BIOS 6222: Biostatistics II

Instructors:
Qingzhao Yu
Don Mercante
Cruz Velasco

1

Outline

- Course Presentation
- Review of Basic Concepts
- Why Nonparametrics
- The sign test

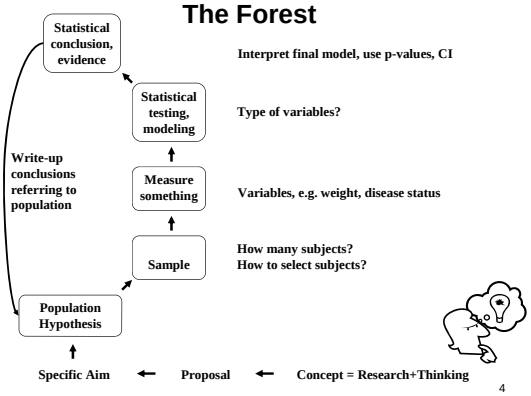
2

Course Presentation

- Contents Justification
- Evaluation
- Goals

3

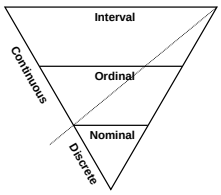
The Forest



4

Classification of Variables and Choice of Analysis

- By level of measurement: Nominal, Ordinal, Interval
- By gappiness: Discrete, Continuous
- By descriptive orientation: response, predictor



5

Selecting a statistical technique

Predictors	Response	Method
Any type	Continuous	Linear regression • Simple (one predictor) • Multiple (several predictors)
Any type	Dicotomous	Logistic regression
Any type	Discrete (counts)	Poisson regression
All nominal	Continuous	ANOVA • Two-way (two predictors) • Repeated measures (measurements over time)
All nominal	Nominal	Log-linear analysis

See:
http://www.graphpad.com/www/Book/Choose.htm
http://www.wadsworth.com/psychology_d/templates/student_resources/workshops/stat_workshp/choose_stat/chose_stat_01.html
For exact methods only see:
http://www.cytel.com/Products/StatXact/RoadMap_SX7_FINAL.pdf
CAUTION:
Nominal, Ordinal, Interval, and Ratio Typologies are Misleading
http://www.spss.com/research/wilkinson/Publications/Stevens.pdf

6

Review of Basic Statistics Concepts

Data description

- Qualitative data
  - Graph bars
  - Pie charts
- Quantitative data
  - Graphically
  - Histogram
  - Stem-and-leaf
  - Box plots
- Numerically
  - Central tendency (mode, median, mean, other)
  - Dispersion (range, variance/SD, CV, other)

7

Review of Basic Statistics Concepts

- Some important distributions
  - Discrete
    - Bernoulli, Binomial
  - Continous
    - Normal, Chi-square, Student's t, F

8

Review of Basic Statistics Concepts

- The Central Limit Theorem
  - There are several versions of the central limit theorem
  - The idea is that we can approximate probabilities with a very well known and easy to use distribution, provided we have a large sample size

•For large sample size:

- Binomial approaches normal
- Student's t approaches normal

9

Review of Basic Statistics Concepts

- Testing hypothesis
  - A single population mean

$H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$

$H_0 : \mu \leq \mu_0$ vs. $H_a : \mu > \mu_0$ $\qquad t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$H_0 : \mu \geq \mu_0$ vs. $H_a : \mu < \mu_0$

Can we conclude that average BMI of the population from which 14 subjects were randomaly selected is different to 35?

10

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
BMI	14	30.50	10.639	2.843

$$t = \frac{30.50 - 35}{10.639/\sqrt{14}} = -1.583$$

One-Sample Test

	Test Value = 35					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
BMI	-1.583	13	.138	-4.500	-10.64	1.64

Can we conclude that average BMI of the population from which 14 subjects were randomaly selected is different to 35?

11

- Difference of two independent means

- Large sample sizes or known variances

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

- Small sample sizes/unknown variances

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s_p \dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

12

- Difference of two independent means

Group Statistics

	V1	N	Mean	Std. Deviation	Std. Error Mean
V2	LOWFAT	100	9.310	4.6682	.4668
	REGULAR	100	7.400	4.0352	.4035

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
V2	Equal variances assumed	1.367	.244	3.095	198	.002	1.9100	.6170	.6932	3.1268
	Equal variances not assumed			3.095	193.940	.002	1.9100	.6170	.6930	3.1270

- Difference of two Dependent means

$$t = \frac{\bar{y}_d - 0}{s_d / \sqrt{n_d}}$$

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	New	76.0000	8	6.92820	2.44949
	Standard	71.6250	8	7.00892	2.47803

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	New & Standard	8	.971	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	New - Standard	4.37500	1.68502	.59574	2.96629	5.78371	7.344	7	.000

- Interval estimation

$$Estimator \pm (c_\alpha) SE(Estimator)$$

If the CI contains the value under the null hypothesis then the Null Hypothesis is not rejected

• Binomial test

A quality of life study found that 36 out of 85 patients made use of available spiritual councelor at the hospital. Can we conclude that the rate of use of spiritual counceling services is less than 40%

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

$H_0 : p \geq 0.4$ vs. $H_a : p < 0.4$

16

• Binomial test $z = \frac{0.4375 - 0.3}{\sqrt{0.3(0.7)/80}} =$

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
USE	80	.44	.499	0	1

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
USE	Group 1	1	35	.4	.3	.006[a]
	Group 2	0	45	.6		
	Total		80	1.0		

a. Based on Z Approximation.

Reject $H_0 : p \geq 0.4$ at $\alpha = 0.05$ level, since p-value=2(0.006)<0.05

17