F DISTRIBUTIONS AND SOME APPLICATIONS

1. Definition of F distributions; relation to t distributions

An F distribution with n_1 degrees of freedom in the numerator and n_2 in the denominator, or $F(n_1, n_2)$, is the distribution of $\frac{U/n_1}{V/n_2}$ where U and V are independent, U has a $\chi^2(n_1)$ distribution and V a $\chi^2(n_2)$ distribution.

It's easily seen that if T has a $t(d)$ distribution then T^2 has an $F(1, d)$ distribution.

2. Testing equality of variances in normal samples

Suppose $X_1, ..., X_m$ are i.i.d. $N(\mu, \sigma^2)$ and $Y_1, ..., Y_n$ are i.i.d. $N(\nu, \tau^2)$ and independent of the X_i. Let H_0 be the hypothesis that $\sigma^2 = \tau^2$. We can test H_0 using F distributions as follows. We know that $(m-1)s_X^2/\sigma^2$ has a $\chi^2(m-1)$ distribution, and likewise $(n-1)s_Y^2/\tau^2$ has a $\chi^2(n-1)$ distribution. If $\sigma^2 = \tau^2$, then s_X^2/s_Y^2 has an $F(m-1, n-1)$ distribution and s_Y^2/s_X^2 an $F(n-1, m-1)$ distribution. Thus we would reject H_0 in a 2-sided test at level α if either ratio of sample variances is larger than the $1-(\alpha/2)$ quantile of the respective F distribution. Rice, Table 5, tabulates some quantiles. For $\alpha = .05$ one would need 0.975 quantiles which Rice gives on p. A12.

Now suppose we have k different, independent samples with $k \geq 3$. For convenience, let's assume that each sample contains the same number m of elements, so that for each $i = 1, \ldots, k$, $X_{i1}, X_{i2}, \ldots, X_{im}$ are i.i.d. $N(\mu_i, \sigma_i^2)$. Suppose we want to test the hypothesis H_0 that all the σ_i^2 are equal. For each pair (i, i') with $1 \leq i < i' \leq k$, we could test the hypothesis that $\sigma_i^2 = \sigma_{i'}^2$ as in the last paragraph, using ratios of s_i^2's where s_i^2 is the sample variance of the ith sample. Then we would be doing $\binom{k}{2} = k(k-1)/2$ 2-sided tests. If we set the size (significance level) of each individual test at some $\gamma > 0$, and reject H_0 if any one of the hypotheses $\sigma_i^2 = \sigma_{i'}^2$ is rejected, then if H_0 is true, the probability of rejecting it is $\leq \binom{k}{2}\gamma$. For this to equal a predetermined α such as 0.05, we can set $\gamma = \alpha/\binom{k}{2}$. (This is called a Bonferroni correction.) One does not actually need to perform $\binom{k}{2}$ tests. Let $S^2 = s_i^2$ be the largest

Date: 18.650, Oct. 17, 2015.

of the k sample variances and $s^2 = s^2_{i'}$ the smallest. Then reject H_0 if $F := S^2/s^2$ is larger than the $1 - \gamma/2$ quantile of the $F(m-1, m-1)$ distribution, which can be found via R as qf$(1 - \gamma/2, m-1, m-1)$.

3. One-way analysis of variance (ANOVA)

. Here we're given observations X_{ij} for $i = 1, ..., I$ and $j = 1, ..., J$ where $I \geq 2$ and $J \geq 2$. The model, H_1 say, is that $X_{ij} = \mu_i + \varepsilon_{ij}$ where ε_{ij} are all i.i.d. $N(0, \sigma^2)$. For each i, the X_{ij} could be tested by the Shapiro–Wilk test for being i.i.d. normal with the same mean and variance, using a Bonferroni correction for I tests: is the smallest p-value less than α/I, where $\alpha = 0.025$ say. If normality is not rejected one could then test for equality of the variances for different i, also with a Bonferroni correction with $\gamma = \alpha/\binom{I}{2}$ as above, again with $\alpha = 0.025$. (The default $\alpha = 0.05$ is divided in two in a Bonferroni correction, one half for the Shapiro–Wilk tests of normality and the other half for the F-tests for equal variances.) If equality of variances is also not rejected one could decide not to reject H_1 and then proceed with the following traditional analysis of variance.

Let $\mu = \frac{1}{I}\sum_{i=1}^{I} \mu_i$ and $\alpha_i = \mu_i - \mu$. Then $\sum_{i=1}^{I} \alpha_i = 0$ and $X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$. Here α_i is the ith treatment effect. In one-way ANOVA, the hypothesis H_A is that in addition to H_1, $\alpha_i = 0$ for all i, in other words all μ_i are equal to the same μ, is tested. For each i let $X_{i\cdot} := \frac{1}{J}\sum_{j=1}^{J} X_{ij}$. Let SSE $:= \sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - X_{i\cdot})^2$. Here SSE abbreviates "sum of squared errors" although it is not the true errors ε_{ij} but rather estimates of them that are being squared and summed. Under H_1, SSE$/\sigma^2$ has a $\chi^2(I(J-1))$ distribution. Under H_A, since $\alpha_i \equiv 0$, the $X_{i\cdot}$ are i.i.d. $N(\mu, \sigma^2/J)$ and therefore for

$$X_{\cdot\cdot} = \frac{1}{I}\sum_{i=1}^{I} X_{i\cdot} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij},$$

$SSA = J\sum_{i=1}^{I}(X_{i\cdot} - X_{\cdot\cdot})^2/\sigma^2$ has a $\chi^2(I-1)$ distribution. Next it will be shown that under H_A, SSA and SSE are independent. This is an extension of the theorem that for $X_1, ..., X_n$ i.i.d. $N(\mu, \sigma^2)$, $\overline{X}$ and s_x^2 are independent, as seen early in the course. The variables X_{ij} are independent for different i. For each i, $X_{i\cdot}$ is independent of $\sum_{j=1}^{J}(X_{ij} - X_{i\cdot})^2$ by the theorem just mentioned. Thus under H_A, $\frac{SSA/(I-1)}{SSE/(I(J-1))}$ has an $F(I-1, I(J-1))$ distribution. So H_A will be rejected if this ratio is too large. This is a one-sided test, so for level $\alpha = 0.05$ one would use the 0.95 quantile.

If H_A is rejected, so one decides that not all μ_i are equal, one can then ask which μ_i are significantly different from one another. Comparing pairs $X_{i\cdot}$ for different values of i, there is again a Bonferroni correction. In this case one is not only interested in whether the largest sample mean is significantly different from the smallest but whether other means may differ significantly. (Actually, that could be true for variances also.) Rice, Table 7, gives a table for the "Bonferroni t statistic."

4. Regression

Suppose we have observed (x_j, Y_j), $j = 1, ..., n$, where x_j are design points, and we have a general linear regression model

$$Y_j = \beta_0 + \sum_{i=1}^{k} \beta_i f_i(x_j) + \varepsilon_j \tag{1}$$

where f_i are some given functions of x, such as $f_i(x) = x^i$, and ε_j are assumed to be i.i.d. $N(0, \sigma^2)$ for some unknown $\sigma > 0$. The functions f_i, including $f_0 \equiv 1$, restricted to the set of x_j, are assumed to be linearly independent, which requires that $n \geq k + 1$. The model is called a linear model because it is linear in the β_i although the functions f_i can be rather general. By minimizing

$$\sum_{j=1}^{n} (Y_j - \beta_0 - \sum_{i=1}^{k} \beta_i f_i(x_j))^2$$

we get estimates $\hat{\beta}_i$ of the β_i. If $n = k + 1$ one can get an exact fit (for example in simple linear regression with $n = 2$) so in practice we will require that $n > k + 1$, preferably by a substantial margin.

Recalling the residuals $\hat{\varepsilon}_j = Y_j - \hat{\beta}_0 - \sum_{i=1}^{k} \hat{\beta}_i f_i(x_j)$, let $\text{SSE} = \sum_{j=1}^{n} \hat{\varepsilon}_j^2$. If the model (1) holds, then SSE/σ^2 has a $\chi^2(n - k - 1)$ distribution. (The degrees of freedom equal n minus the number of parameters estimated from the data.) Similarly as in analysis of variance, one can show that under (1), differences $\hat{\beta}_i - \beta_i$, suitably normalized, have $t(n - k - 1)$ distributions. As you may have noticed, the R output for linear models mentions both t and F distributions, where t tests are used for whether $\hat{\beta}_i$ are significantly different from 0, which we used for the quadratic coefficient in quadratic regression to test the simple linear regression hypothesis. An F test is used to test whether $\beta_1 = \cdots = \beta_k = 0$, which seems to be a less interesting hypothesis.

5. Historical notes

The leading statistician R. A. Fisher (1890–1962) invented analysis of variance. According to Miller, he tabulated probabilities not for what is now called an F statistic but for $z = (\log F)/2$. Snedecor (1934) gave the name "F distribution" in honor of Fisher and tabulated F itself.

From the Wikipedia article on Fisher, he worked at the Rothamsted Experimental Station in England from 1919 to 1933, designing and performing agricultural experiments and analyzing them statistically. He published books and became a world-famous statistician during that time. In 1931 and later in 1936 he visited Iowa State College in Ames, Iowa, where there was also great interest in agricultural experiments and statistics. Fisher met Snedecor there.

According to Wikipedia on Snedecor, the book by Snedecor and Cochran (1937 and later editions) was at least for a while the most often cited book in all of science. On "F-distribution" it's said sometimes to be called "Snedecor's F distribution" or the "Fisher–Snedecor distribution," but "F distribution" is the standard terminology in textbooks.

Snedecor lived from 1881 to 1974 and Cochran from 1909 to 1980. So even the 8th edition of their book (1989) was posthumous for both of them.

REFERENCES

*Koehler, Kenneth J., *Snedecor and Cochran's Statistical Methods*, projected publication Oct. 10, 2016 (according to Amazon).

Miller, Jeff, maintainer. Website "Earliest known uses of some of the words of mathematics," accessed Oct. 25, 2014, then said to be last revised Oct. 17, 2014. The whole site has a great many entries and contributors. The entry "F distribution" describes Fisher's and Snedector's contributions. Miller is said to be a teacher at Gulf High School in New Port Richey, Florida.

*Snedecor, George W. (1934), *Calculation and Interpretation of Analysis of Variance and Covariance*, Collegiate Press, Ames, Iowa. A Google Book.

*Snedecor, George W., and Cochran, William G., *Statistical Methods*, Blackwell, Ames, Iowa, first ed. 1937, 8th ed. 1989, 9th ed.: see Koehler.

* – I have not seen these items in the original.