Final Examination Answers

1. There is self-selection bias in that those who visited France for pleasure and returned are less likely to consider the French unfriendly than are people who visited once and never returned.

2. a. The chance of being chosen every week is $(6/15)^{10}$
   b. The probability of never being chosen is $(9/15)^{10}$

3. a. Calculate the standard deviation s of the 7,000 starting salaries in order to calculate a 95% confidence interval:

$$\$48{,}000 \pm t_{6{,}999} \frac{s}{\sqrt{7{,}000}}$$

   b. Do a matched-pair test of the null hypothesis that the average prediction error is zero by calculating each student's difference X between the actual and predicted starting salary:

$$t_{6{,}999} = \frac{\bar{X} - 0}{s/\sqrt{7{,}000}}$$

   where s is the standard deviation of X.

4. Regression toward the mean. There is considerable uncertainty in whether a movie will be a success, and the most successful movies no doubt tended to have more good luck than bad.

5. The probability of at least one player is equal to 1 minus the probability that no player does something with such a low probability. The probability of a player not doing something with a 0.0000477 probability is 1 – 0.0000477. The probability that all 124 players will not do such a thing is $(1 - 0.0000477)^{124}$, and the probability that at least one player will do such a thing is $1 - (1 - 0.0000477)^{124} = 0.0059$.

6. For the 2 categories, half the balls are solid and half striped, implying these observed and expected values:

	Observed	Expected
Solid	31	(1/2)(54) = 27
Striped	23	(1/2)(54) = 27
Total	54	54

   The chi-square value is 1.1852:

$$\chi^2 = \frac{(31-27)^2}{27} + \frac{(23-27)^2}{27} = 1.1852$$

   With 2 – 1 = 1 degree of freedom, the P value is 0.276, which is not nearly sufficient to reject the null hypothesis at the 5 percent level.

7. a. expected value = 0(1) = 0
   b. expected value = $(2.5(50) - 0.01(50)^2)(0.5) + (2.5(-50) - 0.01(-50)^2)(0.5) = 100(0.5) - 150(0.5) = -25$
   This risk-averse investor prefers a safe 0% returns to a 50% chance of winning or losing 50%.

8. This specification for D_2 assumes an unrealistic linear relationship, with the effect of law school attended on salary being twice as big for Yale as for Harvard, and three times as big for Stanford as for Harvard. With

four school categories, there should be three 0-1 dummy variables for law school attended; for example,

$D_2 = 1$ if attended Harvard, $D_2 = 0$ otherwise
$D_3 = 1$ if attended Yale, $D_3 = 0$ otherwise
$D_4 = 1$ if attended Stanford, and $D_4 = 0$ otherwise

9. The prisoner can improve his chances from 50/50. If he rearranges the marbles, putting one white marble in one box and the other 99 marbles in the other box, his chances are improved to 0.5(1) + 0.5(49/99) = 148/198.

10. a. A random number generator was used to determine which page the visitor went to, so we shouldn't expect the number of visitors to be exactly equal.
    b. We could use a chi-square test or a difference-in-proportions test:
    The overall success proportion is

$$\hat{p} = \frac{6{,}927 + 5{,}851}{77{,}729 + 77{,}858} = 0.0821$$

The Z-value is

$$Z = \frac{\frac{6{,}927}{77{,}729} - \frac{5{,}851}{77{,}858}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{77{,}729} + \frac{\hat{p}(1-\hat{p})}{77{,}858}}} = 10.03$$

The two-sided p-value is 1.09710-23.

11. We can assume they play all 5 games and calculate the binomial probability of A winning at least 3 of 5:

$$P[X \geq 3] = \binom{5}{3} 0.6^3 0.4^2 + \binom{5}{4} 0.6^4 0.4^1 + \binom{5}{5} 0.6^5 0.4^0 = 0.68256$$

Alternatively, we can work through the possibilities
There is one way to win the first 3 games, with a probability 0.6(0.6)(0.6)

WWW

There are 3 ways to win 3 of 4 without wining the first 3 games, each with a probability 0.4(0.6)(0.6)(0.6)

LWWW
WLWW
WWLW

There are 6 ways to win 3 of 5 without winning the first 3 games or 3 of the first 4 games, each with a probability 0.4(0.4)(0.6)(0.6)(0.6):

LLWWW
LWLWW
LWWLW
WLLWW
WLWLW
WWLLW

The overall probability is 0.6(0.6)(0.6) + 3(0.4)(0.6)(0.6)(0.6) + 6(0.4)(0.4)(0.6)(0.6)(0.6) = 0.68256

12. Bayes' Rule gives us the probability that it is the fair coin:

$$P[\text{fair if 2 heads}] = \frac{P[\text{fair}]P[\text{2H if fair}]}{P[\text{fair}]P[\text{2H if fair}] + P[\text{not fair}]P[\text{2H if not fair}]}$$

$$= \frac{(2/3)(1/4)}{(2/3)(1/4) + (1/3)(1)}$$

$$= 1/3$$

Or we can think of it this way. Suppose the experiment is repeated 300 times, with each coin selected 100 times. The 3-headed coin will land heads twice in a row 100 times. Each fair coin can be expected to land heads twice in a row 25 times, a total of 50 times that 2 heads in a row comes from a fair coin. So, of the 150 times that a coin lands heads twice in a row, only 50 (one-third) are because the coin is fair and 100 (two-thirsds) is because the coin has heads on both sides.

Therefore, the probability of a heads on a third toss is (1/3)(1/2) + (2/3)(1) = 5/6

13. Twenty percent wear one boot, and the remaining 80 percent average one boot. So, the average number of boots per person is 1. The population of Townberg is 20,000.

14. This is the fallacious law of averages. [Marotta, Stefanie, Horizons looks to revamp Canada's first AI ETF after performance in first year a 'complete disappointment,' The Globe and Mail, October 2019.]

15. Self-selection bias. People who choose to drive sports cars may do so because they want to drive fast. People who choose to drive minivans may do so because they have children; they may also choose to drive cautiously.

16. The natural null hypothesis is H_0: $\pi = 0.5$
Using the binomial distribution (and doubling the p-value for a two-sided test)

$$P[X \geq 62 \text{ if } \pi = 0.5] = \sum_{X=62}^{100} \binom{100}{X} 0.5^X 0.5^{100-X} = 0.0105$$

17. a. matched-pair
    b. multiple regression
    c. simple regression
    d. binomial or one-sample proportions test
    e. chi-square

18. a. All of them since the t-values (estimate divided by standard error) are all larger than 2.
    b. The coefficient of X5 might be negative because, holding square footage constant, people do not want (for example) an extra bedroom at the cost of a smaller living room, or smaller bedrooms.
    c. The parameters of housing regression equations change over time because construction costs, land values, and tastes change.

19. a. 0
    b. 49.49
    c. Yes, t = 2.62 is well above 2
    d. A one-standard deviation increase in tweets is predicted to increase the high temperature by 2°F
    e. This is clearly data mining, as there is no reason for these variables to be related.

20. a. An unusually large number of Californians like reporting UFO sightings.
    b. Some people think fireworks are UFOs.