Statistical versus Practical Significance *M. C. Spruill*

Sometimes when one wishes to utilize a statistical model to generate predictions or to study the key determining factors of some phenomenon, especially when there is a large amount of data, one can demonstrate with a convincing degree of statistical significance that a parameter in a model is non-zero; yet, it may be that such a parameter, although non-zero, is of no particularly great use in predicting the value of an observable or in achieving a clear understanding of the key variables involved in determination of the observable.

1 An example

In Figure 1 find a plot of some (x, y) pairs, $-10 \leq x \leq 10$ and in Table 1 the ANOVA for a regression, $Y(x) = \beta_0 + \beta_1 x + \epsilon$. Note that β_1 is significantly different from 0 (p-value < 0.0001) but that the data plot shows it is of little use in predicting the value of $Y(x)$; it is β_0 which is important.

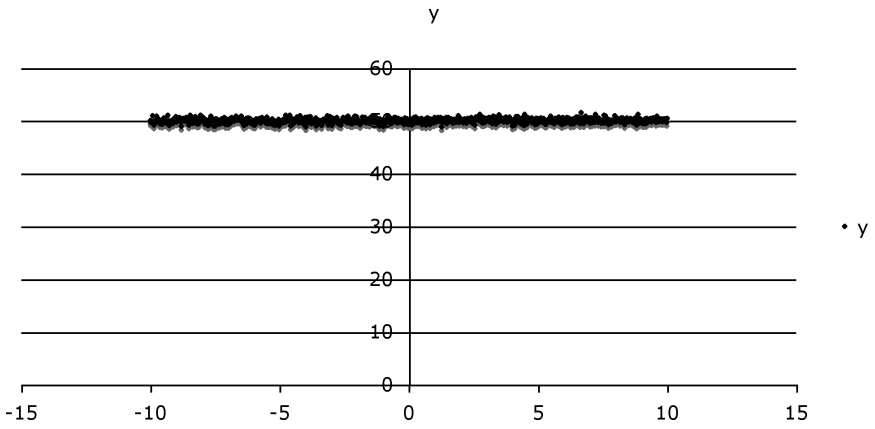

Figure 1: Plot of 900 data points, independent $N(50 + 0.01x_i, (0.44)^2), x_i \in [-10, 10]$.

Variable	DF	Parameter Estimate	Standard Error	t Value	$Pr > \|t\|$
Intercept	1	49.99285	0.01515	3299.09	$< .0001$
x	1	0.01371	0.00262	5.23	$< .0001$

Table 1: The slope β_1 is significantly different from 0.

Thus it is β_0 that dictates the size of $Y(x)$, no matter what x is in the interval of possible values. The fact that 900 observations reveal a high degree of significance that β_1 is non-zero is of little practical importance.