

Marking Notes on Rounding

- 0.25 pts. for each instance of incorrect rounding
- Default for parameter estimates is two decimal places unless specified otherwise, e.g. Q3 specifically requests that answers be reported to four decimal places
- P-values should be reported to exactly three decimal places unless $p < .001$ (no deduction for leading zeros, as APA wasn't specified; only rounding requirement)

Loading Packages:

```
install.packages(c("psych", "latex2exp", "pracma", "ggplot2"))
library(psych)
library(pracma)
library(ggplot2)
library(latex2exp)
```

Question 1 (3 pts.)

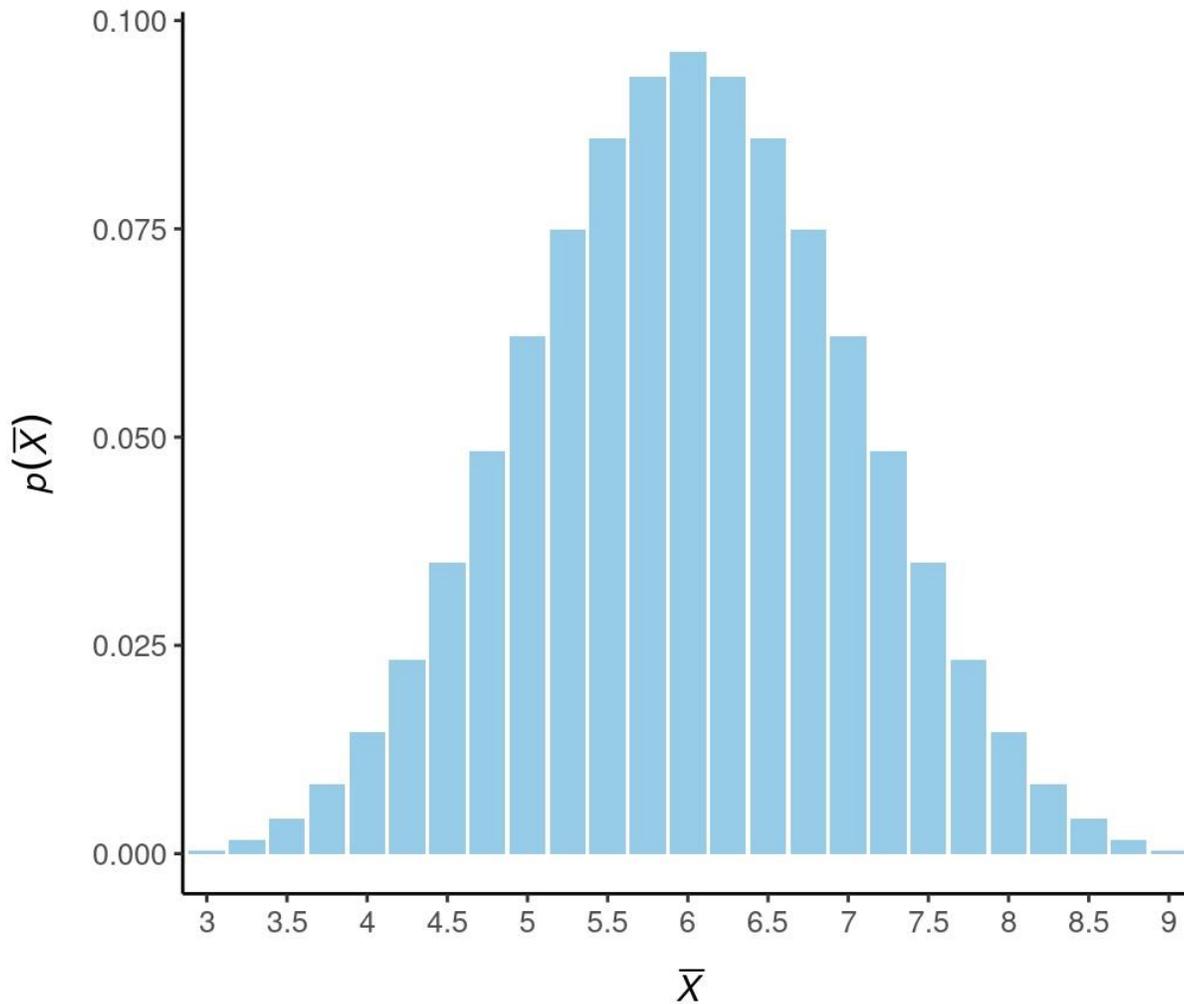
Given the population set of scores {3,4,5,6,7,8,9}, use base R graphics or ggplot to generate a bar graph representing the sampling distribution of mean for samples of $N = 4$. Assume sampling is one at a time with replacement. Give your graph a descriptive title, and be sure to include appropriate labels for the y- and x-axes.

Code2:

```
scores <- 3:9
samp_dist_m <- data.frame(expand.grid(rep(list(scores), 4)))
colnames(samp_dist_m) <- c('score.1', 'score.2', 'score.3', 'score.4')
samp_dist_m$mean <- rowMeans(samp_dist_m)
round(samp_dist_m$mean, 2)
t <- table(samp_dist_m$mean)
print(t)
pct <- as.numeric(t)/sum(t)
xlab <- names(t)
df <- data.frame(xlab, pct)
samp_dist_p <- ggplot(data = df, aes(x = xlab, y = pct)) +
  geom_bar(stat = "identity", fill = 'skyblue') +
  labs(x = TeX("$\\bar{\\textit{X}}$"),
       y = TeX("$\\textit{p}\\bar{\\textit{X}}$"),
       title = TeX("Sampling distribution of $\\bar{\\textit{X}}$ with $\\textit{N} = 4$")) + theme_classic() +
  scale_x_discrete(breaks = seq(from = 3, to = 9, by = 0.5)) +
  scale_y_continuous() +
  theme(axis.title.x = element_text(margin = unit(c(3.5,0,0,0),
                                                    "mm"), size = 11),
        axis.title.y = element_text(margin = unit(c(0,3.5,0,0),
                                                    "mm"), size = 11),
        axis.text = element_text(size = 9.5))
samp_dist_p
```

Output:

Sampling distribution of \bar{X} with $N = 4$



Question 2 (2 pts.)

A professor has been teaching statistics for many years. His records show that the overall mean for final exam scores is 73, with a standard deviation of 11. The professor believes that this year's class is superior to his previous ones. The mean for final exam scores for this year's class of 68 students is 76. Using $\alpha = 0.05$ -tail, state your conclusion and report Z_{obs} and the associated p-value.

Code:

```

SE_mean<-11/sqrt(68)
zobt <-(76-73)/SE_mean
round(zobt,digits=2)
zcrit<-qnorm(0.05, lower.tail = FALSE)
abs(zobt)>=abs (zcrit)
p<-pnorm(q = 76,mean = 73, sd = SE_mean, lower.tail =FALSE)
round(p,digits=3)

```

Output:

```

[1] 2.25
[1] TRUE
[1] 0.012

```

Answer:

From the given information, it appears that the professor's current class outperformed the previous years' classes by a significant margin, and this difference in performance holds statistical significance at a 0.05 level of significance ($p < .05$). This implies that the likelihood of observing such a difference in performance due to chance is less than 5%. Rejecting the null hypothesis means that there is proof to support the alternative hypothesis that the current class has performed better than the previous years' classes. However, it's important to consider that statistical significance alone may not be indicative of practical significance. Therefore, it's advisable to assess the effect size to determine the practical significance of the gap in performance between the current and previous years' classes. The value of Z_{obt} was 2.25, which is greater than the critical value of $Z_{crit} = 1.64$. This implies that the null hypothesis can be rejected. The mean of final exam scores for this year's class was considerably higher than the overall mean, with a p-value of .012, indicating statistical significance.

Question 3 (4 pts.)

Pagano's (2013, p. 316) Practice Problem 12.2 presented hypothetical data testing a new gasoline additive. A large number of mileage measurements on the gasoline without the additive showed a mean of 24.7 miles per gallon and a standard deviation of 4.8. An experiment was performed in which 75 cars were tested using the gasoline plus the additive. The results showed a sample mean of 26.5 miles per gallon. To evaluate these data, a directional test with $\alpha = 0.05$ -tail was used. Suppose that before doing the experiment, the manufacturer wants to determine the probability that he will be able to detect a real mean increase of 2.0 miles per gallon with the additive if the additive is at least that effective.

a. If he tests 25 cars, what is the power to detect a mean increase of 2.0 miles per gallon? Report your answer to exactly four decimal places. (1 pt.)

Codes:

```

n <- 25
xobt <- 26.5
m <- 24.7
sigma <- 4.8
sem <- sigma / sqrt(n)
m_real <- 26.7
print(zcrit.gas <- qnorm(p = .05, lower.tail = FALSE))
print(mcrit <- (m + zcrit.gas*sem))
zobt.gas <- ((mcrit - m_real)/sem)
power <- round(pnorm(zobt.gas, lower.tail = FALSE), 4)
print(power)

```

Outout:

```

[1] 0.6695

```

b. If he increases the N to 75 cars, what is the power to detect a mean increase of 2.0 miles per gallon?

Report your answer to exactly four decimal places. (1 pt.)

Code:

```
SE_msev<-4.8/sqrt(75)
zcrit_gas<-qnorm(0.05, lower.tail = FALSE)
m_real<-(24 + 2)
print(mcrit_two<-24+zcrit_gas*SE_msev)
zobt_real_two<-(mcrit_two-m_real)/SE_msev
round(pnorm (zobt_real_two,lower.tail = FALSE), digits =4)
```

Output:

```
[1] 0.9752
```

c. Again using only 25 cars, what is the probability the manufacturer would make a type-II error rate if there was a mean increase of 2.0 miles per gallon? Report your answer to exactly four decimal places. (1 pt.)

Code:

```
n <- 25
xobt <- 26.5
m<- 24.7
sigma <- 4.8
sem <- sigma / sqrt(n)
m_real <- 26.7
print(zcrit_gas <- qnorm(p = .05, lower.tail = FALSE))
print(mcrit <- (m + zcrit_gas*sem))
zobt_gas <- ((mcrit - m_real)/sem)
power <- pnorm(zobt_gas, lower.tail = FALSE)
round((1 - power),digits=4)
```

Output:

```
[1] 0.3305
```

d. How many cars should he use if he wants to have a 95% chance of detecting a mean increase of 2.0 miles per gallon? Use $\alpha = 0.05$ -tail and report your answer to the nearest integer that satisfies the power requirement. (1 pt.)

Code:

```
Sigma<-4.8
m_real<-(24+2)
mean_poulation<-24
zcrit<-qnorm(p =0.95, lower.tail =F)
zobt.need<-qnorm(p =0.95, lower.tail =F)
numer<-Sigma*(abs(zcrit)+abs(zobt.need))
denom<-m_real- mean_poulation
N<-(numer/denom)^2
print(round(N,digits =0))
```

Output:

```
[1] 62
```

Answer:

Based on the provided information, the original answer is 62.34, which is below 0.5. Normally, we would round down, but if we want to achieve a power of at least 95%, we need to round up to 63 cars.