

Final Examination Answers

- [Robert Gebeloff and Shaila Dewan, Measuring the Top 1% by Wealth, Not Income, *New York Times*, January 17, 2012.] Perhaps the causation runs the other way, in that people in relatively poor health are less likely to accumulate wealth.
- As in E. Scott Geller's study of beer drinking in bars near the Virginia Tech campus, people who buy alcohol in large-volume containers presumably plan to consume large amounts of alcohol. If forced to buy smaller-volume containers, they will simply buy more containers. Some Stanford students created the mocking web site "Make Stanford Safe Again" with this graphic:

Stanford's new alcohol policy:



- [From Daniel Kahneman, *Thinking Fast and Slow*, p. 166.] Most people guess the answer is around 80%, but the correct Bayesian probability is 41%:

$$\begin{aligned}
 P[\text{Blue if "Blue"}] &= \frac{P[\text{Blue}]P[\text{"Blue" if Blue}]}{P[\text{Blue}]P[\text{"Blue" if Blue}] + P[\text{Green}]P[\text{"Blue" if Green}]} \\
 &= \frac{0.15(0.80)}{0.15(0.80) + 0.85(0.20)} \\
 &= 0.41
 \end{aligned}$$

- With a 1/4-point penalty, the expected value from guessing is $(+1)(1/5) + (-1/4)(4/5) = 0$. Without the penalty, the expected value is $(+1)(1/5) + (0)(4/5) = 1/5$
- Letting S stand for success and F for failure on any shot, Blaine's winning sequences are

sequence	probability
S	$(1/3) = 1/3$
FFS	$(2/3)(1/3)(1/3) = (1/3)*b$
FFFFS	$(2/3)(1/3)(2/3)(1/3)(1/3) = (1/3)*b^2$
FFFFFFFS	$(2/3)(1/3)(2/3)(1/3)(2/3)(1/3)(1/3) = (1/3)*b^3$

...

where $b = (2/3)(1/3) = 2/9$. The total probability is

$$P = \frac{1}{3}(1 + b + b^2 + \dots) = \frac{1}{3}\left(\frac{1}{1-b}\right) = \frac{1}{3}\left(\frac{1}{1-2/9}\right) = \frac{3}{7}$$

6. There should be separate 0-1 dummy variables for each state, instead of one dummy variable for all 50 states. There is no reason why D going from 1 to 2 (from Alabama to Alaska) has the same effect on household income as does D going from 2 to 3 (from Alaska to Arizona).
7. The coefficient for the roommate dummy measures the effect on GPA of having a roommate, holding constant happiness.
8.
 - a. The standard deviation of the price level across time does not measure the diversity of prices within the index. Index 1 below has a standard deviation of 138.9 while Index 2 has a standard deviation of 26.6, even though there is no variation in the monthly increase for either. It would be better to look at the standard deviation of the monthly differences from trend or the standard deviation of the monthly changes.
 - b. A luxury ratio of 150 means that prime prices have increased 50% more than non-prime prices, and tells us nothing about the relative cost of buying prime and non-prime properties. This is likely saying that an index of car prices is 200 and an index of apples prices is 400; therefore an apple costs twice as much as a car.
 - c. The price of prime property relative to non-prime property could have increased even while the price of prime property is falling. Suppose prime is \$1 million and non-prime \$500,000 initially, and prime falls to \$500,000 while non-prime falls to \$200,000. The ratio has increased from 2.0 to 2.5 even while prime was a disastrous investment (though not as disastrous a nonprime). Buying at \$1,000,000 was not “the right moment to buy prime property in London if one meant to do so.” In addition, the return from housing depends on the rent saving, property taxes, etc.
9. The values *are* the same. If we substitute AA = 1 into the multiple regression equation, we get

$$\begin{aligned}
 Y &= 138.97 + 0.062*M - 47.83 + 0.021M \\
 &= (138.97 - 47.83) + (0.062 + 0.021)M \\
 &= 91.14 + 0.083M
 \end{aligned}$$

the same as the AA simple regression.

10. There is perfect multicollinearity in that 2015 income is always equal to 2014 income plus the change in income between 2014 and 2015. The ceteris paribus coefficients of the income variables cannot be interpreted; for example, the effect on spending of an increase in 2015 income, holding constant 2014 income and the change in income between 2014 and 2015.
11. Bluto had many more at-bats after the All-Star break and Popeye had many more before the break. For example, Popeye was 5 for 20 before the break and Bluto was 30 for 100; after the break, Popeye was 30 for 80 and Bluto was 4 for 10. Overall, Popeye was 35 for 100 and Bluto was 34 for 110.
12.
 - a. Chi-square
 - b. ANOVA
 - c. Multiple regression
 - d. Matched-pair t-test
 - e. Simple regression.

13. [Charles P. Pierce, “Just Maybe...Right?,” *Sports Illustrated*, May 16, 2016, pp. 47 - 49.] This is clear evidence of regression to the mean. These 13 teams generally won more than half their games the remainder of the season (they were above-average teams), but none of them did as well after the first 30 games as they did during the first 30 games.
14. This is like a pick-6 lottery game:
- $$\frac{5}{21} \frac{4}{20} \frac{3}{19} = \frac{1}{133} = 0.0075188$$
15. This is an example of over-fitting the data. The R^2 is 1 because there are 9 explanatory variables for 10 observations, which guarantees a perfect fit, just like two observations for the simple regression model guarantees a perfect fit.
16. Imagine that everyone draws a number out without looking and then everyone turns over their slip of paper simultaneously. Because every student has the same chances, each has a 1-out-of-21 probability of not being selected, whether the student picks the first or last number out of the hat.
17. Coincidental clusters can be found even in random data, including these data from *Standard Deviations* regarding exposure to electromagnetic fields (EMFs) from power lines.
18. (Their actual numbers are incorrect, but that is not the issue here!)
- They should use the total number of suicides to adjust the expected values, not to adjust the observed values.
 - By increasing the total number of observed values from 76 to 126, they increased the chi-square value.
 - 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 $76(638/\text{total number of players})$.
19. Only the cars registered in 1974 lasted 12 years. Most of their cars were registered more recently.
20. This is the fallacious law of averages.