Econ 57
Gary Smith
Fall 2010

Final Examination Answers

1. If the null hypothesis is true, the probability of 8 out of 8 correct predictions is $0.5^8 = 0.004$. The two-sided p-value is 2(0.004) = 0.008. There is undoubtedly data grubbing in that Paul may have predicted many other games, perhaps in other sports, that were not reported because he was not so successful. In addition, many animals may have been tested by many people, with only the lucky successes reported. In the 2010 World Cup, there were news stories about several psychic animals, including Mani the Parakeet, Pino the Chimp, Apfelsin the Red River Hog, and Lin Ping the Psychic Panda.

2. a. No. The appropriate null hypothesis is that there is no effect. How would you specify a particular null hypothesis that there is an effect, since there are so many possible values?
   b. No. You reject the null hypothesis at the 5 percent level if the p-value is less than 0.05.
   c. No. The t-test is used when the population standard deviation is estimated and is more appropriate than a Z-test if the sample size is small. For large samples, they are identical.
   d. No. A t-value can be negative, for example, in a one-sample test where the sample mean is less than the null-hypothesis value of the population mean.

3. a. Yes. The expected wait is $1/\pi = 2$ for $\pi = 0.5$
   b. No. The success probability is a fixed parameter; it does not have a distribution.
   c. No. For example, a 0.8 probability of a 1-point increase and an 0.2 probability of a 4-point decline has an expected value of zero.
   d. No. As the number of flips increases, the probability of exactly 50% heads declines.

4. a. Yes. $F = t^2$ and the p-value for the F-test is equal to the two-sided p-value for the t-test.
   b. No. The null hypothesis concerns the population mean, not the sample mean.
   c. No. We reject the null hypothesis at the 5% level if the p-value is less than 0.05. If the p-value is larger than 0.05, then we do not reject the null hypothesis, but this is not the same as rejecting the alternative hypothesis.
   d. No. An F-value can never be negative.

5. Yes, because there is a 2/3 probability that the other side of the card is black. This can be shown with Bayes' Rule, letting BB be the double-black card, WW be the double-white card, and BW be the card that is black in one side and white on the other:

$$P[\text{BB if "B"}] = \frac{P[\text{BB}]P[\text{"B" if BB}]}{P[\text{BB}]P[\text{"B" if BB}] + P[\text{BW}]P[\text{"B" if BW}] + P[WW]P[\text{"B" if WW}]}$$

$$= \frac{(1/3)(1)}{(1/3)(1) + (1/3)(1/2) + (1/3)(0)} = 2/3$$

Similarly, a contingency table with 300 plays of the game shows that of the 150 times that a black card is picked, it comes from BB two-thirds of the time:

	pick B	pick W	Total
BB	100	0	100
BW	50	50	100

WW	0	100	100
Total	150	150	300

6. The omission of the rate of inflation will bias the estimated coefficient of the unemployment towards zero, reducing its absolute value. When unemployment increases, this tends to reduce the incumbent vote but also tends to reduce the rate of inflation, which increases the incumbent vote. These two effects on the incumbent vote tend to offset each other, so that it appears that an increase in the unemployment rate does not have as large a negative effect on the incumbent vote as it actually has.

7. a. No, There is a perfect positive linear relation, but the slope need not be one.
   b. No, if the error terms are mostly positive, the estimated slope and intercept might both be above their true values.

8. John Bohannon, Robin Goldstein, and Alexis Herschkowitsch, Can People Distinguish Pate from Dog Food?," Chance, (23), 2010, 43-46.
   The observed values:

	First	Second	Third	Total
Duck	13	3	2	18
Spam	3	12	3	18
Dog food	2	3	13	18
Total	18	18	18	54

   The expected values if the rows and columns are independent:

	First	Second	Third	Total
Duck	6.0	6.0	6.0	18
Spam	6.0	6.0	6.0	18
Dog food	6.0	6.0	6.0	18
Total	18	18	18	54

   The chi-square value with (3 - 1) (3 - 1) = 4 degrees of freedom is 33.67.

$$\chi^2 = \frac{(13-6)^2}{6} + \frac{(3-6)^2}{6} + \ldots + \frac{(13-6)^2}{6} = 33.67$$

   The p-value is minuscule.

9. The binomial distribution gives the probability of 3 or fewer successes in 18 independent trials, each with an 0.2 success probability:

$$P[x \le 3] = \binom{18}{3} 0.2^3 0.8^{15} + \binom{18}{2} 0.2^2 0.8^{16} + \binom{18}{1} 0.2^1 0.8^{17} + \binom{18}{0} 0.2^0 0.8^{18} = 0.1646$$

   The two-sided p-value is 2(0.1646) = 0.3292

10. a. This equation predicts that a 1-pound increase in the engine size reduces the MPG by 0.0022, or a 1000 pound increase reduces the MPG by 2.2. This seems reasonable.
    b. The 0.65 R-squared means that 65 percent of the variation is used car prices is explained by differences in mileage; this seems reasonable.

c. The variable S is statistically significant at the 5 percent level because the 13.2987 t-value is much larger than 2.
d. No. Multiple regression estimates the effects of a change in any explanatory variable holding the other variables constant, here the effect of engine size on miles per gallon holding car weight constant.

11. The correct answer is (e). The margin for error depends on the size of the sample, not the size of the population (unless the sample is a substantial part of the population, which is not the case here).

12. The p-value < 5% means that, if the null hypothesis (that GPA does not depend on hours of sleep) is true, there is less than a 0.05 probability of observing as much variation in GPA across sleep times as were observed.

13. Here are the expected values:

	Sprain	Strain	Total
Practice	15.556	8.444	24
Game	19.444	10.556	30
Total	35	19	54

There are (2 - 1)(2 - 1) = 1 degree of freedom. The chi-square value is 14.1334, with a p-value of 0.00006.

$$\chi^2 = \frac{(9-15.556)^2}{15.556} + \frac{(15-8.444)^2}{8.444} + \frac{(26-19.444)^2}{19.444} + \frac{(4-10.556)^2}{10.556} = 14.1334$$

14. Instead of calculating the probability of fewer than 2 deaths, they should have calculated the probability of 2 or fewer deaths, which is 0.5:

$$P = \binom{5}{0} 0.5^0 0.5^5 + \binom{5}{1} 0.5^1 0.5^4 + \binom{5}{2} 0.5^2 0.5^3 = 0.5$$

and then doubled this for a two-sided p-value. (Yes, the two-sided p-value is 1.0!)

15. Write the model for the two possible values of the dummy variable

$$\text{incumbent } (D = +1)\text{: } Y = (\alpha + \beta_1) + (\beta_2 + \beta_3)X + \varepsilon$$
$$\text{challenger } (D = -1)\text{: } Y = (\alpha - \beta_1) + (\beta_2 - \beta_3)X + \varepsilon$$

We are essentially estimating two separate simple regression equations relating vote to the unemployment rate, one equation for incumbents and one for other candidates. The difference in the intercepts is equal to $2\beta_1$ and the difference in the slopes is equal to $2\beta_3$. The parameter β_2 is the effect of a one-percentage-point change in the unemployment rate on the vote percentage before we consider whether the candidate is the incumbent or not.

16. In general, suppose that a fraction λ of the null hypotheses tested by a researcher are false, there is a probability α of rejecting the null hypothesis if it is true, and there is a probability β of not rejecting a null hypothesis if it is false. In this exercise, $\lambda = 0.01$, $\alpha = 0.05$, and $\beta = 0.05$.
a. The probability of rejecting a null hypothesis is

$$\begin{aligned} P[\text{reject } H_0] &= P[H_0 \text{ true}]P[\text{ reject } H_0 \text{ if } H_0 \text{ true}] + P[H_0 \text{ not true}]P[\text{ reject } H_0 \text{ if } H_0 \text{ not true}] \\ &= (1-\lambda)\alpha + \lambda(1-\beta) \end{aligned}$$

For $\lambda = 0.01$, $\alpha = 0.05$, and $\beta = 0.05$, this probability is 0.059

b. Using Bayes' Rule,

$$P\left[H_0 \text{ true if reject } H_0\right] = \frac{P\left[H_0 \text{ true}\right]P\left[\text{ reject } H_0 \text{ if } H_0 \text{ true}\right]}{P\left[H_0 \text{ true}\right]P\left[\text{ reject } H_0 \text{ if } H_0 \text{ true}\right] + P\left[H_0 \text{ not true}\right]P\left[\text{ reject } H_0 \text{ if } H_0 \text{ not true}\right]}$$

$$= \frac{(1-\lambda)\alpha}{(1-\lambda)\alpha + \lambda(1-\beta)}$$

For $\lambda = 0.01$, $\alpha = 0.05$, and $\beta = 0.05$, this probability is 0.839.

Using a contingency table:

	reject H0	do not reject H0	Total
H0 true	495	9,495	9,900
H0 false	95	5	100
Total	590	9,500	10,000

a. P[reject H0] = 590/10,000 = 0.059
b. P[H0 true if reject H0] = 495/590 = 0.839

17. The winners are part of all players; we need two independent samples.

18. We deflate dollar values by the price level in order to measure purchasing power. The population is not a dollar value, and the population's purchasing power makes no sense.

19. (Their actual numbers are incorrect, but that is not the issue here!)
    a. They should use the total number of suicides to adjust the expected values, not to adjust the observed values.
    b. By increasing the total number of observed values from 76 to 126, they increased the chi-square value.
    c. I would leave the original 76 observed values as is and use the total number of births in each month to calculate the expected values; for example, for January, the expected value is 76(638/total number of players).

20. [Cole Craddick, "Big Bucks for a Brand Name," December 2010.]
    a. The matched-pair test since it had the higher t-value.
    b. The matched-pair test controls for the possibility that, with two independent samples, one sample might, by chance, choose items that are more expensive because of what they are rather than what store they are sold in.
    c. The correct interpretation of a 95% confidence interval is that there is a 95% probability that a confidence interval calculated in this way will include the actual value of the population mean.