KEY

Introduction to Biostatistics (BIOS 4120)
Breheny

Final (Practice)

1. (17 points) A study of ozone exposure and asthma conducted at the University of Southern California states that "3535 children with no history of asthma were recruited from schools in 12 communities in southern California and were followed up for 5 years."

   (a) Is this a case-control, cohort, or cross-sectional study?

   Cohort

   (b) Suppose the researchers want to compare the likelihood of developing asthma for children who live in high-ozone areas vs. low-ozone areas. Can they estimate relative risks from this study, or is this type of study incompatible with the estimation of relative risks?

   Yes, they can estimate relative risks

   (c) Can they estimate odds ratios, or are odds ratios incompatible with this type of study?

   Yes, they can estimate odds ratios

   (d) In the study, the authors find that children who play three or more sports in high-ozone communities are 3.3 times more likely to develop asthma than children who play no sports. If the investigators wanted to test whether this observed increase could be due to chance, name two tests they might perform.

   χ^2 test; Fisher's exact test

   (e) Of the two tests in part (d), which test would you perform in this situation? Why?

   I would prefer Fisher's test, because it is exact, although with n = 3535, it won't matter

   (f) The investigators do not report a p-value for the comparison in part (d), but they do report a 95% confidence interval of (1.9-5.8). If they were to carry out a test, what can you say about the p-value they would obtain?

   $p < 0.05$

2. (12 points) In an effort to determine whether two drugs used in treatment of thyroid disorders differed in terms of increasing the risk of cancer, researchers at Cambridge performed a study in which rats were randomly assigned to receive one of the drugs. The rats were then exposed to a known carcinogen, and the time until each rat died of cancer was recorded. The first few outcomes for one of the groups of rats is given below. There were 21 rats in this group.

Day	Cancer Death	$n(t)$
142	Yes	21
156	Yes	20
163	Yes	19
198	Yes	18
204	No	17
205	Yes	16
...		

(a) On day 205, how many rats were at risk for dying of cancer?

16

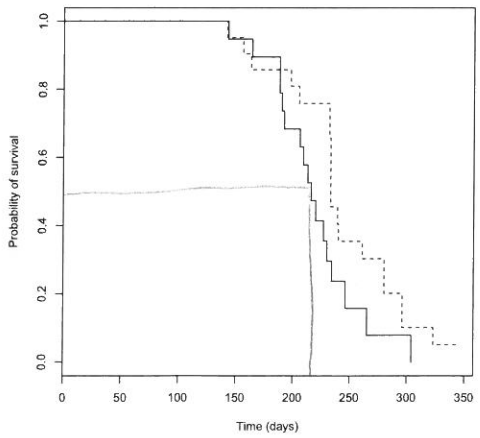
(b) What was the observed probability of dying from cancer on day 205?

$\frac{1}{16}$

(c) Estimate the probability of surviving until day 160. Show your work.

$\frac{20}{21}\left(\frac{19}{20}\right) = .905$

(d) Below are the Kaplan-Meier curves for this study:



What was the median survival time for group with the solid black line?

about 215 days

3. (16 points) Nausea and vomiting are frequent side effects of cancer chemotherapy, which can contribute to the decreased ability of patients to undergo long-term chemotherapy schedules. To investigate the capacity of marijuana to reduce these side effects, researchers at the National Cancer Institute performed a double-blind, randomized, cross-over trial. Fifteen cancer patients on chemotherapy schedules were randomly assigned to receive either a marijuana treatment or a placebo treatment after their first three chemotherapy sessions, and then crossed over to the opposite treatment after their next three sessions. The treatments, which involved both cigarettes and pills, were made to appear the same whether in active or placebo form.

   The outcome of interest was the number of episodes of vomiting and wretching that the subjects experienced. The patients averaged 6.2 episodes while receiving marijuana, compared with 38.1 episodes on the placebo. The mean of the difference between the number of episodes the patients suffered on placebo and on marijuana was 31.9 and the standard deviation of that difference is 35.9.

   (a) Perform an appropriate hypothesis test to address whether or not the observed change in episode frequency could be due to chance.

$$SE = \frac{35.9}{\sqrt{15}} = 9.27$$

$$t = \frac{31.9}{9.27} = 3.44$$

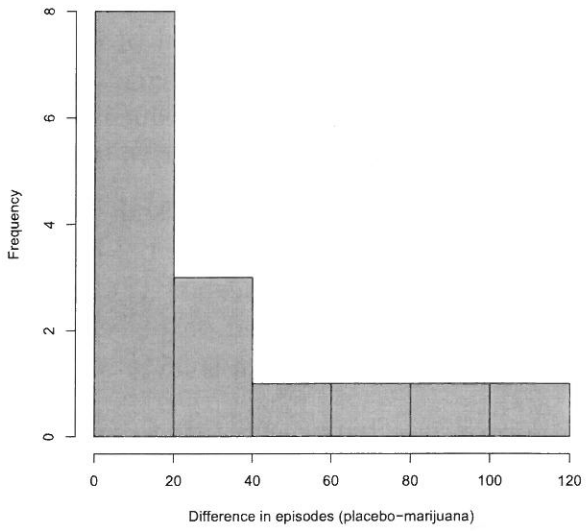
$df = 14$, so $p \approx 2(.002) = .004$

   (b) By how much does the treatment reduce the number of episodes? Provide a 95% confidence interval.

$$t_{95\%} \approx 2.15$$

$$31.9 \pm 2.15(9.27) = (12.0, 51.8)$$

(c) Below is a histogram of the differences in the number of episodes between the treatment and placebo for each patient.



Taking into account this histogram and the description of the study, do you have any concerns about how you analyzed the data in (a) and (b)? Why?

Yes; the data is highly skewed

(d) What might be a better way to analyze these data?

transform the data or use the Wilcoxon signed rank test.

4. (21 points) The H1N1 virus was the first pandemic of the 21st century. Development of effective vaccines for this virus was a top priority in the field of public health in 2009. In one study, researchers randomized adults 18-50 years of age into four groups, with each group receiving vaccines for the H1N1 virus in a different way:

- A: One dose
- B: Two doses at the same time
- C: One dose, then a second dose 1 week later
- D: One dose, then a second dose 2 weeks later

One hundred adults were involved in the study, with 25 subjects randomized to each group. To measure the immune response provoked by the vaccine, the researchers performed an assay which measured the level of H1N1 antibodies in each subject's blood. The outcome of interest was the ratio of this antibody level three weeks after their first dose compared with baseline (before receiving any vaccine).

(a) Instead of analyzing the ratio directly, the researchers took the log of the ratio before conducting any hypothesis tests. Why would they have done this?

presumably to correct for skewness

(b) Suppose the investigators wanted to test whether the mean levels of the log-ratio from (a) were identical in the four groups. How could they have tested this?

ANOVA

(c) The investigators performed all possible pairwise comparisons between the groups above at a nominal significance level of .05. They found that groups B, C, and D all had significantly higher immune responses than group A. What is the false discovery rate of this procedure?

# of tests: 6

# of significant results: 3

$$FDR = \frac{6(0.05)}{3} = 10\%$$

(d) For the series of hypothesis tests described in (e), suppose that the investigators wanted to limit the rate of making a single type I error to 5%. What nominal significance level would they have to use for each individual hypothesis test?

$$\frac{.05}{6} = .0083$$

(e) For the entire sample (all four groups combined), 42 people experienced muscle aches as a side-effect of the vaccination. If you had to construct by hand a confidence interval for the percent of the population that would experience muscle aches after receiving the H1N1 vaccine, what distribution would you use to construct the interval?

normal (z)

(f) With a computer, you could construct an exact confidence interval for the parameter of interest in part (e). What distribution would that interval be based on?

Binomial

(g) Would you expect the intervals from parts (e) and (f) to be close to each other? Why or why not?

Yes; n is fairly large and $\hat{p}$ is not close to 0 or 1

5. (3 points) One study of 22 patients who received gastric bypass surgery found that the average systolic blood pressure of patients before the surgery was 124 mm Hg, and 114 mm Hg after. The investigators tested whether or not this reduction was due to chance; what test do you suppose they used?

paired t-test

6. (9 points) According to a 1986 study, 96.7% of babies received prenatal care, 6.9% of babies had a low birth weight, and 6.4% of babies received prenatal care and had a low birth weight. P = Prenatal care L = Low birth weight

(a) What is the probability that a baby either received prenatal care or had a low birth weight?

$$P(P \cup L) = P(P) + P(L) - P(P \cap L)$$
$$= .967 + .069 - .064 = 0.972$$

(b) What is the probability that a baby had low birth weight, given that the baby received prenatal care?

$$P(L \mid P) = \frac{P(L \cap P)}{P(P)} = \frac{.064}{.967} = .0662$$

(c) What is the probability that a baby had low birth weight, given that the baby did not receive prenatal care?

	L	L^c
P	64	903
P^c	5	28

$$P(L \mid P^c) = \frac{P(L \cap P^c)}{P(P^c)} = \frac{P(L) - P(L \cap P)}{1 - P(P)} = \frac{.069 - .064}{1 - .967} = .1515$$

7. (8 points) Consider a study which examined the association between sleep apnea, as measured by the apnea-hypopnea index (AHI), and diabetes, as measured by fasting insulin levels in the blood.

(a) Is the above study an observational study or a controlled experiment?

Observational; you cannot control who has sleep apnea

(b) What statistical method might the researchers use to analyze their data?

linear regression / correlation

(c) Obesity is positively correlated with both sleep apnea and diabetes. If the above study makes no effort to control for the effect of obesity, will the effect of sleep apnea on diabetes be overestimated or underestimated?

over estimated

8. (15 points) A 1980 study compared the indoor air quality of offices where smoking was permitted with offices where it was forbidden. In the offices where smoking was permitted, the average carbon monoxide concentration was 11.6 parts per million (ppm), with standard deviation 7.3 ppm. In the no-smoking offices, the carbon monoxide concentration was 6.9 ppm, with standard deviation 2.7 ppm.

(a) Suppose that carbon monoxide levels follow a normal distribution; what two tests would be appropriate for determining whether the difference in carbon monoxide levels is larger than you would expect by chance alone?

Welch two-sample t-test

Student's two-sample t-test

(b) Of the two tests in (a), which test seems more appropriate in this case? Why?

Welch's; the SDs are quite different.

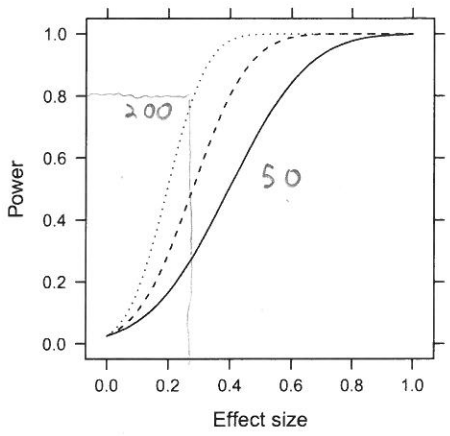
(c) If large outliers are present in the carbon monoxide levels, what test would be better than the tests in part (a)?

Wilcoxon rank-sum test

(d) What kind of plot might you draw in order to inspect whether the data seemed to follow a normal distribution or not?

histogram / box plot

9. (6 points) The two questions below refer to this graph, which plots the power of a one-sample t-test versus effect size:



(a) The three lines correspond to different sample sizes of 50, 100, and 200. Which sample size corresponds to the solid black line?

$n = 50$

(b) For the most powerful curve, what is the smallest effect size that can be detected with 80% power?

≈ 0.25