Econ 57
Gary Smith
Spring 2011

Final Examination Answers

1. These are not random samples. People who choose to drive sports cars may like to drive fast, while people who choose to drive minivans may be concerned about the safety of their children.

2. The lines inside the boxes show that they have the same median. Because outliers pull the mean away from the median, Boxplot 1 has the higher mean. The greater dispersion shows that Boxplot 1 has the higher standard deviation.

3. This the St Petersburg paradox. The probability of the game ending after 1 flip (tails on the first flip) is 1/2; the probability of the game ending after two flips (HT sequence) is 1/4; the probability of the game ending after three flips (HHT sequence) is 1/8; and so on.
   a. The expected value of the payoff is infinite
   $$\mu = 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) + 4\left(\frac{1}{8}\right) + \ldots = \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) + \ldots = \infty$$
   b. But few people would pay more than a few dollars to play, demonstrating that people do not simply maximize expected return.

4. Imagine a scatter diagram with new points placed on top of each of the original points.
   a. The estimates of the slope and intercept do not change.
   b. The standard error of the estimate of the slope (and intercept) declines because the number of observations increased.
   c. The t-value of the estimate of the slope (and intercept) increases because the standard error decreased.
   d. The correlation is unchanged, so the R^2 is unchanged.

5. To win two consecutive games, Smith must win the middle game. So he is better off playing Ernst in the second game. Here is a formal analysis, letting A be the probability of defeating Andrabi and E be the probability of defeating Ernst:

   Andrabi-Ernst-Andrabi: $AE + (1 - A)EA = 2AE - AEA$

   Ernst-Andrabi-Ernst: $EA + (1 - E)AE = 2AE - EAE < 2AE - AEA$ if $E > A$

6. The probability of correctly not rejecting the null hypothesis in a single test is $1 - 0.05 = 0.95$. The probability of incorrectly rejecting the null hypothesis in at least one case is equal to one minus the probability of not rejecting the null hypothesis in all six cases: $1 - 0.95^6 = 0.2649$.

7. If p is small then F is large. A statistical test that fails to reject the null hypothesis does not prove the null hypothesis.

8. Regression to the mean.

9. For the three categories (corner, adjacent, and other), 3/15 of the balls are in the corners, 6/15 are adjacent, and 6/15 are other, implying these expected values:

	Observed	Expected
Corner	25	(3/15)(54) = 10.8
Adjacent	22	(6/15)(54) = 21.6
Other	7	(6/15)(54) = 21.6
Total	54	54

Far more corner balls were sunk and far fewer in the other category than would be expected if each ball were equally like to be sunk. The chi-square value is 28.55,

$$\chi^2 = \frac{(25-10.8)^2}{10.8} + \frac{(22-21.6)^2}{21.6} + \frac{(7-21.6)^2}{21.6} = 28.55$$

which decisively rejects the null hypothesis. Statistical software shows that, with 3 - 1 = 2 degrees of freedom, the P value is 0.0000007.

10. A 95% confidence interval is:

$$\frac{X}{n} \pm 1.96\sqrt{\frac{\pi(1-\pi)}{n}} = \frac{12{,}012}{24{,}000} \pm 1.96\sqrt{\frac{(12{,}012/24{,}000)(1-(12{,}012/24{,}000))}{24{,}000}}$$
$$= 0.5005 \pm 0.0063$$

11. Using the binomial distribution,

$$P[11{,}988 \le X \le 12{,}012] = \sum_{x=11{,}988}^{12{,}012} \binom{24{,}000}{X} 0.5^X 0.5^{24{,}000-X} = 0.1282$$

12. The first two equations show a positive statistically significant relationship between SAT scores and graduation rates and between GPAs and graduation rates. These positive effects persist when we run a multiple regression equation that includes both explanatory variables, although the SAT coefficient and t value drop slightly and the GPA coefficient drops dramatically and is no longer statistically significant. Evidently, SAT scores and grades are positively correlated, so that the omission of either from the multiple regression equation biases the estimated coefficient of the included variable upward.
    a. Yes, the omission of an important explanatory variable that is correlated with an included explanatory variable can bias the estimated coefficient of the included variable, thereby making it statistically significant.
    b. I believe that SAT scores and high school GPAs are positively correlated.
    c. If the multiple regression equation is correct, so that SAT scores and high school GPAs both have positive effects on a school's graduation rate, and these two explanatory variables are positively correlated, then the omission of either variable from the regression equation makes it appear the other variable has a larger effect.

13. Imagine a scatter diagram. The order in which the data are arranged does not matter.

14. The first year is part of the 10-year period, which makes it more likely that there is a positive relationship. He should have compared 1930 with 1931-1940, and so on.

15. a. The intercept α is the salary of a male with no teaching experience; β_1 is the difference between female and male salaries with no experience ($\alpha + \beta_1$ is the salary of a female with no teaching experience); β_2 is the increase in male salaries for each additional year of experience; β_3 is the difference between female and male salary increases for each additional year of experience ($\beta_2 + \beta_3$ is the increase in female salaries for each additional year of experience).
    b. A comparison of the group means doesn't take experience into account.
    c. Suppose salaries rise with experience. If there is no discrimination and males happen to have more experience than females, a regression equation can show no discrimination while a difference-of-means test does.
    d. If there is discrimination and males happen to have less experience than females, a regression equation can show discrimination while a difference-of-means test does not.

16. The probability of a particular Venus sequence is (0.39)(0.37)(0.12)(0.12). There are (4)(3)(2)(1) = 24 possible Venus sequences, since any of the four sides can appear on the first astragalus, any of the three remaining sides on the second astragalus, two on the third astragalus, and only the last side on the fourth astragalus. Thus, the probability of a Venus is 24(0.39)(0.37)(0.12)(0.12) = 0.0499, almost exactly 5%.

17. His procedure requires multiple tests. He should have used the chi-square distribution for a single test.

18. The sample sizes don't add up, standard errors cannot be negative, and R-squared cannot be larger than 1.

19. Only if the probability of being struck by a car is equal to the probability of wearing dark clothing. Using Bayes' Rule:

$$P[S \text{ if } D] = \frac{P[S]P[D \text{ if S}]}{P[D]}$$
$$= P[D \text{ if S}] \text{ if } P[S] = P[D]$$

This can also be demonstrated by a contingency table using a hypothetical population of 1,000:

	Struck by Car	Not Struck by Car	Total
Dark clothes	A	D – A	D
Not Dark Clothes	S – A	1000 + A – S – D	1000 – D
Total	S	1000 – S	1000

We want A/D = A/S = 0.8, which is only true if D = S

20. They are different. Suppose a student gets a very high score on the midterm. Regression to the mean implies that this student's score on the final exam is likely to be closer to the average score. The fallacious law of averages says that the score on the final exam will likely be below average.