Lecture 4
Statistical background review (III)

MCB 416A/516A
Statistical Bioinformatics and Genomic Analysis

Prof. Lingling An
Univ of Arizona

■ **Last time we reviewed:**

— Hypothesis testing
  ❖ One- two- sided tests
  ❖ One- two- sample test
  ❖ Pvalue and interpretation
— Relationship between hypothesis test and confidence interval
— Type I and II errors

Summary of hypothesis tests

Tests on Means with Variance Known

Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection	P-Value
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$|Z_0| > Z_{\alpha/2}$	$P = 2[1 - \Phi(|Z_0|)]$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$Z_0 = \dfrac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$	$Z_0 < -Z_\alpha$	$P = \Phi(Z_0)$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$Z_0 > Z_\alpha$	$P = 1 - \Phi(Z_0)$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$		$|Z_0| > Z_{\alpha/2}$	$P = 2[1 - \Phi(|Z_0|)]$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	$Z_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$Z_0 < -Z_\alpha$	$P = \Phi(Z_0)$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$		$Z_0 > Z_\alpha$	$P = 1 - \Phi(Z_0)$

One sample (rows 1–3)

two samples (rows 4–6)

Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection	P-Value
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$		$|t_0| > t_{\alpha/2, n-1}$	sum of the probability above t_0 and below $-t_0$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$t_0 = \dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$	$t_0 < -t_{\alpha, n-1}$	probability below t_0
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$		$t_0 > t_{\alpha, n-1}$	probability above t_0
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$	if $\sigma_1^2 = \sigma_2^2$ $t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $v = n_1 + n_2 - 2$	$|t_0| > t_{\alpha/2, v}$	sum of the probability above t_0 and below $-t_0$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$	if $\sigma_1^2 \neq \sigma_2^2$ $t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	$t_0 < -t_{\alpha, v}$	probability below t_0
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$v = \dfrac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_2 - 1}}$	$t_0 > t_{\alpha, v}$	probability above t_0

One sample

two samples

or min(n_1, n_2)

4

- We discussed method for comparing ***two*** conditions or treatments.

- But how about compare more than two conditions or levels of a factor?

- The *t*-test does not directly apply
  — Pairwise comparisons will inflate type I error

Why pairwise comparisons inflates type I error?

- Each time a hypothesis test is performed at significance level α, there is probability α of rejecting in error.

- Performing multiple tests increases the chances of rejecting in error *at least once*.

- For example:
  - — if you did 3 independent hypothesis tests at the α = 0.05
  - — If, in truth, H_0 were true for all three.
  - — The probability that *at least one* test rejects H_0 is 14.3%
    - ❖ P(at least one rejection) = 1-P(no rejections) = $1-.95^3 = 0.143$
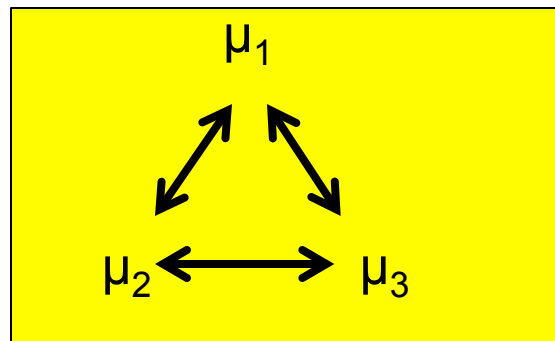
ANOVA - Analysis of Variance

- Extends independent-samples t test
- Compares the means of groups of independent observations.
  - —— Don't be fooled by the name.
  - —— ANOVA does not compare variances.
  - —— The name "analysis of variance" stems from a **partitioning** of the total variability in the response variable into components
  - —— Can compare more than two groups
- The ANOVA was developed by Fisher in the early 1920s, and initially applied to agricultural experiments.

ANOVA – Null and Alternative Hypotheses

Say the sample contains *K* independent groups

- ANOVA tests the null hypothesis

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

—— That is, "the group means are all equal"

- The alternative hypothesis is

$$H_1: \mu_i \neq \mu_j \text{ for some } i, j$$
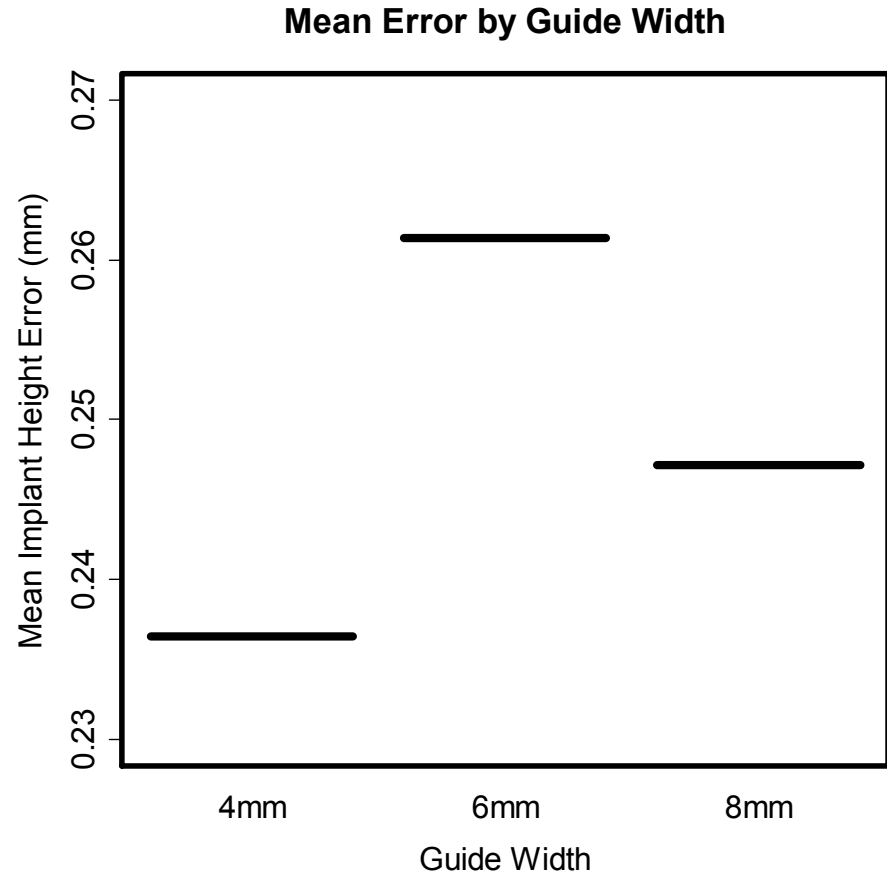
—— or, "the group means are *not* all equal"

Example: Accuracy of Implant Placement

Implants were placed in a manikin using placement guides of various widths.

15 implants were placed using each guide.

Error (discrepancies with a reference implant) was measured for each implant.
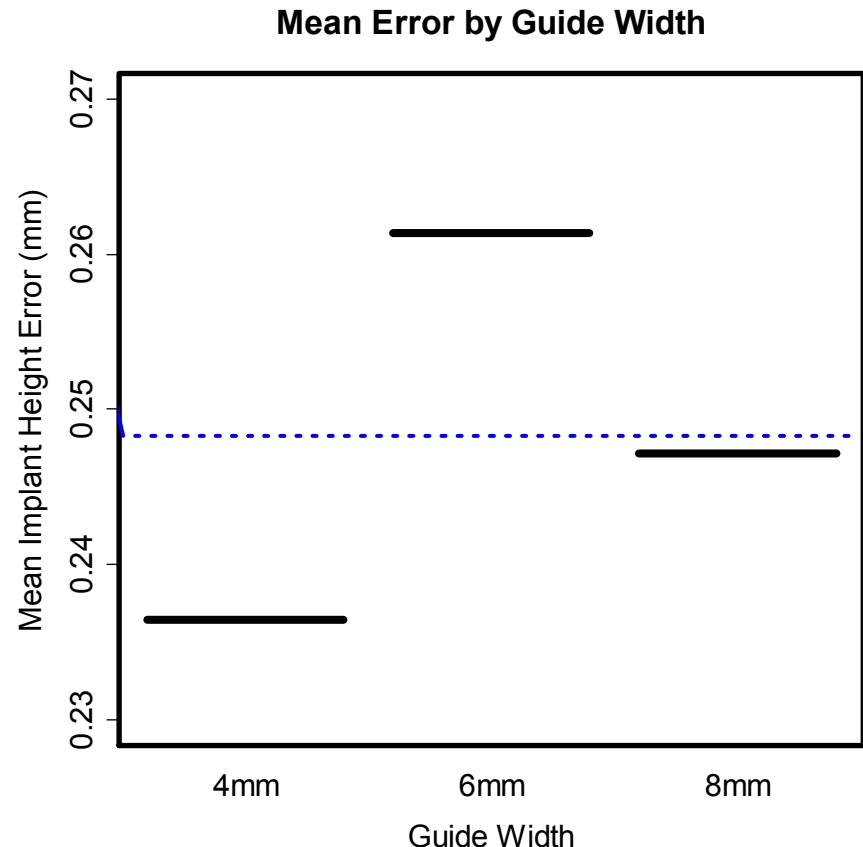
Mean Error by Guide Width

Example: Accuracy of Implant Placement -2

The overall mean of the entire sample was 0.248 mm.

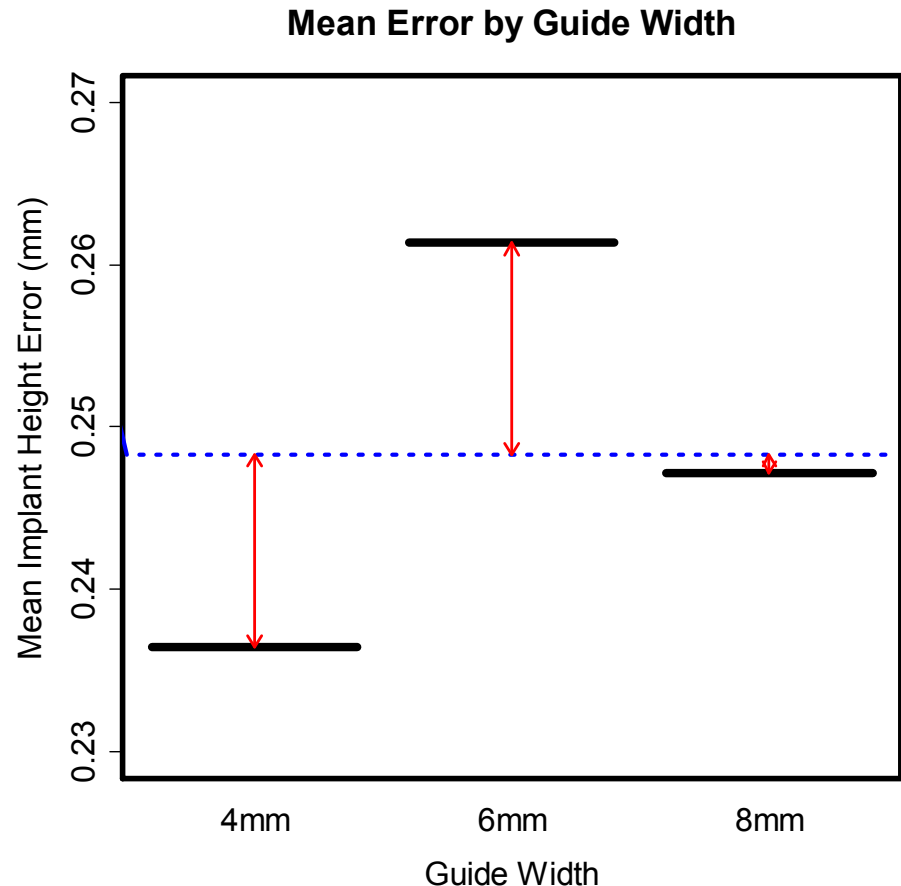This is called the "grand" mean, and is often denoted by $\overline{X}.$ or $\overline{\overline{X}}$

If H_0 were true then we'd expect the group means to be close to the grand mean.

Mean Error by Guide Width

Example: Accuracy of Implant Placement -3

The ANOVA test is based on the combined distances from $\bar{\bar{x}}$.

If the combined distances are large, that indicates we should reject H_0.

Mean Error by Guide Width

The Anova Statistic

To combine the differences from the grand mean we
- —— Square the differences
- —— Multiply by the numbers of observations in the groups
- —— Sum over the groups

$$SSB = 15\left(\overline{X}_{4mm} - \overline{\overline{X}}\right)^2 + 15\left(\overline{X}_{6mm} - \overline{\overline{X}}\right)^2 + 15\left(\overline{X}_{8mm} - \overline{\overline{X}}\right)^2$$

where the $\overline{X}_*$ are the group means.

"*SSB*" = **S**um of **S**quares **B**etween groups

Note: This looks a bit like a variance.

How big is big?

- For the Implant Accuracy Data, *SSB* = 0.0047

- Is that big enough to reject H_0?

- As with the *t* test, we compare the statistic to the variability of the individual observations.

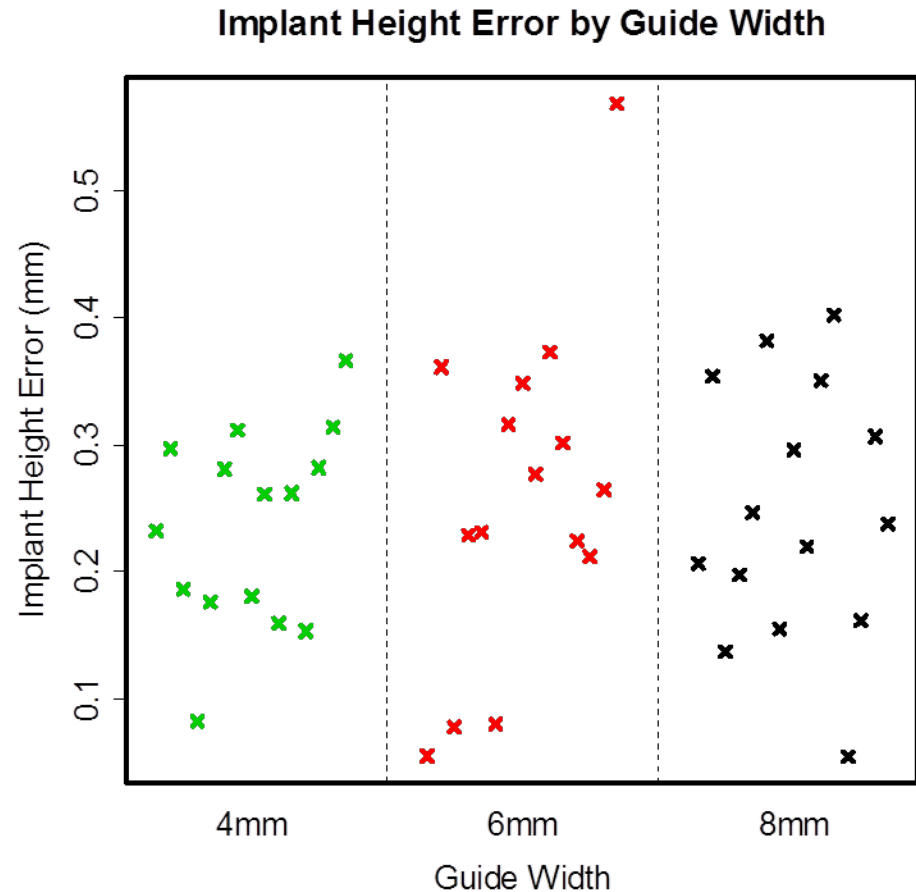- In ANOVA the variability is estimated by the Mean Square Error, or MSE

MSE: Mean Square Error

The Mean Square Error is a measure of the variability after the group effects have been taken into account.

$$MSE = \frac{1}{N-K}\sum_{j}\sum_{i}\left(x_{ij} - \overline{X}_j\right)^2$$

where x_{ij} is the ith observation in the jth group.

K = number of groups
N = total number of obs

Implant Height Error by Guide Width

MSE: Mean Square Error -2

The Mean Square Error is a measure of the variability after the group effects have been taken into account.

$$MSE = \frac{1}{N-K} \sum_{j} \sum_{i} \left(x_{ij} - \overline{X}_j \right)^2$$

where x_{ij} is the ith observation in the jth group.

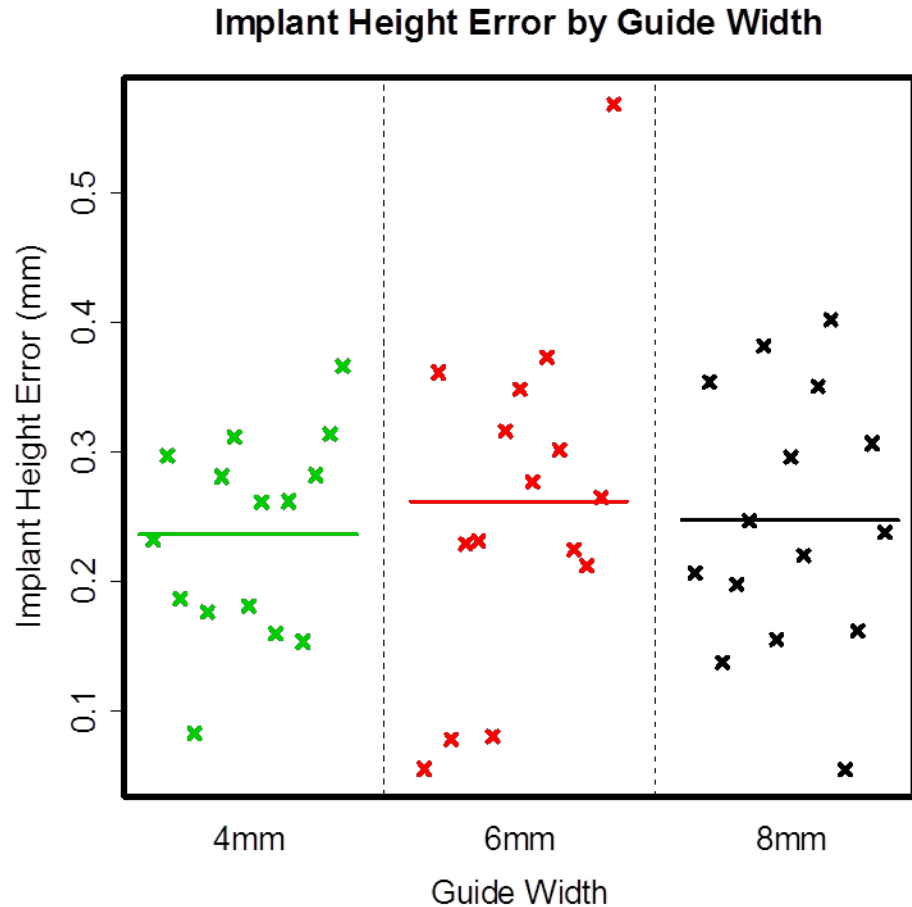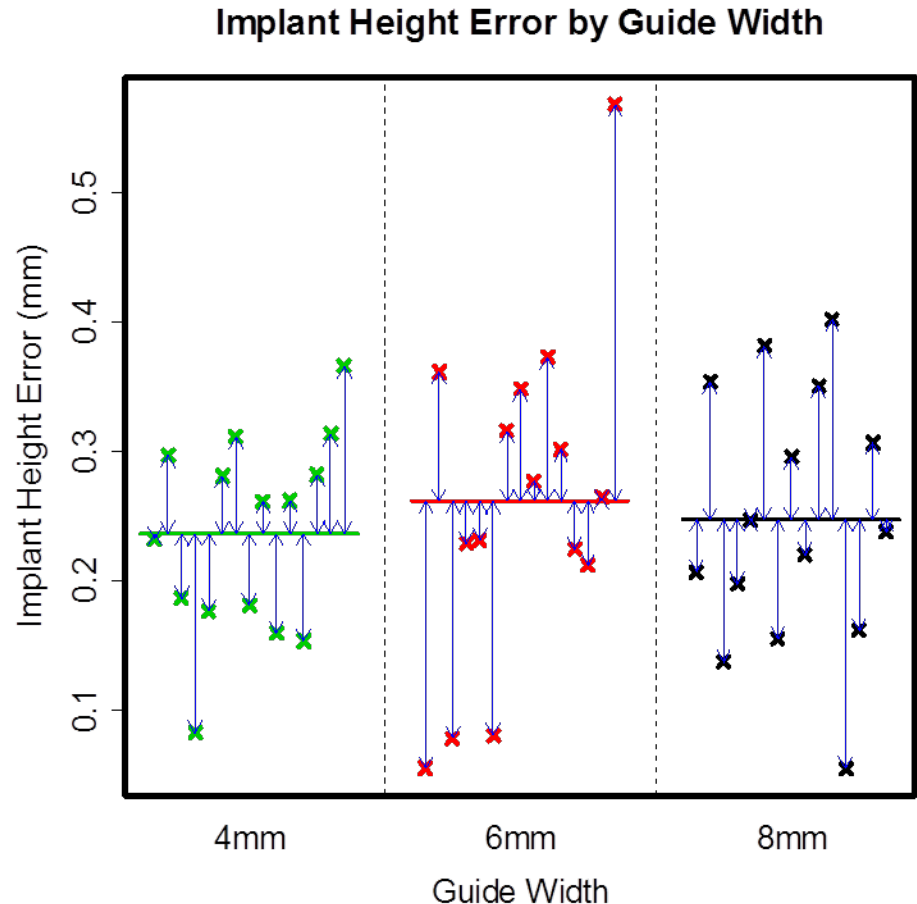

Implant Height Error by Guide Width

MSE: Mean Square Error -3

The Mean Square Error
is a measure of the
variability after the group
effects have been taken
into account.

$$MSE = \frac{1}{N-K}\sum_{j}\sum_{i}\left(x_{ij} - \bar{X}_j\right)^2$$

Note that the variation of
the means seems quite
small compared to the
variance of observations
within groups

Implant Height Error by Guide Width

Notes on *MSE*

- If there are only two groups, the *MSE* is equal to the pooled estimate of variance used in the equal-variance *t* test.

- ANOVA assumes that all the group variances are equal.

- Other options should be considered if group variances differ by a factor of 2 or more.

ANOVA F Test

- The ANOVA F test is based on the *F* statistic

$$F = \frac{SSB/(K-1)}{MSE}$$

  where *K* is the number of groups.

- Under H_0 the *F* statistic has an "F" distribution, with K-1 and N-K degrees of freedom (N is the total number of observations)
  - —— In this case N=45

Implant Data: F test p-value

To get a p-value we compare our *F* statistic to an F(2, 42) distribution.

F(2,42) distribution

F

Implant Data: F test p-value -2

To get a p-value we compare our *F* statistic to an F(2, 42) distribution.

In our example

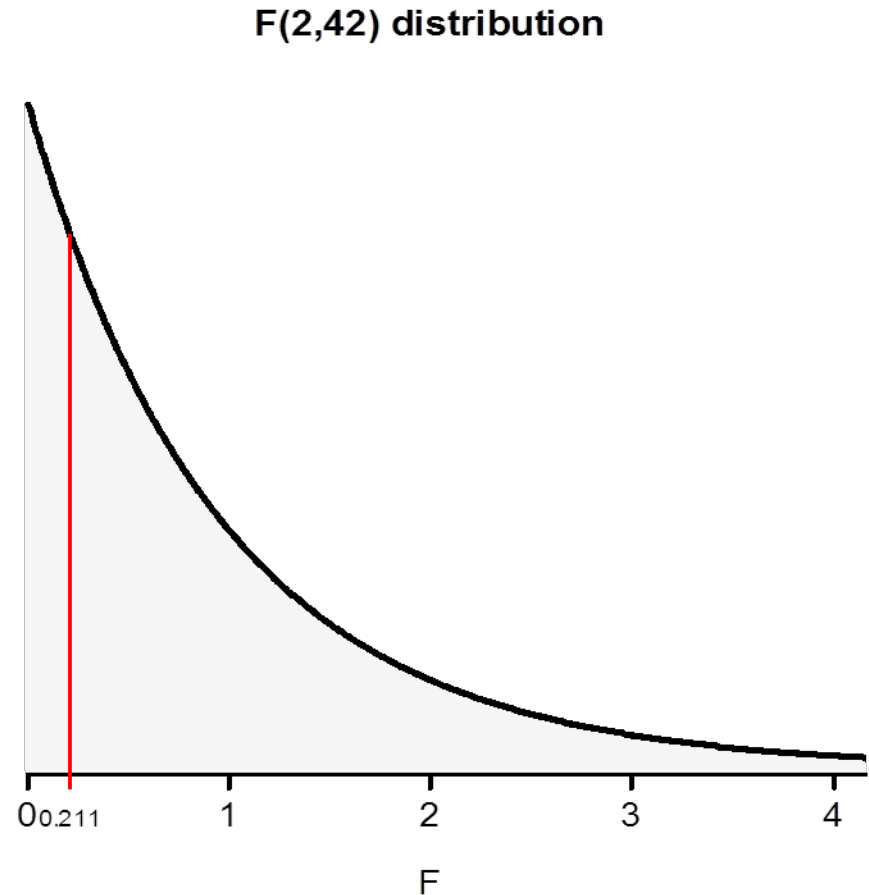$$F = \frac{.0047/2}{.0467/42} = .211$$

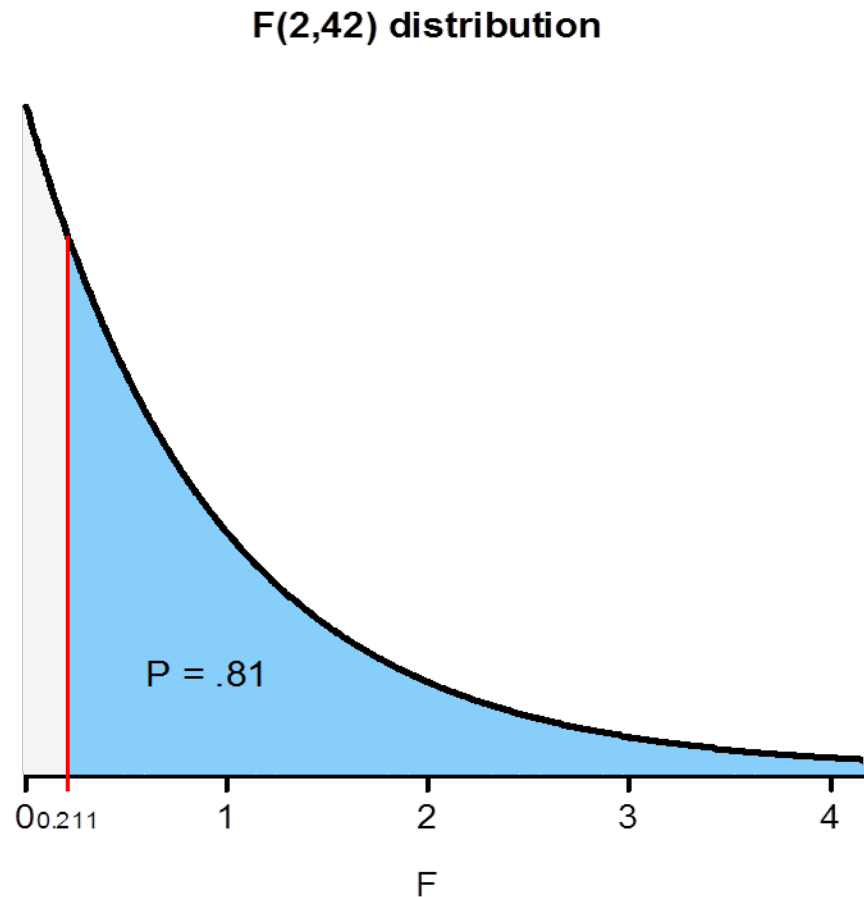**F(2,42) distribution**

Implant Data: F test p-value -3

To get a p-value we compare our *F* statistic to an F(2, 42) distribution.

In our example

$$F = \frac{.0047/2}{.0467/42} = .211$$

The p-value is

$$P\big(F(2,42) > .211\big) = 0.81$$

F(2,42) distribution

P = .81

0 0.211   1   2   3   4

F

ANOVA Table

Results are often displayed using an ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	F	P-value
Between Groups	.005	2	.002	.211	.811
Within Groups	.466	42	.011		
Total	.470	44			

Pop Quiz!:   Where are the following quantities presented in this table?

Sum of Squares Between (*SSB*)   Mean Square Error (*MSE*)   F Statistic   p value

ANOVA Table -2

Results are often displayed using an ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	F	P-value
Between Groups	.005	2	.002	.211	.811
Within Groups	.466	42	.011		
Total	.470	44			

Sum of Squares Between (*SSB*)

Mean Square Error (*MSE*)

F Statistic

p value

fail to reject at .811

Post Hoc Tests

Mean Serum Cholesterol by Perio. Class.

Women between 40 and 60 years of age

Source of Variation	Sum of Squares	df	Mean Square	F	P-value
Between Groups	33383	3	11128	5.1	**.002**
Within Groups	4417119	2007	2201		
Total	4450502	2010			
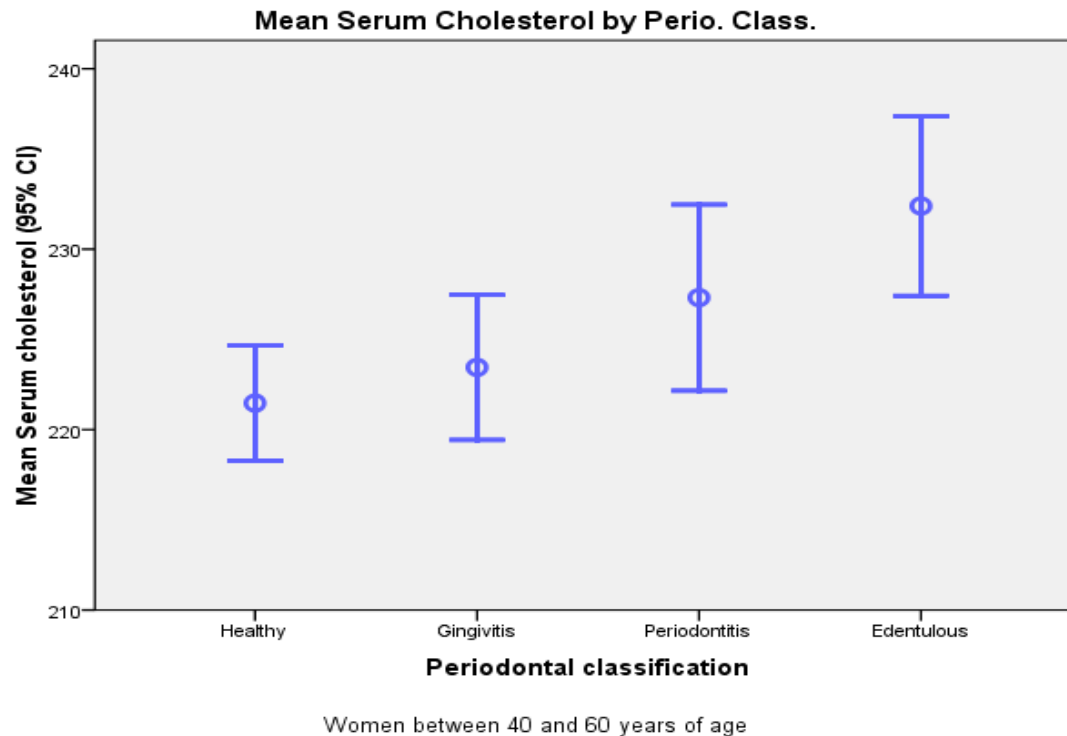
Compare cholesterol between periodontal groups. 40-60 yrs old women.

The ANOVA shows good evidence (p = 0.002) that the means are not all the same.

Which means are different?

Can directly compare the subgroups using "post hoc" tests.

Least Significant Difference test

	N	Mean	Std. Deviation
Healthy	802	221.5	46. 2
Gingivitis	490	223.5	45.3
Periodontitis	347	227.3	48.9
Edentulous	372	232.4	48. 8

	Sum of Squares	df	Mean Square	F	pvalue
Between Groups	33383	3	11128	5.1	**.002**
Within Groups	4417119	2007	2201		
Total	4450502	2010			

The most simple post hoc test is called the *Least Significant Difference* Test.

The computation is very similar to the equal-variance *t* test.

Compute an equal-variance *t* test, but replace the pooled variance (s^2) with the *MSE.*

Least Significant Difference Test: Examples

	N	Mean	Std. Deviation
Healthy	802	221.5	46. 2
Gingivitis	490	223.5	45.3
Periodontitis	347	227.3	48.9
Edentulous	372	232.4	48. 8

	Sum of Squares	df	Mean Square	F	pvalue
Between Groups	33383	3	11128	5.1	.002
Within Groups	4417119	2007	2201		
Total	4450502	2010			

Compare Healthy group to Periodontitis group:

$$T = \frac{221.5 - 227.3}{\sqrt{2201(1/802 + 1/347)}} = -1.92$$

$$p = 2 \cdot P(t_{1147} > 1.92) = 0.055$$
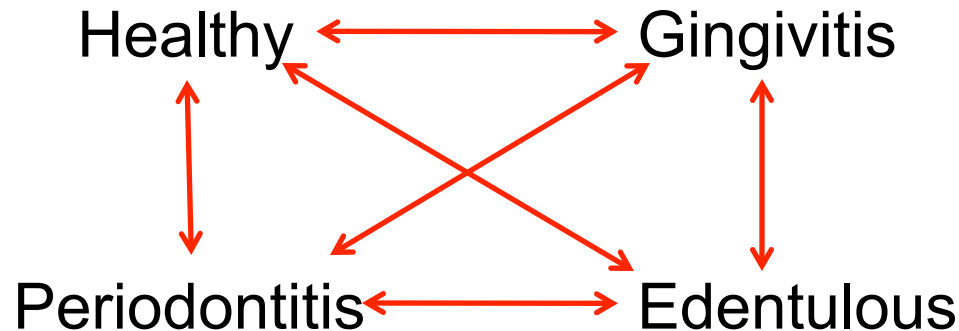
Compare Gingivitis group to Periodontitis group:

$$T = \frac{223.5 - 227.3}{\sqrt{2201(1/490 + 1/347)}} = -1.15$$

$$p = 2 \cdot P(t_{835} > 1.15) = 0.25$$

Post Hoc Tests: Multiple Comparisons

- Post-hoc testing usually involves multiple comparisons.

- For example, if the data contain 4 groups, then 6 different pairwise comparisons can be made

Post Hoc Tests: Multiple Comparisons -2

- Each time a hypothesis test is performed at significance level α, there is probability α of rejecting in error.

- Performing multiple tests increases the chances of rejecting in error *at least once*.

- For example:
  - if you did 6 independent hypothesis tests at the $\alpha = 0.05$
  - If, in truth, H_0 were true for all six.
  - The probability that *at least one* test rejects H_0 is 26%
    - ❖ P(at least one rejection) = 1-P(no rejections) = $1-.95^6 = .26$

Bonferroni Correction for Multiple Comparisons

■ The Bonferroni correction is a simple way to adjust for the multiple comparisons.

very conservative only good for small numbers

Bonferroni Correction

•Perform each test at significance level α.

•Multiply each p-value by the number of tests performed.

•The overall significance level (chance of any of the tests rejecting in error) will be less than α.

Example: Cholesterol Data post-hoc comparisons

Group 1	Group 2	Mean Difference (Group 1 - Group 2)	Least Significant Difference p-value	Bonferroni p-value	
Healthy	Gingivitis	-2.0	.46	1.0	max = 1
Healthy	Periodontitis	-5.8	.055	.330	
Healthy	Edentulous	-10.9	.00021	.00126	
Gingivitis	Periodontitis	-3.9	.25	1.0	
Gingivitis	Edentulous	-8.9	.0056	.0336	
Periodontitis	Edentulous	-5.1	.147	.88	

Conclusion: The Edentulous group is significantly different than the Healthy group and the Gingivitis group ($p < 0.05$), after adjustment for multiple comparisons