Homework #4

Glenn Clapp

6.2.2 (Using population standard deviation)(3 pts)

We wish to estimate the mean serum indirect bilirubin level of 4-day-old infants. The mean for a sample of 16 infants was found to be 5.98 mg/100cc. Assume that bilirubin levels in 4-day-old infants are approximately normally distributed with a standard deviation of 3.5 mg/100cc

n\_infants = 16  
xbar\_infants = 5.98  
sd\_infants = 3.5 # population standard deviation.  
  
sd\_xbar = (sd\_infants^2/n\_infants)^0.5 # standard deviation of sample means  
  
u\_infants = xbar\_infants + c(-1,1)\*sd\_xbar  
print(u\_infants) # Lower and upper limits of approximate 95% confidence interval

## [1] 5.105 6.855

6.3.2 (7 pts)

In a study of the effects of early Alzheimer’s disease on nondeclarative memory, Reber et al. used the Category Fluency Test to establish baseline persistnece and semantic memory and language abilities. The eight subjects in the sample had Category Fluency Test scores of 11, 10, 6, 3, 11, 10, 9, 11. Assume that the eight subjects constitute a simple random sample from a normally distributed population of similar subjects with early Alzheimer’s disease.

CFT = data.frame(  
 scores = c(11,10,6,3,11,10,9,11)  
)  
n = 8  
# normally distributed population

1. What is the point estimate of the population mean?

u\_estimate = mean(CFT$scores)  
print(u\_estimate)

## [1] 8.875

1. What is the standard deviation of the sample?

s = sd(CFT$scores)  
print(s)

## [1] 2.900123

1. What is the estimated standard error of the sample mean

standard\_error = s/n^0.5  
print(standard\_error)

## [1] 1.025348

1. Construct a 95 percent confidence interval for the population mean category fluency test score.

ts <- -qt((1-.95)/2,df=n-1,lower.tail=TRUE)  
lower.ci=u\_estimate-ts\*standard\_error  
upper.ci=u\_estimate+ts\*standard\_error  
  
print(lower.ci)

## [1] 6.450436

print(upper.ci)

## [1] 11.29956

1. What is the precision of the estimate?

precision = upper.ci - lower.ci  
print(precision)

## [1] 4.849127

1. State the probabilistic interpretation of the confidence interval you constructed.

We can be 95% certain that the true population mean is somewhere between 6.450 and 11.300

1. State the practical interpretation of the confidence interval you constructed.

In repeated sampling, 95% of of the intervals constructed in this way will straddle the mean.

6.3.4 (4pts)

The concern of a study by Beynnon et al. were nine subjects with chronic anterior cruciate ligament (ACL) tears. One of the variables of interest waws the laxity of the anteroposterior, where higher values indicate more knee instability. The researchers found that among subjects with ACL-deficient knees, the mean laxity value was 17.4mm with a standard deviation of 4.3mm.

1. What is the estimated standard error of the mean?

x\_bar\_ACL = 17.4  
s\_ACL = 4.3  
n\_ACL = 9  
  
standard\_error = s\_ACL/n^0.5  
print(standard\_error)

## [1] 1.52028

1. Construct the 99 percent confidence intreval for the mean of the population from which the nine subjects may be presumed to be a random sample.

ts <- -qt((1-.99)/2,df=n-1,lower.tail=TRUE)  
lower.ci=u\_estimate-ts\*standard\_error  
upper.ci=u\_estimate+ts\*standard\_error  
  
print(lower.ci)

## [1] 3.554807

print(upper.ci)

## [1] 14.19519

1. What is the precision of the estimate

precision = upper.ci - lower.ci  
print(precision)

## [1] 10.64039

1. What assumptions are necessary for the validity of the confidence interval you constructed?

The samples must be drawn from a random, approximately normal population distribution.

6.3.6 (3pts)

The subjects of a study by Dugoff et al. were 10 obstetrics and gynecology interns at the University of Colorado Health Sciences Center. The researchers wanted to assess competence in performing clinical breast examinations. One of the baseline measurements was the number of such examinations performed. The following data give the number of breast examinations performed for this sample of 10 interns.

Construct a 95 percent confidence interval for the mean of the population from which the study subjects may be presumed to have been drawn.

breast\_exams = data.frame(  
 Intern\_number = c(1,2,3,4,5,6,7,8,9,10),  
 No\_exams = c(30,40,8,20,26,35,35,20,25,20)  
)  
n\_breast = 10  
ci\_breast = 0.95  
  
mean\_breast = mean(breast\_exams$No\_exams)  
sd\_breast = sd(breast\_exams$No\_exams)  
  
stderr\_breast = sd\_breast/n\_breast^0.5  
  
ts <- -qt((1-ci\_breast)/2,df=n-1,lower.tail=T)  
lower\_breast = mean\_breast-ts\*stderr\_breast  
upper\_breast = mean\_breast+ts\*stderr\_breast  
print(lower\_breast)

## [1] 18.81971

print(upper\_breast)

## [1] 32.98029

6.4.2 (3pts)

Chan et al. developed a questionnaire to assess knowledge of prostate cancer. There was a total of 36 questions to which respondents could answer “agree,” “disagree,” or “don’t know.” Scores could range from 0 to 36. The mean scores for Caucasian study participants was 20.6 with a standard deviation of 5.8, while the mean scores for African-American men was 17.4 with a standard deviation of 5.8. The number of Caucasian study participants was 185, and the number of African-Americans was 86.

Construcct the 90, 95, and 99 percent confidence intervals for the difference between population means. Where appropriate, state the assumptions that make your method valid. State the practical and probabilisitc interpretations of each interval that you construct. Consider the variables under consideration in each exercise, and state what use you think reseaerchers migh make of your results.

uCaucasian = 20.6  
sdCaucasian = 5.8  
nCaucasian = 185  
  
uAfrican = 17.4  
sdAfrican = 5.8  
nAfrican = 86  
  
s\_pooled = ((nCaucasian-1)\*sdCaucasian^2+(nAfrican-1)\*sdAfrican^2)/(nCaucasian+nAfrican-2)  
# This calculates a pooled variance estimate. We can do this because the variances of the two samples are equal.  
  
stderr\_prostate = (s\_pooled/nCaucasian + s\_pooled/nAfrican)^0.5  
  
diff\_estimate = uCaucasian-uAfrican  
ts90 <- -qt((1-0.90)/2,df=nAfrican-1,lower.tail=T)  
CI90\_prostate = diff\_estimate + c(-1,1)\*ts90\*stderr\_prostate  
  
ts95 <- -qt((1-0.95)/2,df=nAfrican-1,lower.tail=T)  
CI95\_prostate = diff\_estimate + c(-1,1)\*ts95\*stderr\_prostate  
  
ts99 <- -qt((1-0.99)/2,df=nAfrican-1,lower.tail=T)  
CI99\_prostate = diff\_estimate + c(-1,1)\*ts99\*stderr\_prostate  
  
print(CI90\_prostate)

## [1] 1.941178 4.458822

print(CI95\_prostate)

## [1] 1.694945 4.705055

print(CI99\_prostate)

## [1] 1.205454 5.194546

Probabilistic: We can be 90% certain that the true population mean is somewhere between 1.94 and 4.46 We can be 95% certain that the true population mean is somewhere between 1.69 and 4.71 We can be 99% certain that the true population mean is somewhere between 1.21 and 5.19

Practical In repeated sampling, 90% of of the intervals constructed in this way will straddle the mean. In repeated sampling, 95% of of the intervals constructed in this way will straddle the mean. In repeated sampling, 99% of of the intervals constructed in this way will straddle the mean.

Assumptions:

The scores for each of the populations are randomly selected from a population that is approximately normally distributed.

Results:

In this case, because even the 99% confidence interval did not straddle zero, the researchers can be 99% certain that there was a mean difference between the caucasian and African populations and that the average Caucasian scored between 1.21 and 5.19 points higher on the prostate test.

# Resume work here.

6.4.4 (3pts)

The purpose of a study by Nozawa et al. was to determine the effectiveness of segmental wire fixation in athletes with spondylolysis. Between 1993 and 2000, 20 athletes (6 women and 14 men) with lumbar spondylolysis were treated surgically with the technique. The followign table gives the Japanese Orthopaedic Association (JOA) evaluation score for lower back pain syndrome from men and women prior to the surgery. The lower score indicates less pain.

Construct the 90, 95, and 99 percent confidence intervals for the difference between population means. Where appropriate, state the assumptions that make your method valid. State the practical and probabilisitc interpretations of each interval that you construct. Consider the variables under consideration in each exercise, and state what use you think reseaerchers migh make of your results.

#JOA\_scores = data.frame(  
# female = c(14,13,24,21,20,21,'','','','','','','',''),  
# male = c(21,26,24,24,22,23,18,24,13,22,25,23,21,25)  
#)  
  
female\_scores = c(14,13,24,21,20,21)  
male\_scores = c(21,26,24,24,22,23,18,24,13,22,25,23,21,25)  
  
mean\_female = mean(female\_scores)  
sd\_female = sd(male\_scores)  
n\_female = 6  
  
mean\_male = mean(male\_scores)  
sd\_male = sd(male\_scores)  
n\_male = 14  
  
s\_pooled = ((n\_male-1)\*sd\_male^2+(n\_female-1)\*sd\_female^2)/(n\_male+n\_female-2)  
# This calculates a pooled variance estimate. We can do this because the variances of the two samples are equal.  
  
stderr\_spondy = (s\_pooled/n\_male + s\_pooled/n\_female)^0.5  
  
diff\_estimate = mean\_male-mean\_female  
ts90 <- -qt((1-0.90)/2,df=n\_female-1,lower.tail=T)  
CI90\_spondy = diff\_estimate + c(-1,1)\*ts90\*stderr\_spondy  
  
ts95 <- -qt((1-0.95)/2,df=n\_female-1,lower.tail=T)  
CI95\_spondy = diff\_estimate + c(-1,1)\*ts95\*stderr\_spondy  
  
ts99 <- -qt((1-0.99)/2,df=n\_female-1,lower.tail=T)  
CI99\_spondy = diff\_estimate + c(-1,1)\*ts99\*stderr\_spondy  
  
print(CI90\_spondy)

## [1] 0.08184755 6.68005722

print(CI95\_spondy)

## [1] -0.8276905 7.5895952

print(CI99\_spondy)

## [1] -3.220607 9.982512

Probabilistic: We can be 90% certain that the true population mean is somewhere between 0.082 and 6.680 We can be 95% certain that the true population mean is somewhere between -0.828 and 7.590 We can be 99% certain that the true population mean is somewhere between -3.221 and 9.983

Practical In repeated sampling, 90% of of the intervals constructed in this way will straddle the mean. In repeated sampling, 95% of of the intervals constructed in this way will straddle the mean. In repeated sampling, 99% of of the intervals constructed in this way will straddle the mean.

Assumptions:

The scores for each of the populations are randomly selected from a population that is approximately normally distributed.

Results:

In this case, the 90% confidence interval excluded a mean difference of 0. In other words, the stronger confidence intrevals would lead to the conclusion that no difference between the male and female populations was detected, and that it is possible that either population was slightly higher or slightly lower than the other. This would be an inconclusive result.

6.4.6 (3pts)

Transverse diameter measurements on the hearts of adult males and females gave the following results:

Construct the 90, 95, and 99 percent confidence intervals for the difference between population means. Where appropriate, state the assumptions that make your method valid. State the practical and probabilisitc interpretations of each interval that you construct. Consider the variables under consideration in each exercise, and state what use you think reseaerchers migh make of your results.

n\_male = 12  
mean\_male = 13.21 #cm  
sd\_males = 1.05 #cm  
  
n\_female = 9  
mean\_female = 11.00 #cm  
sd\_females = 1.01 #cm  
# Assume normally distributed populations with equal variances.  
  
s\_pooled = ((n\_male-1)\*sd\_male^2+(n\_female-1)\*sd\_female^2)/(n\_male+n\_female-2)  
# This calculates a pooled variance estimate. We can do this because the variances of the two samples are equal.  
  
stderr\_heart = (s\_pooled/n\_male + s\_pooled/n\_female)^0.5  
  
diff\_estimate = mean\_male-mean\_female  
ts90 <- -qt((1-0.90)/2,df=n\_female-1,lower.tail=T)  
CI90\_heart = diff\_estimate + c(-1,1)\*ts90\*stderr\_heart  
  
ts95 <- -qt((1-0.95)/2,df=n\_female-1,lower.tail=T)  
CI95\_heart = diff\_estimate + c(-1,1)\*ts95\*stderr\_heart  
  
ts99 <- -qt((1-0.99)/2,df=n\_female-1,lower.tail=T)  
CI99\_heart = diff\_estimate + c(-1,1)\*ts99\*stderr\_heart  
  
print(CI90\_heart)

## [1] -0.5413159 4.9613159

print(CI95\_heart)

## [1] -1.201875 5.621875

print(CI99\_heart)

## [1] -2.754502 7.174502

Probabilistic: We can be 90% certain that the true population mean is somewhere between -0.541 and 4.961 We can be 95% certain that the true population mean is somewhere between -1.202 and 5.622 We can be 99% certain that the true population mean is somewhere between -2.755 and 7.175

Practical In repeated sampling, 90% of of the intervals constructed in this way will straddle the mean. In repeated sampling, 95% of of the intervals constructed in this way will straddle the mean. In repeated sampling, 99% of of the intervals constructed in this way will straddle the mean.

Assumptions:

The scores for each of the populations are randomly selected from a population that is approximately normally distributed.

Results:

In this case, none of the confidence intervals excluded 0. In other words, we would conclude that there was no difference between the male and female populations was detected, and that it is possible that either population was slightly higher or slightly lower than the other. This would be an inconclusive result.

6.7.4 (3pts)

For multiple sclerosis patients we wish to estimate the mean age at which the disease was first diagnosed. We want a 95 percent confidence interval that is 10 years wide. If the population variance is 90, how large should our sample be.

zs = qnorm(0.975)  
SD\_age = 90^0.5  
ci\_interval = 10  
ci\_half = ci\_interval/2  
n = (zs\*SD\_age/ci\_half)^2  
print(ceiling(n)) # round up to the nearest integer

## [1] 14