Final Exam

12/11/12

Instructions: Answer all questions. Point values are as marked. You should support your answers as well as you are able to. You may use a calculator, and one sheet of notes.

Problem 1 (10 points) Suppose that the percentage returns on stock in Al's Apples and Betty's Bourbon are described by the following parameters:

$$\mu_A = .05 \quad \sigma_A = .04 \quad \rho_{12} = 0$$
$$\mu_B = .12 \quad \sigma_B = .07$$

a. Is there a portfolio of stocks A and B whose return has a standard deviation which is lower than both σ_A and σ_B? Be specific.

b. Do the returns on the $(p_A, p_B) = (\frac{1}{2}, \frac{1}{2})$ portfolio have a standard deviation which is lower than $\frac{1}{2}\sigma_A + \frac{1}{2}\sigma_B$?

Problem 2 (20 points) Using team-level data for the 2012 MLB season, Alice estimates the following regression equation:

$$Wins = \beta_0 + \beta_1 * RUNS\_SCORED + \beta_2 * HITS + \beta_3 * HOME\_RUNS + \beta_4 * RUNS\_BATTED\_IN + \beta_5 * BATTING\_AVERAGE + \beta_6 * RUNS\_ALLOWED + \epsilon$$

where $RUNS$ is the number of runs scored by each team's offense, $HITS$ is the number of hits each team's offense produced, $HOME\_RUNS$ is the number of home runs each team's offense hit, $RUNS\_BATTED\_IN$ is the number of RBI's each team's offense produced,[1] $BATTING\_AVERAGE$ is the fraction of the team's plate appearances that resulted in a hit, and $RUNS\_ALLOWED$ is the total number of runs each team allowed their opponents to score.

The regression results are below:

Regression	**Statistics**
Multiple R	.952705
R Square	.907647
Adjusted R Square	.883555
Standard error	4.072272
Observations	30

	coefficients	Standard error	t stat	P-value	Lower 95%	Upper 95%
Intercept	84.33761	27.43993	3.073536	.005374	27.57379	141.1014
RUNS	.201274	.138988	1.448134	.161075	-.08625	.488793
HITS	.033766	.05902	.572107	.572799	-.08833	.155859
HOME_RUNS	.03065	.043232	.708976	.485464	-.05878	.120081
RUNS_BATTED_IN	-.12493	.139605	-.89486	.380134	-.41372	.163869
BATTING_AVERAGE	-113.967	400.4509	-.2846	.7785	-942.363	714.4286
RUNS_ALLOWED	-.12055	.010088	-11.9495	2.4 E-11	-.14142	-.09968

a. Which variables are significant explainers of a team's win total?

[1] An RBI is credited to a batter when the result of his appearance is a run being scored. For example, a home run with 3 men on base results in 4 RBI's being credited to the batter.

b. Interpret the estimated coefficient on BATTING_AVERAGE. Note that batting average is measured as a fraction (e.g. a batting average of .250 means that a team gets a hit in 25% of its plate appearances, while a batting average of .300 means a hit occurs in 30% of plate appearances). The lowest team batting average in 2012 was .240, while the highest was .271.

c. The minimum number of runs allowed in 2012 was 577 (Tampa Bay), while the maximum was 890 (Colorado). How many additional wins do the regression estimates predict Tampa Bay will have relative to Colorado, based on allowing fewer runs?

d. Alice examines the above estimates and concludes that good pitching is essential to winning a high number of games, but offense is not particularly important. Assess Alice's conclusion.

Problem 3 (10 points) A doctor's family practice clinic treated 100 patients with flu-like symptoms in 2012. 50% of those who were advised only to rest at home got better within 2 days. 40% of those who were given Fluaway, an experimental new drug, got better within two days.

a. You are hired as a consultant to the clinic. You disaggregate the data so that you can look at children and adults separately. Describe, using the table below, data that would lead you to conclude that Fluaway is a *more* effective treatment than rest alone (there are many correct answers!).

Children only	Improved	Not improved	% improved
Fluaway			
Rest			

Adults only	Improved	Not improved	% improved
Fluaway			
Rest			

b. Describe, using the table below, data that are consistent with the conclusion that Fluaway is a *less* effective treatment than rest alone (again, there are many correct answers).

Children only	Improved	Not improved	% improved
Fluaway			
Rest			

Adults only	Improved	Not improved	% improved
Fluaway			
Rest			

Hint: One of parts a-b is hard, one is easy.

Problem 4 (10 points) Consider the two-asset CAPM model studied in class. Suppose that Asset 1 has an expected return of .08 and a standard deviation of .04, while Asset 2 has an expected return of .12 and a standard deviation of .1.

a. Draw the feasible set of (σ, μ) pairs from a portfolio of these two assets if the correlation between the two assets is -1. Be as precise as you can.

b. Draw the feasible set of (σ, μ) pairs from a portfolio of these two assets if the correlation between the two assets is 1. Be as precise as you can.

c. Draw the feasible set of (σ, μ) pairs from a portfolio of these two assets if the correlation between the two assets is 0. Be as precise as you can, but don't worry about drawing an exact picture.

d. What is the maximum value of $\rho_{1,2}$, the correlation between assets 1 and 2, for which the feasible set is backward bending? What value of $\sigma_{1,2}$, the covariance between assets 1 and 2, is implied by the value of $\rho_{1,2}$ in your answer?

Extra credit (6 points) Revisit a-c above. In each case, solve for the minimum variance portfolio.

Problem 5 (10 points) The following is an abstract of an economics research paper:

> This paper compares corruption in China over the past 15 years with corruption in the U.S. between 1870 and 1930, periods that are roughly comparable in terms of real income per capita. Corruption indicators for both countries and both periods are constructed by tracking corruption news in prominent U.S. newspapers. The comparison indicates that corruption in the U.S. in the early 1870s when its real income per capita was about $2,800 (in 2005 dollars) was 7 to 9 times higher than Chinas corruption level in 1996, the corresponding year in terms of income per capita. By the time the U.S. reached $7,500 in 1928 — approximately equivalent to Chinas real income per capita in 2009 — corruption was similar in both countries. The findings imply that, while corruption in China is an issue that merits attention, it is not at alarmingly high levels, compared to the U.S. historical experience.

What objections or concerns is the author most likely to run into when he presents his paper to other economists?

Problem 6 (20 points) A researcher is interested in the effects of smoking and body mass on cardiovascular health. He finds a random sample of 2,300 30 year old smokers, and estimates the following regression equation:

$$\begin{aligned} CV\_HEALTH =& \beta_0 + \beta_1 * CIGS + \beta_2 * CIGS^2 \\ &+ \beta_3 * MALE + \beta_4 * BMI + \beta_5 * BMI * MALE + \epsilon \end{aligned}$$

CV_HEALTH is a measure of cardiovascular health determined by a series of tests such as running on a treadmill, and measured on a scale of 0 to 100, with 0 being the lowest possible score, and 100 the highest. $CIGS$ measures the number of cigarettes a person smokes on a typical day. $MALE$ is a dummy variable equal to 1 if the person is male, and 0 if female. BMI equals mass divided by height squared, with higher values meaning someone weighs more.

He obtains the following results:

$$R^2 = .452 \tag{1}$$

variable	coefficients	standard error	t-stat	P-value
Intercept	80	16	5	5.733 E-7
CIGS	.02	.008	2.5	.0124
$CIGS^2$	-.02	.005	-4	6.3E-5
MALE	2	2.5	.8	.424
BMI	-.8	.267	-3	.0027
BMI*MALE	-.2	.1	-2	.0455

a. According to the regression results, which variables are significant explainers of CV_HEALTH at the 5% level?

b. According to the regression results, what is the average CV_HEALTH score for a non-smoking male with a BMI of 24?

c. How much do the results predict that smoking 10 cigarettes per day decreases CV_HEALTH? How much does smoking 20 cigarettes per day decrease CV_HEALTH?

d. Interpret the estimate of the coefficient β_5.

Problem 7 (10 points) Lana Sociologist is interested in the effect of breast feeding on the cognitive development of infants. She randomly samples 10,000 US 40-year old adults, finds that 6,000 of them were breast fed as babies, while 4,000 of them were not. She finds that the average annual income of the first group is \$64,000, while the average annual income of the second group is \$52,000. The standard deviations are low enough to reject a hypothesis test with the null hypothesis that $\mu_1 = \mu_2$, so she concludes that breast feeding does indeed increase cognitive development, as reflected in the labor market earnings of adults (a wide variety of prior studies have found a positive link between intelligence and labor market earnings).

Give a detailed analysis of the additional information you would need to evaluate Lana's claim. Your answer should include information on what concerns you have about her study, and what additional data you would need to be shown to address those concerns.

Problem 8 (10 points) The percentage returns on asset 1 are normally distributed with mean .14 and variance .0004, while the percentage returns on asset 2 are normally distributed with mean .10 and variance .0016. You have $10,000 to invest, and for legal reasons you must either invest all of your money in asset 1 or invest all of your money in asset 2.

a. Suppose your preferences are like those of investors described by the CAPM. Which asset would you rather invest in? Why?

b. Suppose that you only care about generating a return of $2,000 or more; you don't care about any return less than $2,000. Which asset should you invest in? Why?

Bonus question: (5 points) How many people need to be in a room before the probability that (at least) two of them have the same birthday exceeds one-half? To answer this question, make reasonable assumptions about the joint distribution of birthdays of randomly selected people, and pretend that no one is born on February 29th. Do not attempt to solve this problem unless you have finished the rest of the exam.