# Stat 511 Hw8

*Amy Fox*

*10/27/2019*

Read in the necessary packages

```
library(tidyverse)
library(emmeans)
```

## 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.**

```
# tidy data
# reorder so S (standard is first for Dunnett's comparison)

weight_loss_data <- read.csv("../Data/OTT_Final/ASCII-comma/CH09/ex9-13.txt") %>%
  rename(A1 = X.A1.,
         A2 = X.A2.,
         A3 = X.A3.,
         A4 = X.A4.,
         S = X.S.) %>%
  gather(key = "Trt", value = "Loss") %>%
  mutate(Trt = as.factor(Trt) )%>%
 mutate(Trt = fct_relevel(Trt, "S"))

str(weight_loss_data)
```

```
## 'data.frame':    50 obs. of  2 variables:
##  $ Trt : Factor w/ 5 levels "S","A1","A2",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ Loss: num  12.4 10.7 11.9 11 12.4 12.3 13 12.5 11.2 13.1 ...
```

**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 pvalue for each contrast.**

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

```
one_way_fit_weight_loss <- lm(Loss ~ Trt, data = weight_loss_data)
emmout_weight_loss <- emmeans(one_way_fit_weight_loss, "Trt")

# order is S, A1, A2, A3, A4
contrast(emmout_weight_loss, list(S_v_all = c(1, -1/4, -1/4, -1/4, -1/4)))
```

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

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

```r
contrast(emmout_weight_loss, list(S_v_all = c(0, 1/2, 1/2, -1/2, -1/2)))
```

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

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

```r
contrast(emmout_weight_loss, list(S_v_all = c(0, 1/2, -1/2, 1/2, -1/2)))
```

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

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

```r
contrast(emmout_weight_loss, list(S_v_all = c(-1, 1/2, 0, 1/2, 0)))
```

```
##   contrast estimate    SE df t.ratio p.value
##   S_v_all      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.**

**A. Mean and standard deviation of Y.**

```r
n = 22
pi= 0.7

# mean
mu_Y <- n*pi
mu_Y
```

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

```r
# sd
sigma_y <- sqrt(n*pi*(1-pi))
sigma_y
```

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

**B. P(Y $\leq$ 15)**

```r
pbinom(15, size = n, prob = pi)
```

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

**C. P(Y $<$ 15)**

```r
pbinom(14, size = n, prob = pi)
```

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

**D. P(Y = 15)**

```r
dbinom(15, size = n, prob = pi)
```

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

**E. $P(15 \leq Y < 18)$**

```r
pbinom(17, size = n, prob = pi) - pbinom(14, size = n, prob = pi)
```

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

**F. $P(Y \geq 18)$**

```r
1-pbinom(17, size = n, prob = pi)
```

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

**G. The normal approximation to $P(Y \geq 18)$ without continuity correction.**

therefore, 1- $P(Y \leq 17)$

P((Y-mu_y)/sigma_y)

P((Y-15.4) / 2.149)

P((17-15.4) / 2.149)

```r
(17-15.4) / 2.149
```

```
## [1] 0.7445323
```

```r
1-pnorm(0.744)
```

```
## [1] 0.2284382
```

$P(0.744) = 0.7715$

Answer: 0.228

**H. The normal approximation to $P(Y \geq 18)$ with continuity correction.**

Note: add 0.5 and then do pnorm

P((Y-mu_y)/sigma_y)

P((Y-15.4) / 2.149)

P((17.5-15.4) / 2.149)

```r
(17.5-15.4) / 2.149
```

```
## [1] 0.9771987
```

```r
1-pnorm(0.9771987)
```

```
## [1] 0.1642354
```

# 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".**

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

```r
Y = 242
n = 800
pi_hat = 424/800
```

proportion estimate = 0.53

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

```r
z_alpha = 1.96 # got from table value in notes

pi_hat - z_alpha*sqrt(pi_hat*(1-pi_hat)/n)
```

```
## [1] 0.4954142
```

```r
pi_hat + z_alpha*sqrt(pi_hat*(1-pi_hat)/n)
```

```
## [1] 0.5645858
```

CI: (0.495, 0.564)

**C. 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: $\pi \leq 0.50$ versus HA: $\pi > 0.50$. Give the Z statistic, p-value and conclusion. (4 pts)**

```r
pi_not = 0.5

# by hand
# Z test statistic
z= (pi_hat - pi_not)/sqrt(pi_hat*(1-pi_hat)/n)
z
```

```
## [1] 1.700119
```

```r
# rejection region
qnorm(1-0.05)
```

```
## [1] 1.644854
```

```r
# p-value
1-pnorm(z)
```

```
## [1] 0.04455425
```

```r
# large sample normal approx - just to check
prop.test(424, 800, p = 0.5, alternative = "greater", correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  424 out of 800, null probability 0.5
## X-squared = 2.88, df = 1, p-value = 0.04484
## alternative hypothesis: true p is greater than 0.5
## 95 percent confidence interval:
##  0.5009229 1.0000000
## sample estimates:
##    p
## 0.53
```

Because the p-value $< 0.05$, we reject the null hypothesis and assume that the true probability $\pi$ is greater than 0.5

# 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.**

**A. Give an estimate for the proportion of defective items.**

```
4/50
```

```
## [1] 0.08
```

**B. Using R, calculate a 90% confidence interval for the true proportion of defective items using the normal approximation. NOTES: (1) Use correct = TRUE (default). (2) The R CI will not match a hand calculation for this problem because R uses a different formula.**

```
prop.test(4, 50, conf.level = .9, correct = TRUE)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  4 out of 50, null probability 0.5
## X-squared = 33.62, df = 1, p-value = 6.7e-09
## alternative hypothesis: true p is not equal to 0.5
## 90 percent confidence interval:
##  0.0301942 0.1792166
## sample estimates:
##    p
## 0.08
```

90% CI: (0.030, 0.179)

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

```
binom.test(4, 50, conf.level = 0.9)
```

```
##
##  Exact binomial test
##
## data:  4 and 50
## number of successes = 4, number of trials = 50, p-value =
## 4.462e-10
## alternative hypothesis: true probability of success is not equal to 0.5
## 90 percent confidence interval:
##  0.02778767 0.17379116
## sample estimates:
## probability of success
##                   0.08
```

90% CI: (0.028, 0.174)

**D. 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).**

```
pi_hat = 4/50
n = 50

pi_hat > 3*sqrt((pi_hat*(1-pi_hat))/n)
```

```
## [1] FALSE
```

```r
pi_hat < 1- 3*sqrt((pi_hat*(1-pi_hat))/n)
```

```
## [1] TRUE
```

Both of these equations must be true to be considered valid. Pi hat needs to be more than 3 SE from 0 and 1, but it is not based on the first equation. Therefore, the sample size is not large enough to be valid for the normal approximation.