

به نام خدا

پروژه درس روش های آماری ترم دوم 1401-1402

سوال 1) داده های موجود در فایل STM.xlsx را فراخوانی کنید، این فایل داده های مربوط به بازیکنان برتر اروپا را به شما نشان میدهد که در این فایل شما در هر ستون اطلاعات خاص خود را مشاهده میکنید که بدون در نظر گرفتن ستون اول که اسم بازیکنان است

از ستون دوم به ترتیب سن بازیکنان (Age)، ملیت آن ها (Nationality)، پست تخصصی آن ها (Pos)، امتیاز فصل (Season Score)، طول قرارداد باقی مانده (Contract Length)، معروفیت (Popularity)، وزن (Weight)، قد (Height)، پای تخصصی (Foot)، قیمت یا ارزش بازیکنان در مارکت (Market Value)، حقوق سالانه (Wage)، لیگ (League)، وضعیت مالی تیم (TFS)، در لیست فروش قرار دارند یا خیر (Transfer List)

الف) خلاصه اطلاعات آماری شامل میانگین، واریانس، چارک ها را برای ستون های عددی به دست آورید و گزارش کنید.

ب) جدول فراوانی مربوط به ملیت، معروفیت، پای تخصصی، لیگ، وضعیت مالی تیم را گزارش کرده و تحلیل مناسب آن را بنویسید.

پ) نمودار های مربوط به امتیاز فصل، طول قرارداد، وزن و قد بازیکنان، قیمت بازیکنان، حقوق سالیانه آن ها، رسم نموده و در آخر نمودار مربوط به وضعیت مالی تیم و معروفیت را رسم کنید و تحلیل مناسب برای آن بنویسید.

ث) نرمال بودن ستون هایی که اطلاعات امتیاز فصل، قیمت و حقوق بازیکنان را با تمامی روش های ممکن بررسی کرده و نتیجه را برای هر کدام گزارش کنید.

ج) آزمون کنید که میانگین امتیاز فصل بازیکنان لیگ انگلستان بیشتر از 7.3 است یا خیر؟

د) آیا میانگین قد بازیکنان پست دفاع وسط (CB) از میانگین قد بازیکنان پست فولبک بیشتر است؟

ه) آزمون کنید آیا میانگین قیمت بازیکنانی که در لیست فروش هستند (Yes) با میانگین بازیکنانی که در لیست فروش نیستند (No) متفاوت است یا خیر؟

و) آزمون کنید که آیا میزان متوسط حقوق در سطوح لیگ متفاوت است؟

د) آزمون کنید که آیا میزان متوسط حقوق بازیکنان در پست های مختلف متفاوت از یکدیگر هست یا خیر؟

ر) اگر بازیکنان پست هافبک بازی ساز (Cam) و فورواردها (FW) دو جامعه مستقل از هم باشند آزمون کنید که آیا میانگین تعداد گل های فورواردها از هافبک های بازی ساز بیشتر است یا خیر و یک فاصله اطمینان 99٪ برای آن بدست آورید؟

ز) بازیکنان تمامی لیگ ها و با تمامی پست هارا در نظر بگیرید اگر این بازیکنان به دو جامعه مستقل از هم وضعیت عالی و خوب تقسیم شوند آزمون کنید که میانگین امتیاز فصل آن هایی که وضعیتشان عالی است اختلاف امتیاز 2 دارند و یک فاصله اطمینان 90 درصد برای آن به دست آورید؟

راهنمایی کلی در مورد این سوال: ابتدا برای هر آزمون باید توزیع را برای خودتان مشخص کنید و در تحلیلتان بنویسید سپس به سراغ روش های ناپارامتری یا پارامتری بروید.

در قسمت های (ج، ر، ز) آزمون را هم با روش های پارامتری و هم با روش ناپارامتری آزمون کنید و سپس تحلیل خود را بنویسید

سوال 2)

1000 عدد تصادفی از توزیع نرمال تولید کرده با میانگین 6 و واریانس 3 و در X بریزید سپس به ترتیب 700 و 500 عدد از توزیع های باینومیال $(0.4, 35)$ و کای دو با پارامتر 4 تولید کنید در Y و Z بریزید.

الف) احتمالات زیر را با کمک از نرم افزار به دست آورید.

$$P(-1 < X < 2) = ?$$

$$P(X > 2.765) = ?$$

$$P(Y < 1.5) = ?$$

$$P(1 < Z < 2) = ?$$

ب) چندک های 0.1 و 0.35 و 0.6247 و 0.975 را برای دو بردار X و Y گزارش کنید.

سوال 3) در یک حلقه با تکرار 300 از نمونه تصادفی شبیه سازی شده با توزیع نرمال با میانگین 5 و واریانس 16 :

الف) با تمامی روش هایی که میدانید نرمال بودن $Y = \left(\frac{\bar{x} - 5}{\sqrt{16}}\right)$ را بررسی کرده و نتیجه را گزارش کنید.

ب) نشان دهید که $\left(\bar{x} - z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right)$ برای $\alpha = 0.1$ یک فاصله اطمینان 90 درصد برای میانگین جامعه است.

ج) اگر اندازه نمونه را در قدم اول به 500 سپس به 800 و در آخر به 960 افزایش دهیم فاصله اطمینان چه تغییری خواهد کرد؟