Day 51 Review for Hypothesis Test Test

2/13/2024

1. A team of young scientists trying to make a name for themselves decides to study possums in Australia and New Guinea to determine if global climate change has effected the marsupials. The biologist think that weather conditions have decimated the possums' food supply. The file `possum.csv` is the results of their sample. In the year 2000 the average length of a possum (total_l) was known to be 87.8 mm. Is there evidence that possums have shrunk?

```
possum.data <- read.csv("possum.csv")
```

State hypotheses and alpha

mu = mean length of possum
H0: mu = 87.8
Ha: mu < 87.8

Explore data and collect sample statistics

```
#define your sample statistics; use R functions to define these when possible
#consider printing out the values to make sure everything is working
mu <- 87.8
xbar <- mean(possum.data$total_l) #sample mean
sx <- sd(possum.data$total_l) #sample standard deviation
n <- nrow(possum.data)
SE.xbar <- sx/sqrt(n) #standard error
```

Check assumptions and conditions

We are assuming that this sample of possums is random and representative, though the problem does not state it. Is the sample size n > 30? Yes, n = 104. Is n < 10% of population size? Yes, it seems safe to assume that there are over 1040 possums in Australia and New Guinea.

```
n
```

```
## [1] 104
```

Perform the test: pnorm or pt depending on type of test

```
#t-statistic
t <- (xbar - mu)/SE.xbar
t
```

```
## [1] -1.683381
```

```
#use pt to find pvalue (if testing >, then do 1 - pt)
pt(t, n-1)
```

```
## [1] 0.04766501
```

p-value = 0.0487

Make a conclusion using alpha

Because our p-value of 0.0487 is less than our alpha of .05, we can reject the null hypothesis that the mean length of possums is 87.8mm and accept the alternative hypothesis that the mean length of possums has decreased and is less than 87.8mm.

Confidence Interval

```
xbar - qt(.975,n-1)*SE.xbar
```

```
## [1] 86.25017
```

```
xbar + qt(.975,n-1)*SE.xbar
```

```
## [1] 87.92676
```

95% CI: 86.25 to 87.93 mm

Interpret CI and summarize findings

We are 95% confident that true mean length of possums in Australia and New Guinea is between 86.25 and 87.93mm. This complicates our hypothesis test results; we rejected the null hypothesis and accepted that the mean length of possums is shorter than 87.8mm, but 87.8mm falls in our confidence interval. If the scientists are hoping to prove a point that possums are suffering from the effects of climate change, then they will likely highlight their p-value and downplay the confidence interval they found. If scientists want to be thorough, they should probably just conduct further testing and not try to claim emphatically that possums have gotten shorter.

  2. What is a type I error in the context of problem 1? (closed note on paper)

The true mean length of possums is still 87.8 mm but the researchers conclude that it has decreased.

  3. What is a type II error in the context of problem 1, and what is a potential consequence of this type of error? (closed note on paper)

The researchers fail to reject the null hypothesis that the true mean length is 87.8mm, when in fact the true mean length of possums is less than 87.8mm. A consequence could be that no intervention is made to help solve the issue of the possums' decimated food supplies and the possums continue to get smaller.

  2. The scientists believe that there are more males than female possums in the population, and they decide to use their sample to test this as well. They will use an alpha of .05.

State hypotheses and alpha

p = proportion of possums that are male
H0: p = 0.5 (even proportion of males and females)
Ha: p > 0.5 (more males than females)

Explore data and collect sample statistics

```r
x <- sum(possum.data$sex == "m") #sum up the number of males in the sample
n <- nrow(possum.data)
p <- 0.5 #null hypothesis p
q <- 1-p
SD.p.hat <- sqrt(p*q/n)

p.hat <- x/n
q.hat <- 1 - p.hat

#print values to check that it worked:
p.hat
```

```
## [1] 0.5865385
```

```r
n
```

```
## [1] 104
```

Check assumptions and conditions

Again, we are assuming that this sample of possums is random and representative, though the problem does not state it, so therefore we have a constant p. We have a success/failure situation (male or female). We already checked that n is less than 10% of the population in the problem above. Now we just need to check the Large Enough condition: are np, nq > 10? Yes, as shown in the code chunk below.

```r
n*p
```

```
## [1] 52
```

```r
n*q
```

```
## [1] 52
```

Perform the test

```r
#testing if p > .5, so we use 1-pnorm
1 - pnorm(p.hat, p, SD.p.hat)
```

```
## [1] 0.03877808
```

p-value: 0.0388

3

Make a conclusion using alpha

Because our p-value of 0.0388 is less than alpha = .05, we can reject the null hypothesis that there are an even proportion of male and female possums and accept the alternative hypothesis that there is a greater proportion of male possums than female possums, which is the same as saying that the proportion of possums that are male is greater than .5.

Confidence Interval

```
#define the standard error of p.hat
SE.p.hat <- sqrt(p.hat*q.hat/n)

#determine the z* value for your chosen level of confidence
z.star <- qnorm(.975)

#check Large Enough conditions for a CI for proportions (use p.hat instead of p)
n*p.hat
```

```
## [1] 61
```

```
n*q.hat
```

```
## [1] 43
```

```
#construct your confidence interval using the formula we know well now!
p.hat - z.star*SE.p.hat
```

```
## [1] 0.4918936
```

```
p.hat + z.star*SE.p.hat
```

```
## [1] 0.6811834
```

Note that we checked the Large Enough condition for a confidence interval for proportions in the code chunk above and found that np.hat and nq.hat are both greater than 10, so we can proceed with constructing our CI.
95% CI: (.492, .681)

Interpret CI and summarize findings

We are 95% confident that between 49 and 67.6% of possums are male. Again, our CI has complicated our decision based on the p-value we found. We rejected the null hypothesis that 50% of possums are male, but 50% falls in the CI we constructed. As in the previous problem, the thorough thing to do would be to conduct further testing in hopes of coming to a more conclusive answer. But if the goal is to make a big discovery and get published, these researchers may choose to emphasize the p-value they found and put the CI in a well-hidden footnote!

4. For problem 3, if the scientists had instead used an alpha of .01, would they have made the same decision? Explain.

No, if the alpha value had been .01 the scientists would have failed to reject the null hypothesis that there are an even proportion of male and female possums. The p-value they found was .0388 which is greater than .01. Using a smaller alpha means that more convincing evidence is required in order to reject a null hypothesis.

5. Imagine that a new law was created that would impose a significant fine on airlines if more than 50% of their flights departed late. In order to test this the government takes a random sample of flights from each airline and uses it to determine if they deserve the fine. The file `flights.sample.csv` is a sample of 100 flights from Delta airlines. The variable `is.late` is TRUE if the flight departed late and FALSE if it departed on time. Does Delta deserves a fine? Perform a hypothesis test and construct a CI to answer this question.

```
flights.data <- read.csv("flights.sample.csv")
head(flights.data)
```

```
##   year month day dep_time dep_delay arr_time arr_delay carrier tailnum flight
## 1 2013     1  10     1955        -5     2055       -48      DL  N322NB    402
## 2 2013     8  25     1552        -7     1825       -18      DL  N981DL    847
## 3 2013     5  17     1459         0     1727        -5      DL  N660DL   2347
## 4 2013     5   5     2023        -2     2253       -32      DL  N3730B   1729
## 5 2013     5  15     1833         4     2025       -14      DL  N320US   2019
## 6 2013     8   2     1131        -4     1439       -12      DL  N944DL    954
##   origin dest air_time distance hour minute X1.nrow.fly. is.late
## 1    JFK  BOS       36      187   19     55         3417   FALSE
## 2    LGA  ATL       97      762   15     52         3331   FALSE
## 3    LGA  ATL      103      762   14     59         3142   FALSE
## 4    JFK  LAS      288     2248   20     23         1115   FALSE
## 5    LGA  MSP      151     1020   18     33         3033   FALSE
## 6    LGA  FLL      147     1076   11     31         1059   FALSE
```

State hypotheses and alpha

p = proportion of Delta flights that departed late
H0: p = .5
Ha: p > .5

###Explore data and collect sample statistics

```
x <- sum(flights.data$is.late == TRUE) #sum up the number late flights in sample
n <- nrow(flights.data)
p <- 0.5 #null hypothesis p
q <- 1-p
SD.p.hat <- sqrt(p*q/n)

p.hat <- x/n
q.hat <- 1 - p.hat

#print values to check that it worked:
p.hat
```

```
## [1] 0.42
```

```
n
```

```
## [1] 100
```

Check assumptions and conditions

We are told that this is a random sample of flights. We have a success/failure situation (late or not). n = 100, and we can assume that 100 is less than 10% of all Delta flights. Now we need to check the Large Enough condition: are np, nq > 10? Yes, both are equal to 50, as shown in the code chunk below.

```
n*p
```

```
## [1] 50
```

```
n*q
```

```
## [1] 50
```

Perform the test

```
#testing if p > .5, so we use 1-pnorm
1 - pnorm(p.hat, p, SD.p.hat)
```

```
## [1] 0.9452007
```

p-value: 0.945

Make a conclusion using alpha

Our p-value of 0.945 is greater than alpha = .05, so we fail reject the null hypothesis that 50% of Delta flights are late. We did NOT find convincing evidence to conclude that more than 50% of Delta flights are late. Looks like Delta will not be getting fined as we could not prove that they are in violation of this new law.

Confidence Interval

```
#define the standard error of p.hat
SE.p.hat <- sqrt(p.hat*q.hat/n)

#determine the z* value for your chosen level of confidence
z.star <- qnorm(.975)

#check Large Enough conditions for a CI for proportions (use p.hat instead of p)
n*p.hat
```

```
## [1] 42
```

```
n*q.hat
```

```
## [1] 58
```

```
#construct your confidence interval using the formula we know well now!
p.hat - z.star*SE.p.hat
```

```
## [1] 0.3232643
```

```
p.hat + z.star*SE.p.hat
```

```
## [1] 0.5167357
```

Note that we checked the Large Enough condition for a confidence interval for proportions in the code chunk above and found that np.hat and nq.hat are both greater than 10, so we can proceed with constructing our CI.
95% CI: (0.323, 0.517)

Interpret CI and summarize findings.

We are 95% confident that the true proportion of Delta flights that are late is between 32.3 and 51.7%. This CI supports the result of our hypothesis test; we failed to reject the null that the true proportion of late flights is 50%, and we can see here that 50% is indeed in our confidence interval.

6. What is a type I error in the context of problem 5, and a potential consequence?

A type I error here would be if we rejected the null hypothesis and conclude that more than 50% of Delta flights are late and they are violating the new law, when in reality that is not true. The consequence is that Delta would have to pay a fine they do not deserve! They may also get unfairly bad press as a result of this.

7. What is a type II error in the context of problem 5, and a potential consequence?

A type II error would be that more than 50% of Delta's flights are late, so they ARE in violation of this new law, but we fail to find sufficient evidence to reject the null hypothesis that they are not violating this law. A consequence here is that Delta gets away with inconveniencing many customers without having to pay a fine or get bad press.

8. Would Delta prefer the government use an alpha of .1 or .001 in this context? Make sure to explain why.

Delta would definitely prefer that the government use a smaller alpha, so .001. A smaller alpha means a lower chance they would be caught if they were in violation; the government would have to have more convincing evidence (a smaller p-value) than if the alpha was .05. A lower alpha will also mean a lower chance of type I error if they were not in violation of the rule, so a lower chance that they are falsely accused of violating the law.

9. The government is also looking into assessing fines to airlines based on average arrival delay. The variable `arr_delay` contains information about the number of minutes each flight arrived late (or early, if it's a negative number). Run the code below to make a subset of the flight data that only contains flights that arrived late.

```r
late.flights <- flights.data[flights.data$is.late == TRUE,]
```

The government is considering imposing a fine on airlines that have a mean arrival delay greater than 15 minutes. Perform a hypothesis test using an alpha of .05 and construct a confidence interval to see whether or not Delta would be fined under this new rule.

State hypotheses and alpha

mu = mean departure delay for Delta flights H0: mu = 15 minutes Ha: mu > 15 minutes

Explore data and collect sample statistics

```r
#define your sample statistics; use R functions to define these when possible
#consider printing out the values to make sure everything is working
mu <- 15
xbar <- mean(flights.data$arr_delay) #sample mean
sx <- sd(flights.data$arr_delay) #sample std dev
n <- nrow(flights.data)
SE.xbar <- sx/sqrt(n)

#check that code is working
sx
```

```
## [1] 54.14713
```

```r
xbar
```

```
## [1] 7.13
```

Check assumptions and conditions

We already discussed in the previous problem that the sample is random and than the sample size is less than 10% of the population. Is n > 30? n = 100, so yes, we have a Large Enough sample.

```r
n
```

```
## [1] 100
```

Calculate t statistic then perform the test

Perform the hypothesis test, and state the p-value OUTSIDE OF CODE CHUNK!

```r
#calculate t statistic and print it out
t <- (xbar - mu)/SE.xbar
t
```

```
## [1] -1.453447
```

```
#perform test: Ha is greater than, so use 1 - pt
1 - pt(t, n-1)
```

[1] 0.9253689

p-value: 0.925

Make a conclusion using alpha

Our p-value of .925 is greater than an alpha of .05. This means that we fail to reject the null hypothesis that Delta flights mean arrival delay of 15 minutes. We have NOT found convincing evidence to support our alternative hypothesis that the mean arrival delay of Delta flights is greater than 15 minutes. Looks like Delta will not have to pay this fine, either!

Confidence Interval

```
xbar - qt(.975,n-1)*SE.xbar
```

[1] -3.613966

```
xbar + qt(.975,n-1)*SE.xbar
```

[1] 17.87397

95% CI: -3.61 to 17.87 minutes

Interpret CI and summarize findings.

We are 95% confident that on average, Delta flights arrive between 3.61 minutes early and 17.87 minutes late. This supports the result of our hypothesis test, as 15 minutes falls in the confidence interval so it is reasonable to believe that the mean arrival delay could be 15 minutes. Delta would not have to pay this fine, based on our test and CI.

10. If the government does choose to impose this rule and wants to catch and fine as many airlines as possible, how can it increase the *power* of this test without changing alpha?

We did not discuss power.