Agenda

1. Interpreting simple linear regression coefficients
2. Residuals
3. Conditions for regression
4. Outliers– leverage and influence

Interpreting coefficients Let's talk about the homework questions due last night.

```
require(mosaic)
require(Stat2Data)
data(Cereal)
m1 <- lm(Calories~Sugar, data=Cereal)
coef(m1)

## (Intercept)       Sugar
##   87.427690    2.480813
```

How can we interpret the intercept in this model? The slope?

Residuals

- Residuals: $y - \hat{y}$

$$SSE = \sum_{i=1}^{n} (y - \hat{y})^2, \qquad \hat{\sigma}_\epsilon = \sqrt{\frac{SSE}{n-2}}$$

- Least Squares: technique for minimizing SSE
- Finds *unique* straight line between scatterplot of points

Conditions for regression

- **L**inearity:

- **I**ndependence:

- **N**ormality of Residuals:

- **E**qual Variance of Residuals:

The LINE conditions are the same as the conditions from the book, just differently stated. Lets look at some examples of the conditions failing to be met in the residuals lab.

Lab Some code from the residuals lab

```
classdata %>% slice(12)
classdata %>% slice(10)

mod = lm(num_languages~num_countries, data=classdata)
summary(mod)
xyplot(num_languages~num_countries, data=classdata, type=c("p", "r"))

new_ds = data.frame(num_countries = c(3, 11))
predict(mod, newdata = new_ds)
1 - 2
4 - 2

residuals(mod)
sum(residuals(mod)^2)
mean(residuals(mod))

plot(mod, which=c(1,2))
histogram(~residuals, data=mod, fit="normal")
```

Transformations What if the assumptions for regression are not met? Apply transformations! This process is as much art as science... The most common transformation to apply is a log transformation.

Leverage and influence Not all outliers have high leverage!