Univariate Normal Distribution; GLM with the Univariate Normal; Least Squares Estimation

PSYC 943: Fundamentals of Multivariate Modeling

Lecture 3

Today's Class

- The building blocks: The basics of mathematical statistics:
  - Random variables: definitions and types
  - Univariate distributions
    - General terminology
    - Univariate normal (aka, Gaussian)
    - Other (continuous) univariate distributions
  - Expected values: means, variances, and the algebra of expectations
  - Linear combinations of random variables

- The finished product: How the GLM fits within statistics
  - The GLM with the normal distribution
  - The statistical assumptions of the GLM

- Estimation of GLMs with Least Squares

RANDOM VARIABLES AND STATISTICAL DISTRIBUTIONS

Random Variables

Random: situations in which the certainty of the outcome is unknown and is at least in part due to chance

+

Variable: a value that may change given the scope of a given problem or set of operations

=

Random Variable: a variable whose outcome depends on chance

(possible values might represent the possible outcomes of a yet-to-be-performed experiment)

Today we will denote a random variable with a lower-cased: x

Key Features of Random Variables

- Random variables each are described by a **probability density/mass function (PDF)** $f(x)$ that indicates relative frequency of occurrence
  - A PDF is a mathematical function that gives a rough picture of the distribution from which a random variable is drawn

- The type of random variable dictates the name and nature of these functions:
  - Continuous random variables:
    - $f(x)$ is called a probability density function
    - Area under curve must equal 1 (found by calculus – integration)
    - Height of curve (the function value $f(x)$):
      - Can be any positive number
      - Reflects relative likelihood of an observation occurring
  - Discrete random variables:
    - $f(x)$ is called a probability mass function
    - Sum across all values must equal 1
    - The function value $f(x)$ is a probability (so must range from 0 to 1)

Uses of Distributions in Data Analysis

- Statistical models make distributional assumptions on various parameters and/or parts of data

- These assumptions govern:
  - How models are estimated
  - How inferences are made
  - How missing data may be imputed

- If data do not follow an assumed distribution, inferences may be inaccurate
  - Sometimes a very big problem, other times not so much

- Therefore, it can be helpful to check distributional assumptions prior to (or while) running statistical analyses

CONTINUOUS UNIVARIATE DISTRIBUTIONS

Continuous Univariate Distributions

- To demonstrate how continuous distributions work and look, we will discuss two:
  - Uniform distribution
  - Normal distribution
- Each are described a set of **parameters**, which we will later see are what give us our inferences when we analyze data
- What we then do is put constraints on those parameters based on hypothesized effects in data

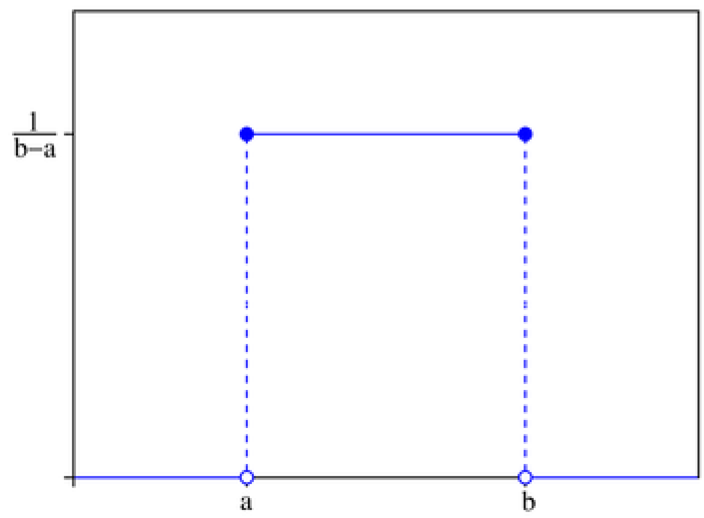
Uniform Distribution

- The uniform distribution is shown to help set up how continuous distributions work

- For a continuous random variable x that ranges from (a, b), the uniform probability density function is:

$$f(x) = \frac{1}{b - a}$$

- The uniform distribution has two parameters:
  - a – the lower limit
  - b – the upper limit



- $x \sim U(a, b)$

More on the Uniform Distribution

- To demonstrate how PDFs work, we will try a few values:

x	a	b	$f(x)$
.5	0	1	$\frac{1}{1-0}=1$
.75	0	1	$\frac{1}{1-0}=1$
15	0	20	$\frac{1}{20-0}=.05$
15	10	20	$\frac{1}{20-10}=.1$

- The uniform PDF has the feature that all values of x are **equally likely** across the sample space of the distribution
  - Therefore, you do not see x in the PDF $f(x)$

- The mean of the uniform distribution is $\frac{1}{2}(a+b)$
- The variance of the uniform distribution is $\frac{1}{12}(b-a)^2$

Univariate Normal Distribution

- For a continuous random variable x (ranging from $-\infty$ to ∞) the univariate normal distribution function is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)$$

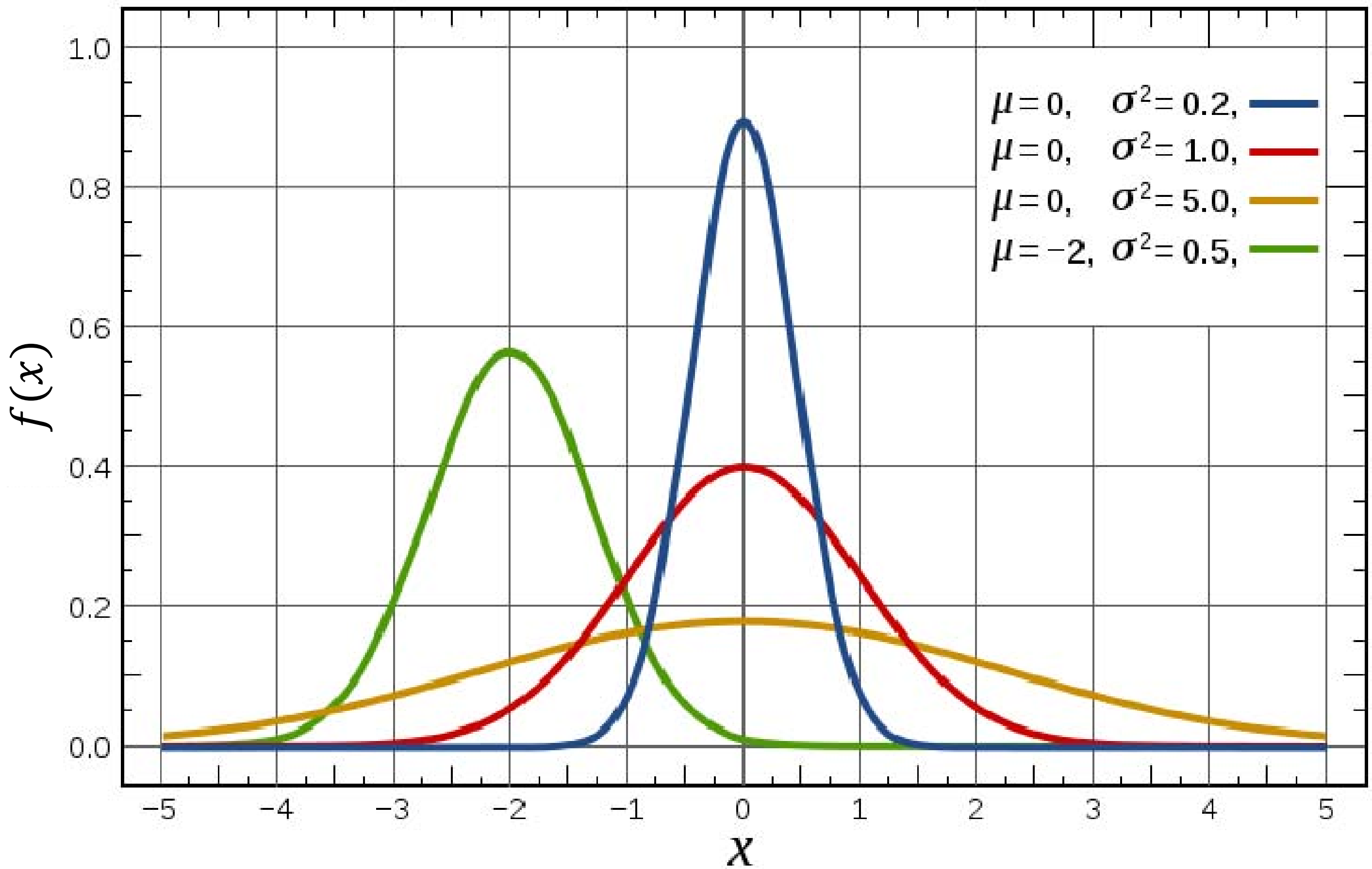
- The shape of the distribution is governed by two parameters:
  - The mean μ_x
  - The variance σ_x^2
  - These parameters are called **sufficient statistics** (they contain all the information about the distribution)

- The skewness (lean) and kurtosis (peakedness) are fixed

- Standard notation for normal distributions is $x \sim N(\mu_x, \sigma_x^2)$
  - Read as: "x follows a normal distribution with a mean μ_x and a variance σ_x^2"

- Linear combinations of random variables following normal distributions result in a random variable that is normally distributed
  - You'll see this later with respect to the GLM…

Univariate Normal Distribution



$f(x)$ gives the height of the curve (relative frequency) for any value of x, μ_x, and σ_x^2

More of the Univariate Normal Distribution

- To demonstrate how the normal distribution works, we will try a few values:

x	μ_x	σ_x^2	$f(x)$
.5	0	1	0.352
.75	0	1	0.301
.5	0	5	0.079
.75	-2	1	0.009
-2	-2	1	0.399

- The values from $f(x)$ were obtained by using Excel
  - The "=normdist()" function
  - Most statistics packages have a normal distribution function

- The mean of the normal distribution is μ_x
- The variance of the normal distribution is σ_x^2

EXPECTED VALUES AND THE ALGEBRA OF EXPECTATIONS

Expected Values

- Expected values are statistics taken the sample space of a random variable: they are essentially weighted averages

- The weights used in computing this average correspond to the densities for a continuous random variable (the heights of the distribution)

- Notation: the expected value is represented by: $E(x)$
  - *The actual statistic that is being weighted by the PDF is put into the parentheses where* x *is now*

- Expected values allow us to understand what a statistical model implies about data, for instance:
  - How a GLM specifies the (conditional) mean and variance of a DV

Expected Value Calculation

- For discrete random variables, the expected value is found by:

$$E(x) = \sum_x xP(X = x)$$

- For example, the expected value of a roll of a die is:

$$E(x) = (1)\frac{1}{6} + (2)\frac{1}{6} + (3)\frac{1}{6} + (4)\frac{1}{6} + (5)\frac{1}{6} + (6)\frac{1}{6} = 3.5$$

- For continuous random variables, the expected value is found by:

$$E(x) = \int_x xf(x)dx$$

- We won’t be calculating theoretical expected values with calculus…we use them only to see how models imply things about our data

Variance and Covariance…As Expected Values

- A distribution's theoretical variance can also be written as an expected value:

$$V(x) = E\left(x - E(x)\right)^2 = E(x - \mu_x)^2$$

  - This formula will help us understand predictions made GLMs and how that corresponds to statistical parameters we interpret

- For a roll of a die, the theoretical variance is:

$$V(x) = E(x - 3.5)^2 = \frac{1}{6}(1 - 3.5)^2 + \frac{1}{6}(2 - 3.5)^2 + \frac{1}{6}(3 - 3.5)^2 + \frac{1}{6}(4 - 3.5)^2 + \frac{1}{6}(5 - 3.5)^2 + \frac{1}{6}(6 - 3.5)^2 = 2.92$$

  - Likewise, the SD is then $\sqrt{2.92} = 1.71$

- Likewise, for a pair of random variables x and z, the covariance can be found from their joint distributions:

$$Cov(x, z) = E(xz) - E(x)E(z) = E(xz) - \mu_x\mu_z$$

LINEAR COMBINATIONS OF RANDOM VARIABLES

Linear Combinations of Random Variables

A **linear combination** is an expression constructed from a set of terms by multiplying each term by a constant and then adding the results

$$x = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

- The linear regression equation is a linear combination

- More generally, linear combinations of random variables have specific implications for the mean, variance, and possibly covariance of the new random variable

- As such, there are predicable ways in which the means, variances, and covariances change
  - These terms are called the algebra of expectations

- To guide us through this process, we will use the descriptive statistics from the height/weight/gender example from our 1st class

Descriptive Statistics for Height/Weight Data

Variable	Mean	SD	Variance
Height	67.9	7.44	55.358
Weight	183.4	56.383	3,179.095
Female	0.5	0.513	0.263

Diagonal: Variance

Above Diagonal: Covariance

Correlation /Covariance	Height	Weight	Female
Height	55.358	334.832	-2.263
Weight	.798	3,179.095	-27.632
Female	-.593	-.955	.263

Below Diagonal: Correlation

Algebra of Expectations

Here are some properties of expected values: x and z are random variables, c and d constants

Sums of Constants:

$$E(x + c) = E(x) + c$$
$$V(x + c) = V(x)$$
$$Cov(x + c, z) = Cov(x, z)$$

Products of Constants:

$$E(cx) = cE(x)$$
$$V(cx) = c^2V(x)$$
$$Cov(cx, dz) = cdCov(x, z)$$

Sums of Random Variables:

$$E(cx + dz) = cE(x) + dE(z)$$
$$V(cx + dz) = c^2V(x) + d^2V(z) + 2cd\big(Cov(x, z)\big)$$

Examples for Algebra of Expectations

- Image you wanted to convert weight from pounds to kilograms (where 1 pound = 0.453 kg)

$$Weight_{kg} = .453Weight_{lb}$$

- The mean (expected value) of weight in kg:

$$E\left(Weight_{kg}\right) = E(.453Weight_{lb}) = .453E(Weight_{lb})$$
$$= .453\overline{Weight_{lb}} = .453 * 183.4 = 83.08\text{kg}$$

- The variance of weight in kg:

$$V\left(Weight_{kg}\right) = V(.453Weight_{lb}) = .453^2\,V(Weight_{lb})$$
$$= .453^2 * 3{,}179.095 = 652.38kg^2$$

- The covariance of weight in kg with height in inches:

$$Cov\left(Weight_{kg}, Height\right) = Cov(.453Weight_{lb}, Height)$$
$$= .453Cov(Weight_{lb}, Height) = .453 * 334.832$$
$$= 151.68\text{kg} * \text{inches}$$

Don’t Take My Word For It…

SAS syntax for transforming weight in a DATA step:

```
DATA htwt;
INPUT id Gender $ height weight ;
IF Gender = 'F' THEN female=1; IF Gender = 'M' THEN female=0;
heightMC = height-67.9;
weightKG = 0.453*weight;
```

SAS syntax for marginal descriptive statistics and covariances:

```
*NEW SAMPLE STATISTICS FOR WEIGHT;
PROC MEANS DATA=htwt MEAN VAR;
VAR weight weightKG;
RUN;

PROC CORR DATA=htwt COV;
VAR weight weightKG height;
RUN;
```

SAS output:

The MEANS Procedure

Variable	Mean	Variance
weight	183.4000000	3179.09
weightKG	83.0802000	652.3788519

Covariance Matrix, DF = 19

	weight	weightKG	height
weight	3179.094737	1440.129916	334.831579
weightKG	1440.129916	652.378852	151.678705
height	334.831579	151.678705	55.357895

Where We Use This…The Dreaded ESTIMATE Statement

- The ESTIMATE statement in SAS computes the expected value and standard error (square root of variance) for a new random variable
  - ➢ The new random variable is a linear combination of the original model parameters (the fixed effects)
  - ➢ The original model parameters are considered "random" here as their sampling distribution is used (assuming normal errors and a large N)

```
MODEL score = Dgroup2 Dgroup3 Dgroup4 experience4 enthusiasm
              Dgroup2*experience4  Dgroup3*experience4  Dgroup4*experience4 / SOLUTION;
ESTIMATE 'experience for mini' experience4 1 dgroup2*experience4 1;
```

$$Estimate = 1 * \beta_{experience4} + 1 * \beta_{G2*experience4}$$

- Where:
  - ➢ $\beta_{experience4}$ has mean $\widehat{\beta_{experience4}}$ and variance $se\left(\widehat{\beta_{experience4}}\right)^2$
  - ➢ $\beta_{G2*experience4}$has mean $\widehat{\beta_{G2*experience4}}$ and variance $se\left(\widehat{\beta_{G2*experience4}}\right)^2$
  - ➢ There exists a covariance between $\widehat{\beta_{experience4}}$ and $\widehat{\beta_{G2*experience4}}$
    - We'll call this $Cov\left(\widehat{\beta_{experience4}}, \widehat{\beta_{G2*experience4}}\right)$

More ESTIMATE Statement Fun

- So…if the estimates are:

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	75.49934727	0.38707620	195.05	<.0001
Dgroup2	-10.07267266	0.54896179	-18.35	<.0001
Dgroup3	4.17623925	0.54961852	7.60	<.0001
Dgroup4	-6.04195829	0.54912685	-11.00	<.0001
experience4	-0.38518388	0.29569936	-1.30	0.1943
enthusiasm	-5.00727782	0.18730609	-26.73	<.0001
Dgroup2*experience4	-0.63103823	0.39198136	-1.61	0.1091
Dgroup3*experience4	-0.10925920	0.41111045	-0.27	0.7907
Dgroup4*experience4	0.16959725	0.41917025	0.40	0.6862

  - And $Cov\left(\widehat{\beta_{experience4}}, \widehat{\beta_{G2*experience4}}\right) = -.08756$

…What is:

$E(Estimate) = E\left(1 * \beta_{experience4} + 1 * \beta_{G2*experience4}\right)$
$= 1 * E\left(\beta_{experience4}\right) + 1 * E\left(\beta_{G2*experience4}\right) = -.385 - .631 = -1.016$
$V(Estimate) = V\left(1 * \beta_{experience4} + 1 * \beta_{G2*experience4}\right)$
$= 1^2 V\left(\beta_{experience4}\right) + 1^2 V\left(\beta_{G2*experience4}\right) + 2 * 1 * 1 Cov\left(\beta_{experience4}, \beta_{G2*experience4}\right) =$
$.296^2 + .391^2 - 2 * .08756 = .0653$

Parameter	Estimate	Standard Error	t Value	Pr > |t|
experience for mini	-1.0162221	0.25685951	-3.96	0.0001

$$se(Estimate) = \sqrt{V(Estimate)} = .257$$

THE GENERAL LINEAR MODEL WITH WHAT WE HAVE LEARNED TODAY

The General Linear Model, Revisited

- The general linear model for predicting Y from X and Z:

$$Y_p = \beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p + e_p$$

In terms of random variables, under the GLM:

- e_p is considered random: $e_p \sim N(0, \sigma_e^2)$

- Y_p is dependent on the linear combination of X_p, Z_p, and e_p

- The GLM provides a model for the **conditional distribution** of the dependent variable, where the conditioning variables are the independent variables: $f(Y_p | X_p, Z_p)$
  - There are no assumptions made about X_p and Z_p - they are constants
  - The regression slopes $\beta_0, \beta_1, \beta_2, \beta_3$ are constants that are said to be fixed at their values (hence, called fixed effects)

Combining the GLM with Expectations

- Using the algebra of expectations predicting Y from X and Z:

The expected value (mean) of $f(Y_p|X_p, Z_p)$:

$$\hat{Y}_p = E(Y_p) = E(\underbrace{\beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p}_{\text{Constants}} + \underbrace{e_p}_{\text{Random Variable with } E(e_p)=0})$$

$$= \beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p + E(e_p)$$

$$= \beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p$$

The variance of $f(Y_p|X_p, Z_p)$:

$$V(Y_p) = V(\beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p + e_p) = V(e_p) = \sigma_e^2$$

Distribution of $f\left(Y_p \middle| X_p, Z_p\right)$

- We just found the mean (expected value) and variance implied by the GLM for the conditional distribution of Y_p given X_p and Z_p

- The next question: what is the distribution of $f\left(Y_p \middle| X_p, Z_p\right)$?

- Linear combinations of random variables that are normally distributed result in variables that are normally distributed

- Because $e_p \sim N(0, \sigma_e^2)$ is the only random term in the GLM, the resulting conditional distribution of Y_p is normally distributed:

$$Y_p | X_p, Z_p \sim N\left(\underbrace{\beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p}_{\text{Model for the means}}, \underbrace{\sigma_e^2}_{\text{Model for the variances}}\right)$$

Model for the means: from fixed effects; literally gives mean of $f\left(Y_p \middle| X_p, Z_p\right)$

Model for the variances: from random effects; gives variance of $f\left(Y_p \middle| X_p, Z_p\right)$

Examining What This Means in the Context of Data

- If you recall from our first lecture, the final model we decided to interpret: Model 5

$$W_p = \beta_0 + \beta_1\left(H_p - \bar{H}\right) + \beta_2 F_p + \beta_3\left(H_p - \bar{H}\right)F_p + e_p$$

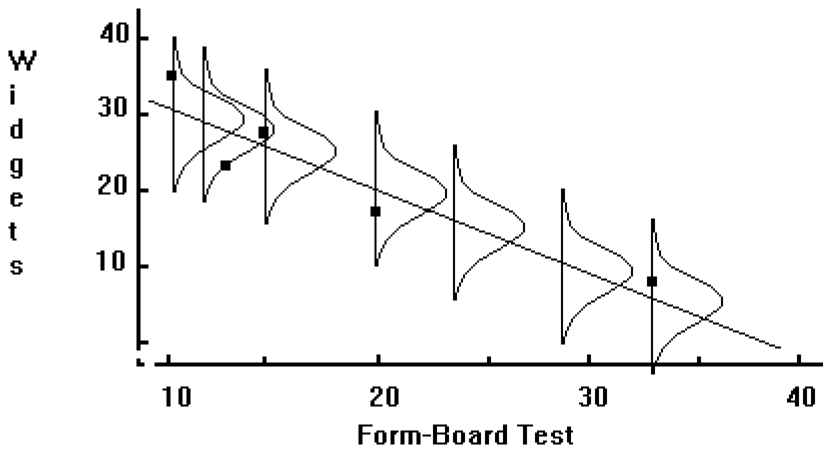
where $e_p \sim N(0, \sigma_e^2)$

- From SAS:

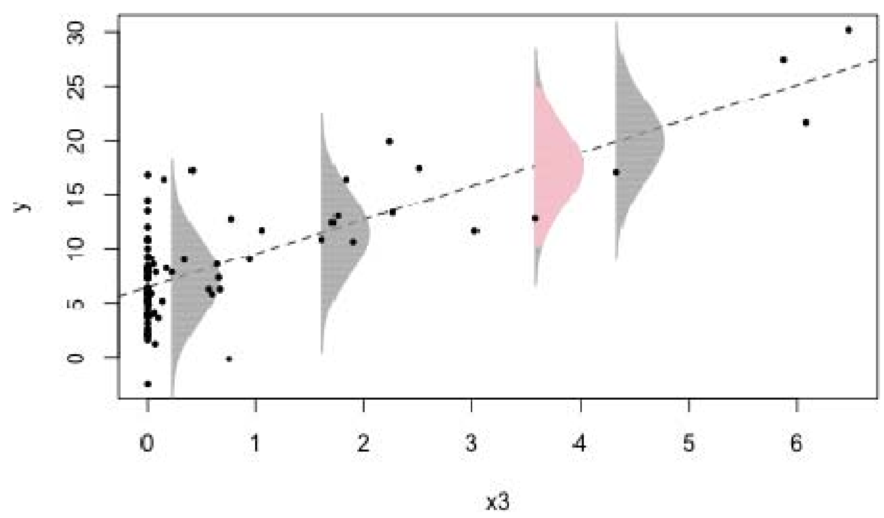
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	222.1841719	0.83809108	265.11	<.0001
heightMC	3.1897275	0.11135027	28.65	<.0001
female	-82.2719216	1.21109969	-67.93	<.0001
heightMC*female	-1.0938553	0.16777741	-6.52	<.0001

Picturing the GLM with Distributions

The distributional assumptions of the GLM are the reason why we do not need to worry if our dependent variable is normally distributed



Our dependent variable should be **conditionally** normal

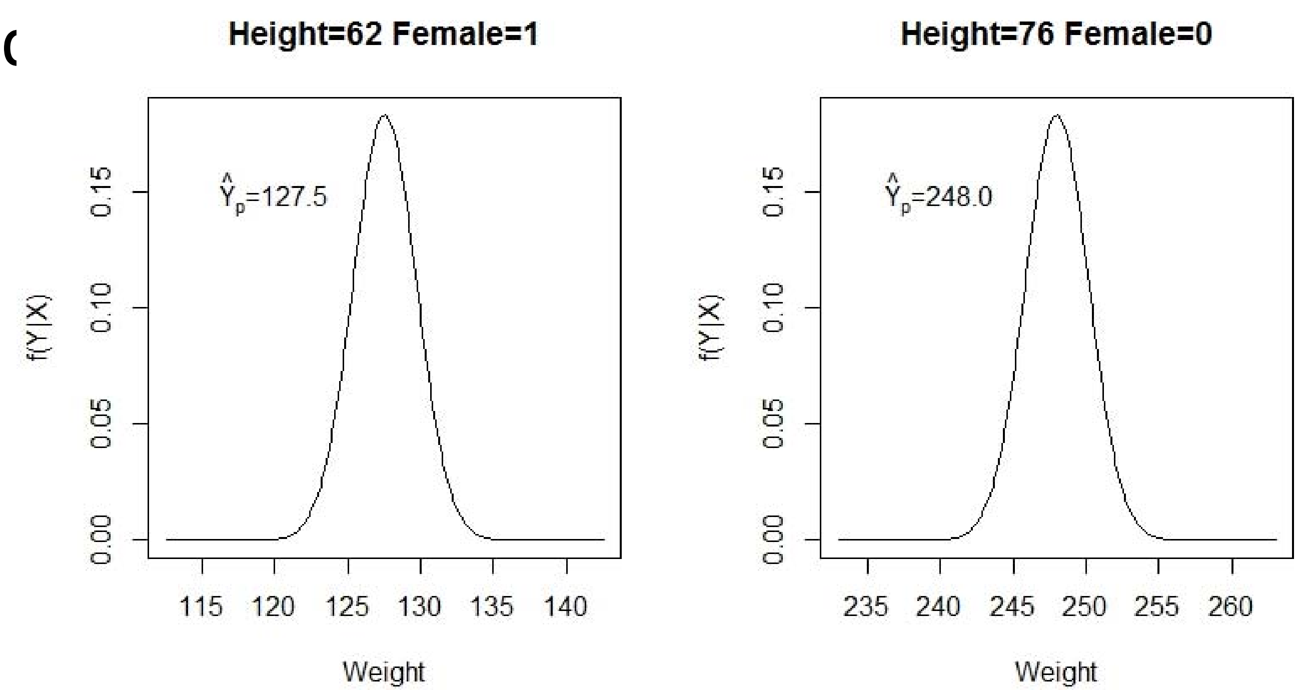


We can check this assumption by checking our assumption about the residuals, $e_p \sim N(0, \sigma_e^2)$

More on this soon…

More Pictures of the GLM

- Treating our estimated values of the slopes $(\beta_0, \beta_1, \beta_2, \beta_3)$ and the residual variance (σ_e^2) as the true values* we can now see what the theoretical* distribution of $f(Weight_p | Height_p, Female_p)$ looks like for a given set of pre



*Note: these distributions change when sample estimates are used (think standard error of the prediction)

Behind the Pictures…

- To emphasize the point that PDFs provide the height of the line, here is the normal PDF (with numbers) that produced those plots:

$$f\left(W_p \middle| H_p, F_p\right) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{\left(W_p - \widehat{W}_p\right)^2}{2\sigma_e^2}\right)$$

Model for the Means

$$= \frac{1}{\sqrt{2\pi\boxed{\sigma_e^2}}} \exp\left(-\frac{\left(W_p - \boxed{\left(\beta_0 + \beta_1\left(H_p - \bar{H}\right) + \beta_2 F_p + \beta_3\left(H_p - \bar{H}\right)F_p\right)}\right)^2}{2\boxed{\sigma_e^2}}\right)$$

$$= \frac{1}{\sqrt{2\pi(4.73)}} \exp\left(-\frac{\left(W_p - \left(222.18 + 3.19\left(H_p - \bar{H}\right) - 82.27F_p - 1.09\left(H_p - \bar{H}\right)F_p\right)\right)^2}{2(4.73)}\right)$$

Model for the Variance

The plots were created using the following value for the predictors:

$\bar{H} = 67.9$

Left plot: $H_p = 62; F_p = 1$

Right plot: $H_p = 76; F_p = 0$

ESTIMATION OF GLMS USING LEAST SQUARES

Why Estimation is Important

- In "applied" statistics courses, estimation is not discussed very frequently
  - Can be very technical…very intimidating

- Estimation is of critical importance
  - Quality and validity of estimates (and of inferences made from them) depends on how they were obtained
  - New estimation methods appear from time to time and get widespread use without anyone asking whether or not they are any good

- Consider an absurd example:
  - I say the mean for IQ should be 20 – just from what I feel
  - Do you believe me? Do you feel like reporting this result?
    - Estimators need a basis in reality (in statistical theory)

How Estimation Works (More or Less)

- Most estimation routines do one of three things:

1. **Minimize Something:** Typically found with names that have “least” in the title. Forms of least squares include “Generalized”, “Ordinary”, “Weighted”, “Diagonally Weighted”, “WLSMV”, and “Iteratively Reweighted.” Typically the estimator of last resort…

2. **Maximize Something:** Typically found with names that have “maximum” in the title. Forms include “Maximum likelihood”, “ML”, “Residual Maximum Likelihood” (REML), “Robust ML”. Typically the gold standard of estimators (and next week we’ll see why).

3. **Use Simulation to Sample from Something:** more recent advances in simulation use resampling techniques. Names include “Bayesian Markov Chain Monte Carlo”, “Gibbs Sampling”, “Metropolis Hastings”, “Metropolis Algorithm”, and “Monte Carlo”. Used for complex models where ML is not available or for methods where prior values are needed.

Estimation of General Linear Models

- Recall our GLM (shown here for the prediction of a dependent variable Y_p by two independent variables X_p and Z_p):

$$Y_p = \beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p + e_p$$

- Traditionally (dating to circa 1840), general linear models can be estimated via a process called least squares

- Least squares attempts to find the GLM parameters (the βs) that minimize the **squared residual** terms:

$$\min_{\{\beta_0, \beta_1, \beta_2, \beta_3\}} \left\{ \sum_{p=1}^{N} e_p^2 \right\}$$

Where We Are Going (and Why We Are Going There)

- Because the basics of estimation are critical to understanding the validity of the numbers you will use to make inferences from, we will detail the process of estimation
  - Today with Least Squares and then ending with Maximum Likelihood

- The LS estimation we will discuss is to get you to visualize functions of statistical parameters (the βs here) and data in order to show which estimates we would choose
  - To be repeated: In practice LS estimation for GLMs does not do this (by the magic of calculus and algebra)

- In the end, we would like for you to understand that not all estimators are created equally and that some can be trusted more than others
  - We would also like for you to see how estimation works so you can fix it when it goes wrong!

How Least Squares Estimation Works

- How Least Squares works is through minimizing the squared error terms…but its what goes into error that drives the process:

$$e_p = Y_p - \hat{Y}_p = Y_p - \left(\beta_0 + \beta_1 X_p + \beta_2 Z_p + \beta_3 X_p Z_p\right)$$

- If you were to do this (and you wouldn't), the process called optimization would go like this:

1. Pick values for regression slopes
2. Calculate $\hat{Y}_p$ and then e_p for each person p
3. Calculate $OF = \sum_{p=1}^{N} e_p^2$ (letters OF stand for **objective function**)
4. Repeat 1-3 until you find the values of regression slopes that lead to the smallest value of OF

Today's Example Data

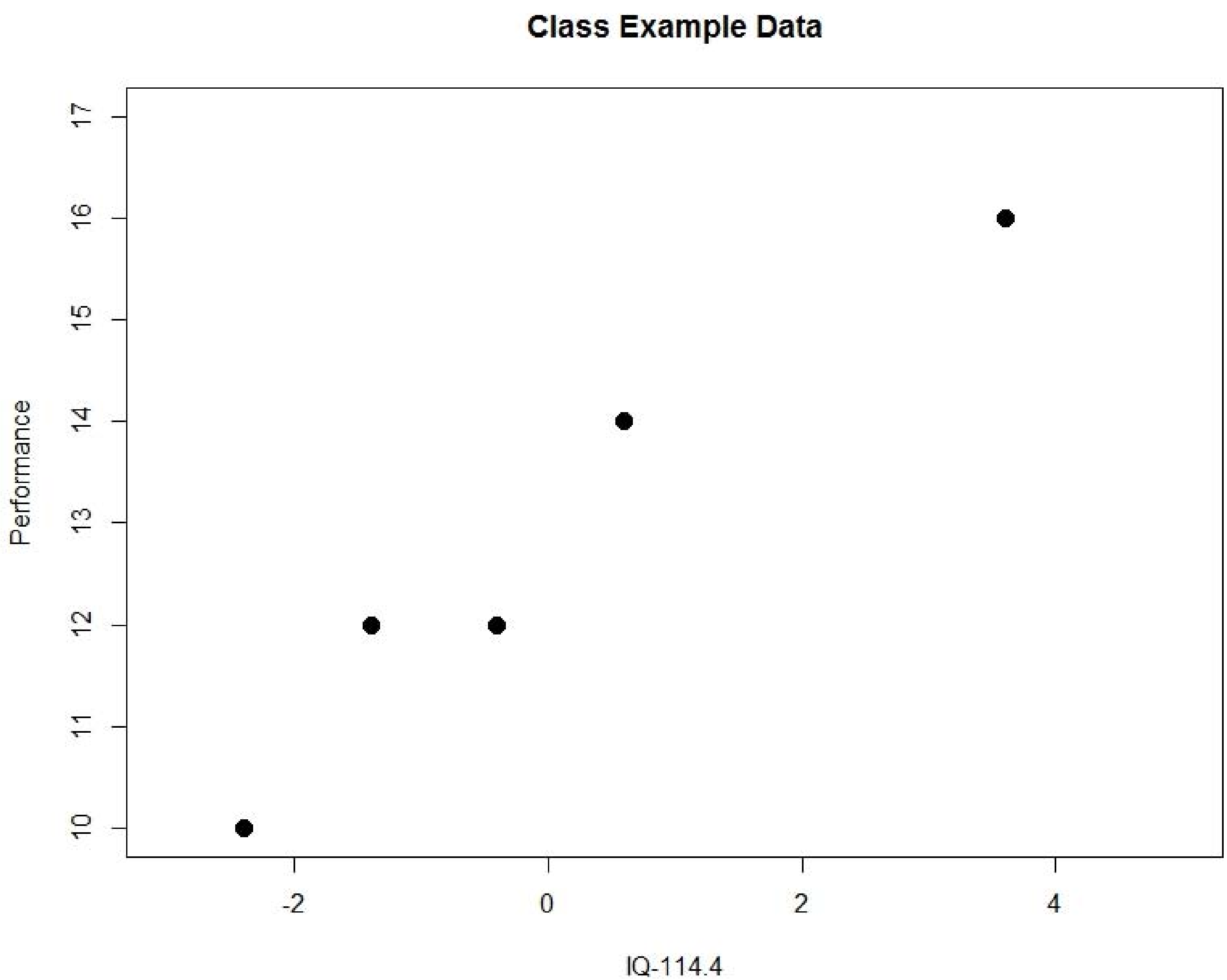
- Imagine an employer is looking to hire employees for a job where IQ is important
  - We will only use 5 observations so as to show the math behind the estimation calculations
- The employer collects two variables:
  - IQ scores
  - Job performance
- Descriptive Statistics:

Variable	Mean	SD
IQ	114.4	2.30
Performance	12.8	2.28

Covariance Matrix		
IQ	5.3	5.1
Performance	5.1	5.2

Observation	IQ	Performance
1	112	10
2	113	12
3	115	14
4	118	16
5	114	12

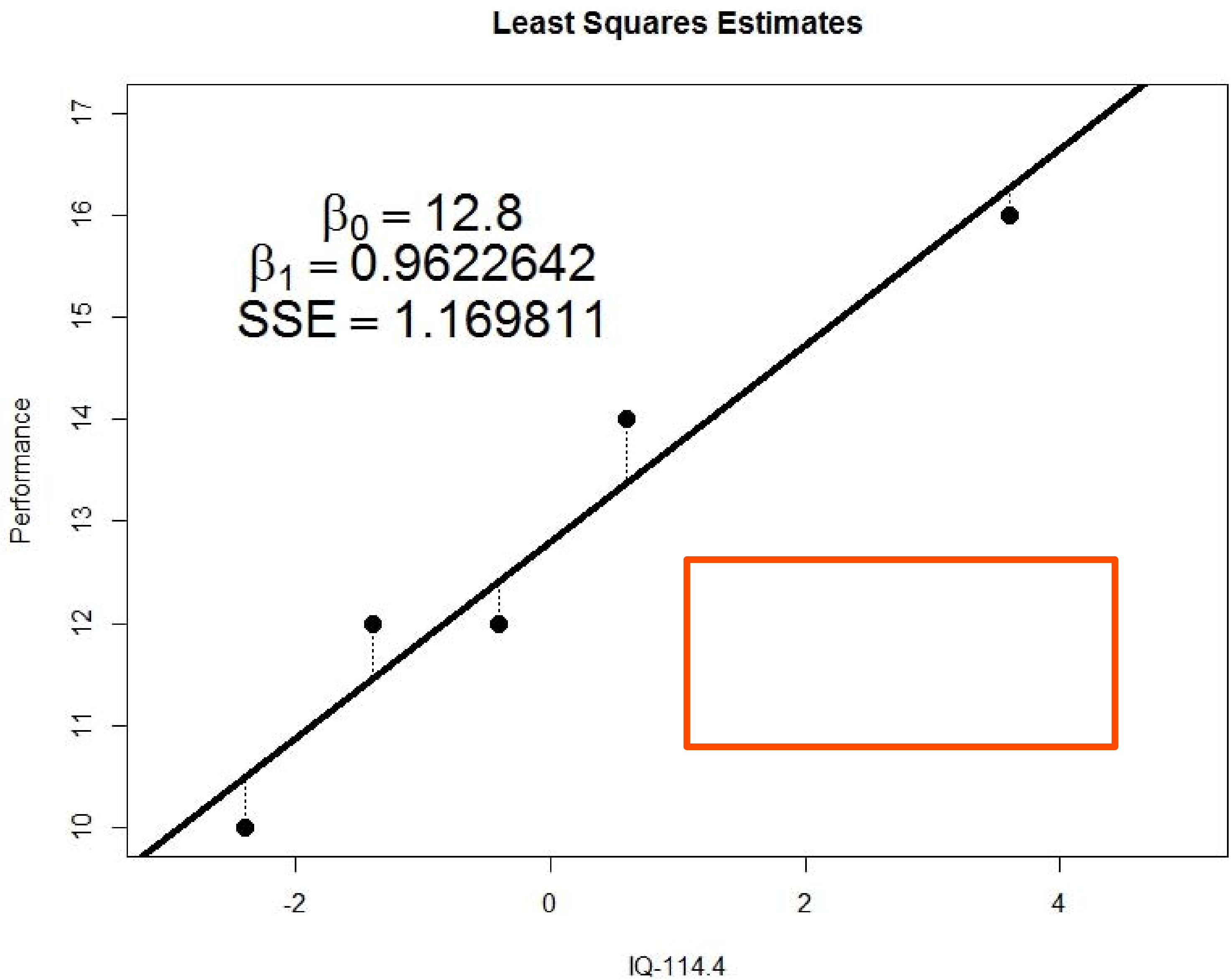
Visualizing the Data



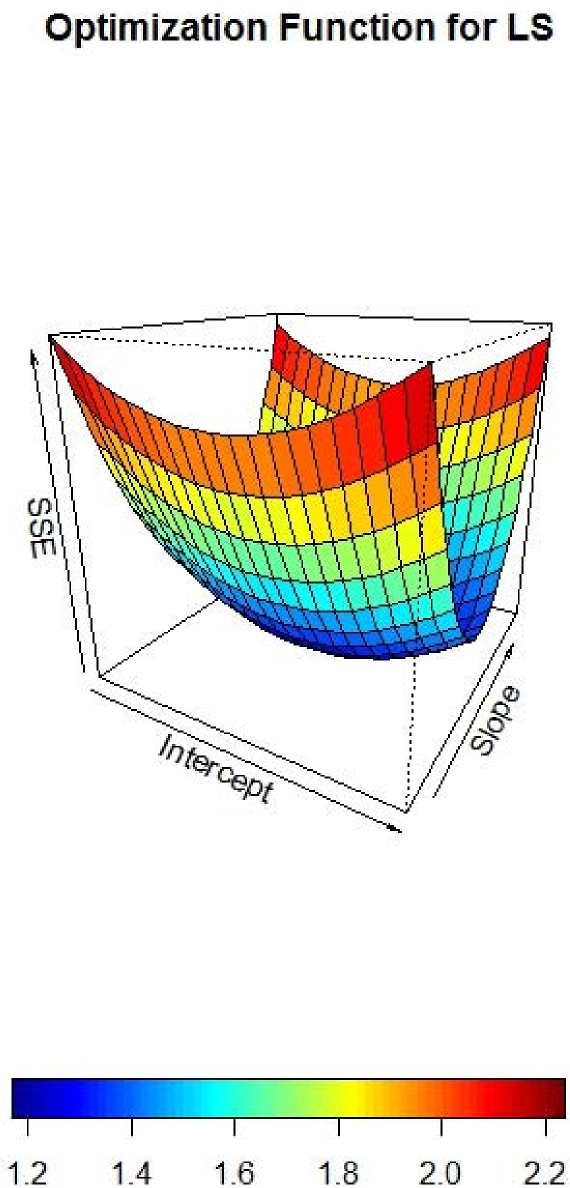
Let’s Play…Pick the Parameters…

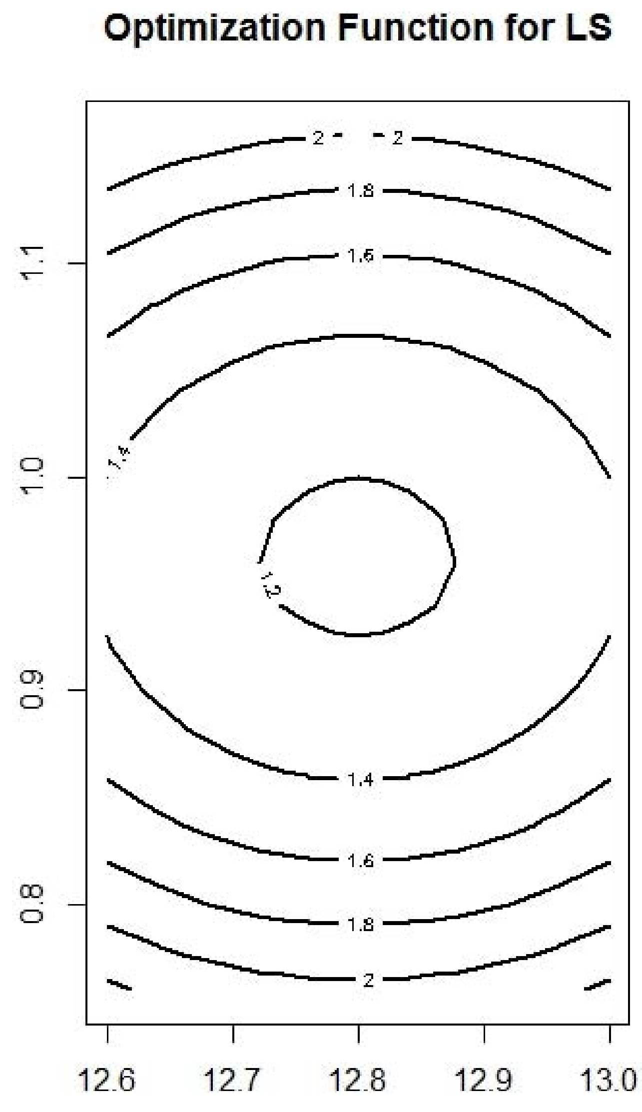
- This slide is left as a placeholder for the Camtasia recording – we will now do a demonstration in R

And…The Winner Is…



Examining the Objective Function Surface





LS Estimates of GLMs

- The process of least squares estimation of GLMs does not need an iterative search

- Using calculus, a minimum of the objective function can be found
  - This involves taking the first derivative of the objective function with respect to each parameter
    - Derivative = slope of the tangent line for a given point
  - The first derivative is then set equal to zero
    - Flat slope = minimum (or maximum or saddle point – neither apply here)
  - The equation is then solved for the parameter
    - Producing the equations you know and love

- For simple linear regression (one predictor):
  - $\beta_X = \frac{\frac{1}{\mathrm{N-k}}\sum_{p=1}^{N}(X_p - \bar{X})(Y_p - \bar{Y})}{\frac{1}{\mathrm{N-k}}\sum_{p=1}^{N}(X_p - \bar{X})(X_p - \bar{X})} = \frac{covariance\ of\ X\ and\ Y}{covariance\ of\ X\ and\ X} = \frac{sum\ of\ cross{-}products}{sum\ of\ squared\ Xs}$

- When we get to matrix algebra, you will know this as

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Ramifications of Closed-Form LS Equations

- Because least squares estimates have a **closed form** (equations that will provide statistics directly), they will work nearly every time
  - Only fail when collinearity is present (soon you'll know this to mean $\mathbf{X}^T\mathbf{X}$ is singular and cannot be inverted)
- Virtually all other estimators you will encounter in statistics will not have a closed form
  - Even least squares estimators for other types of data (not continuous)
- Without a closed form, least squares estimates are found by search the objective function for its minimum
  - Like finding the drain of a pool….

Why LS is Still in Use

- Least squares estimates still make up the bulk of GLM analyses because:
  - They are easy to compute
  - They pretty much always give you an answer
  - They have been shown to have good statistical properties

- The good statistical properties actually come because LS estimates of GLMs match the **Maximum Likelihood Estimates**
  - We will learn more about maximum likelihood estimation next
  - For now, know that MLEs are the gold standard when it comes to estimates

Where LS Fails

- For all their flexibility, least squares estimates are somewhat limited
  - Only have good properties for basic univariate GLM for continuous data
    - Normally distributed error terms with homogeneous variance

- When data are not continuous/do not have normally distributed error terms, least squares estimates are not preferred

- For multivariate models with continuous data (repeated measures, longitudinal data, scales of any sort), least squares estimates quickly do not work
  - Cannot handle missing outcomes (deletes entire case)
  - Limited in the types of ways of modeling covariance between observations

WRAPPING UP

Wrapping Up

- Today discussed estimation, and in the process showed how differing estimators can give you different statistics

- The key today was to shake your statistical view point:
  - There are many more ways to arrive at statistical results than you may know

- The take home point is that **not all estimators are created equal**
  - If ever presented with estimates: ask how the numbers were attained
  - If ever getting estimates: get the best you can with your data

- Next week your world will further be expanded when we introduce maximum likelihood estimators