Econ 57
Gary Smith
Fall 2013

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. Answer this question to Ask Marilyn, assuming that each birth is an independent event with an equal chance of being male or female: "If you have four children, they may all be of one sex, there may be three of one sex and one of the other sex, or there may be two of each. Which is most likely?"

2. *Discover* magazine once reported that 90% of the passengers who survived airplane crashes had thought ahead of time about how they would exit the plane if it crashed. The magazine recommended that all passengers do likewise. Why do the data they cite not provide persuasive support for their recommendation?

3. A 1950s study found that married men were in better health than men of the same age who never married or were divorced, suggesting that the healthiest path is for a man to marry and never divorce.
   a. Explain why there might be sampling bias in this study.

   b. Suppose that marriage is generally bad for a man's health. Explain how it could still be true that men who marry and stay married are in better health than: (i) men who don't marry; and (ii) men who marry and then divorce.

5. Long ago, the astragali (heel bones) of animals were used as dice. An astragalus has four sides. Experiments have shown that the probabilities of each of these four sides are 0.39, 0.37, 0.12, and 0.12. A game played in ancient Greece was to roll four astragali simultaneously, with the best outcome being a "Venus," in which each of the four different sides appears. What is the probability of rolling a Venus?

6. A study in one state compared the traffic fatality rates (number of fatalities per miles driven) on highways with 55, 65, and 75 miles per hour speed limits. They found that highways with a 75 mph speed limit had the lowest fatality rate and highways with a 55 mph speed limit had the highest fatality rate. Traffic fatalities could evidently be reduced by raising speed limits, perhaps because people pay more attention when they are driving fast. What is the biggest *statistical* problem with this study?

7. Least squares regression minimizes which of the following sums? Explain your reasoning.

   a. $\sum_{i=1}^{n}(Y_i - X_i)^2$

   b. $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

   c. $\sum_{i=1}^{n}(X_i - \bar{X})^2$

   d. $\sum_{i=1}^{n}(Y_i - a - bX_i)^2$

8. The life of a Rolling Rock tire is normally distributed with a mean of 40,000 miles and a standard deviation of 5,000 miles. Assuming independence, if you buy four tires, what is the probability that

   a. at least one of the tires will last 50,000 miles?

   b. all four tires will last at least 50,000 miles?

   c. the average life of these four tires will be at least 50,000 miles?

9. Identify the most serious statistical problem with this model of household income

$$Y = \alpha + \beta_1 D + \beta_2 E + \beta_3 A + \varepsilon$$

where Y = average household income in state
D = state dummy variable, = 1 for Alabama, 2 for Alaska, 3 for Arizona, etc.
E = average education in state
A = average age in state

10. This regression equation was estimated using data for 481 California Census blocks that were more than 95% Hispanic in 2008

$$P = 0.843 + 0.00023Y - 0.00417A, \quad R^2 = 0.0025$$
$$[97.12] \quad [1.18] \quad [8.24]$$

where P = percentage of the two-party vote received by Barack Obama in the 2008 Presidential election; Y = median household income of Hispanics, $1000s of dollars; A is the percentage of the population over the age of 65; and the t values are in brackets. Identify the logical error in each of these two conclusions:

a. Income is not statistically significant; the p value is 0.27, meaning that there's a 27% chance we accept the null hypothesis that income has no effect on the Hispanic vote.

b. The effect of A on P is substantial because the t-value is 8.42.

11. Data from a random sample of 30 first-year college students who lived in single rooms and 30 who lived in doubles were used to estimate this equation.

$$Y = 2.31 - \underset{[0.439]}{0.027G} + \underset{[3.036]}{0.144H} + \underset{[0.716]}{0.045R}, \quad R^2 = 0.243$$

where Y = grade point average (4-point scale); G = 1 if male, 0 if female; H = happiness (6-point scale), R = 1 if have roommate, 0 if not, and the t-values are in brackets. Explain the error in this interpretation of the results: "I expected that students who did not have roommates would score low on the happiness scale and consequently have low GPAs. However, the coefficient of the roommate dummy variable is not close to being statistically significant."

12. Explain why you either agree or disagree with this explanation of regression to the mean by Barry B. Bannister, Trust Investment Officer, AmSouthBank, Birmingham Alabama:

Key financial ratios of companies tend, over time, to revert to the mean for the market as a whole. The thesis is easily defended. High returns eventually invite new entrants, driving down profitability, while poor returns cause the exit of competitors, leaving a more profitable industry for the survivors.

13. J. B. Rhine's book *New Frontiers of the Mind* reported results from his ESP experiments. The *New York Times* science editor was enthusiastic. He noted that some people dismiss Rhine's results by arguing that, just like in gambling, people can have a run of luck. The writer rejected this suggestion: "The run-of-luck theory has also been invoked with some disregard of the obvious fact that runs of bad luck will cause runs of good luck in the course of 100,000 trials." Do you agree that a run of bad luck will cause a run of good luck? Explain your reasoning.

14. One battle in the board game Risk involves the attacker rolling two standard six-sided dice and the defender rolling one die. The attacker's highest die is then compared to the defender's die. The defender loses one army if the attacker's highest die is higher than the defender's die (for example, the attacker rolls 6 and 2 and the defender rolls a number less than 6). The attacker loses one army if the attacker's highest die is equal to or lower than the defender's die (for example, the attacker rolls 4 and 2 and the defender rolls 4, 5, or 6). In a single battle,
    a. what is the probability that the attacker will lose an army?

    b. what is the probability that the defender will lose an army?

    c. what is the expected value of the net gain for the attacker?

15. A Russian statistician named Ladislaus Bortkiewicz used the Poisson distribution to investigate whether the number of soldiers killed by horse kicks were randomly distributed among Prussian cavalry corps or seemed to happen to some corps in some years more often than would be expected by chance alone. He collected data for 122 deaths over a 20-year period in 12 cavalry corps. Using the Poisson distribution he calculated the expected number of cases in which a corps would have no deaths in a single year (108.67), the expected number of cases in which a corps would have a single death (66.29) and so on. He compared these to the actual number of cases. It has been suggested that the correspondence between the predicted and observed instances is too good to be true. How would you test this?

Number of Deaths by	Predicted Instances	Observed Instances
0	108.67	109
1	66.29	65
2	20.22	22
3	4.11	3
4	0.63	1
5	0.08	0
6	0.01	0

16. A student used ordinary least squares to estimate the relationship between an automobile's weight and its miles per gallon. She was concerned that she may have omitted an important explanatory variable, like horsepower. Identify two distinct problems that can arise if an important explanatory variable is omitted.

17. Mr. Smith is the father of two. We meet him walking along the street with a young girl whom he proudly introduces as his daughter. What is the probability that Mr. Smith's other child is also a girl? (Assume that boys and girls are equally likely and independent.) Don't just guess. Prove your answer.

18. The Motley Fool's "Foolish Four" investment strategy is, at the beginning of January each year: (a) identify the 10 Dow Jones Industrial Average stocks with the highest dividend yields (dividend/price); (b) of these 10, identify the five stocks with the lowest prices; (c) drop the stock with the lowest price; and (d) invest 40% of your wealth in the second-to-lowest priced stock and 20% each in the other three stocks. The Motley Fool's leaders, Tom and David Gardner, report that during the 20-year period, 1973-1993, the Foolish Four strategy returned an annual average return of 25 percent and that it "should grant its fans the same 25 percent annualized returns going forward that it has served up in the past." Why are you skeptical?

19. Provide a purely statistical explanation for this observation: "Highly intelligent women tend to marry men who are less intelligent than they are."

20. What do you find misleading about this graph, published on the front page of the *Ithaca Times* on December 7, 2000, evidently showing that Cornell students were paying more and getting less?