Lesson 14: Hopelessness and Exercise

LTC J.K.Starling

Last compiled on 16 February, 2022

A recent study (Dunn et al., J. Cardiopulm. Rehab. Prevent., 2017) gathered data on patients who had recently had a heart attack requiring hospitalization. Before patients were discharged from the hospital, patients took a series of surveys which included measures of chronic hopelessness. Chronic hopelessness was measured based on a series of Likert scale questions which were averaged together yielding a Chronic Hopelessness score of 1–4. Higher scores indicate a more hopeless feeling. Researchers were interested in learning about whether exercise after a heart attack was associated with hopelessness levels eight months later. Two-hundred and twenty-three (n = 223) patients have data measuring hopelessness eight months after hospitalization, as well as information about their exercise habits during the eight-month period following hospitalization. Exercise was classified as “Regular home exerciser” (exercised at home at least three times per week) or “Not regular home exerciser” (did not exercise at home at least three times per week).

```
### Loading packages
library(tidyverse)
library(ggResidpanel)
library(car)
## Load data
# hope <- read.table('https://raw.githubusercontent.com/jkstarling/MA376/main/hopelessness8months_2group.txt', header = TRU
E, stringsAsFactors = TRUE)

setwd('C:/Users/james.starling/OneDrive - West Point/Teaching/MA376/JimsLessons/Block II/LSN14/')
hope <- read.table('hopelessness8months_2group.txt', header = TRUE, stringsAsFactors = TRUE)

# add effect coding for Exercise and BaselineGroup
contrasts(hope$Exercise) <- contr.sum
contrasts(hope$BaselineGroup) <- contr.sum
```

1. State the null and alternative hypothesis for the research question to learn whether exercise after a heart attack was associated with hopelessness levels eight months later. Identify the explanatory and response variables, and whether they are categorical or quantitative.

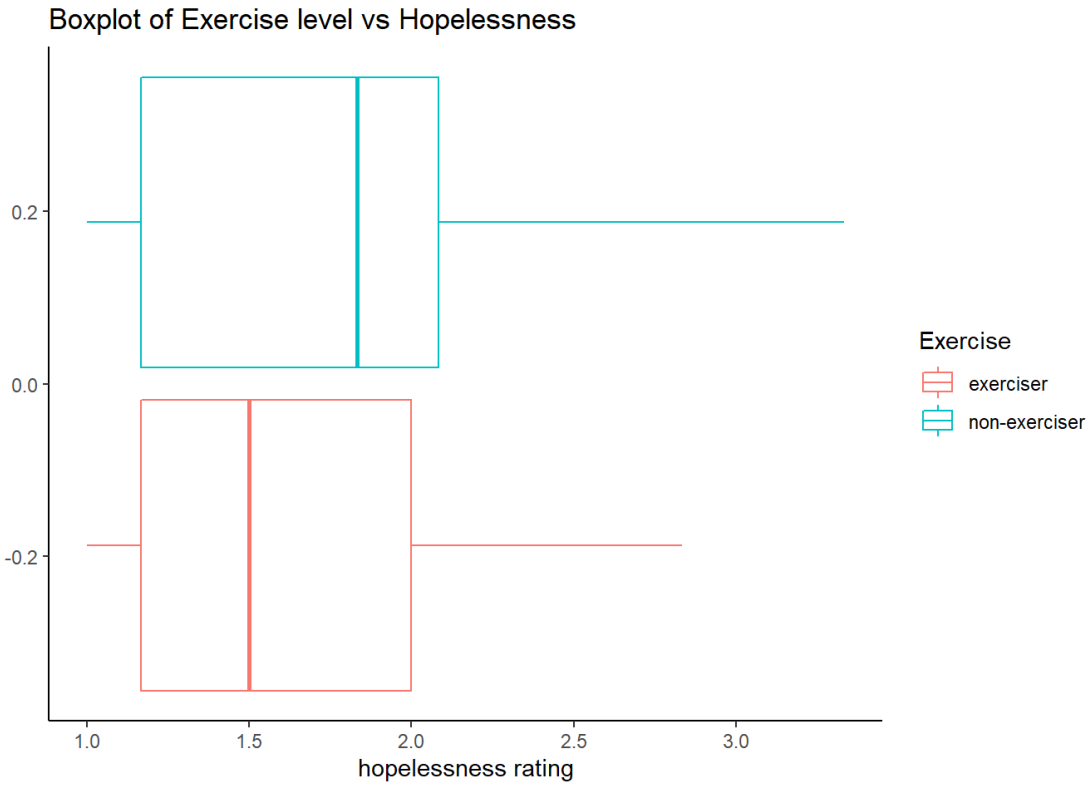
H_0 : There is not a genuine association between exercise and 8 month hopelessness in the population. H_a : There is a genuine association between exercise and 8 month hopelessness in the population. Explanatory variable – exercise, categorical Response variable – 8month hopelessness, quantitative

2. Was this an observational study or a randomized experiment? If the latter, are there any blocking variables to account for?

Observational study, data was simply collected and no assignment was employed.

3. Plot the distributions of hopelessness 8 months after a heart attack between the two exercise levels. Also provide the mean, standard deviation, and count for each level of exercise. Is there preliminary evidence to suggest that there is a relationship between exercise status and hopelessness level? What are the estimated effects for exercisers and non-exercisers?

```
hope %>% ggplot(aes(x=hopelessness8, color=Exercise)) +
  geom_boxplot() +
  labs(title="Boxplot of Exercise level vs Hopelessness", x='hopelessness rating') +
  theme_classic()
```




```
hope %>% group_by(Exercise) %>% summarise(n=n(), mn = mean(hopelessness8), sd = sd(hopelessness8))
```

Exercise <fct>	n <int>	mn <dbl>	sd <dbl>
exerciser	91	1.545557	0.4600434
non-exerciser	83	1.777557	0.6260262

2 rows

```
1.78-1.55
```

```
## [1] 0.23
```

```
(1.78-1.55) / 2
```

```
## [1] 0.115
```

Yes, the average exerciser has a hopelessness score that is 0.23 points lower than the average nonexerciser. Exercisers have an effect of -0.115 while non-exercisers have an effect of 0.115.

Single-variable model

4. Conduct a one-variable analysis evaluating the association between exercise and hopelessness 8 months after a heart attack and provide an ANOVA table. How much of the variation in hopelessness scores is explained by whether or not the individual exercises? Save the sum of squared error total as `SST_ex` .

```
hope.lm <- lm(hopelessness8 ~ Exercise, data=hope)
summary(hope.lm)
```

```
##
## Call:
## lm(formula = hopelessness8 ~ Exercise, data = hope)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.77756 -0.44422 -0.04556  0.43811  1.55578
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.6616     0.0414  40.135  < 2e-16 ***
## Exercise1    -0.1160     0.0414  -2.802  0.00566 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5455 on 172 degrees of freedom
## Multiple R-squared:  0.04365,    Adjusted R-squared:  0.03809
## F-statistic: 7.851 on 1 and 172 DF,  p-value: 0.005661
```

```
anova(hope.lm)
```

	Df <int>	Sum Sq <dbl>	Mean Sq <dbl>	F value <dbl>	Pr(>F) <dbl>
Exercise	1	2.336396	2.3363960	7.851266	0.005661068
Residuals	172	51.184116	0.2975821	*NA*	*NA*

2 rows

```
SST_ex <- 2.336 + 51.184
2.336 / SST_ex
```

```
## [1] 0.04364723
```

Exercise explains about 4.4% of the variation in hopelessness scores.

Two-variable model

5. Conduct a two-variable analysis evaluating the association between exercise and hopelessness 8 months after a heart attack and provide an ANOVA table. Give the prediction equation and residual standard error. What variables are significant (at the $\alpha = 0.05$ level)? (Load the package `car` and use `Anova(hope.lm2, type="3")`)

```
hope.lm2 <- hope %>% lm(hopelessness8 ~ Exercise+BaselineGroup, data = .)
summary(hope.lm2)
```

```
##
## Call:
## lm(formula = hopelessness8 ~ Exercise + BaselineGroup, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97686 -0.34560 -0.06331  0.32107  1.48773
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.66123    0.03839  43.275  < 2e-16 ***
## Exercise1      -0.10885    0.03841  -2.834  0.00515 **
## BaselineGroup1  0.20677    0.03837   5.389 2.32e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5058 on 171 degrees of freedom
## Multiple R-squared:  0.1825, Adjusted R-squared:  0.1729
## F-statistic: 19.09 on 2 and 171 DF,  p-value: 3.298e-08
```

```
# anova(hope.lm2)
Anova(hope.lm2, type="3")
```

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
(Intercept)	479.168107	1	1872.706891	4.829797e-94

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
Exercise	2.054982	1	8.031374	5.151001e-03
BaselineGroup	7.430475	1	29.040123	2.321737e-07
Residuals	43.753642	171	*NA*	*NA*

4 rows

pediction hopelessness = 1.66 + ((-0.1089 if exercise) (0.1089 if not exercise) + ((-0.2068 if not hopeless) (0.2068 if moderate to severe))

The residual standard error is 0.5058.

Both variables are significant at the $lpha = 0.05$ level.

6. Provide the model R^2. What is the value of the covariation? (Hint: You will need to use `SST_ex` calculated above.)

```
SSexercise <- 2.05
SSbasegroup <- 7.43
SSE <- 43.75
SScovariation <- SST_ex - SSexercise - SSbasegroup - SSE
SScovariation
```

```
## [1] 0.29
```

```
# rsquared
(SSexercise + SSbasegroup + SScovariation) / SST_ex
```

```
## [1] 0.1825486
```

```
SScovariation
```

```
## [1] 0.29
```

R^2 is 0.1825; SScovariation = 0.29.

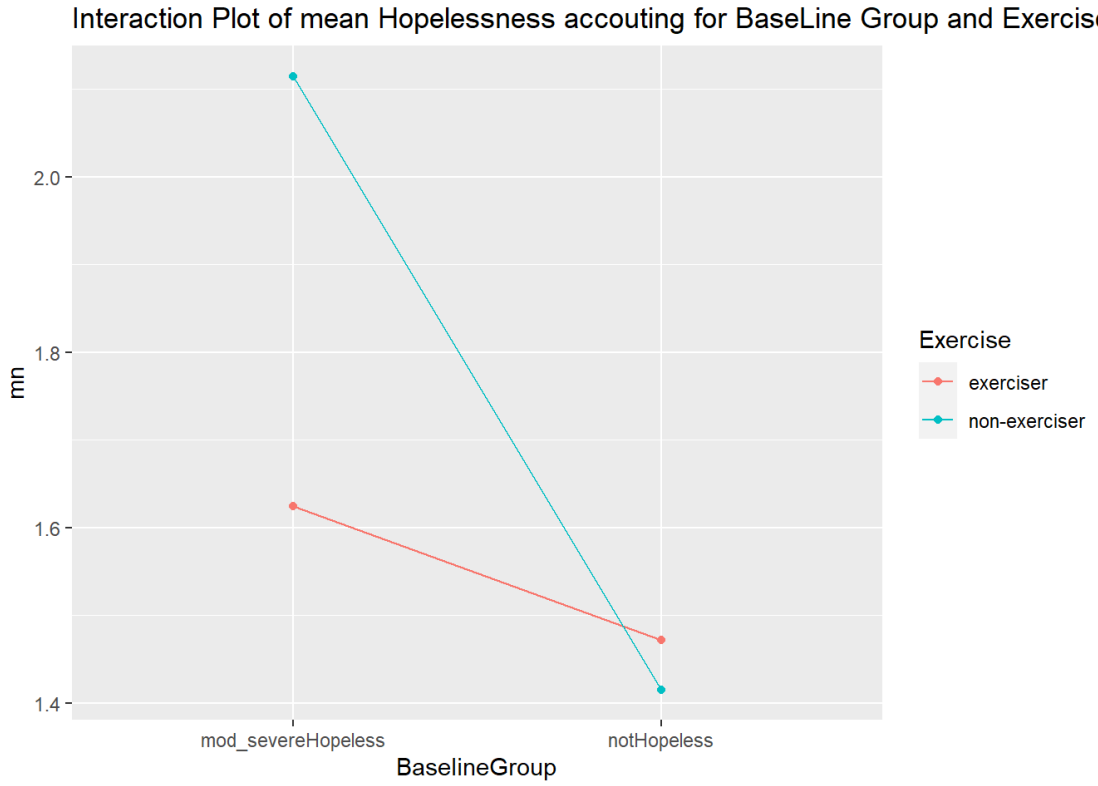
Interaction

7. Create an interaction plot between Baseline hopelessness (x-axis) and Exercise (group & color). What story is being told with this graph?

```
hope.int <- hope %>% group_by(Exercise, BaselineGroup) %>% summarise(mn=mean(hopelessness8))
```

```
## `summarise()` has grouped output by 'Exercise'. You can override using the `.groups` argument.
```

```
hope.int %>% ggplot(aes(x=BaselineGroup, y = mn, group=Exercise, color=Exercise)) +
  geom_line() + geom_point() +
  labs(title="Interaction Plot of mean Hopelessness accouting for BaseLine Group and Exercise")
```



There is a larger difference in the 8 month hopelessness scores between exercisers and nonexercisers of those who are moderately to severely hopeless than those who are not hopeless. There appears to be an interaction between Exercise and BaselineGroup.

8. Conduct a two-variable analysis evaluating the association between exercise and hopelessness 8 months after a heart attack WITH INTERACTION. Provide an ANOVA table. Give the prediction equation and residual standard error. What variables are significant (at the $\alpha = 0.05$ level)? What is the R^2 value?

```
hope.lm3 <- hope %>% lm(hopelessness8 ~ Exercise*BaselineGroup, data = .)
summary(hope.lm3)
```

```
##
## Call:
## lm(formula = hopelessness8 ~ Exercise * BaselineGroup, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1146 -0.3050 -0.1146  0.3755  1.4181
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.65650    0.03707  44.684  < 2e-16 ***
## Exercise1                 -0.10842    0.03707  -2.925 0.003919 **
## BaselineGroup1             0.21307    0.03707   5.748  4.1e-08 ***
## Exercise1:BaselineGroup1 -0.13662    0.03707  -3.685 0.000307 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4882 on 170 degrees of freedom
## Multiple R-squared:  0.243,  Adjusted R-squared:  0.2296
## F-statistic: 18.19 on 3 and 170 DF,  p-value: 2.78e-10
```

```
Anova(hope.lm3, type="3")
```

	Sum Sq <dbl>	Df <dbl>	F value <dbl>	Pr(>F) <dbl>
(Intercept)	475.871352	1	1996.668706	7.069690e-96
Exercise	2.038583	1	8.553521	3.919184e-03
BaselineGroup	7.873143	1	33.034259	4.101988e-08
Exercise:BaselineGroup	3.237090	1	13.582237	3.067372e-04
Residuals	40.516551	170	*NA*	*NA*

5 rows

With the addition of the interaction term the model still accounts for a significant amount of variation in the model. In context, the model utilizing exercise, baseline hopelessness, and the interaction between the two explains a significant amount of variation in 8 month hopelessness.

$R^2 = 0.243$; SE resid = 0.488.

9. Explain how to get the degrees of freedom for SSM (not shown), SSexercise, SSBaselineGroup, SSinteraction, and SSresiduals. Four total treatments, df model = 3 (4-1); Two levels of exercise, df exercise = 1 (2-1); Two levels of baseline of hopelessness, df baseline = 1 (2-1); Interaction uses the remaining df of the model, df interaction = 1 (3-1-1); df residuals = 170 (174 - 1 - 3).

10. What are the null and alternative hypotheses for the “model” F-test (with interaction)? What do you conclude about the model? H_0: The model does not explain variation in 8 month hopelessness.; H_a: The model explains variation in 8 month hopelessness. The F-statistic of 18.19 and p-value of < 0.0001 give very strong evidence to reject the null hypothesis.

11. What are the null alternative hypotheses for the interaction F-test? What do you conclude about the interaction between baseline hopelessness and exercise? What percentage of the variation in eight-month hopelessness is explained by this interaction? H_0: The model does not explain variation in 8 month hopelessness.; H_a: The interaction term explains variation in 8 month hopelessness, after adjusting for exercise and baseline hopelessness.

The F-statistic of 13.58 and p-value of 0.0003 for the interaction term give strong evidence to reject the null hypothesis. We have convincing evidence of an interaction effect in the population, with the interaction explaining approximately 6% of the variation in 8-month hopelessness.

12. Calculate the SS for the Model (SSM). What is the covariation of this model? How does it compare to the previous model?

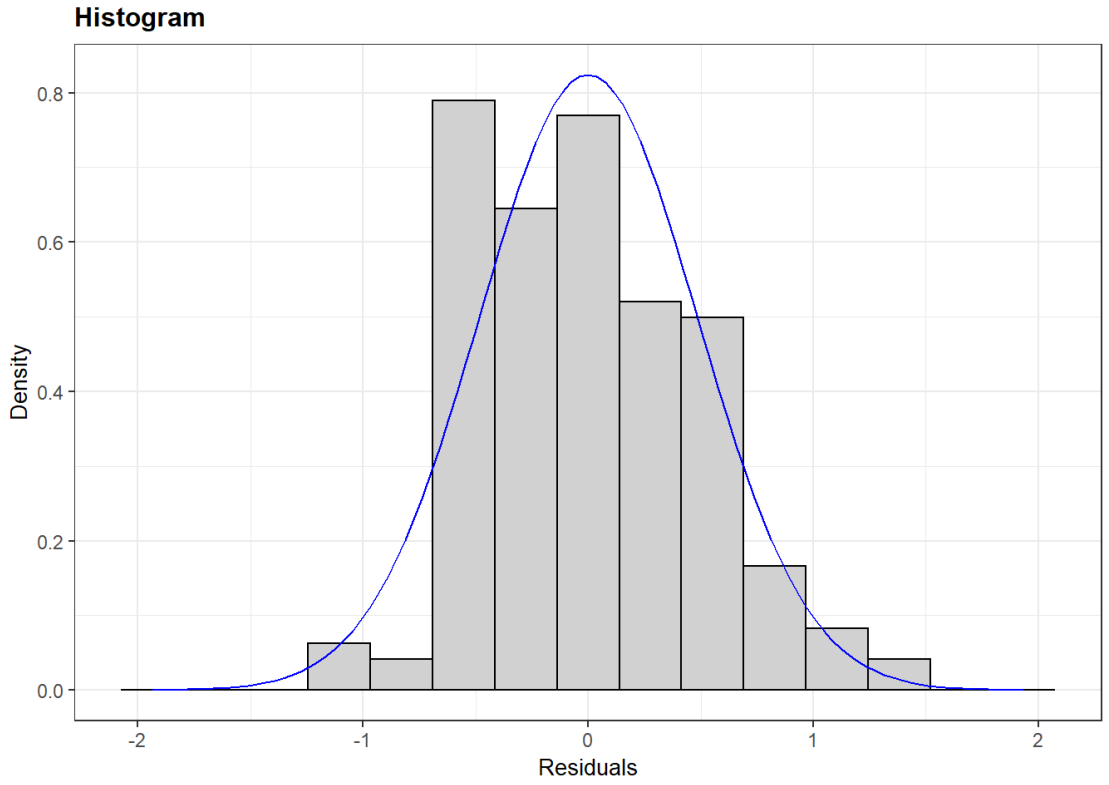
```
SScovariation2 <- SST_ex - 40.52 - 3.24 - 7.87 - 2.04; SScovariation2
```

```
## [1] -0.15
```

13. How does the SSError change when comparing the two-variable model with and without the interaction? Why? The SSError decreases from 43.75 to 40.52 due to the increased amount of explained variation.

14. Discuss whether or not the validity conditions are met for this analysis. Cite specific characteristics of graphs or numbers from the output to support your statements.

```
resid_panel(hope.lm3, plots = 'hist', bins = 15)
```




```
hope %>% group_by(Exercise, BaselineGroup) %>%
  summarise(n=n(), mn=mean(hopelessness8), sd = sd(hopelessness8))
```

```
## `summarise()` has grouped output by 'Exercise'. You can override using the `.groups` argument.
```

Exercise <fct>	BaselineGroup <fct>	n <int>	mn <dbl>	sd <dbl>
exerciser	mod_severeHopeless	44	1.624523	0.4532048
exerciser	notHopeless	47	1.471631	0.4588359
non-exerciser	mod_severeHopeless	43	2.114609	0.5827913
non-exerciser	notHopeless	40	1.415226	0.4453482

4 rows

```
# hope <- hope %>% unite('ExBase', 'Exercise', 'BaselineGroup', remove=FALSE)
# hope %>% ggplot(aes(x=hopelessness8, color=ExBase)) + geom_boxplot()
```

1) The individuals were only recorded once in the study confirming independence. 2) Standard deviations are similar. 3) The distribution of the residuals is appx. normal. There is a slight right skewing though.

References

Nathan Tintle et al.(2019). Intermediate Statistical Investigations.