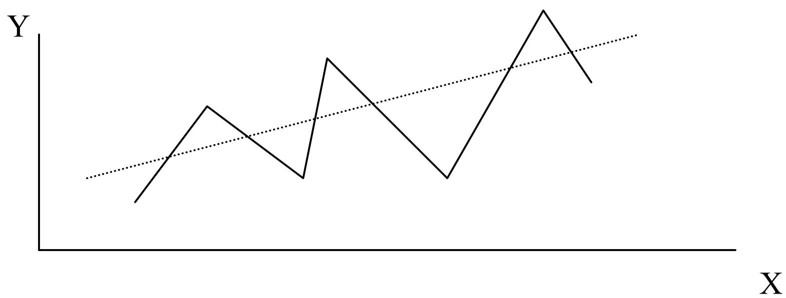
ECON 120B, SPRING 2003 --- ANSWERS TO HOMEWORK #1

Exercise 3.8

The slope of the line connecting the first and second point is $(Y_2 - Y_1)/(X_2 - X_1)$. The slope of the line connecting the second and third points is $(Y_3 - Y_2)/(X_3 - X_2)$. By proceeding similarly, the slope of the line connecting point t-1 with t is given by $(Y_t - Y_{t-1})/(X_t - X_{t-1})$.

The scatter diagram for this is given below.



The average of all the slopes in the scatter diagram is given by

$$\tilde{\beta} = \frac{1}{n-1} \sum_{t=2}^{n} \left[\frac{Y_t - Y_{t-1}}{X_t - X_{t-1}} \right]$$

To compute the expected value of $\tilde{\beta}$, we have,

$$\frac{Y_t - Y_{t-1}}{X_t - X_{t-1}} = \frac{\alpha + \beta X_t + u_t - \alpha - \beta X_{t-1} - u_{t-1}}{X_t - X_{t-1}} = \beta + \frac{u_t - u_{t-1}}{X_t - X_{t-1}}$$

Hence,

$$\tilde{\beta} = \beta + \frac{1}{n-1} \sum_{t=2}^{n} \left[\frac{u_t - u_{t-1}}{X_t - X_{t-1}} \right]$$

X_t is nonrandom and $E(u_t) = E(u_{t-1}) = 0$. Therefore, $E(\tilde{\beta}) = \beta$, which means that $\tilde{\beta}$ is unbiased. By the Gauss-Markov Theorem, OLS estimates are most efficient among unbiased linear estimators. This implies that any other such estimator, in particular $\tilde{\beta}$, is inefficient (or at least is no more efficient) than the OLS estimator.

Exercise 3.10

From the model, $\overline{Y} = \alpha + \beta\overline{X} + \overline{u}$. Therefore, $\tilde{\beta} = \dfrac{\overline{Y}}{\overline{X}} = \beta + \dfrac{\alpha + \overline{u}}{\overline{X}}$. Taking the expected value and noting that X is non-random and that E ($\overline{u}$) = 0 because E(u_t) = 0, we have,
$E(\tilde{\beta}) = \beta + \dfrac{\alpha}{\overline{X}} \neq \beta$. Therefore $\tilde{\beta}$ biased.

Exercise 3.28

a) In Section 3.5 we described how to test the model as a whole with an F-test. The null hypothesis is that X and Y are uncorrelated (that is, $\rho_{XY} = 0$) and the alternative is that they are correlated. The test statistic is $F_c = R^2(n-2)/(1- R^2)$. In our example, n = 427 and the F-statistics for the three models are, respectively, 84, 32, and 60. For a 1 percent-level of significance, the critical $F^*_{1,425}(0.01)$ is approximately 6.7. Because the calculated F-values are well above this, we reject the null hypothesis of lack of correlation between X and Y and conclude that they are correlated. This means that all three models are significant overall.

b) The null hypothesis is that a particular regression coefficient is zero. The alternative for a two-tailed test is that it is nonzero. The critical $t^*_{425}(0.005)$ is below 2.617. If an observed t-value exceeds this (in absolute terms) we reject the null hypothesis and conclude that the coefficient is statistically significant. The calculated t-values are:

0.92058/0.20463	=	4.50
0.52417/0.05712	=	9.18
1.99740/0.14128	=	14.14
0.00157/0.00028	=	5.61
1.62845/0.15135	=	10.76
0.00204/0.00026	=	7.85

Because all the t-statistics exceed the critical value, every regression coefficient in every model is statistically significantly different from zero.

c) The low values for R^2 indicate that the independent variables HSGPA, VSAT, and MSAT do not explain much of the variance in COLGPA. We will see in later chapters that a more extended model does better. A multiple regression model specification would be

$$\text{COLGPA} = \beta_1 + \beta_2\,\text{HSGPA} + \beta_3\,\text{VSAT} + \beta_4\,\text{MSAT} + u$$