The University of Georgia

Applied Analysis of Variance Methods in Education

Fall 2009

Educational Research 8310

Today's Lecture

- Syllabus and course information.
- Information sheet, introduction.
- ANOVA Quiz.
- Brief review of basic concepts:
  - Basic mathematics we will be using.
  - Univariate statistical terminology.
  - Bivariate statistical terminology.
  - Simple linear regression.
- Lab time: Introduction to SPSS.

Review of Basic Math

- As you know, statistics is a field of applied mathematics.
  - Much is written in mathematical formulas, necessitating the understanding of some fundamental mathematics concepts.
- To help remember many of the key concepts of math used in this class, we will now review several mathematics terms:
  - Summation.
  - Exponents.
  - e - Euler's number.
  - Logarithms.
- Please refer to these terms as you go along in the class and keep these slides available as a quick reference.

Rules of Summation

- First, we discuss the summation operator.

$$\sum$$

- The summation symbol dictates what to add.
  - This typically is over a variable, x - say someone's test score.
  - The variable typically is subscripted: x_i.
- Say you have five people, each with a number of items correct out of 20: {10, 12, 15, 7, 20}.
- Let x be a score on a test.
- We will say that x_i is the score person of i on that test, where $i = 1, \ldots, 5$.

Rules of Summation, Continued

- The sum symbol then tells you which of the values to add up:

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5 = 10 + 12 + 15 + 7 + 20 = 54$$

- Often times, we have multiple subscripts.
  - Say j will be for a classroom ($j = 1, \ldots, J$ - there are J classrooms).
  - Say i will be for a person in a classroom ($i = 1, \ldots, N_j$ - there are N_j people in classroom j).
- To sum across all people and classrooms we have:

$$\sum_{j=1}^{J} \sum_{i=1}^{N_j} x_{ij} = x_{11} + x_{21} + \ldots + x_{N_1 1} + x_{12} + x_{22} + \ldots + x_{N_1 2} + \ldots + x_{1J} + x_{2J} + \ldots + x_{N_1 J}$$

Rules of Summation

There are a few rules for sums:

1. If a is some constant value over the N observations i, then:

$$\sum_{i=1}^{N} a = Na$$

2. Given the value a, which is constant over all people entering into the summation, then:

$$\sum_{i=1}^{N} ax_i = a\sum_{i=1}^{N} x_i$$

3. Operations on the variables summed must be carried out before the sum:

$$\sum_{i=1}^{N} x_i^2 = x_1^2 + x_2^2 + \ldots + x_N^2$$

Rules of Exponents

- Exponents occur frequently in statistics, and represent a quantity multiplied by itself a number of times:

$$x^2 = x \times x$$

- Here are some basic exponential rules:
  - $x^m \times x^n = x^{m+n}$
  - $\frac{x^m}{x^n} = x^{m-n}$
  - $x^{-n} = \frac{1}{x^n}$
  - $\left(x^m\right)^n = x^{mn}$
  - $(xy)^n = x^n y^n$
  - $\left(\frac{x}{y}\right)^n = \left(\frac{x^n}{y^n}\right)$

Euler's Number

- From time to time, the letter e or the phrase $\exp$ appears in statistical formulas.
- This is short hand for a number called Euler's number.
  - $e = e^1 \approx 2.718$
- Euler's number comes from more complicated mathematics (beyond the scope of this course).
- If you are interested in learning more, consult the Wikipedia entry on the number at:

  `http://en.wikipedia.org/wiki/E_(mathematical_constant).`

Rules of Logarithms

- The definition of a logarithm (also called a log function) is:

  The logarithmic function with base b, where $b > 0$ and $b \neq 1$, is denoted by $y = \log_b x$ if and only if $b^y = x$.

- What? What? What?

- Logarithms are another way to write an exponent.

- They are used frequently in statistics to transform data or to make complicated mathematical formulas (with exponents) easier to work with.

  - We will see how to transform data in SPSS later in lab.

Rules of Logarithms

Here are some rules of logarithms:

1. $\log_a a^x = x$.
2. $a^{\log_a x} = x$.
3. $\log_a(xy) = \log_a x + \log_a y$.
4. $\log_a \frac{x}{y} = \log_a x - \log_a y$.
5. $\log_a (x^p) = p \log_a x$.
6. $\log_a x = \frac{\log_b x}{\log_b a}$ (change of base).

However:

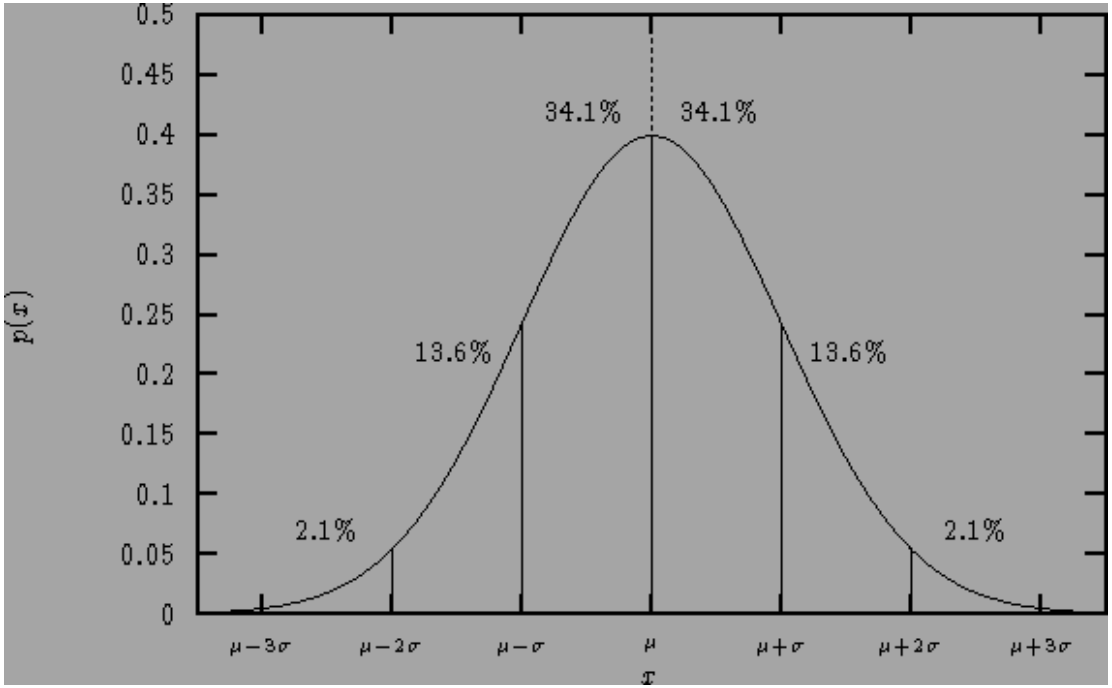
- $\log_a(x + y) \neq \log_a x + \log_a y$.
- $\log_a(x - y) \neq \log_a x - \log_a y$.
- $\log_e$ is often called the natural log or $\ln$.

Univariate statistics

A single random variable can be described by several characteristics of it's distribution:

- Measures of central tendency (or location).
  - Mean (a.k.a. expected value).
  - Median (point where 50% of distribution falls below)
  - Mode (most likely value)
- Measures of spread.
  - Variance and/or standard deviation.
  - Range(s).
- Measures that most normal people don't think about.
  - Skewness - a measure of the symmetry of a distribution.
  - Kurtosis - a measure of the peakedness of a distribution.

Our Two Favorites



The mean:

$$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N}$$

The variance:

$$s_x^2 = \frac{\sum_{i=1}^{N}(X - \bar{X})^2}{N-1} = \frac{\sum_{i=1}^{N} X^2 - \frac{(\Sigma_{i=1}^{N} X)^2}{N}}{N-1}$$

Additional Notes

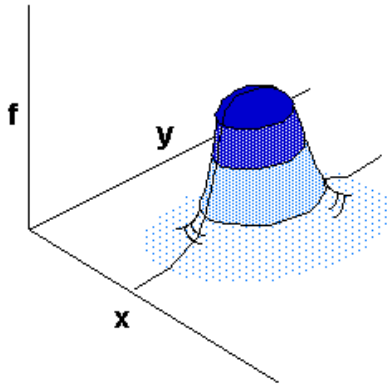
- Units associated with variance are squared (e.g., a variance of a distribution of height is given in $feet^2$).
- More understandable is something where units aren't squared: standard deviation:

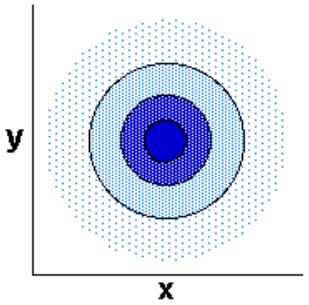
$$s_x = \sqrt{s_x^2}$$

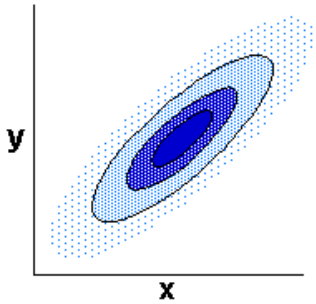
- We divide the sum of squares by $N - 1$ to obtain an unbiased estimate of the variance.
- Variance/Standard Deviation are positive numbers.

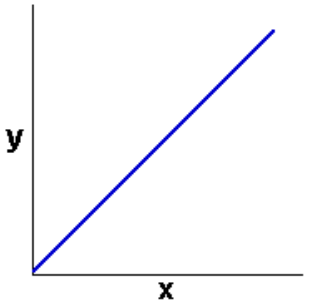
Bivariate Distributions

Each of these images shows the distribution of a pair of variables:









Bivariate Descriptive Statistics

- Covariance: the joint covariation of two sets of variables from their respective means.

$$s_{xy} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{N-1} = \frac{\sum_{i=1}^{N} XY - \frac{(\Sigma_{i=1}^{N} X)(\Sigma_{i=1}^{N} Y)}{N}}{N-1}$$

  - Units are products of units used to create measure (e.g., the covariance of a distribution of height and weight might be reported in *foot-pounds*).
  - Covariance of a variable and itself is the variance.
  - Covariance can take the value of any real number.

Bivariate Descriptive Statistics

- Because products of units can be hard to interpret, consider the correlation:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

  - Correlation is “unitless.”
  - Range of correlation is from $[-1, 1]$.

A Single Hand Calculated Example

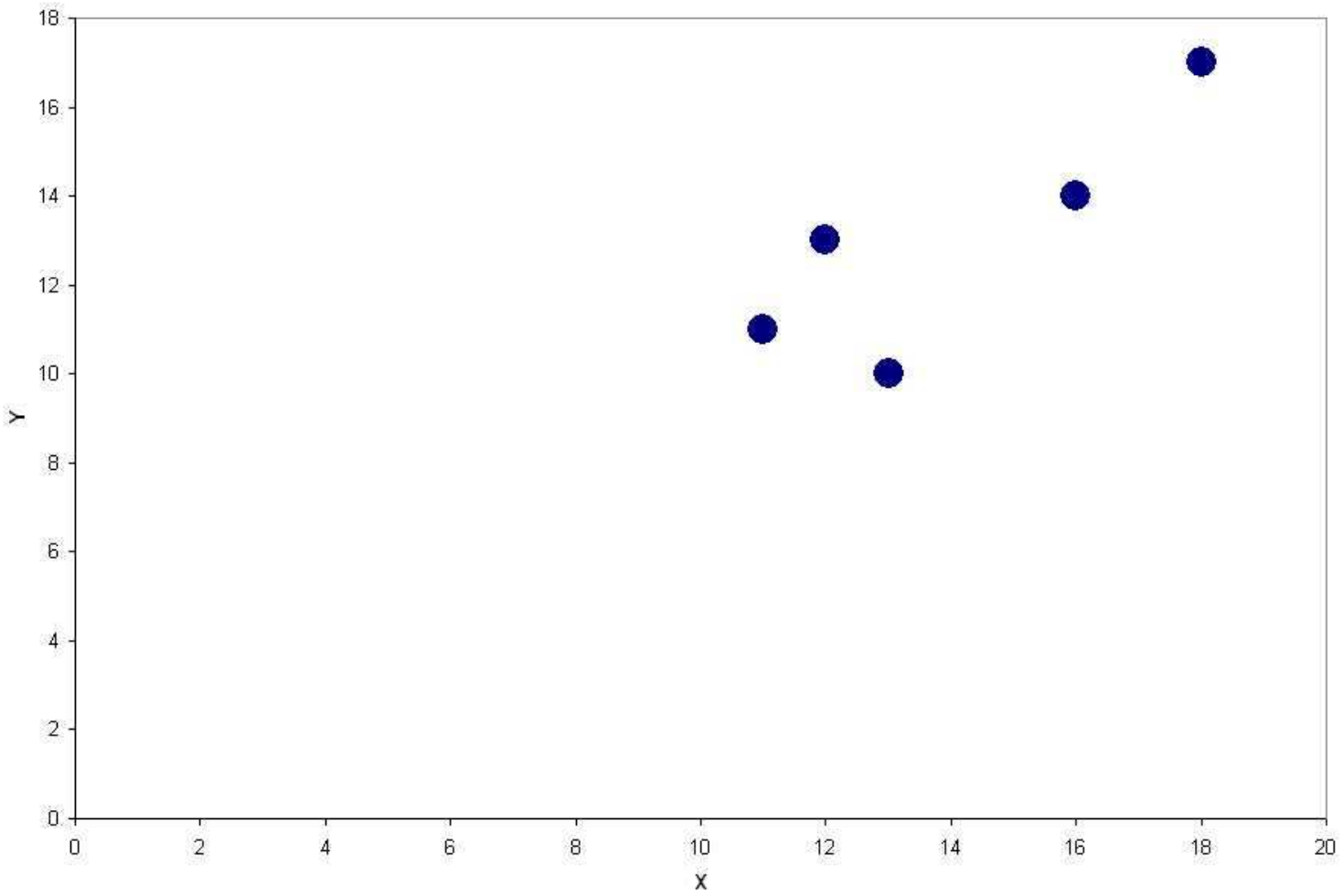
Consider some scores of students on Algebra and Geometry tests:

Person	Algebra (X)	Geometry (Y)	X^2	Y^2	XY
A	11	11	121	121	121
B	13	10	169	100	130
C	18	17	324	289	306
D	12	13	144	169	156
E	16	14	256	196	224
N=5	70	65	1014	875	937

$\sum X = 70$	$\sum Y = 65$	$\sum XY = 937$
$\sum X^2 = 1014$	$\sum Y^2 = 875$	

$\bar{X} = 14$	$\bar{Y} = 13$
$s_X^2 = 8.5$	$s_Y^2 = 7.5$
$s_{XY} = 6.75$	$r_{XY} = 0.845$

A Single Hand Calculated Example



$$r_{XY} = 0.845$$

Summary

The statistics reviewed today are descriptive measures of the distribution of one or two variables.

- No inferences.
- Just descriptions.

Seeing The Future

- Tonight's lab: Introduction to SPSS.
- Homework due prior to the start of class next week.
- Read for next week:
  - Chapter 1: Experimental Design.
  - Chapter 2: Simple Linear Regression.