Nemours Biomedical Research Statistics Course

Session 3

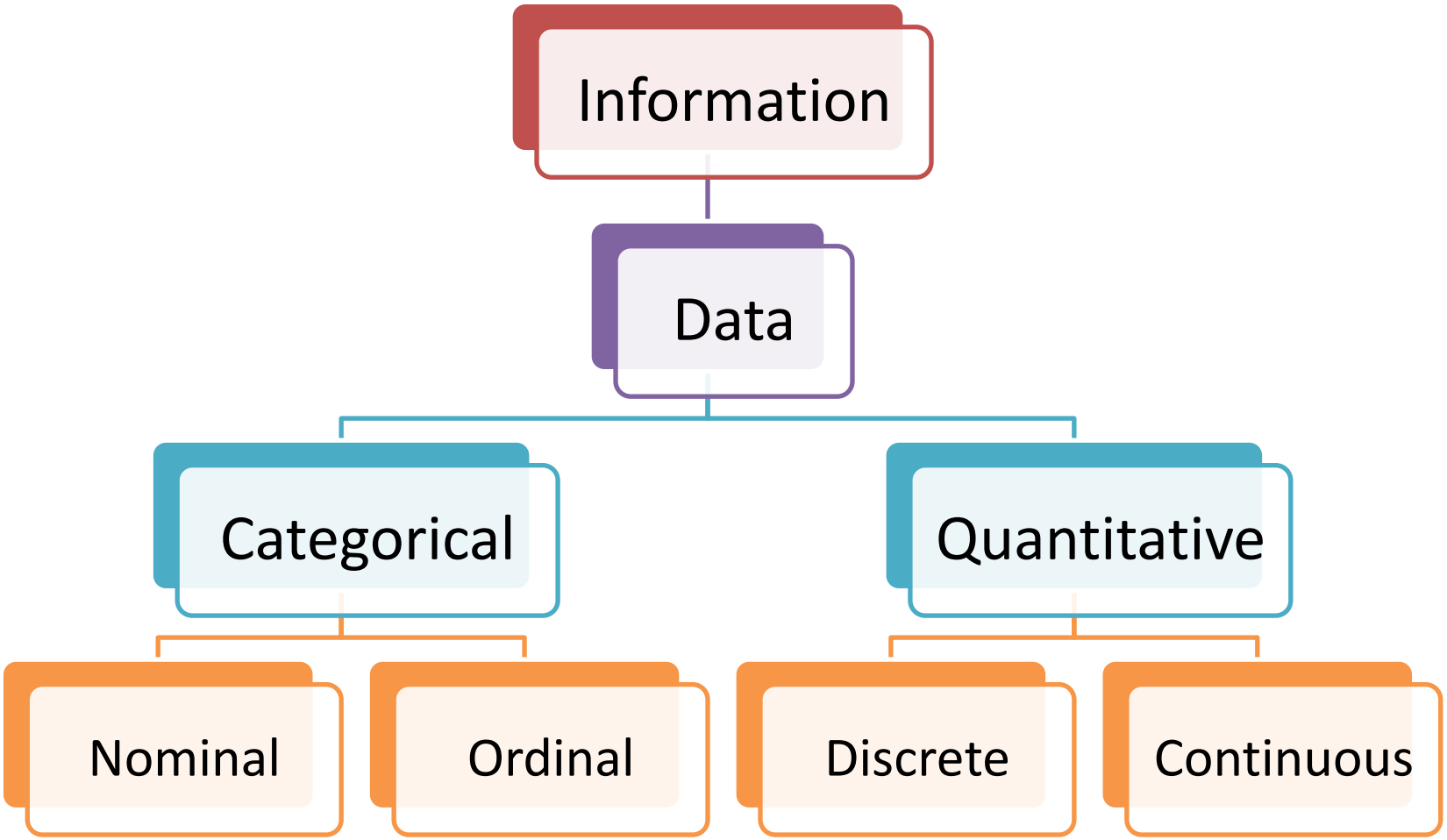
Li Xie

February 26, 2014

Outline

- Recap: Variable Typing, Descriptive Statistics
- Standard Error vs Standard Deviation
- Introduction to Hypothesis Test

Variable Types



Descriptive Statistics

Descriptive statistics are numbers that are used to summarize and describe data.

- Categorical variable: proportion
- Quantitative variable: mean, median, variance, standard deviation

$$\text{Median} = \frac{1}{2}(n+1)\text{th value, where } n \text{ is the number of data values in the sample}$$

Sample Mean

$$\bar{x} = \frac{\sum x}{n}$$

Sample Variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$$

Sample Standard Deviation

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

Descriptive Statistics - Exercise

Data points are as follows:

Race Track:	99	98	100	96	99	88	99
PET Scan:	10	52	29	69	67	92	87

Calculate & interpret the following statistics:

- Sample Median
- Sample Mean $\bar{x}$
- Sample Variance s^2
- Sample Standard Deviation s

Descriptive Statistics - Exercise

Race Track:	99	98	100	96	99	88	99
PET Scan:	10	52	29	69	67	92	87

What % of data points are above 95?

What % of data points are above 90?

What would the inference be if only were observed?

Possible Results Representation

Conditions	Mean (SD)	Median (Range)
Race Track		
PET Scan		

Conditions	≥ 95	<95
Race Track		
PET Scan		

Conditions	≥ 90	<90
Race Track		
PET Scan		

Standard Error (SE) vs SD (*s*)

- SE and SD are mathematically related but conceptually different

$$SE = \frac{s}{\sqrt{n}}$$

- SD tells us the distribution of individual data points around the mean WITHIN ONE SAMPLE, and SE informs us how precise our sample mean is as an estimate of the population mean, given the sample size and sample distribution--sample mean and sample variability (*s*).

❖*Remember SD for "**S**ample **D**escription"*

Hypothesis Test - Glossary

- http://www.stats.gla.ac.uk/steps/glossary/hypothesis_testing.html

Hypothesis Test - Logic

- Overall logic:

  “Probable impossibilities are to be preferred to improbable possibilities.” -- Aristotle
- By convention, most tests test a null hypothesis (H0) that states no difference, no association, etc.
- Mathematical assumptions are indispensible to all hypothesis tests and statistical models. These assumptions must be taken into consideration prior to and during study execution.

"To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of."

– R. A. Fisher

Hypothesis Test - Procedure

Empirical distribution → Test statistic → Theoretical distribution → Theoretical result → Empirical inference

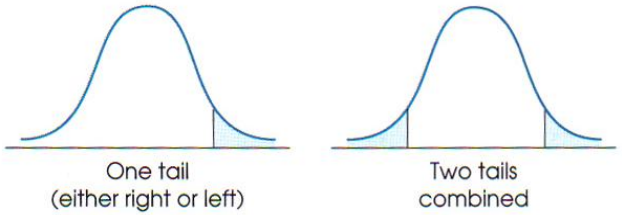
$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

$$\text{d.f.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}.$$

TABLE B.2 THE *t* DISTRIBUTION

Table entries are values of *t* corresponding to proportions in one tail or in two tails combined.



	PROPORTION IN TWO TAILS COMBINED					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250

Example: Two-Sample Unpaired Student's t-test

- Step-by-step test procedure:

http://psychology.ucdavis.edu/faculty_sites/sommerb/sommerdemo/stat_inf/tutorials/ttesthand.htm

Two-Sample Unpaired Student’s t-test - Exercise

Mann-Whitney Test* (if time allows)