L E C T U R E 6

§7. HOTELLING DISTRIBUTION.

Hotelling's T-squared distribution (T^2) is a multivariate distribution proportional to the F-distribution and arises importantly as the distribution of a set of statistics which are natural generalizations of the statistics underlying Student's t-distribution. Hotelling's T-squared statistic (T^2) is a generalization of Student's t-statistic that is used in multivariate hypothesis testing.

Multivariate method that is the multivariate counterpart of Student's t and which also forms the basis for certain multivariate control charts is based on Hotelling's T^2 distribution, which was introduced by Hotelling (1947). Recall that

$$t = \frac{\overline{x} - \mu}{S/\sqrt{n}}$$

has a t-distribution, provided that η is normally distributed. If we wanted to test the hypothesis that $=_0$, we would then have

$$t = \frac{\overline{x} - \mu}{S/\sqrt{n}}$$

so that

$$t^2 = \frac{(\overline{x} - \mu)^2}{S^2/n} = n(\overline{x} - \mu)(S^2)^{-1}(\overline{x} - \mu).$$

When T^2 is generalized to m variables it becomes

$$T^2 = n(\overline{\mathbf{x}} - \mu_0)S^{-1}(\overline{\mathbf{x}} - \mu_0),$$

with

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ . \\ . \\ . \\ \bar{x}_m \end{bmatrix} \qquad \bar{\mu} = \begin{bmatrix} \mu_1^0 \\ \mu_2^0 \\ . \\ . \\ . \\ \mu_m^0 \end{bmatrix}.$$

S^{-1} is the inverse of the sample variance-covariance matrix, S, and n is the sample size upon which each $\overline{\mathbf{x}}$, $i = 1, 2, ..., m$, is based. (The diagonal elements of S are the variances and the off-diagonal elements are the covariances for the m variables).

The continuous probability distribution, concentrated on the positive semi-axis $(0, \infty)$, with density

$$f_k(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right) x^{k/2-1} \left(1+\dfrac{x}{n}\right)^{-(n+1)/2}}{\Gamma\left(\dfrac{n-k+1}{2}\right) \Gamma(k/2) \cdot n^{k/2}}. \tag{7.1}$$

depending on two integer parameters (the number of degrees of freedom) n and k. If a k-dimensional random vector Y has the normal distribution with null vector of means and covariance matrix Σ and if

$$S = \frac{1}{n} \sum_{i=1}^{n} Z_i' Z_i$$

where the random vectors Z_i are independent, distributed as Y and also independent of Y, then the random variable $T^2 = Y' S^{-1} Y$ has the Hotelling -distribution with n degrees of freedom (Y is a column vector and Y' means transposition). If $k = 1$, then

$$T^2 = \frac{Y^2}{\frac{1}{n}\chi_n^2} = t_n^2,$$

where the random variable t_n has the Student distribution with n degrees of freedom. If in the definition of the random variable T^2 it is assumed that Y has the normal distribution with parameters (ν, Σ) and Z_i has the normal distribution with parameters $(0, \Sigma)$, then the corresponding distribution is called a non-central Hotelling T^2-distribution with degrees of freedom and non-centrality parameter ν.

Hotelling's T^2-distribution is used in mathematical statistics in the same situation as Student's t-distribution, but then in the multivariate case. If the results of observations X_1, X_2, ...,X_n are independent normally-distributed random vectors with mean vector ν and non-degenerate covariance matrix Σ, then the statistic

$$T^2 = n(\overline{X} - \mu)' S^{-1} (\overline{X} - \mu),$$

where

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and

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

has the Hotelling T^2-distribution with $n-1$ degrees of freedom. This fact forms the basis of the Hotelling test. For numerical calculations one uses tables of the beta-distribution or of the Fisher F-distribution, because the random variable

$$\frac{n-k+1}{nk}T^2$$

has the F-distribution with k and $n-k+1$ degrees of freedom.

The Hotelling -distribution was proposed by H. Hotelling [1] for testing equality of means of two normal populations.

Let us verify, that the function in (7.1) is density function. Recall definition of the Beta-function:

$$B(m,n) = \int_0^1 x^{m-1}\,(1-x)^{n-1}\,dx.$$

It is well-known that the values of this function can be calculated using Γ-function:

$$B(m,n) = \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(n+m)}.$$

Therefore, using Beta-function, (7.1) we can rewrite in the following form:

$$\frac{1}{B\left(\frac{k}{2}, \frac{n-k+1}{2}\right)}\,\frac{x^{k/2-1}\left(1+\frac{x}{n}\right)^{-(n+1)/2}}{n^{k/2}}.$$

If we make the change of variables

$$x = \frac{y}{1+y} \quad \text{and} \quad dx = \frac{1}{(1+y)^2}\,dy$$

we get

$$B(m,n) = \int_0^1 x^{m-1}\,(1-x)^{n-1}\,dx = \int_0^{+\infty} y^{m-1}\,(1+y)^{-(m+n)}\,dy.$$

Therefore,

$$\int_0^{+\infty} f_k(x)\,dx = \frac{1}{B\left(\frac{k}{2}, \frac{n-k+1}{2}\right)} \int_0^{+\infty} \frac{x^{k/2-1}\left(1+\frac{x}{n}\right)^{-(n+1)/2}}{n^{k/2}}\,dx =$$

If we make change of variable $\frac{x}{n} = z$, we obtain

$$= \frac{1}{B\left(\frac{k}{2}, \frac{n-k+1}{2}\right)} \int_0^{+\infty} z^{k/2-1}(1+z)^{-(n+1)/2}\,dz = \frac{1}{B\left(\frac{k}{2}, \frac{n-k+1}{2}\right)}\,B\left(\frac{k}{2}, \frac{n-k+1}{2}\right) = 1.$$