Homework 5 - Key

Manuel Villarreal

6/3/2022

```
link <- "https://raw.githubusercontent.com/ManuelVU/psych-10c-data/main/homework5.csv"
grades <- read_csv(file = link)
```

1. **0.4 points**

```
betas_null <- lm(formula = grade ~ 1, data = grades)$coef
grades <- grades %>%
  mutate("prediction_null" = betas_null,
         "error_null" = (grade - prediction_null)^2)
```

ANS: The prediction of the null model is equal to the estimate of β_0

Model	Prediction
Null	73.84

2. **0.8 points**

```
betas_sw <- lm(formula = grade ~ study_week, data = grades)$coef

grades <- grades %>%
  mutate("prediction_sw" = betas_sw[1] + betas_sw[2] * study_week,
         "error_sw" = (grade - prediction_sw)^2)
```

ANS: The values of the parameters in the model are:

Parameter	Estimate
β_0	69.11
β_1	1.62

3. **0.8 points**

```
betas_cm <- lm(formula = grade ~ classes_missed, data = grades)$coef

grades <- grades %>%
  mutate("prediction_cm" = betas_cm[1] + betas_cm[2] * classes_missed,
         "error_cm" = (grade - prediction_cm)^2)
```

ANS: The estimated values of the parameters in the model are:

Parameter	Estimate
β_0	75.44
β_1	-0.94

4. **0.8 points**

```
betas_swcm <- lm(formula = grade ~ study_week + classes_missed, data = grades)$coef

grades <- grades %>%
  mutate("prediction_swcm" = betas_swcm[1] + betas_swcm[2] * study_week +
           betas_swcm[3] * classes_missed,
         "error_swcm" = (grade - prediction_swcm)^2)
```

ANS: The estimated value of the parameters of this model are:

Parameter	Estimate
β_0	70.97
β_1	1.5
β_2	-0.9

5. **0.8 points**

```
grades <- grades %>%
  mutate("classes_id" = case_when(classes_missed == 0 ~ 0,
                                  classes_missed > 0 ~ 1))
```

ANS: The proportion of students that missed at least one class was **0.38**.

6. **0.8 points**

```
betas_swcid <- lm(formula = grade ~ study_week + classes_id, data = grades)$coef

grades <- grades %>%
  mutate("prediction_swcid" = betas_swcid[1] + betas_swcid[2] * study_week +
           betas_swcid[3] * classes_id,
         "error_swcid" = (grade - prediction_swcid)^2)
```

ANS: The estimated values of the parameters of the model are:

Parameter	Estimate
β_0	71.47
β_1	1.46
β_2	-4.99

7. **0.8 points**

```
betas_int <- lm(formula = grade ~ study_week + classes_id + study_week * classes_id, data = grades)$coef

grades <- grades %>%
  mutate("prediction_int" = betas_int[1] + betas_int[2] * study_week +
           betas_int[3] * classes_id +
           betas_int[4] * classes_id * study_week,
         "error_int" = (grade - prediction_int)^2)
```

ANS: The estimated values of the parameters of the model are:

Parameter	Estimate
β_0	69.96
β_1	1.96

Parameter	Estimate
β_2	-0.78
β_3	-1.47

8. **0.8 points (code is the answer)**

```
sse_null <- sum(grades$error_null)
sse_sw <- sum(grades$error_sw)
sse_cm <- sum(grades$error_cm)
sse_swcm <- sum(grades$error_swcm)
sse_swcid <- sum(grades$error_swcid)
sse_int <- sum(grades$error_int)
```

9. **0.8 points (code is the answer)**

```
n_total <- nrow(grades)

mse_null <- n_total * sse_null
mse_sw <- n_total * sse_sw
mse_cm <- n_total * sse_cm
mse_swcm <- n_total * sse_swcm
mse_swcid <- n_total * sse_swcid
mse_int <- n_total * sse_int
```

10. **0.8 points**

```
bic_null <- n_total * log(mse_null) + 1 * log(n_total)
bic_sw <- n_total * log(mse_sw) + 2 * log(n_total)
bic_cm <- n_total * log(mse_cm) + 2 * log(n_total)
bic_swcm <- n_total * log(mse_swcm) + 3 * log(n_total)
bic_swcid <- n_total * log(mse_swcid) + 3 * log(n_total)
bic_int <- n_total * log(mse_int) + 4 * log(n_total)
```

ANS: The values of the BIC are the following (take 0.2 points if they miss the labels for at least one BIC value).

Parameter	Estimate
Null	1236.63
Hours	1222.34
Classes	1210.44
Hours + Classes	1192.58
Hours + ID	1189.19
Interaction	1188.05

11. **0.8 points**

ANS: According to the BIC the best model is the interaction model, its parameters can be interpreted as (order of the values depends on the order they assign in the lm function):

- β_0: expected grade of students that missed no classes and studied for 0 hours each week.

- β_1: impact of an additional hour of study on the expected grade of students that missed 0 classes.

- β_2: difference on the expected grade of students who missed at least one class when they study for 0 hours a week in comparison to students that missed no classes.

- β_3 difference in the impact of an additional hour of study on the expected grade of students that missed at least one class in comparison to students that missed no classes.

12. **0.8 points**

```
r2_int <- (sse_null - sse_int) / sse_null
```

ANS: The proportion of variance accounted for by the model was **0.46**.

13. **0.8 points**

ANS: Something along the lines of more hours of study are associated with better grades, however, the impact of each additional hour depends on whether students missed at least one class. Students that missed at least one class have a lower grade on average in comparison to students that missed no classes when they spend 0 hours a week studying.