17.872
Fall 2003
Final Exam

Question 1.

The department of Housing and Urban Development audits rental markets to measure the extent of racial discrimination. They send pairs of individuals with similar income but different apartments to measure the rents offered to blacks and whites, controlling for income. Consider the following data.

Pair	Income	Race	Rent
1	$65,000	B	$1400
1	65,000	W	1300
2	50,000	B	1210
2	50,000	W	1210
3	35,000	B	1250
3	35,000	W	1210
4	25,000	B	900
4	25,000	W	850
5	20,000	B	740
5	20,000	W	700

(a) Calculate the means and standard deviations of rent charged to blacks and of rents charged to whites. Calculate the 95 % confidence interval for the difference between black and white rents.

(b) Calculate the mean difference and standard deviation of the differences across pairs. Calculate a 95 percent confidence interval for the difference, using the matched pairs. Compare your result with part (a).

(c) Regress Rent on Income for whites and for blacks (separately). What is the difference between the intercepts? Do the slopes appear to differ?

Question 2.

A study is done to measure the effect of smoking on blood pressure. 111 smokers and 69 non-smokers volunteered to participate in the study. Their blood pressures were measured and recorded as high or normal. Of the smokers, 66 had high blood pressure and 45 had normal blood pressure. Of the non-smokers, 21 had high blood pressure and 48 percent had normal blood pressure.

(a) Conduct a Chi-squared test of the hypothesis that smoking and blood pressure are unrelated.

(b) High blood pressure (versus normal blood pressure) is the dependent variable; smoking (or not) is the independent variable. What is the effect of smoking on high blood pressure? Give a 95 percent confidence interval for the effect of smoking on blood pressure.

(c) Before the study was performed the researchers calculated the minimum number of cases needed for their study. They wished to be able to distinguish among effects of 5 percentage points or more (i.e., differences in the percent with high blood pressure of .05 or more). How many people in each group must be studied in order to be able to detect such differences between the smokers' and non-smokers' blood pressure rates? Assume that the researchers will use the same number of cases in each group (smokers and non-smokers).

Question 3.

Investment decisions are based on rates of returns from securities (stocks) and the risk associated with those returns. Specifically, the price of any stock j on day t is $p_{j,t}$. The percent change in prices is called the rate of return on an investment:

$$R_{j,t} = (p_{j,t} - p_{j,t-1})/\, p_{j,t-1}$$

The average rate of return of any stock (expected value of R) is μ_j and the risk of the stock (the variance) is σ_j^2.

(a) Here are the average and standard deviations of the daily returns of Microsoft and Intel over the last 15 years (n = 3784). Also shown is the average return and standard deviation of all stocks traded on the New York Stock Exchange.

	Average Daily Rate of Return	Standard Deviation
Microsoft	.00116	.0249
Intel	.00116	.0286
NYSE	.00046	.0099

Construct a 95 percent confidence interval for the true rate of return for Microsoft, for Intel, and for the NYSE.

(b) If you bought 10 shares of Microsoft 15 years ago at $25 a share and held them until today what would they be worth? (Hint: This is a compound interest problem; use the average rate of return for the "interest rate.")

(c) Microsoft and Intel have had the same average rate of return, but Microsoft has a lower standard deviation, suggesting that it has smaller day-to-day variation in price. A problem with this as a measure of risk is that it does not correct for variation in a stock's price that is attributable to the market as a whole and variation that is not. A better way to measure the riskiness of a security uses regression of the rate of return of a given stock's price on the rate of return of the overall market. That is:

$$R_{j,t} = a + b\,X_t + e_{j,t},$$

where X is the market's rate of return on day t. The regression coefficient b measures the riskiness of the security relative to the market. If b>1 the security is said to be "riskier" than the market, and if b<1 it is less risky than a portfolio that is diversified into the entire market.

The following tables present the formula for the regressions for Microsoft and Intel against the market. I have omitted several elements of the analysis of variance table and the regression estimates.

(c) Fill in the missing numbers using the formula for the ANOVA tables.

```
. reg return mktr

      Source |       SS       df       MS              Number of obs =    3784
-------------+------------------------------           F(  1,  3777) = 2039.19
       Model |  xxxxxxxxxx     1  1.08832729          Prob > F      =  0.0000
    Residual |  2.01580918  xxxx     xxxxxx           R-squared     =  0.3506
-------------+------------------------------           Adj R-squared =  0.3504
       Total |  3.10413647  3783  .000821635           Root MSE      =   .0231

----------------------------------------------
      return |      Coef.   Std. Err.      t
-------------+--------------------------------
        mktr |    1.72063    xxxxxxx     xxxx
       _cons |   .0003903   .0003762     1.04
----------------------------------------------

. reg return mktr

      Source |       SS       df       MS              Number of obs =    3784
-------------+------------------------------           F(  1,  3777) = 1740.77
       Model |  .739449093     1  .739449093           Prob > F      =  0.0000
    Residual |  1.60440442  xxxx  xxxxxxxxxx           R-squared     =  0.3155
-------------+------------------------------           Adj R-squared =  0.3153
       Total |  2.34385351  3783  .000620395           Root MSE      =  .02061

----------------------------------------------
      return |      Coef.   Std. Err.      t
-------------+--------------------------------
        mktr |   xxxxxxxx   .0340055    41.72
       _cons |   .0005364   xxxxxxxx     xxxx
----------------------------------------------
```

(d) Fill in the missing coefficients and standard errors and t-statistics. (Hint: Use the formulas for the sums of squares of X, for the t-statistic (for the hypothesis that b=0), and for the standard errors as well as the information in the tables and the variance of the market rate of return above.) Which stock has higher b?

(e) For each stock test the hypothesis that b = 1.