Exam2 Spring 2020

Natalie Schmer

Signature for honor pledge: Natalie Schmer

Multiple Choice

(hint: two spaces at end of a line starts new line when knitting to pdf. otherwise numbered lines like below do so automatically.)

1 - 7 True/False

1. T
2. F
3. T
4. T
5. F
6. T
7. F

```
n7 <- 55
pi7 <- 0.11

3*sqrt(pi7*(1-pi7)/n7)

## [1] 0.1265701
```

8-13 A,B,C,or D

8. B
9. C
10. B
11. B
12. D
13. D

Matching (place a unique letter next to each)

Scenario1: B
Scenario2: A
Scenario3: G
Scenario4: F
Scenario5: E
Scenario6: D
Scenario7: C

R Code Questions

1. Sleep Data

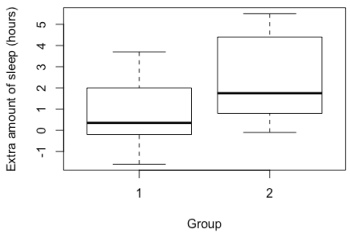
```
sleep <- sleep
str(sleep)

## 'data.frame':    20 obs. of  3 variables:
##  $ extra: num  0.7 -1.6 -0.2 -1.2 -0.1 3.4 3.7 0.8 0 2 ...
##  $ group: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ ID   : Factor w/ 10 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 9 10 ...

?sleep
summary(sleep)

##      extra        group        ID
##  Min.   :-1.600   1:10   1      :2
##  1st Qu.:-0.025   2:10   2      :2
##  Median : 0.950          3      :2
##  Mean   : 1.540          4      :2
##  3rd Qu.: 3.400          5      :2
##  Max.   : 5.500          6      :2
##                          (Other):8
```

1A. Hypotheses

The parameters are the null and alternative hypotheses. The null hypothesis is that there is no difference in how the 2 drugs increase sleep compared to the control– the means of extra sleep in hours for each group are not different. The alternative hypothesis is that one of the drugs increases sleep more than the other as compared to the control– the mean extra sleep in hours for each group are not equal to eachother.

1B. Boxplot

```
boxplot(extra ~ group, data = sleep, xlab = "Group", ylab = "Extra amount of sleep (hours)")
```



The boxplot may be misleading because it it showing that the means of the groups are different, when they may not be when tested statistically. Also, the two groups are both compared to a control, with which the sleep in hours is not shown here.

1C. t.test can be done at least 3 different ways

```
t.test(extra ~ group, data = sleep, paired = T)
```

```
##
##  Paired t-test
##
## data:  extra by group
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.4598858 -0.7001142
## sample estimates:
## mean of the differences
##                   -1.58
```

1D. Conclusion

Since p < 0.05, we reject the null hypothesis and conculde that the mean increased sleep is not the same between the 2 groups. Based on the boxplot, it appears that the drug given to the second group increased sleep more than the drug in the first group.

2. Proportion of lefties then and now

2A. Hypotheses

This is a proportion, so the parameters of interest include: number of trials, n = 150, and Y, number of successes = 18, with the main parameter being pi hat, the proportion of successes, in this case being 18/ 150 or 0.12. The null hypothesis is that the proprotion of Americans who are left handed has stayed the same since the 1980's, and the alternative is that the proprtion of left handed Americans has increased since the 1980's.

2B. Test Statistic (multiple ways)

```
#z- score for alpha = 0.05
qnorm(0.95) #1.645
```

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

```
#test stat:
(0.12 - 0.08)/(sqrt((0.08*(1-0.08))/150)) #1.81
```

```
## [1] 1.805788
```

2C. p-value (multiple ways)

```
prop.test(18, 150, p = 0.05, correct = T)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  18 out of 150, null probability 0.05
```

```
## X-squared = 14.035, df = 1, p-value = 0.0001794
## alternative hypothesis: true p is not equal to 0.05
## 95 percent confidence interval:
##  0.0746155 0.1855432
## sample estimates:
##    p
## 0.12
```

2D.

From part b, z=1.81 > z(alpha/2)= 1.96, so part b would say to reject the null hypothesis that the proprotion of americans who are left handed has stayed the same since the 1980's. From the prop.test in part c, the pvalue is < 0.05, which is also sufficient evidence that the proportion of left handed americans is higher at the time of the study than in the 1980's.

3 Cuckoos

```
{
  library(tidyverse)
  library(car)
  library(emmeans)
}

eggs.wide <-
read.csv('/Users/natalieschmer/Desktop/GitHub/stats_511/data/cuckoo.csv')

str(eggs.wide)
```

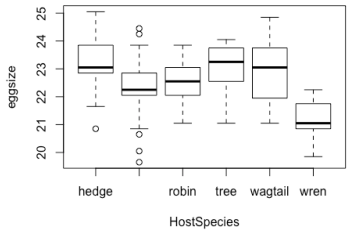
```
## 'data.frame':    45 obs. of  6 variables:
##  $ meadow : num  19.6 20.1 20.6 20.9 21.6 ...
##  $ tree   : num  21.1 21.9 22.1 22.4 22.6 ...
##  $ hedge  : num  20.9 21.6 22.1 22.9 23.1 ...
##  $ robin  : num  21.1 21.9 22.1 22.1 22.1 ...
##  $ wagtail: num  21.1 21.9 21.9 21.9 22.1 ...
##  $ wren   : num  19.9 20.1 20.2 20.9 20.9 ...
```

```
eggs <- gather(eggs.wide,"HostSpecies","eggsize",na.rm=TRUE) #missing values
stacked
str(eggs)
```

```
## 'data.frame':    120 obs. of  2 variables:
##  $ HostSpecies: chr  "meadow" "meadow" "meadow" "meadow" ...
##  $ eggsize    : num  19.6 20.1 20.6 20.9 21.6 ...
```

3A. boxplot

```
boxplot(eggsize ~ HostSpecies, data = eggs)
```



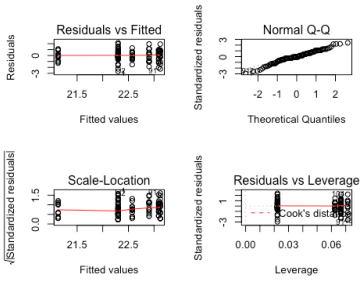
3B. Sumstats

```
eggs %>%
            group_by(HostSpecies) %>%
            summarise(n = n(),
                      mean = mean(eggsize),
                      sd = sd(eggsize),
                      se = sd/sqrt(n))
```

```
## # A tibble: 6 x 5
##   HostSpecies     n  mean    sd    se
##   <chr>       <int> <dbl> <dbl> <dbl>
## 1 hedge          14  23.1 1.07  0.286
## 2 meadow         45  22.3 0.921 0.137
## 3 robin          16  22.6 0.685 0.171
## 4 tree           15  23.1 0.901 0.233
## 5 wagtail        15  22.9 1.07  0.276
## 6 wren           15  21.1 0.744 0.192
```

3C. Diagnoatics

```
Fit_eggs = lm(eggsize ~ HostSpecies, data = eggs)

#Set up plot space and plot
par(mfrow= c(2, 2))
plot(Fit_eggs)
```




```
#Levene's test
car::leveneTest(eggsize ~ HostSpecies, data = eggs, )

## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   5  0.6397 0.6698
##       114
```

```
#Shapiro - Wilk
(shapiro.test(eggs$eggsize))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  eggs$eggsize
## W = 0.98241, p-value = 0.1193
```

This data appears to be ok for an anova. Residuals are generally very close to 0, except for the last two groups, and the QQ plot has a line that is very close to straight but slightly deviates at the bottom left. Additionally, the Levene p value is > 0.05 so would not be non-normal, and the shapiro-wilk p value= 0.12, also not non-normal. If anything, the differences in sample size where meadow has many more observations could be a problem, but the summary stats are very similar to the other groups.

3D. ANOVA table

```
anova(Fit_eggs)
```

```
## Analysis of Variance Table
##
## Response: eggsize
##              Df Sum Sq Mean Sq F value    Pr(>F)
## HostSpecies   5 42.940  8.5879  10.388 3.152e-08 ***
## Residuals   114 94.248  0.8267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3E.

Since F < 0.05, at least one population group mean appears to be different.

3F. CLD

```
(eggs_cld<- emmeans::emmeans(Fit_eggs, pairwise ~ HostSpecies))
```

```
## $emmeans
##  HostSpecies emmean    SE  df lower.CL upper.CL
##  hedge         23.1 0.243 114     22.6     23.6
##  meadow        22.3 0.136 114     22.0     22.6
##  robin         22.6 0.227 114     22.1     23.0
##  tree          23.1 0.235 114     22.6     23.6
##  wagtail       22.9 0.235 114     22.4     23.4
##  wren          21.1 0.235 114     20.7     21.6
##
## Confidence level used: 0.95
```

```
##
## $contrasts
##  contrast         estimate    SE  df t.ratio p.value
##  hedge - meadow     0.8225 0.278 114  2.956  0.0429
##  hedge - robin      0.5464 0.333 114  1.642  0.5726
##  hedge - tree       0.0314 0.338 114  0.093  1.0000
##  hedge - wagtail    0.2181 0.338 114  0.645  0.9872
##  hedge - wren       1.9914 0.338 114  5.894  <.0001
##  meadow - robin    -0.2761 0.265 114 -1.043  0.9022
##  meadow - tree     -0.7911 0.271 114 -2.918  0.0475
##  meadow - wagtail  -0.6044 0.271 114 -2.230  0.2325
##  meadow - wren      1.1689 0.271 114  4.312  0.0005
##  robin - tree      -0.5150 0.327 114 -1.576  0.6160
##  robin - wagtail   -0.3283 0.327 114 -1.005  0.9155
##  robin - wren       1.4450 0.327 114  4.422  0.0003
##  tree - wagtail     0.1867 0.332 114  0.562  0.9932
##  tree - wren        1.9600 0.332 114  5.903  <.0001
##  wagtail - wren     1.7733 0.332 114  5.341  <.0001
##
## P value adjustment: tukey method for comparing a family of 6 estimates
```

```
CLD(eggs_cld)
```

```
##  HostSpecies emmean    SE  df lower.CL upper.CL .group
##  wren          21.1 0.235 114     20.7     21.6  1
##  meadow        22.3 0.136 114     22.0     22.6   2
##  robin         22.6 0.227 114     22.1     23.0   23
##  wagtail       22.9 0.235 114     22.4     23.4   23
##  tree          23.1 0.235 114     22.6     23.6    3
##  hedge         23.1 0.243 114     22.6     23.6    3
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 6 estimates
## significance level used: alpha = 0.05
```

Based on the above, it apears that the wren host group tends to have smaller egg sizes than the others.

3G Contrasts

3G part i.

The null hypothesis would be that there is not a significant difference in the mean sizes of wren and meadowlark eggs together as compared to robins and wagtails together.

3G part ii.

```
eggs_g_model <- lm(eggsize ~ HostSpecies, data = eggs)

eggs_emmeans <- emmeans(eggs_g_model, "HostSpecies")
```

```r
#find order of factors
levels(factor(eggs$HostSpecies))
```

```
## [1] "hedge"   "meadow"  "robin"   "tree"    "wagtail" "wren"
```

```r
#don't want hedge or tree, positions 1 or 4

contrast(eggs_emmeans, list(partg = c(0, 0.5, -0.5, 0, -0.5, 0.5)))
```

```
##  contrast estimate    SE  df t.ratio p.value
##  partg       -1.02 0.212 114 -4.827  <.0001
```