Biostatistics for Health Care Researchers: A Short Course

Comparison of Means

Presented by:

Susan M. Perkins, Ph.D.

Division of Biostatistics

Indiana University School of Medicine

Objectives

- Understand common tests for comparing means
  - t-tests for Paired vs. Independent groups data
  - One-way ANOVA
  - Non-parametric methods
- Know when each test is most appropriate
- Understand simple sample size calculations for two-group problems

Choice of Test

- Study Design

  Paired or Independent groups ?
- Distribution (continuous data)

  Normal or skewed ?
- Number of Groups

  Two or more than two?

Examples

- Two **paired** samples:

  Are patients' systolic blood pressure decreased significantly after a new medication?

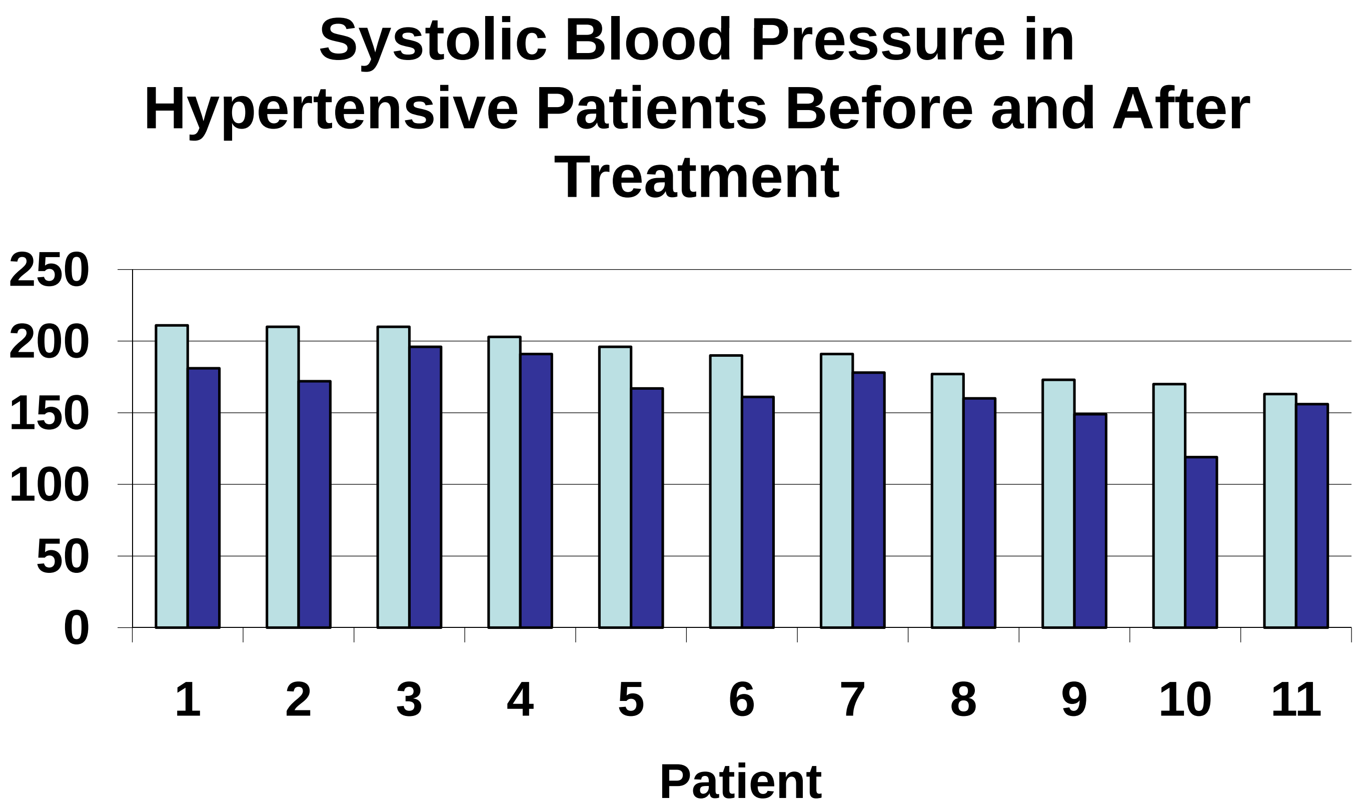
- Two **independent** samples:

  Is number of reproductive years related to bone mineral density?

- More than two independent samples:

  Do CD4+ counts differ between those with PTSD and/or depression in those living with HIV?

Paired Design



Student's t-Test: Paired Design

- Null Hypothesis: The intervention has no effect on systolic blood pressure

$$H_0: \ \mu_1 = \mu_2 \ (\ \mu_d = 0 \ )$$

- Alternative Hypothesis:

$$H_a: \ \mu_1 \neq \mu_2 \ (\mu_d \neq 0 \ )$$

- Assuming the systolic blood pressure are approximately normally distributed.
- Test Statistic: $t = \dfrac{\bar{d}}{s_d / \sqrt{n}}$

Data Example: Paired Design

Paired Differences d : 30, 38, 14, 12, 29, 29, 13, 17, 24, 51, 7

Mean Difference: $\bar{d}$ = 24.0

Standard Deviation: s_d = 13.1

Number of Subjects: n = 11

Test Statistic T:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{24}{13.1 / \sqrt{11}} = 6.1$$

Compare to a t-distribution with n -1 = 10 degrees of freedom

P-value < 0.001

Conclusion: the difference in SBP is statistically significant.

Student's t-Test for Independent Groups

- Used when observations in the two samples are on different subjects from two *independent population groups.*
  - Null Hypothesis: $\mu_1 = \mu_2$
  - Alternative Hypothesis: $\mu_1 \neq \mu_2$
  - Test Statistic *t*:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{1/n_1 + 1/n_2}}$$

  - Degrees of freedom are $n_1 + n_2 - 2$
  - Assuming X_1, X_2 are approximately normally distributed and the variances in two groups are close.

Pooled Variance

- Pooled Variance:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Data Example: Independent Groups

- Study of bone mineral density (BMD) in the hips of postmenopausal women. Interest focuses upon a potential relationship between the number of reproductive years and BMD.

Key Question: Does hip BMD in women with fewer than 30 reproductive years differ from those with 30 or more reproductive years?

$H_0: \mu_1 = \mu_2$ vs. $H_a: \mu_1 \neq \mu_2$

Data Example: Independent Groups

Mean (SD)	Women with <30 reproductive years (n=7)	Women with 30+ reproductive years (n=23)
Hip BMD (g/cm^2)	0.682 (0.133)	0.770 (0.136)

TS: $$t = \frac{\overline{X}_1 - \overline{X}_2}{s\sqrt{1/n_1 + 1/n_2}} = \frac{0.682 - 0.770}{.135\sqrt{1/23 + 1/7}} = -1.51$$

RR: reject if t < -2.05 or t > 2.05 (compare to a t-distribution with $n_1 + n_2 - 2 = 28$ d.f., two-sided)

P-value $= 0.14$

Conclusion: the two groups are not significantly different at alpha=0.05 level.

Re-Cap

- Differences between 2 types of t-test
  - One for paired groups ("cross-over designs")
  - One for independent groups

ASSUMPTIONS:

1) Observations from approximately normal distributions.
2) In the independent group design, the variances of the two groups should be close. If not, then similar sample sizes are needed.

Two Groups -- Non-normal Data

- Use a non-parametric test:

PAIRED	*INDEPENDENT*
Wilcoxon Signed Rank	Wilcoxon Rank-Sum
Rank differences	Rank values
Compare + and - ranks	Compare ranks in groups
	(Also called Mann-Whitney U)

Summary
(Two-Groups)

Study Design

Type of Data		Paired	Independent Group
	App. Normal	Paired t-test	Independent t-test
	Not Normal	Wilcoxon signed-ranks test	Wilcoxon rank-sum test

Sample Size Calculations

- Are there enough patients to answer the question?

Negative results can be due to three reasons:

- – No difference
- – Sample size is too small
- – Confounding variables

Choice of Sample Size

Depends on:

Question:	Difference of interest
Variability:	Pilot or historical data
Analysis:	Statistical methodology

Sample Size Calculation

- Minimal difference to be detected (Δ)
- Estimate of variance (s^2)
- Significance level (α)
- Power ($1-\beta$)

Independent:

$$n = \frac{2s^2\left(z_{\alpha/2}+z_{\beta}\right)^2}{\Delta^2}$$

per group

Paired:

$$n = \frac{s_d^2\left(z_{\alpha/2}+z_{\beta}\right)^2}{\Delta^2}$$

total pairs

Relationship to Sample Size

- Minimal difference to be detected (Δ) -

  Inverse relationship: Need large sample size to detect small difference

- Estimate of variance (s^2) -

  Direct relationship: Large variance means that large sample size is needed

- Significance level (α) – *Inverse relationship*

  e.g. Larger sample sizes needed for significance level of 1% than for a significance level of 5%

- Power ($1-\beta$) – *Direct relationship*

  e.g. Larger sample sizes are needed for 95% power than for 80% power.

Sample Size Calculation: An Example

Study:

Comparison of 2 oral hypoglycemic drugs

Measure:	Glycosylated hemoglobin
Minimal difference:	1%
Standard deviation:	s = 2% (two group design),
	s_d = 1% (paired design)
Significance level:	0.05
Power:	0.90

Sample Size Calculation: An Example

- Group Design

$$n = \frac{2(4)(1.96 + 1.28)^2}{1^2} = 84 \quad \text{per group}$$

- Paired Design

$$n = \frac{1(1.96 + 1.28)^2}{1^2} = 11 \qquad \text{total}$$

Sample Size Calculation for One Sample Problem

- Question: How many subjects are needed to detect a 1.5 unit change in processing speed from 10 units?
- Want 80% power and two sided α=0.05

$$n = \frac{s^2 (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

$$= \frac{9 \times (1.96 + 0.84)^2}{1.5^2} = 32$$

Analysis of Variance (ANOVA)

- Used for comparing means of three or more groups

- Tests if *at least one* group mean is different from the others.

Analysis of Variance (ANOVA)

H_0: $\mu_1 = \mu_2 = ... = \mu_m$

versus

H_A: at least one μ is not equal to one other

Only address "two-tailed" differences; ignore direction (higher, lower) of differences.

The ANOVA procedure uses observed variability in the data to quantify size of mean differences.

Assumptions for ANOVA

- *Normality* -- values of the dependent variable are assumed to be normally distributed within each group
- *Equal variances* - population variance is the same in each group
- *Independent observations* - observations are independent, i.e. not related to each other

(Don’t need equal variances if group sizes are close)

Analysis of Variance: An Example

Study of the relationship between PTSD and Depression on $CD4^+$ counts in people living with HIV.

PTSD measured by Impact of Life Events, Depression by CESD.

Subjects divided into 4 groups: control (low PTSD and CESD score), PTSD (high PTSD, low CESD), Depression (low PTSD, high CESD), and mixed (high on both)

Sledjeski et al. Incidence and Impact of Posttraumatic Stress Disorder and Comorbid Depression on Adherence to HAART and $CD4^+$ Counts in People Living with HIV. AIDS Patient Care and STDs; 19(11), 2005.

Analysis of Variance: An Example

Research question: Are mean $CD4^+$ counts different among the four groups?

	Mean	SD	N
Control:	399.64	241.40	22
PTSD:	535.27	274.93	11
Mixed:	469.00	334.91	24
Depression:	252.25	250.93	12

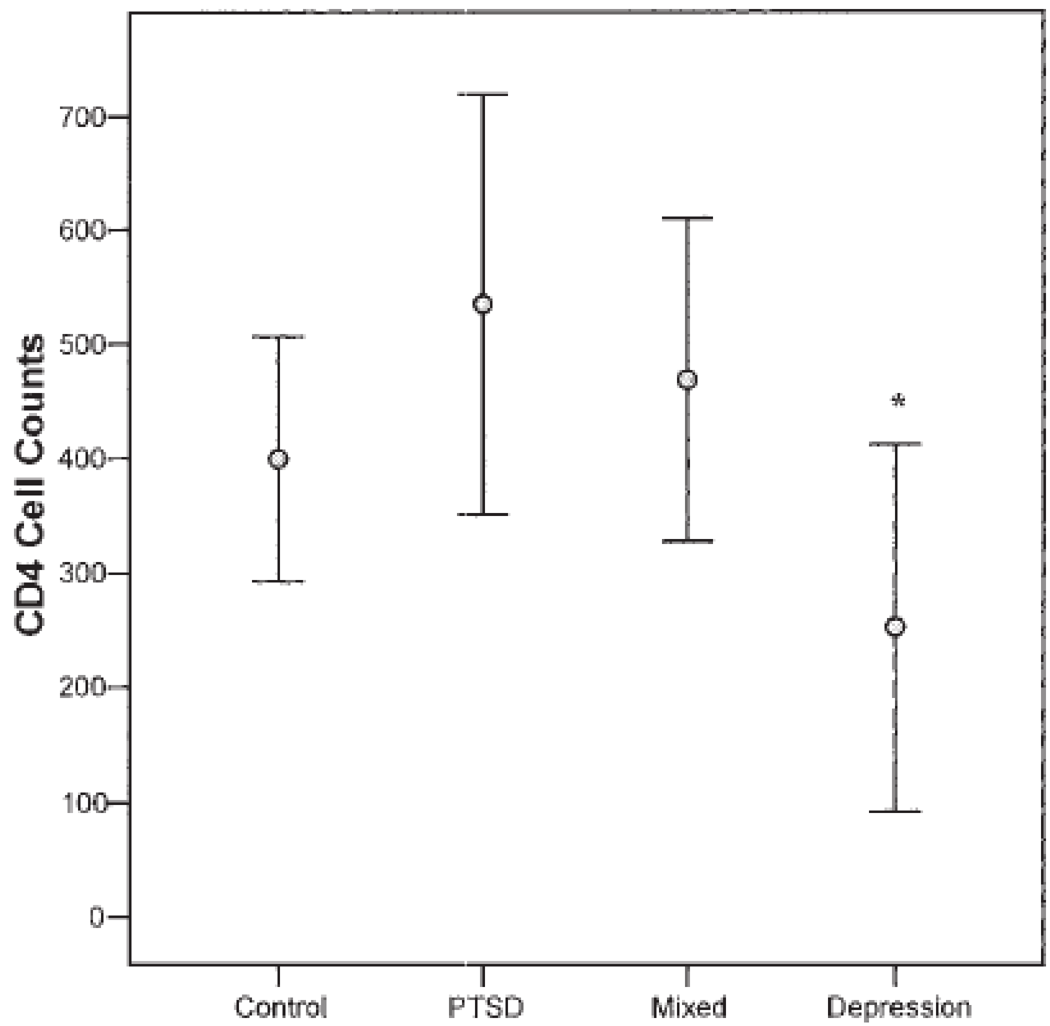


FIG. 1. Group differences in mean (+standard error [SE]) CD4 cell count levels. *Depression group is significantly different from all other groups. $p < 0.05$.

ANOVA Formulas

Source of Variation	Sums of Squares	Degrees of Freedom
Among Groups	$SS_A = \sum_{i,j} (\bar{X}_j - \bar{\bar{X}})^2$	*k - 1*
Error	$SS_E = \sum_{i,j} (X_{ij} - \bar{X}_j)^2$	*N - k*
Total	$SS_T = \sum_{i,j} (X_{ij} - \bar{\bar{X}})^2$	*N - 1*

X_{ij} : the observation of individual *i* in group *j*
$\bar{X}_j$: the mean of all observations in group *j*
$\bar{\bar{X}}$: the grand mean of all observations

Analysis of Variance: Test Statistic

$$F = \frac{SS_A / (k-1)}{SS_E / (N-k)}$$

We compare the F statistic above, to an F distribution with *k-1* and *N-k* degrees of freedom.

If F test is significant ($p < 0.05$), we conclude that there are differences amongst the groups.

Analysis of Variance: Example (cont'd)

$CD4^+$ data analysis:

$F = 4.404$, p-value = .007

We conclude that there is a difference in $CD4^+$ counts among the 4 comorbidity groups.

Which groups are different?

Multiple Comparisons

Suppose we have three groups: A, B, C

Could do multiple two group tests:

A vs. B   B vs. C   A vs. C ....

BUT: If α = 0.05 for each test, then the overall level is α = 0.14 for three tests. When there are no true differences, one would expect to erroneously decide that there is at least one difference 14% of the time.

Multiple Comparison Techniques

- Use these techniques to make multiple comparisons among pairs (all or some subset) of group means.

- Well-known tests:
  - Bonferroni correction
  - Scheffe
  - Student-Neuman-Keuls
  - Tukey’s HSD
  - Duncan Multiple Range

Multiple Comparisons: An Example

Using a Tukey HSD adjustment the CD4+ example:

- Depressed group lower than PTSD and Mixed
- PTSD and Mixed not different
- Control group not different than any group

Other Designs and Methods in ANOVA

- Repeated-measures design
  - Use the ANOVA counterpart to the paired t-test, R-M ANOVA.

- Non-parametric ANOVA
  - Kruskal-Wallis (for independent groups)
  - Friedman two-way ANOVA (for repeated measures)

Summary

- Most commonly-used methods for comparing means are:
  - t-test (for use with 2 groups)
  - ANOVA (for use with 3+ groups)

- Nonparametric methods are good alternatives when data are skewed (differ significantly from a normal distribution)