Effects Coding a Categorical Predictor

2019-10-19

Preparation

In this set of notes, we will examine the question of whether there is an effect of activity-based learning on course grade points. To do so, we will use the *activity-based-courses.csv* data (see the data codebook). To begin, we will load several libraries and import the data into an object called `activity`.[1]

```
# Load packages
library(broom)
library(corrr)
library(tidyverse)
library(educate)
library(ggridges)
library(ungeviz)

# Read in data
activity = read_csv(file = "~/Documents/github/epsy-8251/data/activity-based-courses.csv")
head(activity)
```

```
# A tibble: 6 x 3
  grade_points condition    act
         <dbl> <chr>      <dbl>
1            0 Control       15
2            0 Control       18
3            0 Control       14
4            1 Control       18
5            1 Control       17
6            1 Control       15
```

Exploration

Initially, we plot the data and compute summary statistics (means, standard deviations, and sample sizes) to compare the grade points for the two conditions. (Note: The syntax to examine the summary statistics is not shown.) Rather than looking at a scatterplot, we compare the density plots of the two distributions using the `geom_density_ridges()` function from the **ggridges** package. (In this function, the x= aesthetic mapping must be to a continuous variable.)

```
ggplot(data = activity, aes(x = grade_points, y = condition)) +
  geom_density_ridges() +
  theme_bw() +
  ylab("Experimental Condition") +
  xlab("Grade points")
```

---

[1]The `tidyverse` library is a meta-package that loads dplyr, forcats, ggplot2, purrr, readr, stringr, tibble, and tidyr.
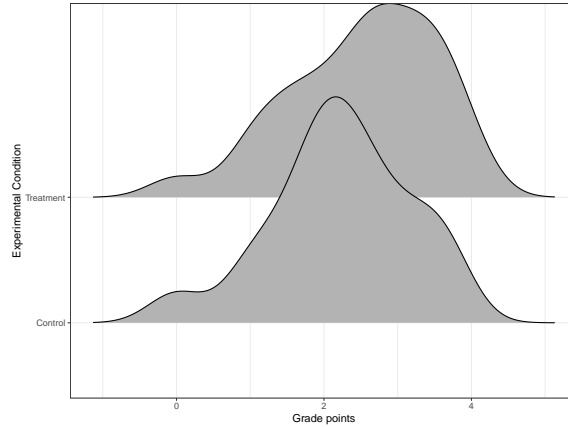
Figure 1. Density plots of the grade points conditioned on experimental treatment.

Table 1
Mean (M), Standard Deviation (SD), and Sample Size (N) of the Grade Points for Activity-Based (Treatment) and Lecture-Based (Control) Sections of a Course

Condition	M	SD	N
Control	2.22	0.94	55
Treatment	2.52	0.99	55

The two distributions of grade points seem quite similar. Students in the activity-based section seem to have a slightly higher average grade point than students in the lecture-based section (0.3 grade points higher).

Treatment Differences in ACT Scores

In theory, the randomization to conditions should "equalize" the groups on all possible covariates; it is why random assignment to groups is the Gold Standard in creating equivalent groups. With small samples, it is often worthwhile checking to see that the random assignment worked to "equalize" the covariates. Here we will check that the distribution of ACT scores is equivalent for the students in the two conditions. (The syntax to examine the summary statistics is not shown.)

```
ggplot(data = activity, aes(x = act, y = condition)) +
  geom_density_ridges() +
  theme_bw() +
  ylab("Experimental Condition") +
  xlab("ACT score")
```
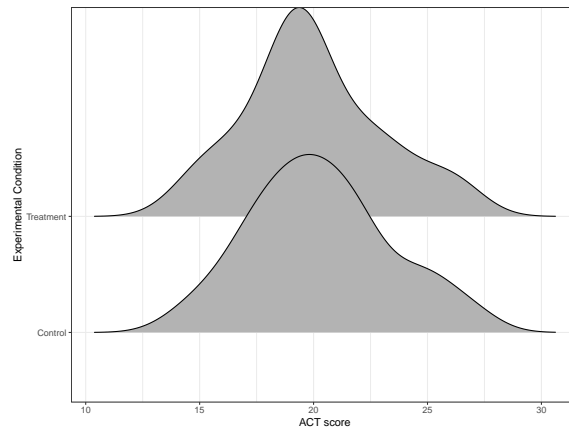
Figure 2. Density plots of the ACT scores conditioned on experimental treatment.

Table 2
Mean (M), Standard Deviation (SD), and Sample Size (N) of the ACT sSores for Activity-Based (Treatment) and Lecture-Based (Control) Sections of a Course

Condition	M	SD	N
Control	20.29	3.07	55
Treatment	20.00	3.03	55

We can also carry out a hypothesis test to see if there are differences in the mean ACT scores between the two conditions. To do this, we mutate() a new column (called treatment) into our dataset that takes the value of 0 if condition is Control and 1 if condition is Treatment. Here we use the if_else() function to carry out this logic.

```
# Create new column that dummy-codes condition
activity = activity %>%
  mutate(
    treatment = if_else(condition == "Control", 0, 1)
  )

head(activity)
```

```
# A tibble: 6 x 4
  grade_points condition   act treatment
         <dbl> <chr>     <dbl>     <dbl>
1            0 Control      15         0
2            0 Control      18         0
3            0 Control      14         0
4            1 Control      18         0
5            1 Control      17         0
6            1 Control      15         0
```

We can then use this new column as the predictor in our regression model.

```
lm.act = lm(act ~ 1 + treatment, data = activity)
tidy(lm.act)
```

3

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   20.3      0.411     49.4   6.12e-76
2 treatment     -0.291    0.581     -0.501 6.18e- 1
```

The empirical evidence suggests that although there is a sample difference in average ACT scores (the activity-based section has an average ACT score that is 0.29-points lower than the students in the lecture-based section), the sampling uncertainty associated with this difference ($SE = 0.58$) suggests that the observed difference may be entirely due to chance ($p = .617$). The empirical evidence here is consistent with the null hypothesis of no difference in the average ACT score between the two experimental conditions.

All of this analysis suggests that the students in the two conditions have "statistically equivalent" ACT scores. This implies that the randomization was effective in creating "equal" groups (at least on ACT scores) and that we can rule out differences in ACT scores as a reason the two groups may differ on grade points.

Treatment Differences in Grade Points

To examine whether the observed difference in average grade point between the lecture and activity-based courses is more than we would expect because of chance, we can fit a regression model using condition to predict variation in grade point. We can again use the dummy-coded `treatment` column as a predictor in a regression model that predicts variation in students' grade-points.

```
# Fit model
lm.2 = lm(grade_points ~ 1 + treatment, data = activity)

# Model-level output
glance(lm.2)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
      <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>  <dbl> <dbl> <dbl>
1    0.0249        0.0159 0.964      2.76  0.0995     2  -151.  308.  316.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```
# Coefficient-level output
tidy(lm.2)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    2.22     0.130      17.1  1.96e-32
2 treatment      0.305    0.184       1.66 9.95e- 2
```

The output indicates that differences in experimental condition explain roughly 2.5% of the variation in students' grade-points. The estimated slope coefficient is 0.305 which indicates that the average grade-point for students in the activity-based section is 0.305-points higher than the average grade-point for students in the activity-based section. The p-value associated with the effect of experimental condition is .0995. This evidence is consistent with the null hypothesis that there are no differences in average grade-point between students taking the lecture-based and the activity-based sections of the course.

Controlling for ACT Differences: Adding Statistical Power

Although we believe that the distributions of ACT scores are statistically equivalent, it doesn't prevent us from including ACT as a covariate in our model. Many times, if you have an important covariate, it should be included in the model. This helps in two ways. First, in the case of any sample differences in the covariate it equates the two groups by controlling for those differences in the model. Secondly, controlling for major covariates often increases statistical power for the effect of treatment (produces less uncertainty). We illustrate this by fitting a model that includes not only the effect of treatment, but also the effect of ACT score to predict variation in students' grade-points.

```
# Fit model
lm.3 = lm(grade_points ~ 1 + treatment + act, data = activity)

# Model-level output
glance(lm.3)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic  p.value    df logLik  AIC   BIC
      <dbl>         <dbl> <dbl>     <dbl>    <dbl> <int>  <dbl> <dbl> <dbl>
1     0.336         0.323 0.799      27.0 3.14e-10     3  -130.  268.  279.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```
# Coefficient-level output
tidy(lm.3)
```

```
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    -1.40     0.523     -2.68 8.42e- 3
2 treatment       0.357     0.153      2.34 2.10e- 2
3 act             0.179    0.0252      7.08 1.62e-10
```

The output indicates that differences in experimental condition and ACT scores explain roughly 3.36% of the variation in students' grade-points. The estimated partial slope coefficient for treatment is 0.357 which indicates that the average grade-point for students in the activity-based section is 0.35-points higher than the average grade-point for students in the activity-based section, after controllng for differences in ACT score. The p-value associated with the effect of experimental condition is now .0.021, which reflects the decrease in uncertainty after controlling for differences in ACT ($SE = 0.153$, a reduction from the previous SE of 0.581). The smaller SE and p-value reflect the increased statistical power.

The empirical evidence now is much less consistent with the null hypothesis that there are no differences in average grade-point between students taking the lecture-based and the activity-based sections of the course.

Effects-Coding Experimental Condition

Recall that how we numerically code the categories is irrelevant to the results we get. We will obtain the same results no matter how we code the predictor. What changes is the interpretation of the coefficients. One useful coding scheme, especially for experimental data, is to use *effects-coding*. Effects-coding for a binary categorical variable uses the values $+1$ and -1 rather than 0 and 1 to encode the categories. Here I will encode the control group using -1 and the treatment group using $+1$.

```
# Create new column that effect-codes condition
activity = activity %>%
  mutate(
    effect = if_else(condition == "Control", -1, 1)
  )

head(activity)
```

```
# A tibble: 6 x 5
  grade_points condition    act treatment effect
         <dbl> <chr>      <dbl>     <dbl>  <dbl>
1            0 Control       15         0     -1
2            0 Control       18         0     -1
3            0 Control       14         0     -1
4            1 Control       18         0     -1
5            1 Control       17         0     -1
6            1 Control       15         0     -1
```

We then fit the regression model using this new `effect` predictor. Initially we will fit the simple regression model using `effect` to predict variation in students' grade-points (without including ACT scores).

```
# Fit model using effects-coded predictor
lm.4 = lm(grade_points ~ 1 + effect, data = activity)

# Model-level output
glance(lm.4)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
      <dbl>         <dbl> <dbl>     <dbl>   <dbl> <int>  <dbl> <dbl> <dbl>
1    0.0249        0.0159 0.964      2.76  0.0995     2  -151.  308.  316.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```
# Coefficient-level output
tidy(lm.4)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)     2.37    0.0919      25.8 5.60e-48
2 effect          0.153   0.0919       1.66 9.95e- 2
```

The model-level information is identical to when we used the dummy-coded predictor, namely that that differences in experimental condition explain roughly 2.5% of the variation in students' grade-points. The F-value and p-value for the model (not shown) would also be identical to those from the dummy-coded model.

Because we used a different coding scheme, the estimates of the coefficients and standard errors of those estimates are different. However the inferential evidence (t- and p-value) for the effect of experimental condition is the same. We would come to the same conclusion about whether there is a difference in the average grade-points between the two course styles.

The fitted equation is now:

$$\hat{\text{Grade-Point}}_i = 2.37 + 0.15(\text{Effect}_i)$$

Interpeting these coefficients:

- The model predicted average grade-point for students in the experimental condition coded as 0 is 2.37.
- Each one-unit difference in `Effect` is associated with a grade-point difference of 0.15, on average.

While we did not explicitly code either one of the experimental conditions as 0, this nonetheless can be interpreted. It turns out that 2.37 is the marginal mean grade-point.

```
activity %>%
  summarize(
    M = mean(grade_points)
  )
```

```
# A tibble: 1 x 1
      M
  <dbl>
1  2.37
```

The slope estimate now represents the difference in average grade-points between the marginal mean (an x-value of 0) and the treatment group (an x-value of 1). This is often referred to as the *effect of treatment* in an experimental study; how much more does the treatment give you above and beyond the mean.

In our study, the effect of treatment is 0.15. That is, the students who are in the activity-based course have an average grade-point that is 0.15 higher than the mean grade-point for all students. Similarly, students who are in the lecture-based course have an average grade-point that is 0.15 lower than the mean grade-point for all students.

We can verify this by computing the conditional mean grade-point for both groups. To do so, we substitute the appropriate predictor values into the fitted equation:

$$\textbf{Control}: \hat{\text{Grade-Point}}_i = 2.37 + 0.15(-1)$$
$$= 2.22$$

$$\textbf{Treatment}: \hat{\text{Grade-Point}}_i = 2.37 + 0.15(1)$$
$$= 2.52$$

These are the same values we obtained earlier when we computed the conditional mean grade-point for each condition.

Statistical Model for the ANOVA

In an experimental design class, you would be introduced to the method of ANOVA for testing mean differences. The statistical model that is presented for this methodology is:

$$Y_i = \mu + \alpha_j + \epsilon_{ij}$$

where

- Y_i is the outcome for person i;
- μ is the marginal mean of Y;
- α_j is the effect on Y for condition j; and
- ϵ_{ij} is the error for person i in condition j

This is model is the same as the regression model, but with different notation,

$$Y_i = \beta_0 + \beta_1(X_i) + \epsilon_{ij}$$

When we use effects-coding to encode the predictor and substitute them in to the regression model, this can be simplified to:

$$\textbf{Control}: \; Y_i = \beta_0 + \beta_1(-1) + \epsilon_{ij}$$
$$= \beta_0 - \beta_1 + \epsilon_{ij}$$

$$\textbf{Treatment}: \; Y_i = \beta_0 + \beta_1(1) + \epsilon_{ij}$$
$$= \beta_0 + \beta_1 + \epsilon_{ij}$$

In this representation, the intercept parameter is equivalent to the mu parameter from the ANOVA model and the slope parameter is equivalent to the alpha parameter from the ANOVA model. The two models are identical when we use effects-coding in the regression model!

Inference for the ANOVA Model

Under the ANOVA model, the primary null hupothesis of interest is:

$$H_0 : \alpha_j = 0 \quad \text{for all } j$$

In words, there is no effect on Y (beyond the marginal mean) for any condition. Looking at the coefficient-level output we find that for the slope, $t(107) = 1.66$, $p = .0995$. This implies that the empirical evidence is consistent with the null hypothesis of no effect on grade-points for any of the conditions. In other words, the average for students in the activity-based course and that for the students in the lecture-based course does not seem to be statistically different than the marginal average.

Controlling for Differences in ACT Scores

Again, because ACT score is an important covariate when it comes to predicting students' grade-points, we will control for ACT score in our regression model. In the experimental design literature this is called an Analysis of Covariance (ANCOVA). The statistical model for ANCOVA is:

$$Y_i = \mu + \alpha_j + \beta_1(X_{ij}) + \epsilon_{ij}$$

where

- Y_i is the outcome for person i;
- μ is the marginal mean of Y;
- α_j is the effect on Y for condition j;
- X_{ij} is the value of covariate X for person i in condition j;
- β_1 is the effect of covariate X on Y; and

- ϵ_{ij} is the error for person i in condition j

This is again absolutely equivalent to the regression model:

$$Y_i = \beta_0 + \beta_1(X_{1i}) + \beta_2(X_{2i}) + \epsilon_{ij}$$

where

```
# Fit model using effects-coded predictor
lm.5 = lm(grade_points ~ 1 + effect + act, data = activity)

# Model-level output
glance(lm.5)
```

```
# A tibble: 1 x 11
  r.squared adj.r.squared sigma statistic  p.value    df logLik  AIC   BIC
      <dbl>         <dbl> <dbl>     <dbl>    <dbl> <int>  <dbl> <dbl> <dbl>
1     0.336         0.323 0.799      27.0 3.14e-10     3  -130.  268.  279.
# ... with 2 more variables: deviance <dbl>, df.residual <int>
```

```
# Coefficient-level output
tidy(lm.5)
```

```
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    -1.23     0.514     -2.39 1.88e- 2
2 effect          0.179    0.0763     2.34 2.10e- 2
3 act             0.179    0.0252     7.08 1.62e-10
```

The output of the indicates that differences in experimental condition and ACT scores explain roughly 3.36% of the variation in students' grade-points, $F(2, 107) = 27.04$, $p < .001$. The estimated partial slope coefficient for treatment is 0.178 which indicates that the average grade-point for students in the activity-based section is 0.178-points higher than the average grade-point for all students, after controllng for differences in ACT score. The p-value associated with the effect of experimental condition is 0.021, which reflects the decrease in uncertainty after controlling for differences in ACT. The smaller p-value reflects the increased statistical power. The empirical evidence is not very consistent with the null hypothesis that there is no effect of treatment on average grade-points, after controlling for ACT scores.