Homework 3

Write your name here

24 April, 2022

```
library(tidyverse)
```

Section 1

Background: For this section, we will work with a new data set from an experiment regarding the effect of naps in the consolidation of information in memory. A group of researchers is interested in understanding the effect that taking a nap can have on memory consolidation. They designed an experiment where participants are assigned randomly to one of 3 conditions. In the control condition, participants first study a list of 100 words and, after one hour, they are asked to write down all the words they can remember. In the first experimental condition, participants are given the same list. After studying the list, each participant takes a 15 min nap and then waits 45 min (so that a total of 1 hr has passed since they studied the list) before they are asked to write down all the words they can remember from the list. In the second experimental condition, participants are given the same list and after a study time they are asked to take a 1 hour nap. Once the hour has passed, participants are asked to write down all the words from the list that they can remember.

We have access to the data of the experiment which includes an *id* assigned to each participant, the number of correctly *recalled* words and the experimental condition that they are assigned to. Due to the nature of the experiment we are interested in 4 models: (1) the Null model which assumes that there are no differences in the total number of recalled words between conditions, (2) a "**15 minutes**" model which assumes that the total number of correctly recalled words is different only for participants that had the 15 nap, (3) a "**60 minutes**" model which assumes that the total number of correctly recalled words is only different for participants who had a 1 hour nap, and (4) the effects model which assumes that each group is different.

```
link <- "https://raw.githubusercontent.com/ManuelVU/psych-10c-data/main/homework3-p1.csv"
nap <- read_csv(link)
```

```
## Rows: 135 Columns: 3
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): condition
## dbl (2): id, recalled
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

1. Obtain the predictions of the Null model, add the required values as a new variable (column) to the data set (remember to give this new variable a name!). What is the total number of correctly recalled words predicted by the Null model for each group?

```
# total sample size
n_total <- nrow(nap)
# predictions of null model
pred_null <- nap %>%
```

```
  summarise("pred" = mean(recalled)) %>%
  pull(pred)
# add predictions as a new variable
nap <- nap %>%
  mutate("prediction_null" = pred_null)
```

ANS: 26.2148148 [**0.7 points for a correct value**]

2. Calculate the squared error (SE) of each observation under the Null by adding a new column to our data set. What is the mean Sum of Squared Errors (m-SSE) of the Null model?

```
# obtain error by observation
nap <- nap %>%
  mutate("error_null" = (recalled - prediction_null)^2)
# obtain sse null
sse_null <- sum(nap$error_null)
# obtain mse null
mse_null <- 1/n_total * sse_null
```

ANS: 10.1390398 [**0.7 points for the correct answer**]

3. What is the value of the Bayesian Information Criterion (BIC) for the Null model?

```
bic_null <- n_total * log(mse_null) + log(n_total)
```

ANS: 317.6183699 [**0.7 for correct value**]

4. Obtain the predictions of the Effects model and add those values as a new variable (column) to the data set (remember to give this new variable a DIFFERENT name!). What are the total number of correctly recalled words predicted by the Effects model for each group?

```
# prediction of the effects model
pred_effects <- nap %>%
  group_by(condition) %>%
  summarise("pred" = mean(recalled))
# add predictions as new column
nap <- nap %>%
  mutate("prediction_effects" =
           case_when(condition == "control" ~ pred_effects$pred[3],
                     condition == "15_min" ~ pred_effects$pred[1],
                     condition == "60_min" ~ pred_effects$pred[2]))
```

ANS: The prediction for the control condition was 23.7777778 for the 15 min nap condition it was 26.8666667 and for the 60 minutes group it was 28 [**0.3 for each correct value**]

5. Calculate the squared error (SE) for each observation under the Effects model by adding a new column to the data set (again, be careful with the names you give to the new column!). What is the mean Sum of Squared Errors (m-SSE) of the Effects model?

```
# add error by observation
nap <- nap %>%
  mutate("error_effects" = (recalled - prediction_effects)^2)
# obtain sse effects model
sse_effects <- sum(nap$error_effects)
# obtain mse effects model
mse_effects <- 1/n_total * sse_effects
```

ANS: 6.9553909 [**0.7 for correct value**]

6. What is the value of the Bayesian Information Criterion (BIC) for the Effects model?

```
bic_effects <- n_total * log(mse_effects) + 3 * log(n_total)
```

ANS: 276.550624 [**0.7 points for the correct value**]

7. Obtain the predictions of the 15 minutes model and add those values as a new variable (column) to the data set (remember to give this new variable a name!). What are the total number of correctly recalled words predicted by the 15 minutes model for each group?

```
# add indicator of 15 vs other
nap <- nap %>%
  mutate("id_15min" = ifelse(test = condition == "15_min",
                             yes = "15_min", no = "other"))
# obtain predictions
pred_15min <- nap %>%
  group_by(id_15min) %>%
  summarise("pred" = mean(recalled))
# add predictions to data
nap <- nap %>%
  mutate("prediction_15min" = ifelse(test = id_15min == "15_min",
                                     yes = pred_15min$pred[1],
                                     no = pred_15min$pred[2]))
```

ANS: The prediction for the 15 min nap group was 26.8666667 for the other 2 groups it was 25.8888889 [**0.4 for one correct value, 0.7 for two correct values**]

8. Calculate the squared error (SE) of each observation under the 15 minute model by adding a new column to our data set. What is the mean Sum of Squared Errors (m-SSE) of the 15 minutes model?

```
# add observation error
nap <- nap %>%
  mutate("error_15min" = (recalled - prediction_15min)^2)
# obtain sse for 15 min model
sse_15min <- sum(nap$error_15min)
# obtain mse for 15 min model
mse_15min <- 1/n_total * sse_15min
```

ANS: 9.9265844 [**0.7 points for correct value**]

9. What is the value of the Bayesian Information Criterion (BIC) for the 15 minutes model?

```
# obtain bic for the 15 min model
bic_15min <- n_total * log(mse_15min) + 2 * log(n_total)
```

ANS: 319.6647699 [**0.7 points for correct value**]

10. Obtain the predictions of the 60 minutes model, add the required values as a new variable (column) to the data set (remember to give this new variable a name!). What are the total number of correctly recalled words predicted by the 60 minutes model for each group?

```
# create indicator variable for 60 min group
nap <- nap %>%
  mutate("id_60min" = ifelse(test = condition == "60_min",
                             yes = "60_min", no = "other"))
# obtain predictions by group
pred_60min <- nap %>%
  group_by(id_60min) %>%
  summarise("pred" = mean(recalled))
# add prediction to data file
```

```r
nap <- nap %>%
  mutate("prediction_60min" = ifelse(test = id_60min == "60_min",
                                     yes = pred_60min$pred[1],
                                     no = pred_60min$pred[2]))
```

ANS: The prediction for the 60 min group was 28 and for the other groups it was 25.3222222. **[0.4 for one correct value, 0.7 for two correct values]**

11. Calculate the squared error (SE) of each observation under the 60 minutes by adding a new column to our data set. What is the mean Sum of Squared Errors (m-SSE) of the 60 minutes model?

```r
# obtain error by observation
nap <- nap %>%
  mutate("error_60min" = (recalled - prediction_60min)^2)
# obtain sse for 60 min model
sse_60min <- sum(nap$error_60min)
# obtain mse for 60 min model
mse_60min <- 1/n_total * sse_60min
```

ANS: 8.5455967 **[0.7 for correct value]**

12. What is the value of the Bayesian Information Criterion (BIC) for the 60 minutes model?

```r
# obtain bic for 60 min model
bic_60min <- n_total * log(mse_60min) + 2 * log(n_total)
```

ANS: 299.4417292 **[0.7 for correct value]**

13. Which of the four models would you select and why?

ANS: However they phrase it the key is to choose the **effects model** because it has the lowest BIC. **[0.7 for correct value]

14. What does the Model that you selected tell us about our research question?

ANS: Something on the lines of:

- Interpret the results according to model selected in (13) but they don't use means to talk about the order of groups **[0.5 points]**
- Interpret the results according to selected model in (13) and use predictions to say that some groups have higher recall than others **[0.7 points]**

15. Interpret the value of R^2 associated with the model that you selected.

ANS: Something on the lines of:

- Interpret result as proportion when reporting a percentage or vice versa **[0.4 points]**
- Interpret the results as proportion or percentage of explained variance and present correct values (according to model selected on 13) **[0.7 points]**
  - Effects $R^2 = 0.314$
  - 15 min $R^2 = 0.021$
  - 60 min $R^2 = 0.157$