Commonly Used Terms, Symbols, and Expressions

The lectures, handouts, readings, and problem sets use many terms and symbols which may have been used in different ways in your previous statistical courses. This document includes most of the non-standard symbols and terms that will be used during the term; I will work hard to insure that we remain faithful to the definitions outlined below.

Please note: Most of this will not make sense to you right now unless you have had a linear regression course before. The contents will become clearer as we go along.

Mean and Expected Value
In general, the symbols used for the mean and expected value are straightforward.

\[ \mu = \text{population mean} \]
\[ E(x) = \text{the expected value of } x \text{ (e.g., } \mu \text{)} \]
\[ E(x|y) = \text{the conditional mean of } x; \text{ the expectation of } x \text{ given } y. \]
\[ \bar{x} = \text{the sample mean} \]

Variances: Residual variance, standard deviation, sample standard deviation, and standard error of coefficient estimates
The variance symbols are perhaps the most confusing in the course. Unfortunately, authors use the lower case sigma, \( \sigma \) to represent all of the concepts listed above. In RPAD 705, the lower case sigma, \( \sigma \), will in almost all cases, refer to the standard deviation of the error term in a specification like \( y = X\beta + \varepsilon \). Sometimes it may also represent the residual variance. The residual variance is sometimes also called the error variance. The residual variance is the variation around the “true” regression line that exits due to random fluctuations.

In most cases, we will also assume that the residual variance is homoskedastic — constant across all levels of the independent variables (non-constant variance, or heteroskedasticity, is a common problem that affects coefficient standard error in regression and other estimation techniques). The sampling equivalent of error variance is the residual variance, indicated by the root mean squared error (called “Root MSE” in most Stata output).
Thus:

\[ \sigma = \text{error or residual variance} \]
\[ \sigma_x = \text{standard deviation of a variable (here } x) \]
\[ \sigma_{\beta x} = \text{standard error of the coefficient on } x \]
\[ s = \text{sample standard deviation of a variable} \]

Note that \( s \) and \( s^2 \) always refer to sample standard deviation and variance.

For reference, the Root Mean Square Error can be calculated once a regression is run by using the estimated coefficients – here called “\( b \).” The formula is:

\[
\left\{ \frac{\sum(y_i - x_i b)^2}{N-k} \right\}^{1/2}
\]

where

\( b = \) estimated coefficients
\( y_i = \) dependent variable
\( x_i = \) independent variables
\( N = \) number of observations
\( k = \) number of variables in the regression (including the constant)

**Probability Distributions**

The distribution of a variable is often denoted by a set of symbols that include the parameters that define the distribution. For instance, if \( \varepsilon \) is distributed normally with a mean of \( \mu \) and standard deviation of \( \sigma_\varepsilon \) then we would write: \( \varepsilon \sim N(\mu, \sigma_\varepsilon^2) \), or simply \( \varepsilon \sim N(\mu, \sigma^2) \). If the mean and standard deviation are known, then \( \mu \) and \( \sigma_\varepsilon \) would be replaced with the actual values: for instance, if \( \varepsilon \) is distributed as a unit normal, then the expression would be: \( \varepsilon \sim N(0,1) \)

**Unit Normal Probability Density Function (PDF)**

The unit normal PDF assumes that a variable \( X \) is distributed normally with \( \mu = 0 \) and \( \sigma = 1 \). This function is used to find the probability of a value \( x \) occurring, given these assumptions.

Symbol: \( \phi(x) \)  read as the “probability” or likelihood that a unit normal random variable takes the value \( x \).
Examples: $\phi[y]$ is the probability that a unit normal random variable takes the value $y$ across all observations $i$.

$\phi[y_i - x_i \beta]$ is the probability that a unit normal random variable takes the value $y_i - x_i \beta$ across all observations $i$. (This usually arises in a context where $E(y_i) = x_i \beta$, so that $[y_i - x_i \beta]$ has a unit normal distribution if the variance of $y$ is 1).

$\phi[x_i \beta]$ is the probability that a unit normal random variable takes the value $x_i \beta$. (Note that because the unit normal distribution is symmetrical with mean 0, $\phi[0 - x_i \beta] = \phi[-x_i \beta] = \phi[x_i \beta]$)

$(1/\sigma) \phi[(y_i - x_i \beta)/\sigma]$ is the probability of $y$ across all observations $i$, given that $y$ is distributed normally with mean and standard deviation $\mu = x_i \beta$ and $\sigma = 1$. This is simply short hand for the normal PDF (not the unit normal), given $\mu$ and $\sigma$. Note carefully why the $(1/\sigma)$ appears in this statement. $\phi[x]$ represents the equation $(1/\sqrt{2\pi \sigma^2}) e^{-[1/2(y - x \beta)^2]}$, but the expression we wish to replicate is $(1/\sqrt{2\pi \sigma^2}) e^{[1/2((y_i - x_i \beta)/\sigma)^2]}$. In the short-hand statement, we have already divided by the residual variance, which takes care of the $\sigma$ term in the exponent on $e$. However, this does not account for the $\sigma$ in the first term — $(1/\sqrt{2\pi \sigma^2})$. Thus, we must multiple by $(1/\sigma)$ to use this short-hand properly. It turns out that the probability that the random variable $Y$ takes the value $y$ is given by $1/\sigma$ times the probability that a unit normal distribution takes the value $(y_i - x_i \beta)/\sigma$. 

Unit Normal Cumulative Distribution Function (CDF)

The unit normal CDF finds the probability that a unit normal random variable (i.e., \( \mu = 0 \) and \( \sigma = 1 \)) is less than a given value \( x \).

Symbol: \( \Phi[x] \) read as “the probability that a unit normal random variable takes a value less than \( x \)."

Examples: \( \Phi[y_i] \) is the probability a unit normal random is less than \( y_i \).

\( \Phi[y_i - x_i \beta] \) is the probability that a unit normal random variable is less than \( y_i - x_i \beta \).

\( \Phi[x_i \beta] \) is the probability that a unit normal random variable is less than \( x_i \beta \). (Note that because the unit normal distribution is symmetrical with mean zero, \( \Phi[-x_i \beta] = 1 - \Phi[x_i \beta] \))

\( \Phi[(y_i - x_i \beta)/\sigma] \) is the probability that a unit normal random variable takes a value less than \( (y_i - x_i \beta)/\sigma \).

All of the examples above give the probability that a unit normal random variable is less than some value. The probability that the unit normal is greater than this value is simply given by \( 1 - \Phi[.] \). For example, the probability of finding a value \( y \) greater than \( (y_i - x_i \beta)/\sigma \) is \( 1 - \Phi[(y_i - x_i \beta)/\sigma] \).
Frequently Used Letters (Greek and otherwise)
In general, we will stick to the conventional notation for dependent and independent variables. However, in many cases, the letters will refer to vectors rather than scalars.

\( y \) = dependent variable (usually a scalar)

\( X \) = independent variables (usually a vector)

\( \beta \) = coefficients on independent variables (usually a vector)

\( \varepsilon \) = error term in an estimation

\( \rho \) = correlation

\( L \) = likelihood function

\( Pr \) = generic notation for probability

\( f(y) \) = generic notation for a probability density function (NOT necessarily the normal PDF)

\( F(y) \) = generic notation for a cumulative distribution function (NOT necessarily the normal CDF)

\( M \) = Mills ratio (i.e., \( \varphi[ ] / \Phi[ ] \))

\( \Pi \) = Multiplicative series (i.e., \( \Pi f(y_i) = f(y_1) \times f(y_2) \times f(y_3) \times f(y_3) \ldots \))
**Sums of Squares**

For calculation of F tests, root mean squared error, and other measures, one must often use sums of squares. Unfortunately, computer programs and textbooks also tend to use abbreviations inconsistently in this area.

**TSS = total sum of squares**

This measures the total variance of the dependent around its *unconditional* mean. The formula is:

\[ \sum (y_i - \bar{y})^2 \]

Regression sum of squares (RSS) or Model sum of squares (MSS)

This measures the amount of variance the estimated coefficients explain. If “b” is the vector of estimated coefficients, then the RSS or MSS formula is:

\[ \sum (\mathbf{x}_i \mathbf{b} - \bar{y})^2 \]

Residual sum of squares (RSS) or Error sum of squares (ESS)

This measures the amount of variance that is left unexplained by the regression estimate. It is part of the calculation for the root mean squared error. If “b” is the vector of estimated coefficients, then the RSS or ESS formula is:

\[ \sum (y_i - \mathbf{x}_i \mathbf{b})^2 \]