STAT 511A Homework 8

Kathleen Wendt

11/01/2019

Load packages

```
library(readxl)
library(tidyverse)
library(emmeans)
library(broom)
```

Question 1

Return to the Problem 9.13 data (from HW7) which concerns a weight loss study. For consistency, perform the same "preprocessing" as we did last time.

Weight data

```
weight_data <- read_xlsx("ex9-13.xlsx") %>%
  gather(key = "trt", value = "loss") %>%
  mutate(trt = as_factor(trt)) %>%
  mutate(trt = fct_relevel(trt, "S"))
glimpse(weight_data)
```

```
## Observations: 50
## Variables: 2
## $ trt  <fct> A1, A1, A1, A1, A1, A1, A1, A1, A1, A1, A2, A2, A2, A2, A...
## $ loss <dbl> 12.4, 10.7, 11.9, 11.0, 12.4, 12.3, 13.0, 12.5, 11.2, 13....
```

```
levels(weight_data$trt)
```

```
## [1] "S"  "A1" "A2" "A3" "A4"
```

Use the following additional information about the Trts (called agents in the book):

S = Standard A1 = Drug therapy with exercise and with counseling A2 = Drug therapy with exercise but no counseling A3 = Drug therapy no exercise but with counseling A4 = Drug therapy no exercise and no counseling

Estimate and test the following contrasts. You just need to provide the estimate and p-value for each contrast.

Linear model

```r
weight_lm <- lm(loss ~ trt, data = weight_data)
```

Estimated means

```r
weight_em <- emmeans(weight_lm, "trt")
weight_em
```

```
##  trt emmean    SE df lower.CL upper.CL
##  S     9.27 0.313 45     8.64      9.9
##  A1   12.05 0.313 45    11.42     12.7
##  A2   11.02 0.313 45    10.39     11.7
##  A3   10.27 0.313 45     9.64     10.9
##  A4   12.24 0.313 45    11.61     12.9
##
## Confidence level used: 0.95
```

Part 1A through 1D Preparation

```r
weight_list <- contrast(weight_em, list(
    sVa =c(1, -.25, -.25, -.25, -.25),
    exerVno =c(0, 0.5, 0.5, -0.5, -0.5),
    counVno =c(0, 0.5, -0.5, 0.5, -0.5),
    counVs =c(1, -0.5, 0, -0.5, 0)
    ))
weight_list
```

```
##  contrast estimate    SE df t.ratio p.value
##  sVa         -2.12 0.350 45 -6.064  <.0001
##  exerVno      0.28 0.313 45  0.893  0.3764
##  counVno     -0.47 0.313 45 -1.500  0.1407
##  counVs      -1.89 0.384 45 -4.924  <.0001
```

Part 1A

Compare the mean for the standard agent versus the average of the means for the four other agents.

```r
weight_list[1]
```

```
##  contrast estimate   SE df t.ratio p.value
##  sVa         -2.12 0.35 45 -6.064  <.0001
```

Part 1B

Compare the mean for the agents with exercise versus those without exercise. (Ignore the standard.)

```r
weight_list[2]
```

```
##  contrast estimate    SE df t.ratio p.value
##  exerVno      0.28 0.313 45 0.893   0.3764
```

Part 1C

Compare the mean for the agents with counseling versus those without counseling. (Ignore the standard.)

```r
weight_list[3]
```

```
##  contrast estimate    SE df t.ratio p.value
##  counVno     -0.47 0.313 45 -1.500  0.1407
```

Part 1D

Compare the mean for the agents with counseling versus the standard.

```r
weight_list[4]
```

```
##  contrast estimate    SE df t.ratio p.value
##  counVs      -1.89 0.384 45 -4.924  <.0001
```

Question 2

Suppose Y is a binomial random variable with n = 22 and $\pi = 0.7$. Compute the following.

Part 2A

Mean and standard deviation of Y.

μ of Y = n x π = 22 x 0.7 = 15.4.

σ of Y = sqrt((n x π) x (1 - π)) = 2.1494185.

Part 2B

P(Y less than or equal to 15)

```r
pbinom(q = 15, size = 22, prob = 0.7)
```

```
## [1] 0.5058237
```

Part 2C

P(Y less than 15) = P(Y less than or equal to 14)

```r
pbinom(q = 14, size = 22, prob = 0.7)
```

```
## [1] 0.3287493
```

Part 2D

P(Y equal to 15)

```r
dbinom(x = 15, size = 22, prob = 0.7)
```

```
## [1] 0.1770744
```

Part 2E

P(15 less than or equal to Y less than 18) =

P(Y greater than or equal to 15) AND P(Y less than 18) = P(Y less or equal to 17)

```r
pbinom(q = 17, size = 22, prob = 0.7) - pbinom(q = 14, size = 22, prob = 0.7)
```

```
## [1] 0.5067019
```

Part 2F

P(Y greater than or equal to 18)

```r
1 - pbinom(17, size = 22, prob = 0.7)
```

```
## [1] 0.1645488
```

Part 2G

The normal approximation to P(Y greater than or equal to 18) without continuity correction.

Hand calculation

μ of Y = n x π = 22 x 0.7 = 15.4.

```r
mu <- 22*0.7
```

σ of Y = sqrt((n x π) x (1 - π)) = 2.1494185.

```r
sigma <- sqrt((22*0.7)*(1-0.7))
```

P(Y less than or equal to 17) = P((Y - mu)/sigma)) less than or equal to ((17-mu)/sigma).

((17-mu)/sigma) = 0.7443874

```
z_2g <- (17-mu)/sigma
```

P(Z less than or equal to 0.7443874) = 0.7704.

Therefore, P(Y greater than or equal to 18) is approximately 1 - 0.7704 = 0.2296.

Using `pnorm()` (exact)

```
1 - pnorm(z_2g)
```

```
## [1] 0.2283211
```

Part 2H

The normal approximation to P(Y greater than or equal to 18) with continuity correction.

Hand Calculation

```
mu
```

```
## [1] 15.4
```

```
sigma
```

```
## [1] 2.149419
```

```
z_2h <- (17.5 - mu) / sigma
z_2h
```

```
## [1] 0.9770084
```

P(Y less than or equal to 17.5) = P((Y - mu)/sigma)) is less than or equal to ((17.5 - mu)/sigma) = 0.9770084.

Therefore, P(Z less than or equal to 0.9770084) = 0.8340

Therefore, P(Y greater than or equal to 18) is approximately 1 = 0.8340 = 0.166.

Using `pnorm()` (exact)

```
1 - pnorm(z_2h)
```

```
## [1] 0.1642825
```

which is closer to the exact calculation for P(Y greater than or equal to 18):

```r
1 - pbinom(17, size = 22, prob = 0.7)
```

```
## [1] 0.1645488
```

Question 3

Problem 10.14: Many articles have written about the relationship between chronic pain and age of the patient. A survey of a random cross section of 800 adults who suffer from chronic pain found that 424 of them were above age 50. For this question, (1) use large sample normal approximation and (2) do the calculations "by hand".

Part 3A

Give an estimate for the proportion of persons suffering from chronic pain that are over 50 years of age.

$\hat{\pi}$ = y (events) / n (trials) = 424/800 = 0.53.

Part 3B

Give a 95% confidence interval on the proportion of persons suffering from chronic pain that are over 50 years of age.

$\hat{\pi}$ plus/minus Z(alpha/2) x (sqrt($\hat{\pi}$(1-$\hat{\pi}$)/n))

```r
pi_hat <- 424/800
n <- 800
z_crit <- qnorm(0.95)
se <- sqrt(pi_hat*(1-pi_hat)/n)
```

The 95% confidence interval for the proportion of persons suffering from chronic pain who are over 50 years of age is (0.5009752, 0.5590248).

Part 3C

Using the data in the survey, is there substantial evidence (alpha = 0.05) that more than half of persons suffering from chronic pain are over 50 years of age? In other words, test H0: pi less than or equal to 0.50 versus HA: pi greater than 0.50. Give the Z statistic, p-value and conclusion. (4 pts)

Test Statistic

```r
pi_null <- 0.50

z_ts <- (pi_hat-pi_null)/sqrt(pi_null*(1-pi_null)/n)
z_ts
```

```
## [1] 1.697056
```

P-value

```r
p_value <- 1 - pbinom(423, size = 800, prob = 0.50)
p_value
```

```
## [1] 0.04825579
```

Conclusion

Because 1.6970563 is greater than critical z-score of 1.6448536 and 0.0482558 is less than $\alpha = 0.05$, we reject the null hypothesis that the proportion of persons suffering from chronic pain is less than or equal to 0.50. There is evidence to suggest that π is greater than 0.50.

Question 4

A factory manager decided to estimate the proportion of defective items. A random sample of 50 items was inspected and it was found that 4 of them are defective.

```r
pi_hat_fac = 4/50
n_fac = 50
```

Part 4A

Give an estimate for the proportion of defective items.

$\hat{\pi} = 4/50 = 0.08$.

Part 4B

Using R, calculate a 90% confidence interval for the true proportion of defective items using the normal approximation.

NOTES: (1) Use correct = TRUE (default) and (2) The R CI will not match a hand calculation for this problem because R uses a different formula.

```r
factory_ci <- tidy(prop.test(x = 4,
                             n = 50,
                             conf.level = 0.90,
                             correct = FALSE))
factory_ci
```

```
## # A tibble: 1 x 8
##   estimate statistic p.value parameter conf.low conf.high method
##      <dbl>     <dbl>   <dbl>     <int>    <dbl>     <dbl> <chr>
## 1     0.08      35.3 2.86e-9         1   0.0364     0.167 1-sam~
## # ... with 1 more variable: alternative <chr>
```

The 90% CI for π of defective items is (0.036422, 0.1666979).

Part 4C

Using R, calculate a 90% confidence interval for the true proportion of defective items using the exact binomial method.

```
fact_ci_exact <- tidy(binom.test(x = 4,
                                 n = 50,
                                 conf.level = 0.90))
fact_ci_exact
```

```
## # A tibble: 1 x 8
##   estimate statistic  p.value parameter conf.low conf.high method
##      <dbl>     <dbl>    <dbl>     <dbl>    <dbl>     <dbl> <chr>
## 1     0.08         4 4.46e-10        50   0.0278     0.174 Exact~
## # ... with 1 more variable: alternative <chr>
```

Using the exact binomial method, the 90% CI for π of defective items is (0.0277877, 0.1737912).

Part 4D

Is the sample size large enough for the normal approximation to be valid? Justify your response using the criteria discussed in the notes (CH10 slide 23).

$\hat{\pi} = 4/50 = 0.08$

$\hat{\pi}$ should be more than 3 SEs away from 0 and 1 to justify the use of the large sample normal approximation.

```
n_check <- 3*(sqrt(pi_hat_fac*(1-pi_hat_fac)/n_fac))
n_check
```

```
## [1] 0.1151
```

$\hat{\pi}$ (0.08) is less than 0.1151, indicating that $\hat{\pi}$ is less than 3 standard errors from 0.

$\hat{\pi}$ (0.08) is less than 0.8849, indicating that $\hat{\pi}$ is less than 3 standard errors from 1.

The first condition is not satisfied, thus the large sample normal approximation is not appropriate.