Day 52 Hypothesis Test Assessment

2/15/2024

Honor Code

This is a test. You are expected to follow the honor code. You may not give or receive unauthorized aid.

Authorized Aid for this assessment includes: Rmd and pdf files that you have completed as well as answers and examples I have provided. Your notes.

Unauthorized Aid for this assessment includes: You may not use: internet sources (blog posts, webpages, manuals, etc.). You may not use generative AI for any part of this test. You may not consult any other person during this test.

On my honor I have given nor received unauthorized aid. *Type your name here to indicate you pledge of the honor code.*

Ryan Cheng

Hypotheses Tests and Confidence Intervals

1. In-N-Out Burger recently published an advertisement that stated that 70% of Americans preferred their burgers over the competition. A different burger company doesn't believe the claim and hires you to prove that In-N-Out is lying. You start by taking a random sample of 500 people and ask them what their favorite burger joint is. The file `burger.csv` contains the results of that survey. Set up an appropriate hypothesis test and confidence interval to determine if In-N-Out Burger is lying. Use an alpha of .05.

```
burgers <- read.csv("burger.csv")
```

a. State appropriate hypotheses and alpha

p = proportion of Americans that prefer In-N-Out Burgers over the competition H0: $p = 0.7$ Ha: $p < 0.7$ alpha = 0.05

b. Explore data and collect sample statistics

```
p <- 0.7
q <- 1 - p

x <- sum(burgers$best_burger_place == "In-N-Out Burger")
n <- nrow(burgers)

SD.p.hat <- sqrt(p*q/n)

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

c. Check assumptions and conditions

```r
n*p
```

```
## [1] 350
```

```r
n*q
```

```
## [1] 150
```

```r
n
```

```
## [1] 500
```

Success/failure: Yes, success = best burger place is In-N-Out, fail = best burge place is not In-N-Out np > 10: Yes, 350 > 10 nq > 10: Yes, 150 > 10 n < 10% of total population: Yes, 500 is less than 10% of total population Question says the sample is random.

d. Perform the test and find a p-value.

```r
pnorm(p.hat, p, SD.p.hat)
```

```
## [1] 0.2472623
```

p-value: 0.2472623

e. Make a conclusion using alpha

p-value of 0.2472623 is less than alpha of 0.05, so we failed to reject the null hypothesis. We did not find sufficient evidence that less than 70% of the American public prefers In-N-Out Burgers over the competition.

f. Construct a 95% confidence interval

```r
SE.p.hat <- sqrt(p.hat*q.hat/n)
z.star <- qnorm(0.975)
```

```r
n*p.hat
```

```
## [1] 343
```

```r
n*q.hat
```

```
## [1] 157
```

n*p.hat > 10: Yes, 343 > 10* n*q.hat > 10: Yes, 157 > 10

```r
p.hat - z.star*SE.p.hat
```

```
## [1] 0.6453192
```

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

```
## [1] 0.7266808
```

95% CI: (0.645, 0.727)

g. Interpret your confidence interval and summarize your findings. Is your employer happy?

We are 95% confident that the true proportion of Americans that prefer In-N-Out Burgers over the competition is between 64.5 and 72.7%. This supports our hypothesis test because In-N-Out's advertised 70% is within our CI, meaning that there is a possibility that 70% of American public truly prefer In-N-Out Burgers over the competition. Our employer may be happy to hear that our CI includes proportions below 70% but will probably not be happy to hear that the In-N-Out advertised 70% is still in our CI and that we could not find sufficient evidence to prove that In-N-Out Burger's claim was false in our hypothesis test.

2. The file `run10samp.csv` contains a sample of observations from the 2010 Cherry Blossom Run. The annual 10 mile run is very popular and regularly draws over 14,000 participants. In 2009 the participants average time was 90.87 minutes. In 2010 it rained for a few days before the race which lead to muddy conditions and in theory would make it slower. From the sample in `run10samp.csv` does it appear that the mean time in 2010 was slower than the mean time in 2009? Use an alpha of .05.

```r
run.data <- read.csv("run10samp.csv")
```

a. State appropriate hypotheses and alpha

mu = mean time H0: mu = 90.87 Ha: mu > 90.87 alpha = 0.05

b. Explore data and collect sample statistics

```
mu <- 90.87
xbar <- mean(run.data$time)
sx <- sd(run.data$time)
n <- nrow(run.data)
SE.xbar <- sx/sqrt(n)
```

c. Check assumptions and conditions

```
n
```

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

n > 30: Yes, 100 > 30 n < 10% of total population: Yes, 100 is less than 10% of 14,000. We are assuming that the sample is random and/or representative because the problem does not say so.

d. Perform the test and find a p-value.

```
t <- (xbar-mu)/SE.xbar
1 - pt(t, n-1)
```

```
## [1] 0.001670632
```

p-value: 0.001670632

e. Make a conclusion using alpha

p-value of 0.001670632 is less than alpha of 0.05, so we reject the null hypothesis and accept the alternative hypothesis. We found sufficient evidence to say that the average time of participants in 2010 was greater than that of participants in 2009 (90.87 minutes).

f. Construct a 95% confidence interval

```r
xbar - qt(0.975, n-1)*SE.xbar
```

```
## [1] 92.48412
```

```r
xbar + qt(0.975, n-1)*SE.xbar
```

```
## [1] 98.74508
```

95% CI: (92.5, 98.7)

g. Interpret your confidence interval and summarize your findings.

We are 95% confident that the mean time of participants in 2010 is between 92.5 and 98.7 minutes. This supports our hypothesis test because the 2009 average participant time of 90.87 minutes is not in and is below our CI. Yes, it appears that the mean participant time in 2010 was slower than the mean time in 2009.