

Mathematics 231

Lecture 29

Liam O'Brien

Announcements

- Reading

- Today M&M 10.2 579-583
- Next class M&M 10.1 559-576

ANOVA vs Linear Regression

- The ANOVA is analogous to running a linear regression model with indicator functions for the group categories.
- An **indicator function** is a function that equals 1 if a certain condition is true, and 0 otherwise.
- For example, if measurement i came from group 1, then we could denote this by $I_i(\text{group } 1)=1$.

ANOVA vs Linear Regression

- If we have K groups, we can describe those with $K-1$ indicator functions.

Let us denote the rating measurement for subject i by y_i . Then a regression model could be specified relating y_i to the ad from which measurement i came.

$$\hat{y}_i = \beta_0 + \beta_1 I_i(\text{Affirm}) + \beta_2 I_i(\text{Control})$$

For Affirming ad: $\hat{y}_i = \beta_0 + \beta_1$

For Control ad: $\hat{y}_i = \beta_0 + \beta_2$

For Undermining ad: $\hat{y}_i = \beta_0$

Thus, each ad has its own expression for the average rating.

ANOVA vs. Linear Regression

- This model is an example of a multiple regression model.
- **Multiple regression** is a regression model that has a single outcome that is predicted by more than one predictor.
- In this case, we have two predictors – both indicators of specific groups.
- We'll cover how to do this in Stata later, but it's the same procedure you used before.

Example: Ad Ratings

```
. regress rating adA adC
```

Source	SS	df	MS	Number of obs = 127		
Model	18.828255	2	9.4141275	F(2, 124)	=	15.28
Residual	76.3843434	124	.61600277	Prob > F	=	0.0000
Total	95.2125984	126	.755655543	R-squared	=	0.1977
				Adj R-squared	=	0.1848
				Root MSE	=	.78486

rating	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
adA	.5464646	.1682594	3.25	0.001	.2134321	.8794972
adC	.9075758	.1682594	5.39	0.000	.5745432	1.240608
_cons	4.509091	.1058303	42.61	0.000	4.299623	4.718559

Example: Ad Ratings

```
. oneway rating group
```

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	18.828255	2	9.4141275	15.28	0.0000
Within groups	76.3843434	124	.61600277		
Total	95.2125984	126	.755655543		

```
Bartlett's test for equal variances:  chi2(2) = 2.6669  Prob>chi2 = 0.264
```

ANOVA vs Linear Regression

- Note that the ANOVA table and the regression output are identical.
- ANOVA is simply a linear regression with categorical predictors.
- What we called “within groups MS” is the MSE (mean square error, or mean square residual) in the regression model.
- What we called “between groups MS” in the mean square model in the regression model.