Choice of Statistical Tests

Nominal Data

- χ^2 test of independence
  - check whether one variable influences another
  - organize one set of variable values into columns, the second into rows
  - (reformulated) question: is the distribution roughly the same in the different rows?
  - H_0: no influence

Pet Ownership

Owner	**Type of Pet**		
	dogs	cats	other
boys	39	55	6
girls	26	65	9

- no cell with expectation of zero's
- small 2×2 tables require $Yates$ correction

Sign Test

- can be used, e.g.,
  a. to test agreement in judgement
  b. to indicate improvement
- interpretation
  a. $+$ indicates agreement, $-$ disagreement
  b. $+$ indicates improvement, $-$ none
- question: is the breakdown of $+$'s and $-$'s a chance breakdown?
- relatively insensitive
- almost always applicable (even when original data in not normally distributed, and even asymmetrical)

Averages

To check for a difference in averages where the standard deviation is known, the most sensitive test is the z-test.

- how many standard errors separate the two sample averages?

$$z = \frac{m_1 - m_2}{\sigma / \sqrt{n_1 + n_2}}$$

- averages are always normally distributed (no need to check for normality
- H_0: no difference in averages
- interpret using standard-normal tables

RuG

Averages, σ unknown

To check for a difference in averages where the standard deviation is unknown, first ask whether are testing two different groups (unrelated samples), or two scores from one group (paired data).

t-test for unrelated samples

- t statistic like z, uses s instead of σ
- H_0: no difference in averages
- sample size is important
  - if $n \leq 15$, dist. must be normal
  - if $n \geq 40$, z is almost identical
- if t cannot be used, try **Mann-Whitney U test**
- with 3 or more groups, use **ANOVA**

RuG

Mann-Whitney U-Test

alternative to t-test (unpaired data)

- H_0 : samples from same population
- often applied to Likkert scale data
- **generalization** to several groups: **Kruskal-Wallis**

Averages, σ unknown

To check for a difference in averages where the standard deviation is unknown and two scores from one group (paired data).

t-test for paired data

- uses same t statistic as t-test for unrelated samples, uses s instead of σ
- H_0: no difference in averages
- sample size is important
  - if $n \leq 15$, dist. must be normal
  - if $n \geq 40$, z is almost identical
- if t cannot be used, try **Wilcoxon test**

Wilcoxon

Wilcoxon's Signed Rank Test

- also applicable to ordinal data
- fallback for paired t-test
- distribution should be roughly symmetric, not skewed
- if data asymmetric, try **sign test**

ANOVA

To check for a difference in averages where the standard deviation is unknown and there are three or more groups (unpaired data).

ANOVA

- roughly normal distribution per subgroup
- same variance in subgroups
  least sd $>$ one-half of greatest sd
- **independent** observations
  watch out for test-retest situations!
- if there is one than one dimension of comparison, then use **multiple ANOVA**
- if conditions on normality or variance are violated, consider **Kruskal-Wallis**

Correlation and Regression

To check for statistical dependence among numeric variable, investigate the CORRELATION. If one variable is postulated (perhaps) to explain, even partially another variable, perform a REGRESSION analysis

Correlation and Regression

- roughly normal distributions
- examine and illustrate dependence using SCATTERPLOTS
- r^2 characterizes how much of one variable's variance may be explained by another
- if conditions on normality are violated, consider **Spearman's Rank Correlation**

Correlation