STA305/1004 - Class 14

March 6, 2017

Today's class

- ANOVA demonstrtation
- Estimating Treatment Effects in ANOVA using Regression
- Coding Qualitative Predictors in Regression Models

ANOVA Demonstration

Figure 1:

►

- ► Count the total number of each colour (e.g., yellow, purple, pink, green).
- ► Eat the Smarties.

ANOVA Data Setup

- How should the data be setup?

Box	Colour	Count
1	Green	
1	Pink	
1	Purple	
1	Yellow	
2	Green	
2	Pink	
2	Purple	
2	Yellow	
3	Green	
3	Pink	
3	Purple	
3	Yellow	
4	Green	
4	Pink	
4	Purple	
4	Yellow	
5	Green	
5	Pink	
5	Purple	
5	Yellow	

-

Figure 2:

Smarties Data from 3 boxes

```r
count <- c(4,3,4,3,1,4,2,5,1,1,2,4)
colour <- as.factor(c(rep("Yellow",3),rep("Purple",3),
                      rep("Green",3),rep("Pink",3)))
#Get means for each flavour
sapply(split(count,colour),mean)
```

```
   Green      Pink    Purple    Yellow
2.666667  2.333333  2.666667  3.666667
```

Estimating Treatment Effects in ANOVA using Regression

- y_{ij} is the jth observation under the ith treatment.
- The model for smarties $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, $\epsilon_{ij} \sim N(0, \sigma^2)$ can be written in terms of the dummy variables X_1, X_2, X_3 as:

$$y_{ij} = \mu + \tau_1 X_{1j} + \tau_2 X_{2j} + \tau_3 X_{3j} + \epsilon_{ij}.$$

- What is $y_{ij}, \mu, \tau_i, X_{ij}, \epsilon_{ij}$?

The ANOVA Table

```
#ANOVA table
anova(lm(count~colour))


Analysis of Variance Table

Response: count
          Df Sum Sq Mean Sq F value Pr(>F)
colour     3  3.000  1.0000  0.4286 0.7381
Residuals  8 18.667  2.3333
```

Dummy coding

- ▶ Dummy coding compares each level to the reference level. The intercept is the mean of the reference group.
- ▶ Dummy coding is the default in R and the most common coding scheme. It compares each level of the categorical variable to a fixed reference level.

```
contrasts(colour) <- contr.treatment(4)  #Treatment contrast
contrasts(colour) # print dummy coding
```

```
       2 3 4
Green  0 0 0
Pink   1 0 0
Purple 0 1 0
Yellow 0 0 1
```

```
lm(count~colour)
```

```
Call:
lm(formula = count ~ colour)

Coefficients:
(Intercept)      colour2      colour3      colour4
  2.667e+00   -3.333e-01    4.710e-16    1.000e+00
```

Deviation coding

▶ This coding system compares the mean of the dependent variable for a given level to the overall mean of the dependent variable.

```
contrasts(colour) <- contr.sum(4)  # Deviation contrast
contrasts(colour) # print deviation coding
```

```
       [,1] [,2] [,3]
Green     1    0    0
Pink      0    1    0
Purple    0    0    1
Yellow   -1   -1   -1
```

```
lm(count~colour)
```

```
Call:
lm(formula = count ~ colour)

Coefficients:
(Intercept)      colour1      colour2      colour3
     2.8333      -0.1667      -0.5000      -0.1667
```