F test example problems with solutions pdf

Years ago, statisticians found that when pairs of samples are taken from a normal population, the ratio of sample abnormalities in each pair will always follow the same distribution. Unsurprisingly, over the years, statisticians have found that the ratio of sample deviations collected in various ways follows the same distribution, F-distribution. Because we know that the distribution of the variance sampling is followed by a known distribution, we can conduct hypothesis tests using the variance ratio. The F-stats are simple: latexf s21 /s'2'2/latex, where s12 is the variance of sample 1. Remember that the sampling variance is that: latex2 (s2 (x - overlinex) (n-1) / latex Think of the form, If s12 and s22 are taken from samples of the same population, if you take a lot of pairs of samples and calculate F-scores, most of these F-points will be close to one. which will also be positive. Thinking about ratios requires some care. With most values around one, it's obviously not symmetrical; there is a long tail on the right, and a steep descent to zero on the left. There are two uses of F-distribution that will be discussed in this chapter. First, it is a very simple test to see if the two samples come from populations with the same variance. Second, it is a single-1G variance analysis (ANOVA) that uses F-distribution to test to see if three or more samples come from populations with the same average. Simple test: Do these two samples come from populations with the same variance? Because F-distribution is generated by drawing two samples from the same normal population, it can be used to test the hypothesis that two samples are taken from populations with the same deviations. You would have two samples (one n1 size and one n2 size) and a variance sample from each. Obviously, if these two deviations are very close to being equal, these two samples could easily come from groups with equal deviations. Because F-stats are a ratio of two sample deviations, where two sampling deviations are close to equal, the F-score is close to one. If you calculate an F-account and it is close to one, you accept your hypothesis that samples come from populations with the same variance. This is the basic method of the F-test. Suppose the samples come from populations from that Dispersion. Calculate the F-score by finding a sample ratio If the F-score is close to one, conclude that your hypothesis is correct and that the samples come from populations with equal deviations. If the F-score is far from one, then conclude that the population probably have different differences. The basic method should be fleshed out with some details if you are going to use this test at work. There are two sets of details: first, formally writing hypotheses, and secondly, using F-distribution tables, so you can tell if your F-score is close to one or not. Formally, two hypotheses are needed for completeness. First, the zero hypothesis is that there is no difference (hence the zero). It's usually referred to as Ho. Secondly, there is a difference, and it is called an alternative, and is designated H1 or Ha. Using F-tables to decide how close to one is close enough to adopt a zero hypothesis (really formal statistics would say not to reject null) is quite difficult because the F-distribution tables are quite difficult. Before using the tables, the researcher must decide how much chance he or she is willing to accept that zero will be rejected when indeed so. The usual choice is 5 percent, or as statisticians say, α - 0.05. If more or less the chance is wanted, α can be varied. Choose α and go to the F-tables. The first notification is that there are several F-tables, one for each of several different levels of α (or at least a table for each of the two α with F-values for one α in bold and values for the other in the usual type). There are strings and columns on each F-table, and both for degrees of freedom. Since two separate samples are taken to calculate the F-score and the samples do not have to be the same size, there are two separate degrees of freedom, one for each sample. For each sample, the number of degrees of freedom is n-1, which is one less than the sample size. When you move to the table, how do you decide which degrees of sample freedom (df) are for a line and which are for a column? While you can put any of them anywhere, you can save yourself a step if you place a sample with a larger variance (not necessarily a larger sample) in the numerator, and then that sample df identifies the column and df another sample determines the string. The reason why this saves a step is that the tables show only the F values that leave α in the right tail, where the F zgt;1, the image at the top of most F-tables shows that. Finding a critical F-point for left tails requires another step, which is outlined in the interactive Excel template in figure 6.1. Simply change the numerator and the denominator of the degree of freedom, and α in the right tail of the F-distribution in the yellow cells. Figure 6.1 Interactive Excel F-Table Pattern - see Appendix 6. F-tables are almost always printed as single-tailed tables, The F-value separates the right tail from the rest of the distribution. In most statistical F-distribution applications, only the right tail is of interest because most applications are tested to see if the deviation from a particular source is greater than the variance from another source, so the researcher is interested in finding if the F-score is larger than one. In the peer-to-peer deviation test, the researcher is interested in figuring out if the F-score is close to one, so either a large F-score or a small F-score will lead the researcher to conclude that the deviations are not equal. Because the critical F that separates the left tail from the rest of the distribution is not printed, not just the negative value of the printed, the researchers often simply divide the larger sampling variance into a smaller sample variance and use printed tables to see whether the more than one ratio is effectively rigged into a one-tail format. For purists, and sometimes cases, left-tailed critical value can be calculated fairly easily. Left-tailed critical value for x, y degrees of freedom (df) is simply reverse right tail (table) critical value for y, x df. Looking at the F-table, you will see that the F-value that leaves α - 0.05 in the right tail when there are 10, 20 df is F.2.35. To find an F-value that leaves α - 0.05 in the left tail with 10, 20 df, look up F'2.77 for α - .05, 20, 10 df. Divide one into 2.77 by finding .36. This means that 5 percent of F-distribution for 10, 20 df is below the critical value of 0.36, and 5 percent above the critical value of 2.35. Putting it all together, here's how to conduct a test to see if the two samples come from populations with the same variance. First, collect two samples and calculate the variance sample of each, s12 and s22. Second, write your hypotheses and choose α. The third is to find the F-account of your samples by dividing more s2 into smaller ones, so that's the F'gt;1. Fourth, go to the tables, find a table for α/2 and find a critical (table) F-score for the correct degree of freedom (n-1 and n-1). Compare it to the F-rated samples. If F samples are more critical F, F samples are not close to one, and Ha differences in population are not equal, is the best hypothesis. If the F samples are smaller than the critical F, Ho, that the differences in the population are equal, should be taken. An example #1 Lin Xiang, a young banker, moved from Saskatoon, Saskatchewan, to Winnipeg, Manitoba, where she was recently promoted to the manager of City Bank, a newly established bank in Winnipeg with branches across the prairie. A few weeks later, she discovered that maintaining the right number of writers seems to be more difficult than it was when she was an assistant manager at a branch in Saskatoon. Some days very long, but on other days, the teters seem to have little to do. She wonders if the number of customers in her new branch is just more variable than the number of customers in the branch where she worked. Because the tellers work all day or afternoon (morning or afternoon), it collects the following data on the number of transactions per day from its branch and branch where it worked: Winnipeg branch: 156, 278, 134, 202, 236, 198, 187, 199, 143, 165, 223 Suscathun Branch: 345, 332, 309, 367, 388, 312, 355, 363, 381 It assumes: H_o: sigma 2_W sigma 2_S/latex (latex) H_a: Sigma 2_W Sigma 2_S/latex) She decides to use α - .05 On this. She calculates the sampling deviations and finds: latex 2_W (1828.56)/latex latex (latex) 2_S 795.19/latex After the rule, to put a large variance in the numerator, so it saves a step, she finds: latexFs2_W/s'2_S 1828.56/795.19 2.30/latex Figure 6.2 Interactive Excel template for F-Test - see Appendix 6. Using the interactive Excel pattern in figure 6.2 (and not forgetting to use the α table - .025, because the table is one-tailed and the test is two-tailed), it considers that the critical F for 10.8 df is 4.30. Since her F-rated score with a figure of 6.2 is less than a critical score, she concludes that her F-score is close to one, and that the customer variance in her office is the same as it was in the old office. She will need to look further to solve her personnel problem. Dispersion Analysis (ANOVA) A more important use of F-distribution is to analyze the variance to see if three or more samples come from populations with equal means. This is an important statistical test not so much because it is often used, but because it is a bridge between non-offensive statistics and multivariate statistics, and because the strategy it uses is used in many multivariate tests and procedures. One-way ANOVA: Are all three (or more) samples taken from populations with the same average? This seems wrong - we will test the hypothesis about the means, analyzing the variance. This is not the case, but a very clever understanding of what some stats were years ago. This idea - looking at variance to learn about differences in tools - is the basis for much of the multivariate statistics used by researchers today. ANOVA ideas are used when we look at the relationship between two or more variables, a big reason why we use multivariate statistics. Testing to see if three or more samples come from populations with the same average can often be a kind of multivariate exercise. If three samples came from three different plants or were the subject of different procedures, we effectively see if there is a difference in results due to different plants or procedures - whether there are between the plant (or (or And the result? Think of three samples. Group x have been assembled, and for some good reasons (except for their x value) they can be divided into three groups. You have x from the group (sample) 1, some from the group (sample) 2, and some from the group (sample) 3. If the samples were combined, it would be possible to calculate a large average and complete discrepancy around this great average. You can also find between and (sample) variances in each group. Finally, you can take three samples of the tools, and find a variance between them. ANOVA is based on an analysis of where the general variance comes from. If you chose one x, the source of its variance, its distance from the large average, would have two parts: (1) how far it is from the average of its sample, and (2) how far the average of its sample is from the large average. If the three samples do come from populations with different means, then for most x's, the distance between the average sample and the great average is likely to be greater than the distance between x and its average groups. When these distances are come together and turned into deviations, you can see that if the population means different, the difference between the sample means will probably be greater than the variance in the samples. By this point in the book, it shouldn't surprise you to learn that statisticians have found that if three or more samples are taken from a normal population, and the difference between the samples is divided into variance in samples, the distribution of the sample is formed by doing so over and over again will have a known form. In this case it will be distributed as F with m-1, n-m df, where m is the number of samples and n size m samples in general. The difference between is: where xj is the average J sample, and x is a great average. The variance number between the sum of the distance squares between the average sample of each x and the grand average. It's just summing up one of these sources of variance in all observations. The difference inside is: Double amounts should be handled with care. First (working on the internal or second amount mark) find the average of each sample and the amount of the distances of each x in the sample from its average. Second (works on the external amount mark), mix the results from each sample. The strategy for one-to-one variance analysis is simple. Collect samples m. Calculate the variance between the samples, the variance in the samples and the ratio between them and the inside, which will give an F-assessment. If the F-score is less than one or not much larger than one, the difference between the samples is no greater than the difference in samples and the samples probably come from populations with the same average. If the F-score is much larger than one, the difference between The source of most of the variance in the overall sample, and the samples probably come from populations with different means. Details of one-way ANOVA fall into three categories: (1) writing hypotheses, (2) maintaining calculations organized, and (3) using an F-table. The zero hypothesis is that all means for the population are equal, and the alternative is that not all funds are equal. Often enough, although two hypotheses are really necessary for completeness, only Ho is written: latex H_o: m_1m_2'ldots'm_m/latex) Preservation of organized calculations is important when you find variance inside. Remember that the variance inside is by quadring and then summing up, the distance between each observation and the average sample of it. While different people do calculations differently, I believe the best way to keep it all straight is to find sample remedies, find square distances in each of the samples, and then add them together. It is also important that the calculations are organized in the final calculations of the F-account. If you remember that the goal is to see if the difference between the big one, then it is easy to remember to split the variance between the variance inside. Using F-tables is the third detail. Remember that F-tables are single-tail tables and that ANOVA is a one-tail test. Although the zero hypothesis is that all means are equal, you test this hypothesis, seeing if the difference between them is smaller or equal to the variance inside. The number of degrees of freedom m-1, nm, where m is the number of samples and n - the total size of all samples combined. An example of #2 A young bank manager in Example 1 is still struggling to find the best way to get the best possible way for the staff of his branch. She knows she should have more thinkers on Fridays than on other days, but she tries to find if the need for tact is constant for the rest of the week. It collects data on the number of transactions each day for two months. Here is her data: Monday: 276, 323, 298, 256, 277, 309, 312, 265, 311 Tuesdays: 243, 279, 301, 285, 274, 243, 228, 298, Wednesdays 255 Wednesdays: 288, 292, 310, 267, 243, 293, 255, 273 Thursdays: 254, 279, 241, 227, 278, 276, 256, 262 It checks zero hypothesis: latex: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: H_o: The tua m_w m_th/latex is m_m m_ and decides to use the α -05. She finds: m 291.8 that 267.3 in 277.6 thousand 259.1 and a large average of 274.3 She calculates the variance inside: (276-291.8)2 (323-29 1.8)2... (243-267.6)2... (288-277,6)...» (254-259.1)2/34-4/30-529.6 Then it calculates the variance between: 9 (291.8-274,3)2'9 (267.3-274.3)2'8 (277.6 -274.3)3)3 2'8 (259.1-274.3)2'/4-1' - 5151.8/3 - 1717.3 It calculates its F-account: Figure 6.3 Interactive Excel template for One-Way ANOVA - see Appendix 6. You can enter the number of transactions each day in the yellow cells in figure 6.3 and choose α. See 6.3, the calculated F is 3.24, while the F-Critical for α is .05 and 3, 30 df is 2.92. Because its F-rating is greater than the critical F-value, or as an alternative, since the p-value (0.036) is less than α - .05, it concludes that the average number of trades is not equal on different days of the week, or at least there is one day that differs from others. She will want to adjust her staffing schedule so that she has more of those on some days than on others. The F-distribution summary is the distribution of the sample ratio of variances of two samples taken from a normal population. It is used directly to test to see if the two samples come from populations with the same variance. While you will sometimes see it used to test the equality of deviations, the more important is the use in variance analysis (ANOVA). ANOVA, at least in its simplest form, presented in this chapter, is used for testing to see if three or more samples come from populations with the same average. By checking whether the difference in observations is greater from the variation of each observation from the average sample or from variations of means samples from the large medium, ANOVA is testing to see if the samples come from populations with equal means or not. It is the basis for regressive analysis, a statistical method that has many business applications; it is covered in later chapters. F-tables are also used in testing regression hypotheses. This is also the beginning of multivariate statistics. Note that in one side of the ANOVA, each observation is for two variables: x variable and group from which the observation is part. In later chapters, observations will have two, three or more variables. The F-test for variance equality is sometimes used before using a t-test for equality of funds, since t-test, at least in the form presented in this text, requires that samples come from populations with equal deviations. You will find that it is used along with t-tests when the stakes are high or the researcher is a bit compulsive. Compulsive. f test example problems with solutions pdf