Midterm Output

Lillian Chen

10/29/2021

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.5      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
#check of anova table for question 4
#numbers do not match up exactly - calculations for the table utilized the printed numbers
#on the partially filled out table

options(knitr.kable.NA = '')
k <- 4
N <- 13+13+14+14
df_b <- k-1
df_w <- N-k
df_tot <- N-1

mean11 <- 341.69
mean12 <- 312.92
mean13 <- 321.43
mean14 <- 302.00
(grandmean <- (13*mean11 + 13*mean12 + 14*mean13 + 14*mean14)/N)
```

```
## [1] 319.2213
```

```
ssb <- 13*(mean11-grandmean)^2 + 13*(mean12-grandmean)^2 +
       14*(mean13-grandmean)^2 + 14*(mean14-grandmean)^2
ssw <- 12*(18.12)^2 + 12*(23.20)^2 + 13*(21.72)^2 + 13*(17.02)^2
msb <- ssb/df_b
msw <- ssw/df_w
Fstat <- msb/msw
p <- 1 - pf(Fstat, k-1, N-k)
```

```
q4 <- data.frame(Source = c("Between groups (Treatment)",
                            "Within groups (Error)",
                            "Total"),
                 df = c(k-1, N-k, N-1),
                 SS = c(ssb, ssw, ssb+ssw),
                 MS = c(msb, msw, (ssb+ssw)/(N-1)),
                 F = c(Fstat, NA, NA),
                 p = c(p, NA, NA))

kbl(q4, booktabs = T,
    caption = "ANOVA Table Question 4") %>%
  kable_styling(latex_options = c("hold_position"))
```

Table 1: ANOVA Table Question 4

Source	df	SS	MS	F	p
Between groups (Treatment)	3	11299.46	3766.4855	9.278156	5.54e-05
Within groups (Error)	50	20297.60	405.9519		
Total	53	31597.05	596.1708		

```
(samplemean <- grandmean)
```

```
## [1] 319.2213
```

```
(s_p <- sqrt(405.942))
```

```
## [1] 20.148
```

```
q4g <- data.frame(groupmeans = c(341.69,312.92,321.43,302.00),
                  groupsds = c(18.12,23.20,21.72,17.02))
q4g %>%
  summarise(across(c(groupmeans,groupsds), sd, .names = "{.col}.sd"))
```

```
##   groupmeans.sd groupsds.sd
## 1      16.78955    2.921889
```

```
(r <- (sum((q4g$groupmeans-samplemean)*(q4g$groupsds-s_p)))/(4-1))/(16.78955*2.921889))
```

```
## [1] -0.05067413
```

```
#check
cor(q4g$groupmeans, q4g$groupsds, use="pairwise.complete.obs")
```

```
## [1] -0.0496305
```

```
q5data <- data.frame(x=c(1,2,4,5,10),
                     y=c(4,3,2,1,10))

#check for question 5a
cor(q5data$x, q5data$y, use="pairwise.complete.obs")
```

```
## [1] 0.7056681
```

```
#check for question 5b
q5lm <-lm(y~x,data=q5data)
summary(q5lm)
```

```
##
## Call:
## lm(formula = y ~ x, data = q5data)
##
## Residuals:
##       1       2       3       4       5
##   2.4187  0.7073 -1.7154 -3.4268  2.0163
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.8699     2.2284   0.390    0.722
## x             0.7114     0.4124   1.725    0.183
##
## Residual standard error: 2.893 on 3 degrees of freedom
## Multiple R-squared:  0.498,  Adjusted R-squared:  0.3306
## F-statistic: 2.976 on 1 and 3 DF,  p-value: 0.183
```

```
#question 5c
(s_e <- summary(q5lm)$sigma)
```

```
## [1] 2.892613
```

```
(se_b1 <- s_e/sqrt(sum((q5data$x-4.4)^2)))
```

```
## [1] 0.4123896
```

```
(se_b0 <- s_e*sqrt(1/5 + (4.4)^2/sum((q5data$x-4.4)^2)))
```

```
## [1] 2.228431
```

```
#remove 10,10
q5ddata <- data.frame(x=head(q5data$x,-1),
                      y=head(q5data$y,-1))
```

```
#calculate new means and sds
q5ddata %>%
  summarise(across(c(x,y), mean, .names = "{.col}.smean"),
            across(c(x,y), var, .names = "{.col}.svar"))
```

```
##   x.smean y.smean   x.svar   y.svar
## 1       3     2.5 3.333333 1.666667
```

```
#new sample correlation
(r <- cov(q5ddata$x,q5ddata$y)/(3.333*1.667))
```

```
## [1] -0.419958
```

```
#check for question 5e
q5dlm <- lm(y~x, data=q5ddata)
summary(q5dlm)
```

```
##
## Call:
## lm(formula = y ~ x, data = q5ddata)
##
## Residuals:
##    1    2    3    4
##  0.1 -0.2  0.2 -0.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.60000    0.23979  19.183  0.00271 **
## x           -0.70000    0.07071  -9.899  0.01005 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2236 on 2 degrees of freedom
## Multiple R-squared:   0.98,  Adjusted R-squared:   0.97
## F-statistic:    98 on 1 and 2 DF,  p-value: 0.01005
```