Chapter 25. Comparing several means.

1. What are some typical problems where we might want to do this?
2. Which procedures from earlier chapters does this build upon?
3. Is there a reasonable way to just make these comparisons by looking at graphs?
4. Why is the name of this procedure "Analysis of Variance" when our hypotheses are about means?
5. What are the hypotheses?
6. How do we use software to obtain a P-value for our test?
7. How do we write the conclusion of this ANOVA test?
8. What is the follow-up data analysis we do? (Remember the follow-up data analysis after the chi-squared test of independence. How is this similar and how is it different from that?)
9. How are they getting that P-value? What is our test statistic? What values of it support Ha and why?
10. What are the conditions needed to do this procedure?
11. What's in this chapter that we're not covering?
12. From this material in Ch. 25, what homework and quiz problems are students expected to do?
13. From this material in Ch. 25, what will be covered on the Final Exam?
14. Why is this material covered in the course?

Discussion:

1. What are some typical problems where we might want to do this?
   Answer: Look at the beginning of Chapter 25. Example 25.1. We saw this example about the lengths of three different varieties of flowers in Chapter 2. Are the graphs alone adequate for us to conclude that the three population means are not equal? Is it possible that three groups could have means that were significantly different where the graphs would not make it this obvious? (That is, do you suppose we need some more precise method than looking at graphs?)

2. Which procedures from earlier chapters does this build upon?

Answer: This is a "multiple comparisons" method, just as we learned in Ch. 23 for categorical variables. It is also a question about comparing means, and we compared two means in Chapter 19. The calculations we will do in this chapter do not resemble those in either of those chapters.

3. Is there a reasonable way to just make these comparisons by looking at graphs?

   Answer: Often there is. You are advised to ALWAYS make a set of comparative graphs of your data to begin your analysis of data like this.

4. Why is the name of this procedure "Analysis of Variance" when our hypotheses are about means?

   Below are results from two hypothetical studies. (Study A and Study B.) Each study compared three medical treatments (Treatments 1, 2, and 3.) The data give are the time it took the patient to respond to the treatment, in minutes.

   Activity 1a. By hand, make reasonably careful comparative boxplots of the data for Study A. All of your boxplots should have a scale of 20 to 50. As Minitab does, you should "stack" them so that it is easy to see which of the three treatments has larger numbers or smaller numbers. What do your comparative dotplots suggest about whether the mean length of time to respond is different for the three treatments.

   Activity 1b. Follow the same instructions for Study B.

   Activity 1c. In fact, in both studies, the mean of the values for Treatment 1 is 30 minutes, the mean for Treatment 2 is 36 minutes, and the mean for Treatment 3 is 27 minutes. Since the question here is whether the means are significantly different, if all we had to look at is the sample means, the answer to this question in these two studies should be the same. Yet you probably don't think that the answer to this question is the same. Try to describe what you see. Think of the word "variability."

A Treatment	A time	B Treatment	B time
1	30	1	24
1	32	1	22
1	28	1	36
1	30	1	38
2	36	2	30
2	37	2	34
2	35	2	38
2	36	2	46
3	24	3	20
3	26	3	24
3	28	3	30
3	30	3	34

Activity 2: Open the document file of this handout from the Course Calendar page, copy the data, and then paste it in to Minitab. Use Minitab to compute the means and standard deviations of the three treatments in each study and confirm the numbers given to you in Activity C. Fill in this table.

Study A				Study B		
Treatment	Mean	St Dev		Treatment	Mean	St Dev
1				1		
2				2		
3				3		

Discuss with your group. What do you see in both your graphs and in these numerical standard deviations that supports the idea that Study A and Study B give different answers to our question?

5. What are the hypotheses?

   Answer:
   Ho: The population means are all equal.
   Ha: At least two of the population means are different.

6. How do we use software to obtain a P-value for our test?

   Answer:

   Minitab: Stat > ANOVA > One-way
   In "Response" put the variable with the numerical values that you are getting the means from.
   In "Factor" put the variable that identifies the groups.

   Because we have so little time to look at this material, all I expect you to interpret is the P-value. Interpret that in the same way we have always done.

   (You can also understand the graphs provided in the Minitab ANOVA output – they are simply comparative graphs of confidence intervals for the mean. So the less they overlap, the stronger the evidence for differences in the means.)

7. How do we write the conclusion of this ANOVA test?

Answer: In the same way as before: P-value less than or equal to the significance level means that you reject Ho. Etc.

8. What is the follow-up data analysis we do? (Remember the follow-up data analysis after the chi-squared test of independence. How is this similar and how is it different from that?)

   Answer: If the P-value indicates there is a significant difference, then say what you see about which means appear to be most different and by how much.

   Activity 3. Groups. Each of you pick one of the studies (A or B). For that study write hypotheses, use Minitab to do the ANOVA test, and write the conclusion. Then ask someone in your group who analyzed the other study what their conclusion was. Compare the P-values.

9. How are they getting that P-value? What is our test statistic? What values of it support Ha and why?

   Answer: The question is – how far apart are the sample means compared to the variability of the individual observations. So, in our Studies A and B, the same means were the same distance apart in both studies, but in Study B, the variability in the individual observations was much larger, so the sample means didn't look so far apart in Study B, relatively speaking. That's why you decided that Study A gave good evidence that the three treatments differed in the mean time to respond, where Study B did not give good evidence that the three treatments differed in the mean time to respond.

   In our text, this is discussed in the section "The idea of analysis of variance." The statistic that measures this is $F = \frac{\text{variation among the sample means}}{\text{variation among the individual observations in the same sample}}$.

   Thus large values of F tell us that the sample means are relatively far apart. Those give P-values near zero, which tells us to reject Ho.

   There is an applet on One-way ANOVA which allows you to change each of these and see how the change affects the F. I like Exercises 25.5 and 25.6, but I believe they were written for a previous version. Here are the needed changes: Use "New Samples" instead of reset. Also, ignore any specific numbers in the problems or the answer because it no longer always opens with the same data. Follow the general idea as stated in each part of each question.

10. What are the conditions needed to do this procedure?

    Answer: Read the section "Conditions on ANOVA." As in our other chapters, we discuss robustness – how important is each condition and how much can we violate it and still have faith that our P-value is reasonably close to correct? As they say, the main condition to check is that the variability in the samples is about the same – or at least the largest standard deviation is no more than twice as large as the smallest standard deviation. If that is not true, then your P-value should not be trusted.

Example 25.4 in this section is a good example to read and to see the output. You can see from the comparative stemplots that the three samples aren't very far apart, although it is somewhat clear that the Control group has a lower mean. Look at the Minitab output for this problem and focus on the p-value and also on the graphs of the confidence intervals for the sample means. What would your conclusion to this test be at the 1% level? At the 5% level?

11. What's in this chapter that we're not covering?

    We're skipping the last two sections about details of the F distribution and how the F statistic is calculated. All of it is fairly easy to read if you need to know this. They are based on ideas we already had in the course – degrees of freedom, standard deviations means, etc. But because the formulas are rather long, and they don't look just like our other formulas (in the way that the Ch. 18-21 formulas all had the same pattern) you won't find them very useful in remembering the basic ideas of how to test whether a group of means are all the same or not.

12. From this material in Ch. 25, what homework and quiz problems are students expected to do?

    See the end of this page.

13. From this material in Ch. 25, what will be covered on the Final Exam?

    Answer: None.

14. Why is this covered if it isn't going to be on the test?

    Answer: . As you recall, I think that the statistics techniques where you compare several populations are commonly used in real-world applications. Those definitely include inference on two-way tables, regression, and inference on comparing several means.

    I also think it is a good idea to allow you more than four days or so between covering the material for the final exam and actually taking the final exam. So this is what I choose to do in that last class or two.

Quiz 15. NO QUIZ.

Homework: For each of these homework problems, use software. If graphs are requested make comparative dotplots, since Minitab does not make comparative stemplots. If the results of a test are requested, give only the P-value, and do not try to give the value of the F-statistic.

HW: 25.1, 25.3, 25.5, 25.6, 25.7, 25.9. (Remember the revised instructions for 25.5 and 25.6.)