Econ 57
Gary Smith
Fall 2021

Final Examination Answers

1. survivor bias

2. a. H_0: average price change = 0
   test: matched-pair t-test

   b. H_0: average difference in daily returns is 0
   test: matched-pair t-test

   c. H_0: chance of bradycardia for people named Brady is 0.0036
   test: 1-sample binomial test

   d. H_0: average rating is the same for both groups
   test: difference in means t-test

   e. H_0: average chance of COVID-19 infection is the same for both groups
   test: difference-in-proportions Z test or chi-square test

3. a. We can do a matched-pair test of the null hypothesis that the average difference is equal to zero. The steps: calculate the difference between each pair of prices; calculate the mean and standard deviation of these differences; calculate the t-value $t = \frac{\bar{X} - 0}{s/\sqrt{n}}$ and the two-sided p-value for this t-value with 20 – 1 = 19 degrees of freedom. The answers are shown below

	US	UK	US-UK
Big Mac	550	508	42
Quarter Pounder w/ Cheese	520	518	2
Double Quarter Oounder w/ cheese	740	750	10
Hamburger	250	250	0
Cheeseburger	300	301	-1
Double Cheeseburger	450	445	5
McChicken	400	388	12
Filet-o-Fish	380	329	51
Sausage & Egg McMuffin	480	430	50
6-piece chicken nuggets	255	259	-4
French Fries (regular)	320	337	-17
Pancakes	580	477	103
Apple Pie	240	250	-10
McFlurry Oreo	340	258	82
Chocolate Milkshake	520	468	52
Vanilla Milkshake	510	469	41
Strawberry Milkshake	530	458	72
Cappuccino	120	128	-8
Caramel Frappuccino	420	399	21
Hot Chocolate	370	231	139

mean	413.75	382.65	32.1
SD	145.5740	139.5890	42.0925

$$t = \frac{\bar{X}-0}{s/\sqrt{n}} = \frac{32.1-0}{42.0925/\sqrt{20}} = 3.41$$

2p = 0.0029

b. The 413.75 average U.S. calories is 32.1 calories higher than the 382.65 average U.K. calories, an 8.4 percent difference. I think this is substantial, but you may disagree.

4. They are all false.

5. Using Bayes' rule, with "Epic+" meaning that ESM generated a sepsis alert and "Epic–" meaning it did not:

$$P\left[\text{no sepsis if ESM+}\right] = \frac{P\left[\text{no sepsis}\right]P\left[\text{ESM+ if no sepsis}\right]}{P\left[\text{no sepsis}\right]P\left[\text{ESM+ if no sepsis}\right]+P\left[\text{sepsis}\right]P\left[\text{ESM+ if sepsis}\right]}$$

$$= \frac{\left(\frac{35{,}903}{38{,}455}\right)\left(\frac{35{,}903-29{,}775}{35{,}903}\right)}{\left(\frac{35{,}903}{38{,}455}\right)\left(\frac{35{,}903-29{,}775}{35{,}903}\right)+\left(\frac{2{,}552}{38{,}455}\right)\left(\frac{843}{2{,}552}\right)}$$

$$= \frac{\left(\frac{6{,}128}{38{,}455}\right)}{\left(\frac{6{,}128}{38{,}455}\right)+\left(\frac{843}{38{,}455}\right)}$$

$$= \frac{6{,}128}{6{,}128+843}$$

$$= \frac{6{,}128}{6{,}971}$$

$$= 0.88$$

A two-way table can also be used:

	Epic+	Epic–	Total
Sepsis	843	1,709	2,552
No sepsis	6,128	29,775	35,903
Total	6,971	31,484	38,455

The false positive frequency is 6,128/6,971 = 0.88

6. A chi-square test is appropriate here:

	Actual	Expected
rock	27	40
scissors	54	40
paper	39	40
Total	120	120

$$\chi^2_2 = \frac{(27-40)^2}{40} + \frac{(54-40)^2}{40} + \frac{(39-40)^2}{40} = 9.15$$

The p-value is 0.0102

7. Dropping variables because of low *t*-values is generally a bad idea because it is likely to bias the estimated coefficients of the variables left in the model.

8. Regression to the mean, just like the tallest parents tend to be taller than their adult children and the tallest children tend to be taller than their parents.

9. These are inverse probabilities. Instead of looking at the percent of hospitalized persons who were fully vaccinated, we want to compare the fractions of the fully vaccinated and not fully vaccinated who were hospitalized.

   Let V represent fully vaccinated and H represent hospitalized

$$P[H \text{ if } V] = \frac{301}{5{,}634{,}634} = 0.00005342$$

$$P[H \text{ if not } V] = \frac{515-301}{6{,}937{,}546-5{,}634{,}634} = 0.00016425$$

10. There are two ways of being dealt a blackjack: an ace followed by a king, queen, jack, or 10; or a king, queen, jack, or 10 followed by an ace, with respective probabilities (4/52)(16/51) and (16/52)(4/51). The total probability is 2(4/52)(16/51) = 0.0482655

11. Using the binomial distribution with $\pi = 0.5$,

$$P[X \geq 410] = \binom{600}{410} 0.5^{410} 0.5^{600-410} + \binom{600}{411} 0.5^{411} 0.5^{600-411} + \ldots + \binom{600}{600} 0.5^{600} 0.5^{600-600}$$

$$= 4.09 \text{ x } 10^{-20}$$

   The two-sided p-value is 8.18 x 10-20.

12. Each estimated coefficient is *ceteris paribus*, holding the other explanatory variables constant.

13. These statements are all false. (Adapted from Gigerenzer, G. (2004), "Mindless Statistics," *Journal of Socio-Economics*, 33, 587–606.)

14. (1) With 4 seasons and 20 years, there are 80 observations, not 20; (2) R^2 is not a percent but a number between 0 and 1; (3) p-values cannot be negative; (4) for t = 2.62, the p-value is much lower than 0.9894; and (5) with 4 seasonal categories, there should be only 3 dummy variables.

15. a. $p = (1/2)^{10} = 1/1024$

    b. $w = 1/p = 2^{10} = 1024$

    c. 1 – probability that it never happens = $1 - (1-p)^w = 1 - \left(1 - \left(\frac{1}{2}\right)^{10}\right)^{2^{10}} = 1 - \left(\frac{1023}{1024}\right)^{1024} = 0.6323$

   In general:

   a. $p = (1/2)^m$

   b. $w = 1/p = 1/(1/2)^m = 2^m$

c. $1-(1-p)^w = 1-\left(1-\frac{1}{2^m}\right)^{2^m}$

If m = 1 then p = 1/2, w = 2, and $1-(1-p)^w = 1-\left(1-\frac{1}{2}\right)^2 = 0.75$

As m increases, so does x = 2m and

$$\frac{\partial\left(1-\left(1-\frac{1}{x}\right)^x\right)}{\partial x} = \left(1-\frac{1}{x}\right)^x \ln\left(1-\frac{1}{x}\right)\frac{1}{x^2} < 0$$

In the limit,

$$\lim_{m\to\infty} 1-\left(1-\frac{1}{2^m}\right)^{2^m} \to 1-\frac{1}{e} = 1-0.367879 = 0.632121$$

so the answer to Part c is always > 0.632121

16. This is a multiplication problem. The probability that the first tile will not be a bonus tile is 136/144. The probability that the second tile will not be a bonus tile, given that the first tile picked was not a bonus tile is 135/143. Thus the probability that the first two tiles are not bonus tiles is (136/144)(135/143). The probability of picking 144 – 16 = 128 tiles without any being a bonus tile is

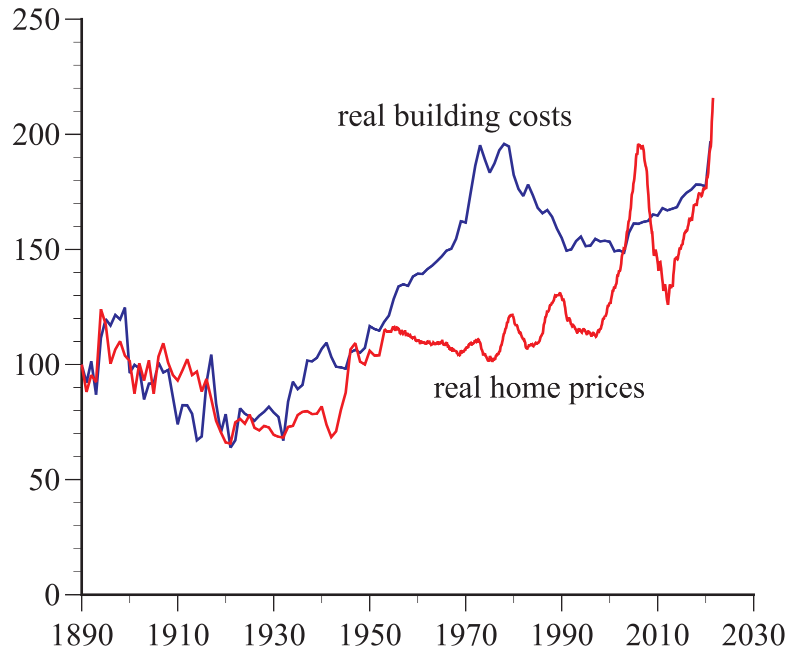
$$\frac{136}{144}\frac{135}{143}\frac{134}{142}\cdots\frac{9}{17}$$

17. First, we want to compare these stocks to other stocks, not to the S&P 500 which includes these stocks. Second, we should identify the five biggest stocks in 2009 and see how they did subsequently. The biggest stocks today may be the ones that performed the best, rather than vice versa.

18. Walker left the 5-hour category out of his graph, evidently because it contradicted his argument.

19. The table shows that there was indeed a 24 percent increase after the spring change and a 21 percent decrease after fall change, as Walker stated, but these fluctuations did not happen, as Walker claimed, the "following day." Saturday night is when we have one less or one more hour of sleep and the following day is Sunday. Instead, the spring increase in the Michigan data was on a Monday and the fall decrease was on a Tuesday, two seemingly random days—except for the fact that they were the only days to have had p-values (0.011 anD 0.044, respectively) below 0.05. On the Sunday after the time changes, the number of heart attacks went in the opposite direction.

20. The two lines would have to start at the same point (for example, 100) in order to see which increased more:



real building costs
real home prices
250
200
150
100
50
0
1890
1910
1930
1950
1970
1990
2010
2030