غیر اینصورت یک شکست رخ داده است یک متغیر تصادفی که تعداد پیروزیها یعنی تعداد دفعات وقوعA در N آزمایش نشان می دهد متغیر دو جمله ای Binomial نامیده می شود.

اگر آزمایشی که خاصیت فوق را داراست دوبار تکرار شود حالتهای زیر و احتمال مربوط با آنها کلیه حالات ممکن خواهد بود اگر وقوع حادثه A موفقیت و احتمال آن برابر p باشد آنگاه :

P (هر دو موفقیت) $= P(A).P(A)=P.P=P^2$

P(بار اول موفقیت و بار دوم شکست) $= P(A).P(\bar{A})=P.q=P(1-P)$

P(بار اول شکست و بار دوم موفقیت) $= p(\bar{A}).P(A)=q.P=(1-P).P$

P (هر دو بار شکست باشد) $=p(\bar{A}).p(\bar{A})=q.q=q^2$

بنابر این احتمال صفر موفقیت برابر q^2 و احتمال یک موفقیت برابر$2pq$ و احتمال دو موفقیت برابر خواهد بود ، که کلیه حالات ممکن در عبارت زیر مورد توجه قرار می گیرد.

$$(p+q)^2=p^2+2pq+q^2$$

اگر این آزمایش ۳ بار تکرار شود کلیه حالتها و احتمالات مربوط به آنها بصورت زیر خواهد بود :

$P(AAA)=P^3$ $\quad\quad$ $P(\bar{A}\,\bar{A}A)=Pq^2$

$P(AA\,\bar{A}=p^2q$ $\quad\quad$ $P(A\,\bar{A}\,\bar{A}=Pq^2$

$P(A\,\bar{A}A) =p^2q$ $\quad\quad$ $P(\bar{A}A\,\bar{A})=Pq^2$

$P(\bar{A}AA)= p^2q$ $\quad\quad$ $P(\bar{A}\,\bar{A}\,\bar{A})=q^3$

در آنصورت کلیه حالتها در عبارت زیر مورد بررسی قرار می گیرد :

$$(P+q)^3=P^3 + 3p^2q +3pq^2+q^3$$

اگر آزمایش از نوع فوق n بار تکرار شود و احتمال وقوع r بار موفقیت مورد توجه باشد در آنصورت تعداد حالتهائی که r از n پیروزی اتفاق افتد برابر ترکیب ($\frac{n}{r}$)است که بصورت C_n^r نشان داده می شود این ترکیب عبارت است از :

$$C_n^r = \frac{n!}{r!(n-r)!} = \frac{n\times(n-1)(n-2)\ldots(n-r)!}{r!(n-r)!}$$

عبارت !nکه n فاکتوریل خوانده می شود عبارت است از حاصلضرب اعداد از ۱ تا n و نیز صفر فاکتوریل و یک فاکتوریل برابر یک تعریف می شود.

بنابر این اگر در حالت کلی n بار آزمایش بصورت مستقل تکرار شود احتمال r موفقیت بصورت زیر خواهد بود.

$$(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1\times2}p^{n-2}q^2 + \frac{n(n-1)(n-2)}{1\times2\times3}p^{n-3}q + \ldots$$
$$+q^n = \sum_{r=0}^{n}\binom{n}{r}p^r q^{n-r}$$

به توزیعی که از عبارت فوق تبعیت کند توزیع *دو جمله ای* گفته می شود.

مثال : اگر احتمال آمدن روی ۵ در یک تاس پس از پرتاب کردن موفقیت بحساب آید و احتمال آمدن روی ۵ برابر $\frac{۱}{۶}$ باشد حساب کنید که اگر تا سه بار پرتاب شود احتمال اینکه ۲ بار عدد ۵ ظاهر شود چقدر خواهد بود ؟

$$p(\text{۲ بار ۵ ظاهر شود}) = p^{۲}\, q^{۳-۲} = \left(\frac{۳}{۲}\right)\left(\frac{۱}{۶}\right)^{۲}\left(\frac{۵}{۶}\right)^{۱} = ۳۰\times\frac{۱}{۳۶}\times\frac{۵}{۶}$$

چون : $\frac{3!}{2!1!}=\frac{3\times2\times1}{2\times1\times1}=3$

$P(\text{۲ بار ۵ ظاهر شود})=\frac{۵}{۷۲}$

مثال) اگر احتمال پسر یا دختر بودن نوزاد $\frac{۱}{۲}$ باشد احتمال های زیر را برای یک خانواده ۳ فرزندی محسابه نمائید؟

الف) احتمال اینکه تها یک فرزند این خانواده پسر باشد.

ب) احتمال اینکه حداکثر ۲ فرزند این خانواه پسر باشد.

حل الف) $\binom{۳}{۱}p^{۱}q^{۳-۱}=۳\times\frac{۱}{۲}\times(\frac{۱}{۲})^{۲}=\frac{۳}{۸}$

حل ب) $p(\text{حداکثر۲ پسر})=۱-p(\text{هر سه پسر})=۱-p^{۳}=۱-(\frac{۱}{۲})^{۳}=\frac{۷}{۸}$

۶-۵- توزیع پوآسون

محدودیت های توزیع دوجمله ای:

توزیع دو جمله ای در مواردی که n بسیار بزرگ و p (یا q) خیلی کوچک (یا بزرگ) باشند محاسبه ی احتمالات بر اساس توزیع دو جمله ای با محدودیت هایی مواجه است لذا توزیع پوآسن با تقریب بهتری احتمال وقوع پیشامدها را مشخص می کند.

پوآسن ابتدا این توزیع را در زمینه تعیین تصادفی بودن برخورد موشک ها در نواحی مختلف شهر به کار برد که بعد ها در زمینه کلیه فرآیند هایی که در واحد سطح، واحد طول(زمان) و یا واحد حجم به صورت تصادفی واقع می شوند مورد استفاده قرار گرفت.

در همه ی این موارد λ ضریب پوآسن عبارت است از متوسط تعداد وقایع در واحد سطح، طول یا حجم که پارامتر توزیع پوآسن نیز نامیده می شود.

فرمول توزیع پوآسن به صورت زیر است:

$$p(x)=\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x=0,1,2,\ldots,n$$

که در آن e مبنای لگاریتم طبیعی (نپری) و تقریبا برابر با ۲/۷۱۸ است.
λ ضریب پوآسن که به صورت زیر برآورد می شود.

$$\lambda = n.p$$

که در آن n تعداد آزمایش و p احتمال موفقیت در هر آزمایش می باشد که هردو مقداری مثبت هستند بنابر این ضریب پواسون همواره غیر منفی است.
احتمال x موفقیت در n آزمایش را با $P(x)$ نشان می دهند.

میانگین و واریانس توزیع پوآسن:

یکی از ویژگی های مهم این توزیع آن است که میانگین و واریانس آن با هم برابر و مقدار پارامتر توزیع یعنی λ است.

$$\delta^2 = \mu = n.p = \lambda$$

مثال:

فرض کنید احتمال وقوع یک حادثه رانندگی (تصادف) برای هر اتومبیل در یک بزرگراه برابر ۰/۰۰۱ باشد. اگر تعداد ۲۰۰۰ اتومبیل در این مسیر تردد نمایند مطلوب است:
الف- احتمال اینکه ۳ تا از اتومبیل ها دچار حادثه شوند.
ب- احتمال اینکه حداکثر ۲ تصادف واقع شود.

حل: الف- ابتدا مقدار ضریب پواسون محاسبه می شود.

چون $\lambda = n.p$ پس : $\lambda = ۰/۰۰۱ * ۲۰۰۰ = ۲$

$$p(x) = \frac{e^{-2} 2^3}{3!} = \frac{8}{(2.718)^2 * 6}$$

$$= \frac{8}{44.325} = 0.1805$$

ب- احتمال وقوع حداکثر ۲ تصادف برابر است با جمع احتمالات مربوط به صفر، ۱ و ۲ تصادف که با قرار دادن $x=0, 1, 2$ در فرمول توزیع پواسون و جمع آنها به دست می آید.

$$\sum_{x=0}^{2} p(x) = \frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}$$

$$= \frac{1}{(2.718)^2 * 1} + \frac{2}{(2.718)^2 * 1} + \frac{4}{(2.718)^2 * 2}$$

$$= 0.14 + 0.27 + 0.54 = 0.95$$

۷-۵- توزیع نرمال

توزیع نرمال دارای فرمول ریاضی زیر می باشد :

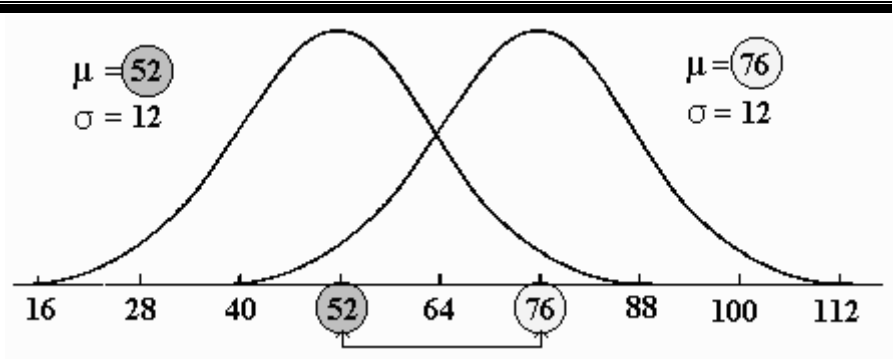
$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\delta^2}}$$

که اعداد πو e به ترتیب ۳/۱۴۱۵۹و۲/۷۱۸۲۸می باشد.

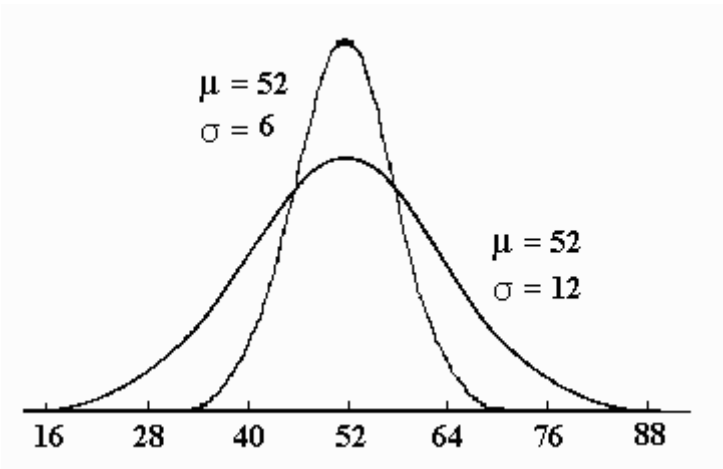
پارامترهای توزیع نرمال عبارتند از: μ میانگین و δ انحراف معیار

با توجه به اینکه برای متغیرهای مختلف میانگین و واریانس های مختلفی وجود دارد پس منحنی های نرمال می تواند به یک شکل کلی ولی با میانگین و واریانسهای متفاوتی رسم شود. به عبارت دیگر به ازای هر مقدار میانگین در هر متغیر مرکز توزیع مشخص می شود وبه ازای هر مقدار واریانس، ارتفاع توزیع تعیین می شود. به این دلیل می گویند خانواده ای از توزیع های نرمال خواهیم داشت.

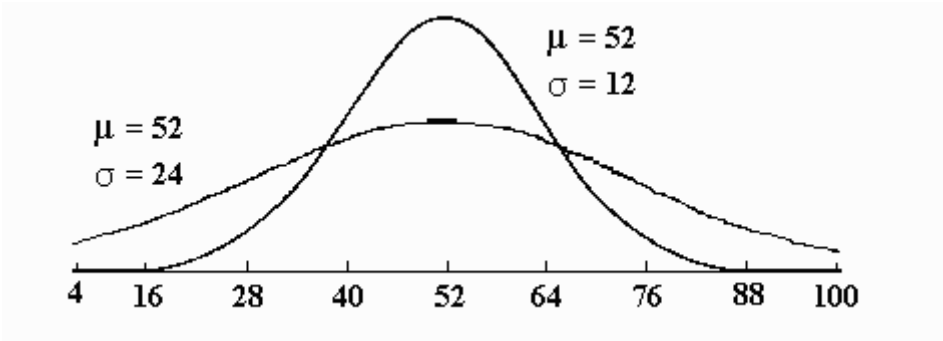
در زیر دو نمونه از توزیع های نرمال به ترتیب با میانگین های ۵۲ و۷۶ و انحراف معیارهای مساوی ۱۲ نشان داده می شود.



ودر زیر دو نمونه از توزیع های نرمال به ترتیب با میانگین های مساوی ۵۲ ولی با انحراف معیارهای مختلف به ترتیب ۶ و۱۲ نشان داده می شود.



اما در زیر دو نمونه از توزیع های نرمال به ترتیب با میانگین های مساوی ۵۲ ولی با انحراف معیارهای مختلف به ترتیب ۱۲و۲۴ نشان داده می شود.



از مقایسه توزیع های فوق مشخص می شود که هر چقدر واریانس توزیع کمتر شود ارتفاع آن بیشتر می شود وهر چقدر واریانس آن بیشتر شود ارتفاع توزیع کمتر می شود. این امر به دلیل این است که سطح زیر منحنی یک توزیع احتمال باید همواره برابر یک واحد مربع ثابت باشد.

خصوصیات عمده توزیع نرمال:

۱- توزیع نرمال زنگوله ای شکل وحول میانگین μ متقارن است.

۲- در توزیع نرمال میانگین ،میانه و نما همگی بر هم منطبق و با هم برابرند.

۳- سطح کل زیر منحنی نرمال برابر یک واحد مربع است.

۴- احتمال فاصله ای که باندازه

یک انحراف معیار در هر طرف میانگین امتداد دارد برابر است با :

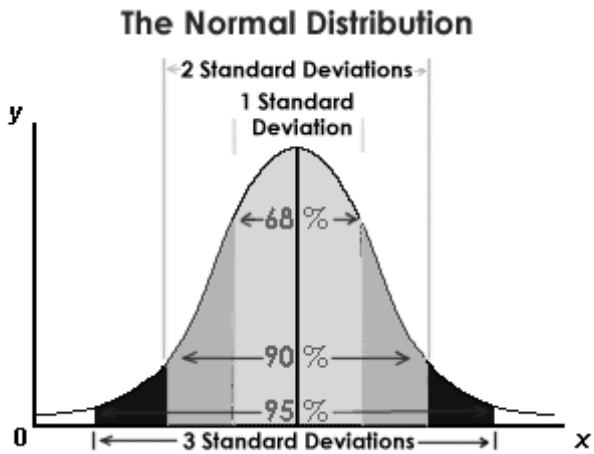
$$p[\mu-\delta\langle x\langle\mu+\delta]=۰/۶۸$$

ودو انحراف معیار در هر طرف میانگین امتداد دارد برابر است با :

$$p[\mu-2\delta\langle x\langle\mu+2\delta]=۰/۹۵$$

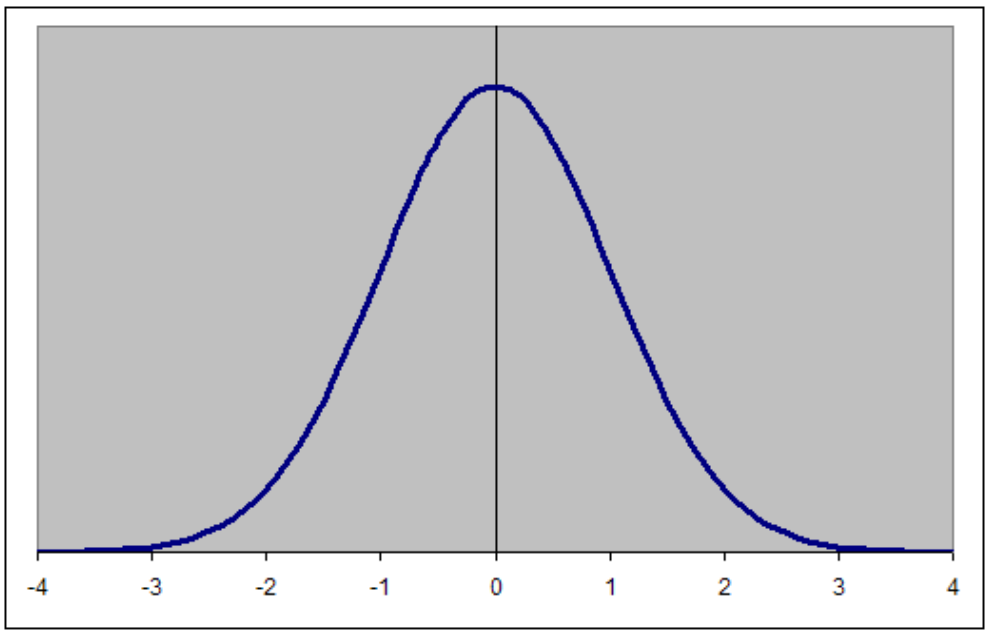
سه انحراف معیار در هر طرف میانگین امتدادداردبرابر است با :

$$p[\mu-3\delta\langle x\langle\mu+3\delta]=۰/۹۹$$



توزیع نرمال استاندارد ،دارای یک منحنی زنگوله ای با :

$\mu=0$(میانگین) و $\delta=1$(انحراف معیار)

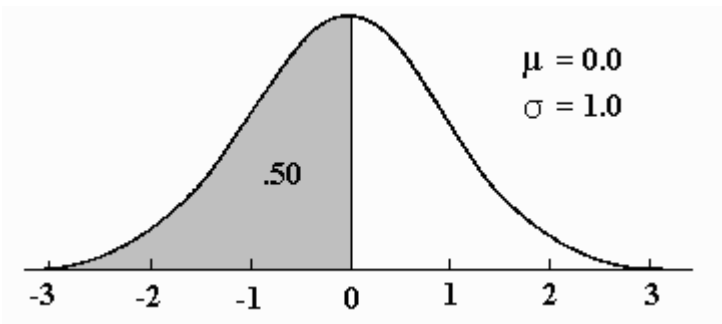


توزیع نرمال استاندارد بصورت N(۰و۱) نشان داده می شود که از تغییر متغیر $z=\frac{x-\mu}{\delta}$ در فرمول توزیع نرمال می توان آنرا بدست آورد:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}} \qquad -\infty\langle z\langle+\infty$$

خواص زیر را می توان از روی خاصیت تقارن حول صفر بدست آورد:

الف) $p[z\leq 0]=0/5$



ب) $p[z\leq -z]=1-p[z\leq z]$

ج) اگر Z>0 داریم :

$$p[z\leq z]=0/5+[0\langle z\leq z]$$
$$p[z\leq -z]=0/5+[0\langle z\leq z]$$

مثال : ۱- با استفاده از جدول نرمال استاندارد $p(z\geq$۱/۵۲) و $p(z\leq$۱/۵۲) را پیدا کنید ؟

حل : احتمال یا مساحت واقع در سمت چپ ۱/۵۲ برابر با ۹۳۵۷/. است در نتیجه :

$p(z\leq$۱/۵۲)=۰/۹۳۵۷

بعلاوه چون Z>۱/۵۲ متمم $Z\leq$۱/۵۲ است داریم :

$$p(z\geq 1/52)=1-p(z)\leq 1/52)=1-0/9357=0/643$$

مثال ۲- $P(۰/۱۵ <Z<۱/۶۰)$ را پیدا کنید ؟

با استفاده از جدول :

$P(z \leq ۱/۶۰)=۰/۹۴۵۲$

$p(z \leq -۰/۱۵)=-۰/۱۵$ = مساحت واقع در سمت چپ $=۰/۴۴۰۴$

$p(-۰/۱۵ < z < ۱/۶۰ =۰/۹۴۵۲-۰/۴۴۰۴=۰/۵۰۴۸$

مثال ۳- در صورتیکه x دارای توزیع $N(۱و۴)$ باشد $P(X>-۳)$ و $(-۳<x<۱۰)$ p را محاسبه کنید ؟

حل : منظور از $N(۱و۴)$ یعنی میانگین برابر ۱ و انحراف معیار برابر ۴ است. بنابراین

$$P(-۳<X<۱۰) = P\left(\frac{-۳-۱}{۴} \leq \frac{x-۱}{۴} \leq \frac{10-۱}{۴}\right) = P(-۱<Z<۲/۲۵)=۰/۸۲۹۱$$

$P(-۱<Z<۲/۲۵)=۰/۸۲۹۱$

نمونه سؤالات فصل پنجم:

۱- منظور از دو پیشامد مستقل چیست؟

الف- احتمال توام آنها صفر باشد. ب- احتمال حاصل جمع آنها صفر باشد.

ج- احتمال وقوع یکی در دیگری تاثیر نداشته باشد. د- ناسازگار باشند.

۲- با فرض اینکه احتمال دختر و پسر بودن فرزند۱/۲ باشد احتمال اینکه یک خانواده سه فرزندی حداکثر یک فرزند پسر داشته باشد چقدر است؟

الف- ۱/۲ ب- ۱/۴ ج- ۱/۸ د- ۳/۸

۳- اگر احتمال وقوع یک پیشامد ناچیز باشد در تکرار زیاد آزمایش برای محاسبه احتمال n بار وقوع آن حادثه کدام توزیع مناسبتر است؟

الف- دوجمله ای ب- نرمال

ج- پواسون د- توزیع t

۴- در یک توزیع نرمال استاندارد همواره ..... .

الف- میانگین و واریانس برابرند.

ب- میانگین برابر یک و واریانس برابر صفر می باشد.

ج- میانگین کمتر از واریانس است. د- میانگین بیشتر از واریانس است.

۵- کدام یک از ویژگیهای زیر در مورد توزیع نرمال صادق نیست؟
الف- توزیع نرمال حول میانگین متقارن است.
ب- در توزیع نرمال، میانگین و میانه و نما تقریبا برابرند.
ج- سطح زیر منحنی نرمال برابر یک است.
د- در توزیع نرمال، میانگین و انحراف معیار برابرند.
۶- فرض کنید توزیع فشارخون افراد یک جامعه، نرمال با میانگین ۱۲ وانحراف معیار ۱ باشد، مطلوب است: احتمال اینکه یکی از افراد به تصادف انتخاب شود دارای فشارخون کمتر از ۱۳ باشد؟

الف- ۰/۸۴۱۳ ب- ۰/۳۴۱۳

ج- ۰/۵ د- ۰/۷۸۵۱

تمرین:

۱- در جعبه ای ۶ مهره سفید و ۴ مهره سیاه وجود دارد. به تصادف ۲ مهره را با هم از جعبه خارج می کنیم:
الف- احتمال اینکه هر دو مهر سفید باشد چقدر است؟
ب- احتمال اینکه هر دو مهره یک رنگ باشد چقدر است؟

۲- در یک پدیده احتمال موفقیت هر بار آزمایش ۰/۰۰۲ باشد، اگر این آزمایش هزار بار تکرار شود:
الف- احتمال اینکه دو بار موفقیت حاصل شود چقدر است؟
ب- احتمال اینکه تعداد موفقیت ها از دو بار بیشتر نباشد چقدر است؟
ج- احتمال اینکه تعداد موفقیت ها از دو بار کمتر باشد چقدر است؟

۳- شیوع یک بیماری یک در هزار است. اگر دو هزار نفر از افراد جامعه به تصادف بررسی شوند احتمال اینکه دو نفر به این بیماری مبتلا باشند چقدر است؟

۴- احتمال مرگ کودکان زیر یک سال در جامعه ای ده درصد باشد، و احتمال مرگ کودکان بین یک سالگی تا پنج سالگی در همان جامعه ده درصد باشد:
الف- احتمال اینکه یک نوزاد تازه متولد شده به پایان پنج سالگی برسد چقدر است؟
ب- احتمال اینکه نوزاد مذکور بین یک تا پنج سالگی فوت نماید چقدر است؟

۵- امید ریاضی و انحراف معیار تعداد موفقیت در یک توزیع پوآسون کدام است؟اگر p برابر ۰/۰۰۱ و n برابر ۲۰۰۰ باشد.

۶- اگر $E(x)=2.5$ آنگاه $E(2x)$ چقدر است؟

۷- فرض کنید توزیع قند خون افراد یک جامعه، نرمال با میانگین ۱۰۰ وانحراف معیار ۳ باشد، مطلوب است:

الف- احتمال اینکه یکی از افراد به تصادف انتخاب شود دارای قند خون کمتر از ۹۴ باشد؟

ب- چند درصد افراد جامعه دارای قند خونی در فاصله ۹۷ تا ۱۰۶ خواهند بود؟

۸- فرض کیند قد دانشجویان پسر یک دانشگاه توزیع نرمال با میانگین ۱۶۸ و انحراف استاندارد ۵ سانتی متر باشد

الف- چند درصد این دانشجویان قدی بلندتر از ۱۷۸ سانتیمتر دارند؟

ب- اگر x نمایش قد دانشجویان باشد $P(X<۱۷۸)$ را محاسبه نمائید.

فصل ششم:
برآورد

هدفهای فصل ششم: دانشجو پس از مطالعه این فصل باید بتواند:

۱- توزیع نمونه ای میانگین و نسبت را مشخص کند.

۲- قضیه حدمرکزی را توضیح دهد.

۳- فاصله اطمینان را برای میانگین و نسبت تعیین کند.

۴- حجم نمونه را در برآورد های میانگین و نسبت برآورد کند.

۱-۶- مبانی یک بررسی نمونه ای:

وقتی بیان مسئله تحقیق به شکل یک مسئله آماری در آمده است بی گمان درصدد آن هستیم که با استفاده از داده های مربوط به نمونه، اطلاعات لازم در باره مشخصه های (پارامترهای) جمعیت را برآورد کنیم. یعنی آمارهایی که یک جمعیت را توصیف می کنند با این اطمینان که تا حدی از

دقت معقولی برخوردارند برآورد گردند. بنابر این در یک مطالعه نمونه گیری باید به نکات زیر توجه کرد:

الف- جمعیت مورد نمونه گیری چقدر قابل شناسایی است. گاهی اوقات ممکن است بین جمعیت نمونه گیری و جمعیت هدف اختلاف کمی وجود داشته باشد. برای شناسایی اعضاء جمعیت نمونه گیری ممکن است نیاز به یک چارچوب نمونه گیری باشد.مثلا لیست خانوارهای یک منطقه، اسامی یا شماره های دانشجویان یک دانشکده و یا نقشه جغرافیایی یک منطقه از چارچوبهای نمونه گیری محسوب می شوند. بر اساس این چارچوبها است که واحدهای نمونه مشخص می شوند.

ب- تعیین اطلاعات یا پارامترهای مورد نظر مانند میانگین یا نسبت در جامعه

ج- مشخص نمودن دقت لازم برای برآوردها

د- تعیین حجم نمونه مورد نیاز

انجام موارد مذکور تنها بخشی از یک مطالعه توصیفی نیست بلکه ارتباط بین داده های نمونه ای و پارامتر های جامعه را امکان پذیر می کند که از مبانی آمار استنباطی خواهد بود که در فصل برآورد مورد توجه قرار می گیرد.

نمونه گیری ،آماره ،توزیع نمونه ای:

به طور کلی نمونه گیری بر دو نوع است ، نمونه گیری احتمالی و نمونه گیری غیر احتمالی . برای بسیاری از مطالعات نمونه گیری احتمالی (تصادفی) کاربرد دارد.

نمونه احتمالی به نمونه ای گفته می شود که از جمعیت به گونه ای انتخاب گردد که هر عضو جمعیت احتمال مشخصی برای انتخاب شدن در نمونه مورد نظر را داشته باشد.

اگر نمونه ای nتایی از جمعیتی به اندازهN انتخاب شود به طریقی که هر یک از انتخابهای ممکن درنمونه گیری با اندازه n ، احتمال انتخاب شدن یکسان داشته باشد نمونه را نمونه تصادفی ساده می گویند.

نمونه گیری ساده به دو صورت با جایگزینی و یا بدون جایگزینی انجام می گیرد.

اگر نمونه گیری بدون جایگزینی انجام گیرد همه نمونه های ممکن به تعداد $\binom{N}{n}$ حالت می تواند اتفاق افتد که :

$$\binom{N}{n}=\frac{N!}{n!(N-n)!}$$

(بدون توجه به ترتیب واحد نمونه)

و اگر نمونه گیری با جایگزینی انجام گیرد تعداد همه نمونه های ممکن برابر N^n خواهد بود.

۲-۶- آماره :

در توصیف خصوصیات مشاهدات نمونه ای در انجام استنباط دربارهٔ جامعه با تعدادی معیار سر و کار داریم که از روی نمونه محاسبه می شوند . این معیارها ممکن است میانگین نمونه ، میانه و یا واریانس نمونه باشد .

هر یک از این معیارها تابع معینی از مشاهدات نمونه هستند . به طور کلی هر تابعی از مشاهدات نمونه ، آماره نامیده می شود.

۳-۶- توزیع های نمونه ای :

آماره که تابعی از نمونه تصادفی است متغیری تصادفی خواهد بود که دارای یک توزیع می باشد . توزیع احتمال آماره را توزیع نمونه ای آماره می گویند اگر $\bar{X}$ میانگین نمونه باشد به دلیل اینکه $\bar{X}$ متغیری به حساب می آید که برای هر نمونه n تائی ممکن, مقداری خاص را انتخاب می کند لذا توزیع میانگین نمونه یعنی توزیع نمونه ای $\bar{X}$ و واریانس آن مورد توجه است. این نکته مهم است که به منظور استنباط درباره جامعه از این توزیعهای نمونه ای در بر آورد و آزمون فرضها استفاده می شود.

توزیع میانگین نمونه :

وقتی نمونه گیری با جایگزینی انجام پذیرد یا از جمعیت نامحدود صورت گیرد به منظور تعمیم نتایج دو حالت مختلف در نظر گرفته می شود.

حالت اول : نمونه گیری از جمعیت برخوردار از توزیع نرمال انجام گیرد.

حالت دوم: نمونه گیری از جمعیت برخوردار از توزیع غیر نرمال صورت گیرد.

در حالت اول توزیع میانگین نمونه خواص زیر را دارا است .

الف – توزیع $\bar{X}$ نرمال خواهد بود.

ب- میانگین $\bar{X}$ برابر μ میانگین جمعیت اصلی است که نمونه از آن بوده است.

ج – واریانس توزیع $\bar{X}$ برابر نسبت واریانس جمعیت به اندازهٔ نمونه است.

در حالت دوم به کمک قضیه ریاضی بسیار مهم به نام "قضیه حد مرکزی" مجوز لازم برای نتیجه گیری فراهم می شود.

۴-۶- قضیه حد مرکزی :

در نمونه گیری تصادفی از جامعهٔ دلخواهی با میانگین μ و انحراف معیار σ ، وقتی کـه n بـزرگ است توزیع $\bar{x}$ توزیعی است تقریبا نرمال با میانگین μ و انحراف معیار $\frac{\sigma}{\sqrt{n}}$.

$$z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

به عبارت دیگر تقریباً N(۰ و ۱) است. (در عمل نمونه بیشتر از ۳۰ تا کافی است).

مثال ۱ : جمعیتی به طور نرمال با میانگین ۵۰ و انحراف معیار ۲ توزیع شده است . احتمالات زیـر را بر اساس اندازهٔ نمونه ای ۱۶ تایی پیدا کنید:

الف ـ $P(\bar{x}\geq 50)$

ب ـ $P(49\leq\bar{x}\leq 50)$

X → N (۵۰ و ۲)

حل :

(الف

$$\bar{X}\to N(50,\frac{2}{\sqrt{16}})$$

$$P(\bar{x}\geq 50)=P(\frac{\bar{x}-50}{\frac{2}{\sqrt{16}}}\geq\frac{50-50}{\frac{2}{\sqrt{16}}})=P(z\geq 0)=0.5$$

(ب

$$P(49\leq\bar{x}\leq 50)=P(\frac{49-50}{\frac{2}{\sqrt{16}}}\leq\frac{\bar{x}-50}{\frac{2}{\sqrt{16}}}\leq\frac{50-50}{\frac{2}{\sqrt{16}}})=P(-2\leq z\leq 0)$$

به کمک جدول توزیع نرمال استاندارد خواهیم داشت :

$$=P(0\leq z\leq 2)=0.477$$

مثال ۲: در جامعه ای با میانگین ۵۰ و انحراف معیار ۹ ، مطلوب است احتمال اینکه میـانگین یـک نمونه ۳۶ تائی :

الف – از ۴۸ کمتر باشد

ب – در فاصله ۴۹ تا ۵۳ باشد

حل : طبق قضیه فوق چون نمونه بالای ۳۰ تاست میانگین های حاصل از نمونه ها دارای یک توزیع

$$\frac{\delta}{\sqrt{n}}=\frac{9}{\sqrt{36}}=1.5$$

نرمال با میانگین ۵۰ و انحراف معیار است پس :

الف)
$$P(\bar{x}<48)=P(z<\frac{48-50}{1.5})=P(z<-1.33)=0.5-0.4082=0.0918$$

ب)
$$P(49<\bar{x}<53)=P(\frac{49-50}{1.5}<z<\frac{53-50}{1.5})=P(-0.67<z<2)=$$

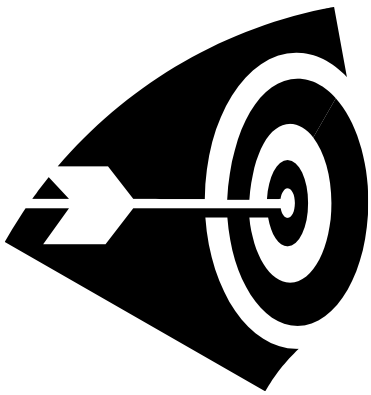
$$=0.2486+0.4772=0.7258$$

۵-۶- استنباط درباره جامعه :

هر خصوصیت عددی از توزیع جامعه ، یک پارامتر نامیده می شود. استباط آماری عبارت است از استخراج نکات کلی دربارهٔ پارامترهای جامعه ، از طریق تحلیل داده های نمونه .

$\bar{x}$ و δ نمونه هایی از آماره و پارامتر هستند و نیز S^2 واریانس نمونه آماره و δ^2 واریانس جامعه به عنوان یک پارامتر محسوب می شود.

بحث برآورد پارامترها و آزمون فرضهای آماری دو مبحث مهم آمار استنباطی است که مناسب با هدف تحقیق مورد استفاده قرار می گیرد.



۶-۶- برآورد میانگین :

محققی می خواهد در مورد میانگین تعداد ضربان قلب در دقیقه برای جمعیت معینی اطلاعاتی داشته باشد با انتخاب نمونه ای تصادفی می تواند

میانگین نمونه یعنی $\bar{x}$ را به عنوان برآورد میانگین جامعه μ به حساب آورد (*برآورد نقطه ای*).

و ممکن است فاصله ای از مقادیر حول میانگین نمونه ($\bar{X}$) را *برآوردی فاصله* ای از μ ارائه نماید.

۷-۶- فاصله اطمینان میانگین

فاصله اطمینان $(1-\alpha)$ درصد برای μ عبارت است از :

حد بالا $\mu=\bar{X}+z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

حد پایین $\mu=\bar{X}-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

با فرض اینکه توزیع جامعه نرمال یا حجم نمونه به اندازه کافی بزرگ باشد.

اگر n به قدر کافی بزرگ و انحراف معیار جامعه نامعلوم باشد فاصله اطمینان میانگین در سطح $(1-\alpha)$ برابر است با :

$$(\bar{x}-t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \quad , \quad \bar{x}+t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}})$$

ولی اگر حجم نمونه کوچک باشد فاصله اطمینان میانگین در سطح $(1-\alpha)$ به صورت زیر است :

$$(\bar{x}-t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \quad , \quad \bar{x}+t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}})$$

که در آن مقدار $t_{1-\frac{\alpha}{2}}$ از روی جدول توزیع t ی استودنت با n-۱ درجه آزادی تعیین می گردد.

مثال – نمونه ای ۱۰۰ تائی از پسران بالغ ۲۵ ساله که ظاهراً دارای توزیع نرمال فشار خون سیستولیک با میانگین ۱۲۵ می باشند تحت مطالعه است. اگر انحراف معیار جمعیت ۱۵ باشد .

الف – حدود اطمینان ۹۵ درصد را برای μ

ب- حدود اطمینان ۹۰ درصد را برای μ

پیدا کنید

$\bar{x}$=۱۲۵

حل : الف -۹۵٪ = $(1-\alpha)$ ٪بنابر این ۵٪=a

δ=۱۵

$z_{۰/۹۷۵}=۱/۹۶$ حد بالای $\mu=\bar{x}+z_{1-\frac{\alpha}{2}}\frac{\delta}{\sqrt{n}}=125+1/96\times\frac{15}{\sqrt{100}}=127/9$

حد پائین μ (با استفاده از جدول z)

$$= \bar{x} - z_{1-\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} = 125 - 1/96 \times \frac{15}{\sqrt{100}} = 122/06$$

ب- $\mathbf{Z}_{1-\frac{\alpha}{2}}$ = z۰/۹۵=۱/۶۵ → با استفاده از جدول z a=۰/۱۰ → a=٪۱۰ =(۱-a)٪۹۰

$$\mu \text{ حد بالای} = \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} = 125 + 1/65 \times \frac{15}{\sqrt{100}} = 127/4$$

$$\mu \text{ حد پایین} = \bar{x} - z_{1-\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} = 125 - 1/65 \times \frac{15}{\sqrt{100}} = 122/5$$

۸-۶- فاصله اطمینان بزرگ نمونه برای P

مثال : فرض کنید محققی می خواهد نسبت کودکان زیر ۵ سال را که مبتلا به سوء تغذیه هستند بر آورد نماید. برای این منظور یک نمونه ۵۰۰ تائی انتخاب و مشاهده می کند که ۵۰ نفر از کودکان مورد مطالعه دچار سوء تغذیه هستند . یک فاصله اطمینان ۹۵ درصد برای نسبت کودکان مبتلا به سوء تغذیه در جمعیت کودکان زیر ۵ سال محاسبه کنید.

حل: چون حجم نمونه (n=۵۰۰) بزرگ است بنابر این تقریب نرمال برای توزیع نسبت نمونه قابل توجیه است .

بنابر این از جدول نرمال داریم: $z_{1-\alpha/2} = 1.96$

و چون برآورد کننده نسبت جامعه $\hat{P} = \frac{x}{n}$ و انحراف استاندارد آن $\sqrt{\frac{\hat{P}(1-P)}{n}}$ می باشد

پس فاصله اطمینان ۹۵ درصد برای P عبارت است از :

$$\hat{P} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \quad \text{و} \quad \hat{P} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

بنابراین فاصله اطمینان ۹۵٪ برای مبتلایان به سوء تغذیه برابر است با :

$$\frac{50}{500}-1/96\sqrt{\frac{\frac{50}{500}(1-\frac{50}{500})}{500}},\frac{50}{500}+1/96\sqrt{\frac{\frac{50}{500}(1-\frac{50}{500})}{500}}$$

یعنی می توان ۹۵٪ اطمینان داشت که نسبت P در فاصله (۰/۱۳ و ۰/۰۷) قرار دارد.

۹-۶- برآورد حجم نمونه:

الف- در مطالعاتی که هدف آنها برآورد میانگین یک متغیر کمی در جامعه است برای تعیین حجم نمونه از فرمول زیر استفاده می شود:

$$n=\frac{(Z_{1-\frac{\alpha}{2}})^2\sigma^2}{d^2}$$

برای برآورد واریانس در این حالتها به یکی از صورت های زیر عمل می شود:
۱- با توجه به مطالعات مشابه دیگران، واریانس متغیر مورد نظر را برآورد می کنند.
۲- محقق با استفاده از تجارب قبلی خودش در باره واریانس متغیر اطلاع دارد.
۳- با انجام یک مطالعه راهنما یا مقدماتی، برآوردی از واریانس به دست می آورد.

مثال: فرض کنید محققی می خواهد میانگین وزن نوزادان در هنگام تولد را در استان مازندران با دقتی در حدود ۱۰ گرم برآورد نماید. با فرض اینکه انحراف معیار وزن نوزادان در هنگام تولد برابر ۵۰ گرم باشد درسطح اطمینان ۹۵درصد چه تعداد نمونه لازم است مطالعه شود؟

$$n=\frac{(۱/۹۶)(۵۰)^2}{(۱۰)^2}=9604$$

تعداد۹۷ نمونه انتخاب نمایند.

ب- در مطالعاتی که هدف آنها برآورد نسبت در جامعه است برای تعیین حجم نمونه از فرمول زیر استفاده می شود:

$$n=\frac{(Z_{1-\frac{\alpha}{2}})^2 P(1-P)}{d^2}$$

برای برآورد نسبت (p) در این حالتها به یکی از صورت های زیر عمل می شود:

۱- با توجه به مطالعات مشابه دیگران، نسبت را برآورد می کنند.
۲- محقق با استفاده از تجارب قبلی خودش در باره نسبت متغیر کیفی اطلاع دارد.
۳- با انجام یک مطالعه راهنما یا مقدماتی، برآوردی از نسبت به دست می آورد.
۴- در صورت عدم امکان حالت های فوق با قرار دادن P =۰/۵ می توان برآورد تقریبی از حجم نمونه ارائه داد که در صورتی که نسبت کسانی که ویژگی مورد نظر را دارند کمتر یا بیشتر از ۵/. باشد این تعداد نمونه کافی خواهد بود.

مثال: در برآورد شیوع بیماری خاصی که بر اساس مطالعات قبلی از ۲۰ درصد تجاوز نمی کند در سطح اطمینان ۹۵ درصد, وبا دقت ۲ درصد کمتر یا بیشتر, چه تعداد نمونه لازم است؟

$$n=\frac{(۱/۹۶)^2(۰/۲۰)(۰۰/۸}{(۰/۰۲)^2}=1536$$

پس ۱۵۳۷ نمونه لازم است تا مطالعه ای کاملتر برای برآورد شیوع بیماری مورد نظر انجام گیرد.

نکته: در این نوع برآورد ها پیشنهاد می شود مقدار d را حداکثر تا 0.1p در نظر بگیریم بدیهی است با مقادیر کمتر برآورد دقیقتری حاصل خواهد شد.

فرمولها برای تحقیقات تحلیلی یا تجربی و کارآزماییهای بالینی دو گروهی:

ج- تعیین حد اقل حجم نمونه برای مقایسه یک صفت کیفی در دو جامعه

$$n=\frac{(z_{1-\alpha/2}+z_{1-\beta})^2*[p_1(1-p_1)+p_2(1-p_2)]}{(p_1-p_2)^2}$$

د- تعیین حد اقل حجم نمونه برای مقایسه یک صفت کمی در دو جامعه

$$n=\frac{(z_{1-\alpha/2}+z_{1-\beta})^2*(\sigma_1^2+\sigma_2^2)}{(\bar{x}_1-\bar{x}_2)^2}$$

نکته: در حالتی که بخواهیم بیش از دو گروه را مقایسه کنیم همین فرمول ها کفایت می کند به شرط این که میانگین و واریانس آن دو گروهی که به هم نزدیک ترند را در فرمول استفاده کنیم.

نمونه سؤالات فصل ششم:

۱- شیوع یک بیماری حدود ۱۰ درصد است. حجم نمونه ای که با ۹۵ درصد اطمینان این شیوع را با خطای کمتر از ۲ درصد برآورد می کند تقریبا برابر است با:

الف- ۱۸۰ ب- ۱۸۰۰ د- ۹۰ د- ۹۰۰

مسئله:

۱- فرض کنید با توجه به تجارب قبلی ، شیوع بیماری سوء تغذیه در کودکان بیشتر از ۱۰ درصد نیست ولی برای برآورد دقیقتر نیاز به یک مطالعهٔ تحقیقاتی می باشد. اگر بخواهیم با حدود اطمینان ۹۵ درصد و با یک درصد کمتر یا بیشتر شیوع را برآورد کنیم تعداد نمونه لازم را محاسبه کنید.

۲- اگر با انتخاب نمونه ای به حجم ۱۰۰ از جامعه ای مشخص شود که تعداد ۲۰ نفر دارای بیماری A هستند :

الف- مطلوب است تعیین حدود اطمینانی که نسبت مبتلایان به بیماری را در سطح خطای ۵ درصد برآورد کند.

ب- برآورد نقطه ای نسبت مبتلایان چقدر است؟

۳- فرض کنید با توجه به تجارب قبلی ، بیشتر از ۱۰ درصد پرونده ها ناقص تکمیل شده اند ولی برای برآورد دقیقتر نیاز به یک مطالعهٔ تحقیقاتی می باشد. اگر بخواهیم با حدود اطمینان ۹۵ درصد و با یک درصد کمتر یا بیشتر این نسبت را برآورد کنیم تعداد نمونه لازم را محاسبه کنید.

۴- یک کارشناس بهداشت عمومی می خواهد نسبت خانمهای همسر دار واجب شرایط را که از روش پیشگیری از بارداری استفاده می نمایند بر آورد کند، بر اساس مطالعات قبلی این نسبت کمتر از ۴۰ درصد نباشد تعداد نمونه ی لازم با اطمینان ۹۵ درصد، این نسبت را با حد اکثر خطای ۴ درصد بر آورد نماید ، چه تعداد خواهد بود؟

۵- یک کارشناس بهداشت حرفه ای می خواهد میانگین ظرفیت باز دمی سریع (FVC) افراد سالم بالاتر۱۸سال یک منطقه را بر آورد کند اگر بر اساس مطالعات گذشته واریانس FVC برابر۰/۲۵ باشد با دقت ۰/۰۱ لیتر چه تعداد نمونه لازم است تا در سطح اطمینان ۹۵ درصد میانگینFVC برآورد گردد؟

در سطح اطمینان ۹۹درصد چه تعداد نمونه لازم است؟

فصل هفتم:
آزمون فرضیه

هدفهای فصل هفتم: شما باید بتوانید با مطالعه این فصل به هدفهای زیر دست یابید:

۱- آزمون فرضیه را توضیح دهید.

۲- انواع مهم آزمون فرضیه را بیان کنید.

۳- مراحل آزمون فرضیه را نام ببرید.

۴- آزمونهای مهم t را انجام دهید.

۵- آزمون کای دو را در فرضیه های مربوط به کار ببرید.

۶- خروجی کامپیوتری spss را در مورد آزمون های مهم تفسیر کنید.

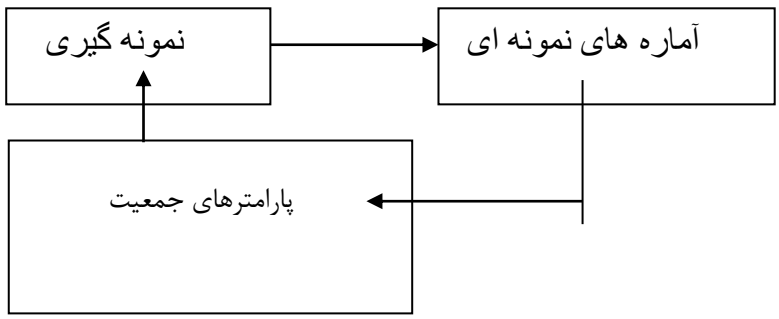
فرضیه (Hypothesis)

۱. یک پیشنهاد ثابت نشده

۲. یک راه حل ممکن برای یک مسئله

۳. حدس

استنباط آماری (STATISTICAL INFERENCE)



آزمون فرضیه

در آزمون آماری دو نوع فرضیه دخالت دارند ، اول ، فرضیه ای است که باید آزمون شود و آنرا فرضیه صفر می نامند و با علامت H_0 شان می دهند. فرضیه صفر را گاهی فرضیه عدم تفاوت نیز می نامند چون این فرضیه حاکی از موفقیت (یا عدم تفاوت) باشرایط واقعی جمعیت مورد نظر بوده و وضع موجود را بیان می دارد.

فرضی که در مقابل فرض صفر اقامه میگردد به فرض جانشینی یا فرض مقابل (فرض دیگر) معروف بوده و معمولاً آنرا با علامت H_a نشان می دهند. باید توجه داشت که به طور کلی آزمون فرضیه ها و استنباط آماری هیچ کدام به اثبات فرضیه انجامد ، بلکه تنها بر این نکته دلالت دارد که آیا فرضیه بوسیله داده های موجود تأئید می شود یا نه ؟

بنابر این اگر فرضیه صفر رد نشود نمی توان گفت که این فرضیه در حقیقت فرضیه درستی است بلکه می توان اظهار داشت که امکان صحت این فرضیه وجود دارد.

خطای نوع اول و دوم

در فرایند تصمیم گیری در جریان آزمون فرض امکان دو نوع خطا وجود دارد. رد کردن H_0 وقتی این فرض صحیح است خطای نوع I نامیده می شود احتمال آن البته برابر با x یعنی سطح آزمون است اما گیریم که فرض صفر غلط است یعنی فرض مقابل H_a درست است و مشخصاً در این موارد تصمیم درست آن است که فرض صفر غلط رد شود . خطای نوع دوم (II) که احتمال آن β خوانده می شود پذیرفتن H_0 هنگامی که غلط باشد به حساب می آید.

نتایج ممکن یک آزمون فرض

تصمیم / وضعیت	H_0 پذیرفته می شود	H_0 رد می شود
اگر H_0 درست باشد	تصمیم درست $(1-\alpha)$٪=احتمال سطح اطمینان=	خطای نوع I α = احتمال سطح آزمون =

اگر H_0 غلط باشد	خطای نوع II β = احتمال	تصمیم درست β -۱= توان آزمون

نکته مهم آن است که کم بودن خطای نوع I به قدر مطلوب ممکن است باعث افزایش خطای نوع II گردد به همین دلیل باید هر چه ممکن استα ی کوچکتری در نظر گرفت که خطای نوع دوم به قدر زیاد و نامعقولی افزایش نیابد تنها با افزایش نمونه می توانα وβ را کاهش داد. در این قسمت به چند نمونه از آزمونهای مهم آماری اشاره می کنیم.

انواع آزمونهای مهم:

۱- آزمون برابری(یا اختلاف) میانگین در یک جامعه با یک عدد ثابت

۲- آزمون برابری(یا اختلاف) نسبت در یک جامعه با یک عدد ثابت

۳- آزمون برابری(یا اختلاف) میانگین در دو جامعه آزمون T-TEST یا STUDEUT T-TEST

۴- آزمون برابری(یا اختلاف) نسبت در دو جامعه

۵- آزمون برابری(یا اختلاف) مشاهدات دوتائی (زوجی) -PAIRED T) (TEST یا t مزدوج

۶- آزمون بستگی صفات کیفی (مربع کای) یاآزمون CHI-SQUARE

مراحل آزمون فرضیه:

برای هرآزمون فرضیه باید مراحل زیر را اجرا نمائیم :

۱- تبیین کردن فرضیه صفر(H_0) و فرضیه مقابل آن (H_1)

۲- تعیین آماره مناسب برای آزمون فرضیه

۳- انتخاب تصادفی نمونه ها ونمونه گیری

۴- تعیین سطح معنی داری (LEVEVL OF SIGNIFICANCE) و انتخاب اعدادبحرانی (CRITICAL VALUE) از جدول آماری.

۵ - مقایسه مقادیر محاسبه شده با مقادیر بحرانی

۶- تصمیم گیری در مورد رد یا قبول فرضیه

آزمونهایt

در فرضیه هایی که مبتنی بر اختلاف میانگین داده های کمی در جامعه با یک عدد ثابت می باشد و یا فرضیه هایی که مبتنی بر اختلاف میانگین داده های کمی در دو جامعه هستند باید انحراف معیار جامعه مشخص باشد یا حجم نمونه به اندازه کافی بزرگ باشد تا برآورد خطای معیار ملاک

آزمون از دقت کافی برخوردار باشد. در غیر این صورت برای آزمون این فرضیه ها از آزمون های t استفاده می شود. آزمون های t بر مبنای کارهای آقای W. S. Gosset است که به دلیل استخدام در یک شرکت خود را در هنگام ارائه مقاله با نام مستعار دانشجو student معرفی می کرد. لذا این آزمونها به t-student نیز معروف شدند,که به شرح زیر می باشند.

الف- آزمون برابری (یا اختلاف) میانگین در یک جامعه با یک عدد ثابت

مراحل آزمون t یک نمونه ای: (one sample t-test)

مرحله ۱- تبیین کردن فرضیه مقابل H_0

تبیین کردن فرضیه مقابل H_1

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

مرحله۲- تعیین ملاک آزمون

مرحله ۳- با انجام نمونه گیری مقدار ملاک t محاسبه می شود.

مرحله ۴- مقدار $t_{1-\frac{\alpha}{2}}$ جدول t را با df=n-1 برای خطای مورد نظر یا سطح معنی داری α تعیین می کنیم.

مرحله ۵- اگر قدر مطلق مقدار t محاسبه شده از $t_{1-\frac{\alpha}{2}}$ جدول کمتر بود فرضیه H_0 رد نمی شود.

واگر بیشتر بود فرضیه H_0 رد می شود.

مرحله ۶- تفسیر وتصمیم گیری, یعنی منظور از رد یا عدم رد فرضیه صفر به چه معناست و در باره استفاده از نتایج آزمون تصمیم گیری شود.

نکته :

فرض های این آزمون عبارتند از :

الف- داده ها کمی هستند.

ب- نمونه ها از جمعیت با فرض توزیع نرمال برای متغیر مورد نظر انتخاب شده است.

ج- نمونه ها از هم مستقل اند.

مثال:

محققی مدعی است که توزیع فشار خون سیستولیک در یک جامعه نرمال با میانگین ۱۲۵ و انحراف معیار ۱۵ است. برای آزمون این فرضیه یک نمونه تصادفی ۲۵ تایی از افراد این جامعه انتخاب کرده است که میانگین فشار خون سیستولیک آنها برابر ۱۳۰ میلیمتر جیوه می باشد. آیا در سطح اطمینان ۹۵ درصد می توان نتیجه گرفت که ادعای او در باره میانگین جامعه مورد تایید است؟

حل: مرحله ۱- تبیین کردن فرضیه صفر H_0

تبیین کردن فرضیه مقابل H_1

$$H_0 : \mu = 125$$
$$H_1 : \mu \neq 125$$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

مرحله۲- تعیین ملاک آزمون

مرحله ۳- با انجام نمونه گیری مقدار ملاک t محاسبه می شود.

$$t = \frac{130 - 125}{15/\sqrt{25}} = 1.66$$

مرحله ۴- مقدار $t_{1-\frac{0.5}{2}} = t_{0.975}$ جدول t ضمیمه کتاب را با df=25-1 برای خطای مورد نظر یا سطح معنی داری ۹۵ درصد تعیین می کنیم که برابر ۲/۰۶۴ است.

مرحله ۵- قدر مطلق مقدار t محاسبه شده برابر ۱/۶۶ است که از $t_{1-\frac{\alpha}{2}}$ جدول یعنی ۲/۰۶۴ کمتر است بنابر این فرضیه H_0 رد نمی شود.

مرحله ۶- تفسیر وتصمیم گیری: منظور از رد فرضیه صفر یعنی ادعای محقق در باره میانگین توزیع بر اساس داده های نمونه ای رد نمی شود. و در باره استفاده از نتایج آزمون تصمیم گیری می شود.

ب- آزمون برابری (یااختلاف) میانگین در دو جامعه

آزمون T دو نمونه ای مستقل (independent two sample T-TEST یا STUDEUT T-TEST)

وقتی انحراف معیار دو جامعه معلوم باشد یا حجم نمونه های انتخابی از دو جامعه به اندازه کافی بزرگ باشد برای آزمون برابری میانگین در جامعه از آزمون z استفاده مس شود. ولی اگر نمونه ها کوچک باشند برای افزایش دقت برآورد خطای معیار در ملاک آزمون از انحراف معیار آمیخته از دو نمونه استفاده می شود و از آزمون t به شرح زیر استفاده می شود.

مراحل آزمون T دو نمونه ای مستقل:

مرحله ۱- تبیین کردن فرضیه صفر H_0

تبیین کردن فرضیه مقابل H_1

$$H_0 : \mu_1 = \mu_2$$
$$H_1 = \mu_1 \neq \mu_2$$

مرحله ۲- ملاک آزمون

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

که در آن :

$$s_p = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

مرحله ۳- با انجام نمونه گیری از دو جمعیت و محاسبه میانگین و انحراف معیار هر دو نمونه مقدار t محاسبه می شود.

مرحله ۴- مقدار $t_{1-\frac{\alpha}{2}}$ جدول t را با درجه آزادی $df.=n_1+n_2-2$ برای خطای مورد نظر یا سطح معنی داری αتعیین می کنیم.

مرحله ۵- اگر قدر مطلق مقدار t محاسبه شده از $t_{1-\frac{\alpha}{2}}$ جدول کمتر بود فرضیه H_0 رد نمی شود. واگر بیشتر بود فرضیه H_0 رد می شود.

مرحله ۶- تفسیر وتصمیم گیری, یعنی منظور از رد یا عدم رد فرضیه صفر به چه معناست و در باره استفاده از نتایج آزمون تصمیم گیری شود.

فرضیات آزمون t دو نمونه ای مستقل:

۱- متغیر وابسته کمی است.

۲- دو نمونه از توزیع تقریبا نرمال گرفته شدند.

۳- مشاهدات نمونه ای از هم مستقل هستند.

۴- دو نمونه از توزیع هایی گرفته شده اند که ممکن است در میانگین متفاوت ولی در واریانس ها یکسان هستند.

نکته: اگر فرضیات مطرح شده برقرار نباشند مثلا واریانس ها برابر نباشند در برآورد خطای معیار ممکن است از فرمول های دیگری استفاده شود (Armitage) .

مثال:

فرض کنید بخواهیم این فرضیه که "میانگین غلظت هموگلوبین خون افراد یک جامعه مردان بزرگسال در ناحیه ساحل دریا و ناحیه کوهستانی برابر است،" را آزمون کنیم. برای این کار دو نمونه از این دو جامعه در ساحل و مناطق کوهستانی به ترتیب به تعداد ۲۵ و ۱۶ نفر انتخاب کردیم که میانگین و انحراف معیار غلظت هموگلوبین به ترتیب در ناحیه ساحلی ۱۴/۳ و ۱/۱۲ و در ناحیه کوهستانی به ترتیب ۱۵/۳ و ۱/۱۷ بوده است. آیا داده ها دلیلی بر رد فرضیه محسوب ارائه می کنند؟ در سطح اطمینان ۹۵ درصد آزمون کنید.

آزمون فرضیه یک دامنه:

با توجه به کاهش فشار اکسیژن در نواحی کوهستانی افزایش غلظت هموگلوبین قابل انتظار است بنابر این محققین از قبل احتمال اینکه غلظت هموگلوبین در کوهستان کمتر از ساحل باشد را ناچیز می دانند بنابر این در تبیین فرضیه تساوی را در مقابل نامساوی یک طرفه مقایسه می کنند و این همان مفهوم *فرضیه یک دامنه* است که به شکل زیر نوشته می شود.

مرحله ۱- تبیین کردن فرضیه صفر H_0

تبیین کردن فرضیه مقابل H_1

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

مرحله ۲- ملاک آزمون

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

که در آن :

$$s_p = \sqrt{\frac{(17-1)(1.12)^2 + (30-1)(1.17)^2}{17+30-2}} = 1.153$$

مرحله ۳- با انجام نمونه گیری از دو جمعیت و محاسبه میانگین و انحراف معیار هر دو نمونه مقدار t محاسبه می شود.

$$t=\frac{147-153}{1.153\sqrt{\frac{1}{17}+\frac{1}{30}}}=-1.71<$$

مرحله ۴- چون آزمون یک دامنه است مقدار $t_{1-\alpha}$ جدول t را با درجه آزادی df.=17+30-2 برای خطای مورد نظر یا سطح معنی داری α تعیین می کنیم که این مقدار در جدول ضمیمه برای ۴۵ درجه آزادی برابر ۱/۶۸ است.

مرحله ۵- اگر قدر مطلق مقدار t محاسبه شده از $t_{1-\alpha}$ جدول کمتر بود فرضیه H_0 رد نمی شود. واگر بیشتر بود فرضیه H_0 رد می شود که در این مسئله ۱/۷۱۴ بیشتر از ۱/۶۸ است. بنابر این فرضیه H_0 رد می شود.

مرحله ۶- تفسیر وتصمیم گیری:

منظور از اینکه فرضیه H_0 رد می شود به این معناست که میانگین غلظت هموگلوبین خون در ناحیه کوهستانی در سطح اطمینان ۹۵ درصد بیشتر از مناطق ساحلی است.

ج- آزمون برابری (یا اختلاف) میانگین در مشاهدات دوتائی (زوجی) (PAIRED T-TEST) یا t مزدوج

اگر بخواهیم نتایج مشاهدات یک متغیرکمی را که در یک نمونه در دو مقطع زمانی جمع آوری شده اند مقایسه نماییم. به دلیل اینکه بین مشاهدات دو تایی برای هر فرد، استقلال وجود ندارد به نمی توان از آزمون t مستقل استفاده کرد و باید از آزمون t زوجی استفاده کرد.

فرضیات :

۱- داده ها کمی هستند.

۲- توزیع اختلاف بین دو مشاهده برای همه افراد تقریبا نرمال است.

۳- اختلافات از یکدیگر مستقل هستند.

مراحل آزمون t زوجی:

چون در این موارد از یک نمونه استفاده می شود, اگر اختلاف مشاهدات دو تایی برای هر فرد را به عنوان یک مقدار در نظر بگیریم ان آزمون حالت خاصی از آزمون t یک نمونه ای تبدیل خواهد شد که در آن فرضیه میانگین اختلاف ها μ_d برابر صفر است را مورد آزمون قرار می دهد. پس:

$$H_0:\mu_d=0$$
$$H_a:\mu_d\neq 0$$

مرحله ۱-

مرحله ۲- ملاک آزمون

مرحله ۳- با انجام نمونه گیری مقدار ملاک t محاسبه می شود.

$$t=\frac{\bar{d}}{S/\sqrt{n}}$$

مرحله ۴- مقدار $t_{1-\frac{\alpha}{2}}$ جدول t را با df=n-1 برای خطای مورد نظر یا سطح معنی داری α تعیین می کنیم.

مرحله ۵- اگر قدر مطلق مقدار t محاسبه شده از $t_{1-\frac{\alpha}{2}}$ جدول کمتر بود فرضیه H_0 رد نمی شود.

واگر بیشتر بود فرضیه H_0 رد می شود.

مرحله ۶- تفسیر وتصمیم گیری، یعنی منظور از رد یا عدم رد فرضیه صفر به چه معناست و در باره استفاده از نتایج آزمون تصمیم گیری شود.

مثال: آزمون t زوجی برای ارزیابی نتایج آزمایش crossover از اثر ارتفاع روی غلظت هموگلوبین سرم(g/dl) که در جدول زیر آمده است انجام می شود.

شماره نمونه i	در کنار دریاHb (X_{1i})	در ارتفاع زیاد Hb (X_{2i})	اختلاف $(d_i=X_{1i}-X_{2i})$
۱	۱۳,۸	۱۴,۱	-.۳
۲	۱۶,۱	۱۶,۴	-.۳
۳	۱۴,۶	۱۵,۱	-.۵

۴	۱۵,۰	۱۵,۸	-.۸
۵	۱۳,۱	۱۳,۸	-.۷
۶	۱۵,۶	۱۶,۶	-۱,۰
			$\Sigma d_i = -3.6$

$$\bar{d} = \frac{-3.6}{6} = -0.6$$

$$sd = \sqrt{\frac{\Sigma(d_i - \bar{d})^2}{n-1}} = 0.28$$

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{0.6}{\frac{0.28}{\sqrt{6}}} = -5.249$$

مقدار t جدول با درجه آزادی ۵ برابر ۱/۷۱۴ است که این مقدار t محاسبه شده از t جدول بیشتر است پس فرضیه صفر رد می شود. به عبارت دیگر حدود اطمینان برآورد شده برای اختلاف غلظت هموگلوبین در دو منطقه شامل صفر نیست پس اختلاف در سطح اطمینان ۹۵ درصد معنی دار است یعنی با تغییر ارتفاع ارتباط دارد.

$$df = n - 1 = 5$$

$$|t| \rangle 1.714$$

$$\Delta = \bar{d} \pm t\left(\frac{s_d}{\sqrt{n}}\right) = -0.6 \pm 2.571\left(\frac{0.28}{\sqrt{6}}\right)$$

$$= 0.6 \pm 0.3$$

$$= (-0.9 \quad -0.3) g/dl$$

آزمون کای دو (CHI-SQUARE)

برای آزمون بستگی صفات کیفی از آزمون کای دو استفاده می شود.

مراحل آزمون کای دو:

مرحله ۱- فرضیه صفر مبنی بر اینکه رابطه ای بین دو صفت وجود ندارد در مقابل این فرضیه که ارتباط وجود دارد آزمون می شود.

مرحله ۲- ملاک آزمون به صورت زیر تعیین می گردد.

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

که در آن O_{ij} و E_{ij} به ترتیب فراوانی های مشاهده شده و فراوانی های مورد انتظار تحت فرضیه صفر می باشد.

ابتدا بر اساس عبارت زیر فراوانی های مورد انتظار برای هر خانه جدول فراوانی مشاهده شده محاسبه می شود.اگر تعداد حالات متغیر کیفی اول در r سطر و تعداد حالات متغیر کیفی دوم در c سطر طبقه بندی شود وفراوانی مشاهده شده در هر حالت از ترکیات سطر و ستونهای مختلف در جدول آمده باشد فراوانی های مورد انتظار به صورت زیر محاسبه می شود.

$$E = \frac{\text{جمع ستون} \times \text{جمع سطر}}{\text{جمع کل}}$$

سپس به صورت زیر مجذور اختلافات نسبی فراوانی های مشاهده شده و فراوانی مورد انتظار را محاسبه می کنند.

یعنی برای هر خانه، فراوانیهای مشاهده شده را منهای فراوانیهای موردانتظار کرده به توان دو می رسانند. و حاصل را تقسیم بر فراوانی های مورد انتظار همان خانه می کنند. این مقادیر همان مجذور اختلاف نسبی بین مشاهدات نمونه ای و انتظاری است که تحت درست بودن فرضیه صفر حاصل می شود. پس از انجام این عمل برای همه خانه ها ، آنها را با هم جمع می کنند. نتایج حاصل همان مقدار ملاک کای دو می باشد.

که بزرگ بودن این مقدار برحسب مقادیر بحرانی جدول نشان دهنده عدم توافق مشاهدات نمونه ای و فرضیه مورد نظر است.

مرحله ۳- با انجام نمونه گیری مقدار ملاک کای دو χ^2 محاسبه می شود.

مرحله ۴- مقدار χ^2 جدول را با $df. = (r-1)(c-1)$ برای خطای مورد نظر یا سطح معنی داری α تعیین می کنیم که در آن r تعداد سطر و c تعداد ستون جدول داده ها است.

مرحله ۵- اگر قدر مطلق مقدار χ^2 محاسبه شده از $\chi^2_{1-\alpha}$ جدول کمتر بود فرضیه H_0 رد نمی شود. واگر بیشتر بود فرضیه H_0 رد می شود.

مرحله ۶- تفسیر وتصمیم گیری, یعنی منظور از رد یا عدم رد فرضیه صفر به چه معناست و در باره
استفاده از نتایج آزمون تصمیم گیری شود.

مثال: اطلاعات مربوط به اثر سه دارو در درمان بیماری معین در جدول زیر ثبت شده است برای مقایسه آن از آزمون کای دو استفاده شود تا مشخص شود آیا بین نتیجه درمانی و نوع دارو ارتباطی وجود دارد؟

نتیجه	نوع دارو			جمع
	داروی۳	داروی۲	داروی۱	
بهبود یافته	۱۶	۶	۸	۳۰
بهبود نیافته	۷	۶	۱۲	۲۵
جمع	۲۳	۱۲	۲۰	۵۵

مرحله ۱-

فرضیه H_0: سه نسبت با هم برابر می باشد(نتیجه درمانی مستقل از نوع دارو است)

فرضیه H_1: سه نسبت با هم برابر نمی باشد(نتیجه درمانی مستقل از نوع دارو نیست)

مرحله ۲-

جمع ستون × جمع سطر
---------- = E
جمع کل

^{2}(فراوانهای مشاهده شده- فراوانی های مورد انتظار)
---------- = ملاک کای دو
فراوانی های مورد انتظار

$$E_{16}=\frac{30\times23}{55}=1254 \quad E_6=\frac{12\times25}{55}=2227$$

$$E_7=\frac{23\times25}{55}=1045 \quad E_8=\frac{20\times30}{55}=109$$
$$E_6=\frac{12\times30}{55}=6.54 \quad E_{12}=\frac{20\times25}{55}=9.09$$

مرحله ۳-

مقدار ملاک کای دو χ^2 محاسبه شده:

0.95+1.13+0.04+11.88+0.77+0.93=15.7

مرحله ۴-

مقدار α =**0.05**

مرحله ۵-

عدد بحرانی جدول کای دو با دو درجه آزادی برابر 5.99 است.

مرحله ۶-

تصمیم گیری:

چون آماره محاسبه شده 15.7 بزرگتر از مقدار کای دو جدول می باشد فرضیه H_0 را رد می نمائیم و فرضیه H_1 را می پذیریم و چنین تفسیر مینمائیم که سه نسبت بـــا هـــم برابرنمی

باشند یا به عبارت دیگر وقوع بهبودی بستگی به نوع دارو دارد و بر اساس داده ها داروی ۳ بهتر از دو داروی دیگر عمل کرده است.

نمونه سؤالات فصل هفتم:

۱- منظور از خطای نوع اول چیست؟

الف- احتمال قبول فرضیه صفر وقتی اشتباه باشد

ب- احتمال رد فرضیه صفر وقتی درست باشد

ج- احتمال رد فرضیه صفر

د- احتمال رد فرضیه مقابل

۲- در تحقیقی که هدف آن تعیین ارتباط بین جنس و نتیجهٔ تست است چه آزمونی مناسب است؟

الف-آزمون t ب- آزمون کای دو

ج- ازمونF د- آزمون z

۳- برای مقایسه میانگین یک متغیر کمی در دوجامعه از کدام آزمون استفاده می شود؟

الف- آزمون مربع کای ب- آزمون t زوجی

ج- آزمون t مستقل د- آزمون نسبت

مسئله:

۱- یک کارشناس بهداشت حرفه ای مدعی است ؛ " بین نتیجه یک تست آزمایشگاهی و واحد محل خدمت کارگران، رابطه وجود دارد" برای آزمون این فرضیه، یک نمونه ۵۰ تائی از کارگران مثبت و یک نمونه ۵۰ تایی از کارگرانی که نتیجه تست آنها منفی بود از نظر محل کار مورد سوال قرار داد. مشخص شده است که تعداد ۳۰ تا از مثبتها در واحد تولید و۲۰ تا در واحد حمل و نقل مشغول بودند و در منفی ها به ترتیب ۱۸ نفر در واحد تولید و۳۲ نفر در حمل و نقل کار می کردند. در سطح اطمینان ۹۵ درصد نسبت به تائید یا رد ادعای مذکور اظهار نظر نمائید.

۲- مدیر واحد بهداشت محیط شهرستانی مدعی است که ۸۰ درصد مراکز تهیه مواد غذائی از کیفیت مطلوب بهداشتی برخوردارند، برای آزمون این ادعا، یک نمونه ۳۶ تائی از این مراکز به صورت تصادفی مورد بازدید قرار گرفتند. مشخص شده است که تعداد ۲۵ تا از وضعیت مطلوب برخوردار هستند. در سطح اطمینان ۹۵ درصد نسبت به تائید یا رد ادعای مذکور اظهار نظر نمائید.

۳- مقادیر میانگین و انحراف معیار پارامترهای اسپیرومتری اندازه گیری شده ی افراد سالم ۱۸ تا ۲۵ سال در یکی از شهرهای شمال کشور بر حسب جنس در جدول زیر آمده است.

به منظور تعیین ارتباط جنس با این پارامترها چه نوع آزمونی مناسب است. نتایج آزمون را بر حسب سطح معنی داری(P-value) ، فاصله ی اطمینان برای اختلاف میانگین ها در دو جنس گزارش و تفسیر کنید.

خروجی برنامه کامپیوتری SPSS پس از انجام آزمون مربوط در جدول زیر آمده است نتایج را با محاسبات خودتان مقایسه کنید.

Group Statistics

	SEX	N	Mean	Std. Deviation
m-Fvc	1	67	4.9775	.6310
	2	59	3.6786	.4902
m-Fev1	1	67	4.5073	.5463
	2	59	3.4069	.4330
m-mmef	1	67	5.5403	.9860
	2	59	4.3924	.8841

Independent Samples Test

	t-test for Equality of Means					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
m-Fvc	12.774	124	.000	1.2988	1.0976	1.5001
	12.978	122.168	.000	1.2988	1.1007	1.4969
m-Fev1	12.413	124	.000	1.1004	.9249	1.2758
	12.596	122.701	.000	1.1004	.9274	1.2733

m-mmef	6.842	124	.000	1.1479	.8159	1.4800
	6.890	123.955	.000	1.1479	.8182	1.4777

۴- توزیع فراوانی زنان همسردار برحسب نظرات آنان درباره ضرر روشهای پیش گیری در جدول زیر آمده است. این ها نمونه ای تصادفی از جمعیت زنان همسر دار شهرستان بابل هستند که در سال ۱۳۸۳ مورد پرسش قرار گرفتند. آیا ارتباطی بین نظرات آن ها با نوع روش های برنامه های تنظیم خانواده وجود دارد؟ در سطح اطمینان ۹۹ درصد آزمون کنید.

روش ضرر	قرص	IUD	آمپول تزریقی	کاندوم	بستن لوله	طبیعی
کم زیاد	۱۵۶	۴۳	۹	۲۶۷	۱۳۰	۴۴
	۱۸۴	۷۹	۲۲۷	۲۲	۱۰۲	۶

۵- توزیع فراوانی زنان بر حسب انتخاب نوع روش و محل سکونت (شهری یا روستایی بودن) در جدول زیر آمده است. این ها نمونه ای تصادفی از جمعیت زنان همسر دار شهرستان بابل هستند که در سال ۱۳۸۳ مورد پرسش قرار گرفتند. آیا ارتباطی بین انتخاب نوع روش های برنامه های تنظیم خانواده با محل سکونت وجود دارد؟

الف- در سطح اطمینان ۹۵ درصد آزمون کنید.

ب- در سطح اطمینان ۹۹ درصد آزمون کنید.

محل تولد	شهر	روستا	جمع
انتخاب روش	تعداد	تعداد	تعداد
قرص	۵۳	۶۲	۱۱۵
IUD	۱۴	۱۲	۲۶
آمپول	۳	۴	۷
کاندوم	۷۲	۳۶	۱۰۸
بستن لوله	۵۹	۵۶	۱۱۵
طبیعی	۱۲۷	۸۷	۲۱۴
هیچ روش	۵۷	۲۹	۸۱

۶- فرض کنید می خواهیم نسبت میزانهای عفونت را در بیمارانی که laparoscopy (جراحی شکم) انجام داده اند و آنتی بیوتیک مصرف کرده اند با دارونما مقایسه کنیم. اطلاعات در جدول زیر آمده است. آیا نسبت بیمارانی که مبتلا به عفونت شده اند در دو گروه متفاوت بوده است؟ آزمون مناسبی انجام دهید.

	بدون عفونت	عفونت	جمع
آنتی بیوتیک	۲۳۳	۷	۲۴۰
پلاسبو	۲۳۹	۲۱	۲۶۰
جمع	۴۷۲	۲۸	۵۰۰

فصل هشتم
تحلیل همبستگی و رگرسیون

هدف های این فصل:

با مطالعه این فصل دانشجو باید قادر باشد:

۱- نسبت به تعیین همبستگی دو متغیر کمی اقدام کند.

۲- رابطه رگرسیون خطی بین متغیر ها را شناسایی و معادله آن را بنویسد.

۳- خروجی کامپیوتری همبستگی و رگرسیون ساده خطی را تفسیر کند.

فرض ها:

به منظور بررسی ارتباط بین دو یا چند متغیر کمی، از روشهای تحلیل رگرسیون و همبستگی استفاده می شود.

در تحلیل رگرسیون و همبستگی برای دو متغیر x و y باید فرض های زیر برقرار باشد:

۱- نرمال بودن توزیع های xو y

۲- تصادفی بودن

۳- کمی بودن متغیرها

تحلیل همبستگی

روشی برای تحلیل ارتباط بین دو یا چند متغیر کمی است. قبل از انجام تحلیل همبستگی می توان با رسم نمودار پراکنش بین متغیر های x و y به صورت شهودی نسبت به وجود یا عدم وجود همبستگی آگاه شد. در صورت وجود همبستگی مقدار و شدت آن را بر اساس ضرایب مختلف تعیین کرد. یکی از مهمترین این ضرایب همان ضریب همبستگی پیرسون است.

اگر با افزایش مقادیر متغیر x ، مقادیرy نیز افزایش یابد ویا برعکس با کاهش مقادیر متغیر x ، مقادیرy نیز کاهش یابد می گوییم همبستگی مستقیم بین x و y وجود دارد که در آن صورت ضریب همبستگی، مقداری مثبت خواهد بود.

اگر با افزایش مقادیر متغیر x ، مقادیرy کاهش یابد ویا برعکس با کاهش مقادیر متغیر x ، مقادیرy افزایش یابد می گوییم همبستگی معکوس بین x و y وجود دارد که در آن صورت ضریب همبستگی، مقداری منفی خواهد بود.

اگر تغییرات x ، مستقل از تغییرات y باشد ضریب همبستگی بین آنها صفر خواهد بود. ولی همواره نمی توان نتیجه گرفت که اگر ضریب همبستگی بینx و y برابر صفر باشد آنگاهx و y مستقل اند، چون ممکن است رابطه غیر خطی بین آنها برقرار باشد، مگر در مواردی که توزیع x و y نرمال باشند.

فرمول ضریب همبستگی پیرسون:

$$r=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}=\frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)\,\mathrm{var}(y)}=\frac{\sigma_{xy}}{\sigma_x\cdot\sigma_y}$$

فرمول فوق را می توان به شکل زیر نیز نوشت:

$$r=\frac{\sum xy-\frac{\Sigma x\Sigma y}{n}}{\sqrt{(\sum x^2-\frac{(\Sigma x)^2}{n})(\sum y^2-\frac{(\Sigma y)^2}{n})}}$$

که پس از استاندارد کردن به شکل زیر در می آید:

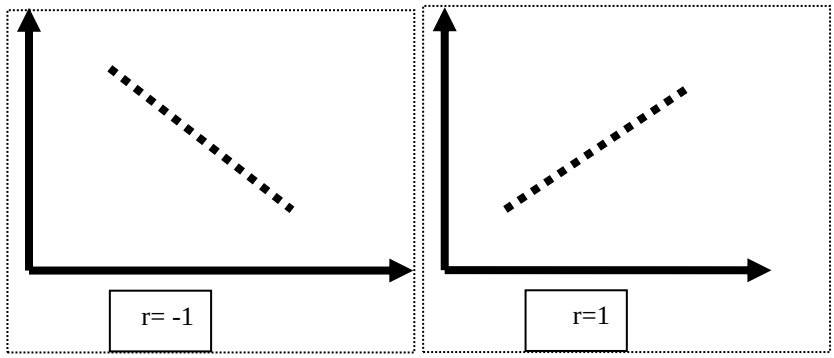
$$r=\frac{\Sigma z_x z_y}{n}$$

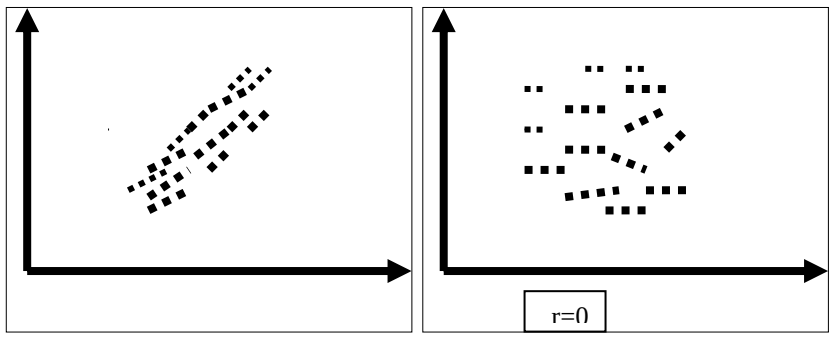
مقدار ضریب همبستگی همواره در فاصله ۱- تا ۱+ است. ضریب همبستگی را برای داده های مربوط به نمونه با r و برای جامعه با ρ نشان می دهند.

مجذور ضریب همبستگی، یعنی r^2 را ضریب تعیین می گویند که همواره مقداری مثبت یا صفر خواهد بود ونمی تواند منفی باشد.

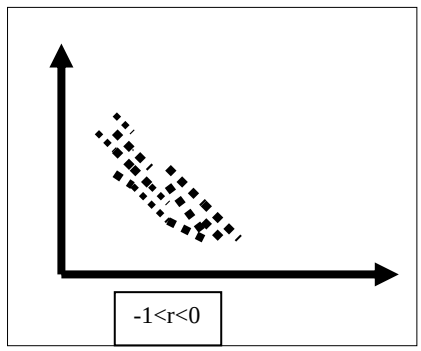
ضریب همبستگی در شکل های مختلف:

حالتهای مختلف همبستگی بین دو متغیر کمی که به ترتیب مقادیر آنها برروی محور های افقی و عمودی مشخص شده باشند برحسب ضریب همبستگی پیرسون r نشان داده شد.





0<r<1



آزمون اختلاف (برابری) ضریب همبستگی با صفر:

اگر بخواهیم فرضیه عدم وجود همبستگی یعنی $\rho=0$ را آزمون کنیم برآورد انحراف معیار r به صورت زیر است:

$$se(r)=\sqrt{\frac{(1-r^2)}{(n-2)}}$$

که آماره آزمون به صورت $t=\frac{r}{se(r)}$ از توزیع t با $n-2$ درجه آزادی برخوردار است که مقدار آن با جدول t مقایسه وتصمیم گیری می شود.
برای تبدیل r به t از جدول ضمیمه نیز می توان استفاده کرد.

مثال: ضریب همبستگی پیرسون بین متغیر های قد، وزن، و مقادیر اسپیرومتری اندازه گیری شده برای ۲۴۵ نفر از مردان بالای ۱۸ سال با استفاده از نرم افزار spss در زیر آمده است و برابری آن ها با صفر نیز آزمون شده است نسبت به تفسیر آن ها اقدام کنید.

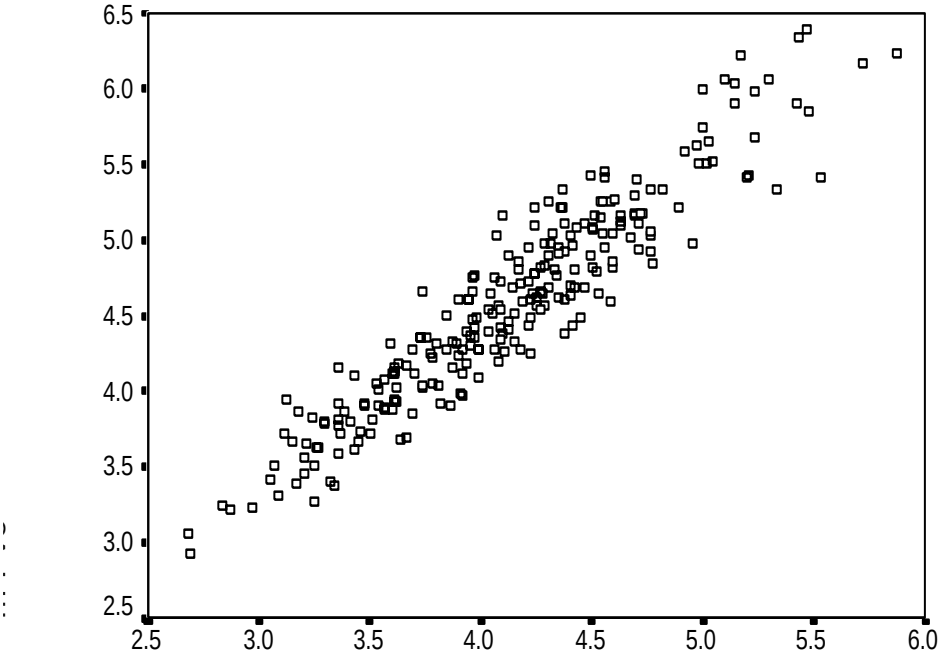
Correlations

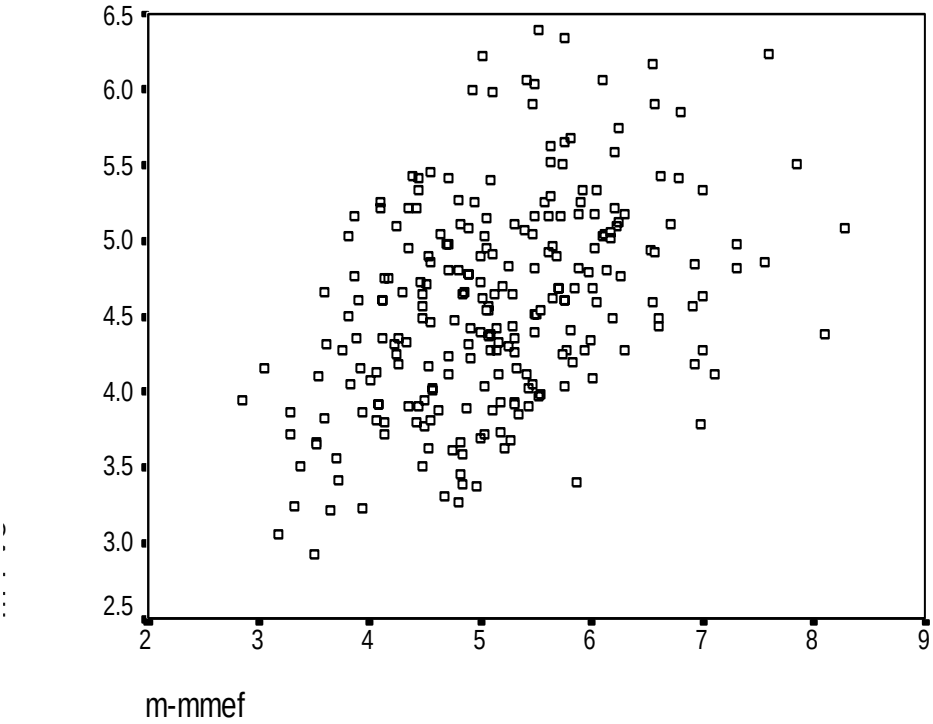
		HT	WT	m-Fvc	m-Fev1	m-Fev1%
HT	Pearson Correlation	1.000	.326	.753	.741	-.094
	Sig. (2-tailed)	.	.000	.000	.000	.143
	N	245	245	245	245	245
WT	Pearson Correlation	.326	1.000	.162	.159	-.018
	Sig. (2-tailed)	.000	.	.011	.013	.781
	N	245	245	245	245	245
m-Fvc	Pearson Correlation	.753	.162	1.000	.942	-.257
	Sig. (2-tailed)	.000	.011	.	.000	.000

	N	245	245	245	245	245
m-Fev1	Pearson Correlation	.741	.159	.942	1.000	.066
	Sig. (2-tailed)	.000	.013	.000	.	.306
	N	245	245	245	245	245
m-Fev1%	Pearson Correlation	-.094	-.018	-.257	.066	1.000
	Sig. (2-tailed)	.143	.781	.000	.306	.
	N	245	245	245	245	245

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

مثال: نمودار پراکنش مقادیر اسپیرومتری اندازه گیری شده با استفاده از نرم افزار spss در زیر رسم شده است بر اساس آن می توانید ضریب همبستگی پیرسون را در دو حالت مثبت و صفر مقایسه کنید.


6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
m-Fev1



تحلیل رگرسیون:

تحلیل رگرسیونی انواع مختلفی دارد که ساده ترین شکل آن تحلیل رگرسیون خطی است که در مواردی که یک رابطه خطی بین دو صفت مورد مطالعه به صورت $y_i = a + bx_i$ وجود دارد نسبت به انجام آن اقدام می شود. بیان ریاضی این رابطه را مدل هم می نامند.

در این معادله x متغیر مستقل یا ورودی وy متغیر وابسته یا خروجی نامیده می شود که با تغییر x متغیرy تغییر می کند. ضریب a به عنوان ثابت عرض از مبدا و b شیب خط یا ضریب خط رگرسیون نامیده می شود.

ضرایب خط رگرسیون یعنی aو b با فرمول های زیر برآورد می شوند:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})} = \frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} = \frac{\sigma_{xy}}{\sigma^2_x}$$

انحراف معیار شیب خط رگرسیون یعنی b به صورت زیر برآورد می گردد.

$$se(b)=\sqrt{\frac{E(xy)}{E(xx)}}$$

که در آن:

$$E(xy)=\sum(y-\bar{y})^2-b^2\sum(x-\bar{x})^2$$

$$E(xx)=(n-2)\sum(x-\bar{x})^2$$

که برای آزمون فرضیه صفر بودن شیب خط از آماره $t=\frac{b}{se(b)}$ از توزیع t با $n-2$ درجه آزادی برخوردار است که مقدار آن با جدول t مقایسه وتصمیم گیری می شود.

مثال: به منظور تعیین ارتباط بین قد افراد بر حداکثر ظرفیت حیاتی بازدمی مردان بالای ۱۸ سال شهرستان ساری مقادیر داده ها در جدول شماره آمده است. تحلیل رگرسیون خطی مربوط با نرم افزار آماری spss انجام گردید و نتایج بدست آمده به همراه نمودار پراکنش مربوط بین متغیرها در زیر ارائه گردید.

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	HT[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: m-Fvc

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.848[a]	.719	.719	.4786

a. Predictors: (Constant), HT

b. Dependent Variable: m-Fvc

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	296.127	1	296.127	1292.734	.000[a]
	Residual	115.680	505	.229		
	Total	411.807	506			

a. Predictors: (Constant), HT

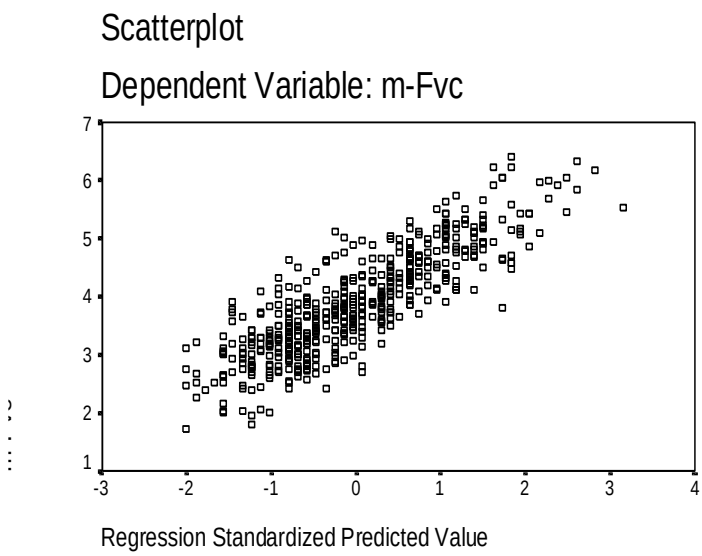
b. Dependent Variable: m-Fvc

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-9.883	.384		-25.738	.000
	HT	389E-02	.002	.848	35.955	.000

a. Dependent Variable: m-Fvc

Charts



مثال رگرسیون چندگانه:

برای تعیین ارتباط عوامل سن، وزن و قد افراد بر حداکثر ظرفیت حیاتی بازدمی مردان بالای ۱۸ سال شهرستان ساری مقادیر داده ها در جدول شماره آمده است. مدل رگرسیون چندگانه مربوط را با نرم افزار آماری spss ارائه و نتایج بدست آمده به همراه ماتریس همبستگی بین متغیرها ارائه گردید.

Regression

Correlations

		m-Fvc	AGE	HT	WT
Pearson Correlation	m-Fvc	1.000	-.619	.753	.162
	AGE	-.619	1.000	-.413	.062
	HT	.753	-.413	1.000	.326
	WT	.162	.062	.326	1.000
Sig. (1-tailed)	m-Fvc	.	.000	.000	.006

	AGE	.000	.	.000	.168
	HT	.000	.000	.	.000
	WT	.006	.168	.000	.
N	m-Fvc	245	245	245	245
	AGE	245	245	245	245
	HT	245	245	245	245
	WT	245	245	245	245

Model Summary

Model	R	R Square	Adjusted R Square	F Change	df1	df2	Sig. F Change
1	.753	.566	.565	317.501	1	243	.000
2	.825	.681	.678	86.387	1	242	.000

a Predictors: (Constant), HT
b Predictors: (Constant), HT, AGE
c Dependent Variable: m-Fvc

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64.808	1	64.808	317.501	.000
	Residual	49.601	243	.204		
	Total	114.408	244			
2	Regression	77.856	2	38.928	257.727	.000
	Residual	36.552	242	.151		
	Total	114.408	244			

a Predictors: (Constant), HT
b Predictors: (Constant), HT, AGE
c Dependent Variable: m-Fvc

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Coefficients Sig.
1	(Constant)	-7.994	.706		-11.322	.000
	HT	7.339E-02	.004	.753	17.819	.000
2	(Constant)	-4.642	.706		-6.573	.000
	HT	5.844E-02	.004	.599	15.018	.000
	AGE	-2.346E-02	.003	-.371	-9.294	.000

a Dependent Variable: m-Fvc

Excluded Variables

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	AGE	-.371	-9.294	.000	-.513	.829
	WT	-.093	-2.099	.037	-.134	.894
2	WT	-.012	-.308	.759	-.020	.847

a Predictors in the Model: (Constant), HT
b Predictors in the Model: (Constant), HT, AGE
c Dependent Variable: m-Fvc

Regression

Descriptive Statistics

	Mean	Std.	N

			Deviation	
	m-Fvc	3.2710	.5599	262
	AGE	34.64	11.43	262
	HT	157.81	5.23	262
	WT	68.00	12.47	262

Correlations

		m-Fvc	AGE	HT	WT
Pearson Correlation	m-Fvc	1.000	-.646	.556	-.041
	AGE	-.646	1.000	-.309	.356
	HT	.556	-.309	1.000	.114
	WT	-.041	.356	.114	1.000
Sig. (1-tailed)	m-Fvc	.	.000	.000	.255
	AGE	.000	.	.000	.000
	HT	.000	.000	.	.033
	WT	.255	.000	.033	.
N	m-Fvc	262	262	262	262
	AGE	262	262	262	262
	HT	262	262	262	262
	WT	262	262	262	262

Model Summary

Model	R	R Square	Adjusted R Square	F Change	df1	df2	Sig. F Change
1	.646	.417	.415	186.032	1	260	.000
2	.746	.557	.554	81.799	1	259	.000
3	.755	.569	.564	7.435	1	258	.007

a Predictors: (Constant), AGE
b Predictors: (Constant), AGE, HT
c Predictors: (Constant), AGE, HT, WT
d Dependent Variable: m-Fvc

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.126	1	34.126	186.032	.000
	Residual	47.695	260	.183		
	Total	81.821	261			
2	Regression	45.574	2	22.787	162.822	.000
	Residual	36.247	259	.140		
	Total	81.821	261			
3	Regression	46.589	3	15.530	113.723	.000
	Residual	35.232	258	.137		
	Total	81.821	261			

a Predictors: (Constant), AGE
b Predictors: (Constant), AGE, HT
c Predictors: (Constant), AGE, HT, WT
d Dependent Variable: m-Fvc

Coefficients

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	4.367	.085		51.631	.000
	AGE	-3.163E-02	.002	-.646	-13.639	.000
2	(Constant)	-2.488	.761		-3.267	.001
	AGE	-2.567E-02	.002	-.524	-12.051	.000

	HT	4.213E-02	.005	.393	9.044	.000
3	(Constant)	-2.258	.757		-2.983	.003
	AGE	-2.828E-02	.002	-.577	-12.234	.000
	HT	3.886E-02	.005	.363	8.173	.000
	WT	5.532E-03	.002	.123	2.727	.007

a Dependent Variable: m-Fvc

Excluded Variables

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	HT	.393	9.044	.000	.490	.904
	WT	.216	4.418	.000	.265	.873
2	WT	.123	2.727	.007	.167	.818

a Predictors in the Model: (Constant), AGE
b Predictors in the Model: (Constant), AGE, HT
c Dependent Variable: m-Fvc

Descriptive Statistics

	Mean	Std. Deviation	N
HT	171.26	7.02	245
WT	76.91	12.14	245

Correlations

		HT	WT
HT	Pearson Correlation	1.000	.326
	Sig. (2-tailed)	.	.000
	N	245	245
WT	Pearson Correlation	.326	1.000
	Sig. (2-tailed)	.000	.
	N	245	245

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
HT	164.31	9.12	507
WT	72.31	13.08	507

Correlations

		HT	WT
HT	Pearson Correlation	1.000	.396
	Sig. (2-tailed)	.	.000
	N	507	507
WT	Pearson Correlation	.396	1.000
	Sig. (2-tailed)	.000	.
	N	507	507

** Correlation is significant at the 0.01 level (2-tailed).

تمرین همبستگی:

۱- از ۱۰ نفر از کارکنان بهداشتی که در یک کلاس بازآموزی یک ماهه شرکت کرده اند در ابتدا و انتهای دوره آزمون به عمل آمد. نمرات کسب شده در جدول زیر آمده است به سوالات مربوط پاسخ دهید:

الف- ضریب همبستگی پیرسون (r) را برای داده ها محاسبه کنید.

ب- چه آزمون آماری باید انجام شود تا معلوم شود که این مقدار همبستگی معنی دار است یا نه؟

ردیف	پیش آزمون	آزمون نهایی
۱	۵۰	۹۸
۲	۳۰	۸۴
۳	۳۴	۸۵
۴	۴۵	۹۲
۵	۳۳	۸۷
۶	۲۳	۸۶
۷	۵۰	۹۷
۸	۱۳	۷۲
۹	۲۱	۷۶
۱۰	۴۶	۹۵

ج- آیا همبستگی محاسبه شده در سطح اطمینان ۹۵ درصد معنی دار است

د- نمودار پراکنش داده ها را رسم کنید. آیا داده ها به صورت خطی هم بسته اند. آیا داده ها نقاط پرت (outlier) دارند؟

ه- فرض کنید کارمند یازدهم به این جمع اضافه شده باشد به ترتیب با نمرات ۴۰ و ۷۰ در پیش آزمایی و آزمون نهایی، آیا این نمرات بر r تاثیر دارند؟

۲- در یک بررسی، وزن ۲۰ نوزاد در دوران جنینی به وسیله ی پارامتر های سونوگرافی تخمین زده شد (sonowei.) و پس از تولد نوزاد مقدار وزن اندازه گیری شده (weight) ثبت گردید. نتایج مطالعه بر حسب متغیر هایی مانند سن مادر(age) و وزن مادر(m weight) و سن حاملگی بر حسب هفته و روز در جدول زیر آمده است.

Gesage	weight	Sonoweight	mweight	Age	Row
37.43	3100	3000	58.0	25	1
41.00	4100	3130	70.0	34	2
39.14	2400	2640	68.5	27	3
40.86	3750	3580	59.0	33	4
37.00	3000	3074	72.0	26	5
39.71	3250	3021	66.5	19	6
39.00	3300	3400	66.0	21	7
38.71	3700	3326	90.0	29	8
38.86	3000	3182	86.0	29	9
37.43	2700	3150	61.0	22	10
37.00	3000	3074	71.5	26	11
37.43	2700	3150	61.0	22	12
38.57	3200	3384	80.0	25	13
37.57	3350	3293	75.0	21	14
40.86	3750	3580	59.0	33	15
37.14	2350	2364	50.0	25	16
41.29	2740	2600	70.0	23	17
37.71	3500	3517	62.0	21	18
38.43	3680	3324	91.0	21	19
38.57	2800	2655	73.0	19	20

به سوالات زیر پاسخ دهید:

الف- نمودار پراکنش را برای وزن تخمینی بر اساس سونوگرافی و وزن واقعی نوزاد در هنگام تولد رسم کنید.

ب- میزان همبستگی بین وزن تخمینی بر اساس سونوگرافی و وزن واقعی نوزاد در هنگام تولد را محاسبه کنید.

ج- میزان همبستگی بین وزن مادر و وزن نوزاد را محاسبه کنید.

د- میانگین میانه و نما را محاسبه کنید.

ه- شاخص های پراکندگی را محاسبه کنید.

خروجی کامپیوتری SPSS برای محاسبه ی همبستگی در جدول زیر آمده است نتایج را با محاسبات خودتان مقایسه کنید.

۳- فرض کنید نمرات میان ترم و پایان ترم درس آمار حیاتی دانشجویان بهداشت محیط در یک کلاس در جدول زیر نشان داده شده است.

ضمن رسم نمودار پراکنش، میزان همبستگی بین نمرات میان ترم و نمرات پایان ترم را محاسبه کنید.

ردیف	میان ترم	پایان ترم	ردیف	میان ترم	پایان ترم
۱	۱۷	۱۸	۱۱	۸	۱۱
۲	۱۵/۵	۱۴	۱۲	۱۷	۱۶/۵
۳	۱۶	۱۶/۵	۱۳	۷	۶
۴	۱۷	۱۸	۱۴	۱۴	۱۷
۵	۱۸	۲۰	۱۵	۱۹	۱۸
۶	۱۴/۵	۱۶	۱۶	۱۶	۱۷
۷	۱۶	۱۵	۱۷	۱۶	۱۵
۸	۲۰	۱۹	۱۸	۲۰	۲۰
۹	۱۲	۱۳	۱۹	۱۲	۱۰
۱۰	۱۳	۱۷	۲۰	۱۵	۱۶/۵

۴- به منظور تعیین ارتباط عوامل سن، وزن و قد افراد بر حداکثر ظرفیت حیاتی بازدمی مردان بالای ۱۸ سال شهرستان ساری مدل رگرسیون خطی مربوط که در قسمت قبل ارائه شد، تفسیر کنید.

فصل نهم
تجزیه و تحلیل واریانس

هدف های این فصل:

با مطالعه این فصل دانشجو باید قادر باشد:

۱- نسبت به تعیین مقایسه های میانگین یک متغیر کمی در چند جمعیت اقدام کند.

۲- محاسبات تحلیل واریانس را انجام دهد.

۳- خروجی کامپیوتری آنالیز واریانس را تفسیر کند.

تجزیه و تحلیل واریانس (ANOVA)

اگر بخواهیم میانگینهای بیش از دو جامعه را با هم مقایسه کنیم و برابری آنها را بر پایه میانگینهای نمونه ای آزمون کنیم از روش تجزیه و تحلیل واریانس (آنالیز واریانس) استفاده می کنیم.

ممکن است پرسیده شود که برای مقایسه میانگین چند جامعه چرا میانگین آنها را بر اساس آزمون t مستقل به صورت دو به دو مقایسه نمی کنند؟ در پاسخ باید گفت در انجام هر بار آزمون t اگر احتمال خطای آزمون α باشد در انجام آزمونهای دو به دو میزان خطای کلی افزایش می یابد. ولی اگر از روش آنالیز واریانس استفاده شود میزان خطا به مقدار αn کاهش می یابد که n برابر تعداد گروهها است.

فرض کنیم بخواهیم k جمعیت با میانگین های $\mu 1$ و$\mu 2$ و . . . و μk و واریانس مشترک $\delta 2$ را مقایسه کنیم، آنالیز واریانس روش مناسبی برای آزمون فرضیه زیر خواهد بود:

$H0: \mu 1= \mu 2= \mu 3= . . .= \mu k$

$H1: \mu 1\neq \mu 2\neq \mu 3\neq . . . \neq \mu k$ ⟶

(حداقل میانگین دو جمعیت مخالف هستند)

اگر نمونه تصادفی ni تایی از هر یک از جمعیت ها انتخاب نماییم و میانگین آنها را به ترتیب با $\bar{X}_i$ نشان دهیم میانگین کل نمونه ها برابر $\bar{X}_i$ باشد. شکل کلی مشاهدات و آنالیز واریانس در جدولهای زیر خلاصه می شود.

جدول : تنظیم مشاهدات نمونه های مختلف از جمعیت ها

	جمعیت های آماری					
	۱	۲	...	i	...	k
مشاهدات	X_{11} X_{21} ... $X_{n_1 1}$	X_{12} X_{22} ... $X_{n_2 2}$		X_{1i} X_{2i} ... $X_{n_i i}$		X_{1k} X_{2k} ... $X_{n_k k}$
جمع	$\sum_{i=1}^{n_1} X_{i1} = X_{\cdot 1}$	$\sum_{i=1}^{n_2} X_{i2} = X_{\cdot 2}$		$\sum_{j=1}^{n_i} X_{ji} = X_{\cdot i}$		
میانگین	$\bar{X}$	$\bar{X}_{01}$	...	$\bar{X}_{02}$		$\bar{X}_{0k}$

روش محاسبات:

پس از محاسبه مجموع هر ستون، میانگین های هر نمونه از هر جمعیت (میانگین ستونها)به صورت زیر محاسبه می شود:

$$\bar{X}_{0i} = \frac{X_{0i}}{n_i}$$

میانگین کل مشاهدات:

$$\bar{X}_{..} = \frac{\sum_{i=1}^{n_i} \sum_{j=1}^{k} x_{ii}}{\sum n_i} = \frac{x_{...}}{n}$$

فاکتور تصحیح:

$$CF = \text{correction factor} = \frac{X_{...}^2}{n}$$

مجموع کل مربعات یا کل انحرافات:

مربعات کل

$$\begin{cases} T_{SS} = \sum_i \sum_i x_{ij}^2 - CF \\ V_{ss} = \dfrac{\sum_i \sum_i x_{ij}^2}{n_i} - CF \\ E_{ss} = T_{ss} - V_{ss} \end{cases}$$

مربعات ستون

مربعات خطا

محاسبه واریانس:

$$\begin{cases} VMS = \dfrac{V_{ss}}{k-1} \\ EMS = \dfrac{E_{ss}}{n-k} \end{cases}$$

محاسبه F یا آمارافیشر:

$$F = \frac{VMS}{EMS}$$

جدول : تجزیه و تحلیل واریانس

منابع تغییرات	درجه آزادی(df)	مجموع مربعات ss	میانگین مربعات (Ms)	F
بین گروهها(تیمار treatment)	k-1	Vss	VMS	$F=\frac{VMS}{EMS}$
داخل گروهها (خطا)	n-k	ESS	EMS	
مجموع کل	n-1	TSS		

تصمیم گیری:

پس از محاسبه F و مقایسه آن با Ft جدول در سطح α و درجه آزادی مربوطه تصمیم گیری می شود. Ft جدول از طریق جدول ضمیمه تعیین می گردد. چون F از تقسیم واریانس بین گروهها بر واریانس داخل گروهها بدست می آید پس درجه آزادی k-1 در صورت و درجه آزادی n-k در مخرج بر اساس α های تعیین شده مقدار Ft جدول مشخص می شود. اگر Ft بزرگتر باشد

فرض H0 را در سطح اطمینان (1-α) رد می کنیم. در غیر اینصورت با مشاهدات مذکور H0 رد نمی شود.

تفسیر:

وقتی H0 رد شود یعنی بین میانگین های حداقل دو جامعه اختلاف معنی داری وجود دارد. حال باید به این سوأل پاسخ داد که کدام یک از دو جامعه دارای میانگین های مختلفی هستند؟

برای پاسخ این سوال, آزمون t چندگانه (multiple t) متعاقب آنالیز واریانس که اختلاف را نشان دهد انجام می گیرد.

مثال: وزن ۲۰ نوزاد به کیلوگرم و بر حسب سن حاملگی مادران آنها در سه گروه در جدول آمده است. آیا میانگین وزن نوزادان این مادران با هم اختلاف دارد؟

کمتر یا مساوی ۳۷ هفته	۳۸ تا ۳۹ هفته	۴۰ هفته یا بیشتر
۳/۱	۲/۴	۴/۱
۳	۳/۲	۳/۸
۲/۷	۳/۳	۳/۸
۳	۳/۷	۲/۸
۲/۷	۳	
۳/۴	۳/۲	
۲/۴	۳/۵	
	۳/۷	
	۲/۸	
$N_1=7$	$N_2=9$	$N_3=4$

$\sum_{i=1}^{n_i} X_i$	۱۴/۵	۲۸/۸	۲۰/۳	۶۳/۶
$\overline{X}$	۳/۶	۳/۲	۲/۹	۳/۱۸

$\sum_{i=1}^{n_i} X_i^2$	۵۳/۵۳	۹۳/۶	۵۹/۵۱	۲۰۶/۶۴

$$CF=\frac{(636)^2}{n}=\frac{404496}{20}=20224$$

$$TSS=20664-20224=439$$

$$Vss=\frac{(145)^2}{4}+\frac{(282)^2}{9}+\frac{(203)^2}{7}-20224=134$$

$$Ess=439-134=305$$

$$VMS=\frac{134}{3-1}=0671$$

$$EMS=\frac{305}{20-3}=0179$$

$$F=\frac{0671}{0179}=374$$

با F جدول در سطح ۰/۰۵ با ۲ درجه آزادی در صورت و ۱۷ درجه آزادی در مخرج کسر برابر ۳/۵۹ است پس مقدار F محاسبه شده از مقدار جدول بزرگتر است . پس فرضیه H_0 رد می شود.

تفسیر: یعنی اختلاف میانگین های حداقل دو تا از گروه ها معنی دار است. با نگاه کردن به مقادیر میانگین ها در جدول مشخص می شود که وزن هنگام تولد نوزادان مادرانی که کمتر از

۳۷ هفته حاملگی آنها طول کشیده است از گروه مادران با سن حاملگی ۴۰ هفته یا بالاتر کمتر بوده است. یعنی ۲/۹ کیلو گرم در مقابل ۳/۶۲۵ کیلوگرم تفاوت معنی داری را در سطح اطمینان ۹۵ درصد نشان می دهد.

ولی با F جدول در سطح ۰/۰۱ با ۲ درجه آزادی در صورت و ۱۷ درجه آزادی در مخرج کسر برابر ۶/۱۱ است پس مقدار F محاسبه شده از مقدار جدول کوچکتر است . پس نمی توان فرضیه H_0 در سطح اطمینان ۹۹ درصد رد کرد. یعنی اختلاف میانگین های حداقل دو تا از گروه ها با اطمینان ۹۵ درصد معنی دار نیست.

$$(X_{ij}-\overline{X..})=(\overline{X_{.j}}-\overline{X_{..}})+(X_{ij}-\overline{X_{.j}})$$

تجزیه واریانس:
واریانس داخل گروه ها + واریانس بین گروه ها = مجموع واریانس ها

پس از مجذور کردن دو طرف عبارت و محاسبات ریاضی:

مجموع مجذورات واریانس های داخل گروه ها + مجموع مجذورات واریانس بین گروه ها = مجموع مجذورات کل واریانس ها

$$SSTO=SSTR+SSE$$

پس از تقسیم این ها بر درجه آزادی شان، میانگین مجذورات به دست می آید.

$$MSTR=\frac{SSTR}{r-1}$$
$$MSE=\frac{SSE}{n_t-r}$$

$$F^* = \frac{MSTI}{MSE}$$

مقدار F محاسبه شده با مقدار جدول مقایسه می شود. اگر:

$$F^* \leq F(1-\alpha, r-1, n_t - r)$$

آنگاه فرضیه صفر رد نمی شود و در غیر این صورت رد می شود.

فصل دهم
شاخص های بهداشتی و جمعیتی

هدف های این فصل:

دانشجو پس از مطالعه این فصل می تواند:

۱- بسیاری از شاخصهای مهم بهداشتی را تعریف کند.
۲- شاخصهای بهداشتی مهم را محاسبه کند.
۳- با روش های جمع آوری داده ها برای محاسبه شاخصهای بهداشتی آشنا شود.
۴- برخی شاخصهای جمعیتی، اقتصادی و اجتماعی مرتبط با بررسی های بهداشتی و درمانی را محاسبه نماید.

شاخص های بهداشتی (Health Indices)

برای برنامه ریزی صحیح در زمینه مسائل بهداشتی و درمانی، نیاز است وضعیت موجود و وضعیت مطلوب جامعه از نظر بهداشت ودرمان مورد مطالعه قرار گیرد. چون برنامه ریزی برای رسین از وضعیت موجود به وضعیت مطلوب است در هر دو صورت هم برای شناخت وضعیت موجود وهم برای ترسیم وضعیت مطلوب نیازمند به شاخص ها ونشانگرهای بهداشتی هستیم. شاخص ها یا نشانگرهای بهداشتی مجموعه ای از کسرها، نسبت ها، اعداد وارقام هستند که متناسب با نمود ها

وویژگی های مثبت یا منفی در جامعه مانند سلامت و بیماری، خدمات بهداشتی یا درمانی، وضعیت باروری یا مرگ ومیر به صورتهای مختلف تعریف می شوند. علاوه بر شناخت وضع موجود و ترسیم وضع مطلوب برای نشان دادن پیشرفت کار شناخت هر مرحله ای از اجرای برنامه نیاز به شاخص های بهداشتی است. چون بهداشت یک مقوله کلی و یک مفهوم ذهنی است. برای اینکه مشخص شود که جامعه در چه سطحی از بهداشت بر خوردار است باید وجود یا عدم وجود بیماری یا مرگ و میر ، پوشش خدمات، وجود اتفاقات و ... را تعیین کرد. این ها همه بر اساس شاخص های بهداشتی است.

نیاز به اطلاعات و شاخصهای بهداشتی

برنامه ریزی خوب باعث می شود که راه رسیدن از وضعیت موجود به وضعیت مطلوب (هدف) مشخص شود و با مدیریت و نظارت و ارزشیابی برنامه ها، استفاده اصولی از منابع فراهم گردد و شکوفایی استعدادها و مشارکت جامعه را فراهم سازد و سطح سلامت فرد و جامعه را ارتقا بخشد. مدیران، برنامه ریزان، سیاستگزاران و کارشناسان بهداشتی و درمانی باید تمام اجزائ این فرایند را بشناسند. این مهم فراهم نمی شود مگر با استفاده از اطلاعات و شاخصهای بهداشتی درمانی.

اطلاعات و شاخصها را می توان برای بخشهای زیر مورد استفاده قرار داد:

۱- بررسی وضعیت موجود

برای تعیین نیازمندیها و شناخت مشکلات مردم مناطق مختلف باید محیط اجتماعی- فرهنگی و اقتصادی منطقه بررسی و میزان بیماریها، مرگ و میر و پوشش خدمات محاسبه گردد. شناخت وضعیت موجود بدون استفاده از آمار و شاخصها امکان پذیر نیست.

۲- تعیین وضع مطلوب:

هدف بهداشت برای همه در برنامه های سازمان جهانی بهداشت نمونه بارزی است از اینکه شاخصهای مختلفی برای بیان این هدفها مورد استفاده قرار گرفته است. ترسیم وضع مطلوب در ذهن مدیران و برنامه ریزان و مردم منطقه نیازمند استفاده از شاخصها است که آنها را در دستیابی به هدفها تشویق و ترغیب نماید.

۳- نظارت (پایش):

جزئی از فرآیند مدیریت است و بر اندازه گیری منابع مانند پول، نیروی انسانی و امکانات در مقابل فعالیتهای انجام شده مانند کودکان مراقبت شده، بیماران درمان شده و غیره تأکید دارد. برای مشخص شدن وضعیت اجرا برنامه ها و پایش آنها نیاز به اطلاعات و شاخصها در زمانهای مختلف است.

۴- ارزشیابی:

ارزشیابی فرآیندی است که مناسب بودن، موثر بودن و میزان دستیابی به هدفها را مشخص می کند. پوشش خدمات به رضایتمندی مردم، تغییر در رفتار آنها برای ارزشیابی برنامه ها مورد بررسی قرار می گیرند. این بررسی ها به کمک اطلاعات و شاخصهای کمی و کیفی انجام می گیرند.

۵- توسعه علمی و پیشنهاد برنامه های جدید:

پس از ارزشیابی و گزارش نتایج بررسی ها که بر مبنای اطلاعات و شاخصها انجام می گیرد بخشی از نیازهای علمی آمار و روش تحقیق تأمین می گردد. به منظور توسعه و بهبود روشهای علمی در توصیف و تحلیل پدیده های بهداشتی و درمانی، مقایسه و مدل سازی آنها و ارائه تکنولوژی مناسب توجه متخصصین را به خود جلب می کند که نیازهای آینده را تأمین می کنند.

زیج حیاتی (Vital Horoscope)

یکی از مناسبترین ابزارهای جمع آوری داده ها برای محاسبه شاخصهای بهداشتی در خانه بهداشت، زیج حیاتی است که در طول سال باید بر دیوار خانهء بهداشت نصب و در طی سال تعداد وقایع مهم مربوط به مرگ و میر، باروری، برخی بیماریها در محیط روستاهای اصلی و قمر که تحت پوشش خدمات آن خانه بهداشت هستند بر حسب سن، جنس، ماه و خدمات تنظیم خانواده در آن ثبت شود.

هدف از تکمیل آن، داشتن اطلاعات دقیق و به هنگام درباره جمعیت تحت پوشش روستایی در همه نقاط کشور است. این داده ها برای محاسبه انواع شاخصهای بهداشتی که در این بخش گفته شده است مورد استفاده قرار می گیرد.

یک زیج حیاتی شامل تعدادی جداول و دایره اصلی است که برای روستاهای اصلی در محل □ و برای روستاهای قمر در محل O داده های مربوط تکمیل می گردد.(شکل شماره)

۱- جدول جمعیت بر حسب سن و جنس

۲- جدول مرگ و میر بر حسب سن و علت مرگ

۳- جدول متولدین بر حسب وزن، جنس، سن مادر و شرایط زایمان

۴- جدول مرگ و میر بر حسب سن و جنس

۵- جدول پوشش تنظیم خانواده بر حسب روش ها و ماههای سال

۶- جدول مرگ و میر کودکان زیر ۵ سال بر حسب علت مرگ

۷- دایره وسط ذیج برای نشان دادن تولدهای زنده و مرگ ها بر حسب ماههای سال

۱- جدول جمعیت بر حسب سن و جنس

در ابتدای هر سال جمعیت روستای اصلی و قمر توسط بهورزان محترم سرشماری شده ونتایج آن

به تفکیک روستاهای اصلی و قمر در این جدول ثبت شده و به عنوان اطلاعات پایه در محاسبه شاخص ها مورد استفاده قرار می گیرد. اطلاعات جمعیتی هر روستا در پایان سرشماری به مرکز بهداشت شهرستان ارسال شده و جدول جمعیت روستایی شهرستان را تشکیل می دهد .

جمعیت بر حسب سن و جنس

منطقه و جنس / سن	شهر/روستای اصلی ☐		قمر ◯		کل	
	مرد	زن *	مرد	زن *	مرد	زن *
کمتر از یکسال						
۱ تا ۴ سال						
۵ تا ۹ سال						
۱۰ تا ۱۴ سال						
۱۵ تا ۱۹ سال						
۲۰ تا ۲۴ سال						
۲۵ تا ۲۹ سال						
۳۰ تا ۳۴ سال						
۳۵ تا ۳۹ سال						
۴۰ تا ۴۴ سال						
۴۵ تا ۴۹ سال						
۵۰ تا ۵۴ سال						
۵۵ تا ۵۹ سال						
۶۰ تا ۶۴ سال						
۶۵ تا ۶۹ سال						
۷۰ تا ۷۴ سال						
۷۵ تا ۷۹ سال						
۸۰ تا ۸۴ سال						
۸۵ و بیشتر						
جمع						

* آمار جمعیت زنان ۱۵ تا ۴۹ ساله ، ستون سمت راست (سفید) ، کل جمعیت زنان و ستون سمت چپ (رنگی) جمعیت زنان شوهر دار است .

۲- جدول استفاده از نمک یددار

در سرشماری ابتدای هر سال تعداد خانوارهای استفاده کننده از نمک یددار مشخص ودر این جدول ثبت می گردد.

استفاده از نمک ید دار

منطقه	تعداد کل خانوار	تعداد خانوارهایی که از نمک ید دار استفاده میکنند
شهر/روستای اصلی ☐		
◯ قمر		

۳- جدول تولد

نتیجه و مشخصات هر زایمان در طول سال ، توسط بهورز در این جدول با چوب کبریت نمایش داده می شود.

تولد بر حسب وزن و جنس نوزاد ،سن مادر و شرایط زایمان

منطقه	مرده بدنیاآمده		زنده بدنیاآمده		وزن هنگام تولد و جنس نوزاد زنده بدنیاآمده				سن مادر در این تولد زنده						شرایط زایمان			
					کمتر از ۲۵۰۰ گرم		۲۵۰۰ گرم و بیشتر		وزن نشده		۱۰ تا ۱۴	۱۵ تا ۱۹	۲۰ تا ۲۴	۲۵ تا ۲۹	۳۰ تا ۳۴	۳۵ تا ۴۹	بر بیمارستان یا زایشگاه	در منزل زائو توسط		
	پسر	دختر	پسر	دختر	پسر	دختر	پسر	دختر	پسر	دختر								ماماي تحصيلكرده	ماماي دوره ديده	ماماي نديده
شهر/روستای اصلی ☐																				
قمر ◯																				

۴- جدول پوشش تنظیم خانواده

در روز آخر هر فصل تعداد استفاده کنندگان ازوسایل مختلف پیشگیری ازبارداری به تفکیک وسیله در این جدول نوشته می شود.

پوشش تنظیم خانواده بر حسب نوع روش در آخرین روز هر فصل

منطقه و روش / زمان		۳۱ خرداد	۳۱ شهریور	۳۰ آذر	۲۹ اسفند
شهر / روستای اصلی ☐	قرص				
	کاندوم				
	آی - یو - دی				
	بستن لوله در: زن				
	بستن لوله در: مرد				
	آمپول				
	نورپلانت				
	سایر				
	طبیعی				
قمر ○	قرص				
	کاندوم				
	آی - یو - دی				
	بستن لوله در: زن				
	بستن لوله در: مرد				
	آمپول				
	نورپلانت				
	سایر				
	طبیعی				

۵- جدول مرگ مادران باردار

در طول سال هر گونه مرگ مادر در اثر عوارض بارداری در این جدول با چوب کبریت نمایش داده می شود.

مرگ مادران به دلیل عوارض حاملگی و زایمان بر حسب سن مادر و علت مرگ

منطقه و علت مرگ / سن (سال)		۱۰ تا ۱۴	۱۵ تا ۱۹	۲۰ تا ۲۴	۲۵ تا ۲۹	۳۰ تا ۳۴	۳۵ تا ۳۹	۴۰ تا ۴۴	۴۵ تا ۴۹
شهر / روستای اصلی □	خونریزی								
	عفونت بعد از زایمان								
	مسمومیت حاملگی								
	بیماریهای قلبی								
	سایر علل								
قمر ○	خونریزی								
	عفونت بعد از زایمان								
	مسمومیت حاملگی								
	بیماریهای قلبی								
	سایر علل								

۶- جدول مرگ بر حسب سن وجنس

در این جدول هر گونه مرگ اتفاق افتاده در روستا در گروههای سنی مربوطه با چوب کبریت نمایش داده می شود.

مرگ بر حسب سن وجنس

منطقه	سن / جنس	کمتر از یکماه	یکماه تا یکسال	۱ تا ۴	۵ تا ۹	۱۰ تا ۱۴	۱۵ تا ۱۹	۲۰ تا ۲۴	۲۵ تا ۲۹	۳۰ تا ۳۴	۳۵ تا ۳۹	۴۰ تا ۴۴	۴۵ تا ۴۹	۵۰ تا ۵۴	۵۵ تا ۵۹	۶۰ تا ۶۴	۶۵ تا ۶۹	۷۰ تا ۷۴	۷۵ تا ۷۹	۸۰ تا ۸۴	۸۵ و بیشتر	جمع
شهر/ روستای اصلی □	مرد																					
	زن																					
قمر ○	مرد																					
	زن																					
کل	مرد																					
	زن																					

۷- جدول علت مرگ در کودکان کمتر از ۵ سال

در این جدول هر گونه مرگ اتفاق افتاده در کودکان زیر ۵ سال و در قسمت علت مربوطه با چوب کبریت نمایش داده می شود .

علت مرگ در کودکان کمتر از پنج سال

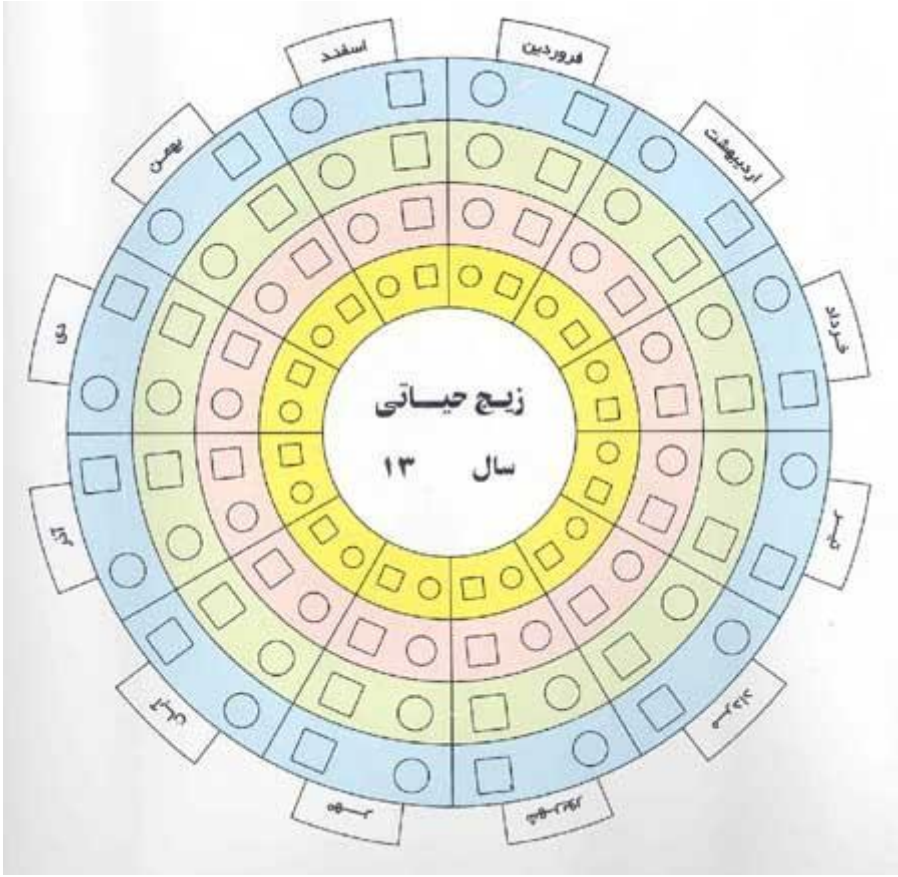
	علت مرگ / منطقه و سن	عفونتهای تنفسی	اسهال و استفراغ	حوادث، مسمومیتها، سرخکیها	عوارض کمبود وزن هنگام تولد *	نارسی نوزاد **	بیماریهای قابل پیشگیری با واکسن	مرگ نوزاد از صدمات زایمانی	ناهنجاریهای مادرزادی	سایر علل
شهر/ روستای اصلی □	کمتر از یکماه									
	یکماه تا کمتر از یکسال									
	یکسال تا کمتر از پنج سال									
قصر ○	کمتر از یکماه									
	یکماه تا کمتر از یکسال									
	یکسال تا کمتر از پنج سال									

* نوزادی که بعد از هفت سی و هفتم حاملگی بدنیا آمده است و وزن هنگام تولد او کمتر از ۲۵۰۰ گرم باشد (نوزاد رسیده) .

** نوزادی که قبل از هفته سی و هفتم حاملگی بدنیا آمده باشد (نوزاد نارس) وزن کودک در اینجا اهمیت ندارد .

۸- گردونه تولد و مرگ

در آخر هر ماه تعداد تولد و مرگ اتفاق افتاده در روستا با از روی جداول تولد و مرگ شمارش شده و در این گردونه ثبت می شود.



دهها شاخص جمعیتی، باروری، مرگ و میر و خدمات رسانی بهداشتی را می توان از مجموع داده های ذیج حیاتی مناطق روستائی بر حسب شهرستان یا استان و مناطق روستائی کل کشور محاسبه کرد.

رشد طبیعی جمعیت

میزان تولد خام

میزان مرگ خام

میزان مرگ نوزادی

میزان مرگ کودکان زیر یکسال

میزان مرگ کودکان زیر ۵ سال

میزان مرگ و میر مادران باردار

پوشش تنظیم خانواده

انواع شاخص های بهداشتی

۱- شاخص های مربوط به باروری
۲- شاخص های مربوط به مرگ و میر
۳- شاخص های مربوط به بیماری ها
۴- شاخص های مربوط به سنجش وضعیت خدمات رسانی در بخش بهداشت و درمان
۵- شاخص های اقتصادی، اجتماعی و فرهنگی که در ارتباط بیشتری با بهداشت هستند.
۶- شاخص های جمعیتی
در بخش های بعدی به جزئیات این شاخص ها می پردازیم.

میزان ها، نسبت ها، ا عداد مطلق

برخی ازشاخص هاهمان شکل کمیت های مطلق و فراوانی ساده دارند مثلا اینکه در کشور چند دانشگاه علوم پزشکی وجود دارد . چند بیمارستان وجود دارد ویا چند نفر جمعیت زندگی می کنند. همه از نوع شاخص های ساده هستند که به صورت اعداد وارقام بیان می شود.
برخی دیگر شاخص هائی هستند که به صورت نسبت دو مقدار که بیان کننده وضعیت دو چیز متفاوت از دو جامعه می باشد بیان می شوند مثلا هزینه یا درآمد یک کشور به نسبت جمعیت آن سنجیده شود.
نوع سوم نسبت هائی هستند که صورت و مخرج کسر حاصل از یک جمعیت آماری بوده وصورت بخشی از مخرج کسر به شمار می رود. به این گروه از نسبت ها معمولا میزانها می گویند که کاربرد های بیش تری در علوم پزشکی و بهداشت دارند.

شاخص های مربوط به باروری

تولد یکی از وقایع مهم حیاتی است که تعیین توزیع آن در برنامه ریزی ، اجرای خدمات بهداشتی و درمانی به ویژه بهداشت مادر و کودک و تنظیم خانواده حائز اهمیت است . بر اساس اطلاعات مربوطه به مادر و اطلاعات مربوط به نوزاد و اطلاعات جامعه ای که تولد ها در آن اتفاق می افتند مطالعات مختلفی انجام می گیرد . کمیت و کیفیت باروری در جامعه مبتنی بر شاخص هایی است که مبنای آن ها تعداد تولد ها و ابعاد آن جامعه است . در این قسمت به چند شاخص مهمتر اشاره می کنیم .

۱- میزان خام ولادت : Crude birth rate

این میزان نشان دهنده آن است که به ازای هر هزار نفر جمعیت یک منطقه در یک سال جند تولد زنده واقع شده است و به صورت زیر محاسبه می شود

$$\text{میزان خام C.B.R.ولادت} = \frac{\text{تعداد تولد ها در یک سال در یک منطقه}}{\text{جمعیت وسط سال}} \times ۱۰۰۰$$

با توجه

به اینکه تولد ها از جمعیت زنانی است که در امر باروری نقش دارند بخشی از جمعیت وسط سال یک منطقه در وقوع این اتفاق نقشی نخواهد داشت لذا این شاخص را خام می گویند چون ممکن است نسبت این عده درجمعیتهای مختلف مناطق یکسان نباشد مقایسه میزان خام ولادت برای مناطق مختلف همواره نتیجه درستی نخواهد داد. لذا برای ارائه شاخص بهتر جمعیتی که در ایجاد ولادت موثرتر هستند مورد توجه قرار گرفتند که جمعیت زنان در سنین باروری گروه موثرتری هستند بنابراین شاخص زیر معرفی شده است .

۲- میزان باروری عمومی : General fertility rate

فرمول این میزان به صورت زیر است :

$$\text{میزان باروری عمومی G.F.R.} = \frac{\text{تعداد تولد ها در یک سال در یک منطقه}}{\text{جمعیت وسط سال زنان در سنین باروری}} \times ۱۰۰۰$$

این میزان نشانگر آن است که به ازای هر ۱۰۰۰زن در سنین باروری چند تولد واقع شده است .

نکته : اگر چه نقش اصلی در تولد ها مربوط به زنان در سنین باروری است ولی با نگاه آماری دقیقتر باید گفت که همه زنان در سنین باروری ممکن است تن به امر زناشویی نداده و هنوز در وقوع ولادت موثر نبوده اند و چون این نسبت در جوامع مختلف یکسان نیست بهتر است که میزان باروری عمومی را برای زنان ازدواج کرده تعیین نمود .

۲- میزان باروری عمومی نکاحی :

$$\text{میزان باروری عمومی نکاحی} = \frac{\text{تعداد تولد ها در یک سال در یک منطقه}}{\text{جمعیت وسط سال زنان همسردار در سنین باروری}} \times ۱۰۰۰$$

۳- میزان باروری اختصاصی سنی: Age specific fertility rate

با توجه به این که زنان در سنین باروری دارای وضعیت متفاوتی از نظر ازدواج و تولد نوزاد هستند برای ملاحظه اثر سن در تعیین وضعیت باروری, میزان باروری اختصاصی سنی به صورت زیر

$$\text{میزان باروری اختصاصی سنی} = \frac{\text{تعداد تولد ها در یک سال مربوط به زنان یک سن یا گروه سنی خاص}}{\text{جمعیت وسط سال زنان همان سن یا گروه سنی در همان سال}} \times ۱۰۰۰$$

محاسبه می شود که برای هر سن یا گروه سنی خاص (مثلا گروه ۵ ساله) قابل محاسبه است:

اگر آمار زنان متال در هر سن یا گروه سنی ر اختیار باشد این میزانها را می توان برای زنان متاهل نیز محاسبه کرد.

۳- میزان باروری کل : Total fertility rate

اگر میزانهای باروری اختصاصی سنی را برای کلیه سنین باروری محاسبه نمائیم و مجموع آنها را برای همه سنین باروری حساب کنیم به حاصل جمع آنها میزان باروری کل می گویند. چون میزانهای باروری اختصاصی سنی در ۱۰۰۰ زن محاسبه می شود این میزان باروری کلی نیز به ازای ۱۰۰۰ زن خواهد بود. اگر آن را بر ۱۰۰۰ تقسیم نمائیم میزان باروری کلی برای یک زن محاسبه می شود. این میزان نشان دهنده آن است که هر زن به طور متوسط در دوران باروری خود چند فرزند به دنیا می آورد (با این فرض که روند باروری در دوران باروری او وضعیت ثابتی ماند جامعه فعلی داشته باشد).

= میزان باروری کل مجموع میزانهای باروری اختصاصی سنی برای کلیه سنین باروری(۱۵ تا ۴۹ سال) تقسم بر ۱۰۰۰

(اگر میزانهای باروری اختصاصی سنی برای گروه های ۵ ساله محاسبه شدند باید مجموع میزانها گروه ها در ۵ ضرب شود به عبارت دیگر از ۵ فاکتور گرفته می شود.)

۶- میزان تجدید نسل: Reproduction rate

اگر بخواهیم در باره وضعیت آینده باروری در جامعه پیش بینی کنیم شاخص تجدید نسل مورد استفاده قرار می گیرد. متوسط تعداد فرزندانی که در دوران باروری از یک مادر متولد می شوند در جامعه به صورت زوج در امر باروری آینده سهیم خواهند بود. اگر فقط یک گروه از آنها یعنی دختران را در نظر بگیریم و متوسط تعداد دخترانی که از یک مادر متولد می شوند را محاسبه کنیم به آن میزان تجدید نسل می گویند. برای محاسبه آن کافی است که نسبت دختر زائی در جامعه را در میزان باروری کل ضرب کنیم.

نسبت دختر زائی برابر است با تعداد متلدین دختر تقسیم بر تعداد کل متولدین آن جامعه در آن سال. که ممکن است تقریبا ۵/. در نظر گرفته شود.

چون در هر جامعه ممکن است به دلیل مرگ و میر همه دخترانی که متولد می شوند به سنین باروری نرسند لذا اگر اثر مرگ ومیر را در نظر نگیریم شاخص زیر محاسبه می شود:

۶-۱- میزان تجدید نسل ناخالص : Gross reproduction rate

نسبت دختر زائی × میزان باروری کل = میزان تجدید نسل ناخالص

۶-۲- میزان تجدید نسل خالص : Net reproduction rate

اگر اثر مرگ ومیر را در متوسط تعداد دخترانی که از یک مادر متولد می شوند در نظر نگیریم یعنی متوسط تعداد دخترانی که از یک مادر متولد می شوند و احتمال دارد تا زنده باشند و به سنین باروری برسند (یعنی جایگزین مادرشان شوند) را محاسبه نمائیم میزان تجدید نسل خالص می گویند.

اگر در محاسبه میزان باروری کل برای میزانهای باروری اختصاصی سنی نسبت دختر زائی در همان سن و احتمال زنده ماندن متولدین دختر تا آن سن را از جدول عمر محاسبه نمایند با محاسبه مجموع آنها میزان تجدید نسل خالص به دست می آید. ولی برا ی پرهیز از محاسبات پیچیده ممکن است احتمال بقا یعنی زنده ماندن متولدین دختر تا متوسط سنین باروری یعنی ۲۹ سالگی را در محاسبات به کار برد که تقریب خوبی از میزان تجدید نسل خالص خواهد بود.

میزان تجدید نسل خالص = احتمال زنده ماندن × نسبت دخترزائی × میزان باروری کل

شاخص های مربوط به مرگ ومیر

اگر چه واقعه مرگ برای هر فرد یک بار رخ می دهد ولی مطالعه فراوانی آن در جامعه می تواند وضعیت بهداشتی ودر مانی آن جامعه را مورد ارزیابی قرار دهد. اندازه گیری مرگ ومیر بر حسب شرایط افرادی که فوت می کنند و همینطور علت مرگ متفاوت است لذا شاخص های گوناگونی برای این منظور ارائه شده است که مهم ترین آنها به شرح زیر می باشد.

۱- **میزان خام مرگ ومیر** Crude death rate

$$\text{میزان خام مرگ ومیر C.D.R.} = \frac{\text{تعداد موارد مرگ در یک سال در یک منطقه}}{\text{جمعیت وسط سال همان منطقه در آن سال}} \times ۱۰۰۰$$

برای نشان دادن میزان شدت اثر واقعه مرگ ومیر در جامعه به شکل کلی از این میزان استفاده می شود که تعداد مرگ و میر به ازای هر ۱۰۰۰ نفر جمعیت یک منطقه را در سال نشان می دهد وبرای محاسبه آن از فرمول زیر استفاده می شود:

چون ساختمان سنی و جنسی و همینطور علت مرگ در مناطق کشور های گوناگون یکسان نیست پس مقایسه وضعیت مرگ ومیر با میزان خام کار درستی نیست لذا باید برحسب هر یک از متغیر های سن, جنس, علت و .... میزانهای مرگ ومیر را به صورت اختصاصی تعریف کرد.

۲- **میزان مرگ ومیر اختصاصی سنی** Age specific death rate

این میزان از تقسیم تعداد مرگ ومیر واقع شده در هر سن یا گروه سنی خاص بر تعداد جمعیت وسط سال همان سن یا گروه سنی به دست می آید:

$$\text{میزان مرگ ومیراختصاصی سنی A.S.D.R.} = \frac{\text{تعداد مرگ درهر سن یا گروه سنی خاص در یک سال}}{\text{جمعیت وسط سال همان سن یا گروه سنی در آن سال}} \times ۱۰۰۰$$

غالبا این میزان را علاوه بر سن بر حسب جنس نیز اختصاصی می کنند که به صورت زیر تعریف می شود:

۳- میزان مرگ ومیر اختصاصی سنی و جنسی

$$\text{میزان مرگ ومیراختصاصی سنی وجنسی} = \frac{\text{تعداد مرگ هرسن یا گروه سنی خاص درهرجنس یک سال}}{\text{جمعیت همان سن یا گروه سنی خاص درهرجنس یک سال}} \times ۱۰۰۰$$

مطالعه مرگ ومیر در جمعیت های آسیب پذیر مانند نوزادان, کودکان و مادران باردار اهمیت ویژه ای دارد لذا برای این گروه ها شاخص های جدا گانه ای تعریف شده است که در این قسمت به آنها اشاره می شود:

۴- میزان مرگ ومیر کودکان زیر یک سال Infant mortality rate

با توجه به این که تعیین تعداد جمعیت وسط سال کودکان زیر یک سال معمولا با خطای بیشتری

$$\text{میزان مرگ ومیرکودکان زیر یک سال .I.M.R} = \frac{\text{تعداد مرگ ومیر کودکان زیر یک سال در سال}}{\text{تعداد موالید زنده همان سال درهمان منطقه}} \times ۱۰۰۰$$

همراه است به جای آن در مخرج کسر فوق از تعدادمتولدین زنده استفاده شده است. پس I.M.R.نشان دهنده این است که به ازای هر ۱۰۰۰ نفر نوزاد زنده متولد شده در سال چه تعداد در زیر یک سالگی فوت می کنند. هر چه قدر وضعیت بهداشتی در جامعه نامناسبتر باشد این میزان افزایش می یابد به عبارت دیگر تاثیر پذیری مرگ ومیر های این گروه از عوامل نامساعد بهداشت محیطی بیشتر است لذا این میزان برای ارزیابی وضعیت بهداشتی کشورها مورد استفاده بیشتری قرار می گیرد.

چون وضعیت مرگ ومیر در زیر یک سالگی نیز برای سنین مختلف یکسان نیست و بیشتر مرگ ومیر های کودکان در ماه اول زندگی اتفاق می افتد شاخص زیر جداگانه تریف می شود:

۵- میزان مرگ ومیر نوزادان زیر یک ماه

$$\text{میزان مرگ ومیرنوزادان زیر یک ماه} = \frac{\text{تعداد مرگ ومیر نوزادان کمتر از یک ماه در سال}}{\text{تعداد موالید زنده همان سال درهمان منطقه}} \times ۱۰۰۰$$

ممکن است مرگ ومیر های بین بیستمین هفته زندگی جنینی که به صورت مرده متولد می شوند و مرگ ومیر های تا پایان هفته اول پس از تولد را تحت عنوان مرگ ومیر های حول تولد (Perinatal) را به ازای هر ۱۰۰۰ تولد زنده محاسبه کرد.

همچنین میزان مرگ ومیر نوزادی (Neonatal) (۲۸-۰ روز) و بعد از آن (Posneonatal) (۳۶۵-۲۹ روز) نیز نشان دهنده وضعیت مرگ ومیر کودکان زیر یک سال به حساب می آیند. معمولا در همه این شاخص ها در مخرج کسر فوق از تعدادمتولدین زنده همان سال استفاده می شود.

۶- میزان مرگ ومیر مادران (Maternal mortality rate)

مرگ ومیر های مادران که ناشی از حاملگی و عوارض زایمان وبارداری باشد به وسیله این میزان مورد بررسی قرار می گیرد که معمولا به ازای هر ۱۰۰۰۰۰ تولد زنده محاسبه می شود :

$$\text{میزان مرگ ومیرمادران به دلیل بارداری وعوارض زایمان} = \frac{\text{تعداد مرگ مادران به دلیل بارداری وعوارض زایمان در سال}}{\text{تعداد موالید زنده همان سال درهمان منطقه}} \times ۱۰۰۰۰۰$$

۷- میزان مرگ ومیر اختصاصی علتی

$$\text{میزان مرگ ومیراختصاصی علتی} = \frac{\text{تعداد مرگ های ناشی از یک علت یا گروهی از علل در یک سال}}{\text{جمعیت وسط سال همان منطقه در آن سال}} \times ۱۰۰۰$$

در مخرج کسر هر چه قدر جمعیت در معرض خطر به جای کل جمعیت را مشخص نمائیم شاخص معتبر تری محاسبه می شود چون ممکن است به صورت طبیعی بعضی از افراد نسبت به مرگ ومیر از یک علت خاص مصونیت داشته باشند یا علت مورد نظر مربوط به گروه جنسی دیگری باشد.

۸- نسبت مرگ ومیر علتی

اگرنسبت مرگ ناشی ا زیک علت به تعدادکل مرگ ها سنجیده شود بهتراست آن را بادرصد بیان نموده و به صورت زیر محا سبه نماییم .

$$\text{نسبت میرائی یامرگ ومیرعلتی} = \frac{\text{تعداد مرگ های ناشی از یک علت یا گروهی از علل}}{\text{تعدادکل مواردمرگ درهمآن دوره زمانی}} \times ۱۰۰$$

معمولااین نسبت را برای سنین مختلف دردوجنس وبرحسب علل مختلف تعیین میکند .

۹- متوسط سن درهنگام مرگ

اگر مجموع سن افراد در هنگام مرگ بر تعداد آنها تقسیم شود متوسط سن در هنگام مرگ محاسبه می شود. اگر چه محاسبه این شاخص برای جامعه بزرگ با تقریب زیادی همراه است ولی در جمعیت های کوچکتر می توان این شاخص را محاسبه کرد. متوسط سن در هنگام مرگ تحت تأثیر ساختمان سنی و جنسی و شغلی افراد هر گروه قرار می گیرد.

۱۰- امید به زندگی در بدو تولد:

متوسط تعداد سالهائی که انتظار می رود یک فرد x ساله با توجه به شرایط مرگ ومیر موجود در جامعه از این سن به بعد زنده بماند را امید به زندگی در سن x می گویند و با e_x نشان می دهند. متوسط تعداد سالهائی که انتظار می رود یک نوزاد متولد شده با توجه به شرایط مرگ ومیر موجود در جامعه از این پس زنده بماند را امید به زندگی در بدو تولد می گویند و با e_0 نشان می دهند.

محاسبه این شاخص بر اساس میزانهای اختصاصی مرگ ومیر در کلیه سنین می باشد که با توجه به این میزانها احتمال مرگ (وهمینطور احتمال زنده ماندن یا بقا) در هر سنی را محاسبه می نمایند وسپس در یک گروه فرضی با احتمالهای زنده ماندن در هر سن مجموع شخص سالهائی که این گروه زنده خواهند ماند را بدست می آورند که با تقسیم آن بر تعداد افراد گروه فرضی, متوسط تعدادسالها یعنی امید به زندگی محاسبه می شود. این روش پایه محاسبات جدول عمر را تشکیل می دهد که معمولا در کتابهای جمعیت شناسی آماری توضیح کامل تری در باره آن داده می شود.

تطبیق میزان ها:

میزان های مرگ و میر خام در دو جامعه به دلیل اینکه توزیع سنی آنها یکسان نیست قابل مقایسه نیستند. برای مقایسه آنها باید میزان ها براساس توزیع سنی جامعه تطبیق داده شود و سپس میزان های تطبیق داده شده با هم مقایسه گردند.

برای این کار از دو روش مستقیم و یا غیر مستقیم استفاده می شود که محاسبات آنها بر اساس انتخاب جمعیت معیار یا میزان های معیار انجام می گیرد که در اینجا از بیان آنها خود داری می شود.

شاخصهای مربوط به بیما ریها

درمطالعات بهداشتی واپیدمیولوژیک از تجزیه وتحلیل بیماریها برحسب ویژگیهای شخص, مکان وزمان استفاده های زیادی به عمل می آید. واقعه تولد و مرگ برای یک فرد فقط یک بار رخ می دهد ولی ممکن است یک فرد چند بار به بیماری خاص یا بیماری های مختلف دچار شود وهمینطور دوره بیماری ممکن است کوتاه (acute) یا طولانی (chronic) باشد بنابر این شاخص های مربوط به بررسی وضعیت بیماری ها در جامعه ابعاد مشکل تری می گیرد.جدید یا قدیم بودن بیماری در زمان مطالعه عاملی برای معرفی میزانهای بروز یا شیوع است که از مهمترین میزانهای بررسی بیماری ها به شمار می رود. رابطه بین بیماری ومیزان مرگ ومیر نیز بعد دیگری است که در شاخص هائی مانند میزان کشندگی بیماری مورد توجه قرار گرفت. در این قسمت به میزانهای مذکور اشاره می کنیم.

۱- میزان بروز: (Incidence rate)

میزان بروز عبارت است از تعداد موارد جدید بیماری که در یک جمعیت معین در یک دوره زمانی مشخص (معمولا یک سال در نظر گرفته می شود) رخ می دهد.

هر چند تعیین افراد در معرض خطر مشکل تر با شد کل جمعیت آن منطقه که به

$$\text{میزان بروز بیماری} = \frac{\text{تعداد موارد جدید بیماری در یک دوره زمانی رخ می دهد}}{\text{جمعیت در معرض خطر}} \times ۱۰۰۰$$

صورت بالقوه در معرض خطر هستند به جای مخرج کسر محاسبه می شود.

۲- میزان شیوع (Prevalence rate)

یکی از مهمترین شاخص بررسی بیماری ها میزان شیوع است که نشان دهنده میزان موارد بیماری در زمان مطالعه است اعم از این که موارد بیماری در آن مقطع زمانی ایجاد شده باشد یا از قبل واقع شده باشد.

$$\text{میزان شیوع بیماری} = \frac{\text{تعداد موارد جدید و قدیم بیماری در یک دوره زمانی خاص}}{\text{جمعیت وسط آن دوره زمانی}} \times ۱۰۰۰$$

شیوع را برای یک دوره زمانی مثلا یک سال حساب می کنند که به آن شیوع دوره ای می گویند. اگر هر چه قدر این دوره زمانی کوتاه تر باشد و موارد جدید و قدیم بیماری در یک مقطع زمانی معین مانند هفته یا روز محدود شود شیوع مقطعی یا لحظه ای خوانده می شود.

۳- میزان (درجه) کشندگی بیماری

اگر نسبت مرگ ومیر های ناشی از یک بیماری را به کل موارد ابتلای به آن بیماری به دست آوریم میزان کشندگی یا درجه کشندگی آن بیماری می گویند که احتمال مرگ

ناشی از آن بیماری را در بین مبتلایان مشخص می کند و معمولا بر حسب درصد بیان می شود.

$$\text{میزان کشندگی بیماری} = \frac{\text{تعداد مرگ ناشی از یک بیماری}}{\text{تعداد مبتلایان به آن بیماری}} \times ۱۰۰۰$$

شاخص های مربوط به سنجش وضعیت خدمات رسانی در بخش بهداشت و درمان

در برنامه ریزی های بهداشتی و در مانی آگاهی از امکانات, تجهیزات, نیروی انسانی و سطح پوشش خدمات لازم وضروری است. برای این منظور ممکن است تعداد بیمارستانها یا تعداد تخت های بیمارستانی ویا تعداد نیروی انسانی مانند پزشک, ماما, کاردان یا کارشناس بهداشتی و سایر گروه های پیرا پزشکی را به عنوان شاخص های ساره در نظر گرفت ولی غالبا بهتر است آنها را به نسبت جمعیت کشور یا منطقه مورد مطالعه در نظر گرفت که در آن صورت به شکل نسبت ویا درصد بیان می شود ومحاسبه آنها چندان پیچیده نیست.

گروهی دیگر از شاخص های بهداشتی و در مانی وضعیت پوشش خدمات را مشخص می کنند. این خدمات ممکن است شامل فعالیت های ایمن سازی (واکسیناسیون), خدمات بهداشت مادر و کودک, خدمات بهداشت محیط یا حرفه ای باشد که در سطح جمعیت های مختلف ارائه می گردد. اگر تعداد افراد یا دفعاتی که این خدمات انجام می شود را بر جمعیتی که که می بایست از این خدمات برخوردار شوند تقسیم کنیم وبه صورت در صد بیان کنیم شاخص های معتبری ارائه کردیم که کاربرد های زیادی در برنامه ریزی ها خواهد داشت. که در این قسمت به چند نمونه از آنها اشاره می کنیم.

پوشش واکسیناسیون

درصد واکسن سرخک, ثلاث۳, هپاتیت۳,پولیو۳, ب ث ژ نسبت به موارد انتظار:

$$\frac{\text{تعداد واکسیناسیون انجام شده برای هر واکسن}}{\text{۱/۱۲ کودکان زیر یک سال}} \times ۱۰۰$$

پوشش تنظیم خانواده

درصد استفاده کنندگان از وسایل پیشگیری از حاملگی مانند قرص, کاندوم, آی یو دی, تزریق سه ماهه نورپلانت, بستن لوله.

$$\frac{\text{تعداد مصرف کنندگان از وسیله مورد نظر}}{\text{تعداد کل زنان واجد شرایط}} \times ۱۰۰$$

پوشش خدمات بهداشت محیط :

در صد خانوار هائی که از آب آ شامیدنی سالم استفاده می کنند:

$$\frac{\text{تعداد خانوار هائی که از آب آ شامیدنی سالم استفاده می کنند}}{\text{تعداد کل خانوارها}} \times ۱۰۰$$

شاخص های درمان:

یکی از شاخصهای مهم بررسی امکانات درمانی ضریب اشغال تخت است :

اگر مجموع تعداد بیماران بستری در ساعت معینی از روزهای مختلف هر ماه را بدست آوریم به آن تخت روز اشغالی می گویند. اگر حاصل جمع تعداد تختهای فعال روزانه را در طول ماه بدست آوریم به آن تخت روز کل می گویند.

$$\text{نسبت تخت روز اشغال شده (به درصد)} = \frac{\text{تخت روز اشغالی}}{\text{تخت روز کل}} \times ۱۰۰$$

شاخص های اقتصادی، اجتماعی و فرهنگی که در ارتباط بیشتری با بهداشت هستند.

۱- شاخص سرانه در آمد ناخالص ملی یا سرانه تولید ناخالص ملی که از تقسیم درآمد یا هزینه بر جمعیت کشور محاسبه می شود.

۲- نسبت سرباری

$$\text{نسبت سرباری} = \frac{\text{جمعیت صفر تا ۱۴ ساله + جمعیت ۶۵ ساله وبالاتر}}{\text{جمعیت ۱۵ ساله تا ۶۴ سال}} \times ۱۰۰$$

۳- میزان باسوادی

$$\text{نسبت باسوادی} = \frac{\text{تعداد جمعیت با سواد ۶ ساله وبالاتر}}{\text{جمعیت فعال (۱۵ تا ۶۴ ساله)}} \times ۱۰۰$$

$$\text{نسبت بیکاری} = \frac{\text{تعداد جمعیت افراد بیکار}}{\text{جمعیت فعال (۱۵ تا ۶۴ ساله)}} \times ۱۰۰$$

۴- میزان بیکاری

۵- مهاجرت خالص که با کم کردن تعداد مهاجرینی که از یک منطقه به جای دیگر مهاجرت کردند از تعداد مهاجرین به آن منطقه بدست می آید.

شاخصهای جمعیتی

رشد طبیعی جمعیت :

$$\text{رشد طبیعی جمعیت} = \frac{\text{تعداد مرگ ها دریک سال} - \text{تعداد تولد ها دریک سال}}{\text{جمعیت وسط سال همان منطقه در آن سال}} \times ۱۰۰۰$$

این فرمول حاصل تفاضل میزانهای خام موالید و مرگ است پس می توان نوشت:

در محاسبه میزان رشد طبیعی جمعیت اثر مهاجرت در نظر گرفته نشده است . اگر مهاجرت خالص

$$\text{رشد طبیعی جمعیت} = \text{میزان خام موالید} - \text{میزان خام مرگ}$$

در یک منطقه در تعیین رشد جمعیت آن منطقه منظور شود معمولا آن را میزان رشد جمعیت می گویند واز کلمه طبیعی خودداری می کنند.

برآورد جمعیت در سال های بین دو سرشماری یا پس از آن:

برای محاسبه جمعیت در بین دو سرشماری یا پس از هر سر شماری میزان رشد جمعیت را براساس فرمول های ریاضی و با فرض های مختلفی در نظر می گیرند. در این قسمت به سه روش مهمتر اشاره می کنیم.

الف- رشد جمعیت بر مبنای تصاعد عددی:

در این روش فرض می شود که هر سال جمعیت پایه به اندازه مقدار ثابت a تغییر می کند (اضافه یا کم می شود) که اگر بخواهیم بعد از n سال جمعیت اولیه p_0 را با این تغییر برآورد کنیم به صورت زیر عمل می کنیم:

$P_n= P_0+na$

ب- رشد جمعیت بر مبنای تصاعد هندسی:

اگر جمعیت با رشد سالانه r در سال مبدا برابر p_0 باشد پس از t سال برابر p_t خواهد بود که به صورت زیر محاسبه می شود:

$$p_t = p_0(1+r)^t$$

ج- رشد جمعیت بر مبنای تابع نمایی:

اگر جمعیت با رشد سالانه λ در سال مبدا برابر p_0 باشد پس از t سال برابر p_t خواهد بود که به صورت زیر محاسبه می شود:

$$p_t = p_0 e^{\lambda t}$$

زمان دو برابر شدن جمعیت یک منطقه یا کشور:

با توجه به این که رشد جمعیت در سالهای آینده بر اساس فرمول نمایی زیر توجیه مناسبتری دارد با فرض این که بخواهیم مدت زمان لازم را برای دو برابر شدن جمعیت محاسبه کنیم باید نسبت جمعیت پایان آن دوره زمانی را به جمعیت اولیه برابر ۲ در نظر بگیریم، بنابر این خواهیم داشت:

$$p_t = p_0 e^{\lambda t}$$

$$2 = \frac{p_t}{p_0} = e^{\lambda t} \Rightarrow \lambda t = \ln 2$$

$$t = \frac{\ln 2}{\lambda}$$

که در آن λ برابر رشد سالانه و نیز لگاریتم ۲ در پایه عدد نپر تقریبا برابر ۰/۷ در نظر گرفته می شود.

$$\ln 2 \cong 0.7$$

نمونه سؤالات فصل هشتم:

۱- میزان تجدید نسل خالص یعنی چه؟

الف- متوسط تعداد فرزندان یک زن ب- متوسط تعداد دختران یک زن

ج- متوسط تعداد دخترانی که یک زن در دوره باروری به دنیا می آ ورد.

د - متوسط تعداد دخترانی که یک زن در دوره باروری به دنیا می آ ورد واحتمال زنده ماندن آ نها تا سن باروری هست.

۲- در تعیین میزان رشد جمعیت کدامیک از توابع ریاضی زیر مناسب تر است؟

الف- تصاعد حسابی ب- تصاعد هندسی ج- تابع نمائی د- تابع خطی

۳- منظور از نرخ رشد جمعیت کدامیک از موارد زیر می باشد؟

الف- تعداد کل جمعیت در وسط سال ب- تعداد جمعیت اضافه شده در سال معین

ج- تعداد جمعیت اضافه شده در سال معین به ازای هر صد نفر د- تعداد موالید زنده ومهاجرت به داخل

- تعداد زنان ۴۹-۱۵ سال یک منطقه ۲۰ در صد جمعیت ۱۰۰۰۰ نفری آن جا را تشکیل می دهد ۰ اگر تعداد تولد های یک سال آن منطقه ۲۵۰ مورد باشد میزان باروری عمومی چقدر است؟

الف- ۲۵ در هزا ر ب- ۴۰ در هزا ر ج- ۱۲۵ در هزا ر د- ۵۰ در هزا ر

۶- با اطلاعات مسئله قبل میزان خام موا لید چقدر است؟

الف- ۲۵ در هزا ر ب- ۴۰ در هزا ر ج- ۱۲۵ در هزا ر د- ۵۰ در هزا ر

اگر جمعیت شهرستانی در میانه سال ۱۰۰۰۰۰ نفر وتعداد زنان در گروه سنی (۴۴-۱۵) سال برابر ۲۵۰۰۰ نفر وتعدا د متولد ین زنده در همان سال ۳۵۰۰ و تعداد کل مرگ ومیر ۸۰۰ مورد باشد که ۲۴۵ مورد از آن مربوط به کودکان زیر یکسال باشد به سؤالات زیر پاسخ دهید۰

۷- با توجه به ا طلاعات فوق میزا ن با روری عمومی در هزا رنفر برابر است با ؟

الف- ۱۴۰ ب- ۳۵ ج- ۷/۱ د- ۱۰۰

۸- با توجه به اطلاعات مذ کور میزان مرگ ومیر اطفال زیر یکسال (I.M.R) برابر است با

الف- ۷۰ ب- ۲/۴۵ ج- ۳۰۰ د- ۸

۹- میزان مرگ و میر خام در شهرستان مذ کور برابر است با ؟

الف- ۸۰۰ ب- ۴/۵ ج- ۱۰ د- ۸

۱۰- میزان رشد طبیعی را در هزا رنفر به دست آورید ۰

الف- ۲۷ ب- ۳۵ ج- ۳۰ د- ۴۳

۱۱- اگر در جمعیتی میزان تجد ید نسل خالص بیشتر ا ز یک باشد رشد جمعیت:

الف- منفی است ب- صفر است ج- مثبت است د- بیشتر از یک است

۱۲- اگر تعداد تولد های مربوط به کلیه زنان در یک سال را بر تعداد زنان سنین ۱۵تا ۴۹ سال تقسیم نماییم ودر هزار ضرب کنیم میزان محاسبه شده چه نام دارد؟

الف- میزان خام موالید ب- میزان باروری کل

ج- میزان باروری عمومی د- میزان باروری اختصاصی سنی

۱۳- اگر تعداد موالید یک سال B و تعدادمرگ ومیر همان سال D و جمعیت اول سال P0 باشد تعداد جمعیت وسط سال عبارت است از :

الف- $P=P0+B-D$ وسط سال ب- $P=P0+D-B$ وسط سال

ج- $P=P0+1/2(B+D)$ وسط سال د- $P=P0+1/2(B-D)$ وسط سال

۱۴- میزان وقوع یا بروز بیماری عبارت است از :

الف- تعداد اشخاصی که در یک مدت معین، جدیدا به بیماری مبتلا می شوند تقسیم بر تعداد جمعیت آن مکان در وسط آن مدت ضربدر عدد ثابت K

ب- تعداد اشخاصی که در یک مدت معین، به بیماری مبتلا می شوند تقسیم بر تعداد جمعیت بیماران منطقه ضربدر عدد ثابت K

ج- تعداد موارد جدید بیماری در یک مکان ودر زمان معین د- هیچ کدام

کدام یک از عبارات زیر درست وکدام یک نادرست است؟ به تر تیب با (د) و (ن) نشان دهید.
الف- در تعیین شاخصهای بهداشتی دو عامل جمعیت و تعداد وقایع مورد نظر است.
ب- بهترین برآورد جمعیت کشور در بین سالهای بین دو سر شماری استفاده از روش تصاعد حسابی است.
ج- میزان مرگ ومیر کودکان حساسیت بیشتری نسبت به سایر میزانها در نشان دادن وضعیت بهداشت جامعه است.
د- میزانها نوعی نسبت هستند.
ه- میزانها و نسبتها تفاوتی ندارند.
و- منظور از شاخصهای بهداشتی همان میزانها هستند.
ز- میزان بروز و وقوع بیماری یکی است.
ح- میزان بروز ومیزان شیوع بیماری یکی است.
ط- میزانهای مرگ ومیر در دو جامعه که دارای توزیع سنی متفاوتی هستند قابل مقایسه نخواهند بود.
ی- رشد سالانه جمعیت در هیچ کشوری نمی تواند منفی باشد.

کدام یک از موارد زیر از ویژگیهای میزان ها به شمار می روند؟
الف- محاسبات پیچیده ب- وجود دوره زمانی خاص
ج- کسری که صورت و مخرج آن از یک جامعه هستند. د- میزانها اعم از نسبتها هستند.

فرمول شاخصهای زیر را بنویسید.
میزان بروز یک بیماری
میزان تجدید نسل خالص
میزان تجدید نسل ناخالص
میزان باروری کل
میزان مهاجرت خالص
رشد طبیعی جمعیت

۹- اطلاعات زیر مربوط به یک شهرستان می باشد.

جمعیت آخر شهریور ۱۰۰۰۰۰ نفر تعداد تولد زنده در سال ۳۵۰۰ مورد

تعداد مرگ در سال ۴۵۰ مورد تعداد زنان ۱۵ تا ۴۴ سال ۲۵۰۰۰ نفر

تعداد مرگ مادران به دلیل بارداری و زایمان ۱ نفر تعداد مرگ نوزاد ۲۵ مورد

تعداد مرگ ناشی از سرطان ۵۰ نفر

با استفاده از اطلاعات فوق شاخصهای زیر را محاسبه کنید:

۹-۱- میزان مرگ ومیر خام

۹-۲- میزان مرگ ومیر نوزادان

۹-۳- میزان باروری عمومی

۹-۴- میزان مرگ مادران به دلیل بارداری وزایمان

۹-۵- میزان مرگ ومیر اختصاصی علیتی سرطان

۹-۶- نسبت مرگ ومیر علیتی سرطان

۹-۷- میزان موالید خام

۹-۸- میزان رشد سالانه جمعیت

جدول عمر زیر را کامل کنید:

گروه های سنی	جمعیت	تعداد مرگ	m	p	q	l
۰ ساله	۱۰۰	۱۰				۱۰۰۰
۱-۴	۴۰۰	۸				
۵-۹	۴۵۰	۳				
۱۰-۱۴	۳۵۰	۲				

فهرست منابع فارسی:

۱- کاظم محمد, ملک افضلی حسین, نهاپتیان وارتکس. روشهای آماری و شاخصهای بهداشتی, چاپ هشتم ۱۳۸۱ .

۲- واین و دانیل- مترجم: آیت اللهی سید محمد تقی, اصول وروش های آمار زیستی, امیر کبیر, ۱۳۶۳.
۳- گوری باتاچاریا, ریچارد جانسون – مترجم: ابن شهر آشوب مرتضی, میکائیلی فتاح, مفاهیم وروش های آماری, جلد اول، مرکز نشر دانشگاهی ۱۳۶۴.
۴- گوری باتاچاریا, ریچارد جانسون – مترجم: ابن شهر آشوب مرتضی, میکائیلی فتاح, مفاهیم وروش های آماری, جلد دوم، مرکز نشر دانشگاهی ۱۳۶۶.
۵- توماس, ووناکات- مترجم: مشکانی, محمد رضا, آمار مقدماتی, جلد اول, مرکز نشر دانشگاهی,۱۳۶۴.
۶- توماس, ووناکات- مترجم: مشکانی, محمد رضا, آمار مقدماتی, جلد دوم, مرکز نشر دانشگاهی,۱۳۶۷.
۷- بت داوسون ساندرز، رابرت ج تراپ، مترجم: سرافراز علی اکبر، غفارزادگان کامران، آمار پزشکی پایه و بالینی، انتشارات دانشگاه علوم پزشکی مشهد، ۱۳۷۶ .
۸- عبادی فرد آذر، کریمی جلال. شاخصهای بهداشتی، انتشارات دانشگاه علوم پزشکی اصفهان، ۱۳۷۵.
۹- دیلی بورک. مک گیلوواری. مترجمان: فقیه زاه سقراط، بمداد جاوید، ریع زاده بابک. تفسیر واستفاده از آمار پزشکی. دانشگاه شاهد، ۱۳۷۶.
۱۰- شیرانی پرویز. نظریه نمونه گیری. مرکز نشر دانشگاهی، ۱۳۶۴.
۱۱- جی ای پارک، ک پارک، مترجم: شجاعی تهرانی حسین، ملک افضلی حسین، درسنامه پزشکی پیشگیری واجتماعی, اصول وروشهای اپیدمیولوژی، جمعیت شناسی وآمارهای بهداشتی. انتشارات دانشگاه علوم پزشکی گیلان، ۱۳۷۲.
۱۲- جان ال فیلیپس، مترجم: ایزدی کاظم، تفکر آماری، انتشارات دانشگاه علامه طباطبائی، تهران، ۱۳۷۱.
۱۳- آیت الهی سیدمحمد تقی، مبانی آمار حیاتی، انتشارات امیر کبیر، تهران، ۱۳۷۲.
۱۴- گروه مؤلفین سازمان جهانی بهداشت، تحقیق در سیستمهای بهداشتی، انتشارات معاونت پژوهشی وزارت بهداشت، ۱۳۷۲.
۱۵- گروه مؤلفین: ملک افضلی حسین وهمکاران، روش شناسی پژوهشهای کاربردی در علوم پزشکی، انتشارات دانشگاه علوم پزشکی تهران، ۱۳۸۳.
۱۶- کای لای چانگ، نظریه مقدماتی احتمال وفرآیندهای تصادفی، مترجم: ابوالقاسم میامی، محمد قاسم وحیدی اصل، مرکز نشر دانشگاهی، چاپ اول، ۱۳۶۴.

۱۷- د دوگه، م ژیرو، تحلیل واریانس وطرح آزمایشها، مترجم: مشکانی علی، مرکز نشر دانشگاهی، چاپ اول، ۱۳۶۵.
۱۸- آر دی مورگان، مقدمه ای بر روشهای نمونه گیری، مترجم: کامران سپهری، چاپ فاروس ایران، ۱۳۵۴.
۱۹- شرکت آمار پردازان، راهنمای کاربران SPSS 6. For Windows ، مرکز فرهنگی انتشاراتی حامی، ۱۳۷۷.
۲۰- رولان پرسا، جمعیت شناسی آماری، مترجم: سید میرزائی محمد، انتشارات آستان قدس رضوی، مشهد،۱۳۷۱.
۲۱- توردا لینوس، مبانی بررسی های نمونه ای، مترجم: صادقی حسن، بزرگ نیا ابوالقاسم، انتشارات آستان قدس رضوی، مشهد، ۱۳۶۸.
۲۲- بهبودیان جواد، آمار ناپارامتری، انتشارات دانشگاه شیراز، ۱۳۶۶.
۲۳- ریچارد شیولسون، استدلال آماری در علوم رفتاری: جلد دوم, استنباط آماری، مترجم: کیامنش علیرضا، انتشارات جهاد دانشگاهی، تهران، ۱۳۶۶.

فهرست منابع انگلیسی

Reference:

1. Snedecor G.W. Cochran W.G. Statistical methods. Iowa state university press, 8th edition, 1989.
2. Neter J. William W. Applied linear statistical models, regression, analysis of variance and experimental designs. Richard D. Irvin INC. 1985.
3. Feinstein A.R. Principles of Medical statistics. Chapman & Hall CRC press. Florida, 2002.
4. Campbell M.J. Statistics at square Two: understanding modern statistical applications in medicine, BMJ. London, 2001.
5. Statistics at square one.
6. Armitage P. Statistical methods in medical research. Blackwell Oxford forth edition , 1977.
7. Bland M. An introduction to medical statistics. Oxford, 3rd edition, 1989.

8. Bailar J.C . And Mosteller F. Medical uses of statistics NEJM books. Massachusetts medical society 1986.
9. Redmond C. and Colton T. Biostatistics in clinical trials. Wiley reference series in biostatistics . west Sussek, 2001
10. Hulley S. B. Cummings S.R. Browner W.S. Grady D. Hearst N, and Newman. Designing clinical research, second edition, Lippincott Williams & Wilkins, Philadelphia, 2001
11. Kramer M.S. Clinical epidemiology and biostatistics. Springer-Verlag, Berlin, 1988.
12. Glaser A. N. High – Yield Biostatistics. Lippincott Williams & Wilking New York, 2001.
13. Casella G. Berger R.L. Statistical inference, Wadsworth & Brooks, California, 1990.
14. Conover W.J. Practical nonparametric statistics. 2ed. John Willy & Sons, Inc. Canada, 1980.
15. Dixon W.J. and Massey F.J. Introduction to statistical analysis, 3rd ed. McGraw Hill, NY, 1969.
16. Mood A. and Graybill F. Introduction to the theory of statistics, McGraw Hill, NY, 1963.
17. Scheffe H. Analysis of variance, Wiley, NY, 1967.

جدول ارقام تصادفی

۳۹۶۳۴ ۶۲۳۴۹ ۷۴۰۸۸ ۶۵۵۶۴ ۱۶۳۷۹ ۱۹۷۱۳ ۳۹۱۵۳ ۶۹۴۵۹ ۱۷۹۸۶ ۲۴۵۳۷
۱۴۵۹۵ ۳۵۰۵۰ ۴۰۴۶۹ ۲۷۴۷۸ ۴۴۵۲۶ ۶۷۳۳۱ ۹۳۳۶۵ ۵۴۵۲۶ ۲۲۳۵۶ ۹۳۲۰۸
۳۰۷۳۴ ۷۱۵۷۱ ۸۳۷۲۲ ۷۹۷۱۲ ۲۵۷۷۵ ۶۵۱۷۸ ۰۷۷۶۳ ۸۲۹۲۸ ۳۱۱۳۱ ۳۰۱۹۶
۶۴۶۲۸ ۸۹۱۲۶ ۹۱۲۵۴ ۲۴۰۹۰ ۲۵۷۵۲ ۰۳۰۹۱ ۳۹۴۱۱ ۷۳۱۴۶ ۰۶۰۸۹ ۱۵۶۳۰
۴۲۸۳۱ ۹۵۱۱۳ ۴۳۵۱۱ ۴۲۰۸۲ ۱۵۱۴۰ ۳۴۷۳۳ ۶۸۰۷۶ ۱۸۲۹۲ ۶۹۴۸۶ ۸۰۴۶۸

۸۰۵۸۳ ۷۰۳۶۱ ۴۱۰۴۷ ۲۶۷۹۲ ۷۸۴۶۶ ۰۳۳۹۵ ۱۷۶۳۵ ۰۹۶۹۷ ۸۲۴۴۷ ۳۱۴۰۵
۰۰۲۰۹ ۹۰۴۰۴ ۹۹۴۵۷ ۷۲۵۷۰ ۴۲۱۹۴ ۴۹۰۴۳ ۲۴۳۳۰ ۱۴۹۳۹ ۰۹۸۶۵ ۴۵۹۰۶
۰۵۴۰۹ ۲۰۸۳۰ ۰۱۹۱۱ ۶۰۷۶۷ ۵۵۲۴۸ ۷۹۲۵۳ ۱۲۳۱۷ ۸۴۱۲۰ ۷۷۷۷۲ ۵۰۱۰۳
۹۵۸۳۶ ۲۲۵۳۰ ۹۱۷۸۵ ۸۰۲۱۰ ۳۴۳۶۱ ۵۲۲۲۸ ۳۳۸۶۹ ۹۴۳۳۲ ۸۳۸۶۸ ۶۱۶۷۲
۶۵۳۵۸ ۷۰۴۶۹ ۸۷۱۴۹ ۸۹۵۰۹ ۷۲۱۷۶ ۱۸۱۰۳ ۵۵۱۶۹ ۷۹۹۵۴ ۷۲۰۰۲ ۲۰۵۸۲

۷۲۲۴۹ ۰۴۰۳۷ ۳۶۱۹۲ ۴۰۲۲۱ ۱۴۹۱۸ ۵۳۴۳۷ ۶۰۵۷۱ ۴۰۹۹۵ ۵۵۰۰۶ ۱۰۶۹۴
۴۱۶۹۲ ۴۰۵۸۱ ۹۳۰۵۰ ۴۸۷۳۴ ۳۴۶۵۲ ۴۱۵۷۷ ۰۴۶۳۱ ۴۹۱۸۴ ۳۹۲۹۵ ۸۱۷۷۶
۹۱۹۴۲ ۹۸۰۷۲ ۸۶۰۱۳ ۷۵۴۶۷ ۵۲۹۲۵ ۰۷۹۷۳ ۸۲۰۰۲ ۹۶۸۲۲ ۵۰۷۹۶ ۶۱۸۸۵
۶۶۵۷۵ ۶۷۳۸۷ ۱۸۷۲۱ ۶۱۲۲۰ ۵۴۸۹۸ ۹۱۴۶۵ ۴۸۲۴۸ ۴۸۶۲۴ ۴۸۱۲۹ ۴۸۹۱۷
۹۱۸۱۶ ۴۵۲۷۶ ۲۸۷۷۵ ۹۹۱۳۲ ۳۹۷۶۱ ۴۹۳۱۴ ۰۳۷۸۵ ۱۲۱۹۳ ۸۴۲۹۹ ۸۸۳۷۸

۱۳۸۲۷ ۰۶۰۲۸ ۴۸۵۷۹ ۸۹۸۴۸ ۱۲۸۷۲ ۰۲۹۵۵ ۹۲۰۸۷ ۰۹۸۰۱ ۲۵۷۳۴ ۷۷۸۰۰
۴۷۱۲۱ ۸۸۶۳۸ ۸۷۲۰۲ ۵۳۶۶۵ ۴۰۱۷۰ ۸۱۹۱۶ ۰۶۳۱۶ ۰۱۱۷۸ ۰۳۴۰۵ ۲۴۰۲۸
۳۰۷۸۱ ۵۲۲۶۱ ۷۱۹۶۴ ۴۵۴۱۶ ۲۷۹۳۷ ۹۶۲۰۵ ۰۱۷۶۰ ۴۳۹۹۴ ۸۴۷۵۰ ۸۶۵۵۸
۵۵۷۹۹ ۱۹۸۱۹ ۳۹۳۰۴ ۴۴۶۸۲ ۹۰۴۷۲ ۱۰۹۷۱ ۱۴۱۸۲ ۰۵۳۲۹ ۴۹۲۰۱ ۷۸۵۴۵
۶۳۹۳۷ ۹۵۴۵۲ ۲۵۴۹۸ ۰۴۱۲۶ ۱۴۶۲۲ ۳۰۹۴۱ ۳۵۶۸۶ ۸۲۷۸۰ ۶۴۶۲۳ ۱۴۹۶۹

۹۹۲۴۱ ۲۴۵۵۸ ۹۸۳۷۵ ۷۹۱۵۷ ۵۶۱۶۴ ۶۲۲۸۷ ۹۴۲۰۲ ۰۶۳۰۳ ۳۱۹۷۳ ۵۸۶۹۷
۰۹۲۵۸ ۴۹۸۳۳ ۹۴۴۹۰ ۲۲۴۰۰ ۰۵۷۶۴ ۷۲۱۴۶ ۰۱۹۰۷ ۹۱۵۷۹ ۴۶۴۳۸ ۳۸۴۴۹
۳۶۱۶۶ ۲۸۶۵۲ ۶۶۹۷۵ ۳۱۰۱۰ ۶۴۷۳۵ ۷۸۴۹۰ ۸۰۱۱۴ ۷۳۳۴۸ ۸۷۲۴۴ ۶۲۱۳۴
۱۳۳۱۷ ۹۴۶۱۷ ۷۴۶۷۴ ۴۹۹۵۳ ۲۷۹۰۴ ۴۹۰۶۷ ۲۶۱۲۸ ۶۵۰۳۰ ۱۳۳۴۷ ۷۲۷۴۹
۷۶۲۴۳ ۴۶۶۹۹ ۶۱۳۰۳ ۴۶۵۴۷ ۱۲۰۲۷ ۵۹۹۴۳ ۳۳۷۱۷ ۴۲۷۰۹ ۳۶۵۶۶ ۸۱۶۳۸

۸۷۴۷۴ ۸۹۴۲۲ ۳۲۵۰۱ ۲۹۵۴۱ ۲۳۴۲۸ ۷۵۰۳۰ ۸۹۵۴۳ ۱۰۳۴۲ ۷۹۶۷۰ ۴۶۵۷۴
۶۱۷۱۵ ۲۳۶۴۲ ۸۳۶۳۸ ۵۹۲۴۵ ۱۲۲۸۸ ۰۷۹۹۰ ۶۷۶۶۳ ۳۲۲۰۹ ۵۷۱۹۶ ۱۱۸۷۳
۶۲۹۴۳ ۳۶۶۰۲ ۳۱۶۹۰ ۱۴۶۳۴ ۱۹۴۷۲ ۰۱۷۹۱ ۲۳۰۹۶ ۰۹۹۴۹ ۷۲۷۷۸ ۱۳۸۶۲
۵۴۲۴۵ ۳۱۸۴۲ ۲۲۹۴۰ ۵۱۰۵۴ ۲۲۵۱۴ ۷۲۹۹۸ ۲۸۶۶۶ ۸۲۳۲۱ ۲۷۸۸۶ ۰۸۳۱۲
۶۱۱۸۵ ۴۱۳۴۰ ۲۳۹۰۴ ۳۲۸۸۲ ۰۵۴۹۴ ۳۵۱۶۳ ۹۱۲۹۴ ۹۴۶۶۴ ۴۴۴۳۰ ۱۱۰۷۱

۱۳۷۶۹ ۸۶۰۷۹ ۴۶۸۵۶ ۹۰۷۱۷ ۳۲۵۴۵ ۰۷۵۱۰ ۵۰۳۰۷ ۸۶۹۶۳ ۱۱۸۴۲ ۸۲۵۰۹
۲۸۲۰۷ ۵۸۱۹۴ ۰۹۳۷۰ ۶۹۴۴۴ ۸۵۷۳۸ ۹۳۹۴۸ ۵۸۹۱۰ ۸۰۳۱۴ ۶۷۳۴۱ ۰۷۴۲۶
۸۲۷۹۸ ۹۳۲۳۳ ۸۰۵۳۵ ۸۹۶۵۹ ۸۴۶۴۵ ۱۵۸۵۷ ۹۵۳۸۶ ۹۱۲۲۱ ۲۵۵۹۲ ۵۷۶۹۶
۷۸۹۷۸ ۹۹۶۲۹ ۲۵۹۵۴ ۷۴۹۲۰ ۸۳۱۱۷ ۰۲۶۸۷ ۹۰۷۴۰ ۴۸۵۲۱ ۸۹۸۱۰ ۰۸۰۷۴
۵۸۵۷۳ ۵۶۰۵۷ ۹۱۳۲۲ ۳۸۹۸۹ ۰۴۷۱۰ ۲۰۸۴۹ ۴۰۶۹۹ ۰۱۵۱۸ ۵۳۷۲۱ ۲۰۱۲۸

۴۵۲۰۵ ۹۴۵۱۸ ۵۵۴۲۴ ۳۸۷۵۲ ۶۱۳۵۷ ۹۲۹۸۷ ۷۹۴۴۶ ۸۳۲۰۸ ۲۷۱۵۷ ۰۰۱۹۰
۳۸۵۶۳ ۳۰۸۱۲ ۴۰۸۳۶ ۷۴۶۹۱ ۳۹۹۸۱ ۳۹۶۰۵ ۳۴۶۱۱ ۳۲۴۵۴ ۵۵۴۲۵ ۲۳۷۹۸
۵۵۴۳۹ ۰۴۹۶۱ ۶۳۶۲۴ ۸۴۸۶۱ ۳۶۹۵۶ ۴۳۸۹۰ ۳۹۰۵۵ ۶۸۲۲۲ ۵۷۹۹۵ ۸۵۳۰۶
۴۲۴۰۳ ۹۳۹۷۷ ۵۷۵۱۴ ۸۹۵۰۰ ۰۶۱۵۷ ۳۴۶۴۳ ۵۳۹۰۱ ۷۴۲۷۴ ۳۶۰۳۶ ۹۹۷۱۹
۳۸۷۴۳ ۲۸۲۴۹ ۴۳۳۹۵ ۱۱۸۸۰ ۴۰۲۶۹ ۵۸۳۴۷ ۰۰۷۸۴ ۴۸۸۷۳ ۸۱۴۵۲ ۹۵۹۷۰

۶۰۰۱۵ ۸۰۰۴۴ ۴۷۶۱۴ ۵۵۰۵۲ ۱۸۵۵۶ ۷۲۰۶۲ ۹۸۵۶۶ ۹۲۴۶۲ ۹۱۴۶۰ ۵۶۶۵۱
۱۷۰۳۷ ۳۴۰۱۴ ۷۹۱۰۰ ۵۵۲۴۹ ۵۵۲۷۲ ۴۷۵۵۶ ۶۷۳۳۷ ۳۵۷۵۰ ۸۰۲۲۰ ۷۱۴۹۹
۲۳۱۲۹ ۱۴۹۷۰ ۷۳۲۳۷ ۷۰۸۳۱ ۷۷۴۸۹ ۶۵۴۱۹ ۷۰۷۳۶ ۴۷۵۴۵ ۷۸۴۴۳ ۶۶۶۶۰
۴۷۲۲۷ ۸۸۸۲۲ ۴۷۵۳۷ ۵۹۷۸۳ ۲۴۸۵۳ ۵۴۶۱۹ ۸۸۶۱۸ ۷۹۹۵۶ ۸۴۵۶۳ ۳۵۴۸۳
۵۲۰۶۴ ۴۳۷۵۷ ۷۰۶۳۹ ۲۳۱۹۸ ۰۲۸۰۰ ۸۶۱۰۳ ۱۹۲۰۳ ۵۷۸۶۲ ۲۵۰۴۱ ۰۹۲۶۲

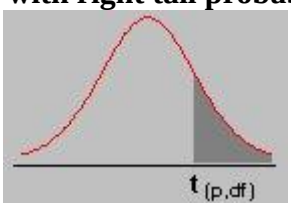
Standard Normal (*Z*) Table: Values in the table represent areas under the curve to the left of *Z* quantiles along the margins.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Examples: $z_{.5000} = 0.00$; $z_{.9750} = +1.96$; $z_{.0250} = -1.96$

t table with right tail probabilities

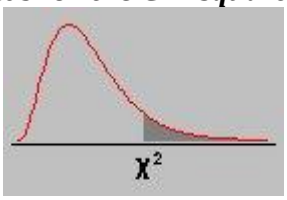


df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12

4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3

27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3

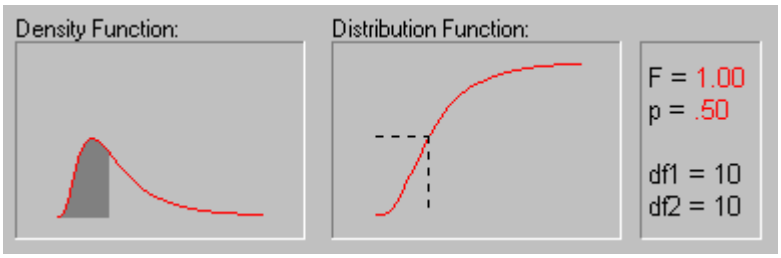
Right tail areas for the *Chi-square* Distribution



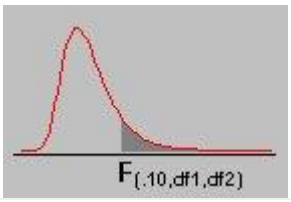
df\area	.995	.990	.975	.950	.900	.750	.500	.250	
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.10153	0.45494	1.32330	
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	1
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	1
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.21885	1
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.38875	1
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.54886	1

11	2.60322	3.05348	3.81575	4.57481	5.57778	7.58414	10.34100	13.70069	1
12	3.07382	3.57057	4.40379	5.22603	6.30380	8.43842	11.34032	14.84540	1
13	3.56503	4.10692	5.00875	5.89186	7.04150	9.29907	12.33976	15.98391	1
14	4.07467	4.66043	5.62873	6.57063	7.78953	10.16531	13.33927	17.11693	2
15	4.60092	5.22935	6.26214	7.26094	8.54676	11.03654	14.33886	18.24509	2
16	5.14221	5.81221	6.90766	7.96165	9.31224	11.91222	15.33850	19.36886	2
17	5.69722	6.40776	7.56419	8.67176	10.08519	12.79193	16.33818	20.48868	2
18	6.26480	7.01491	8.23075	9.39046	10.86494	13.67529	17.33790	21.60489	2
19	6.84397	7.63273	8.90652	10.11701	11.65091	14.56200	18.33765	22.71781	2
20	7.43384	8.26040	9.59078	10.85081	12.44261	15.45177	19.33743	23.82769	2
21	8.03365	8.89720	10.28290	11.59131	13.23960	16.34438	20.33723	24.93478	2
22	8.64272	9.54249	10.98232	12.33801	14.04149	17.23962	21.33704	26.03927	3
23	9.26042	10.19572	11.68855	13.09051	14.84796	18.13730	22.33688	27.14134	3
24	9.88623	10.85636	12.40115	13.84843	15.65868	19.03725	23.33673	28.24115	3
25	10.51965	11.52398	13.11972	14.61141	16.47341	19.93934	24.33659	29.33885	3
26	11.16024	12.19815	13.84390	15.37916	17.29188	20.84343	25.33646	30.43457	3
27	11.80759	12.87850	14.57338	16.15140	18.11390	21.74940	26.33634	31.52841	3
28	12.46134	13.56471	15.30786	16.92788	18.93924	22.65716	27.33623	32.62049	3
29	13.12115	14.25645	16.04707	17.70837	19.76774	23.56659	28.33613	33.71091	3
30	13.78672	14.95346	16.79077	18.49266	20.59923	24.47761	29.33603	34.79974	4

F Distribution Tables



F Table for alpha=.10 .



df2/df1	1	2	3	4	5	6	7
1	39.86346	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531

10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	
11	3.22520	2.85951	2.66023	2.53619	2.45118	2.38907	2.34157	
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.23410	
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.19313	
15	3.07319	2.69517	2.48979	2.36143	2.27302	2.20808	2.15818	
16	3.04811	2.66817	2.46181	2.33274	2.24376	2.17833	2.12800	
17	3.02623	2.64464	2.43743	2.30775	2.21825	2.15239	2.10169	
18	3.00698	2.62395	2.41601	2.28577	2.19583	2.12958	2.07854	
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.05802	
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	
22	2.94858	2.56131	2.35117	2.21927	2.12794	2.06050	2.00840	
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04723	1.99492	
24	2.92712	2.53833	2.32739	2.19488	2.10303	2.03513	1.98263	
25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.97138	
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	
27	2.90119	2.51061	2.29871	2.16546	2.07298	2.00452	1.95151	
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.94270	

29	2.88703	2.49548	2.28307	2.14941	2.05658	1.98781	1.93452
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98033	1.92692
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.87252
60	2.79107	2.39325	2.17741	2.04099	1.94571	1.87472	1.81939
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.76748

To index

inf	2.70554	2.30259	2.08380	1.94486	1.84727	1.77411	1.71672

F Table for alpha=.05 .

df2/df1	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355

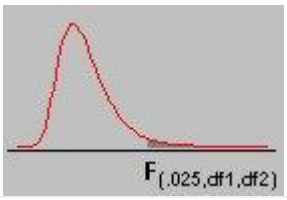
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	

30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	

To index

inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	

F Table for alpha=.025 .



df2/df1	1	2	3	4	5	6	7	
1	647.7890	799.5000	864.1630	899.5833	921.8479	937.1111	948.2169	
2	38.5063	39.0000	39.1655	39.2484	39.2982	39.3315	39.3552	
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	

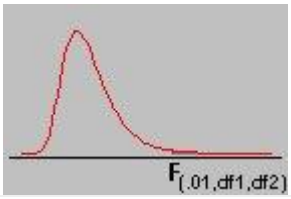
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	
23	5.7498	4.3492	3.7505	3.4083	3.1835	3.0232	2.9023	
24	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	
28	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	

29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	
60	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	
120	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	

To index

inf	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	

F Table for alpha=.01 .



df2/df1	1	2	3	4	5	6	7	
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	

9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3

28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	

To index

inf	6.635	4.605	3.782	3.319	3.017	2.802	2.639	

Mazandaran University of Medical Sciences

Concepts and Methods of Biostatistics & Health Indices

Reza Ali Mohammadpour (Ph. D.)
Mazandaran University of Medical Sciences