The t and F distributions
Math 218, Mathematical Statistics
D Joyce, Spring 2016

Student's t-distribution and Snedecor-Fisher's F-distribution. These are two distributions used in statistical tests. The first one is commonly used to estimate the mean μ of a normal distribution when the variance σ^2 is not known, a common situation. The second is a rather special purpose distribution used to estimate the ratio of the variances of two normal distributions.

Student's t-distribution. We know that if a sample is drawn from a normal distribution with mean μ and variance σ^2 that the scaled sample average

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal distribution. That information can be used to determine just how good $\overline{X}$ is as an estimator for the unknown value μ. Unfortunately, σ^2 is also an unknown. (At least n is known.) So, what can we do?

Student (the nom de plume of William Sealey Gosset (1876–1937)) determined that the unknown σ could be replaced by the sample standard deviation S, which is the square root of the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

That results in a statistic that looks something like the scaled sample average, but isn't identical. It's the variable T defined by

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

The distribution of T is not a standard normal distribution, although for large n, and even moderately sized n, it's very close to a standard normal distribution. The distribution of T is called *Student's t-distribution*. On page 180 of the text, the standard normal distribution is drawn along with Student's t-distribution for $n = 2$ and $n = 10$. Since for large values of n, the t-distribution is so close to the standard normal distribution, the T-distribution is only needed for n small, say $n \leq 30$.

It turns out that the ratio between T and Z (the scaled sample mean described above) is the square root of a scaled χ^2 distribution. Precisely,

$$\frac{Z}{T} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \bigg/ \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{S}{\sigma},$$

and, as we saw last time, $S^2 \sim \dfrac{\sigma^2\chi^2_{n-1}}{n-1}$, so S/σ is the square root of $\dfrac{S^2}{\sigma^2} \sim \dfrac{\chi^2_{n-1}}{n-1}$. That's enough information to compute a t-distribution. Table A.4 has critical values for the t-distribution.

We'll work through example 5.6 in class.

Snedecor-Fisher's F-distribution. This distribution is used to estimate the ratio of the variances of two normal distributions. If we have two random samples, the first $X_1, \ldots, X_{n_1}$ from a $N(\mu_1, \sigma_1)$ distribution, while the second $Y_1, \ldots, Y_{n_2}$ from a $N(\mu_2, \sigma_2)$ distribution, then we know their sample variances are scaled χ^2 distributions. That is,

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2}{n_1 - 1} \sim \frac{\sigma_1^2\chi^2_{n_1-1}}{n_1 - 1}$$

and

$$S_2^2 = \frac{\sum_{i=1}^{n_2}(Y_i - \overline{X})^2}{n_2 - 1} \sim \frac{\sigma_2^2\chi^2_{n_2-1}}{n_2 - 1}.$$

Therefore, for the ratio of these sample variances we have

$$\frac{S_1^2}{S_2^2} \sim \frac{\sigma_1^2\chi^2_{n_1-1}}{n_1 - 1} \bigg/ \frac{\sigma_2^2\chi^2_{n_2-1}}{n_2 - 1},$$

and so

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \frac{\chi^2_{n_1-1}/(n_1 - 1)}{\chi^2_{n_2-1}/(n_2 - 1)}.$$

The random variable on the right, a scaled ratio of two χ^2 distributions, has what is called a Snedecor-Fisher F-distribution with $n_1 - 1$ degrees of freedom in the numerator and $n_2 - 1$ degrees of freedom in the denominator. The notation for an F-distribution with ν_1 and ν_2 degrees of freedom is F_{ν_1,ν_2}. Table A.6 has critical values for this F distribution.

Statistical tests based on the F-distribution can determine if the variances σ_1 and σ_2 of two normal distributions are the same by looking at the ratio of two sample variances S_1^2/S_2^2.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`