

Final Exam

Student Name : Sachid Deshmukh

Date : 12/17/2018

- GitHub Location for rmd file
 - GitHub Location for pdf file
 - RPub's location of published file
-

3. The time taken to complete a statistics final by all students is normally distributed with a mean of 120 minutes and a standard deviation of 10 minutes.

- Find the probability that a randomly selected student will take more than 150 minutes to complete the test.

```
avg = 120
sd = 10
x = 150
1 - pnorm(x, avg, sd)
```

```
## [1] 0.001349898
```

Answer : The probability that a randomly selected student will take more than 150 minutes to complete the test is 0.0013

- Find the probability that the mean time taken to complete the test by a random sample of 16 students would be between 122 and 126 minutes.

```
x1 = 122
x2 = 126
pnorm(x2, avg, sd) - pnorm(x1, avg, sd)
```

```
## [1] 0.1464872
```

Answer : Probability that the mean time taken to complete the test between 122 and 126 mins is 0.146

4. Rh-negative blood appears in 15% of the United States population.

- Find the probability that out of 7 randomly selected U.S. residents at least 3 of them have Rh-negative blood.

```
n = 7
x = 3
p = 0.15
dbinom(x, n, p)
```

```
## [1] 0.06166199
```

Answer : Probability is 0.061

- Use the normal approximation to find the probability that in a group 100 randomly selected people fewer than 17.5% will have a Rh-negative blood.

```
n = 100
p = 0.15
avg = n*p
sd = sqrt(n*p*(1-p))
x = 17.5
prob = pnorm(x, avg, sd)
prob
```

```
## [1] 0.7580801
```

Answer : Probability is 0.758

5. The U.S. Travel Industry estimated that Americans planned to spend an average of 4.8 nights away on vacations in 1995 (U.S. News & World Report, June 12, 1995). Suppose that this mean was based on a random sample of 100 Americans and the population standard deviation was 1.5 nights.

- Construct a 90% confidence interval for the mean length of vacations Americans planned in 1995.

```
sd = 1.5
avg = 4.8
n = 100
std.err = sd/sqrt(n)
crit.value = qnorm(0.95)

confinterval = c((avg-(std.err*crit.value)), (avg+(std.err*crit.value)))
confinterval
```

```
## [1] 4.553272 5.046728
```

Answer: 90% confidence interval is 4.553 to 5.046 days

- A poll of 1226 adults revealed that 49% believe that the devil may sometimes possess earthlings. Find a 95% confidence interval for the population proportion of the adults who hold this opinion. (Source:“Demons Begone,” Asheville Citizen-Times, April 5, 1991).

```
p = 0.49
n = 1226
x = 1226*0.49
prop.test(x,n, conf.level = 0.95)
```

```
##
## 1-sample proportions test with continuity correction
##
## data: x out of n, null probability 0.5
## X-squared = 0.45122, df = 1, p-value = 0.5018
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
## 0.4616864 0.5183770
## sample estimates:
## p
## 0.49
```

Answer: 95% confidence interval for the population proportion who holds above opinion is 0.461% to 0.518%

6. Grocery stores, drugstores, and large supermarkets all use scanners to calculate a customer's bill.

Scanners should be as accurate as possible. A state agency regularly monitors stores by randomly selecting items and comparing with the shelf price with the checkout scanner price. During one check by the agency, 16 items were found to be incorrectly scanned. The amounts of overcharge(in cents) were 200, -99, 100, -50, 40, -60, 20, 30, 50, 300, -120, 100, 50, 30, -70, 40 A negative sign indicates an undercharge-the scanner price was below the shelf price.

- Make a stemplot of the data interpret.

```
data = c(-120, -99, -70, -60, -50, 20, 30, 30, 40, 40, 50, 50, 100, 100, 200, 300)
stem(data)
```

```
##
## The decimal point is 2 digit(s) to the right of the |
##
## -1 | 20
## -0 | 765
## 0 | 2334455
## 1 | 00
## 2 | 0
## 3 | 0
```

Answer: From the stem plot we can see that the range of incorrectly scanned price is from -120 cents to 300 cents. most of the time, scanner is overcharging from 20 cents to 50 cents.

- Compute the mean and the range.

```
avg = mean(data)
range = range(data)
```

Answer: Mean of the data is 35.06. Range of the data is -120 to 300

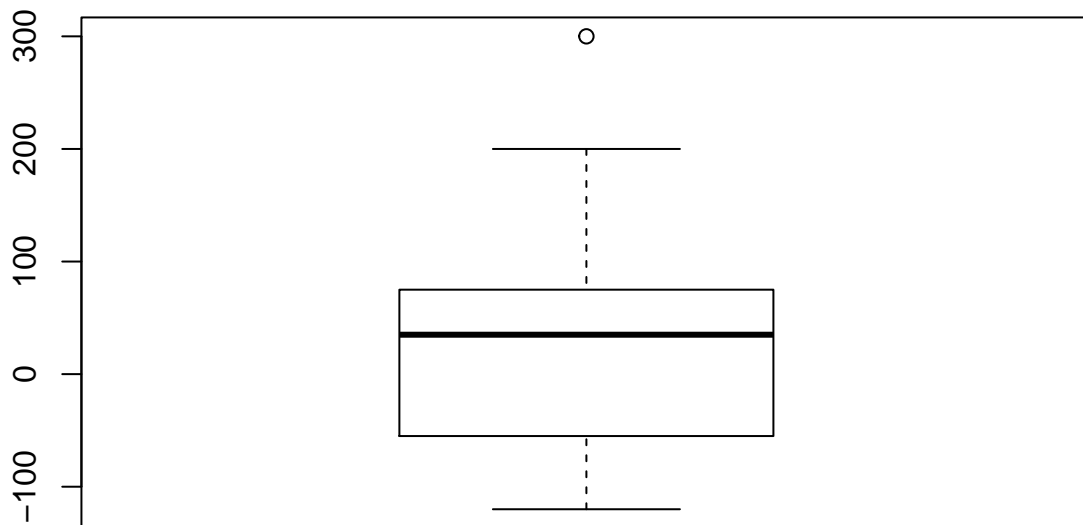
- Give the five-number summary of the data.

```
fivenum(data)
```

```
## [1] -120 -55 35 75 300
```

- Construct a boxplot and interpret.

```
boxplot(data)
```



Answer: From the boxplot, it can be seen that data is right skewed. This indicates scanner have tendency to overcharge the price. The median lies at value 35. The second quartile is tall which indicates there is wide variety in the errors when it comes to under charging. The third quartile is small which indicates that scanner is consistent while making over charging errors. The First and Third quartile numbers stands at -52.5 and 62.5 which indicates that most of the scanner errors are in this range. There is one outlier identified in the box plot with value stamped at 300

- Use the 1.5xIQR criterion to spot suspected outliers.

```
iqr.range = IQR(data)
outlier = data[data > (1.5 * iqr.range)]
outlier
```

```
## [1] 200 300
```

Answer: There are two suspected outliers with value as 200 and 300 respectively

- For this data sample standard deviation is 1.083. Test the hypothesis that the mean overcharge is more than 0 at 0.05 significance level.

```
n = length(data)
sd = 1.083
avg = mean(data)
std.err = sd/sqrt(n)
crit.value = qnorm(0.95)
zvalue = avg/std.err
```

Z value is very high than expected critical value of 1.64. We need to reject null hypothesis

that mean overcharge is less than or equal to 0 at 0.05 significance level. We accept alternate hypothesis that mean overcharge is more than 0 at 0.05 significance level

7. Do cars traveling in the right lane of I-94 travel slower than those in the left lane? The following sample information was obtained. Use the 0.01 significance level to provide an answer to this question.

```
left.size = 6
left.mean = 69
left.sd = 3.22

right.size = 5
right.mean = 65
right.sd = 4.12

left.stderr = left.sd/sqrt(left.size)
right.stderr = right.sd/sqrt(right.size)
tot.stderr = left.stderr + right.stderr

zvalue = (right.mean - left.mean)/tot.stderr

crit.value = qnorm(0.01)
```

Answer : Z Value (-1.266) is lesser than critical value of -2.32 at 0.01 significance level. We can't reject null hypothesis. We need to accept null hypothesis that cars traveling in right lane of I-94 are faster than or equal to left lane. We can conclude that cars travelling in right lane are faster than or equal to cars travelling in left lane

8. A noted medical researcher has suggested that a heart attack is less likely to occur among adults who actively participate in athletics. A random sample of 300 adults is obtained. Of that total, 100 are found to be athletically active. Within this group, 10 suffered heart attacks; among the 200 athletically in active adults, 25 had suffered heart attacks.
- Test the hypothesis that the proportion of adults who are active and suffered heart attacks is different than the proportion of adults who are not active and suffered heart attacks. Use the 0.05 significance level.

```
prop.test(c(10,25), c(100,200))

##
## 2-sample test for equality of proportions with continuity
## correction
##
## data:  c(10, 25) out of c(100, 200)
## X-squared = 0.19811, df = 1, p-value = 0.6562
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.10705274  0.05705274
## sample estimates:
## prop 1 prop 2
##  0.100  0.125
```

Answer: Based on the above proportion test we get p value of 0.65. This is greater than level of significance which is 0.05. We fail to reject null hypothesis. We can conclude that there is no difference in the proportions of adults who are active and suffered heart attacks and adults who are not active and suffered heart attacks

- Construct a 99% confidence interval for the difference between the proportions of all active and inactive adults who suffered heart attacks.

```
prop.test(c(10,25), c(100,200), conf.level = 0.99)
```

```
##
## 2-sample test for equality of proportions with continuity
## correction
##
## data: c(10, 25) out of c(100, 200)
## X-squared = 0.19811, df = 1, p-value = 0.6562
## alternative hypothesis: two.sided
## 99 percent confidence interval:
## -0.13047891 0.08047891
## sample estimates:
## prop 1 prop 2
## 0.100 0.125
```

Answer: 99% confidence interval of difference in the proportion is -13% to 8%. Since 0 falls within this range, we can say that there is no difference in the proportions observed between adults who are athletically active and got heart attack and those who are athletically not active and got heart attack at 99% confidence interval

9. Based on interviews of couples seeking divorces, a social worker compiles the following data related to the period of acquaintanceship before marriage and the duration of marriage.

- Perform a test to determine whether the data substantiate an association between the stability of a marriage and the period of acquaintanceship prior to marriage. Use $\alpha=0.05$.

```
data <- as.table(rbind(c(11, 8), c(28, 24), c(21, 19)))
dimnames(data) <- list(acquaintanceship = c("Under_0.5", "0.5-1.5", "Over_1.5"),
                      Duration = c("Atleast_4", "Morethan_4"))

result = chisq.test(data)
result
```

```
##
## Pearson's Chi-squared test
##
## data: data
## X-squared = 0.15265, df = 2, p-value = 0.9265
```

Answer: From the above chi square test of independence, we get the p value of 0.92. It is greater than significance level of 0.05. We fail to reject null hypothesis. We can conclude that there is no relationship between the stability of a marriage and period of acquaintanceship prior to marriage at 95% confidence interval