ST 411/511 Homework 2

Due on January 23

Vijay Vardhan Tadimeti

Winter 2019

Instructions

This assignment is due by 11:59 PM, January 23, 2019 on Canvas. **You should submit your assignment as either a PDF or Word document, which you can compile (should you choose – recommended) from the provided .Rmd (R Markdown) template.** Please include your code.

Problems (25 points total)

Question 1

A random sample of $n = 500$ books is selected from a library and the number of words in the title of each book is recorded. The sample mean number of words in the title is 6.2 words. The population variance is 40 words^2.

(a) (2 points) Compute the z-statistic for testing the null hypothesis $H_0 : \mu = 7$.

```
Z <- (6.2 - 7) / sqrt(40/500)
Z
```

```
## [1] -2.828427
```

(b) (3 points) Perform a level $\alpha = 0.1$ test of $H_0 : \mu = 7$ vs. the one-sided lesser alternative $H_A : \mu < 7$.

```
pnorm(Z)
```

```
## [1] 0.002338867
```

Since the value is less than alpha, we reject the null hypothesis

(c) (2 points) What is the one-sided lesser p-value for the statistic you computed in part (a)?

```
pnorm(Z)
```

```
## [1] 0.002338867
```

(d) (2 points) What is the two-sided p-value for the statistic you computed in part (a)?

```
2*(1 -pnorm(abs(Z)))
```

```
## [1] 0.004677735
```

(e) (2 points) Construct a 95% confidence interval for the population mean number of words per title. Hint: recall that a 95% confidence interval is formed by the sample mean $\pm 1.96\times$ standard deviation of the sampling distribution.

```
alpha <-qnorm(0.05/2)
alpha
```

```
## [1] -1.959964
```

```
confidence_interva_1 <- 6.2 - (1.96*sqrt(40/500))
confidence_interva_1
```

```
## [1] 5.645628
```

```
confidence_interval_2 <- 6.2 + (1.96*sqrt(40/500))
confidence_interval_2
```

```
## [1] 6.754372
```

(f) (1 point) Based on your confidence interval from part (e), would a level $\alpha = 0.05$ two-sided hypothesis test reject or fail to reject the null hypothesis that the population mean is 6.5 words per title?

Here, we cannot rule out the possibility that the population average mean for words per title is 6.5, and since population mean is in the confidence interval, we fail to reject the null hypothesis.

Question 2

Consider the `rivers` data set in `R`, which is a vector of the lengths (in miles) of 144 "major" rivers in North America, as compiled by the US Geological Survey.

```
data(rivers)
View(rivers)
```

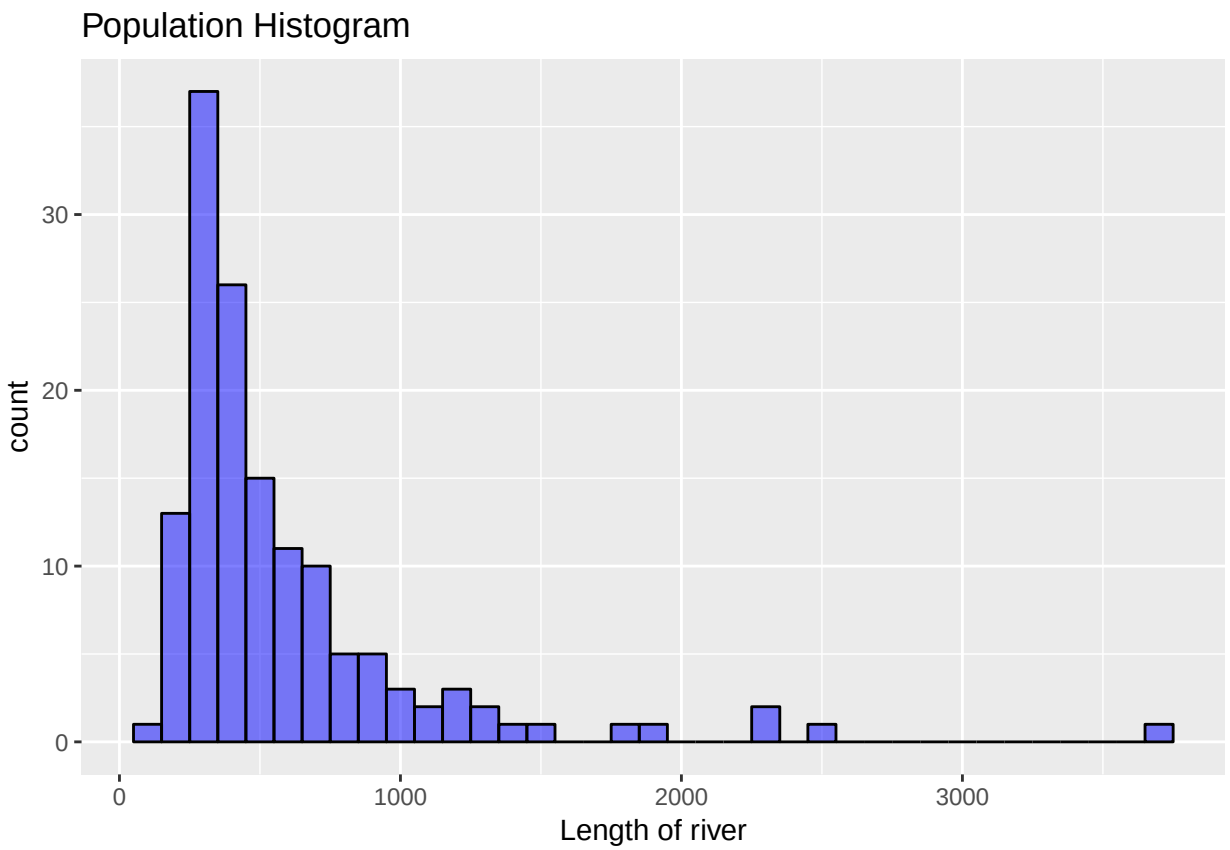
(a) (1 point) What is the length of the longest "major" river in North America? Hint: you can find the maximum of a vector using the `max` function.

```
max(rivers)
```

```
## [1] 3710
```

(b) (2 points) Create a population histogram of the lengths of the rivers. Describe the shape of the distribution. Note: to use ggplot, the data have to be formatted as a data frame. I have given you the line of code that does this.

```
riversdf <- data.frame(rivers)
ggplot(riversdf, aes(x=rivers)) + geom_histogram(binwidth=100, color="black", alpha=0.5, fill="blue") +
```



In the histogram we notice that there are some outliers present. It also looks skewed to the right. The outliers are also bigger/larger than some rivers.

(c) (1 point) Select a random sample of $n = 30$ rivers, using `set.seed(411511)` to make sure you draw the same random sample each time. What is the sample mean?

```
set.seed(411511)
samp <- sample(1:nrow(riversdf), size=30)
Xbar <- mean(riversdf$rivers[samp])
Xbar
```

```
## [1] 525.8667
```

(d) (2 points) Find the test statistic for a z-test of $H_0 : \mu = 600$ versus $H_A : \mu \neq 600$.

```
#this is our sample mean minus $\mu_0$ (the value of $\mu$ assuming the null hypothesis is true), divid
variance <-var(rivers)
variance
```

```
## [1] 243908.4
```

```
Z <- (Xbar - 600)/sqrt(variance/30)
Z
```

```
## [1] -0.8221684
```

(e) (2 points) Find the p-value corresponding to your test statistic from part (d). Recall that you are using a two-sided alternative hypothesis.

```
2*(1 -pnorm(abs(Z)))
```

```
## [1] 0.4109811
```

(f) (2 points) What do you conclude from this hypothesis test? State your conclusion in words.

We assume $\alpha = 0.05$ We fail to reject the null hypothesis,$H_0 : \mu = 600$, as the possibility that the population average mean for lengths of river is 600 since p value is greater than alpha, cannot be ruled out.

Question 3

Researchers are curious about how soil type affects plant growth. To study this, they obtain 100 seeds of a particular plant species from a local seed collector. They randomly choose 50 seeds and plant each in a separate pot filled with soil type A. The remaining 50 seeds are each planted in a separate plot filled with soil type B. The plants receive the same care, and at the end of 3 months the height of each plant is measured.

(a) (1 point) Is this an example of a randomized experiment or an observational study?

Randomized

(b) (2 points) To what group can inference be made?

The inference can be made on the group of seeds of a particular plant species from the local seed collector, which is used for the experiment.