Name Ian Jones

Lab Assignment 5

(40 points)

DIRECTIONS

- ***An uploaded pdf copy is due the Thursday after assignment at 1PM.***

A. Read in the first <u>50</u> samples of mosquito-fish lengths from the data from canvas (unknownData.txt – Lab 4 folder) using R commands. Assume this is your random sample from the population. Calculate a confidence interval from a normal distribution for the following exercises. In this case assume the population standard deviation is known (4.82mm) but the population mean is unknown. Do this question in <u>R only</u>. (10 pts total).

1. Calculate the 50% confidence interval. (2pts)

```
calc_ci <- function(conf_level) {  z_score <- qnorm((1 + conf_level) / 2)  margin_error <- z_score * (pop_sd / sqrt(n))
ci_lower <- sample_mean - margin_error  ci_upper <- sample_mean + margin_error  return(c(ci_lower, ci_upper))}
```

ci_50 <- calc_ci(.5) = 21.32023 - 22.23977

2. Calculate the 90% confidence interval. (2pts)

ci_90 <- calc_ci(.9) = 20.65878 - 22.90122

3. Calculate the 95% confidence interval. (2pts)

ci_95 <- calc_ci(.95) = 20.44399 - 23.11601

4. What happens to the width of the confidence interval as we go from 50% to 90% to 95% confidence? (1pt)

The width of the confidence intervel also increases

5. Use R to construct a probability plot of the data subset (2 pts).

qqnorm(mosquito_fish, main="Plot of Mosquito Fish Lengths")

6. Determine if this variable is approximately normally distributed using R (1 pt).

shapiro.test(mosquito_fish)

p-value = .001482

Normally distributed

B. A population of red-bellied snakes is known to have a ratio of grey color morph to red color morph of 45:55 respectively. Use the normal approximation of the binomial distribution to solve the following exercises. Please show all your work by hand and in R for full credit. (9 pts total)

1. What is the probability of selecting from this population a random sample of 25 snakes containing 4 or fewer **grey** morph individuals. (3 pts)

u = 25 * .45 = 11.25 sd = sqrt(25 * .45 * .55) = 2.487

Z = 4.5 - 11.25 / 2.487 = -2.714

P = .0033

prob1(pnorm(4.5 - (25 * .45) / sqrt(25*.45*(1-.45))

2. What is the probability of selecting from this population a random sample of 20 snakes containing 5 or fewer **red** morph individuals. (3 pts)

u = 20 * .55 = 11 sd = sqrt(20 * .55 * .45) = 2.213

Z = 5.5 - 11/ 2.213 = -2.484

P = .0067

prob2(pnorm(5.5 - (20 * .55) / sqrt(20*.55*(1-.55))

3. What is the probability of selecting a random sample of 45 snakes containing between 15 and 30 **grey** morph individuals. (3 pts)

u = 45 * .45 = 20.25 sd = sqrt(45 * .45 * .55) = 3.345

Z1 = 14.5 - 20.25 / 3.345 = -1.717 Z2 = 30.5 - 20.25 / 3.345 = 3.059

P = .9989 - .0424 = .9565

prob2(pnorm(30.5 - (45 * .45) / sqrt(45*.45*(1-.45)) - pnorm(14.5 - (45 * .45) / sqrt(45*.45*(1-.45))

C. We found the probability that the sample mean length of 10 random bluegill sunfish falls between 110mm and 133mm in the lab. In this case assume the population standard deviation is 35.5mm and the population mean is 125mm. Use the Central Limit Theorem to solve the following exercises. Please show all your work by hand and in R in questions for full credit. (5 pts total)

1. Find the probability that the sample mean length of 20 random bluegill sunfish falls between 110 and 133? (2pts)

SE = 35.5/sqrt(20) = 7.9370

Z1 = 110 - 125/7.9370 =-1.89 Z2 = 133 - 125/7.9370 = 1.0079

P = .8433-.0294 = .8139

pnorm(133,125,35.5/sqrt(20) - pnorm(110,125,35.5/sqrt(20)

2. Find the probability that the sample mean length of 30 random bluegill sunfish falls between 110 and 133? (2pts)

SE = 6.4815 Z1 = -2.314 Z2 = 1.234

P = .8910-.0109 = .8801

pnorm(133,125,35.5/sqrt(30) - pnorm(110,125,35.5/sqrt(30)

3. What happens to the standard error as the sample size increases from n=20 to n=30? (1pts)

The standard error dereases as the sample size increases from n=20 to n=30.

D. We sampled of 14 iguanas and their weights (in grams). In this case assume the population standard deviation and the population mean is unknown and the data is normally distributed. Show your work in R only (10 pts total).

1450,1550,2200,1400,1650,2000,2435,1550,1050,2300,2000,2750,1800,2650

1. Compute the 95% confidence interval (2 pt).

```
weights<-c(1450,1550,2200,1400,1650,2000,2435,1550,1050,2300,2000,2750,1800,2650)
n <- length(weights)          mean_weight<-mean(weights)
SE <- sd(weights) / sqrt(n)          tc95 <- qt(1-.05/2, df = n-1)
ci_95 <- mean_weight + (-1,1) * tc95 * SE          1620.94 - 2205.49
```

2. Compute the 80% confidence interval (2 pt).

```
tc80 <- qt(1-.2/2, df = n-1)
ci_80 <- mean_weight + (-1,1) * tc80 * SE
1730.55 - 2095.88
```

3. In R, use a 'for' loop to generate a sampling distribution of size 2 from this data set (2 pt). Plot this in a histogram (2 pt).

```
num_samples <- 14   sample_means <- numeric(num_samples)
for (i in 1:num_samples) {  sample_means[i] <- mean(sample(weights, size = 2, replace = TRUE))}

hist(sample_means, main = "Weights of Iguanas",  xlab = "Sample Mean")
```

4. In R, plot a t distribution with 13 degrees of freedom. Color the 95% confidence interval. (2 pts)

```
plot(x, dt(-4:4. length.out=14,13, type = "l", main = "t-distribution (df = 13)",      xlab = "t", ylab = "Density")
pt(.95, 3, lower.tail = TRUE)
```

E. <u>In your words</u> define degrees of freedom (3 pts total).

The number of values within a calculation that are able to change. They can be thought of as the number of independent variables.

F. <u>In your words</u> define p-value (3 pts total).

It's a measure of strength or relevance against a null hypothesis, with the smaller the p-value, the more unlikely the results are to happen and vice versa.