Categorical predictors and nonlinear models

Gemma Wallace & Neil Yetz

PSY 653 Module 9 Lab

Apr 8, 2020

Demo Overview

We are practicing three different ways of examining categorical predictors in a regression framework:

1) Dummy coding
2) Effect coding
3) Orthogonal Polynomial Contrasts

Note: our outcome variables are continuous in each of these examples

Load Libraries

````
# Load Libraries
```{r}
library(tidyverse)
library(olsrr)
library(psych)
```
````

Read in Data

```
# Read in dataset
```{r}
slp <- read_csv("slpdata.csv")
```

Parsed with column specification:
cols(
  cond = col_double(),
  prior = col_double(),
  age = col_double(),
  anxiety = col_double(),
  hygiene = col_double(),
  support = col_double(),
  sleep = col_double(),
  lifesat = col_double(),
  sex = col_double(),
  id = col_double()
)
```

This is the same slpdata we've used in previous lab activities.

We'll use these data for parts 1-2 of the demo, for dummy coding and effect coding

Our research question for parts 1-2:

To what extent do treatment condition, sex, and the interaction between these two variables predict sleep hygiene?

We will show two different ways to approach this question using different coding methods for the categorical predictor variables: dummy coding and effect coding

These two methods are similar, but the coding and interpretations are slightly different.

Part 1: Dummy coding

What is dummy coding and why use it?

One of the most common and simplest approaches to evaluating categorical predictors in psychology

Dummy coding allows you to compare the mean difference between two levels of a categorical variable: the level that is coded as a 1 versus the level that is coded as 0

- You can specify any level of the variable to be the reference group (i.e., the level coded as 0)
- Create a new “dummy coded” binary variable for every comparison you want to make between two groups
- You only need to use dummy coding if you have more than two levels of a categorical predictor

Dummy coding

- For dummy coding, we will be converting categorical variables into a series of binary variables.
- For all but one of the levels of the categorical variable, a new variable will be created that has a value of 1 for each observation at that level and 0 for all others.

Level of race	New variable 1 (x1)	New variable 2 (x2)	New variable 3 (x3)
1 (Hispanic)	1	0	0
2 (Asian)	0	1	0
3 (African American)	0	0	1
4 (white)	0	0	0

Reference category

https://stats.idre.ucla.edu/spss/faq/coding-systems-for-categorical-variables-in-regression-analysis/

Specify dummy codes

Syntax of ifelse() function

```
ifelse(test_expression, x, y)
```

TRUE | FALSE

````
## Specify dummy codes for categorical predictors
```{r}
#sex & condition
slp <- mutate(slp,
              sex = ifelse(sex == 1, 0, 1),
              sex.f = factor(sex, levels = c(0,1), labels = c("male", "female")))

slp <- mutate(slp,
              cond2 = ifelse(cond == 2, 1, 0),
              cond3 = ifelse(cond == 3, 1, 0))
```
````

We created two new variables:
<u>cond2</u> is a dummy coded binary variable in which condition 2 is coded as 1 and condition 1 is coded as 0. This variable allows us to compare the mean difference in Y between conditions 2 and 1.

<u>cond3</u> is a is a dummy coded binary variable in which condition 3 is coded as 1 and condition 1 is coded as 0. This variable allows us to compare the mean difference in Y between conditions 3 and 1.

Run model with dummy coded condition variable

```
## Run model
```{r}
m1 <- lm(hygiene ~ cond2 + cond3, data = slp)
ols_regress(m1)
```
```

Note: We aren’t including sex and the interaction term in this first model to make the initial interpretation simpler.

```r
## Run model
```{r}
m1 <- lm(hygiene ~ cond2 + cond3, data = slp)
ols_regress(m1)
```
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.630       RMSE                1.224
R-Squared               0.397       Coef. Var          20.418
Adj. R-Squared          0.395       MSE                 1.498
Pred R-Squared          0.391       MAE                 0.973
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                                ANOVA
----------------------------------------------------------------------
                Sum of
               Squares        DF    Mean Square       F          Sig.
----------------------------------------------------------------------
Regression     589.467         2        294.733    196.774     0.0000
Residual       894.205       597          1.498
Total         1483.671       599
----------------------------------------------------------------------

                                  Parameter Estimates
-------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta      t        Sig     lower    upper
-------------------------------------------------------------------------------------
(Intercept)     4.660         0.087                53.853    0.000    4.490    4.830
      cond2     1.627         0.122        0.488   13.293    0.000    1.386    1.867
      cond3     2.374         0.122        0.712   19.399    0.000    2.134    2.615
-------------------------------------------------------------------------------------
```

Interpretations

Intercept: The predicted sleep hygiene score when all x variables are zero, so participants in Condition 1.

cond2: the predicted difference in sleep hygiene score between participants in condition 2 compared to condition 1.

cond3: the predicted difference in sleep hygiene score between participants in condition 3 compared to condition 1.

Run model with dummy coded condition, sex & interaction

```
## Run model
```{r}


m1 <- lm(hygiene ~ sex.f + cond2 + cond3 + sex.f*cond2 + sex.f*cond3, data = slp)
ols_regress(m1)

```
```

```
## Run model
```{r}


m1 <- lm(hygiene ~ sex.f + cond2 + cond3 + sex.f*cond2 + sex.f*cond3, data = slp)
ols_regress(m1)

```
```

Model Summary

R	0.771	RMSE	1.007
R-Squared	0.594	Coef. Var	16.801
Adj. R-Squared	0.591	MSE	1.014
Pred R-Squared	0.585	MAE	0.799

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	881.213	5	176.243	173.768	0.0000
Residual	602.459	594	1.014		
Total	1483.671	599			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	3.751	0.098		38.346	0.000	3.559	3.943
sex.ffemale	1.935	0.143	0.418	13.562	0.000	1.655	2.215
cond2	2.084	0.132	0.625	15.813	0.000	1.825	2.343
cond3	2.962	0.135	0.888	21.933	0.000	2.697	3.227
sex.ffemale:cond2	-0.643	0.207	-0.119	-3.113	0.002	-1.048	-0.237
sex.ffemale:cond3	-1.160	0.203	-0.223	-5.710	0.000	-1.558	-0.761

Interpretations

Intercept: The predicted sleep hygiene score when all x variables are zero, so males in Condition 1.

female: This variable is involved in an interaction, so it's a simple slope. Specifically, it is the effect of female when both cond2 = 0 and cond3 = 0, so people in Condition 1. Therefore, it is the predicted difference in sleep hygiene between females and males in Condition 1. The slope is positive, meaning that females in Condition 1 tend to have better sleep hygiene than males in Condition 1. This is a statistically significant difference (p-value is less than alpha).

cond2: This variable is involved in an interaction, so it is a simple slope. It is the effect of being in Condition 2 (compared to Condition 1) when female = 0, so among males. It is the predicted difference in sleep hygiene for males in Condition 2 compared to males in Condition 1. This is a statistically significant difference (p-value is less than alpha).

cond3: This variable is involved in an interaction, so it is a simple slope. It is the effect of being in Condition 3 (compared to Condition 1) when female = 0, so among males. It is the predicted difference in sleep hygiene for males in Condition 3 compared to males in Condition 1. This is a statistically significant difference (p-value is less than alpha).

female:cond2: The predicted differential effect of being in Condition 2 compared to Condition 1 for females compared to males. This is a statistically significant difference (p-value is less than alpha). The coefficient for cond2 presents the effect (i.e., benefit) of being in Condition 2 (compared to 1) for males. To get the effect for females, we take the effect for males (2.084) and add the female:cond2 interaction term (-.643). Therefore, the effect (i.e., benefit) of being in Condition 2 (compared to Condition 1) for females is 1.441.

female:cond3: The predicted differential effect of being in Condition 3 compared to Condition 1 for females compared to males. This is a statistically significant difference (p-value is less than alpha). The coefficient for cond3 presents the effect (i.e., benefit) of being in Condition 3 (compared to 1) for males. To get the effect for females, we take the effect for males (2.962) and add the female:cond3 interaction term (-1.160). Therefore, the effect (i.e., benefit) of being in Condition 3 (compared to Condition 1) for females is 1.802.

Part 2: Effect Coding

What is effect coding and why use it?

Similar to dummy coding, except here, you are comparing one level of a categorical predictor to the *mean* of all of the levels.

Instead of asking “are two conditions different from each other?” using dummy coding, effect coding asks “is this condition different from average?”

While the "rule" in dummy coding is that only values of 0 and 1 are valid, the "rule" in effect coding is that all of the values in any new variable must sum to zero.

Specify effect coding variables

```{r}
slp <- mutate(slp,

              cond2.ec = ifelse(cond == 2, 1, 0),
              cond3.ec = ifelse(cond == 3, 1, 0),

              cond2.ec = ifelse(cond == 1, (-1), cond2.ec),
              cond3.ec = ifelse(cond == 1, (-1), cond3.ec))
```

We created two new variables:
<u>cond2.ec</u> is a effect coded variable in which condition 2 is coded as 1, condition 1 is coded as -1, and condition 3 is coded as 0. This variable allows us to compare the mean difference in Y between condition 2 and the average score across all conditions.
<u>cond3.ec</u> is a is a dummy coded binary variable in which condition 3 is coded as 1, condition 1 is coded as -1, and condition 2 is coded as 0. This variable allows us to compare the mean difference in Y between condition 3 and the average score across all conditions.

```r
## Run model without interaction first
```{r}
m2_no_interaction <- lm(hygiene ~ cond2.ec + cond3.ec, data = slp)
ols_regress(m2_no_interaction)
```
```

```
                        Model Summary
-------------------------------------------------------------
R                       0.630       RMSE               1.224
R-Squared               0.397       Coef. Var         20.418
Adj. R-Squared          0.395       MSE                1.498
Pred R-Squared          0.391       MAE                0.973
-------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
---------------------------------------------------------------------
                Sum of
               Squares        DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression     589.467         2        294.733    196.774    0.0000
Residual       894.205       597          1.498
Total         1483.671       599
---------------------------------------------------------------------

                                Parameter Estimates
------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta       t        Sig     lower    upper
------------------------------------------------------------------------------------
(Intercept)     5.994         0.050                 119.969    0.000    5.896    6.092
   cond2.ec     0.293         0.071        0.152      4.149    0.000    0.154    0.432
   cond3.ec     1.041         0.071        0.540     14.726    0.000    0.902    1.179
------------------------------------------------------------------------------------
```

Interpretations

Intercept: In effect coding, the intercept is the grand mean of sleep hygiene across all the three treatment groups

cond2: the predicted difference in sleep hygiene score between participants in condition 2 compared to the mean of all three treatment conditions.

cond3: the predicted difference in sleep hygiene score between participants in condition 3 compared to the mean of all three treatment conditions.

Note: We aren't including sex and the interaction term in this first model to make the initial interpretation simpler.

```r
## Run model with interaction
```{r}
m2 <- lm(hygiene ~ sex.f + cond2.ec + cond3.ec + sex.f*cond2.ec + sex.f*cond3.ec, data = slp)
ols_regress(m2)
```
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.771       RMSE                1.007
R-Squared               0.594       Coef. Var          16.801
Adj. R-Squared          0.591       MSE                 1.014
Pred R-Squared          0.585       MAE                 0.799
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error
```

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	881.213	5	176.243	173.768	0.0000
Residual	602.459	594	1.014		
Total	1483.671	599			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	5.433	0.054		101.005	0.000	5.327	5.539
sex.ffemale	1.334	0.084	0.418	15.880	0.000	1.169	1.499
cond2.ec	0.402	0.074	0.209	5.425	0.000	0.257	0.548
cond3.ec	1.280	0.076	0.665	16.831	0.000	1.131	1.429
sex.ffemale:cond2.ec	-0.042	0.120	-0.014	-0.349	0.727	-0.278	0.194
sex.ffemale:cond3.ec	-0.559	0.118	-0.193	-4.719	0.000	-0.791	-0.326

The model suggests that males benefit more from cond3 compared to average, compared to how much females benefit from cond3 compared to average.

Interpretations

Intercept: The intercept is the grand mean of sleep hygiene among males (i.e., when x = 0) across all of the conditions..

sex.ffemale: This variable is involved in an interaction, so it's a simple slope. It is the predicted difference in average sleep hygiene between females and males across all of the conditions

cond2: This variable is involved in an interaction, so it is a simple slope. It is the effect of being in Condition 2 (compared to the average of all three conditions) when female = 0, so among males. It is the predicted difference in sleep hygiene for males in Condition 2 compared to the grand mean of sleep hygiene (the mean of condition 2 is 0.402 higher than the grand mean, 5.433). This is a statistically significant difference (p-value is less than alpha).

cond3: This variable is involved in an interaction, so it is a simple slope. It is the effect of being in Condition 3 (compared to the average of all three conditions) when female = 0, so among males. It is the predicted difference in sleep hygiene for males in Condition 3 compared to the grand mean of sleep hygiene (the mean of condition 3 is 1.280 higher than the grand mean, 5.433). This is a statistically significant difference (p-value is less than alpha).

female:cond2: female:cond2: The predicted differential effect of being in Condition 2 compared to the grand mean for females compared to males. This is not a statistically significant difference (p-value is greater than alpha). The coefficient for cond2 presents the effect (i.e., benefit) of being in Condition 2 (compared to average) for males. To get the effect for females, we take the effect for males (0.402) and add the female:cond2 interaction term (-.042). Therefore, the effect (i.e., benefit) of being in Condition 2 (compared to average for females is 0.36.

female:cond3: The predicted differential effect of being in Condition 3 compared to average for females compared to males. This is a statistically significant difference (p-value is less than alpha). The coefficient for cond3 presents the effect (i.e., benefit) of being in Condition 3 (compared to average) for males. To get the effect for females, we take the effect for males (1.280) and add the female:cond3 interaction term (-.559). Therefore, the effect (i.e., benefit) of being in Condition 3 (compared to average) for females is 0.721.

Part 3: Orthogonal polynomials contrasts

What are orthogonal polynomial contrasts (aka trend contrasts) and why use them?

Allow us to evaluate non-linear relations between a categorical predictor and an outcome

- With continuous predictors we can model these by squaring X to test a quadratic effect, cubing X to test a cubic effect, etc.

- With categorical predictors, we can use specific polynomial contrasts to test different effects. *These contrast levels can be found online, or in the table on slide 25!*

This type of coding system should be used only with an ordinal variable in which the levels are equally spaced

Efffects we can evaluate with more than 3 levels of a categorical predictor:

Linear: If we increase the dose level the Y values will increase, and we can select the best level based on the highest dose. $Y = a + bX$

Quadratic: If we increase the dose level the Y values will be increased until certain dose after that the level of dosage will have a negative effect. $Y = a + b_1X + b_2X^2$

Cubic: The dosage would increase Y values after certain dosage and then decrease and if we increase more the dosage level, the Y values will increase. $Y = a + b_1X + b_2X^2 + b_3X^3$

New dataset description

Researchers were interested in the effect of time spent in practice on the performance of a visual discrimination task. Subjects were randomly assigned to different levels of practice, following which a test of visual discrimination is administered, and the number of correct responses is recorded for each subject. 40 subjects were randomly assigned to practice 0 minutes, 2 minutes, 4 minutes, 6 minutes, 8 minutes, 10 minutes, 12 minutes, or 14 minutes.

There are two variables:

practice = minutes spent practicing, this was assigned by the experimenter

score = the number of correct answers on the test

~~~~
```{r}
cog <- read_csv("cogtest.csv")
```

Parsed with column specification:
cols(
  subject = col_double(),
  practice = col_double(),
  score = col_double()
)
~~~~

Filter practice to only 4 conditions

````
# Filter prectice
```{r}
cog <- filter(cog, practice == 0 | practice == 4 | practice == 8 | practice == 12)
```
````

We selected four evenly spaced conditions to test non-linear effects between practice and score

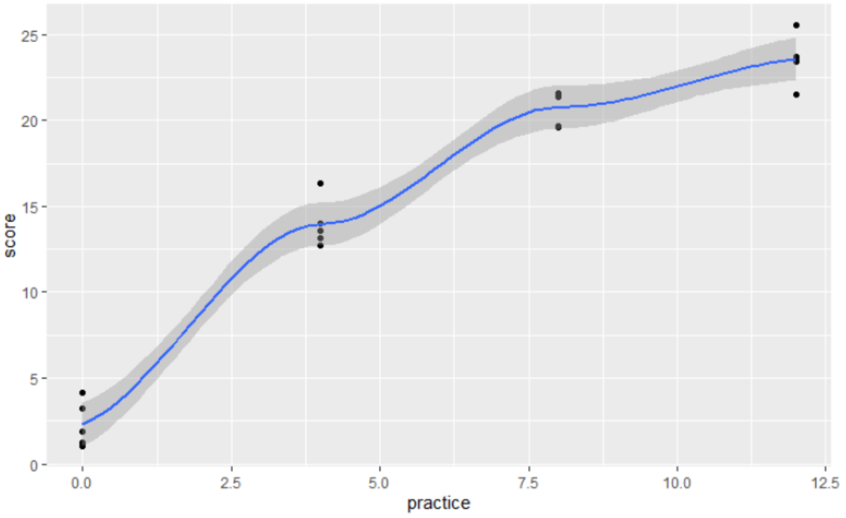
We could include more than four levels of practice if we wanted to, but we use four levels in this example for a simpler interpretation

Visualize the relation between X and Y

```{r}
ggplot(cog, aes(x = practice, y = score)) +
  geom_point() +
  geom_smooth(method = "loess")
```

```
`geom_smooth()` using formula 'y ~ x'

pseudoinverse used at -0.06
neighborhood radius 8.06
reciprocal condition number 4.896e-017
There are other near singularities as well. 64.964
pseudoinverse used at -0.06
neighborhood radius 8.06
reciprocal condition number 4.896e-017
There are other near singularities as well. 64.964
```



The slope appears to become less steep as practice time increases. Since there may be a curve to the regression line, we should test for more than a linear effect. Since we have four levels of our categorical variable, we can test both a quadratic and cubic effect.

The coefficients used for calculating sums of squares are:

Number of treatment	Degree of polynomial	Treatment totals T1	T2	T3	T4	T5	T6	Divisor $k=\sum c_i^2$
2	1	-1	+1					2
3	1	-1	0	+1				2
	2	+1	-2	+1				6
4	1	-3	-1	+1	+3			20
	2	+1	-1	-1	+1			4
	3	-1	+3	-3	+1			20
5	1	-2	-1	0	+1	+2		10
	2	+2	-1	-2	-1	+2		14
	3	-1	+2	0	-2	+1		10
	4	+1	-4	+6	-4	+1		70
6	1	-5	-3	-1	+1	+3	+5	70
	2	+5	-1	-4	-4	-1	+5	84
	3	-5	+7	+4	-4	-7	+5	180
	4	+1	-3	+2	+2	-3	+1	28
	5	-1	+5	-10	+10	-5	+1	252

https://www.ndsu.edu/faculty/horsley/Polycnst.pdf

- We have 4 treatment levels

Specify the orthogonal polynomial contrasts

```
## Orthoganol polynomials
```{r}

cog <- mutate(cog,

              linear = ifelse(practice == 0, -3, practice ),
              linear = ifelse(practice == 4, -1, linear   ),
              linear = ifelse(practice == 8,  1, linear   ),
              linear = ifelse(practice == 12, 3, linear   ),

              quadratic = ifelse(practice == 0,  1, practice ),
              quadratic = ifelse(practice == 4, -1, quadratic),
              quadratic = ifelse(practice == 8, -1, quadratic),
              quadratic = ifelse(practice == 12, 1, quadratic),

              cubic = ifelse(practice == 0, -1, practice),
              cubic = ifelse(practice == 4,  3, cubic   ),
              cubic = ifelse(practice == 8, -3, cubic   ),
              cubic = ifelse(practice == 12, 1, cubic   )
              )

```
```

All of these contrasts come from the table on the previous slide!

We created three new variables:
<u>linear</u> specifies contrasts for testing a linear effect
<u>quadratic</u> specifies contrasts for testing a quadratic effect
<u>cubic</u> specifies contrasts for testing a cubic effect

Step 1: Regress score on linear effect

```{r}
m3 <- lm(score ~ linear, data = cog)
ols_regress(m3)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.953       RMSE                2.635
R-Squared               0.909       Coef. Var          17.400
Adj. R-Squared          0.903       MSE                 6.943
Pred R-Squared          0.886       MAE                 2.207
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
---------------------------------------------------------------------
                Sum of
               Squares        DF    Mean Square       F         Sig.
---------------------------------------------------------------------
Regression    1240.975         1       1240.975    178.727     0.0000
Residual       124.981        18          6.943
Total         1365.957        19
---------------------------------------------------------------------

                                  Parameter Estimates
----------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta      t        Sig      lower    upper
----------------------------------------------------------------------------------------
(Intercept)      15.144         0.589                 25.702    0.000    13.906   16.382
     linear       3.523         0.264        0.953    13.369    0.000     2.969    4.076
----------------------------------------------------------------------------------------
```

The model testing the linear effect between practice and score explained 90.9% of the variance in score, and the linear trend was statistically significant at p<0.001.

This model fits the data pretty well, but since we observed a potential curved relationship when we plotted the data, there could be a better way to examine this relationship.

Evidence of linear effect

Step 2: Regress score on linear & quadratic effect

```{r}
m4 <- lm(score ~ linear + quadratic, data = cog)
ols_regress(m4)
```

```
                        Model Summary
--------------------------------------------------------------
R                       0.990       RMSE               1.274
R-Squared               0.980       Coef. Var          8.411
Adj. R-Squared          0.977       MSE                1.623
Pred R-Squared          0.972       MAE                0.944
--------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                              ANOVA
----------------------------------------------------------------------
                Sum of
               Squares        DF    Mean Square       F          Sig.
----------------------------------------------------------------------
Regression    1338.373         2        669.187    412.425    0.0000
Residual        27.584        17          1.623
Total         1365.957        19
----------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------
      model        Beta    Std. Error    Std. Beta      t        Sig      lower     upper
-----------------------------------------------------------------------------------------
(Intercept)      15.144         0.285                53.168    0.000     14.543    15.745
     linear       3.523         0.127        0.953   27.655    0.000      3.254     3.791
  quadratic      -2.207         0.285       -0.267   -7.748    0.000     -2.808    -1.606
-----------------------------------------------------------------------------------------
```

Evidence of quadratic effect

The model testing the linear and quadratic effects between practice and score explained 98.0% of the variance in score, which is 7.1% higher than the model that only tested the linear relation.

The quadratic term is statistically significant, indicating that there is a substantial curve to the relation between practice and score (i.e., it's not linear). We need to maintain the quadratic term in the model.

Step 3: Regress score on linear, quadratic & cubic effect

```{r}
m5 <- lm(score ~ linear + quadratic + cubic, data = cog)
ols_regress(m5)
```

```
                        Model Summary
----------------------------------------------------------------
R                       0.990       RMSE                  1.309
R-Squared               0.980       Coef. Var             8.641
Adj. R-Squared          0.976       MSE                   1.712
Pred R-Squared          0.969       MAE                   0.954
----------------------------------------------------------------
 RMSE: Root Mean Square Error
 MSE: Mean Square Error
 MAE: Mean Absolute Error

                               ANOVA
-----------------------------------------------------------------------
                Sum of
               Squares        DF    Mean Square       F         Sig.
-----------------------------------------------------------------------
Regression    1338.559         3        446.186    260.572    0.0000
Residual        27.397        16          1.712
Total         1365.957        19
-----------------------------------------------------------------------

                                  Parameter Estimates
-----------------------------------------------------------------------------------------
      model      Beta    Std. Error    Std. Beta      t        Sig      lower      upper
-----------------------------------------------------------------------------------------
(Intercept)    15.144         0.293                 51.755    0.000     14.524     15.764
     linear     3.523         0.131        0.953    26.921    0.000      3.245      3.800
  quadratic    -2.207         0.293       -0.267    -7.542    0.000     -2.827     -1.586
      cubic     0.043         0.131        0.012     0.330    0.746     -0.234      0.321
-----------------------------------------------------------------------------------------
```

No evidence of cubic effect

We tested the cubic term to determine if there is a second bend to the relationship between practice and score. Since we have at least four levels of our categorical predictor, we can test the cubic effect.

This model explains the same amount of variance in score as the previous model that only included the linear and quadratic effects (i.e., adding the cubic effect does not increase the explanatory power of the model).

The cubic term is not significant, indicating that there is not a second bend to the relationship. **Therefore, the quadratic model is the best fit for these data.**

Additional resources on categorical variable coding systems:

- https://stats.idre.ucla.edu/spss/faq/coding-systems-for-categorical-variables-in-regression-analysis/
- http://www.jds-online.com/files/JDS-563.pdf
- https://www.ndsu.edu/faculty/horsley/Polycnst.pdf
- https://www.researchgate.net/post/Linear_quadratic_and_cubic_polynomial_contrasts