**Using One-way ANOVA’s**

**to Calculate p-values with Growth Rate Data**

**Portia Mira**

**Last Updated: February 3, 2014**

One-way analysis of variances (abbreviated ANOVA’s) is a technique that compares the means of two or more samples in a data set using the F-distribution. Here, we will demonstrate how to calculate the F-ratios, and p-values, with growth rate data obtained from E. coli strain DH5αE collected between May 2013-November 2013.

The null hypothesis that we will be testing is:

H0=The growth rate is not significantly different among 16 variants given an antibiotic treatment (CAZ).

We will be using data from the two strains TEM-1 (wild type, 0000) and M69L (1000) treated with the antibiotic Ceftazidime (CAZ) at 0.1 μg/mL. These strains are from the resistance gene TEM-50. The data we have already obtained and will be using to calculate the p-values here is as follows:

|  |  |
| --- | --- |
| TEM-1 | M69L |
| 0.0025 | 0.00029 |
| 0.00218 | 0.00027 |
| 0.00202 | 0.00029 |
| 0.00226 | 0.00024 |
| 0.0023 | 0.00028 |
| 0.00227 | 0.00026 |
| 0.00236 | 0.00037 |
| 0.00167 | 0.00029 |
| 0.00172 | 0.00029 |
| 0.002 | 0.0003 |
| 0.00232 | 0.00029 |
| 0.00201 | 0.00029 |

-Table 1-

There are 12 replicates of each strain that was run during the same experiment. Here we want to measure the impact of the drug concentrations on each strain, and between strains.

One of the equations that we will be using to calculate the F-ratio is:

Total Sum of Squares = Sum of Squares Within Groups + Sum of Squares Between Groups (1)

SST = SSW + SSB

The first thing we need to calculate is the Sum of Squares Within Groups (SSW). To calculate, this we will first calculate the mean of each group:

|  |  |
| --- | --- |
| TEM-1 | M69L |
| 0.002134167 | 0.000288333 |

-Table 2-

Next, subtract this mean from each individual data point, and square the difference:

(For simplicity, I will continue each calculation in detail with just TEM-1 data; the same is done for M69L)

|  |  |  |
| --- | --- | --- |
| TEM-1 | Mean | (TM1 - Mean)2 |
| 0.0025 | 0.002134167 | 1.33834E-07 |
| 0.00218 | 0.002134167 | 2.10066E-09 |
| 0.00202 | 0.002134167 | 1.30341E-08 |
| 0.00226 | 0.002134167 | 1.58339E-08 |
| 0.0023 | 0.002134167 | 2.75006E-08 |
| 0.00227 | 0.002134167 | 1.84506E-08 |
| 0.00236 | 0.002134167 | 5.10005E-08 |
| 0.00167 | 0.002134167 | 2.15451E-07 |
| 0.00172 | 0.002134167 | 1.71534E-07 |
| 0.002 | 0.002134167 | 1.80008E-08 |
| 0.00232 | 0.002134167 | 3.45339E-08 |
| 0.00201 | 0.002134167 | 1.54174E-08 |
|  | **SUM** | **7.16692E-07** |

-Table 3-

Here is a table showing the Sum of Squares for each of the groups:

|  |  |
| --- | --- |
| TEM-1 | M69L |
| 7.16692E-07 | 1.03667E-08 |

-Table 4-

To calculate the Sum of Squares Within Groups (SSW), we will take the sum of these two numbers:

7.16692E-07 + 1.03667E-08 = 7.270587E-07

**SSW = 7.270587E-07** (2)

The next thing we will calculate is the Total Sum of Squares (SST). Here, we will treat all the data points as one group and perform the same steps as described above for each individual group. So we will take the total mean (of all data points within TEM-1 and M69L), subtract each individual data point from the mean, and square the difference:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Group (G) | Mean (M) | (G-M)2 |
| TEM-1 | 0.0025 | 0.00121125 | 1.66E-06 |
| TEM-1 | 0.00218 | 0.00121125 | 9.38E-07 |
| TEM-1 | 0.00202 | 0.00121125 | 6.54E-07 |
| TEM-1 | 0.00226 | 0.00121125 | 1.10E-06 |
| TEM-1 | 0.0023 | 0.00121125 | 1.19E-06 |
| TEM-1 | 0.00227 | 0.00121125 | 1.12E-06 |
| TEM-1 | 0.00236 | 0.00121125 | 1.32E-06 |
| TEM-1 | 0.00167 | 0.00121125 | 2.10E-07 |
| TEM-1 | 0.00172 | 0.00121125 | 2.59E-07 |
| TEM-1 | 0.002 | 0.00121125 | 6.22E-07 |
| TEM-1 | 0.00232 | 0.00121125 | 1.23E-06 |
| TEM-1 | 0.00201 | 0.00121125 | 6.38E-07 |
| M69L | 0.00029 | 0.00121125 | 8.49E-07 |
| M69L | 0.00027 | 0.00121125 | 8.86E-07 |
| M69L | 0.00029 | 0.00121125 | 8.49E-07 |
| M69L | 0.00024 | 0.00121125 | 9.43E-07 |
| M69L | 0.00028 | 0.00121125 | 8.67E-07 |
| M69L | 0.00026 | 0.00121125 | 9.05E-07 |
| M69L | 0.00037 | 0.00121125 | 7.08E-07 |
| M69L | 0.00029 | 0.00121125 | 8.49E-07 |
| M69L | 0.00029 | 0.00121125 | 8.49E-07 |
| M69L | 0.0003 | 0.00121125 | 8.30E-07 |
| M69L | 0.00029 | 0.00121125 | 8.49E-07 |
| M69L | 0.00029 | 0.00121125 | 8.49E-07 |
|  |  | **SUM** | **2.12E-05** |

-Table 5-

**Total Sum of Squares (SST) = 2.12E-05**  (3)

The last thing we need to calculate is the Sum of Squares Between Groups (SSB). Based on equation (1) we can use simple algebra to calculate this value:

2.12E-05 = SSB + 7.270587E-07

**SSB = 2.04729413E-05** (4)

We can also calculate this number as follows:

Take the mean of each strain (Table 4) and subtract the group mean (Table 5); square the differences and sum them up. Then multiply this number by 12, because there are 12 individual measurements for each strain.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Strain Mean(SM) | Group Mean (GM) | (SM – GM)2 |
| TEM-1 | 0.002134167 | 0.00121125 | 8.52E-07 |
| M69L | 0.000288333 | 0.00121125 | 8.52E-07 |
|  |  | **SUM** | **1.70E-06** |

-Table 6-

1.7E-06 X 12 = 2.04E-05

Now we have all of the components needed:

Total Sum of Squares = Sum of Squares Within Groups + Sum of Squares Between Groups

2.12E-05 = 7.270587E-07 + 2.04729413E-05

**Final Calculations: F-Ratio**

Final calculations include dividing the sums of squares by degrees of freedom:

= = 2.04E-05 (5)

degrees of freedom(b) = Number of Groups – 1 = 2 – 1 = 1

= = 3.305E-08 (6)

degrees of freedom(w) = Measurements – Groups = 24 – 2 = 22

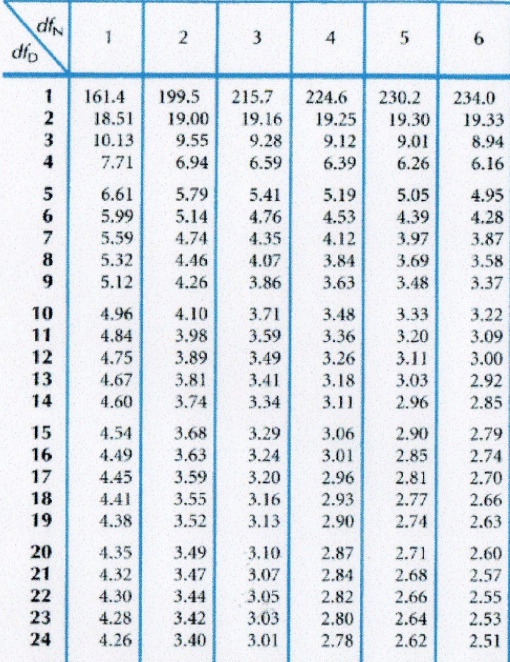
The F-Ratio is calculated by dividing (5) and (6):

F-Ratio = = = 618.18 (7)

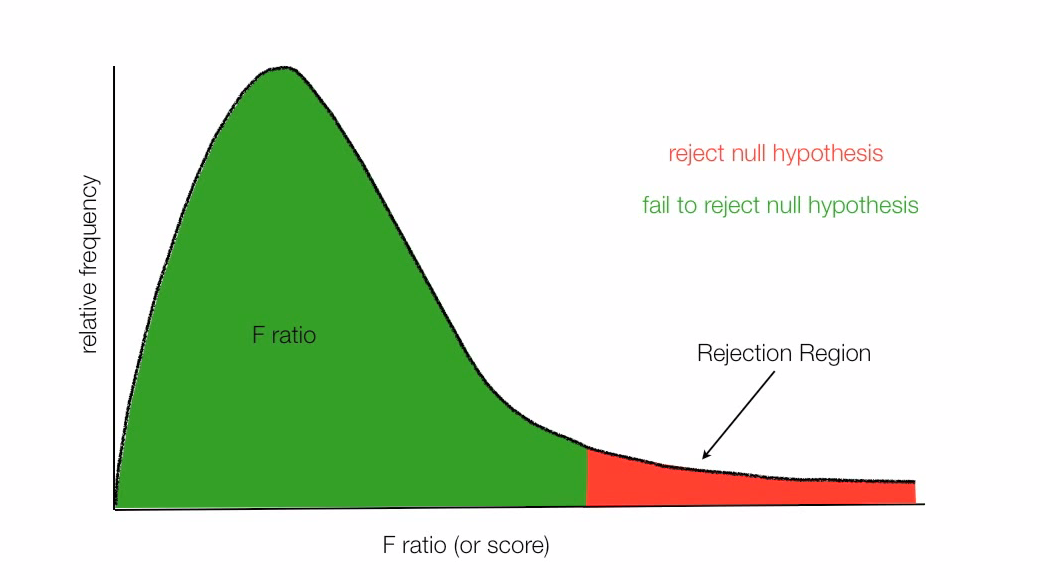
A critical value is obtained by using the F-ratio, the degrees of freedom in the numerator (dfN) of (7), and the degrees of freedom in the denominator (dfD) of (7). The F-ratio can also be written:

F(1,22) = 618.18

We use these coordinates to locate the critical values in the F-distribution table:



**Figure 1:** This figure shows the F-distribution table used to identify the critical values for given F-ratios by the degrees of freedom in the numerator (dfN) and the degrees of freedom in the denominator (dfD)

We can see that the critical value for this F-ratio is 4.30. This critical value indicates the location on the F-distribution curve (Figure 2) where the ‘rejection area’ is located. This rejection area shows where we can reject the null hypothesis, or where the p-value is less than, or equal to 0.05. 

**Figure 2**: This figure represents an F-distribution curve. Once the F-ratio and critical value are calculated, we can determine, based on where the F-ration lies, whether or not the null-hypothesis can be rejected or fail to be rejected (i.e. p-value ≤ 0.05). The line where green meets red is the critical value, if the F-ratio is greater than the critical value, then we can reject the null hypothesis (p ≤ 0.05). If the F-ratio is less than the critical value, then we fail to reject the null hypothesis (p ≥ 0.05).

Looking at Figure 1, the critical value of 4.30 is located where the green meets the red in the F-distribution graph, and the F-ratio (618.18) is much greater than 4.30. Here, the F-ratio lies to the far right of the critical value, or in the red region of the graph, indicating that we can reject the null hypothesis (p ≤ 0.05):

H0=The growth rate is not significantly different among 16 variants given an antibiotic treatment (CAZ).

The F-distribution curve is equal to the probability density function and can be calculated using the following equation:

f(x, d1, d2) = , (8)

where x ≥ 0, B is the Euler integral, and the parameters d1 and d2 are the degrees of freedom (DN and DD in our example) and are positive integers.

**References**

“How To Calculate and Understand Analysis of Variance (ANOVA)” *statisticsfun*.July 2012. <<https://www.youtube.com/user/statisticsfun?feature=watch>>

“Statistics for the Behavioral Sciences” Roger N. Morriessette, PhD. <<https://www2.palomar.edu/users/rmorrissette/Lectures/Stats/OneWayANOVA/OneWayANOVA.htm>>