

UCLA STAT 13
**Introduction to Statistical Methods for
the Life and Health Sciences**

• **Instructor: Ivo Dinov,**
Asst. Prof. In Statistics and Neurology

• **Teaching Assistants: Janine Miller and Ming Zheng**
UCLA Statistics

University of California, Los Angeles, Winter 2003
http://www.stat.ucla.edu/~dinov/courses_students.html

STAT 13, UCLA, Ivo Dinov Slide 1

Chapter 10: Data on a Continuous Variable

- One-sample issues
- Two independent samples
- More than 2 samples
- Blocking, stratification and related samples

STAT 13, UCLA, Ivo Dinov Slide 2

Comparing two means for independent samples

Suppose we have 2 samples/means/distributions as follows: $\{\bar{x}_1, N(\mu_1, \sigma_1^2)\}$ and $\{\bar{x}_2, N(\mu_2, \sigma_2^2)\}$. We've seen before that to make inference about $\mu_1 - \mu_2$ we can use a **T-test for $H_0: \mu_1 - \mu_2 = 0$** with $t_o = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SE(\bar{x}_1 - \bar{x}_2)}$

And **CI($\mu_1 - \mu_2$)** = $\bar{x}_1 - \bar{x}_2 \pm t \times SE(\bar{x}_1 - \bar{x}_2)$

If the 2 samples are **independent** we use the SE formula
 $SE = \sqrt{s_1^2/n_1 + s_2^2/n_2}$ with $df = \text{Min}(n_1 - 1; n_2 - 1)$

This gives a conservative approach for hand calculation of an approximation to the what is known as the **Welch procedure**, which has a complicated exact formula.

Slide 38 STAT 13, UCLA, Ivo Dinov

**Means for independent samples –
equal or unequal variances?**

Pooled T-test is used for samples with assumed equal variances. Under data Normal assumptions and equal variances of $(\bar{x}_1 - \bar{x}_2) / SE(\bar{x}_1 - \bar{x}_2)$, where

$$SE = s_p \sqrt{1/n_1 + 1/n_2}; s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

is **exactly Student's t distributed** with $df = (n_1 + n_2 - 2)$

Here s_p is called the **pooled estimate of the variance**, since it pools info from the 2 samples to form a combined estimate of the single variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$.
The book **recommends** routine use of the Welch unequal variance method.

Slide 39 STAT 13, UCLA, Ivo Dinov

Comparing two means for independent samples

1. How sensitive is the two-sample *t*-test to non-Normality in the data? (The 2-sample T-tests and CI's are even more robust than the 1-sample tests, against non-Normality, particularly when the **shapes** of the 2 distributions are similar and $n_1 = n_2 = n$, even for small n , remember $df = n_1 + n_2 - 2$.)
3. Are there **nonparametric** alternatives to the *two-sample t-test*? (Wilcoxon rank-sum-test, Mann-Witney test, equivalent tests, same P-values.)
4. What **difference** is there between the **quantities tested and estimated** by the **two-sample t-procedures** and the **nonparametric** equivalent? (Non-parametric tests are based on ordering, not size, of the data and hence use median, not mean, for the average. The equality of 2 means is tested and CI($\mu_1 - \mu_2$)).

Slide 40 STAT 13, UCLA, Ivo Dinov

**We know how to analyze 1 & 2 sample data.
How about if we have than 2 samples –
One-way ANOVA, F-test**

One-way **ANOVA** refers to the situation of having one factor (or categorical variable) which defines group membership – e.g., comparing 4 reading methods, effects of different reading methods on reading comprehension, data: 50 – 13/14 y/o students tested.

Hypotheses for the one-way analysis-of-variance F-test
Null hypothesis: All of the underlying true means are identical.
Alternative: Differences exist between some of the true means.

Slide 41 STAT 13, UCLA, Ivo Dinov

Comparing 4 reading methods

Comparing 4 reading methods, effects of different reading methods on reading comprehension, data: 50 – 13/14 y/o students tested.

- Mapping: using diagrams to relate main points in text;
- Scanning: reading the intro and skimming for an overview before reading details;
- Mapping and Scanning;
- Neither.

Table below shows increases in test scores, of 4 groups of students taking similar exams twice, w/ & w/o using a reading technique.

Research question: Are the results better for students using mapping, scanning or both?

Slide 42 STAT 13, UCLA, Tom Dinger

TABLE 10.3.1 Increase in Reading Age

Both:	0.1	3.2	4.3	-0.5	1.9	3.3	2.5	3.6	0.4	2.3	-1.4	-0.7
Map Only:	-0.1	0.2	0.4	0.9	1.2	1.4	1.8	1.8	2.4	3.1		
Scan Only:	1.0	-0.5	1.0	0.6	0.6	1.0	1.0	-1.4	2.2	3.6	3.1	2.6
Neither:	1.0	3.3	1.4	-0.9	1.0	0.0	0.6					
	-0.3	-1.3	1.6	-0.4	-0.7	0.6	-1.8	-2.0	-0.7			

Figure 10.3.1 Increases in reading ages with individual 95% CIs.

Slide 43 STAT 13, UCLA, Tom Dinger

The F-test indicates that there's real evidence true differences exist it does not give indication of where the differences are or how large they are.

Source	DF	SS	MS	F	P
Grp	3	27.06	9.02	4.45	0.008
Error	46	93.35	2.03		
Total	49	120.41			

Figure 10.3.2 Minitab analysis of variance output for reading ages

Slide 44 STAT 13, UCLA, Tom Dinger

Computer output

Source	DF	SS	MS	F	P
Grp	3	27.06	9.02	4.45	0.008
Error	46	93.35	2.03		
Total	49	120.41			

Level	N	Mean	StDev
MapScan	22	1.459	1.544
MapOnly	12	1.233	1.441
ScanOnly	7	0.914	1.302
Neither	9	-0.556	1.135

Figure 10.3.2 Minitab analysis of variance output for reading ages

Slide 45 STAT 13, UCLA, Tom Dinger

Interpreting the P-value from the F-test

(The null hypothesis is that all underlying true means are identical.)

- A **large P-value** indicates that the differences seen between the sample means could be explained simply in terms of sampling variation.
- A **small P-value** indicates evidence that real differences exist between **at least some** of the true means, but gives *no indication* of where the differences are or how big they are.
- **To find out how big** any differences are we need confidence intervals.

Slide 46 STAT 13, UCLA, Tom Dinger

Form of a typical ANOVA table

TABLE 10.3.2 Typical Analysis-of-Variance Table for One-Way ANOVA

Source	Sum of squares	df	Mean sum of Squares ^a	F -statistic	P-value
Between	$\sum n_i(\bar{x}_i - \bar{x})^2$	$k - 1$	s_B^2	$f_0 = s_B^2 / s_W^2$	$pr(F \geq f_0)$
Within	$\sum (n_i - 1)s_i^2$	$n_{tot} - k$	s_W^2		
Total	$\sum \sum (x_{ij} - \bar{x})^2$	$n_{tot} - 1$			

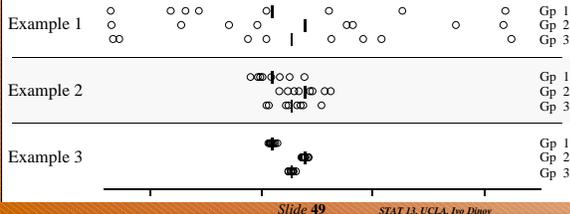
^aMean sum of squares = (sum of squares)/df

- The **F-test statistic**, f_0 , applies when we have independent samples each from k Normal populations, $N(\mu_i, \sigma)$, note same variance is assumed.

Slide 48 STAT 13, UCLA, Tom Dinger

Where did the F-statistics come from?

- Let's look at this example comparing groups. How do we obtain intuitive evidence against H_0 ? Far separated sample means + differences of sample means are large compared to their internal (within) variability! Which of the following examples indicate group diff's are "large"?



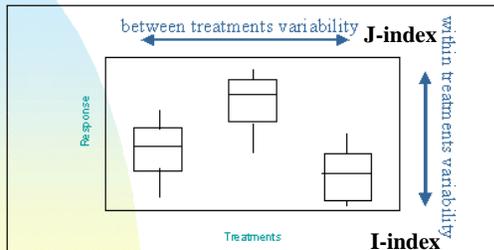
More about the F-test

- s_B^2 is a measure of variability of sample means, how far apart they are.
$$s_B^2 = \frac{\sum n_i (\bar{x}_i - \bar{x})^2}{k-1}$$
- s_W^2 reflects the avg. internal Variability within the samples.
$$s_W^2 = \frac{\sum (n_i - 1) s_i^2}{n_{tot} - k}$$
- The F-test statistic, f_0 , tests H_0 by comparing the variability of the sample means (numerator) with the variability within the samples (denominator).
- Evidence against H_0 is provided by values of f_0 which would be unusually large if H_0 was true.

Slide 50 STAT 13, UCLA, Jon Dinger

What are $x_i, x_{i..}, x_j$, etc.?

One-Way Anova (Sources of Variability)



What are $x_i, x_{i..}, x_j$, etc.?

Need Online reference

Apple juice sales (units per week)

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_A : at least 2 means differ

$$x_{i,j}, 1 < i <= n_j; 1 < j <= 3$$

City 1	City 2	City 3
Quantity	Quantity	Price
628	804	872
863	820	621
783	774	442
614	717	688
883	878	802
718	804	602
711	820	868
808	897	838
481	708	876
628	816	612
483	482	881
883	718	732
804	727	882
486	898	778
426	672	681
667	624	672
663	624	488
667	824	631
642	620	878
814	824	622

What are $x_i, x_{i..}, x_j$, etc.? Sum of Squares for treatments (cities)

$$SST = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2$$

$$\begin{aligned} SST &= 20(577.55 - 613.07)^2 \\ &\quad + 20(653.00 - 613.07)^2 \\ &\quad + 20(608.65 - 613.07)^2 \\ &= 57,512.23 \end{aligned}$$

Slide 54 STAT 13, UCLA, Jon Dinger

What are $x_i, x_{i..}, x_j$, etc.? Sum of squares for the Error

$$\text{Sum of Squares for Error: } SSE = \sum_{j=1}^k \left(\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \right)$$

$$\begin{aligned} SSE &= 19(10,774.44) + 19(7,238.61) + 19(8,669.47) \\ &= 506,967.88 \end{aligned}$$

Slide 55 STAT 13, UCLA, Jon Dinger

What are $x_i, x_{..}, x_{.j}$, etc.? F-test

Test Statistic: $F = \frac{MST}{MSE} = \frac{SST/(k-1)}{SSE/(n-k)}$

$$= \frac{57,512.23/(3-1)}{506,967.88/(60-3)} = 3.23$$

Rejection Region: $F > F_{\alpha; k-1, n-k} = F_{.05, 2, 57} = 3.15$

Conclusion: Reject H_0

Slide 56 STAT 13, UCLA, Jon Dinger

What are $x_i, x_{..}, x_{.j}$, etc.? One-Way Design ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Statistic
Treatments	k-1	SST	MST	MST/MSE
Error	n-k	SSE	MSE	
Total	n-1	SS(Total)		

Note: $MST = SST/(k-1)$
 $MSE = SSE/(n-k)$

$$s_B^2 = \frac{\sum n_i (\bar{x}_i - \bar{x}_{..})^2}{k-1}$$

$$s_W^2 = \frac{\sum (n_i - 1) s_i^2}{n_{tot} - k}$$

Slide 57 STAT 13, UCLA, Jon Dinger

F-test assumptions

1. Samples are **independent**, physically independent subjects, units, objects are being studied.
2. Sample Normal distributions, especially sensitive for small n_i , number of observations, $N(\mu_i, \sigma)$.
3. Standard deviations should be equal within all samples, $\sigma_1 = \sigma_2 = \sigma_3 = \dots = \sigma_k = \sigma$. ($1/2 \leq \sigma_k/\sigma_j \leq 2$)

How to check/validate these assumptions for your data?
For the reading-score improvement data:

- independence is clear since different groups of students are used.
- Dot-plots of group data show no evidence of non-Normality.
- Sample SD's are very similar, hence we assume population SD's are similar.

Slide 59 STAT 13, UCLA, Jon Dinger

Diagnostic plots for the reading-scores improvement data

(a) Original data

(b) Residual plot

Slide 60 STAT 13, UCLA, Jon Dinger

Diagnostic plots for the reading-scores improvement data

(b) Residual plot

(c) Normal prob. plot

Slide 61 STAT 13, UCLA, Jon Dinger

Review

1. What is an **one-way analysis of variance**? (compare means of several groups of independent samples.)
2. When do we use the one-way ANOVA F -test? ($(N(\mu_i, \sigma))^k$ samples).
3. What **null hypothesis** does it test? What is the **alternative hypothesis**? (all underlying true means are identical; at least 2 are different.)
4. Qualitatively, how does the F -test obtain evidence against H_0 ? (separation between sample means/intra-sample variability).
5. Qualitatively, what type of information is captured by the **numerator of the F -statistic**? What about the **denominator**? (variability-of-sample-means/variability-within-samples).

Slide 62 STAT 13, UCLA, Jon Dinger

Review

6. Qualitatively, what values of f_0 provide evidence against H_0 ? (unusually large f_0 if H_0 is true.)
7. What does a large P -value from the F -test tell us about differences between means? How about a small P -value? (diff's between sample means can be explained by sampling variation.)
8. What does a small P -value tell us about which means differ from one another? about how big the differences between means are? (nothing about which/size, only indicates real diff's exist, between at least some sample means.)
9. How do we obtain information about the sizes of differences between means? (need confidence intervals.)

Slide 63 STAT 13, UCLA, Jon Dinger

Review

10. What assumptions are made by the theory on which the F -test is based upon? How important is each of these assumptions in practice? (1. Sample independence – critical; 2. Normal data – robust, if sample-sizes are large; 3. Equal SD's – not too bad if $\sigma_{\max}/\sigma_{\min} < 2$.)
11. What new problem arises when we need to obtain and inspect a large set of confidence intervals? (all need to simultaneously catch, with 95% confidence, their true values, which requires increase of individual levels.)
12. Which is affected worst by departures from the equal-standard-deviations assumption, the F -test or the confidence intervals? Why? [CI, since CI(least-variable groups) = too wide & CI(most-variable-groups)=too narrow.]

Slide 64 STAT 13, UCLA, Jon Dinger

Chapter 10 Summary

STAT 13, UCLA, Jon Dinger

Slide 76

Always plot your data

Always plot your data before using formal tools of analysis (tests and confidence intervals).

- the quickest way to see what the data says
- often reveals interesting features that were not expected
- helps prevent inappropriate analyses and unfounded conclusions
- Plots also have a central role in checking up on the assumptions made by formal methods.

Slide 77

STAT 13, UCLA, Jon Dinger

All formal methods make assumptions

- If the assumptions are false, the results of the analysis may be meaningless.
- A method is **robust** against a specific departure from an assumption if it still behaves in the desired way despite that assumption being violated.
 - e.g. it gives “95% confidence intervals” that still cover the true value of θ for close to 95% of samples taken.
- A method is **sensitive** to departures from an assumption if even a small departure from the assumption causes it to stop behaving in the desired way.

Slide 78

STAT 13, UCLA, Jon Dinger

Assumptions cont.

- Many types of assumption are seldom, if ever, obeyed exactly so that methods which are sensitive to departures from such assumptions are of limited use in practical data analysis.
- You must check whether the data contradicts the assumptions to an extent where the tests and intervals no longer behave properly.
 - (Plots are a useful tool here.)

Slide 79

STAT 13, UCLA, Jon Dinger

Outliers

- If present, try and check back the original sources.
- Any observations which you know to be mistakes should be corrected or removed.
- If in doubt, do the analysis with and without the outliers to see if you come to the “same” conclusions.

Slide 80 STAT 13, UCLA, Tom Dinger

Nonparametric (distribution-free) methods

- less sensitive to outliers
- do not assume any particular distribution for the original observations
- do assume random samples from the populations of interest
- measure of center is the **median** rather than the mean
- tend to be somewhat **less effective** at detecting departures from a null hypothesis and tend to give **wider confidence intervals**

Slide 81 STAT 13, UCLA, Tom Dinger

Normal Theory Techniques

One sample methods

- Two-sided t -tests and t -intervals for a single mean are
 - quite robust against non-Normality
 - can be sensitive to presence of outliers in small to moderate-sized samples
- One-sided tests are reasonably sensitive to skewness.
- Normality can be checked
 - graphically using Normal quantile plots
 - formally, e.g. the Wilk-Shapiro test.

Slide 82 STAT 13, UCLA, Tom Dinger

Paired data

- We have to distinguish between **independent** and **related** samples because they require **different methods of analysis**.
- Paired data (Section 10.1.2) is an example of related data.
- With paired data, we analyze the differences
 - this converts the initial problem into a one-sample problem.
- The **sign test** and **Wilcoxon rank-sum** test are nonparametric **alternatives** to the **one-sample or paired t -test**.

Slide 83 STAT 13, UCLA, Tom Dinger

2-sample t -tests and intervals for differences between means $\mu_1 - \mu_2$

Assume

- statistically independent random samples from the two populations of interest
 - both samples come from Normal distributions
 - Pooled method also assumes that $\sigma_1 = \sigma_2$
Welch method (unpooled) does not
- Two-sample t -methods are
- remarkably robust against non-Normality
 - can be sensitive to the presence of outliers in small to moderate-sized samples
 - One-sided tests are reasonably sensitive to skewness.
- The **Wilcoxon** or **Mann-Whitney** test is a nonparametric **alternative** to the **two-sample t -test**.

Slide 84 STAT 13, UCLA, Tom Dinger

More than two samples and the F -test

- For testing whether **more than two means are different** we use the **F -test**.
- The method of comparing several means is referred to as a **one-way analysis of variance**.
- The formal null hypothesis (H_0) tested is that all k ($k \geq 2$) underlying population means μ_i are identical.
- The alternative hypothesis (H_1) is that differences exist between at least some of the μ_i 's.

Slide 85 STAT 13, UCLA, Tom Dinger

The F -test cont.

- The numerator of the F -statistic f_0 reflects how far apart the sample means are. The denominator reflects average variability within the samples
- Evidence against H_0 is provided by
 - sample means that are further apart than expected from the internal variability of the samples.
 - large values of the F -statistic.
- A small P -value demonstrates evidence that differences exist between some of the true means
 - To estimate the size of any differences we use confidence intervals

Slide 86

STAT 13, UCLA, Im. Dinger

Assumptions of the F -test cont.

- Assumptions of the F -test
 - independent samples;
 - Normality;
 - equal population standard deviations.
- The test
 - is robust to non-Normality
 - is reasonably robust to differences in the standard deviations when there are equal numbers in each sample, but not so robust if the sample sizes are unequal
 - can be used if the usual plots are satisfactory and the largest sample standard deviation is no larger than twice the smallest
 - is not robust to any dependence between the samples.

Slide 87

STAT 13, UCLA, Im. Dinger