Econ 57
Gary Smith
Fall 2017

Final Examination Answers

1. a. There is self-selection bias in that students may have valid reasons for choosing the colleges they attend. Perhaps almost all of the students who choose Eastern Michigan over Michigan do graduate (they are part of the 61%) and would not have graduated if they had gone to the University of Michigan Ann Arbor.
   b. As with medical trials, we could take students accepted to both colleges and randomly choose which students go to which college. Fortunately, we can't do that.

2. They are all incorrect.
   a. We can't put a probability on the null hypothesis being true unless we do a Bayesian analysis. The p value relates to the probability of observing certain data if the null hypothesis is true, not the probability that the null hypothesis is true if we observe certain data.
   b. The t-value gauges statistical significance, not whether the magnitude of the estimated coefficient is substantial.
   c. The R^2 can be low even if some of the explanatory variables are statistically significant.
   d. Which category is given the value D = 1 is arbitrary and does not affect the results.

3. a. The return from Strategy 1 more likely to be positive than negative because it is a normal distribution with a mean of 10%.
   b. The same, because each is a normal distribution with a mean of 10%.
   c. Strategy 2 because the standard deviation on the portfolio of 4 stocks ($20/\sqrt{4}$) is smaller than the standard deviation for any one stock (20)
   d. Again, Strategy 2 because the standard deviation on the portfolio of 4 stocks ($20/\sqrt{4}$) is smaller than the standard deviation for any one stock (20)

4. The returns are normally distorted with a mean of 10%, so there is a 0.5 probability of a return larger than 10%.
   a. $1 - 0.5^4$
   b. Using the binomial distribution,

$$\binom{4}{3} 0.5^3 0.5^1 = 0.25$$

5. A 95% confidence interval describes the confidence we have in our estimate of the population mean, not the dispersion in individual ages. These data do not show that 95 percent (or even a majority) of shoppers are between the ages of 37.7 and 42.5. If the ages have a normal distribution with a mean of 40.1 and a standard deviation of 8.6, the fraction of the shoppers between 37.7 and 42.5 years of age is

$$\begin{aligned} P\left[37.7 < X < 42.5\right] &= P\left[\frac{37.7 - 40.1}{8.6} < \frac{X - \mu}{\sigma} < \frac{42.5 - 40.1}{8.6}\right] \\ &= P\left[-0.279 < Z < +0.279\right] \\ &= 0.22 \end{aligned}$$

6. No. Using the binomial distribution,

$$P[X \geq 89] = \sum_{x=89}^{160} \binom{160}{x} 0.5^x 0.5^{160-x} = 0.0894$$

so that the two-sided p value is 0.1784.

Using a normal approximation to the normal distribution,

$$Z = \frac{\frac{89}{160} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{160}}} = 1.423$$

and P[Z > 1.423] = 0.0774, so that the two-sided p value is 0.1548

7. a. With this many degrees of freedom, the p-value (1-sided or 2-sided) should be much less than 0.03. In fact, the researcher meant 0.03%, or 0.0003.
   b. 0.0383 is the chi-square value, not the p-value.

8. a. In general, multicollinearity among the explanatory variables increases the standard errors and reduces the t values of the estimated coefficients.
   b. There cannot be a multicollinearity problem since there is only one explanatory variable.

9. The first number picked for a group can be anything. Given the first number, the second number picked for that group has to be one of the two remaining numbers that belong in that group. For example, if the first number picked for a group is 7, then the second number selected for that group has to be either a 1 or a 4. If the the second number selected is a 1 or 4, then the third number selected for that group must be the third number that belongs in that group. Thus, the probability that the first three numbers picked will be in the same group is 1(2/8)(1/7). For the second group, given that the first three numbers are correctly placed in the first group, the first number picked can be any of the 6 remaining numbers, the second number picked has to be one of the two numbers that goes with the first number and the third number picked has to be the third number that belongs in that group. This probability is 1(2/5)(1/4). If the first 6 numbers are okay, then the last three numbers must be correct for the third group. Thus the overall probability is

$$1\frac{2}{8}\frac{1}{7}1\frac{2}{5}\frac{1}{4} = \frac{1}{280}$$

10. Theory should come before data. Choosing variables based on their statistical significance can, as here, lead to the inclusion of variables which are coincidentally correlated with the dependent variable and the omission of variables that truly belong in the model, but happen to not be statistically significant with the data used to estimate the model.

11. 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, since Table 6 in the textbook shows that, with 3 – 1 = 2 degrees of freedom, the cutoffs are 5.99 for a test at the 5 percent level and 9.21 for a test at the 1 percent level. (Statistical software shows that the *P* value is 0.0000007.)

12. a. There would be no effect.
    b. There would be no effect. (Think of a scatter diagram for a simple regression model.)
    c. The coefficients are ceteris paribus, and tell us nothing about the effect of A on E.

13. a. We can use a difference-in-means test to determine if the observed difference in the sample means is statistically persuasive evidence against the null hypothesis that there is no difference in the population means.

	Sample Size	Sample Mean	Standard Deviation
Color	20	0.65	0.49
Word	20	0.70	0.57

    Allowing the population variances to differ, the t-value is 0.30 with 37.1 degrees of freedom. The two-sided p-value is 0.768.
    b. The alternative hypothesis is- not rejected.

14. Using Bayes' Rule with this notation: + is labeled criminal, – is labeled non-criminal, C is actually criminal, and NC is actually non-criminal

$$\begin{aligned} P[C \text{ if } +] &= \frac{P[C]P[+ \text{ if } C]}{P[C]P[+ \text{ if } C] + P[NC]P[+ \text{ if } NC]} \\ &= \frac{0.0036(0.895)}{0.0036(0.895) + 0.9964(0.07)} \\ &= 0.044 \end{aligned}$$

    (The 0.36% figure is probably wrong since the authors call this "the crime rate," which is probably the fraction of the population convicted o crimes every year, not the fraction of the male population that has ever been convicted of a crime.)

15. Removing a variable that belongs in an equation simply because it is correlated with another variable is likely to bias the estimated coefficients of the variable left in the model. [Eston Martz, Enough Is Enough! Handling Multicollinearity in Regression Analysis, Minitab blog, 16 April, 2013. http://blog.minitab.com/blog/understanding-statistics/handling-multicollinearity-in-regression-analysis

16. a. A chi-square test with a two-way table works. Here are the expected values, assuming independence:

	Nearsighted	Not Nearsighted	Total
GPA < B+	16.65	12.35	29
GPA ≥ B+	14.35	10.65	25
Total	31	23	54

$$\chi^2 = \frac{(14-16.65)^2}{16.65} + \frac{(17-14.35)^2}{14.35} + \frac{(15-12.35)^2}{12.35} + \frac{(8-10.65)^2}{10.65} = 2.14 \quad , \quad P = 0.114$$

b. There is no assumption of a normal distribution. We can't just invent data by assuming new data would be identical to previous data. That's like saying that I flipped a coin and got a heads; so, if I flip the coin 10 more times, I will get 10 heads. Tripling the numbers triples the observed and expected values and triples the chi-square value, which reduces the p value.

17. Each week, Jill has a 10/18 probability of being selected and an 8/18 probability of not being selected.
    a. The probability of never being selected is $(8/18)^{10}$
    b. The probability of being selected every week is $(10/18)^{10}$
    c. The probability of being selected exactly five times is given by the binomial distribution:

$$P[X=5]=\begin{pmatrix} 10 \\ 5 \end{pmatrix}(10/18)^5(8/18)^5=0.2313$$

18. a. There is no compelling reason why the sample sizes need to be equal. The statistical test will take into account unequal sample sizes.
    b. ANOVA or multiple regression with 2 dummy variables.
    c. A failure to reject the null hypothesis does not prove that the null hypothesis is true.

19. This calculation does not take the sample size into account! The percentage difference are more statistically persuasive, the larger the sample. Here, the chi-square value is 56 times larger: 56*0.3954 = 22.14:

$$\begin{aligned}\chi^2 &= \frac{(10-14)^2}{14}+\frac{(16-14)^2}{14}+\frac{(27-14)^2}{14}+\frac{(3-14)^2}{14} \\ &= 56(0.3954) \\ &= 22.14\end{aligned}$$

20. No. If there is luck involved in famine and plenty, regression toward the mean predicts that an unusually large number of years of plenty will be followed by fewer years of plenty, not an equally large number of years of famine.