

Final Examination (150 minutes)

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

1. Seventy two randomly selected college students were asked to taste three unlabeled chocolate chip cookies from a local grocery, and identify the one they liked best. Thirteen students chose Chips Ahoy; 12 chose Lady Lee, and 47 chose supermarket bakery cookies. Are these results statistically significant at the 1 percent level? What is your null hypothesis?
2. California birth records were used to obtain data on the ZIP code residence of foreign-born mothers when they gave birth to a daughter and the ZIP code residence of these daughters when they gave birth to a child. Nine hundred and ninety six of the adult daughters had the same ZIP code as their mother and 4,298 had different ZIP codes. Estimate a 95% confidence interval for the probability that the adult daughter will have the same ZIP code as her mother.
3. The study in the preceding exercise found that of the 4,298 daughters who lived in different ZIP codes, 2,579 lived in ZIP codes with higher median income than their mother's ZIP code and 1,719 lived in ZIP codes with lower income. Show how to calculate the exact two-sided p value for a test of the null hypothesis that a daughter who lives in a different ZIP code is equally likely to live in one with higher or lower income.
4. Answer the preceding question using a normal approximation.

5. The study in the preceding exercises also looked at the adult daughters of California-born mothers and found that of the 5,935 daughters who lived in different ZIP codes, 3,181 lived in ZIP codes with higher median income than their mother's ZIP code and 2,754 lived in ZIP codes with lower median income. Show how to calculate the two-sided p value for a test of the null hypothesis that, among daughters who live in different ZIP codes, the probability of living in a ZIP code with higher income does not depend on whether the mother is foreign-born or California born.
  
6. In a 1982 racial-discrimination lawsuit, the court accepted the defendant's argument that racial differences in hiring and promotion should be separated into eight job categories. In hiring, it turned out that blacks were underrepresented by statistically significant amounts (at the 5 percent level) in four of the eight job categories. In the other four categories, whites were underrepresented in two cases and blacks were underrepresented in two cases, though the differences were not statistically significant at the 5 percent level. The court concluded that four of eight categories was not sufficient to establish a prima facie case of racial discrimination. Assume that the data for these eight job categories are independent random samples.
  - a. What is the null hypothesis?
  
  - b. Explain why data that are divided into 8 job categories might not show statistical significance in any of these job categories, even though there is statistically significance when the data are aggregated.
  
  - c. Explain why data that are divided into 8 job categories might show statistical significance in each of the 8 categories, even though there is not statistically significance when the data are aggregated.
  
7. In the carnival game Aces, there are six cards—two aces, two kings, and two queens. The cards are turned face down and shuffled. The player picks two cards and wins if neither is a ace. If the game is fair, what is the probability of winning?

8. The following regression equation was estimated using data on college applicants who were admitted both to Pomona College and to another college ranked among the top 20 small liberal arts colleges by *U.S. News & World Report*:

$$\hat{Y} = 0.2935 + 0.0293X, \quad R^2 = 0.562$$

(0.0781)      (0.0065)

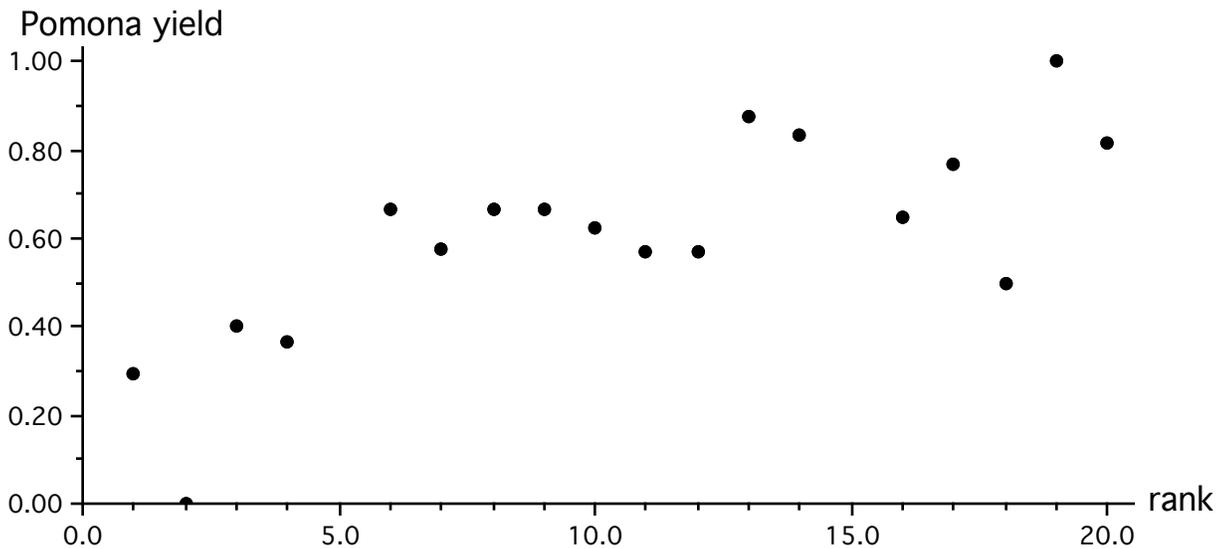
where

( ) = standard deviation

Y = fraction of students who were admitted to both this college and Pomona College that enrolled at Pomona, average value = 0.602

X = *U.S. News & World Report* ranking of this college

- a. Draw the estimated regression line in this scatter diagram of the data:



- b. Use this graph to explain how the estimates 0.2935 and 0.0293 were obtained. (Do not show the formulas; explain in words the basis for the formulas.)
- c. Explain why you are not surprised that the  $R^2$  for this equation is not 1.0.
9. Use the estimated equation reported in the preceding exercise to answer these questions:
- Does the estimated coefficient of X have a plausible value?
  - Is the estimated coefficient of X statistically significant at the 5% level?
  - What is the null hypothesis in question (b)?
  - What is the predicted value of Y for X = 30?
  - Why should we not take the prediction in question (d) seriously?

10. The Australian Bureau of Meteorology uses the monthly air pressure difference between Tahiti and Darwin, Australia, to calculate the Southern Oscillation Index:  $SOI = 10(x - \mu)/\sigma$ , where  $x$  is the air-pressure difference in the current month,  $\mu$  is this particular month's historical average air-pressure difference, and  $\sigma$  is the standard deviation of this month's historical air-pressure difference. Negative values of the SOI indicate an El Niño episode, which is usually accompanied by less-than-usual rainfall over eastern and northern Australia; positive values of the SOI indicate a La Niña episode, which is usually accompanied by more-than-usual rainfall over eastern and northern Australia. Suppose that  $x$  is normally distributed with a mean of  $\mu$  and a standard deviation of  $\sigma$ . Explain why you believe that the probability of an SOI reading as low as -22.8, which occurred in 1994, is closer to  $1.1 \times 10^{-15}$ , 0.011, or 0.110.

11. A study of the effect of driving speed on traffic fatalities estimated the following model by ordinary least squares:

$$Y = \alpha + \beta X + \varepsilon$$

where  $Y$  = deaths per million miles driven and  $X$  = average speed, miles per hour. The researcher used the data shown below:

year	X	Y
1980	57.0	1.6
1984	59.2	1.4
1985	59.5	1.3

- The researcher argued that “supporters of lower speed limits would expect  $\beta$  to be close to one, reflecting a strong, direct effect of  $X$  on  $Y$ .” What is wrong with this argument?
- Do you think that her ordinary least squares estimate of  $\beta$  using these data will be positive or negative?
- Why, no matter what the estimate of  $\beta$ , are you suspicious of her data?

12. Red Lion Hotels ran full-page advertisements claiming that, “For every 50 business travelers who try Red Lion, a certain number don't come back. But 49 of 50 do.” The basis for this claim was a survey of people staying at Red Lion, 98 percent of whom said “they would usually stay in a Red Lion Hotel when they travel.” Use a numerical example to help explain why the survey results do not prove the advertising claim.

13. Answer this letter to Ask Marilyn:

You're at a party with 199 other guests when robbers break in and announce that they are going to rob one of you. They put 199 blank pieces of paper in a hat, plus one marked "you lose." Each guest must draw, and the person who draws "you lose" will get robbed. The robbers offer you the option of drawing first, last or at any time in between. When would you take your turn?

14. A problem of training. The instructors in a flight school adopted a policy of consistent positive reinforcement recommended by psychologists. They verbally reinforced each successful execution of a flight maneuver. After some experience with this training approach, the instructors claimed that contrary to psychological doctrine, high praise for good execution of complex maneuvers typically results in a decrement of performance on the next try. What should the psychologist say in response?

15. Explain why you either agree or disagree with each of the following statements:

- a. A significance test that is significant at the 1% level is also significant at the 5% level.
  
  
  
  
  
  
  
  
  
  
- b. The ANOVA F statistic can be used to test the null hypothesis that the sample means are equal.
  
  
  
  
  
  
  
  
  
  
- c. If the chi-square value is 0, then the p value is 0.
  
  
  
  
  
  
  
  
  
  
- d. If the null hypothesis is  $\pi = 0.5$ , you can use a one-sided p value once you know whether  $x/n$  is larger or smaller than 0.5.
  
  
  
  
  
  
  
  
  
  
- e. In a simple regression, R-squared is equal to the correlation coefficient squared.

16. A survey of 47 sophomores investigated the effect of studying and extracurricular activities on grades:

$$\hat{Y} = 8.809 + 0.069X - 0.085D$$

(0.339)      (0.017)      (0.289)

where  $Y$  = grade point average on a 12-point scale;  $X$  = average hours per week spent studying;  $D = 1$  if the person spends at least 10 hours a week on an extracurricular activity such as work or sports, 0 otherwise; and the standard errors are in parentheses. The researcher concluded that, "The effect of extracurricular activity is not significant and does not lend support to the commonly held notion that extracurricular activity negatively affects grades because it reduces available study time." Explain why the coefficient of  $D$  does not measure the extent to which extracurricular activity affects grades by reducing available study time.

17. Suppose that there are two kinds of households, the Careless and the Careful; 99 percent of households are Careful and 1 percent are Careless. There is a 0.010 probability that fire will destroy a home inhabited by a Careless household, but only a 0.001 probability that fire will destroy a Careful home. If a home is destroyed by fire, what is the probability that it was occupied by a careless household?

18. Explain the error in this Newsweek explanation of the margin of error in a public opinion poll:

The margin of error, calculated according to a textbook statistical formula, varies inversely with the sample size. In general, about 500 responses gives a possible error of 5 percent either way; 2,500 responses decreases it to 1 percent.

19. Explain any possible flaws in this conclusion:

A drinker consumes more than twice as much beer if it comes in a pitcher than in a glass or bottle, and banning pitchers in bars could make a dent in the drunken-driving problem, a researcher said yesterday. Scott Geller, A psychology professor at Virginia Polytechnic Institute and State University in Blacksburg, Va., studied drinking in three bars near campus.... Observers found that, on average, bar patrons drank 35 ounces of beer per person when it came from a pitcher, but only 15 ounces from a bottle and 12 ounces from a glass.

20. Answer this question that a reader asked Marilyn vos Savant, who is listed in the Guinness Book of World Records Hall of Fame for "Highest IQ":

At a lecture on fire safety that I attended, the speaker said: "One in 10 Americans will experience a destructive fire this year. Now, I know that some of you can say you have lived in your homes for 25 years and never had any type of fire. To that I would respond that you have been lucky. But it only means that you are moving not farther away from a fire, but closer to one." Is this last statement correct? Why?