Statistics

One-two sided test, Parametric and non-parametric test statistics: one group, two groups, and more than two groups samples

February 3, 2010

Jobayer Hossain, Ph.D. & Tim Bunnell, Ph.D.

Nemours Bioinformatics Core Facility

Class Objectives -- You will Learn:

1. The difference between one- and two-sided tests of hypotheses and when each is appropriate.
2. The difference between parametric and nonparametric tests and when each is most appropriate.
3. How to select some common parametric and non-parametric tests for quantitative and categorical variables involving:
   - One-group
   - Two groups,
   - More than two groups
4. How to do these tests with SPSS

One- and two-sided tests of hypotheses

- One-sided
  - One direction of effect (e.g. mean efficacy of treatment group is greater than the mean efficacy of placebo group)
  - Greater power to detect difference in expected direction
- Two-sided
  - Effect could be in either direction (e.g. mean efficacy of treatment group is not equal to the mean efficacy of placebo group)
  - More conservative

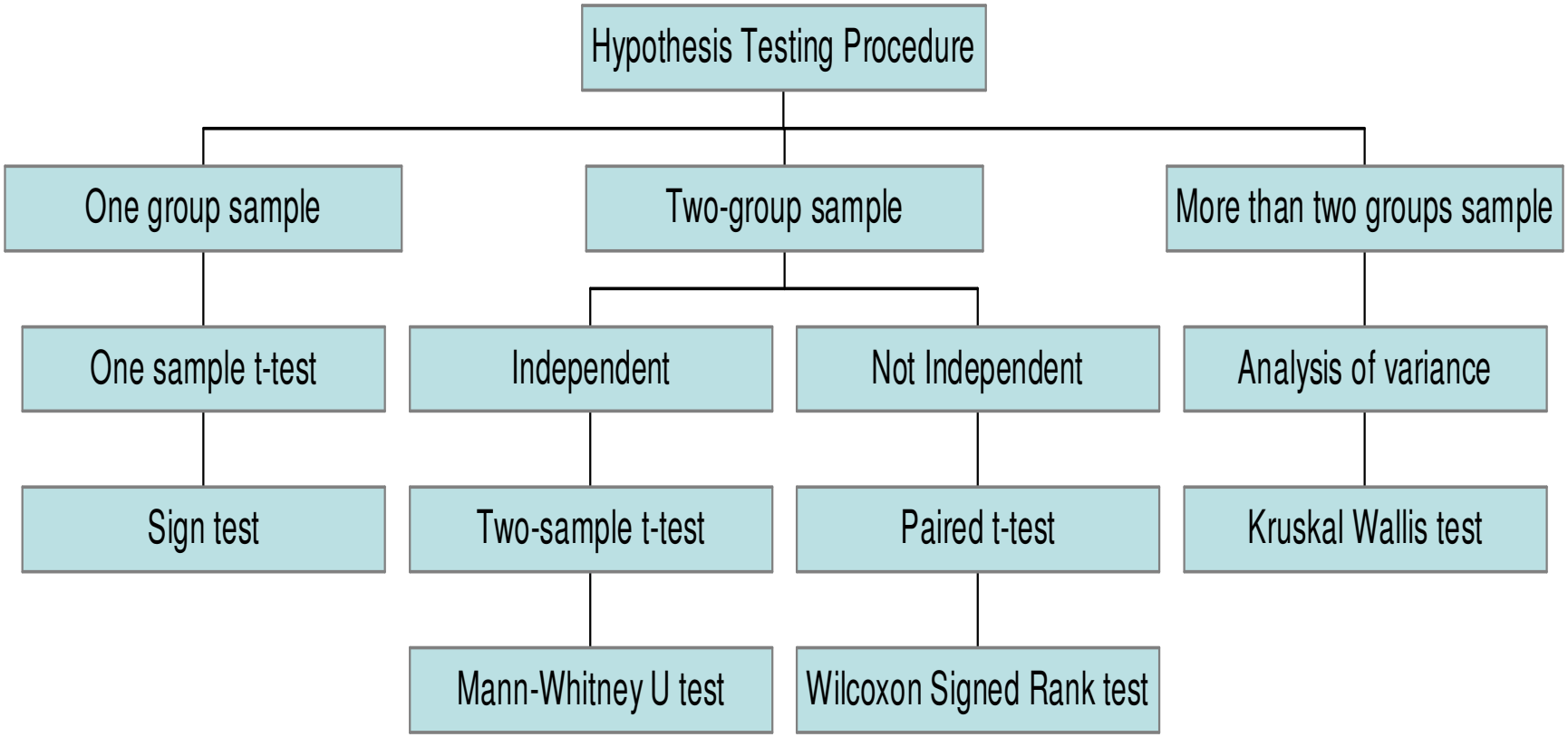
One sided and two sided test of hypothesis

	One group	Two groups
One sided	A single mean differs from a known value in a specific direction. e.g. mean > 0 or median > 0	Two means differ from one another in a specific direction. e.g., $mean_2 < mean_1$ $median_2 < median_1$
Two sided	A single mean differs from a known value in either direction. e.g., mean ≠ 0 median ≠ 0	Two means are not equal. E.g., $mean_1 \neq mean_2$ $median_1 \neq median_2$

Parametric & Nonparametric Tests

- Parametric Test
  - Make certain assumptions about population distribution or parameter of the population from which the sample is taken
  - E.g. A normal population distribution and equality of population variances among all groups being compared
- Non-parametric Test
  - Distribution free methods which do not rely on assumptions that the data are drawn from a given probability distribution.
- If data failed to meet assumptions non-parametric tests are preferred

Parametric & Nonparametric tests for Quantitative Variables



Single group sample - one sample t-test (Parametric)

- Test for value of a single mean
- E.g., to see if mean SBP of all AIDHC employees is 120 mm Hg
- Assumptions
  - Parent population is normal
  - Sample observations (subjects) are independent

Single group sample-one sample t-test (Parametric)

- Formula

  Consider a sample of size n from a normal population with mean μ and variance σ^2, then the following statistic is distributed as Student's t with (n-1) degrees of freedom.

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

  $\bar{x}$ is the sample mean and s is the sample standard deviation

One group sample - Sign Test (Nonparametric)

- Use:

  (1) Compares the median of a single group with a specified value (instead of single sample t-test).

- Hypothesis: H_0:Median = c

  H_a:Median $\neq$ c

- Test Statistic:

  We take the difference of observations from median (x_i - c). The number of positive or negative difference follows a Binomial distribution. For a large sample size, this distribution follows normal distribution.

One group sample: SPSS demonstration

- One-sample t-test (Parametric)
  - Analyze->Compare Means->One-Sample T-test. Then select a test variable and a test value (value for H_0) from this window and click ok.
- One-sample sign-test (Parametric)
  - Analyze->Non-parametric->Binomial Test. Then select a test variable, a cut point, and a value for test proportion (H_0) and then click ok.

Two-group (independent) samples - two-sample t-statistic (Parametric)

- Use
  - Test for equality of two means
- Assumptions
  - Parent populations are normal
  - Sample observations (subjects) are independent.

Two-group (independent) samples - two-sample t-statistic (Parametric)

- Formula (two groups)
  - **Case 1: Equal Population Standard Deviations:**
    - The following statistic is distributed as t distribution with (n1+n2 -2) d.f.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

The pooled standard deviation,

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

n1 and n2 are the sample sizes and S_1 and S_2 are the sample standard deviations of two groups.

Two-group (independent) samples - two-sample t-statistic (Parametric)

- Formula (two groups)
  - **Case 2: Unequal population standard deviations**
    - The following statistic follows t distribution.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

    - The d.f. of this statistic is,

$$v = \frac{\left(s_1^2 / n_1 + s_2^2 / n_2\right)^2}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$$

Two-group (independent) samples - two-sample t-statistic (Parametric) : SPSS demonstration

- Analyze->Compare Means-> Independent-Samples T test. Then select one or more test variables and a grouping variable.

Two-group (independent) samples- Wilcoxon Rank-Sum Test (Nonparametric)

- **Use:** Compares medians of two independent groups.
- Corresponds to t-test for two independent sample means
- **Test Statistic:**

  Consider two samples of sizes m and n. Suppose N=m+n. Compute the rank of all N observations. Then, the statistic,

  W_m= Sum of the ranks of all observations of the first variable (size m).

Two-group (independent) samples- Wilcoxon Rank-Sum Test (Nonparametric): SPSS demo

- Analyze-> Nonparametric tests->Two-Independent-Sample tests. Then select one or more test variables, a group variable, and the test type Mann-Whitney U.

Two-group (matched) samples - paired t-statistic (Parametric)

- Use: Compares equality of means of two matched or paired samples (e.g. pretest versus posttest)
- Assumptions:
  - Parent population is normal
  - Sample observations (subjects) are independent

Two-group (matched) samples - paired t-statistic (Parametric)

- Formula
  - The following statistic follows t distribution with n-1 d.f.

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Where, d is the difference of two matched samples and S_d is the standard deviation of the variable d.

Two-group (matched) samples - paired t-statistic (Parametric): SPSS demo

- Analyze -> Compare Means->Paired-Samples T test. Then select two dependent variables for variable1 (e.g. PLUC_pre) and variable2 (PLUC_post) and then click ok.

Two-group (matched) samples Wilcoxon Signed-Rank Test (Nonparametric)

- **USE:**
  - Compares medians of two paired samples.
- **Test Statistic**
  - Obtain differences of two variables, $D_i = X_{1i} - X_{2i}$
  - Take absolute value of differences, D_i
  - Assign ranks to absolute values (lower to higher), R_i
  - Sum up ranks for positive differences (T_+) and negative differences (T_-)
    - Test Statistic is smaller of T_- or T_+ (2-tailed)

Two-group (matched) samples Wilcoxon Signed-Rank Test (Nonparametric): SPSS demo

- Analyze -> Nonparametric tests -> Two-Related-Samples Test. Then select two dependent variables for variable1(e.g. PLUC_pre) and variable2 (PLUC_post), select Test type ‘Wilcoxon’ and click ok.

More than two independent samples: F statistic (Parametric)

- Use:
  - Compare means of more than two groups
  - Test the equality of two variances.

More than two independent samples: F statistic (Parametric)

- Let X and Y be two independent Chi-square variables with n_1 and n_2 d.f. respectively, then the following statistic follows a F distribution with n_1 and n_2 d.f.

$$F_{n_1,n_2} = \frac{X / n1}{Y / n2}$$

- Let, X and Y are two independent normal variables with sample sizes n_1 and n_2. Then the following statistic follows a F distribution with n_1 and n_2 d.f.

$$F_{n_1,n_2} = \frac{s_x^2}{s_y^2}$$

Where, s_x^2 and s_y^2 are sample variances of X and Y.

More than two independent samples: F statistic (Parametric)

- Hypotheses:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_n$

H_a: $\mu_1 \neq \mu_2 \neq \neq \ldots \neq \mu_n$

- ❑ Comparison will be done using analysis of variance (ANOVA) technique.
- ❑ ANOVA uses F statistic for this comparison.
- ❑ The ANOVA technique will be covered in another class session.

More than two independent samples: F statistic (Parametric) : SPSS demo

- Create a new variable: PLUC.chng = PLUC.post – PLUC.pre
  Transform-> Compute variable, then in this window, type the name of the new variable under the Target variable (e.g. PLUC.chng) and the expression PLUC.post – PLUC.pre under the numeric expression
- One-way ANOVA: Analyze -> Compare means -> One-way ANOVA Select the dependent variable (e.g. PLUC.chng), and factor (e.g. grp).

More than two groups: Nonparametric Kruskal-Wallis Test

- Compares median of three or more groups or (means of ranks of three or more groups)
- Rank the data ignoring group membership
- Perform the one way ANOVA of ranks instead of data itself
- SPSS Demo: Analyze -> Nonparametric Tests ->K Independent Samples, then select one or more test variables and a group variable and click ok.

Test statistics: Categorical variable

- Parametric
  - Proportion Tests
  - Chi-square Tests
- Nonparametric
  - McNemar Test
  - Kendall's Tau
- We will discuss only Proportion and Chi-square tests

Proportion Tests

- Use
  - Test for equality of two Proportions
    - E.g. proportions of subjects in two treatment groups who benefited from treatment.
  - Test for the value of a single proportion
    - E.g., to test if the proportion of smokers in a population is some specified value (less than 1)

Proportion Tests

- Formula
  - One Group:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

  - Two Groups:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

$$\text{where } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

Proportion Test: SPSS demo

- Analyze-> Nonparametric Tests-> Binomial. Then select a test variable (e.g.Ped) and a value for Test proportion and then click ok.

Chi-square Test

- USE
  - Testing the population variance $\sigma^2 = \sigma_0^2$.
  - Testing the goodness of fit.
  - Testing the independence/ association of attributes
- Assumptions
  - Sample observations should be independent.
  - Cell frequencies should be >= 5.
  - Total observed and expected frequencies are equal

Chi-square Test

- Formula: If x_i ($i=1,2,\ldots n$) are independent and normally distributed with mean μ and standard deviation σ^2, then,

$$\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \text{ is a } \chi^2 \text{ distribution with n d.f.}$$

- If we don't know μ, then we estimate it using a sample mean and then, $\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2$ is a χ^2 distribution with (n - 1) d.f.

Chi-square Test

- For a contingency table, we use the following chi- square test statistic,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}, \text{ distributed as } \chi^2 \text{ with (n - 1) d.f.}$$

$$O_i = \text{Observed Frequency}$$

$$E_i = \text{Expected Frequency}$$

Chi-square Test

	Male O(E)	Female O(E)	Total
Group 1	9 (10)	11 (10)	20
Group 2	8 (10)	12 (10)	20
Group 3	11 (10)	9(10)	20
	30	30	60

Chi-square Test– calculation of expected frequency

- To obtain the expected frequency for any cell, use:
- Corresponding (row total X column total) / grand total
- E.g: cell for group 1 and female, substituting: (30 X 20 / 60) = 10

Chi-square Test: SPSS demonstration

- Analyze->Descriptive statistics -> Crosstabs -> Pick row and column variables, select other options and click ok

Thank you