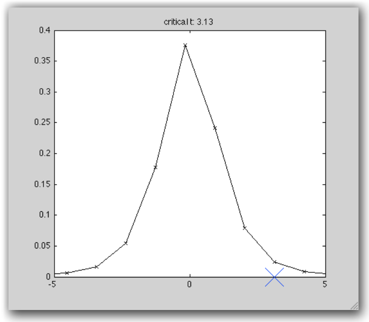
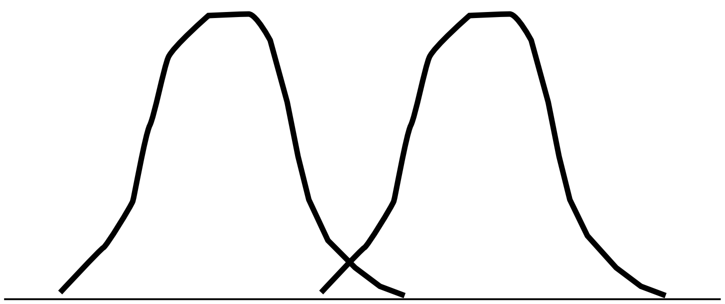
PSYC 7302/8302

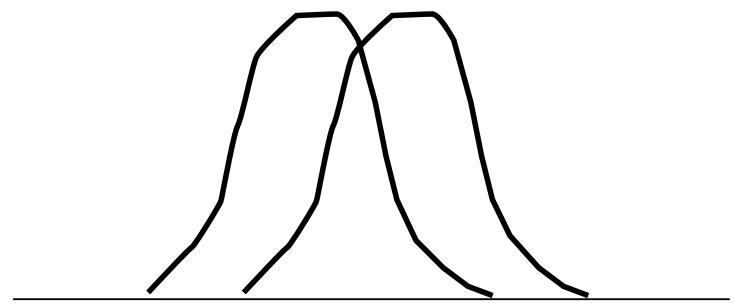
Data, descriptives, and display (II)



Appendix E

- Keith is a wonderful writer...
- “interocular trauma test”





Hypothesis testing

- The methods throughout the course are based on a “logic” called “**null-hypothesis testing**” logic
- Is an obtained difference between two means **significant**?

Logic

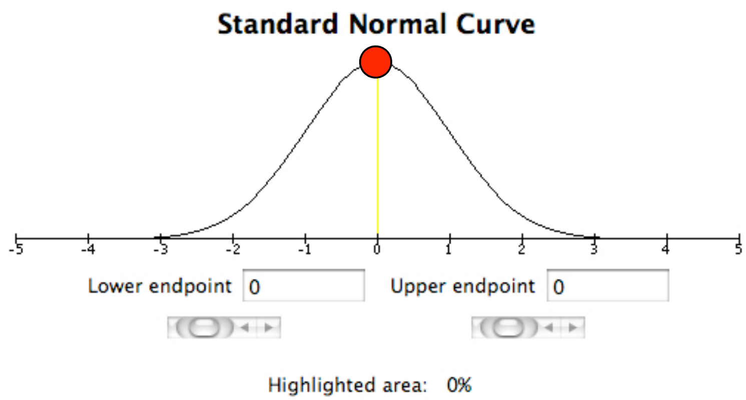
- Let’s assume *no*: This is the **null hypothesis**
- Let’s compute some **statistic of interest**
  - e.g., M1-M2 (i.e., mean of group 1 minus...)
  - e.g., M-3 (group 3’s mean minus 3)
- From certain **assumptions**, can we conclude that this obtained difference could have occurred just **by chance**?

First...

- Mean
- Variance
- Standard deviation
- z scores
- Percentiles

Mean

- Sum of observations divided by number of observations (N)
- “Central tendency”



Variance

- Variance (V) is the average squared variation (from the mean) in a set of observations
  - "Sum of Squares" (SS)
- In fact, we divide by N - 1

Standard deviation

- Is the square root of variance
- This obtains units in the original scale of our observations
- What does standard deviation (SD) do for us?

z-scores

- Because SD is in our original scale, we can say how many SDs a score is from a mean. This is very informative when we examine a particular score in an obtained sample.
- The # of SDs is known as the z score. It is very easy to calculate z scores, once you have M and SD...

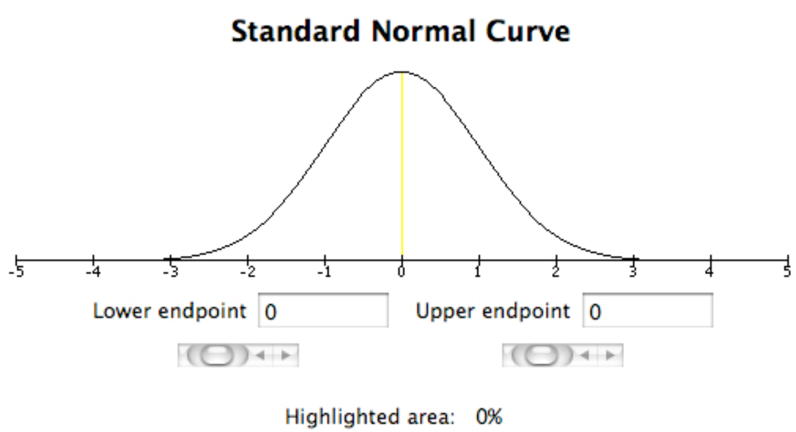
z scores

- First, choose an individual observation from our sample. Call it X.
- Second, take X - M
- Now, divide (X - M) by SD:
  - $z = (X - M)/SD$
- E.g., z could equal: 1, 2, 0, -1

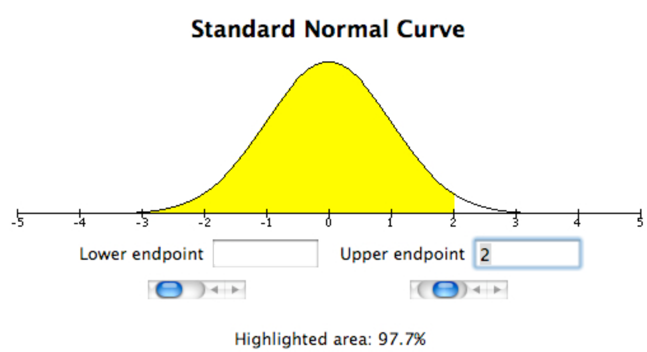
z scores

- Assume, as we often do, that our sample obeys a normal distribution.
- When we convert our sample into z scores, we standardize our sample
- Our z scores have a M of...
  - 0
- Our z scores have a SD
  - 1

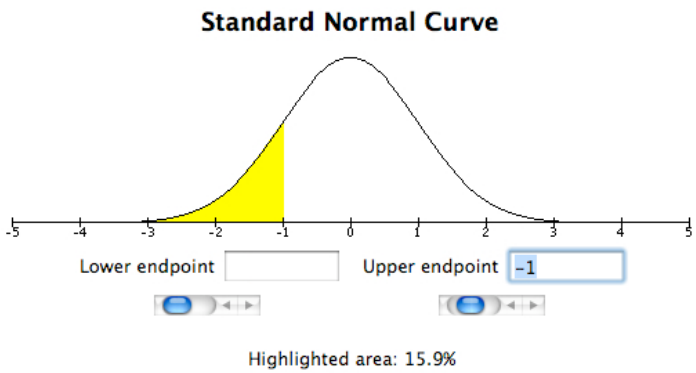
Normal distribution



Normal distribution

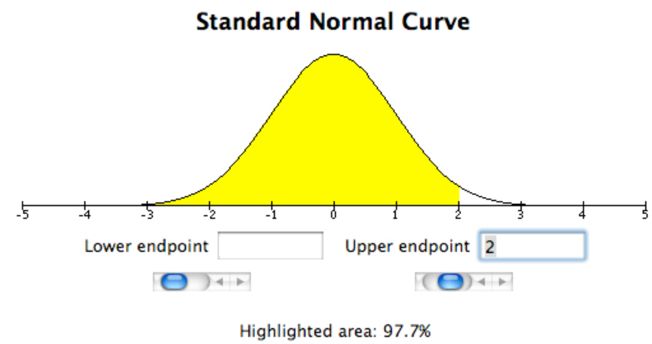


Normal distribution



z score indicates this score is at (approximately) the 16th percentile

Normal distribution



98th percentile

First...

- Mean
- Variance
- Standard deviation
- z-scores
- Percentiles

Second...

- Standard error of the mean
- t-tests
- Confidence interval

A single sample...

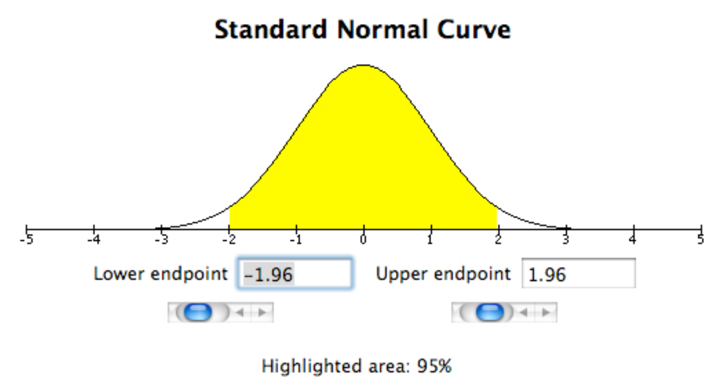
- Our z-scores were based on a single sample. However, how do we determine how much this sample represents the "true population"
- Sample vs. population

Standard error

- We have N observations in our sample.
- Imagine taking multiple samples of this size.
- How much does this "sample of samples" represent the population?
- For any "statistic" (e.g., M) we can compute the standard error (SE) representing the degree of variability our sample will likely have around the "true mean"

Confidence interval

- SE can help us define the confidence interval around our sample mean.
- This confidence interval represents the probability that our interval captures the "true mean" of the population.



Confidence interval

- SE-units instead of SD-units
- Most often, one uses a special statistic called the t-statistic to obtain the desired # of SEs around our sample mean which captures that 95% interval.
- **Warning about confidence interval interpretation... (Keith, p. 11)**

Why t?

- We are now talking about the population, not our sample -- we want to make **inferences** about the "true mean"
- The SE is a score which represents how our sample varies from the population mean, given our SD and the size of our sample.

t

- "Student's" t
  - Actually: William Gosset, 1876-1936
- It is a distribution that emerges from our sampling many times from a population using the equation **statistic / SE**
- E.g., M / SE(of the mean)

z vs. t

z	**t**
(X - M) / SD for one observation	M / SE for a sample
if approximately normal, 95% of **sample** falls within +/-1.96 SD's.	if approximately normal, 95% of such **samples** falls within +/- t_{crit}*SE.

Exercise 2