Stat 401 B – Lecture 33

Categorical Variables

- Response: Highway MPG
- Explanatory: Type of drive
  - All Wheel
  - Rear Wheel
  - Front Wheel

Indicator Variables

- We have used indicator variables so that we can trick JMP into analyzing the data using multiple regression.

Categorical Variables

- There is a more straight forward analysis that can be done with categorical explanatory variables.

Stat 401 B – Lecture 33

Categorical Variables

- The analysis is an extension of the two independent sample analysis we did at the beginning of the semester.
- Body mass index for men and women (Lectures 4 and 5).

4

Analysis of Variance

- Response: numerical, Y
- Explanatory: categorical, X
- Total Sum of Squares

$$SS_{Total} = \sum (y - \bar{y})^2$$

5

Sum of Squares Total

$\bar{y} = 27.7$

$SS_{Total} = \sum (y - \bar{y})^2 = 3669.0$

$df = 99$

6

Stat 401 B – Lecture 33

Analysis of Variance

- Partition the Total Sum of Squares into two parts.
  - Due to differences among the sample means for the categories.
  - Due to variation within categories, i.e error variation.

7

Sum of Squares Factor

$$SS_{Factor} = \sum n_i(\bar{y}_i - \bar{y})^2$$

n_i = number of observations in category i

$\bar{y}_i$ = sample mean for category i

8

Category Sample Means

	Mean	Sample Size
All Wheel	22.608	23
Rear Wheel	26.529	17
Front Wheel	29.983	60

9