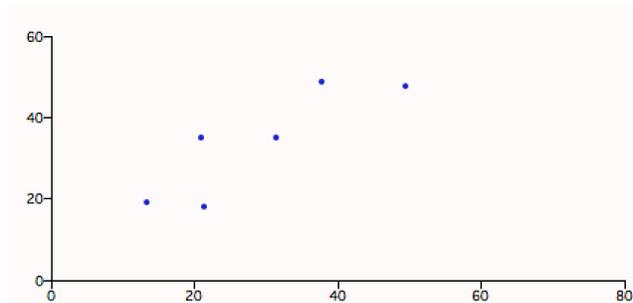
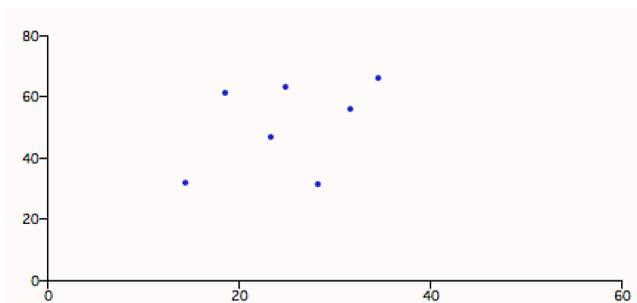
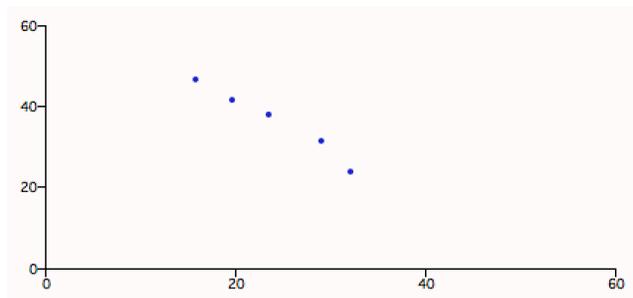
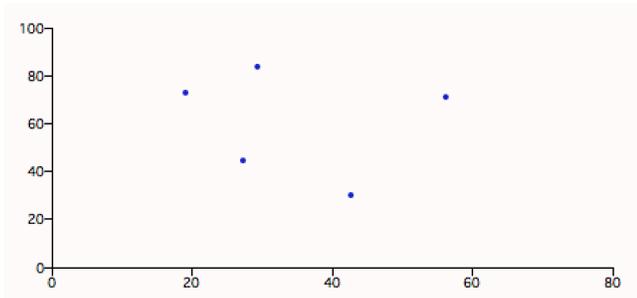


Final Examination (150 minutes)

No calculators allowed. Just set up your answers, for example, $P = 49/52$. BE SURE TO EXPLAIN YOUR REASONING. If you want extra time, you can buy time at a price of 1 point a minute; for example, if your test is handed in 10 minutes after the scheduled finish time, 10 points will be subtracted from the test score.

1. Write the correct correlation value on each graph: -0.165 , -0.988 , 0.411 , 0.863



2. A sports columnist recently wrote that, “Jon Gruden, left [Monday Night Football] to become Oakland’s head coach, a move that has been bad for both the broadcast and the team.” Is it possible for someone who goes from one organization to another to lower the average quality of both organizations?
3. The card game What the Heck! begins with each player being dealt one card from a standard deck of 52 playing cards (4 suits, each suit with 13 cards). If three people (A, B, and C) are playing, what is the probability that the dealt cards will be three different suits? (For example, A dealt a spade, B dealt a heart, and C dealt a club.)

4. Identify the most appropriate statistical test for each of these research hypotheses. You do not need to show any formulas, just identify the test, for example, “difference-in-means t test.”
 - a. People are more likely to have high blood pressure if they eat lots of dairy products and drink very little citrus juice.
 - b. Identical brand-name products cost less at Costco than at Target.
 - c. Pomona College students, on average, get less than 8 hours of sleep a day.
 - d. The Number-1 pick in the Major League Baseball draft is more likely than not to be a Major League All-Stars some day.
 - e. When a company announces a stock split, the return on its stock is usually higher on the day of the announcement than on the day after the announcement.

5. Identify the most appropriate statistical test for each of these research hypotheses. You do not need to show any formulas, just identify the test, for example, “difference-in-means t test.”
 - a. The number of children a woman has (none, 1, 2, 3, 4 or more) is related to the number of children her mother had.
 - b. The ad revenue generated on a travel company’s web page does not depend on whether the main pages shows a picture of a beach, forest, or city.
 - c. A company’s stock price usually rises after it announces the departure of its CEO.
 - d. Home prices in historical Claremont are depend on square footage, number of bedrooms, and number of bathrooms.
 - e. SAT reading and math (no calculator) scores are positively correlated.

6. A monthly survey of consumer sentiment was criticized because only 5,000 people were surveyed every month. In contrast, Twitter tweets can be data mined for millions of people.
 - a. What is the width of a 95% confidence interval for a survey of 5,000 people, if 58% say they are optimistic about the future and 42% say they are not?

 - b. Why might a survey be more reliable than an analysis of Twitter tweets?

7. A sales guru gave an inspirational presentation in which he argued that each *no* brings you closer to a *yes*, because 1 out of every 100 phone calls results in a sale. You just have to get through those 99 *no*’s to get to a *yes*. As a statistician, what do you say?

8. A researcher asked 48 males and 47 females, “Aside from obvious gender differences, who do you resemble more, your biological mother or father?” The responses were

	Males	Females
Biological Mother	19	27
Biological Father	29	20
Total	48	47

What is wrong with each of the following statistical tests?

- a. Test of the null hypothesis that males are equally likely to say “mother” or “father”:

$$P[X \geq 29] = \binom{48}{29} .5^{29} .5^{19}$$

- b. Test of the null hypothesis that males and females are equally likely to say “mother”

$$Z = \frac{\frac{27}{47} - \frac{19}{48}}{\sqrt{\frac{0.5(0.5)}{47} + \frac{0.5(0.5)}{48}}}$$

9. An investor divided her money equally among 100 stocks. She believes that over the next 12 months, each stock’s return will be independently determined from a probability distribution with a mean of 10% and a standard deviation of 30%, What can we say about the probability distribution of the return on her portfolio?

10. Two researchers collected birthday and deathday data on 120 randomly selected deceased celebrities. These data were divided into three categories: deaths that occurred during the 30 days preceding the birthday, deaths that occurred on the birthday or during the 29 days following the birthday, and on other days. A chi-square statistic was used to test the null hypothesis that a person’s deathday is not related to the birthday:

deathdate	Number observed	Number Expected
during 30 days preceding birthday	13	40
on birthday or during 29 days after birthday	8	40
other	99	40
Total	120	120

$$\chi^2 = \frac{(13-40)^2}{40} + \frac{(8-40)^2}{40} + \frac{(99-40)^2}{40} = 130.85$$

What fundamental problem do you see with this statistical analysis?

11. A study of 387 corporate stock splits that occurred during the 10-year period, January 1, 2007, through December 31, 2016, found that 245 of the stocks had price increases on the day the split was announced, while 142 had price decreases. Is this difference substantial and statistically significant at the 5% level?
12. A researcher specified this time-series model of consumer spending,
- $$C = \alpha + \beta_1 Y + \beta_2 D + \varepsilon$$
- where C is consumer spending, Y is disposable income, and $D = 1$ if the stock market went up that year, 0 otherwise. What errors do you see in these reported results?
- $t = 3.64$ for a test of $H_0: Y = 0$
 - The value of β_2 should be positive because when income goes up, the stock market usually goes up too.
 - For β_2 , $t = 2.01$ and the two-sided p-value is 0.95.
 - $R^2 = 1.95$
 - The estimate of ε was 1.13
13. A Yale economics professor and graduate student looked at Bitcoin prices from January 1, 2011, to May 31, 2018 (as far back as there are reliable data), a period during which Bitcoin's price increased from \$0.30 to \$7,530.55, an annual return of 292 percent. Among their findings:
- When Bitcoin's price increased, it tended to keep increasing; so they offered this trading rule: "buy Bitcoin if its price increases more than 20 percent the previous week."
 - Bitcoin's price tended to increase after surges in Google searches for the word *Bitcoin* and decline after a surge in Google searches for the phrase *Bitcoin hack*.
 - They calculated correlations between Bitcoin returns and hundreds of other financial variables, and found that Bitcoin returns are positively correlated with stock returns in the consumer goods and health care industries and negatively correlated with stocks returns in the fabricated products and metal mining industries.
- Why are you skeptical?

14. A researcher argued that, “Multicollinearity inflates the standard errors of the estimated coefficients of the explanatory variables, so that there is a higher likelihood of rejecting the null hypothesis for these coefficients” Why is he wrong?
15. Here’s a stylized example of how polls can be used to predict election outcomes. Suppose that there are only two possibilities: Cameron is preferred by either 51% or 49% of all voters, and our prior probability is that each situation is equally likely. A poll of 1,000 voters finds that 510 (51%) of those surveyed prefer Cameron. What is the posterior probability that 51% of all voters prefer Cameron?
16. In educational testing, a student’s ability is defined as the student’s (theoretical) average score on a large number of tests that are similar with respect to subject matter and difficulty. The student’s score on a particular test is equally likely to be above or below the student’s ability. Suppose that a group of 100 students takes two similar tests and the scores on each test have a mean of 65 with a standard deviation of 14. If a student’s score on the second test is 52, do you predict that this student’s score on the first test was: (a) below 52, (b) 52; (c) between 52 and 65; (d) 65; or (e) above 65? Explain your reasoning.
17. The *Wall Street Journal* and *Washington Post* both reported the results of a 1991 study that estimated the probability that a 40-year-old, sober, seat-belted person driving a heavier-than-average car would have a fatal accident while making a 600-mile automobile trip. The authors of this study calculated this probability by multiplying the overall driver fatality rate by four risk factors. For example, the probability that a heavier-than-average car will have a fatal accident is only 0.77 times the probability that a car of average weight will have a fatal accident. So the overall driver fatality rate was multiplied by 0.77. This adjusted number was then multiplied by 0.68 because the probability that a 40-year-old will have a fatal accident is only 0.68 times the probability that a driver of average age will have a fatal accident. Similar adjustments were made for being sober and wearing a seat belt. What is wrong with this calculation?

18. In 2018 it was reported that people who eat lots of cheese have fewer strokes and less risk of cardiovascular disease. Why might this study be flawed?

19. Explain why you either agree or disagree with this argument:

Regression to the mean follows directly from the law of large numbers, which states that, as the number of trials increases, the observed frequency with which an event occurs will converge to its probability. For example, in a large number of dice rolls, the above-average numbers (4, 5, and 6) will occur half the time and the below-average numbers (1, 2, and 3) will occur half the time. Thus, an above-average roll (like a 5) must be offset by a below-average roll (like a 2), which is regression to the mean.

20. What is misleading about this graph of Affordable Care Enrollment created by a “trusted news source”?

