Econ 57
Gary Smith
Fall 2012

Final Examination (150 minutes)

No calculators allowed. Just set up your answers, for example, P = 49/52. BE SURE TO EXPLAIN YOUR REASONING. If you want extra time, you can buy time at a price of 1 point a minute; for example, if your test is handed in 10 minutes after the scheduled finish time, 10 points will be subtracted from the test score.

1. Identify the most appropriate statistical test for each of these studies (for example, two-sample t test, ANOVA, or multiple regression):
   a. A study compared the December volume of trading for stocks whose prices had decreased between January and November with other stocks.

   b. A study compared the off-season and on-season GPAs of college students who play a varsity sport in either the spring or fall (but not both).

   c. Twenty four students were asked to read a short article and, when they were done, to guess how long it took them to read the article; 21 of 24 students overestimated the time.

   d. The 24 students in part (c), took an average of 74.8 seconds to read the article but their average guess was 108.2 seconds.

2. Identify the most appropriate statistical test for each of these studies (for example, two-sample t test, ANOVA, or multiple regression):
   a. A study compared the change in median home prices in 116 California cities between 2005 and 2010 by looking at these 2005 data for each city: median price/rent ratio, percent of the population that is college educated, and the percentage change in home prices between 2000 and 2005.

   b. A study compared the prices of 20 brand-name health and beauty products with the prices of the Target store-brand "Up and Up" products that have the same ingredients as the brand-name products.

   c. An Internet consulting company tested commercial web page layouts by randomly directing users to different pages; for example, a background that is a light blue, light red, light green, or white. The company then measured the click-through rate for each layout: the fraction of users who click on a link that gives them more information about a product.

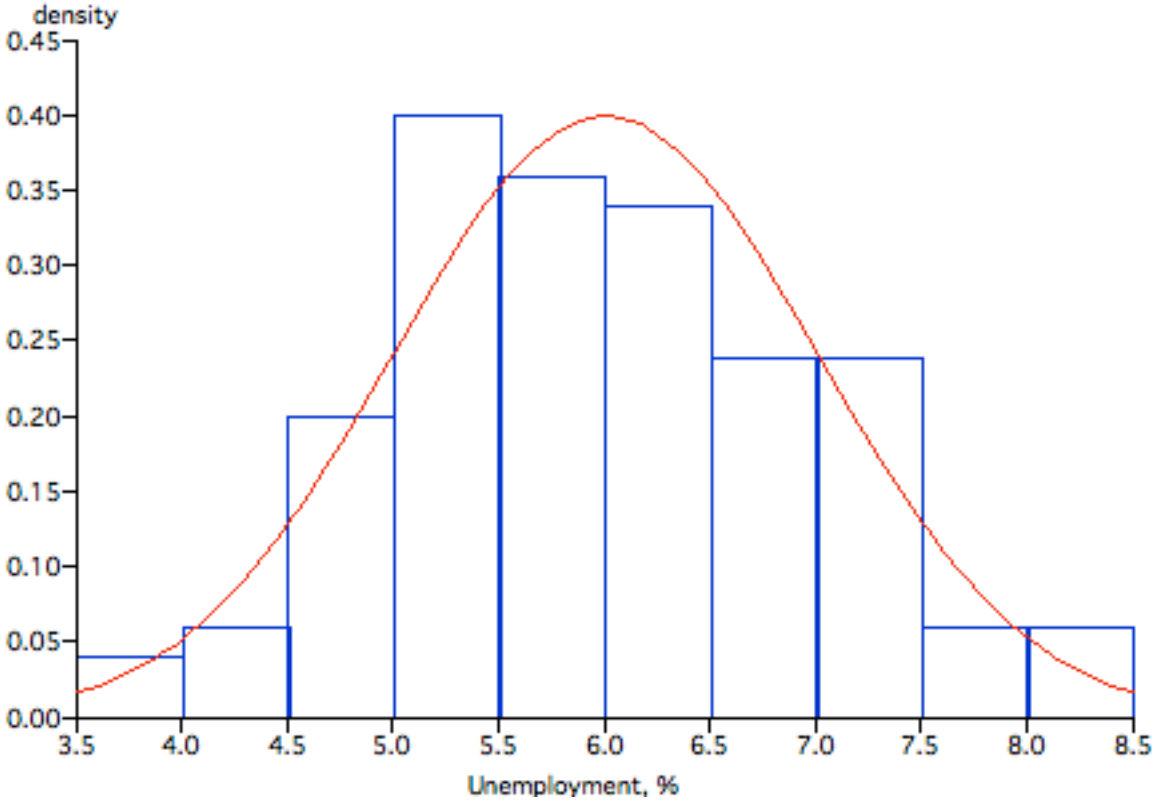
   d. A study tested the theory that the two main determinants of stock prices are the unemployment rate and interest rates.

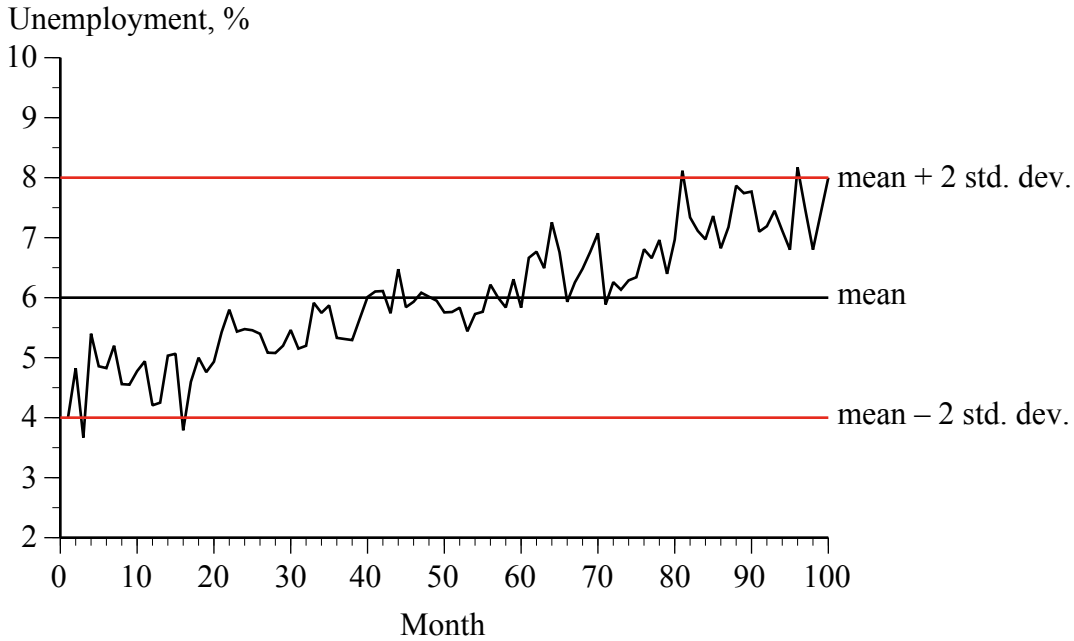
3. Answer this letter to newspaper columnist Marilyn vos Savant:

   I have a really confusing one for you. Let's say my friend puts six playing cards face-down on a table. He tells me that exactly two of them are aces. Then I get to pick up two of the cards. Which of the following choices is more likely? (A) That I'll get one or both of the aces, or (B) That I'll get no aces?

4. Here is an example of "statistics-based economic analysis."

   Statistics-based economic analysis utilizes the known probabilities associated with a bell curve; for example, that there is only a 2.5% chance of getting an observation that is two standard deviations above the mean. Suppose that the unemployment rate over the past 100 months has a mean of 6% with a standard deviation of 1%. If the unemployment rate in month 100 is 8%, there is only 2.5% chance the unemployment rate will go even higher, above 8%, in month 101.





What is the logical error in this reasoning?

5. Evidence from a wide variety of sports indicate that athletic performance is enhanced by our personal assessment of our ability. For example, when both contestants in an arm-wrestling study incorrectly believed the weaker person was stronger than his opponent, the weaker person won ten of twelve matches. When the contestants had correct information about who was the stronger contestant, the stronger person won all twelve matches. Test whether this difference is statistically persuasive.

6. A study looked at the returns on newly issued stock (IPOs) during the two-year period after the stock was issued:

$$R_i = \alpha + \beta_1 R_M + \beta_2 D + \varepsilon$$

where R_i is the return on the IPO stock, R_M is the return on the stock market as a whole during this period, and D = 1 if the stock received a "strong buy" rating from an analyst working for the investment bank handling the IPO, D = 0 otherwise. The t value for the estimate of β_2 was 0.403, leading the author to conclude that, "Based on these results, I reject the null hypothesis that analyst affiliation has a positive effect on two-year returns." Give the correct conclusion.

7. A reviewer for the study described in Exercise 6 suggested that the author instead estimate this equation:

$$R_i = \alpha + \beta_1 R_M + \beta_2 D + \beta_3 DR_M + \varepsilon$$

where DR_M is the dummy variable multiplied by the market return. What advantage (if any) do you see from estimating this equation instead of

a. the equation in Exercise 6, $R_i = \alpha + \beta_1 R_M + \beta_2 D + \varepsilon$?

b. these two separate equations: $R_i = \alpha_3 + \beta_3 R_M + \varepsilon_3$ for stocks that received a strong buy rating and $R_i = \alpha_4 + \beta_4 R_M + \varepsilon_4$ for other stocks?

8. Answer this question to Ask Marilyn:

   I recently returned from a trip to China, where the government is so concerned about population growth that it has instituted strict laws about family size. In the cities, a couple is permitted to have only one child. In the countryside, where sons traditionally have been valued, if the first child is a son, the couple may have no more children. But if the first child is a daughter, the couple may have another child. Regardless of the sex of the second child, no more are permitted. How will this policy affect the mix of males and females?

9. Evergreen Day School is a selective suburban private school with a reputation for academic excellence. Parents often send their children to Evergreen because of the high test scores, and an academic assessment test is an important component of the admissions process. Educational Records Bureau (ERB) tests are administered each year to assess each student's progress. The following table uses percentiles to compare the ERB scores of students admitted to Evergreen in 2008 to the scores of suburban public schools. For example, the Evergreen first-graders scored in the 91.2 percentile compared to suburban public schools. The same students took ERB tests as first, third, and fifth graders.

Grade	Reading Comprehension	Mathematics
1	91.2	90.2
3	67.9	73.2
5	66.1	72.5

   How might a statistician explain these data, other than Evergreen is not as good as public schools? Explain your reasoning in words that the Headmaster could understand.

10. Explain why you agree or disagree: "My null hypothesis for Question 3 (Do you think you will meet your future spouse/life partner in college?) is that P[men] = P[women]. My alternative hypothesis is also my null hypothesis because I do not see reasons to believe that gender would make a difference for this question."

11. Explain why you either agree or disagree:

    The golden theorem proven by the Swiss mathematician Jacob Bernoulli states that a variable will revert to a mean in the long run. For example, a small number of flips of one coin might be mostly heads while a small number of flips of another coin might be mostly tails, but, inevitably, a large number of flips will be 50-50 for both coins. In the same way, earnings might be high for one company and low for another, but, inevitably, will be the same for both.

12. "In the survey, no more than one time in 20 should chance variations in the sample cause results to vary more than plus or minus 3.8 percentage points from the answers that would have been obtained if all adults in the U.S. were surveyed." How many people were surveyed?

13. The following multiple regression equation was used to explain household spending by working families:

$$Y = \alpha + \beta_1 X + \beta_2 D + \varepsilon$$

where Y = annual spending, X = annual income, D = 1 if there are multiple wage earners in the household, 0 if there is only one wage earner. Explain why you either agree or disagree with this interpretation of β_2: "The coefficient of the dummy variable measures the extent to which a household can afford to spend more because families with multiple wage earners tend to have higher income."

14. A highway paint lasts an average of 5 years with a standard deviation of 1 year. A home scientist claims that an additive he discovered will increase the average life to 6 years with no change in the standard deviation. To investigate this claim, the company uses a random sample of 25 products to test these hypotheses:

H_0: $\mu = 5$, $\sigma = 1$

H_1: $\mu = 6$, $\sigma = 1$

The null hypothesis will be rejected if the sample mean is above a cutoff c, and not rejected otherwise.

a. What cutoff c should be used so that the probability of incorrectly rejecting the null hypothesis is $\alpha = 0.05$? (JUST SET UP.)

b. If this cutoff c is used, what is the probability β of incorrectly not rejecting the null hypothesis when the alternative hypothesis is true? (JUST SET UP.)

15. (continuation) Prior to the test, the company believes there is a 50% chance the claim is true and a 50 percent chance that the null hypothesis is true. What is their revised probability that the claim is true if the test rejects the null hypothesis? (JUST SET UP.)

16. Identify the statistical error in the excerpt from the Metro State College of Denver guide to using APA style: "There was not a significant difference between the flattery and no-flattery group in terms of score on the self-esteem measure ($t(34) = 5.20$, $p = .32$)." (Note: $t(34) = 5.20$ means that the t value with 34 degrees of freedom is 5.20.)

17. Pepys asked Newton which of the following three events is most likely:
    a. at least one 6 when six dice are rolled.
    b. at least two 6s when twelve dice are rolled, or
    c. at least three sixes when eighteen dice are rolled?

    Answer Pepys' question.

18. A researcher asked 74 students who drink soda regularly whether they prefer Coke or Pepsi. They were then given a blind taste test with three sodas (Coke, Pepsi, and generic) and asked which soda they liked best:

	Chose Coke	Chose Pepsi	Chose Generic
Coke Drinkers	19	12	16
Pepsi Drinkers	10	7	10
Total	29	19	26

He then tested the null hypothesis that soda drinkers overall are equally likely to choose Coke, Pepsi, or the generic soda: "I used the t distribution to see how much Coke was chosen above the mean. Since this would be a 1-sample numerical test, I used the t distribution:

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} = \frac{29 - (74/3)}{5.132 / \sqrt{3}} = 1.463$$

where s was calculated from

$$s = \sqrt{\frac{(29 - 74/3)^2 + (19 - 74/3)^2 + (26 - 74/3)^2}{3 - 1}} = 5.132$$

and the two-sided p = 0.2810.

Identify the most important error in his analysis.

19. Suppose that the percentage of the popular vote for U. S. President received by the incumbent party's candidate is related negatively to both the unemployment rate and the rate of inflation. A student learned in a macroeconomics class that there is a negative correlation between unemployment and inflation, and consequently omitted the rate of inflation in order to avoid a multicollinearity problem:

$$V = \alpha + \beta U + \varepsilon$$

Predict the effect of this omission on the estimated coefficient of the unemployment rate. Explain your reasoning so that a novice will understand your argument.

20. A student used census data to compare the fraction of the New York City population that was white in 1900 and 2000:

	White Population	Total Population
1900	3,368,458	3,437,202
2000	3,579,700	8,008,278

$$Z = \frac{\dfrac{3{,}368{,}458}{3{,}437{,}202} - \dfrac{3{,}579{,}700}{8{,}008{,}278}}{\sqrt{\dfrac{p(1-p)}{3{,}437{,}202} + \dfrac{p(1-p)}{8{,}008{,}278}}}$$

where

$$p = \frac{3{,}368{,}458 + 3{,}579{,}700}{3{,}437{,}202 + 8{,}008{,}278}$$

What fundamental logical difficulty do you see with applying a statistical test to these data?