APPENDIX 1

Useful Concepts

Mills' Ratio

At 1926 Mills mentioned that "the area A of the tail of a frequency curve could be found approximately by the formula $A = \phi(\cdot) \times R_x$, *where* $\phi(\cdot)$ *is the density function at x, and noted that this had appeared in approximating binomial and geometric probabilitiesin term of the standard Normal distribution. For a standard Normal distribution with cumulative distribution function Φ(·) and density function* $\phi(\cdot)$, *this becomes:*

$$1 - \Phi(x) = \phi(x) \times R_x$$

$$\Leftrightarrow R_x = \frac{1 - \Phi(x)}{\phi(x)}$$

R_x *is termed as Mills' ratio and is the reciprocal of the failure rate or hazard rate. Since some expansions for Φ(x) involve its derivatives* $\phi(x), \phi'(x)$ *and so on, and since*

$$\frac{d^m}{dx^m}\phi(x) = (-1)^m \times H_m(x) \times \phi(x), \qquad m = 1,2,...,$$

where $H_m(x)$ *is a Chebyshev–Hermite polynomial, one would expect that approximations to* R_x *frequently correspond to approximations to Φ(x) and vise versa, and such is indeed the case.*

Approximations for Mills' ratio are largely derived from expansions and inequality bounds. Among many approximations, Kotz, Johnson and Campell (1986, Vol. 4) outline a common one:

$$R_x = \frac{t(x)}{x \times t(x) + \sqrt{2}}, \qquad \text{x} \geq 0$$

where $t(x) = \sqrt{\pi} + x \times (2 - a(x)/b(x))$,

$$a(x) = 0.858{,}407{,}657 + x[0.307{,}818{,}193 + x\{0.063{,}832{,}389{,}1 - x(0.000{,}182{,}405{,}075)\}]$$

$$b(x) = 1 + x[0.650{,}974{,}265 + x\{0.229{,}485{,}819 + x(0.034{,}030{,}182{,}3)\}]$$

If this is used as an approximation to the Mill's ratio, the error is less than 12.5×10^{-9} *for the range* $0 \leq x \leq 6.38$*. Detailed listings of expressions and bounds for* R_x*, along with sources for tables, appear in Johnson and Kotz (1972).*

Chebyshev – Hermite Polynomial

A Chebyshev-Hermite polynomial is formulated as a function, $H_m(x)$ *defined by the identity:*

$$\left(-\frac{d}{dx}\right)^m \times a(x) = H_m(x) \times a(x)$$

where $a(x) = \left(1 \big/ \sqrt{2\pi}\right) \times e^{-\frac{1}{2}x^2}$

it is easy to show that $H_m(x)$ *is the coefficient of* $t^m / m!$ *in the expansion of* $\exp(tx - \frac{1}{2}t^2)$.

The polynomial of m^{th} *degree is defined by the formula*

$$H_m(x) = (-1)^m \times e^{x^2/2} \times \frac{d^m}{dx^m} e^{-x^2/2}, \qquad \text{m} = 1,2,\ldots,$$

or equivalently by:

$$H_m(x) = m! \sum_{k=0}^{m/2} \frac{(-1)^k x^{m-2k}}{k! \times (m-2k)! \times 2^k}$$

In terms of the standard Normal density $\phi(x)$*, the Chebyshev-Hermite polynomial is represented by:*

$$(-1)^m H_m(x) \times \phi(x) = \frac{d^m}{dx^m}\phi(x)$$

Also holds the reoccurence relation:

$$H_m(x) = xH_{m-1}(x) - (m-1)H_{m-2}(x), \qquad \text{m} = 2, 3, ...,$$

The first five Chevyshev-Hermite polynomial are:

$$\begin{array}{ll} H_0(x) = 1 & H_3(x) = x^3 - 3x \\ H_1(x) = x & H_4(x) = x^4 - 6x^2 + 3 \\ H_2(x) = x^2 - 1 & H_5(x) = x^5 - 10x^3 + 15x \end{array}$$

Draper and Tierney (1973) give expressions for $H_m(x)$ *for* $0 \leq m \leq 27$*. Fisher and Cornish (1960) have tabulated values of* $H_m(x_p)$ *for* $1 \leq r \leq 7$ *and* $0.0005 \leq p \leq 0.5$*, where* x_p *is the quantile of the standard normal distribution having a probability p in the right tail.*

Biserial Correlation

Biserial correlation refers to an association between a random variable D_i *which takes only two values (for convenience 0 and 1), and a random variable* Y_i *measured on a continuum. Choice of a parameter to measure such an association, and a statistic to estimate and test the parameter, depend on the conceptualization of the nature of the* (X_i,Y_i) *population. A common form is the point biserial correlation (see Johnson and Kotz, 1972).*

The point biserial correlation coefficient is probably the earliest statistical approach to this problem because of its close relationship both to the product-moment correlation coefficient and to the two sample t-test. Regarding the case where D_i indicates the participation status and log(Y_i) are the logarithic wages of the sampled persons, if it is assumed that the distributions of log(Y_i), conditional on D_i = 0 and 1, are Normal with different means but with a common variance, the product moment correlation coefficient, ρ, between D_i and log(Y_i) is estimated by the point biserial correlation coefficient:

$$\rho = (p \times q)^{1/2} \times \left(\overline{y}_1 - \overline{y}_0\right) / s_y$$

where (d_1, y_1), (d_2, y_2), ..., (d_n, y_n) is a sample from the (D_i, Y_i) population, $\overline{y}_1$ and $\overline{y}_0$ *are the mean y values of observations having D_i = 1 and D_i = 0, respectively;* s_y^2 *is the sample variance of Y_i; and p is the proportion of the D-sample with D_i = 1 (p= 1-q).*

The t-statistic may be used to test the null hypothesis that ρ = 0, where

$$t = (n-2)^{1/2} \times \rho^2 \times \left(1 - \left[\rho^2\right]^2\right)^{-1/2}$$

and t is distributed as Student's t with n-2 degrees of freedom. This test is equivalent to a two sample t-test of the null hypothesis that the mean of Y_i values with D_i = 1 equals that with D_i = 0.

Fourier Transformation

Fourier series are used in the analysis of generally a periodic function into its constituent sine waves of different frequencies and amplitudes. The series is:

$$\frac{1}{2}\alpha_0 + \sum \left(\alpha_n \times \cos nx + b_n \times \sin nx\right)$$

where the coefficients are chosen so that the series converges to the function of interest, f; these coefficients (the Fourier coefficients) are given by:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\times \cos nx\, dx$$
$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\times \sin nx\, dx$$

for n = 1, 2, 3, …

A method of calculating the Fourier transform of a set of observations y_0, y_1, …, y_n, i.e. calculating $d(\omega_p)$ given by:

$$d(\omega_p) = \sum_{t=0}^{n-1} x_t \times e^{i\omega_p t}, \quad p = 0, 1, 2, ..., \mathrm{n-1}; \quad \omega_\mathrm{p} = \frac{2\pi p}{n}$$

can be operated in the following way:

- *Let n = rs where r and s are integers. Let $t = rt_1 + t_0$, t = 0, 1, 2, …, n-1; t_1 = 0, 1, 2, …, s-1; t_0 = 0, 1, 2, …, r-1. Further let $p = sp_1 + p_0$, ; p_1 = 0, 1, 2, …, s-1; p_0 = 0, 1, 2, …, r-1. The Fourier transformation (Fast Fourier Transformation) can be written as:*

$$d(\omega_p) = \sum_{t=0}^{n-1} x_t e^{i\omega_p t} = \sum_{t_0=0}^{r-1}\sum_{t_1=0}^{s-1} x_{rt_1+t_0} \times e^{\frac{2\pi ir}{n}(rt_1+t_0)}$$
$$= \sum_{t_0=0}^{r-1} e^{\frac{2\pi ir}{n}t_0} \times a(p_0, t_0)$$

where

$$a(p_0, t_0) = \sum_{t_1=0}^{s-1} x_{rt_1+t_0} \times e^{\frac{2\pi ir}{s}p_0}$$

Calculation of $a(p_0, t_0)$ requires only s^2 operations, and $d(\omega_p)$ only rs^2. The evaluation of $d(\omega_p)$ reduces to the evaluation of $a(p_0, t_0)$ which is itself a Fourier transform. Following the same procedure, the computation of $a(p_0, t_0)$ can be reduced in a similar way. The procedures can be repeated until a single term is reached.

Hodges – Lehmann Estimator

Hodges and Lehmann (1963) first proposed an important technique for deriving a point estimator for a parameter θ from a test statistic that is distribution-free under an appropriate null hypothesis about θ. The idea in the one-sample location setting is as follows.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution with cumulative density function (c.d.f.) $F(x-\theta)$, where $F(\cdot)$ is the c.d.f. for a distribution that is symmetric about 0. Let $V(X_1, X_2, \ldots, X_n)$ be a test statistic for testing H_0: $\theta = 0$ against H_1: $\theta > 0$ that satisfies the following three conditions:

1. *H_0: $\theta = 0$ is rejected for large values of $V(X_1, X_2, \ldots, X_n)$.*
2. *$V(x_1 + h, x_2 + h, \ldots, x_n + h)$ is a non-decreasing function of h for each $(x_1, x_2, \ldots, x_n)$.*
3. *When H_0: $\theta = 0$ is true, the distribution of $V(X_1, X_2, \ldots, X_n)$ is symmetric about some value ξ for every continuous distribution $F(\cdot)$ that is symmetric about zero.*

For such a setting Hodges-Lehmann estimator of θ is motivated in the following way. The random variables $X_1-\theta, \ldots, X_n-\theta$ are independent, and each has the same distribution that is symmetric about zero. Thus, it would be desirable for an estimator of θ, say $\hat{\theta}$, to posses the property that the variables $X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta}$ "look as close as possible" to being symmetrically distributed about 0. In order to define better the criterion "look as close as possible" let $V(X_1, X_2, \ldots, X_n)$ enter to the problem. Since $V(X_1, X_2, \ldots, X_n)$ is used to test H_0: $\theta = 0$, one intuitive way to evaluate this "closeness" property would be to choose $\hat{\theta}$ so that $V(X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta})$ assumes a value as near as possible to the median of the null H_0: $\theta = 0$ distribution of $V(X_1, X_2, \ldots, X_n)$. In view of condition (3), this implies that one chooses $\hat{\theta}$ so that $V(X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta})$ is as close as possible to ξ, the point of symmetry for the null distribution of $V(X_1, X_2, \ldots, X_n)$; that is, so that $X_1 - \hat{\theta}, \ldots, X_n - \hat{\theta}$ "look as close as possible" to being symmetrically distributed about 0, when viewed through the $V(X_1, X_2, \ldots, X_n)$ statistic.

Formally, the Hodges – Lehmann estimator for θ based on $V(X_1, X_2, \ldots, X_n)$ satisfying (1), (2) and (3) is given by:

$$\hat{\theta} = \hat{\theta}(X_1,\ldots,X_n) = \frac{\theta^* + \theta^{**}}{2}$$

where $\theta^* = \theta^*(X_1,\ldots,X_n) = \sup\{\theta : V(X_1-\theta,\ldots,X_n-\theta) \succ \xi\}$

$$\theta^{**} = \theta^{**}(X_1,\ldots,X_n) = \inf\{\theta : V(X_1-\theta,\ldots,X_n-\theta) \prec \xi\}$$

Randles and Wolfe (1979) and Draper and Smith (1994) describe other robust estimators, namely the Least Absolute Deviation estimators (L_1 estimators), the M-estimators, the Least Median of Squares estimators (LMS) and the robust estimators with Ranked Residuals (rreg).

Ordered Probit Model

There are circumstances where response is measured by a variable that can be placed in rank order, but cannot be assigned a quantitative value. Estimation of the probability for this ordinal variable to fall in category j, given a vector of attributes X, cannot be performed by a conventional Probit model because it does not exploit the ranking information and thus produces misleading estimates. As a result an alternative method has to be considered.

Ordered Probit model is described to be the appropriate tool for such kind of analysis. Given an ordinal variable Y and an explanatory variable X = x, let us denote with $p_{ij}(x)$ the probability that the Y outcome of the i^{th} person falls in category j, where $\sum_{j=1}^{c} p_{ij}(x) = 1$. When c = 2 the Ordered Probit model can be written as:

$$\Phi^{-1}(p_1(x)) = \alpha + \beta X$$

For a further description of this model, the reader is referred in Johnson and Kotz (1986, Vol 6) and in the web address "http:// www.indiana.edu/~statmath/stat/all/cat/2b2.html". In the case where Y is an ordinal dependent variable in a linear regression, rankits of Ipsen and Jerne (1944) can be used to estimate the corresponding model.