

Final Examination Answers

1. Starting at the top left and going clockwise: $-0.165, -0.988, 0.411, 0.863$
2. Yes, if Gruden was above average at Monday Night Football and below average as a coach. [Hunter Felt, Monday Night Football: Has Jason Witten ruined a US television institution?, The Guardian, December 11, 2018.]
3. A's first card can be anything. The probability that B will be dealt a card from one of the three other suits is $39/51$. If A and B have different suits, the probability that C will be dealt a card from one of the two remaining suits is $26/50$. Therefore the probability of three different suits is $(39/51)(26/50) = 0.3976$.
4.
 - a. Multiple regression.
 - b. Matched-pair
 - c. One-sample mean test
 - d. One-sample probability test
 - e. Matched-pair
5.
 - a. Chi-square
 - b. ANOVA
 - c. One-sample probability test
 - d. Multiple regression
 - e. Simple regression
6.
 - a. $0.58 \pm 1.96 \sqrt{\frac{(.58)(.42)}{5,000}}$
 - b. Tweeters are not a random sample.
7. This is the fallacious law of averages. If anything, if you get 99 *no*'s, maybe something is wrong with your sales pitch.
8.
 - a. The formula should be for the probability of 29 or more, not exactly 29.
 - b. It should be

$$Z = \frac{\frac{27}{47} - \frac{19}{48}}{\sqrt{\frac{0.4842(0.5158)}{47} + \frac{0.4842(0.5158)}{48}}}$$

where the overall estimated success proportion is $(19 + 27)/(19 + 29 + 27 + 20) = 0.4842$.

9. This is like selecting a sample from a probability distribution with a mean of 10 and standard deviation of 30. The probability distribution of the sample mean is approximately normally distributed with a mean of 10% and a standard deviation of $30\% / \sqrt{100} = 3\%$

10. The first two categories each contain 30 days, but the third category contains 305 days (not counting leap year). If deathday is not related to the birthday, every day should be equally likely and we expect far more deaths in the category with 305 days than in a category with only 30 days

11. The proportion $245/387 = 0.633$ seems substantially higher than 50%. Using the binomial distribution,

$$P[X \geq 245] = \binom{387}{245} 0.5^{245} 0.5^{387-245} + \binom{387}{246} 0.5^{246} 0.5^{387-246} + \dots$$

$$= 0.00000009$$

giving a two-sided p value of 0.00000018, which is highly statistically significant. (The price increases and decreases were actually relative to the overall market on those days; if, for example the market went up 2% and the stock went up 1%, this was counted as a decrease.)

12. a. The t test is for the coefficient of Y
 b. The value of β_2 is ceteris paribus, holding income constant
 c. The two-sided p-value is close to 0.05.
 d. R^2 cannot be large than 1
 e. There is no single value of ϵ

13. Data mining.

14. Higher standard errors reduces the t-value, reducing the likelihood of rejecting the null hypothesis.

15. Using Bayes' Rule (with 51%A shorthand for 51% of all voters and 51%S short for 51% of those surveyed):

$$P[51\%A \text{ if } 55\%S] = \frac{P[51\%A]P[51\%S \text{ if } 51\%A]}{P[51\%A]P[51\%S \text{ if } 51\%A] + P[49\%A]P[51\%S \text{ if } 49\%A]}$$

$$= \frac{0.5 \binom{1000}{510} 0.51^{510} 0.49^{490}}{0.5 \binom{1000}{510} 0.51^{510} 0.49^{490} + 0.5 \binom{1000}{510} 0.49^{510} 0.51^{490}}$$

$$= \frac{0.5(0.025230)}{0.5(0.025230) + 0.5(0.011335)}$$

$$= 0.69$$

16. Regression to the mean teaches us that this student's ability is probably below average, but not as far below average as was this test score. The score on the second test is most likely between 52 and 65.
17. This calculation assumes that the four factors are independent, and they might not be. For example, 40-year-olds may be more likely to drive heavy cars and wear seat belts.
18. Maybe many of those who habitually ate lots of cheese in the past are now deceased (survivorship bias). The cheese-eaters who survived may have lower genetic risks, exercise more, or have other characteristics that offset their cheese eating. It is also possible that some people who do not eat cheese today were once big cheese eaters, but gave it up because of health problems. If so, they are more at risk, not because they avoid

cheese, but because they once feasted on it.

19. The quotation is describing the fallacious law of averages. Regression to the mean says that an outcome that is far from average (like a 1 or 6) will probably be followed by something closer to average. See also: <http://www.crazyontap.com/topic.php?TopicId=11017>
20. There are no numerical units on the horizontal axis (and it seems odd that *revised* would come before *original*); the units on the vertical axis are upside down—enrollment seems to be declining when it is actually increasing.