Econ 57
Gary Smith
Spring 2022

Final Examination Answers

1. Two-thirds of the data in each time period are overlapping.

2. a. H_0: The coefficient of each explanatory variable is equal to 0.
   test: multiple regression
   b. H_0: The average difference in daily performance is 0
   test: matched-pair test
   c. H_0: The February-December return is not correlated with the January return.
   test: simple regression
   d. H_0: The score on one hole is independent of the score on the previous hole
   test: chi-square test
   e. H_0: The probability of doubles is 1/6
   test: one-sample binomial test

3. a. The entries in the two-way table should be the number of observations, not the percentage of the population.
   b. An appropriate test would be a one-sample test of the null hypothesis that the probability of being a first born or only child is 0.40. This probability is given by the binomial distribution or can be approximated by a normal distribution (in either case, doubling the probability to give a two-sided p-value).

$$P\left[X \geq 48\right] = \binom{100}{48}.4^{48}.6^{52} + \binom{100}{49}.4^{49}.6^{51} + \ldots + \binom{100}{100}.4^{100}.6^{0} = 0.0638$$

$$2p = 0.1276$$

4. The first child can be either a boy or a girl. The expected wait until having a baby of the opposite sex is 1/0.5 = 2, so the average woman with children has 3 children.

5. The ones who left must have been below-average in Oklahoma and above-average in California.

6. Show the names of the variables and the values of the estimated coefficients.

7. a. We can use the multiplication rule:

$$\frac{6}{42}\frac{5}{41}\frac{4}{40}\frac{3}{39}\frac{2}{38}\frac{1}{37} = 0.000000190629$$

$$= \frac{1}{5,245,786}$$

   b. We can use the subtraction rule:

$$1 - \left(1 - \frac{1}{5,245,786}\right)^{1000-1} = 0.00019$$

8. Regression estimates are not possible if all possibilities (like Frosh, Sophomore, Junior, and Senior) are included as dummy explanatory variables.

9. Letting n be the number of miles, the Z statistic is:

$$Z=\frac{\frac{0}{n}-\frac{1}{100{,}000{,}000}}{\sqrt{\frac{1}{100{,}000{,}000}\left(1-\frac{1}{100{,}000{,}000}\right)/n}}$$

If we set Z equal to the value appropriate for a test at the 5% level, we can solve for n.

The solution for n turns out to be very large compared to the number of miles the Google fleet has driven. Setting $Z = 1.96$:

$$1.96\sqrt{\frac{1}{100{,}000{,}000}\left(1-\frac{1}{100{,}000{,}000}\right)/n}=-\frac{1}{100{,}000{,}000}$$

$$1.96^2\frac{1}{100{,}000{,}000}\left(\frac{99{,}999{,}999}{100{,}000{,}000}\right)/n=\left(\frac{1}{100{,}000{,}000}\right)^2$$

$$n=1.96^2\left(99{,}999{,}999\right)$$

$$=384{,}159{,}996.16$$

10. This was a retrospective study with survivor bias. In any group of people, rich or poor, there are bound to be some common characteristics. A valid study would identify the characteristics ahead of time and then compare two groups over time, one group with these characteristics and one without.

11. a. Sex and choice are independent

b. A chi-square test is appropriate. We use the row and column totals

	Joe Biden	Jill Biden	Michael Jordan	Total
females	12	25	10	47
males	13	0	40	53
Total	25	25	50	100

to determine the expected values

	Joe Biden	Jill Biden	Michael Jordan	Total
females	25(47/100)	25(47/100)	50(47/100)	47
males	25(53/100)	25(53/100)	50(53/100)	53
Total	25	25	50	100

The chi-square statistic is

$$\chi^2_{(3-1)(2-1)}=\frac{\left(12-25\frac{47}{100}\right)^2}{25\frac{47}{100}}+\ldots+\frac{\left(40-25\frac{47}{100}\right)^2}{50\frac{53}{100}}=42.83$$

The p-value is minuscule.

12. Regression toward the mean/paradox of luck and skill.

13. a. If the estimated constant and its standard error are correct, then the t-statistic for the constant is much too large (it should be 0.0152); also, the R-squared cannot be negative.

b. Failure to reject the null hypothesis does not prove that the null hypothesis is true.

14. a. A reasonable model is

$$Y = \alpha + \beta_1 D + \beta_2 X + \beta_3 D^* X + \varepsilon$$

where Y = total fine; D = 1 if driver is under the age of 30, 0 otherwise; and X = miles over the speed limit.

b. Positive values of β_1 and β_3 would indicate discrimination against younger drivers.

c. The average fine might be higher for younger drivers, but this is because they tend to drive more miles over the speed limit.

d. The average fines are the same for younger and older drivers, even though older drivers tend to drive more miles over the speed limit.

15. a. The explanatory variables should not be chosen on the basis of t-values.

b. Oomph is gauged by the magnitudes of the estimated coefficients, not by R-squared or t-values.

c. Unless we are Bayesians, we can't put probabilities on the null hypothesis being true.

d. The multiple regression model does not assume that the explanatory variables are independent.

16. a. The probability of being accepted somewhere is equal to one minus the probability of being rejected everywhere: $1 - 0.8^{10} = 1 - 0.107 = 0.893$.

b. If the results are positively related, then the probability of being rejected from every school is larger than $0.8^{10} = 0.107$. Therefore, the probability of being accepted somewhere is reduced. (In the extreme case of perfect correlation, the probability of being rejected everywhere is equal to the probability of being rejected at any individual college (0.8) and the probability of being accepted somewhere is only 0.2.)

17. A 95% confidence interval for a probability π is

$$\text{95\% confidence interval for } \pi: \ \frac{x}{n} \pm Z\sqrt{\frac{\left(\frac{x}{n}\right)\left(1-\frac{x}{n}\right)}{n}}$$

$$= \frac{387{,}444{,}281}{1{,}000{,}000{,}000} \pm 1.96\sqrt{\frac{\left(\frac{387{,}444{,}281}{1{,}000{,}000{,}000}\right)\left(1-\frac{387{,}444{,}281}{1{,}000{,}000{,}000}\right)}{1{,}000{,}000{,}000}}$$

$$= 0.387444 \pm 0.000030$$

18. The 7-month horizon is suspicious and suggests p-hacking. The growth rate of earnings over the next 5 years is not available to investors making buy/sell decisions on December 31 of each year. (In addition, the return data might be measured as relative to the average return on all 30 Dow stocks each year in order to take into account that some years are better than others.)

19. The estimated coefficients and standard errors should be rounded off. Instead of asterisks, the author should show the p-values. The sign of the coefficient of PEG should be discussed; it was expected to be negative. The magnitude of the coefficient of PEG should be discussed (oomph).

20. Show the results in a table (and show how many people were surveyed).