Relation between the χ^2, F and t distributions

The χ^2 distribution

If Z~N(0,1) is a standard normal variable, then

Z^2 has the χ^2 distribution with 1 degree of freedom.

If X_1, X_2 are *independent* χ^2 variables with m and n degrees of freedom respectively, then X_1+X_2 has the χ^2 distribution with m+n degrees of freedom. In particular, if $Z_1,\ldots,Z_n$ are independent samples from a standard normal distribution, then

$\sum_{i=1}^{n} Z_i^2$ has the χ^2 distribution with n degrees of freedom.

Typical example

If $X_1,\ldots,X_n$ are independent samples from a normal distribution, $X \sim N(\mu,\sigma^2)$, then
$(X-\mu)/\sigma \sim N(0,1)$, and

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Note that one degree of freedom is lost due to taking the sample mean.

For our purposes, the following is particularly relevant:

If $Y_i = \alpha + \beta X_i + u_i$ with $u_i \sim N(0,\sigma^2)$ follows the assumptions of the classical linear regression model, then

$\sum_{i=1}^{n} \frac{\hat{u}_i^{\,2}}{\sigma^2}$ =RSS/σ^2 has the χ^2 distribution with n-2 degrees of freedom

(Where $\hat{u}_i = \hat{\alpha} + \hat{\beta} X_i$).

In general, if we have k explanatory variables, $X_1,\ldots,X_k$ in our regression model, then $\sum_{i=1}^{n}\frac{\hat{u}_i^{\,2}}{\sigma^2} \sim \chi^2_{n-k-1}$.

<u>Relationship between the t-statistic and the χ^2 statistic</u>

The t-distribution is initially defined in terms of the standard normal and the χ^2 distribution.

Let $Z\sim N(0,1)$, and let $X\sim\chi^2_n$, with the two variables independent.

Then $\frac{Z}{\sqrt{X/n}}$ has the t-distribution with n degrees of freedom.

Example: Let $X_1,\ldots,X_n$ be independent samples from a normal distribution, $X_i\sim N(\mu,\sigma^2)$.

Then $\frac{(\bar{X}-\mu)}{(\sigma/\sqrt{n})}\sim N(0,1)$

While $\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.

It can be shown that these two variables are independent. Hence,

$$\frac{(\bar{X}-\mu)}{(\sigma/\sqrt{n})}\Bigg/\frac{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)}}{\sigma} = \frac{(\bar{X}-\mu)}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)}\Big/\sqrt{n}}$$

$= \frac{(\bar{X}-\mu)}{\hat{\sigma}^2/\sqrt{n}} = \frac{(\bar{X}-\mu)}{S.E.(\bar{X})} \sim t_{n-1}$, which is exactly how we introduced the t-distribution in the first place.

For regression purposes, we know that the regression coefficient

$\hat{\beta}$ ~N(β,Var($\hat{\beta}$)), where Var($\hat{\beta}$)= $\dfrac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$

Hence $\dfrac{(\hat{\beta}-\beta)}{\sqrt{\sigma^2 \Big/ \sum (X_i - \overline{X})^2}}$ ~ N(0,1).

We also have that $\sum_{i=1}^{n} \dfrac{\hat{u}_i^{\,2}}{\sigma^2}$ ~ χ^2_{n-2} , and it can be shown that these two variables are independent.

Dividing the standard normal variable by the root of the χ^2 variable over its degrees of freedom, and cancelling the sigma, we get

$$\frac{(\hat{\beta}-\beta)}{\sqrt{(\sum \hat{u}_i^{\,2})/(n-2) \Big/ \sum (X_i - \overline{X})^2}} = \frac{(\hat{\beta}-\beta)}{\sqrt{\hat{\sigma}^2 / \sum (X_i - \overline{X})^2}} = \frac{(\hat{\beta}-\beta)}{S.E.(\hat{\beta})}$$

Has the t-distribution with n-2 degrees of freedom. A similar result holds for the k-variable case.

<u>The F-distribution</u>

Let X_1 and X_2 be independent χ^2 variables with n_1 and n_2 degrees of freedom respectively. Then $(X_1/n_1)/(X_2/n_2)$ has the F-distribution with (n_1,n_2) degrees of freedom.

The F-distribution can be used for comparing the variances of two normal distributions. In regression analysis, it is absolutely crucial, for testing *restrictions* on the regression model, and in particular, testing the restriction that <u>all the explanatory variables are insignificant</u>.

We have already seen that the RSS/σ^2, $\sum_{i=1}^{n} \dfrac{\hat{u}_i^{\,2}}{\sigma^2}$ ~ χ^2_{n-k-1}.

It can be shown that, in the classical regression model, the Explained Sum of Squares (ESS) divided by σ^2, that is $\frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sigma^2}$ has the χ^2 distribution with k degrees of freedom, under the assumption of the null hypothesis, that all the explanatory variables are insignificant, that is that the true values of $\beta_1,\ldots,\beta_k$ are zero.

Hence, dividing and cancelling the σ^2, we get that, under the null hypothesis

$$F = \frac{ESS/k}{RSS/(n-k-1)} \sim F(k,n-k-1)$$

We may therefore compare the resulting F statistic with the 95% or 99% etc. critical value of the appropriate F distribution. If the F-statistic exceeds this critical value, we may reject the null hypothesis that $\beta_1=\beta_2=\ldots=\beta_k=0$, and conclude that the regression as a whole is significant, that is that the variables $X_1,\ldots,X_k$ are jointly significant.