

Practice Final Exam

(corrected formulas, 12/10 11AM)

Answer all questions in the 3 (three) bluebooks provided. Make certain you write your name, your student ID number, and your TA's name on all your bluebooks, as well as noting the bluebook (A, B, or C).

This exam is 90 minutes long, although you will be given 120 minutes to complete it. Point allocations are proportional to time allocations. Partial credit will be awarded if the written material indicates understanding of how to answer the question (i.e., gibberish will not be given credit).

Bluebook A: 30 minutes, hypothesis testing

A.1 (10 minutes) Moving companies are required by the government to publish a Carrier Performance Report each year. One of the descriptive statistics they must include is the annual percentage of shipments on which a \$50 or greater claim for loss or damage was filed. Suppose Company A and Company B each decide to estimate this figure by sampling their records, and they report the data shown in the table.

<u>Company</u>	<u>Co. A</u>	<u>Co. B</u>
Total shipments sampled	900	750
Number of shipments with claims \geq \$50	180	60

Conduct a test of hypothesis to determine if Company A has a higher proportion of shipments in which a \$50 or greater claim for loss or damage is filed than does Company B. Hint: $255/1650$ is approximately 0.1545.

A.2 (10 minutes) A study was commissioned to compare the housing costs of two growing cities. The goal of the study was to estimate the difference in the average housing costs (as measured in price per square foot) of the two cities. Two pilot samples of 50 houses in each city were taken and yielded the following information

	<u>City 1</u>	<u>City 2</u>
Mean of housing cost	\$50.40 per square foot	\$53.70 per square foot
Std. Dev. of housing cost	\$2.00 per square foot	$\$ \sqrt{6}$ per square foot

The study also wanted to determine if the variation in the housing costs for the two cities differed. Use $\alpha = .05$ to conduct the desired test.

A.3 (10 minutes) The University of Minnesota uses thousands of fluorescent light bulbs each year. The brand of bulb it currently uses has a mean life of 900 hours. A manufacturer claims that its new brands of bulbs, which cost the same as the brand the university currently uses, has a mean life of more than 900 hours. The university has decided to purchase the new brand if, when tested, the test evidence supports the manufacturer's claim at the .05 significance level. Suppose 64 bulbs were tested with the following results: $x = 920$ hours, $s = 80$ hours. Conduct the test using $\alpha = .05$.

Bluebook B: 30 minutes, regression

In macroeconomics, the Phillips Curve is an empirical relationship that describes inflation as a function of expected inflation and the output gap (the gap between current GDP and the “normal” level of output of GDP).

$$\pi_t = \beta_0 + \beta_1 \pi_t^e + \beta_2 \hat{y}_t + \varepsilon_t$$

Using data on US inflation, assuming expected inflation is just last period’s inflation, and using the Congressional Budget Office’s (CBO) estimates of the output gap, the following results were obtained.

Dependent Variable: INFLUS

Method: Least Squares

Date: 12/07/03 Time: 22:59

Sample(adjusted): 1973:1 2000:2

Included observations: 110 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011274	0.003376	3.339068	0.0012
INFLUS(-1)	0.820513	0.054323	15.10432	0.0000
GAPUS_CBO	0.196954	0.076009	2.591193	0.0109
R-squared	0.684444	Mean dependent var		0.050853
Adjusted R-squared	0.678545	S.D. dependent var		0.033603
S.E. of regression	0.019052	Akaike info criterion		-5.056410
Sum squared resid	0.038838	Schwarz criterion		-4.982760
Log likelihood	281.1025	F-statistic		116.0418
Durbin-Watson stat	2.189920	Prob(F-statistic)		0.000000

- What is the interpretation of the constant in this context? (inflation and the output gap are measured in “decimal” form, i.e., 10% is recorded as “0.10”).
- Show how you would calculate the R-squared using the data in the regression output.
- Compare the coefficient of determination in this specification versus that in the simple regression:

Dependent Variable: INFLUS
 Method: Least Squares
 Date: 12/08/03 Time: 16:36
 Sample(adjusted): 1973:1 2000:2
 Included observations: 110 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009452	0.003389	2.789305	0.0062
INFLUS(-1)	0.814856	0.055697	14.63025	0.0000
R-squared	0.664642	Mean dependent var		0.050853
Adjusted R-squared	0.661537	S.D. dependent var		0.033603
S.E. of regression	0.019549	Akaike info criterion		-5.013731
Sum squared resid	0.041275	Schwarz criterion		-4.964632
Log likelihood	277.7552	F-statistic		214.0443
Durbin-Watson stat	2.094014	Prob(F-statistic)		0.000000

The coefficient of determination is higher in the three variable specification than in the simple regression. Is this sufficient reason to prefer the three variable to two variable regression? What would be a better decision criterion?

- d) **Returning to the three variable regression results**, consider the following question. If last period's inflation were 0%, and the current output gap were 10%, what would your prediction of the current period's inflation rate be, on the basis of the above equation.
- e) Summary statistics for inflation and the output gap are reported in the table below. How confident are you of the prediction you have made? Explain your reasoning.

Date: 12/07/03
 Time: 23:07
 Sample: 1973:1 2000:4

	INFLUS	GAPUS_CBO
Mean	0.050853	-0.010711
Median	0.040663	-0.009872
Maximum	0.153456	0.041520
Minimum	-0.010053	-0.081369
Std. Dev.	0.033603	0.024028
Sum Sq. Dev.	0.123078	0.062929

Bluebook C: 30 minutes, Comprehensive

C.1 Hypothesis Testing (12 minutes)

Bernoulli is a statistical consultant for Pascal Enterprises. Pascal Enterprises wanted to test the null hypothesis, H_0 , that the proportion p of ledger sheets with errors is equal to .05 versus the alternative, H_a , that the proportion is larger than .05. Bernoulli's first task was to construct a test of these hypotheses. Unfortunately, Bernoulli had a bit too much to drink and proposed the following test:

Select two ledger sheets at random. If both are error free, reject H_0 . If one or more contains an error, look at a third sheet. If the third sheet is error-free, reject H_0 . In all other cases, we accept H_0 .

- What is the value of α (the probability of a Type I error) associated with this test?
- Find β (the probability of a Type II error) in terms of p .

C.2 Regressions (18 minutes)

Bernoulli's second task for the day was to estimate some regressions. First, the Pascal Enterprises wanted to test how sales varied by season. Let y be a measure of daily sales. Bernoulli defined the following dummy variables:

$x_1 = 1$, if the day was in the Winter;	$x_1 = 0$ otherwise
$x_2 = 1$, if the day was in the Spring;	$x_2 = 0$ otherwise
$x_3 = 1$, if the day was in the Summer;	$x_3 = 0$ otherwise
$x_4 = 1$, if the day was in the Autumn;	$x_4 = 0$ otherwise

Then, Bernoulli estimated the following regression to determine the effects of season on daily sales:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4$$

- Explain briefly (one sentence) why Bernoulli's regression is not properly specified.
- Write down the regression that Bernoulli should have estimated.

Second, Pascal Enterprises wanted to test whether a corporation's profits, y , could be predicted from information on:

x_1 = the CEO's annual income

x_2 = the percentage of the company's stock owned by the CEO

Bernoulli determined (correctly) that CEO income and stock holdings could interact to predict company profit. Thus, he estimated his model with an interaction term:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_1 x_2 \quad \text{"regression one"}$$

He conducted t-tests on the coefficients and noticed that while $\hat{\beta}_3$ was significantly different from zero, $\hat{\beta}_2$ was not. Thus, he dropped x_2 and instead estimated the following regression:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_1 x_2 \quad \text{“regression two”}$$

c) True/False/Explain:

Both regression one and regression two are correctly specified.

8.12.2003 11:00PM

In this case, Normal, t and F tables would be provided at the exam.

Economics 310 Fall 2003
Equations and formulas for Final Exam

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1}$$

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \text{ where } N! = (N)(N-1)(N-2)\dots(2)(1)$$

$P(L_1 \cap L_2) = P(L_1)P(L_2)$ if L_1 and L_2 are independent

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

$$z\text{-score} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

$$\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$$

$$\hat{p} \pm z_{\alpha/2} \sigma_{\hat{p}} \text{ where } \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

$$\bar{x} \pm t_{\alpha/2, s_{\bar{x}}} \text{ where } s_{\bar{x}} = s / \sqrt{n}$$

$$\tilde{p} = \frac{x+2}{n+4}$$

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{B^2}$$

$$n = \frac{(z_{\alpha/2})^2 pq}{B^2}$$

$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \text{ where for parameter } \theta = \mu, \sigma_{\bar{x}} = \sigma / \sqrt{n}, \text{ and for parameter } \theta = p, \sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}}$$

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Power = 1- β

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\text{or } \sigma_{(\bar{x}_1 - \bar{x}_2)} \approx \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$\text{where } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$\sigma_{\bar{x}_D} = \frac{\sigma_D}{\sqrt{n_D}} \approx \frac{s_D}{\sqrt{n_D}}$$

$$\sigma_{(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \quad \text{or} \approx \sqrt{\hat{p} \hat{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}; \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$n_1 = n_2 = \frac{(z_{\alpha/2})^2 (\sigma_1^2 + \sigma_2^2)}{B^2} \quad n_1 = n_2 = \frac{(z_{\alpha/2})^2 (p_1 q_1 + p_2 q_2)}{B^2}$$

$$F = \frac{\text{larger } s^2}{\text{smaller } s^2} \quad \text{or} \quad F = \frac{s_1^2}{s_2^2}$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \quad \hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}, \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$SS_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad SS_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$s^2 = \frac{SSE}{n - (k + 1)} \quad SSE = \sum (y_i - \hat{y}_i)^2 = SS_{yy} - \hat{\beta}_1 SS_{xy} \quad s_{\hat{\beta}_1} = \frac{s}{\sqrt{SS_{xx}}}$$

$$\hat{\beta}_i \pm t_{\alpha/2} s_{\hat{\beta}_i} \quad r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \quad R^2 = \frac{SS_{yy} - SSE}{SS_{yy}}$$

$$\hat{y} \pm t_{\alpha/2} s \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \quad \hat{y} \pm t_{\alpha/2} s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}} \quad t = \frac{\hat{\beta}_i - 0}{s_{\hat{\beta}_i}}$$

$$R_a^2 = 1 - \left[\frac{(n-1)}{n - (k+1)} \right] (1 - R^2)$$

$$F = \frac{(SS_{yy} - SSE) / k}{SSE / [n - (k + 1)]} = \frac{\text{MeanSquare}(\text{Model})}{\text{MeanSquare}(\text{Error})}$$

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \quad E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$$