12 The F-test and analysis of variance

The statistician's attitude to variation is like that of the evangelist to sin; he sees it everywhere to a greater or lesser extent.

12.1 F-distribution

If $X \sim \chi^2_m$ and $Y \sim \chi^2_n$, independently of X, then

$$Z = (X/m)/(Y/n) \sim F_{m,n},$$

has the **F-distribution** with (or on) m and n degrees of freedom.

Note that if $T \sim F_{m,n}$ then $1/T \sim F_{n,m}$. Tables for the F-distribution usually only give the upper percentage points. If we want to know x such that $\mathbb{P}(T < x) = 0.05$ we can use $F_{n,m}$ tables to find $1/x$ such that $\mathbb{P}(1/T > 1/x) = 0.05$.

Note that if $X \sim t_n$ then $X^2 \sim F_{1,n}$. It is always nice to recognise connections between distributions.

12.2 Two samples: comparison of variances (F-test)

Suppose $X_1, \dots, X_m$ are IID $N(\mu_1, \sigma_1^2)$ and $Y_1, \dots, Y_n$ are IID $N(\mu_2, \sigma_2^2)$, with the two samples independent. It is required to test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 > \sigma_2^2$, with μ_1, μ_2 unknown nuisance parameters.

Now, by either the generalized likelihood ratio test, or common sense, we are led to consider the statistic

$$F = \frac{\hat\sigma_1^2}{\hat\sigma_2^2} = \frac{S_{XX}/(m-1)}{S_{YY}/(n-1)} \sim \frac{\sigma_1^2\chi^2_{m-1}/(m-1)}{\sigma_2^2\chi^2_{n-1}/(n-1)} = \frac{\sigma_1^2}{\sigma_2^2} F_{m-1,n-1}.$$

Thus, under H_0, $F \sim F_{m-1,n-1}$.

If H_1 is true, F will tend to be greater than when H_0 is true, so we reject H_0 if this ratio is large. A size α test is to reject H_0 if $f > F_\alpha^{(m-1,n-1)}$, the upper α point of $F_{m-1,n-1}$.

Example 12.1 *Suppose we wish to test the hypothesis that the variance of longevity is the same for male fruitflies kept with 1 interested or 1 uninterested female, i.e., $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_0 : \sigma_1^2 \neq \sigma_2^2$.*The test statistic is

$$f = (15.6525)^2/(14.9284)^2 = 1.099,$$

which, as $F_{0.05}^{(24,24)} = 1.98$, is not significant at the 10% level (the test is two-tailed). Notice that in order to use F tables we put the larger of $\hat\sigma_1^2$ and $\hat\sigma_2^2$ in the numerator.

12.3 Non-central χ^2

If $X_1, \dots, X_k$ are independent $N(\mu_i, 1)$ then $Z = \sum_{i=1}^k X_i^2$ has the **non-central chi-squared** distribution, $\chi^2_k(\lambda)$, with non-centrality parameter $\lambda = \sum_{i=1}^k \mu_i^2$. Note that $EW = k + \lambda$; thus a non-central χ^2_k tends to be larger than a central χ^2_k.

To see that it is only the value of λ matters, let A be an orthogonal matrix such that $A\mu = (\lambda^{1/2}, 0, \dots, 0)^\top$, so $(A\mu)^\top(A\mu) = \mu^\top\mu = \lambda$. Let $Y = AX$; then $\sum_{i=1}^k X_i^2 = \sum_{i=1}^k Y_i^2$, with $Y_1^2 = \chi_1^2(\lambda)$ and $\sum_{i=2}^k Y_i^2 = \chi^2_{k-1}$.

12.4 One way analysis of variance

Analysis of variance (ANOVA) is a technique for testing hypotheses about means by looking at sample variances. We consider here the question of testing equality of the means of $k > 2$ groups. The mathematical model is:

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, k.$$

Thus there are n_i observations in the ith group. Let $\sum_{i=1}^k n_i = N$.

It is assumed that the ϵ_{ij} are IID $N(0, \sigma^2)$, and that our data consists of observations x_{ij} which are realisations of random variables X_{ij} satisfying the model.

One-way ANOVA is used to test the null hypothesis $H_0 : \mu_1 = \dots = \mu_k$. The alternative hypothesis H_1 is 'H_0 is not true'. Application of the generalized likelihood ratio test gives

$$\begin{aligned} L_x(H_0, H_1) &= \frac{\max_{\mu_1,\cdots,\mu_k,\sigma^2}(2\pi\sigma^2)^{-N/2}\exp\left[-\sum_{ij}(x_{ij}-\mu_i)^2/2\sigma^2\right]}{\max_{\mu,\sigma^2}(2\pi\sigma^2)^{-N/2}\exp\left[-\sum_{ij}(x_{ij}-\mu)^2/2\sigma^2\right]} \\ &= \left[\frac{s_0}{s_1}\right]^{N/2}, \quad \text{where} \quad s_0 := \sum_{ij}(x_{ij} - \bar x_{\cdot\cdot})^2 \quad \text{and} \quad s_1 := \sum_{ij}(x_{ij} - \bar x_{i\cdot})^2. \end{aligned}$$

Here, $\bar x_{\cdot\cdot} = \sum_{ij} x_{ij}/N = \sum_i n_i \bar x_{i\cdot}/N$ is the overall mean (and the MLE of μ under H_0). Similarly, $\bar x_{i\cdot} = \sum_{j=1}^{n_i} x_{ij}/n_i$ is the mean within the ith group (and the MLE of μ_i under H_1).

Thus we are led to consider rejecting H_0 when s_0/s_1 is large. Now

$$\begin{aligned} s_0 &= \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar x_{i\cdot} + \bar x_{i\cdot} - \bar x_{\cdot\cdot})^2 \\ &= \sum_i \sum_j \left[(x_{ij} - \bar x_{i\cdot})^2 + 2(x_{ij} - \bar x_{i\cdot})(\bar x_{i\cdot} - \bar x_{\cdot\cdot}) + (\bar x_{i\cdot} - \bar x_{\cdot\cdot})^2\right] \\ &= \sum_{ij}(x_{ij} - \bar x_{i\cdot})^2 + \sum_i n_i(\bar x_{i\cdot} - \bar x_{\cdot\cdot})^2 \\ &= s_1 + s_2, \end{aligned}$$

where $s_2 := \sum_i n_i(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$, and thus s_0/s_1 is large when s_2/s_1 is large.

s_1 is called the **within samples sum of squares** and s_2 is called the **between samples sum of squares**.

Now, whether or not H_0 is true, $\sum_j (X_{ij} - \bar{X}_{i\cdot})^2 \sim \sigma^2\chi^2_{n_i-1}$, since $E(X_{ij})$ depends only on i. Hence, $S_1 \sim \sigma^2\chi^2_{N-k}$, since samples for different i are independent.

Also, $\sum_j (X_{ij} - \bar{X}_{i\cdot})^2$ is independent of $\bar{X}_{i\cdot}$, so that S_1 is independent of S_2. If H_0 is true $S_2 \sim \sigma^2\chi^2_{k-1}$, and if H_0 is not true, $S_2 \sim \sigma^2\chi^2_{k-1}(\lambda)$, where

$$E(S_2) = (k-1)\sigma^2 + \lambda, \quad \lambda = \textstyle\sum_{i=1}^k n_i(\mu_i - \bar{\mu})^2, \quad \bar{\mu} = \sum_i n_i\mu_i/N.$$

Intuitively, if H_0 is not true S_2 tends to be inflated.

So, if H_0 is true then $Q = \{S_2/(k-1)\}/\{S_1/(N-k)\} \sim F_{k-1,N-k}$, while if H_0 is not true, Q tends to be larger. So for a size α test we reject H_0 if $q > F_\alpha^{(k-1,N-k)}$.

An interpretation of this is that the variability in the total data set is $s_0 = \sum_{ij}(x_{ij} - \bar{x}_{\cdot\cdot})^2$. Under H_1 we expect x_{ij} to be about $\bar{x}_{i\cdot}$ and so a variability of $s_2 = \sum_{ij}(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$ is 'explained' by H_1. Statisticians say that H_1 'explains $(s_2/s_0)100\%$ of the variation in the data', (where since $s_0 = s_1 + s_2$, we must have $s_2/s_0 \le 1$.) If s_2/s_0 is near 1, or equivalently if s_2/s_1 is large, then H_1 does much better than H_0 in explaining why the data has the variability it does.

Example 12.2 *Partridge and Farquhar did experiments with five different groups of 25 male fruitflies. In addition to the groups kept with 1 interested or 1 uninteresed female, 25 males were each kept with no companions, and groups of 25 were each kept with 8 uninterested or 8 interested females. The 'compliance' of the males who were supplied with 8 virgin females per day varied from 7 inseminations per day at age one week to just under 2 per day at age eight weeks.*

Groups of 25 males kept with	mean life (days)	s.e.
no companions	63.56	16.4522
1 uninterested female	64.80	15.6525
1 interested female	56.76	14.9284
8 uninterested females	63.36	14.5398
8 interested females	38.72	12.1021

Suppose we wish to test equality of means in the three control groups, i.e., those kept with either no companions, or 1 or 8 uninterested females (rows 1, 2 and 4).

First we reconstruct the sums of squares,

$$\textstyle\sum_{j=1}^{25}(x_{1j} - \bar{x}_1)^2 = 24(16.4522^2) = 6496.16$$
$$\textstyle\sum_{j=1}^{25}(x_{2j} - \bar{x}_2)^2 = 24(15.6525^2) = 5880.00$$
$$\textstyle\sum_{j=1}^{25}(x_{4j} - \bar{x}_4)^2 = 24(14.5398^2) = 5073.76$$

then we calculate the within and between sums of squares,

$$\bar{x} = (63.56 + 64.80 + 63.36)/3 = 63.91$$
$$s_1 = 6496.16 + 5880.00 + 5073.76 = 17449.92$$
$$s_2 = \sum_{i=1,2,4} 25(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 = \sum_{i=1,2,4} 25\bar{x}_{i\cdot}^2 - 75\bar{x}_{\cdot\cdot}^2 = 30.427$$

and finally we compute the test statistic,

$$q = \frac{30.427/(3-1)}{17449.92/(75-3)} = 0.0628.$$

It is usual to display this data in an ANOVA table of the following form.

Source of variation	Degrees of freedom		Sum of squares		Mean square		F statistic	
Between groups	$k-1$	2	s_2	30.427	$s_2/(k-1)$	15.213	$\frac{s_2/(k-1)}{s_1/(N-k)}$	0.0628
Within groups	$N-k$	72	s_1	17449.92	$s_1/(N-k)$	242.36		
Total	$N-1$	74	s_0	17480.35				

The value of 0.0628 is not significant compared to $F_{0.05}^{(2,72)} = 3.12$ and hence we do not reject the hypothesis of equal means.

A similar test for equality of all five group means gives a statistic with value 507.5, to be compared to $F_{0.05}^{(4,120)} = 2.45$. Clearly we reject the hypothesis of equal means. It does seem that sexual activity is associated with reduced longevity.

ANOVA can be carried out for many other experimental designs. We might want to investigate more than one treatment possibility, or combinations of treatments. (E.g., in the fruitfly experiments each male fly was kept separate from other males; we might want to do experiments in which males are kept with different numbers of interested females and/or competing males.) If there are k possible treatments which can be applied or not applied, then 2^k different combinations are possible and this may be more than is realistic. The subject of 'experimental design' has to do with deciding how to arrange the treatments so as to gather as much information as possible from the fewest observations. The data is to be analysed to compare treatment effects and this typically involves some sort of ANOVA. The methodology is the same as for the one-way ANOVA considered above; we consider a normalised quotient, such as q above, between the reduction in the residual sums of squares that is obtained when moving from H_0 to H_1 (e.g., $s_0 - s_1$) and the value of the residual sum of squares under H_1 (e.g., s_1). In subsequent lectures we will see further examples of this idea in the context of regression models.