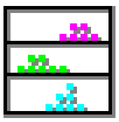
CHAPTER 12



Visualizing Analysis of Variance

CONCEPTS

- Analysis of Variance (ANOVA), Variation within Groups, Variation between Groups, F Ratio (F Statistic), ANOVA Table, Critical Value, Type I Error, Type II Error, Power, Equal Variances Assumption, Normality Assumption

OBJECTIVES

- Be familiar with situations in which one-factor ANOVA is applicable

- Understand how much difference must exist between group means to be detected using an F test

- Appreciate the role of sample size in determining power

- Know the ANOVA assumptions and recognize the effects of violating each of them

Overview of Concepts

When several sample means are compared, the question arises whether the group means are truly different or whether we are merely seeing sample variation. To compare c sample means, we rely on **analysis of variance** (or **ANOVA** for short) to test the hypotheses:

H_0: $\mu_1 = \mu_2 = . . . = \mu_c$ (The true group means are the same.)
H_1: Not all μ_i are equal (The true group means are not all the same.)

Suppose we have c groups with sample sizes n_1 , n_2 , . . . , n_c. Comparison of dot plots for each group may reveal differences in means and/or variances. We can do the ANOVA calculations in a spreadsheet or use specialized software to produce the **ANOVA table**. This table shows us the **variation between groups**, the **variation within groups**, and the **F ratio**. The sample F ratio (also called the **F statistic**) is the ratio of the mean variation between groups to the mean variation within groups. The F ratio is a measure of how strongly the sample contradicts the null hypothesis H_0. A large F ratio suggests that H_0 may be false, while an F ratio near 0 says that the sample evidence is not inconsistent with H_0. We would reject H_0 if the sample F ratio exceeds the F **critical value**.

Critical values of F may be found in a table (or calculated by a computer) for any α level (level of significance). Commonly used α levels are 0.10, 0.05, or 0.01. If the null hypothesis is true and if you do an ANOVA test repeatedly on different samples drawn from the same populations, then you can calculate the empirical **Type I error** as the ratio of the number of rejections to the number of times the test is performed. This should be approximately equal to α. If the null hypothesis is false, the empirical **power** of the test is the ratio of the number of rejections to the number of times the test is performed. Empirical **Type II error** is one minus the empirical power. If the test is able to reject a false null hypothesis every time, then the power of the test is 1. Power is higher when the means are more unequal, when the variances are smaller, when the sample sizes are larger, or when α is larger.

In order to investigate Type I error or power empirically, we must repeat the sampling experiment many times *and* must know whether the null hypothesis is true or false. Under normal situations this is impossible (even if we could conduct the experiment 100, 200, or even 1,000 times, we would not know if the null hypothesis is true or false). However, this type of experiment is possible with computer simulation. Suppose we design an experiment and replicate it 1,000 times with a true H_0. The laws of chance say that we would expect 1000α rejections (that is, if $\alpha = 0.05$, we would expect about 50 rejections, while if $\alpha = 0.01$, we would expect about 10 rejections). For a true H_0, the histogram of replicated sample F statistics should closely resemble the F distribution. But if we replicate the experiments with a false H_0, the histogram will be more symmetrical and will be shifted right. The number of rejections would exceed 1000α. Empirical power would be the number of rejections divided by 1,000. If the means differ by a great deal, power will be near 1, while if they differ very little, the power will be near α. The test's empirical Type II error would equal 1 minus its empirical power.

The ANOVA procedure rests on two assumptions. The **equal variance assumption** says that the variances are the same for all groups. The **normality assumption** says that samples from all groups are drawn from normal populations. Real data often violate one or both of these assumptions. It is therefore useful to study how violations of these assumptions affect Type I error, Type II error, and the power of the F test.

Illustration of Concepts

We would like to know if there are differences in the sizes of California navel oranges sold by three fresh produce retailers. We purchase oranges at each store, taking care to choose them at random from the stack of fresh oranges at each market. Our sample sizes are similar but not identical. We weigh each orange and calculate the sample means and standard deviations.

Retailer	Sample Size	Weights of Sampled Oranges (grams)	Mean	S.D.
Mercado Uno	10 oranges	263 290 284 298 250 308 317 277 357 280	292.51	30.29
Papa Jose	8 oranges	309 317 271 367 242 387 290 268	306.33	49.79
Green Grow	9 oranges	299 288 195 208 231 327 305 226 278	261.79	47.40

The standard deviation of each group is a measure of the **variation within groups**. The difference between the means (relative to their standard errors) suggests that there is also **variation between groups**. In order to test this observation we use **analysis of variance** (or **ANOVA**). Our hypothesis is:

H_0: Mean weight of oranges is the same in all three markets.
H_1: Mean weight of oranges is not the same in all three markets.

Only if the sample information is inconsistent with the null hypothesis will it be rejected. Using a spreadsheet or a statistical analysis program we create an **ANOVA table**:

Source of Variation	Sum of Squares	DF	Mean Square	F Ratio	P-Value
Between groups	8959.1	2	4479.57	2.463	0.106
Within groups	43651.2	24	1818.80		
Total	52610.3	26			

The sample **F ratio** (or **F statistic**) is 2.463. Since the **critical value** is 2.54 using $\alpha = 0.10$ (**Type I error**), we accept (do not reject) the null hypothesis. That is, in this particular sample, there is not enough evidence to refute the null hypothesis. However, the p-value tells us that it was a close decision. The p-value says that, if H_0 were true, an F ratio of 2.463 would arise by chance about 10.6% of the time. By accepting the null hypothesis we know that there is some chance (it could range from 0% to $1-\alpha$, or 90% in this case) that we have made a **Type II error** (the probability that we incorrectly accepted H_0). In this case, since the true means are unknown, we have no insight into the **power** of the test.

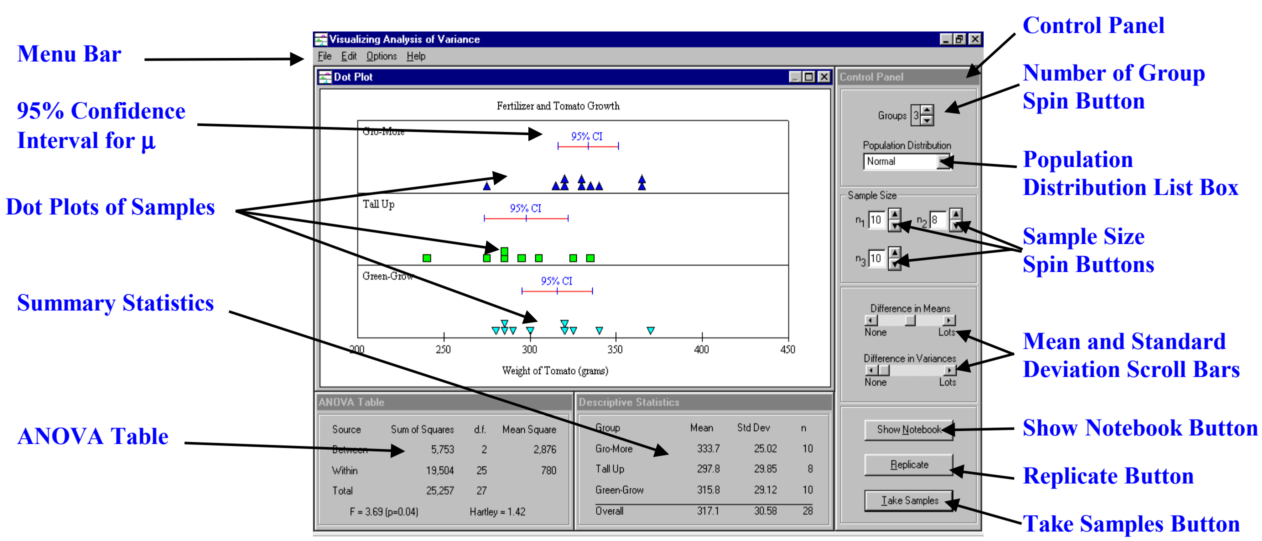
Is the **equal variances assumption** valid? To find out, we use Hartley's F_{max} test statistic. It is the ratio of the largest to the smallest sample variance, or $F_{max} = (49.79)^2/(30.29)^2 = 2.70$. From a table of F_{max} critical values with c = 3 groups we find that $F_{c,n/c-1} = F_{3,26/2} = 4.16$ using a 0.05 level of significance. Since F_{max} is < 4.16 we conclude that this sample does *not* provide enough evidence to reject the equal variances assumption.

A number of questions remain unanswered. Since the decision was close (p-value was 0.106) would we again accept the null hypothesis if the test were redone using a new sample? Would we have obained different results if we had used larger sample sizes? Is the **normality assumption** valid? The only way to answer these questions is to collect additional data, perform new ANOVA tests, and do tests for normality (e.g., probability plots or at least a histogram of each sample).

Orientation to Basic Features

This module allows you to conduct an ANOVA experiment to determine if you would accept or reject the null hypothesis that the means of several groups are equal. In addition, you can replicate the experiment up to 1,000 times to examine Type I error, Type II error, and power.

1. **Opening Screen**
   Start the module by clicking on the module's icon, title, or chapter number in the *Visual Statistics* menu and pressing the Run Module button. When the module is loaded, you will be on the introduction page of the Notebook. Read the questions and then click the Concepts tab to see the concepts that you will learn. Click on the Scenarios tab. Select Agriculture from the table of choices. Select a scenario, read it, and press OK. The upper left of the screen shows dot plots from each sample. The vertical black line represents the overall sample mean. The short blue vertical lines represent the sample mean from a particular group, while the red line connecting the black and blue line is the difference between the group sample mean and the overall sample mean. Below these dot plots are two tables. The ANOVA table is on the left, and a table of Summary Statistics is on the right. To the right is the Control Panel.



2. **Control Panel**
   a. Push the Take Samples button in the lower right of the Control Panel. Note that the dot plot, ANOVA Table and table of Summary Statistics change to reflect the new sample.
   b. Use the Sample Size spin buttons to change the sample size. Push the flashing Take Samples button. The dot plot and both tables reflect the new sample size.
   c. Click on the Difference in Means scroll bar to change the difference between the largest and smallest population means. When set to None there will be no difference in the means, and when set to Lots there will be a large difference in the means. Push the Take Samples button to draw a sample from the new populations. The Difference in Variances scroll bar works in the same way. Remember that an assumption of ANOVA is that the variance of each group is the same.
   d. Click the Groups spin button to change the number of groups in your experiment. You can select two, three, or four groups. Press the flashing Take Samples button so that the displays correspond to your new experiment.

e. Click on the Population Distribution list box. You can select either Normal, Uniform, or Skewed. If you select a skewed distribution, the program will randomly choose either left or right skewness (it will not change the choice unless you make a change in the Number of Groups spin button or Population Distribution list box). Push the Take Samples button to update your displays.

f. Push the Show Notebook button to select a new scenario or change to the Do-It-Yourself controls.

3. **Copying a Display**
   Click on the display you wish to copy. Its window title will be highlighted. Select Copy from the Edit menu (on the menu bar at the top of the screen) or Ctrl-C to copy the display. It can then be pasted into other applications, such as Word or WordPerfect, so it can be printed.

4. **Help**
   Click on Help on the menu bar at the top of the screen. Search for Help lets you search an index for this module, Contents shows a table of contents for this module, Using Help gives instructions on how to use Help, and About gives licensing and copyright information about *Visual Statistics*.

5. **Exit**
   Close the module by selecting Exit in the File menu (or click  in the upper right-hand corner of the window). You will be returned to the *Visual Statistics* main menu.

Orientation to Additional Features

1. **Replication**

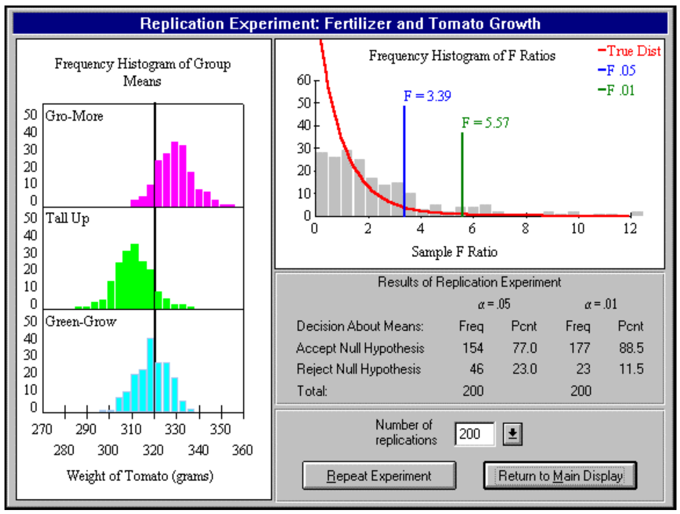
   Push the Replicate button to initiate a replication experiment.

   a. The Number of Replications list box permits 50, 100, 200, 500, or 1,000 replications.

   b. Push the flashing Start Experiment button to start the experiment. Histograms of the sample means from each group appear to the left, a histogram of the F statistics appears at the top right, and a panel showing the Results of Replication Experiment is below the histogram of F statistics.

   c. Finish Experiment and Pause Experiment buttons now appear. Push the Pause Experiment button to pause the experiment. The Pause Experiment button changes to a Continue Experiment button. Press this to resume the experiment. Pressing the Finish Experiment button suspends the building process and updates the display at the end of the experiment.

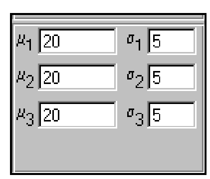
   d. When the experiment is finished, a display similar to the one on the right appears. The histogram of F statistics has an F distribution superimposed upon it. The critical values corresponding to $\alpha = 0.05$ and 0.01 are drawn in blue and green, respectively. The Results of Replication table shows the number and percentage appearing to the right and left of each critical value.



   e. Push the Repeat Experiment button to redo the experiment or the Return to Main Display button to leave replication mode.

2. **Do-It-Yourself**

   Press the Show Notebook button. Select the Do-It-Yourself tab and click OK. This will replace the Difference in Means and Difference in Variances scroll bars with a panel containing μ and σ boxes in which you can type your own values. An example with default values is to the right. Change any of these parameter values using your keyboard. Press the Take Samples button to draw a sample from the populations you have defined.



Basic Learning Exercises Name ________________

Understanding ANOVA

Press the Show Notebook button, select the Scenarios tab, and click on Agriculture. Select the Fertilizer and Tomato Growth scenario. Read the scenario. Click OK.

1. State the null and alternative hypotheses. Are the means equal, based on the Difference in Means scroll bar setting? Does this indicate that the null hypothesis is true or false?

   H_0:
   H_a:

2. What two assumptions are required to do ANOVA? Are they met in this scenario? How do you know? **Hint:** Look at the Population Distribution combo box and the Difference in Variances scroll bar.

3. Based only on the dot plots, would you conclude that the groups differ in central tendency or dispersion? Does any of the three samples have outliers or unusual data points? If so, what might be their effects?

4. Based on the 95% confidence intervals, would you conclude that the means are equal? What decision does this suggest about H_0?

5. Do you see anything in your samples that seems inconsistent with the belief that the population is normal? If you did not know whether or not the population was normal (see exercise 2) what would you conclude about normality?

6. Look at the Summary Statistics. Do the sample means and standard deviations confirm your comparisons of central tendency and dispersion among the groups?

7. To check your understanding of the ANOVA table, verify each of the following calculations. **Hint:** c is the number of groups and n is the total number of observations. If you need to review the formulas, click Help.

DF Between:	c–1 =
DF Within:	n–c =
Total DF:	n–1 =
Mean Square Between:	MSB = SSB / (c–1) =
Mean Square Within:	MSW = SSW / (n–c) =
F ratio:	F = MSB / MSW =

8. Move the Difference in Means scroll bar to None and press Take Samples. Refer to the dot plots to explain what is measured by Mean Square Between (MSB) and Mean Square Within (MSW). Why are they nearly equal if H_0 is true?

9. At $\alpha = 0.05$, does the F statistic (and its p-value) indicate that you should reject the null hypothesis of equal means? Would you expect to reject H_0?

10. Move the Difference in Means scroll bar to Lots and press Take Samples. Use the dot plots to explain why, if H_0 is false (unequal means), MSB is usually greater than MSW.

11. At $\alpha = 0.05$, does the F statistic (and its p-value) indicate that you should reject the null hypothesis of equal means? Would you expect to reject H_0?

Intermediate Learning Exercises Name _______________

Type I Error, Type II Error, and Power

Press the Show Notebook button, select the Scenarios tab, and click on Agriculture. Select the Fertilizer and Tomato Grow scenario. Read the scenario and Click OK.

12. Move the Difference in Means scroll bar all the way to the *left* (so it is set to None). You have made the null hypothesis true (no difference in means). Push the Take Samples button 10 times. Out of the 10 samples, how many times did you reject the null hypothesis? Since the null hypothesis is true in this case, based on these 10 trials what is the empirical Type I error of the ANOVA test? Do you think 10 samples are enough to estimate empirical Type I error? **Hint:** Empirical Type I error is the relative frequency of rejecting a false null hypothesis.

13. Push the Replicate button. Set the number of replications to 1,000 using the Number of Replications list box, and then press Start Experiment. When the experiment is finished, examine the table of results. How many rejections did you get using $\alpha = 0.05$? Using $\alpha = 0.01$? What is the empirical Type I error for $\alpha = 0.05$? For $\alpha = 0.01$? Do you feel that 1,000 samples is enough to yield a reasonable estimate of empirical Type I error?

14. Why is the empirical Type I error generally not exactly equal to the α-level?

15. Describe the histogram of F statistics. Does it closely resemble the theoretical F distribution that is superimposed on the histogram? Should it? Explain.

16. Describe the shapes of histograms of sample means (on the left). Do the means appear to be equal? Do the variances appear equal? Is this what you would expect? Explain.

17. Press the Return to Main Display button. Create a *slightly* false null hypothesis (i.e., slightly unequal means) by moving the Difference in Means scroll bar two clicks to the right of None. Push the Take Samples button 10 times. At $\alpha = 0.05$, how many times did you reject the null hypothesis (based on the p-value)? Since the null hypothesis is false in this case, based on these 10 trials what is the empirical power of the ANOVA test? **Hint:** Empirical power is the relative frequency of rejecting a false null hypothesis.

18. Press the Replication button. Press the Start Experiment button. Watch the table of results. After about 100 replications, press the Pause Experiment button. Look at the histograms of sample means (on the left). What do you notice about them? Press the Finish Experiment button. Describe the final histograms of means.

19. At the conclusion of the experiment, compare the distribution of the sample F statistics to the superimposed F distribution that is based on the assumption of equal means. How does this histogram indicate that the null hypothesis is false? Explain.

20. Refer to the table of results. Using $\alpha = 0.05$, how many rejections did you get? What was the empirical power? What was the empirical Type II error? Why is the empirical power relatively low in this exercise? **Hint:** Empirical Type II error is the relative frequency of accepting a false null hypothesis.

Effects of Sample Size

21. Press the Return to Main Display button. Use the Sample Size spin buttons to double each sample size. Push the Take Samples button to reset all of the displays. Choose 1,000 replications and push the Replicate button. Using an α-level of 0.05, what is the empirical power and empirical Type II error of the test? How does this answer compare with your answer to exercise 20? What generalization can you make about the effect of sample size on power and Type II error in the F test?

Advanced Learning Exercises Name ________________

Investigate Equal Variances Assumption

22. Press the Show Notebook button, select the Do-It-Yourself tab and click OK. At the top of the Control Panel set the number of groups to three and the population to Normal. The Difference in Means and Variances scroll bars have been replaced with boxes to enter the mean and standard deviation of each group. Set all three means to 3,000 and all three standard deviations to 500. Set all three sample sizes to 10 using their spin buttons. Press the Take Samples button. Record the value for the Mean Square Within (MSW), Hartley's statistic, and the p-value of the F statistic. Is its p-value below 0.05? Why is this important? Define Hartley's statistic and tell what it is used for. What is its critical value in this situation? **Hint:** Type Hartley after selecting the Search for Help feature under Help on the menu bar.

    MSW ______________ Hartley's Statistic __________ p-value of F ______

23. Press the Takes Samples button 10 times. Each time record the value of MSW, Hartley's statistic, and the p-value of the F statistic. Could you tell that the variances were equal by looking at the dot plots? What is the median of the 10 MSWs? Circle each value of Hartley's statistic that is greater than the critical value. How many rejections are there? Circle each p-value below 0.05. How many are there? Is the average MSW what you would expect; why or why not? Is the number of circled Hartley's statistics what you would expect? Why or why not? Is the number of circled p-values what you would expect? Explain.

    MSW 1 ________ 2 ________ 3 ________ 4 ________ 5 ________ 6 ________
    7 ________ 8 ________ 9 ________ 10 ________ Median ________

    Hartley's 1 ________ 2 ________ 3 ________ 4 ________ 5 ________
    6 ________ 7 ________ 8 ________ 9 ________ 10 ______

    p-value 1 ________ 2 ________ 3 ________ 4 ________ 5 ________
    6 ________ 7 ________ 8 ________ 9 ________ 10 ________

24. Press the Replications button. Set the Number of Replications list box to 1,000. Press the Start Experiment button (if you wish you can press the Finish Experiment button). What percentage of the time was the null hypothesis rejected at the 5% and 1% α-level? What percentages would you have expected? Are your results plausible given sampling variation?

25. Change two of the standard deviations to 400 and 580 to create *slightly* unequal variances. Verify that the MSW is still about 250,000. Press the Takes Samples button 10 times. Can you tell that the variances are unequal by looking at the dot plots? Each time, record Hartley's statistic and the p-value of the F ratio. Circle each value of Hartley's statistic that is greater than the critical value. How many are there? What does this say about the power of Hartley's statistic to detect *slightly* unequal variances? Circle each p-value below 0.05. How many are there? Is the number of circled p-values what you would expect? Explain.

Hartley's	1 ________	2 ________	3 ________	4 ________	5 ________
	6 ________	7 ________	8 ________	9 ________	10 ________
p-value	1 ________	2 ________	3 ________	4 ________	5 ________
	6 ________	7 ________	8 ________	9 ________	10 ________

26. Press the Replications button. Press the Start Experiment button. What percentage of the time was the null hypothesis rejected at the 5% and 1% α-level (empirical Type I error)? What does this mean about the robustness of one-way ANOVA to slightly unequal variances? Can you tell that the variances are unequal by looking at the histograms of sample means?

27. Change the standard deviations to 300, 500, and 640 to create *moderately* unequal variances. Verify that the MSW is still about 250,000. Press the Takes Samples button 10 times. Can you tell that the variances are unequal by looking at the dot plots? Each time record Hartley's statistic and the p-value of the F ratio. Circle each value of Hartley's statistic that is greater than the critical value; how many are there? What does this say about the power of Hartley's statistic to detect *moderately* unequal variances? Circle each p-value below 0.05; how many are there? Is the number of circled p-values what you would expect? Why or why not? Since comparing dot plots is not a reliable test for equal variances, why look at the dot plots?

Hartley's	1 ________	2 ________	3 ________	4 ________	5 ________
	6 ________	7 ________	8 ________	9 ________	10 ________
p-value	1 ________	2 ________	3 ________	4 ________	5 ________
	6 ________	7 ________	8 ________	9 ________	10 ________

28. Press the Replications button. Press the Start Experiment button. What percentage of the time was the null hypothesis rejected at an α-level of 5% and 1% (empirical Type I error)? Is one-way ANOVA robust to moderately unequal variances? Can you tell that the variances are unequal by looking at the histograms of sample means?

29. Change the standard deviations to 100, 500, and 706 to create *very* unequal variances. Verify that the MSW is still about 250,000. Press the Takes Samples button 10 times. Can you tell that the variances are unequal by looking at the dot plots? Each time, record Hartley's statistic and the p-value of the F ratio. Circle each value of Hartley's statistic that is greater than the critical value; how many are there? Circle each p-value below 0.05. How many are below 0.05? What does this say about the power of Hartley's statistic to detect *very* unequal variances? Is the number of circled p-values what you would expect? Why or why not?

Hartley's	1 ________	2 ________	3 ________	4 ________	5 ________
	6 ________	7 ________	8 ________	9 ________	10 _______
p-value	1 ________	2 ________	3 ________	4 ________	5 ________
	6 ________	7 ________	8 ________	9 ________	10 ________

30. Press the Replications button. Press the Start Experiment button. What percentage of the time was the null hypothesis rejected at the 5% and 1% α-level (empirical Type I error)? What does this mean about the robustness of one-way ANOVA to very unequal variances?

31. Change the standard deviations to 10, 10, and 865 to create *extremely* unequal variances. Verify that the MSW is still about 250,000. Press the Takes Samples button five times. Can you tell that the variances were unequal by looking at the dot plots? Note the size of the Hartley's statistic. Press the Replications button and the Start Experiment button. What percentage of the time was the null hypothesis rejected at the 5% and 1% α-level? What does this mean about the robustness of one-way ANOVA to *extremely* unequal variances?

32. How big is Hartley's statistic in the case of extremely unequal variances? From these experiments, what can you conclude overall about the robustness of one-way ANOVA to the equal variance assumption? *Note this result is only true for one-way ANOVA!*

33. Set all three σ's to 500 and the three μ's to 100, 300, and 500 (so the null hypothesis of equal means is false). Conduct a replication experiment. What is the power of one-way ANOVA in this case? Change the σ's to 300, 500, and 640. What is the empirical power of one-way ANOVA in this case? Why is the power reduced? Do you think this is a general result?

Investigate the Normality Assumption

34. Set all three standard deviations equal to 500 and all three means equal to 100. Set the Population Distribution to Uniform. The normality assumption is now being violated. Conduct a replication experiment. What percentage of the time is the null hypothesis rejected at the 5% and 1% α-level? This is empirical Type I error. What does it say about the robustness of one-way ANOVA to sampling from a uniform distribution?

35. Set the Population Distribution to Skewed. Conduct a replication experiment. What percentage of the time is the null hypothesis rejected at the 5% and 1% α-level? What does this say about the robustness of one-way ANOVA to sampling from a skewed distribution? What is the shape of the histograms of sample means? What do the results in exercise 33 and this exercise tell you about the robustness of one-way ANOVA to *some* violations of the normality assumption? *This result is only true for one-way ANOVA.*

36. Set all three standard deviations to 500 and the three means to 100, 300, and 500. Conduct a replication experiment. What is the power of one-way ANOVA in this case? Compare your results with those in exercise 33. Is this a general result, or is it only true for some violations of the normality assumption?

Individual Learning Projects

Write a report on one of the three topics listed below. Use the cut-and-paste facilities of the module to place the appropriate graphs in your report.

1. Derive a power curve using any scenario or make up your own scenario using the Do-It-Yourself controls (use at least three groups). Set the Difference in Means scroll bar to None (the null hypothesis is true) and the Difference in Variances scroll bar to None (equal population variances). Set the Population Distribution to Normal. Use replicate (with 1,000 replications) to find the empirical Type I error. Copy the F distribution and the histogram of sample means (click on a display and press Ctrl–C on your keyboard to copy the display to the clipboard). Return to the main display. Move the Differences in Means scroll bar two clicks to the right. Repeat the replication experiment finding the power of the test. Continue moving the scroll bar two clicks to the right and repeating the experiment (a total of six experiments will be done). How should you measure the difference in the means ? (**Hint:** The histogram of sample means will give you a good indication of each group's mean.) Place this measure on the horizontal axis and sketch a power curve. Write a report on the power of one-way ANOVA. Describe the scenario, state the null and alternative hypotheses, and discuss how you measured the difference in means. Each point on your power curve should be validated with a display of the F distribution and the histogram of the means. How much difference must there be before there is a 50% chance that the ANOVA test will reject the null hypothesis? Are these results generalizable?
2. Select any scenario or make up your own scenario using the Do-It-Yourself controls (you must use at least three groups). Examine the dot plot of samples, the Summary Statistics and the ANOVA Table. Is the null hypothesis accepted (not rejected) or rejected? How is this reflected in all three displays? Take a second sample. Redo the previous analysis. Do this analysis a total of 10 times. Each time make a copy of your dot plot, the Summary Statistics and ANOVA Table. Write a report on the results of your 10 samples. Describe the scenario, state the null and alternative hypotheses, and discuss how you analyzed all three displays (what you looked for in each display). Based on these 10 samples, are the means equal? If you had used only your first sample, would you have gotten the same result? Why would a statistician always like to replicate an experiment a number of times? Why is it usually infeasible?
3. Sketch two power curves using any scenario or make up your own scenario using the Do-It-Yourself controls (you must use at least three groups). Set the Difference in Means scroll bar to None (the null hypothesis is true) and the Difference in Variances scroll bar to None (equal population variances). Set the Population Distribution to Normal. Follow the process described in project 1, using *three* clicks instead of two (four experiments will be created). Set the population being sampled to skewed and repeat the four experiments. How should you measure the difference in the means? (**Hint:** The histogram of sample means will give you a good measure of each group's mean.) Place this measure on the horizontal axis and sketch both power curves (you have only three points). Write a report discussing how sampling from a skewed distribution affected one-way ANOVA (both Type I error and power). Describe the scenario, state the null and alternative hypotheses, and discuss how you measured the difference in means. Are these results generalizable?

Team Learning Projects

Select one of the three projects listed below. In each case produce a team project that is suitable for an oral presentation. Use presentation software or large poster boards to display your results. Graphs should be large enough for your audience to see. Each team member should be responsible for producing some of the graphs. Ask your instructor if a written report is also expected.

1. Derive a family of power curves to illustrate the effect of sample size. Use any scenario or make up your own scenario using the Do-It-Yourself controls (you must use at least three groups). Set the Difference in Means scroll bar to None (the null hypothesis is true) and the Difference in Variances scroll bar to None (equal population variances). Set the Population Distribution to Normal. Each team member will generate a power curve using different sample sizes. One member will set each group sample size to 5, one member will set each to 50, and all other members will use sample sizes in between (as equally spaced as possible). Each team member will complete the six experiments described in Individual Learning Project 1. After all team members have completed their experiments, one graph showing each team member's power curve should be produced. Discuss the effect sample size has on power. Each team member should have graphs that validate his or her own power curve. Are these results generalizable?
2. Derive a family of power curves to illustrate the effect violating the equal variances assumption has on the power of one-way ANOVA. Use any scenario or make up your own scenario using the Do-It-Yourself controls (use at least three groups). Set the Difference in Means scroll bar to None (the null hypothesis is true). Set the Population Distribution to Normal. Each team member will generate a power curve using a different setting of the Differences in Variances scroll bar. One member will set the scroll bar to None, one member will set the bar to Lots, and all other members will use settings in between (as equally spaced as possible). Each team member will complete the six experiments described in Individual Learning Project 1. After all team members have completed their experiments, one graph showing each team member's power curve should be produced. Discuss the effect on power of the degree of inequality in variances. Each team member should have graphs to validate his or her own power curve. Are these results generalizable?
3. Derive a family of power curves to illustrate how the distribution of sample size among the groups affects the power of one-way ANOVA. Use any scenario or make up your own using the Do-It-Yourself controls. Set the Population Distribution to Normal and select three groups. Use a total sample size of 45. Each team member will generate a power curve using a different distribution of this total sample size. One member will set each group sample size to 15. Another member will set one group sample size to 39, and the other two to 3. Other members will use any sample sizes that sum to 45 (select an interesting distribution of sample sizes). Each team member will complete the six experiments described in Individual Learning Project 1. After the experiments are completed, produce a single graph showing each team member's power curve. Discuss how the distribution of sample size among the groups affects power and explain why. Each team member should have graphs that validate his or her own power curves. Are these results generalizable?

Self-Evaluation Quiz

1. The F statistic for analysis of variance is
   a. used to identify significant differences in means.
   b. either a left-tail or a right-tail test.
   c. based on the assumption of normal populations.
   d. all of the above.
   e. either a or c.

2. A dot plot in connection with ANOVA often
   a. reveals differences in means.
   b. reveals differences in variances.
   c. reveals differences both in means and variances.
   d. reveals nothing about means or variances.
   e. requires a normal population to be effective.

3. The sample F statistic
   a. is always positive.
   b. must be less than or equal to 1.
   c. is based on the ratio of sample means.
   d. goes to zero unless the population is normal.
   e. has more than one of the above characteristics.

4. The effect of increased group sample sizes in ANOVA would be
   a. to improve the utility of the dot plots in assessing the data.
   b. to improve the power of the F test for equality of means.
   c. to improve the power of Hartley's test of homogeneity of variance.
   d. to improve the estimates of the sample means and standard deviations.
   e. all of the above.

5. The F test in analysis of variance
   a. lacks validity unless each group has at least 30 items.
   b. requires equal group sizes in order to be accurate.
   c. suffers a sharp loss of power with slight departures from normality.
   d. is used because computations are easy.
   e. has none of the above characteristics.

6. The p-value for the sample F statistic shows
   a. the probability that the null hypothesis is false.
   b. the probability of obtaining the statistic by chance if H_0 is true.
   c. the probability of committing Type I error if H_0 is rejected.
   d. more than one of the above.
   e. none of the above.

7. If the null hypothesis of equal group means is true, the population is normal, and the group variances are equal, then
   a. the distribution of replicated sample F statistics should be normal.
   b. the distribution of replicated sample F statistics should be chi-square.
   c. the distribution of replicated sample F statistics should be Student's t.
   d. the distribution of replicated sample F statistics should be F.
   e. the distribution of replicated sample F statistics is uncertain.

8. If the null hypothesis of equal group means is false, the population is normal, and the group variances are equal, then
   a. the distribution of replicated sample F statistics should be shifted right.
   b. the distribution of replicated sample F statistics should be shifted left.
   c. the distribution of replicated sample F statistics should be centered on zero.
   d. the distribution of replicated sample F statistics should be the same as the true F.
   e. the distribution of replicated sample F statistics is uncertain.

9. In an ANOVA experiment with four groups of five observations and a true null hypothesis,
   a. the mode of the theoretical F distribution is positive.
   b. the mode of the theoretical F distribution is zero.
   c. the mode of the theoretical F distribution is negative.
   d. the mode of the theoretical F distribution is indeterminate.
   e. the mode of the theoretical F distribution is undefined.

10. If an ANOVA sampling experiment is replicated many times, the histograms of sample means for the groups will
    a. be approximately normal if the sample size is large.
    b. resemble a chi-square distribution with n – 1 degrees of freedom.
    c. reflect differences in group variances, if they exist.
    d. have the same scale from the histogram of F statistics.
    e. have more than one of the above characteristics.

11. In the ANOVA comparison of k groups, which is *not* true of Hartley's F_{max} test statistic?
    a. It is used to assess whether or not the population variances are homogeneous.
    b. It is the ratio of the smallest to the largest sample variance.
    c. It is generally larger when there is inequality of variance.
    d. It requires special tables because it is not the same as Fisher's F.
    e. It tends to be near 1 when the variances are homogeneous.

12. Which is *not* true of heterogeneity of group variances?
    a. One-way ANOVA is robust to a great deal of heterogeneity.
    b. It may lead to non-parametric alternatives (e.g., Kruskal-Wallis test).
    c. It refers to the skewness of each group's distribution.
    d. In general, it reduces the power of the F test.
    e. It is detected through Hartley's F_{max} test statistic.

Glossary of Terms

Analysis of variance Abbreviated ANOVA, this is the partitioning of variation about the overall mean into several additive categories. Part of the variation is attributed to the proposed ANOVA model, and the rest is attributed to error. In one-factor ANOVA there are two categories: *between-group variation* (model) and *within-group variation* (error). This partitioning of variance may be written SSE = SSB + SSW.

ANOVA table Worksheet summarizing the calculations for analysis of variance. In one-factor ANOVA there are three columns: sum of squares, degrees of freedom, and mean square. The first two columns also include a column sum.

Between treatments Refers to variation between groups. See **Analysis of variance**.

Critical value Separates the rejection and acceptance (non-rejection) region in the distribution of the test statistic assuming that the null hypothesis is true. It is a function of α, the level of significance used in a test.

Degrees of freedom For a one-factor ANOVA the variation *between* c treatment groups has $c - 1$ degrees of freedom, and the variation within the groups has $n - c$ degrees of freedom.

Empirical power See **Power of a test**.

Empirical Type I error See **Type I error**.

Empirical Type II error See **Type II error**.

Equal variance assumption To derive the test statistic in one-factor ANOVA we assume that the population of each group being sampled has the same variance. If this assumption is violated, the distribution of the test statistic under the null hypothesis is open to question. If a test is relatively unaffected by a violation of an assumption, the test is said to be robust.

F ratio Ratio of the mean square between groups to the mean square within groups for a sample. If group means are equal, this ratio will tend to be near 0. A large ratio suggests that the group means may be unequal.

F statistic See **F ratio**. These terms are used interchangeably.

Mean square The sum of squares divided by its degrees of freedom.

Normality assumption To derive the test statistic in one-factor ANOVA we assume that the population of each group being sampled is normally distributed. If this assumption is violated the distribution of the test statistic under the null hypothesis is open to question. If a test is relatively unaffected by a violation of an assumption, the test is said to be robust.

Power of a test If the null hypothesis is false, *theoretical* power is the probability of correctly rejecting the null hypothesis $(1 - \beta)$. If the null hypothesis is known to be false, the *empirical* power is the ratio of the number of rejections to the number of times the test is performed. Ideally, the power of the test would be near 1. See **Type II error**.

Robust The quality of being relatively unaffected by a particular problem (e.g., non-normality or unequal variances). For example, one-way ANOVA is moderately robust to non-normality.

Source of Variation See **Analysis of Variance**.

Sum of Squares The sum of squared deviations about the mean.

Theoretical power See **Power of a test.**

Treatment In the context of one-factor ANOVA the "treatment" simply refers to a group (e.g., Freshman, Sophomore, Junior, Senior). The test is looking for possible differences among the treatments (e.g., does GPA vary among the four class "treatments").

Type I error If the null hypothesis is true, the *theoretical* Type I error is the probability of incorrectly rejecting the null hypothesis. Its probability is denoted α. When the null hypothesis is known to be true, the *empirical* Type I error is the ratio of the number of rejections to the number of times the test is performed.

Type II error If the null hypothesis is false, the *theoretical* Type II error is the probability of incorrectly accepting (not rejecting) the null hypothesis. Its probability is denoted β. When the null hypothesis is known to be false, the *Empirical* Type II error is the ratio of the number of acceptances to the number of times the test is performed.

Within treatments Refers to variation within groups. See **Analysis of variance**.

Solutions to Self-Evaluation Quiz

1. e Do Exercises 1, 2, and 7. Read the Overview of Concepts.
2. c Do Exercise 3.
3. a Do Exercises 1–9.
4. e Do Exercises 18–21.
5. e Do Exercises 7, 12–19.
6. d Do Exercises 13–17. Read the Illustration of Concepts.
7. d Do Exercises 13–15. Read the Overview of Concepts.
8. a Do Exercises 17–20.
9. a Do Exercises 12–15.
10. e Do Exercises 2 and 13–16.
11. b Do Exercises 22–23. Read the Illustration of Concepts.
12. c Do Exercises 22–23.