2028: Basic Statistical Methods
Solutions - Homework 4

1. **Mean Confidence Interval Estimation**

(a) a

$$z_{\alpha/2} = z_{0.005} = 2.57, \bar{x} = 13.77, \sigma = 0.5, n = 11$$

$$\bar{x} - z_{0.005}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{0.005}\frac{\sigma}{\sqrt{n}}$$

$$13.77 - 2.57\frac{0.5}{\sqrt{11}} \leq \mu \leq 13.77 + 2.57\frac{0.5}{\sqrt{11}}$$

$$13.383 \leq \mu \leq 14.157$$

b

$$\mu \geq \bar{x} - z_{0.05}\frac{\sigma}{\sqrt{n}} = 13.77 - 1.65\frac{0.5}{\sqrt{11}} = 13.521$$

c

$$n = (\frac{z_{0.025}\sigma}{E})^2 = (\frac{1.96 \times 0.5}{2})^2 = 0.2401 \Rightarrow n = 1$$

d

$$2 \times z_{0.025}\frac{\sigma}{\sqrt{n}} = 1.5$$

$$2 \times 1.96 \times \frac{0.5}{\sqrt{n}} = 1.5$$

$$n = 1.701 \Rightarrow n = 2$$

(b) The lower bound of the confidence interval is:

$$6.668 = \bar{x} - \frac{t_{\alpha,n-1}s}{\sqrt{n}} = 6.861 - \frac{t_{\alpha,15}(0.440)}{\sqrt{16}}$$

so from

$$6.668 - 6.861 = -\frac{t_{\alpha,15}(0.440)}{\sqrt{16}}$$

we obtain that

$$0.193 = \frac{t_{\alpha,15}(0.440)}{\sqrt{16}} \implies t_{\alpha,15} = (0.193)(\sqrt{16})/(0.440) = 1.753.$$

Having the quantile of the t-distribution with 15 degrees of freedom we can find the probability value

$$1 - \alpha = P(X \leq 1.753) = 95\%$$

(c) $(-\infty, c)$ is a one-sided 95% t-interval for the average population mean μ. Since the one-sided 95% t-interval for $n = 30$, $\bar{x} = 14.62$, and $s = 2.98$ is

$$\left(-\infty, \bar{x} + \frac{t_{\alpha,n-1}s}{\sqrt{n}}\right)$$

and $t_{0.05,29} = 1.699$, then

$$c = \bar{x} + \frac{t_{\alpha,n-1}s}{\sqrt{n}} = 14.62 + 1.699\frac{2.98}{\sqrt{30}} = 15.54.$$

Since the upper limit 15.54 is less than 16, then it is plausible that $\mu \leq 16$.

2. **Proportion Parameter Estimation**

(a) i. Here $n = 1010$, and $\widehat{p}$, the sample proportion is given in the poll to be 44%.
Using the 95% C.I. for p that we developed in class:

$$\widehat{p} \pm z_{0.025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

we get that a 95% C.I. for p, the proportion of all US adults who oppose the legalization of marijuana is

$$0.44 \pm 2\sqrt{\frac{.44 \times .56}{1010}} = (0.41, 0.47)$$

Therefore, we are 95% confident that p is between 0.41 and 0.47.

ii. The margin error is a measure of accuracy of our sample result ($\widehat{p} = .44$) as an estimate for p. We are fairly certain (95% confident) that our estimate $\widehat{p}$ is "off" by no more than .031. In other words, we are 95% certain that p, the proportion of <u>all</u> US adults who oppose legalizing marijuana is within .031 of our estimate
$\Rightarrow$ between .44-.031 and .44+.031.

(b) i. A point estimate of the proportion parameter is:

$$\widehat{p} = .88$$

and a 95% CI is given by:

$$\left(\widehat{p} \pm z_{.025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right)$$

and for our observations, the CI is:

$$\left(.88 \pm (1.96)\sqrt{\frac{.88(1-.88)}{1000}}\right)$$

or

$$(.8599, .9001)$$

so .9 falls in the interval.

ii. We want to find n such that:

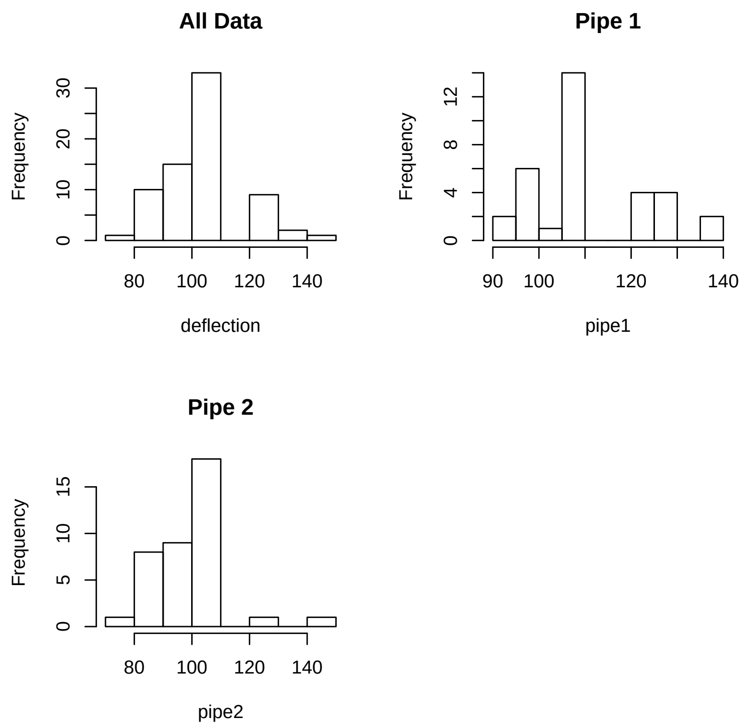
$$2z_{.025}\sqrt{\frac{0.5(1-0.5)}{n}} \leq .02$$
$$\Rightarrow \; n \geq 9604$$

iii. When we know that the true proportion exceeds $p^* = 0.7$, we can approximately obtain a sample size as:

$$n \geq \frac{4z_{\alpha/2}^2 p^*(1-p^*)}{w^2} = \frac{4(1.96)^2 0.7(1-0.7)}{(.02)^2} = 8067.36 \approx 8068.$$

3 **Computer Problem**

(a) Histograms for the full data set and each pipe individually:



The full pipe data has an overall shape that is not normally distributed because of the gap in the data from 110 to 120. The overall shape (without the gap) does appear to be roughly normal though. The histogram for pipe 1 has the same gap, and the tails are not as balanced as in the full data histogram. Pipe 2 has the most uneven distribution, which is the farthest from normal. All three histograms are unimodular, but the histogram for pipe 2 is not symmetric.

(b) $\widehat{\mu}_1 = 111.85$
$\widehat{\mu}_2 = 103.05$
$\widehat{\sigma}_1 = 2.28$
$\widehat{\sigma}_2 = 2.08$

(c) 95% Confidence Interval for mean deflection temperature in type 1 pipes : (107.21,116.49)
95% Confidence Interval for mean deflection temperature in type 2 pipes : (98.84,107.26)
The interval for type 1 pipes is slightly longer than the interval for type 2 pipes (9.28 vs 8.42) but the bigger difference is between the two intervals is how they are centered. The entire interval for type 2 pipes is below the lower bound of the confidence interval for type 1 pipes, indicating that the mean deflection for type 1 pipes is higher than that of type 2 pipes. Using the confidence intervals gives more evidence to back this conclusion than just looking at the sample means.

R code:

```
#read in the data
data = read.table("Deflection.txt",header=FALSE)
index = data[,1]
deflection = data[,2]
index.1 = which(index==1)
index.2=which(index==2)
pipe1 = deflection[index.1]
pipe2 = deflection[index.2]

#Question 1
#generate the histograms of the data
par(mfrow=c(2,2))
hist(deflection, main = "All Data")
hist(pipe1, main = "Pipe 1")
hist(pipe2, main = "Pipe 2")

#Question 2
#find the estimates for the mean for each pipe and the standard
#deviation of the mean estimates
u1 = mean(pipe1)
u2 = mean(pipe2)
n1 = length(pipe1)
n2 = length(pipe2)
#we divide by n1 and n2 because we want the standard deviation of the
#estimates for the sample means - not the standard deviation of the
#data sets alone
sd_u1 = sqrt(var(pipe1))/sqrt(n1)
sd_u2 = sqrt(var(pipe2))/sqrt(n2)

#Question 3
#find 95\% confidence intervals for both pipe types
t_1 = t.test(pipe1, conf.level = .95)
t_2 = t.test(pipe2, conf.level = .95)
```