STAT 515 Spring 2025

Homework 6

Due: Monday, March 17 in class <u>OR</u> electronically by 5:00 PM on Tuesday, March 18

1. The weights of the bananas on sale at a grocery store are normally distributed with a known standard deviation of $\sigma = 15$ grams. We collect a sample of $n = 12$ bananas which has a sample mean of $\bar{X}_n = 135.6$ grams.
   a. Derive the 63% confidence interval (CI) for the mean weight of all bananas on sale at this grocery store.
   b. Derive the 75.4% CI for the mean weight of the bananas on sale.
   c. Derive the 91% CI for the mean weight of the bananas on sale.
   d. Give the interpretation of the confidence interval in part (c).
   e. What do you observe happens to the confidence interval as we increase the confidence level?
   f. What is the width of the 98% CI for the mean weight of all bananas on sale based on our sample?
   g. We collect a new sample of size $n = 33$ and construct a new 98% CI.
      i. What is the width of this interval?
      ii. How does the width of this new interval compare to your answer in part (f)? Explain why.
   h. We collect a new sample and construct a new 98% CI which has a width of 5.66. What is the sample size n of this new sample?

2. The number of customers entering a store per hour has a known standard deviation $\sigma = 7$. A statistician observes $n = 50$ different business hours and records the number of customers that entered the store. The sample mean from the 50 samples is $\bar{X}_n = 44.3$.
   a. Derive the 90% CI for the mean number of customers that enter the store per hour.
   b. Give the interpretation of the CI in part (a).
   c. Derive the 95% CI for the mean number of customers that enter the store per hour.

d. What is the width of the 99% CI for the mean number of customers per hour based on our sample?
e. If we collect 5000 different random samples of size $n = 50$ and construct 99% CI's, approximately how many of these CI's will contain the true mean number of customers per hour?
f. If we collect a new sample of size $n = 820$ and construct a 99% CI, what is the width of this new CI?
g. Give the 100% CI for the mean number of customers entering the store per hour and explain why this CI is not practical.

3. We are interested in determining how many people aged 18-25 are in favor of a proposed law to decriminalize marijuana possession of up to 3 ounces. Suppose we survey a random sample of $n = 48$ individuals between the ages of 18 and 25, and 36 of the respondents say they are in favor of the law.
   a. Obtain an appropriate 75% CI for the population proportion of 18 to 25-year-olds who favor the proposed law.
   b. Obtain an appropriate 95% CI for the population proportion of 18 to 25-year-olds who favor the proposed law.
   c. Give the interpretation of the CI in part (b) and discuss how it compares to the CI in part (a).
   d. If we collect 8000 different random samples of $n = 48$ young adults aged 18 to 25-year-olds, survey them about their opinion of this proposed law, and construct 95% confidence intervals using the method in parts (b), about how many of these CI's will capture the population proportion?
   e. Give the 100% CI for the population proportion of 18 to 25-year-olds who favor the proposed law and explain why this CI is not practical.

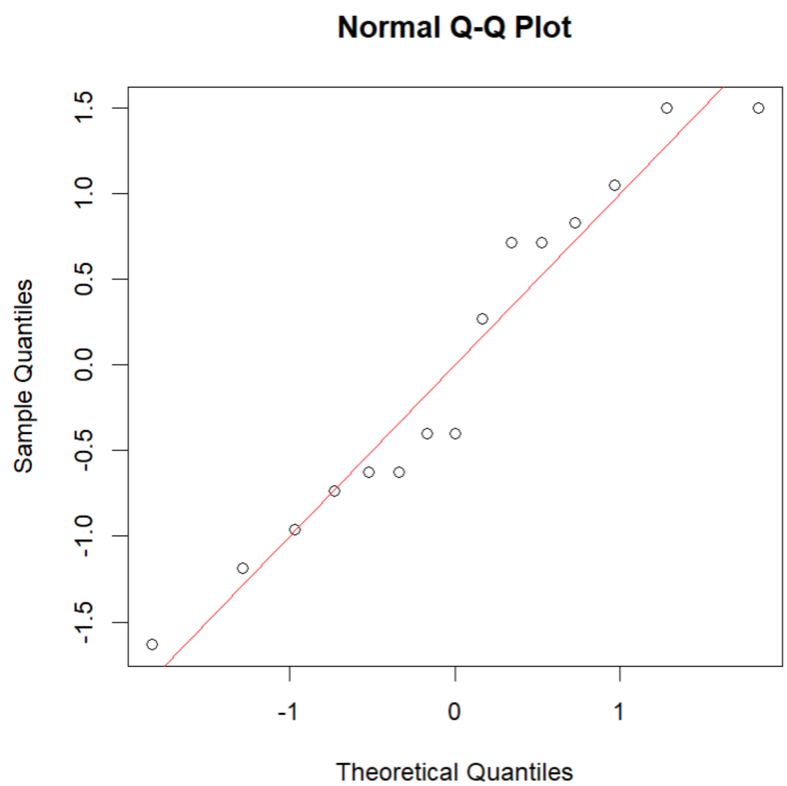
4. A random sample of $n = 242$ workers was collected for a study. The study found that 83 of the workers were employed in STEM professions.
   a. Obtain an appropriate 95% CI for the proportion of the workforce employed in STEM professions.
   b. Obtain an appropriate 82.6% CI for the population proportion of the workforce employed in STEM professions.

c. Give the interpretation of the CI in part (b) and discuss how it compares to the CI in part (a).
d. Based on our sample, what is the width of the 99% CI for the population proportion of the workforce employed in STEM professions?
e. If we increased our sample size to $n = 1210$ and 415 of the workers were employed in STEM professions, what is the width of the 99% CI for the population proportion? How does this compare to your answer in part (d)?

5. Suppose we have a random sample of $n = 15$ IQ scores from a certain population. The data is as follows:

$$\{\,88, 92, 94, 96, 97, 97, 99, 99, 105, 109, 109, 110, 112, 116, 116\,\}$$

The normal Q-Q plot for this data is given below.



a. What can we assume based on the normal Q-Q plot? Explain why.
b. Derive a 90% CI for the variance of the IQ scores in this population.

c. Derive a 95% CI for the variance of the IQ scores in this population.
d. Derive a 98% CI for the variance of the IQ scores in this population.
e. Give the interpretation of the confidence interval in part (d).
f. Assuming that the true population mean and standard deviation for IQ scores in this population are $\mu = 102.5$ and $\sigma = 9.3$ respectively, and the assumption from part (a) holds:
   i. What is the sampling distribution of our sample mean $\bar{X}_n$?
   ii. Is this sampling distribution exact or approximate? Explain.

6. Suppose that we collect many different samples of $n = 15$ IQ scores from a certain population whose IQ scores are normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 12$, where $\boldsymbol{\sigma}$ **is known** but μ is **unknown**. Based on this, we construct many different 95% confidence intervals for μ.

   In this problem, we will explore what happens as we collect more and more samples and construct more and more confidence intervals for μ.

   a. What is the width of any 95% CI constructed for μ using this method?
   b. If we want to achieve a width of **2.5** for 95% confidence, i.e. we want to be 95% confident that we are within 2.5 points of the true IQ score for this population, what is the sample size n that we need? Discuss your results.
   c. If we want to obtain a CI width of **2.5** but we do **NOT** want to collect more data (i.e. we keep our data as is with $n = 15$), what **level of confidence** do we need to use to construct this CI? Discuss your results.
   d. The following `R` code (top of the next page) simulates 10 random samples of size $n = 15$ from the population above and then computes ten 95% CI's for the mean IQ score μ.

      Each time we construct a CI, this code also records whether that CI contains μ or not ("1" is yes, "0" if not). Then it computes the proportion of the 10 CI's that contain the true μ. This is stored in the variable `cp`.

```
set.seed(12345) # To reproduce same results later
n <- 15
mu <- 100
sigma <- 12
# margin of error
m_e <- sigma/sqrt(n)

# niter is number of random samples of size n to draw
niter <- 10
covers_or_not <- rep(0, niter)

# Construct confidence intervals
for(i in 1:niter){
    # Simulate data on n=15 patients and get sample mean
    patient_dat <- rnorm(n=n, mean=mu, sd=sigma)
    xbar <- mean(patient_dat)

    # Construct 95 percent CI
    CI <- c(xbar-1.96*m_e, xbar+1.96*m_e)
    # Check if the CI contains mu
    if( (CI[1]<mu) & (mu<CI[2]) ) covers_or_not[i] <- 1
}

# Coverage probability
cp <- sum(covers_or_not)/niter
cp
```

Run the above code in `R`. What is the coverage probability (`cp`), or the proportion of the 10 CI's that contain the true μ?

e. In the code in part (d), change `niter <- 10` to `niter <- 40`. In other words, we are changing the number of 95% CI's we construct from 10 to 40. Then run the code again. What proportion of these 40 CI's contain the true μ now?
f. In the code in part (d), change `niter <- 10` to `niter <- 1000`. In other words, we are changing the number of 95% CI's we construct from 10 to 1000. Then run the code again. What proportion of these 1000 CI's contain the true μ now?
g. Discuss your findings in parts (d)-(f). What happens as we increase the number of 95% CI's we construct based on different random samples of the population? Why is this occurring?