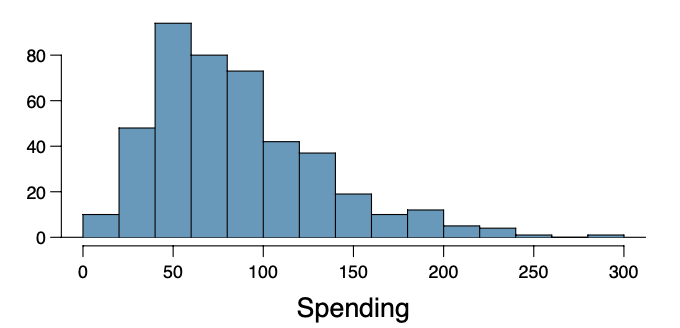
Homework 3: One- and two-sample inference

STAT218

2024-02-22

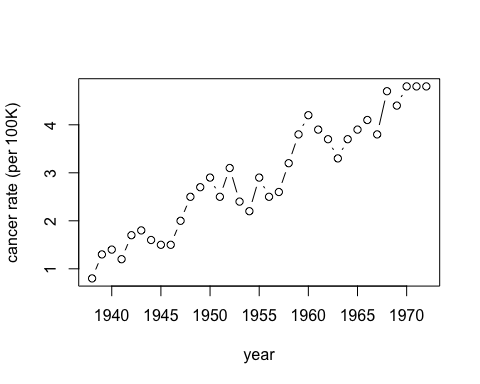
*Instructions: type up your answers and submit your work electronically via Gradescope. Questions with a learning outcome indicated in brackets will be evaluated for credit; other questions are provided for additional practice. You are expected to answer all questions. Note that an R project with datasets and prompts is provided on the class posit.cloud workspace. Please do not submit R codes; show only output or graphics relevant to answering the question. Please skip the question matching step in Gradescope.*

1. **Vu and Harrington exercise 4.4.** The 2010 General Social Survey asked the question: “For how many days during the past 30 days was your mental health, which includes stress, depression, and problems with emotions, not good?” Based on responses from 1,151 US residents, the survey reported a 95% confidence interval of 3.40 to 4.24 days in 2010.
   1. Interpret this interval in context of the data.
   2. What does “95% confident” mean? Explain in the context of the application.
   3. If a new survey were to be done with 500 Americans, would the standard error of the estimate be larger, smaller, or about the same? Assume the standard deviation has remained constant since 2010.
2. [L4] **Vu and Harrington exercise 4.6.** Thanksgiving spending, Part I. The 2009 holiday retail season, which kicked off on November 27, 2009 (the day after Thanksgiving), had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged $84.71. A 95% confidence interval based on this sample is ($80.31, $89.11). Determine whether the following statements are true or false, and explain your reasoning.
   1. We are 95% confident that the average spending of these 436 American adults is between $80.31 and $89.11.
   2. This confidence interval is not valid since the distribution of spending in the sample is right skewed.
   3. 95% of random samples have a sample mean between $80.31 and $89.11.
   4. We are 95% confident that the average spending of all American adults is between $80.31 and $89.11.
   5. A 90% confidence interval would be narrower than the 95% confidence interval.
   6. The margin of error is 4.4.



Answers:

1. False. The average spending of the adults in the sample is a known quantity ($84.71)and does not need to be estimated.
2. False. The sample size is sufficiently large that the model is appropriate.
3. False. The interval is for the population mean, not the sample mean.
4. True. This is the correct interpretation.
5. True. A lower confidence level means lower coverage, which allows interval estimates to be more precise.
6. True. The margin of error is twice the width of the interval. So .
7. **Vu and Harrington exercise 4.21.** Testing for fibromyalgia. A patient named Diana was diagnosed with fibromyalgia, a long-term syn- drome of body pain, and was prescribed anti-depressants. Being the skeptic that she is, Diana didn’t initially believe that anti-depressants would help her symptoms. However after a couple months of being on the med- ication she decides that the anti-depressants are working, because she feels like her symptoms are in fact getting better.
   1. Write the hypotheses in words for Diana’s skeptical position when she started taking the anti-depressants.
   2. What is a Type 1 Error in this context?
   3. What is a Type 2 Error in this context?
8. **Vu and Harrington exercise 4.24.** True or false. Determine if the following statements are true or false, and explain your reasoning. If false, state how it could be corrected.
   1. If a given value (for example, the null hypothesized value of a parameter) is within a 95% confidence interval, it will also be within a 99% confidence interval.
   2. Decreasing the significance level will increase the probability of making a Type 1 Error.
   3. Suppose the null hypothesis is and we fail to reject . Under this scenario, the true population mean is 5.
   4. If the alternative hypothesis is true, then the probability of making a Type 2 Error and the power of a test add up to 1.
   5. With large sample sizes, even small differences between the null value and the true value of the parameter, a difference often called the effect size, will be identified as statistically significant.
9. [L3, L5] Cancer rates and sunspot activity. The dataset cancer contains skin cancer rates per 100,000 people in Connecticut each year from 1938 to 1972. Each year is also classified as following a period of higher than average or lower than average sunspot activity.
   1. Plot the skin cancer rates over time.
   2. Because rates are generally increasing over time, the variable delta measures the deviation of skin cancer rate from the trend. Test the hypothesis that higher-than-average sunspot activity is associated with higher-than-expected skin cancer rates relative to the long term trend. Write a full narrative summary of the test result: state the conclusion; interpret the test result in context; provide supporting statistics; and report and interpret interval and point estimates. Do not provide any R output.
   3. Identify, in words, the population parameter of interest in the test from part (b).



Answers:

1. Plot shown above.
2. Results suggest an association between sunspot activity and cancer rates. We tested the hypothesis that mean deviation of skin cancer rates from long-term trends in Connecticut is the same following years with higher than average and lower than average sunspot activity against the alternative that the mean deviation is higher in years following increased sunspot activity. The data provide strong evidence against the hypothesis of no difference and in favor of the hypothesis that increased sunspot activity is associated with increased cancer rates (*T = 5.0329* on 31.612 degrees of freedom, *p = 0.000009307*). With 95% confidence, the mean deviation following years with high sunspot activity is estimated to be at least 0.294 per 100K greater than the mean deviation following years with low sunspot activity, with a point estimate of 0.443 for the difference.
3. The parameter is mean deviation in cancer rate in CT from the long-term trend.
4. [L4, L5] Comparing self-fertilization with cross-fertilization. The dataset plants contains measurements of plant heights in inches for 15 pairs of plants of the same age; one plant in the pair was grown from a seed from a cross-fertilized flower, and the other was grown from a seed from a self-fertilized flower.
   1. Calculate and interpret an interval estimate for the difference in mean height between plants produced from cross-fertilization and plants produced from self-fertilization.
   2. Test the hypothesis that plants produced from cross-fertilization are taller on average than those produced from self-fertilization. Write a full narrative summary of the test result: state the conclusion; interpret the test result in context; provide supporting statistics; and report and interpret interval and point estimates. Do not provide any R output.

Answers:

1. With 95% confidence, the mean height of plants from cross-fertilized seeds is estimated to be between 0.0032 and 5.2302 centimeters greater than the mean height of plants from self-fertilized seeds.
2. Results suggest that plants from cross-fertilized seeds grow taller. We tested the hypothesis that the mean height of plants grown from cross-fertilized seeds is no greater than the mean height of plants grown from self-fertilized seeds against the alternative that the mean height of plants grown from cross-fertilized seeds is greater than the mean height of plants grown from self-fertilized seeds. The data provide sufficiently strong evidence to reject the hypothesis that the mean height of plants from cross-fertilized seeds is no greater than the mean height of plants from self-fertilized seeds in favor of a positive difference (*T = 2.1474* on 14 degrees of freedom, *p = 0.02488*). With 95% confidence, the difference in mean height between plants from cross-fertilized seeds and plants from self-fertilized seeds is estimated to be at least 0.4704 centimeters, with a point estimate of 2.62 cm.
3. [L4, L5] Tuberculuosis and survival in guinea pigs. The dataset tubercle contains survival times in days of guinea pigs that were randomly assigned to receive a dose of tubercle bacilli, the bacterial pathogen that causes tuberculosis, or to a control group.
   1. Test the hypothesis that exposure to tubercle bacilli causes a decrease in survival time. Interpret the result of the test in context and provide supporting statistics.
   2. Estimate the decrease in survival time using a 99% confidence interval. Interpret the interval in context.

Answers:

1. The data provide sufficiently strong evidence to reject the null hypothesis that mean survival time is no less among rats that received a dose of tubercle bacilli than rats that did not receive a dose against the alternative that survival time decreased (*T = -3.2296* on 97.81 degrees of freedom, *p = 0.0008443*).
2. With 99% confidence, the mean decrease in survival time among rats attributable to the tubercle bacilli dose is estimated to be between 19.161 and 186.239 days.
3. [L5] Medical marijuana. To investigate the capacity of marijuana to reduce the side effects of cancer chemotherapy, researchers performed a double-blind, randomized, crossover trial. Fifteen cancer patients on chemotherapy were randomly assigned to receive either a marijuana treatment or a placebo treatment after their first three sessions of chemotherapy. They were then crossed over to the opposite treatment for their next 3 sessions. The dataset marijuana conatins the resulting measurements.
   1. Test the hypothesis that marijuana treatments cause a reduction in vomiting and retching episodes for cancer patients receiving chemotherapy. Write a full narrative summary of the test results. Do not include any R output.
   2. Assume that you found evidence that medical marijuana caused a reduction in part (a), regardless of your actual result. Would it be advisable to assure a patient that medical marijuana will reduce their vomiting and retching episodes? Why or why not?

Answers:

1. Results suggest that marijuana decreases vomiting and retching episodes among cancer patients receiving chemotherapy. We tested the null hypothesis that the mean number of vomiting and retching episodes was no less while patients were receiving a marijuana treatment compared with a placebo against the alternative that the mean number of episodes was less while patients were receiving the treatment. The data provide sufficiently strong evidence to reject the null hypothesis that the treatments did not decrease vomiting and retching episodes in favor of the alternative of a decrease (*T = -3.4397* on 14 degrees of freedom, *p = 0.001993*). With 95% confidence, the mean decrease in the number of vomiting and retching episodes attributable to the marijuana treatment compared with the placebo is estimated to be at least 15.549 episodes, with a point estimate of 31.867 episodes.
2. No: while we’ve concluded the mean number of episodes decreases due to the treatment for the population, this does not mean that every individual will experience a decrease. While every patient in the study experienced a decrase, and this is highly encouraging, patients cannot be guaranteed a treatment effect.
3. **Vu and Harrington exercise 5.37.** A large farm wants to try out a new type of fertilizer to evaluate whether it will improve the farm’s corn production. The land is broken into plots that produce an average of 1,215 pounds of corn with a standard deviation of 94 pounds per plot. The owner is interested in detecting any average difference of at least 40 pounds per plot.
   1. How many plots of land would be needed for the experiment if the desired power level is 90%? Assume fertilizer treatments (current and new) are allocated in equal proportion among the plots.
   2. Explain in words the meaning of the power level guarantee.
   3. If the experiment will cost $250 per plot and the owner wants to limit the total cost to $10,000, what average difference will they be able to detect at the desired power level?