

Sample Module

a **statsTeachR** resource

These slides were adapted for [statsTeachR](#) by Nicholas Reich from slides written by Jeff Goldsmith and are released under a [Creative Commons Attribution-ShareAlike 3.0 Unported License](#).

Today's lecture

- ▶ Multiple Linear Regression
 - ▶ Assumptions
 - ▶ Interpretation

Motivation

Most applications involve more than one covariate – if more than one thing can influence an outcome, you need multiple linear regression.

- ▶ Improved description of $y|x$
- ▶ More accurate estimates and predictions
- ▶ Allow testing of multiple effects
- ▶ Includes multiple predictor types

Multiple linear regression model

- ▶ Observe data $(y_i, x_{i1}, \dots, x_{ip})$ for subjects $1, \dots, n$. Want to estimate $\beta_0, \beta_1, \dots, \beta_p$ in the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i; \epsilon_i \stackrel{iid}{\sim} (0, \sigma^2)$$

- ▶ Assumptions (residuals have mean zero, constant variance, are independent) are as in SLR
- ▶ Impose linearity which (as in the SLR) is a big assumption
- ▶ Our primary interest will be $E(y|x)$
- ▶ Eventually estimate model parameters using least squares

Multiple linear regression model

```
require(MASS)
data(crabs)
mlr <- lm(FL ~ factor(sex) + RW, data=crabs)
summary(mlr)$coef
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	-2.980	0.38287	-7.784	3.894e-13
## factor(sex)M	2.346	0.14114	16.620	2.400e-39
## RW	1.365	0.02749	49.658	2.384e-113

Multiple linear regression model

```
qplot(RW, FL, facets=~sp, color=sex, data=crabs)
```

