Simple Linear Regression
Week 3 – Thursday
Applied Regression Analysis (STAT 757)

Paul J. Hurtado

Tuesday, 4 Feb, 2016
Announcements

David Quammen

Discover Science Lecture Series
7 p.m., TODAY (Feb. 4)
Redfield Auditorium (DMSC 110)
Random Variables & Probability Distributions

What does it mean for $X$ to be a random variable?

1. $X$ is the outcome of an experiment; a place-holder for a random number.
Random Variables & Probability Distributions

What does it mean for $X$ to be a random variable?

1. $X$ is the outcome of an experiment; a place-holder for a random number.

2. $X$ has a distribution associated with it.
Random Variables & Probability Distributions

What does it mean for $X$ to have a distribution?

1. Distributions describe the propensity for some outcomes to occur more often, or with greater likelihood, than others.
Random Variables & Probability Distributions

What does it mean for $X$ to have a distribution?

1. Distributions describe the propensity for some outcomes to occur more often, or with greater likelihood, than others.

2. When we refer to the distribution, we are referring to a few different, but equivalent, functions!
Random Variables & Probability Distributions

Probabilities of events are calculated from either the PDF (continuous) or PMF (discrete):

Normal PDF

Binomial PMF
Random Variables & Probability Distributions

If you know the CDF, you know the PDF/PMF, and *vice versa*.
Random Variables & Probability Distributions

Histograms of large random samples look like the PDF/PMF!

Normal Data
rnorm(4000)

Binomial Data
rbinom(4000, 10, 0.2)
Estimates vs Estimators?

1. **Estimators** are *functions of random variables*, and thus are themselves random variables. Rules for calculating...
Estimates vs Estimators?

1. **Estimators** are *functions of random variables*, and thus are themselves random variables. Rules for calculating...

2. **Estimates**, which are *single numbers*.

Unfortunately, you must rely on context for which \( \hat{\beta} \) refers to!
Estimators as random variables

3000 Replicated SLR Estimates (N=10)

**Histogram of Intercept**

- Frequency axis ranges from 0 to 150.
- X-axis represents the intercept values ranging from 0 to 8.

**Histogram of Slope**

- Frequency axis ranges from 0 to 300.
- X-axis represents the slope values ranging from 0.0 to 1.5.
Estimators as random variables

3000 Replicated SLR Estimates (N=10)

Histogram of SD