

Packet 12: Analysis of Variance (ANOVA)

Textbook material: Supplemental Chapter 28

After completing this material, you should be able to:

- conduct an analysis of variance f-test with the appropriate StatCrunch output.
- determine when it is appropriate to consider multiple comparisons and interpret the results provided in StatCrunch output.
- state when it is valid to use this procedure

Goal:

Example: For fast-food restaurants, the drive-through window is an increasing source of revenue. The chain that offers the fastest service is likely to attract additional customers. Each month, QSR magazine publishes its results of drive-through service times (from menu board to departure) at fast-food chains. In a recent month, the drive-through service times of random sample of 20 customers from Burger King, McDonald's, Wendy's, and White Castle were recorded.

- What type of samples were selected? What variable was recorded? Why can none of the inferences we've discussed so far be used to analyze this data?

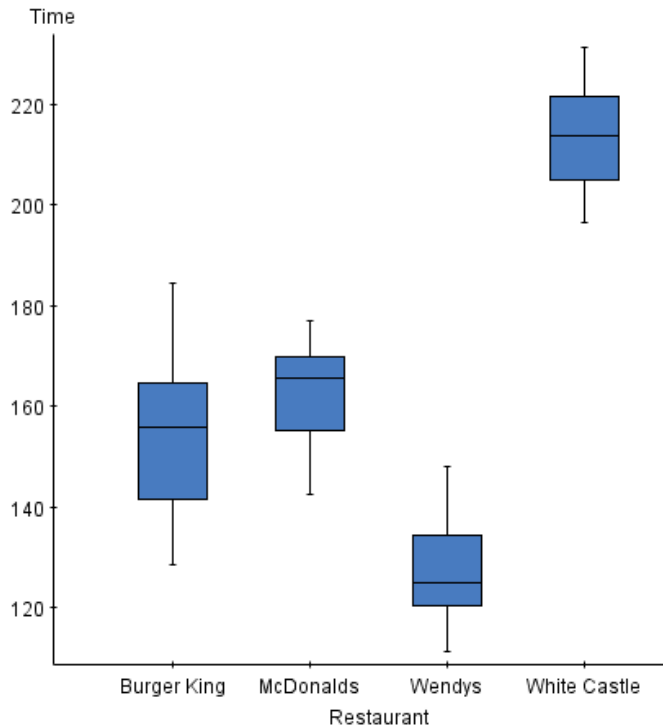
- What hypotheses are we interested in testing?

Now we need a way to test these hypotheses. We cannot simply use a t-test because there is no way to “take the difference” of more than two different means – it just doesn't make sense. To get around this problem, we need to use a different technique – we will use a new test statistic, called an F-statistic. The F-statistic is constructed from two different *variances* (hence, analysis of *variance*) even though our goal is to compare several means.

When comparing two variances, we look at the ratio of those variances and use the F-distribution. So, the test statistic that we're interested in is the following:

Now that we have a test statistic (albeit a difficult test statistic to calculate), we need a way to compute the probability (p -value). Since we have calculated an f -statistic, it makes sense that we would need to use an F -table. However, this table is more difficult to deal with. Instead, we'll rely on StatCrunch for the calculation of both the test statistics and its associated probability.

Back to the example: StatCrunch output is given below comparing the drive through times for the four fast food chains.



Analysis of Variance results:

Responses stored in Time.

Factors stored in Restaurant.

Factor means

Restaurant	n	Mean	Std. Dev.	Std. Error
Burger King	20	154.32375	15.044683	3.3640935
McDonalds	20	162.3761	10.483352	2.3441486
Wendys	20	127.1731	9.7918625	2.189527
White Castle	20	213.11584	10.282395	2.2992134

ANOVA table

Source	df	SS	MS	F-Stat	P-value
Treatments	3	77292.195	25764.064	191.60732	<0.0001
Error	76	10219.177	134.46284		
Total	79	87511.37			

Using $\alpha = 0.05$, determine if there is some difference in the mean drive-through times for the four chains.

Unfortunately, the conclusion from the ANOVA f-test can be rather ambiguous. What additional information would be useful?

Whenever the null hypothesis is rejected, the inference continues in the form of *multiple comparisons* – these are techniques used to determine what specific differences exist among the means. There are a variety of ways to conduct multiple comparisons; we are going to use Tukey’s method which is the method readily available in StatCrunch. This comparison technique tests all possible pairs of means to determine which are significantly different.

In our comparison of fast food drive-through times, how many different pairs of means will be compared?

The output generated by StatCrunch produces both p-values and confidence intervals for all possible pairs. How will we know when a comparison is significant?

Back to the example: The StatCrunch output below shows the results of the multiple comparisons. Use this to determine what significant differences exist in the mean drive-through times. Summarize your results.

Tukey 95% Simultaneous Confidence Intervals

Burger King subtracted from

	Difference	Lower	Upper	P-value
McDonalds	8.05235	-1.5798972	17.684597	0.1336
Wendys	-27.15065	-36.782898	-17.518402	<0.0001
White Castle	58.7921	49.15985	68.42435	<0.0001

McDonalds subtracted from

	Difference	Lower	Upper	P-value
Wendys	-35.203	-44.835247	-25.570753	<0.0001
White Castle	50.73975	41.107502	60.371998	<0.0001

Wendys subtracted from

	Difference	Lower	Upper	P-value
White Castle	85.94275	76.3105	95.575	<0.0001

In order to conduct this type of inference, the following **assumptions** must be satisfied:

Example: Researchers investigated dietary intake of calcium among a cross section of 113 healthy women ages 20 - 88. The researchers formed four age groups as follows: Group A, 20 - 45.9 years; Group B, 46 - 55.9 years; Group C, 56 - 65.9 years; and Group D, 66 and older. Calcium from food intake was measured in mg/day.

Conduct a complete analysis at a significance level of 0.05 to determine if there are significant differences in the mean calcium intake across the four age groups.

Analysis of Variance results:

Data stored in separate columns.

Column means

Column	n	Mean	Std. Error
Group A	22	1448.3182	134.1237
Group B	14	992.6429	167.64308
Group C	29	873.8276	56.58944
Group D	48	851.7083	55.697193

ANOVA table

Source	df	SS	MS	F-Stat	P-value
Treatments	3	5930572	1976857.4		<0.0001
Error	109	2.3024792E7	211236.62		
Total	112	2.8955364E7			

Tukey 95% Simultaneous Confidence Intervals

Group A subtracted from

	Lower	Upper
Group B	-865.6463	-45.704376
Group C	-913.5314	-235.44981
Group D	-905.3508	-287.8689

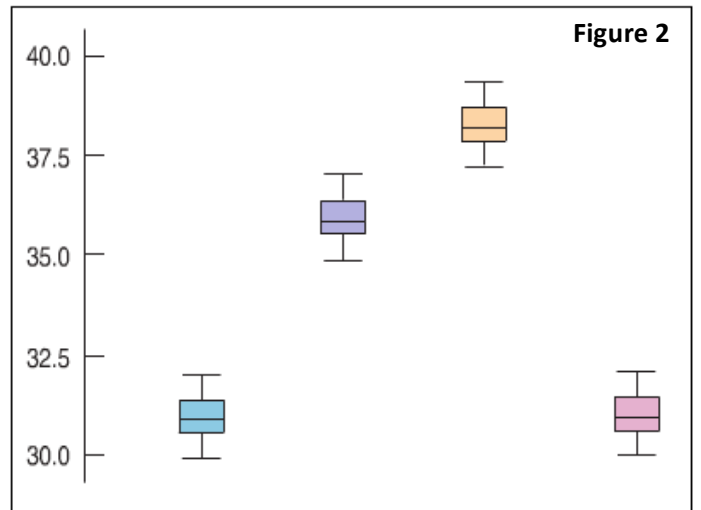
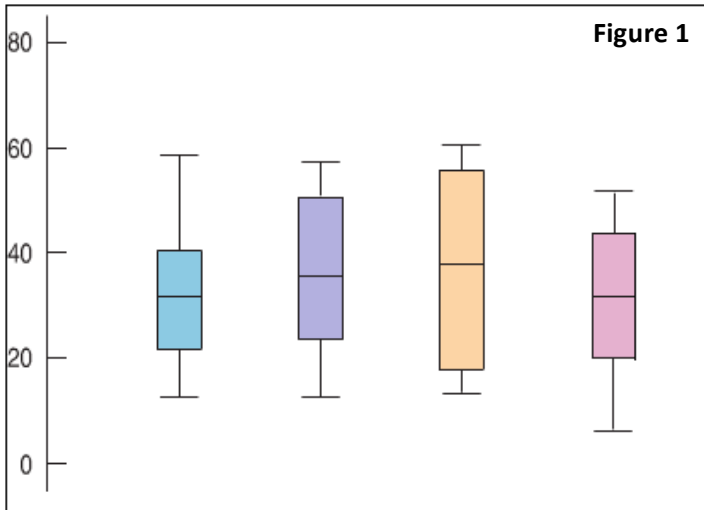
Group B subtracted from

	Lower	Upper
Group C	-509.0702	271.43967
Group D	-505.17517	223.30612

Group C subtracted from

	Lower	Upper
Group D	-304.1542	259.9157

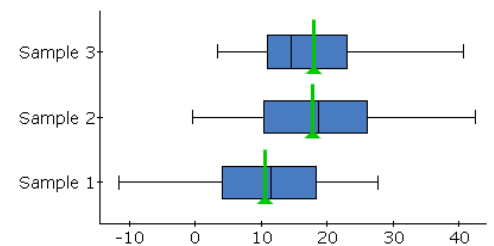
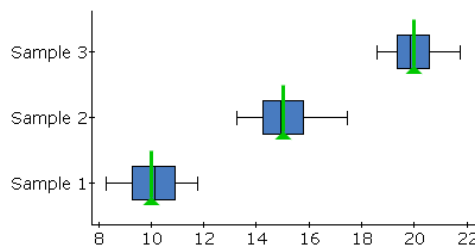
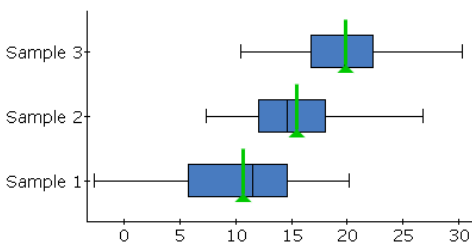
Let's talk a little more about how the test statistic is used to determine if there are differences between the means. Consider the following two boxplots:



Both figures compare four distributions – believe it or not, the sample means are the same in both figures (31,36, 38, and 31 respectively). So, why do the figures look so different?

The F-statistic is constructed from two different **variances** even though our goal is to compare several means.

Use the boxplots below to determine which data would result in the largest f-statistic. Which would result in the smallest?



Example: A science fair project sought to determine which material used in commercially-produced travel mugs did the best job of keeping beverages hot. Three different types of travel mugs were selected for the test: (1) plastic, (2) stainless steel, and (3) a combination of plastic and stainless steel. For the test, boiling water was poured into the mug being tested, and water temperature was measured. The mug's lid was then put in place. The lid was removed exactly 20 minutes later and the temperature of the water was again measured. The change in water temperature was recorded (in °F). The experiment was repeated 15 times for each of the 3 mugs.

- The mean temperature difference for the combination mug was -14.627°F . Explain what this quantity means in context.

- Conduct a *complete* analysis to determine if there is any evidence to support that the mean temperature change is different for any of the 3 travel mugs tested. Test the hypothesis using a 5% level of significance.

Column means

Column	n	Mean	Std. Error
Plastic	15	-17.433	1.034
Stainless Steel	15	-7.133	1.231
Combination	15	-14.627	1.038

ANOVA table

Source	df	SS	MS	F-Stat	P-value
Treatments	2	850.587	425.294		<0.0001
Error	42	768.876	18.307		
Total	44	1619.463			

Tukey 95% Simultaneous Confidence Intervals Plastic subtracted from

	Diff	Lower	Upper	P-value
Stainless Steel	10.3	6.504	14.096	<0.0001
Combination	2.807	-0.989	6.602	0.1831

Stainless Steel subtracted from

	Diff	Lower	Upper	P-value
Combination	-7.493	-11.289	-3.698	<0.0001

Using the boxplots below, discuss if the assumptions have been satisfied for the test which was conducted.

