

1. An investigator tests a few treatments for acne. The response is percentage improvement. Values range from 46 to 72. There are no outliers or strong skewness. An ANOVA table is here.

Source	Sum of Squares	DF	Mean Square	F	p-value
Between	2133.66	2	1066.83	262.12	0.0000
Within	130.30	32	4.07		
Total	2263.96	34			

Circle TRUE or FALSE. If the statement is FALSE, briefly explain why or correct it.

- (a) TRUE or FALSE: There are two different treatments in the study.

Solution: False. There are three treatments (2 degrees of freedom between).

- (b) TRUE or FALSE: There were 35 subjects in the study.

Solution: True.

- (c) TRUE or FALSE: The ANOVA table summarizes a two-way analysis of variance.

Solution: False. This is a one-way ANOVA. There is only a single explanatory categorical variable.

- (d) TRUE or FALSE: The area to the right of 262.12 under an F distribution with 34 degrees of freedom is about 0.

Solution: False. F distributions have both numerator and denominator degrees of freedom. The area to the right of 262.12 under an F distribution with 2 and 32 degrees of freedom is about 0.

- (e) TRUE or FALSE: Chance alone can explain the differences in response to the treatments.

Solution: False. The tiny p-value indicates that the difference in sample means cannot be explained by chance variation.

2. In a study on the effects of carbon monoxide exposure on patients with coronary artery disease, men were selected from three different medical centers. Twenty-one men came from Johns Hopkins Medical Center, sixteen came from Los Amigos Medical Center, and twenty-three came from St. Louis University Medical Center. The response variable is the one second forced expiratory volume (FEV) in liters prior to treatment. Side-by-side boxplots indicate fairly symmetric distributions with similarly sized spread in each sample. The three sample means are 2.63, 3.03, and 2.88, respectively. Is there evidence that the three populations of subjects have different mean FEV values? A partial ANOVA table is below.

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-statistic	P-value
Between	1.58				0.052
Within	14.48				
Total					

- (a) Complete the ANOVA table.

Solution:

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-statistic	P-value
Between	1.58	2	0.79	3.11	0.052
Within	14.48	57	0.254		
Total					

(b) What is the pooled estimate of the common standard deviation?

Solution:  $s_{\text{pooled}} = \sqrt{0.254} = 0.504$ .

(c) Circle the numbers of all correct statements.

- (1) The three population mean FEVs are all equal because the p-value is greater than 0.05.
- (2) The null hypothesis that the population means are all equal is not rejected at the 5% level.
- (3) There is only mild evidence that the population mean FEVs at the three centers are unequal.
- (4) There is strong evidence that the population mean FEV at Johns Hopkins Medical Center is smaller than the population mean FEVs at the other two medical centers.

Solution:

- (1) False. We never conclude that the null hypothesis is true.
- (2) True. The p-value is greater than 0.05.
- (3) True. A p-value near 0.05 is an indicator of mild evidence against the null hypothesis, but this sort of interpretation is rather subjective.
- (4) False. The global  $F$  test does not show strong evidence of any differences.

(d) Suppose that you wanted to use the Bonferroni procedure to find simultaneous 95% confidence intervals for the three pairwise differences in means. Each confidence interval will be of the form

$$\bar{y}_i - \bar{y}_j \pm t_{\text{crit}} \times s_p \times \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

(1) Find a numerical value for  $s_p$ .

Solution: If you missed it the first time,  $s_{\text{pooled}} = \sqrt{0.254} = 0.504$ .

(2) Fill in the blanks. The value  $t_{\text{crit}}$  in the expression for the confidence interval is the \_\_\_\_\_ quantile of a  $t$  distribution with \_\_\_\_\_ degrees of freedom.

Solution: The value  $t_{\text{crit}}$  in the expression for the confidence interval is the  $1 - 0.025/3 = 0.992$  quantile of a  $t$  distribution with  $57$  degrees of freedom.

A quantile is the location that corresponds to a given area to the left under a curve. The usual 95% confidence interval uses a multiplier that is the 0.975 quantile of a distribution. The Bonferroni correction requires a decrease in  $\alpha$  by a factor of 3 when there are three tests.