Econ 57
Gary Smith
Fall 2014

Final Examination Answers

1. The finance group's figure omitted zero from the vertical axis and this magnified the ups and downs in the data, allowing us to detect changes that might otherwise be ambiguous. However, once zero has been omitted, the graph is no longer an accurate guide to the magnitude of the changes. Here the drop in revenue was only 2%, but the height of the line drops by 40%.

2. The probability of any sequence of 4 aces and another card is

$$\frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{1}{49}\frac{48}{48}$$

There are five such sequences, so the probability of 4 aces is

$$5\frac{4}{52}\frac{3}{51}\frac{2}{50}\frac{1}{49}\frac{48}{48}=0.0000184689$$

or 1 in 54,145.

3. Both costs have risen with overall prices over time, and the correlation is cheesy due to the fact that there are many different subway costs and pizza prices. http://gothamist.com/2012/06/28/the_pizza_principle_is_alive_and_we.php

4. There would have been no effect, because the fitted line that minimizes the sum of squared prediction errors does not depend on the ordering of the data.

5. The probability that at least 1 will dealt is equal to 1 minus the probability that none default: 1 - $0.95^5$ = .2262.

6. Converting to a standardized Z value:

$$P[X>5{,}500{,}000]=P\left[\frac{X-\mu}{\sigma}>\frac{5{,}500{,}000-5{,}000{,}000}{2{,}000{,}000}\right]=P[Z>0.25]=0.4013$$

7. We can use the sampling distribution for the sample mean to compare 6-person and 12-person juries. If X is normally distributed with mean μ and standard deviation σ, then the sample mean is normally distributed with mean μ and standard deviation equal to σ divided by the square root of the sample size.

   Thus the awards for both 6-person and 12-person jury systems are normally distributed with the same mean, $5 million, but there is more variation with 6-person juries.

   Here, the standard deviation of the 6-person system is $816,497,

$$\frac{\sigma}{\sqrt{n}}=\frac{\$2{,}000{,}000}{\sqrt{6}}=\$816{,}497$$

compared with $577,350 for a 12-person system:

$$\frac{\sigma}{\sqrt{n}}=\frac{\$2{,}000{,}000}{\sqrt{12}}=\$577{,}350$$

In general, the standard deviation with 6-person juries is 41.4% larger than with a 12-person juries:

$$\frac{\sigma/\sqrt{6}}{\sigma/\sqrt{12}}=\frac{\sqrt{12}}{\sqrt{6}}=\sqrt{2}=1.414$$

8. This is an example of Simpson's Paradox. The actual data were

| | Treatment A | | Treatment B | |
|---|---|---|---|---|
| | treated | successes | treated | successes |
| Small stones | 87 | 81 (93%) | 270 | 234 (87%) |
| Large stones | 263 | 192 (73%) | 80 | 55 (69%) |
| Combined | 350 | 273 (78%) | 350 | 289 (83%) |

As in this real example, Treatment B could have a lower success rate for both large and small kidney stones, but have a higher overall success rate. The explanation is that Treatment A was more often used for large stones (which have a relatively low success rate), while Treatment B was more often used for small stones (which have a relatively high success rate).

9. We can reason that the probability tree has 9 equally likely branches, 1-1, 1-2, 1-3, 2-1, 2-2, 2-3, 3-1, 3-2, and 3-3, with five of the branches giving even sums. Therefore, the probability of an even sum is 5/9. E's expected value is (+\$1)(5/9) + (-\$1)(4/9) = \$1/9. Since this is a zero-sum game, O's expected value is -\$1/9.

10. An ANOVA test is for numerical data. These are categorical data and it should be a chi-square test.

11. Why should we assume that every team has a 50% chance of winning its home games? This should be difference-in-proportions test of the null hypothesis that a team is equally likely to win at home or on the road. (A chi-square test with a 2-by-2 table would also work.)

12. The numbers in each cell of a contingency table should be the number of counties in that category, not the values of the graduation rate and income. The chi-square should be something like this

| | low income | middle income | high income |
|---|---|---|---|
| low graduation | 5 | 7 | 1 |
| high graduation | 0 | 8 | 5 |

It would be better to do a regression model.

13. These are two versions of the Texas sharpshooter fallacy. The first study draws the target after shooting the gun; the second fires at many targets and only reports the one that was hit.

14. Here are all the possible (equally likely) rank orderings: BRG, BGR, RBG, RGB, GBR, GRB. If blue is preferred to red, that leaves BRG, BGR, GBR and in 2 of these 3 cases green is preferred to red. Written as a Bayesian problem

$$
\begin{aligned}
P[G>R \text{ if } B>R] &= \frac{P[G>R]P[B>R \text{ if } G>R]}{P[G>R]P[B>R \text{ if } G>R]+P[R>G]P[B>R \text{ if } R>G]} \\
&= \frac{0.5(2/3)}{0.5(2/3)+0.5(1/3)} \\
&= 2/3
\end{aligned}
$$

15. A firm's profits fluctuate about its average profit. Regression to the mean occurs because firms with observed profits that are far from the mean tend to have average profits that are closer to the mean. Thus, their profits in any other year will be closer to the mean.

16. Omitting important explanatory variables may bias the estimates of the coefficients of the included variables. The substantial drop in R-squared suggests that the omitted variables were important. Although

the coefficient of female secondary education is now statistically significant, it should be troubling that it has the wrong sign. It is better to respond to a multicollinearity problem by adding more data than by dropping explanatory variables.

17. a. There is perfect multicollinearity since X1 + X2 + X3 = 100 for each country. The model ignores the size of country, which surely affects the total amount of CO2 emissions.
    b. If the explanatory variables add to a constant, one can be omitted. The dependent variable might be changed to CO2/GDP, or the explanatory variables might be multiplied by GDP.
    c. The variable Y doesn't have a t-value. The author is apparently referring to the estimate of the intercept. The other t and p values are for the coefficients of the explanatory variables, not the explanatory variables themselves.

18. a. Omitting the draws, there were 35 winning games in each half. The player won 19 of 35 in the first half and 13 of 35 in the second half. To compare win/loss percentages in the first and second halves, we can use a difference-in-proportions test:

$$\hat{\pi} = \frac{19+13}{35+35} = 0.4571$$

$$Z = \frac{\frac{19}{35} - \frac{13}{35}}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{35} + \frac{\hat{\pi}(1-\hat{\pi})}{35}}} = 1.4396$$

The two-sided p-value is 0.150. This difference seems substantial, 0.543 vs. 0.371.
    b. We can use the binomial distribution to test the null hypothesis that the player had a 1/3 probability of repeating a move. The probability of 7 of fewer repeats is 0.0000000005.

19. The data seem to have been tortured by omitting data that did not support this silly theory. Why the letter D? Why 1875-1930? Why only first names?

20. There is no reason why yes and no answers should be equally likely. Plus, this test doesn't get at the central question, whether women who are at women's colleges feel more successful than those at coeducational colleges. A more plausible test is whether the yes/no answers are independent of the college attended. The expected values, assuming independence, are as follows:

| | Women's College | Coed College | Total |
|---|---|---|---|
| Yes | 36.67 | 30.32 | 67 |
| No | 15.32 | 12.67 | 28 |
| Total | 52 | 43 | 95 |

(For example, (52)(67)/95 = 36.67.) The chi-square value is virtually zero, and far from the 3.84 cutoff required to reject the null hypothesis at the 5 percent level (the P value is 0.88):

$$\chi^2 = \frac{(37-36.67)^2}{36.67} + \frac{(15-15.32)^2}{15.32} + \frac{(30-30.32)^2}{30.32} + \frac{(13-12.67)^2}{12.67} = 0.022$$