Lab 007 – Inference for proportions.

EPIB607 - Inferential Statistics[a]

Sampling distribution | Standard error | Normal distribution | Quantiles | Percentiles | Z-scores
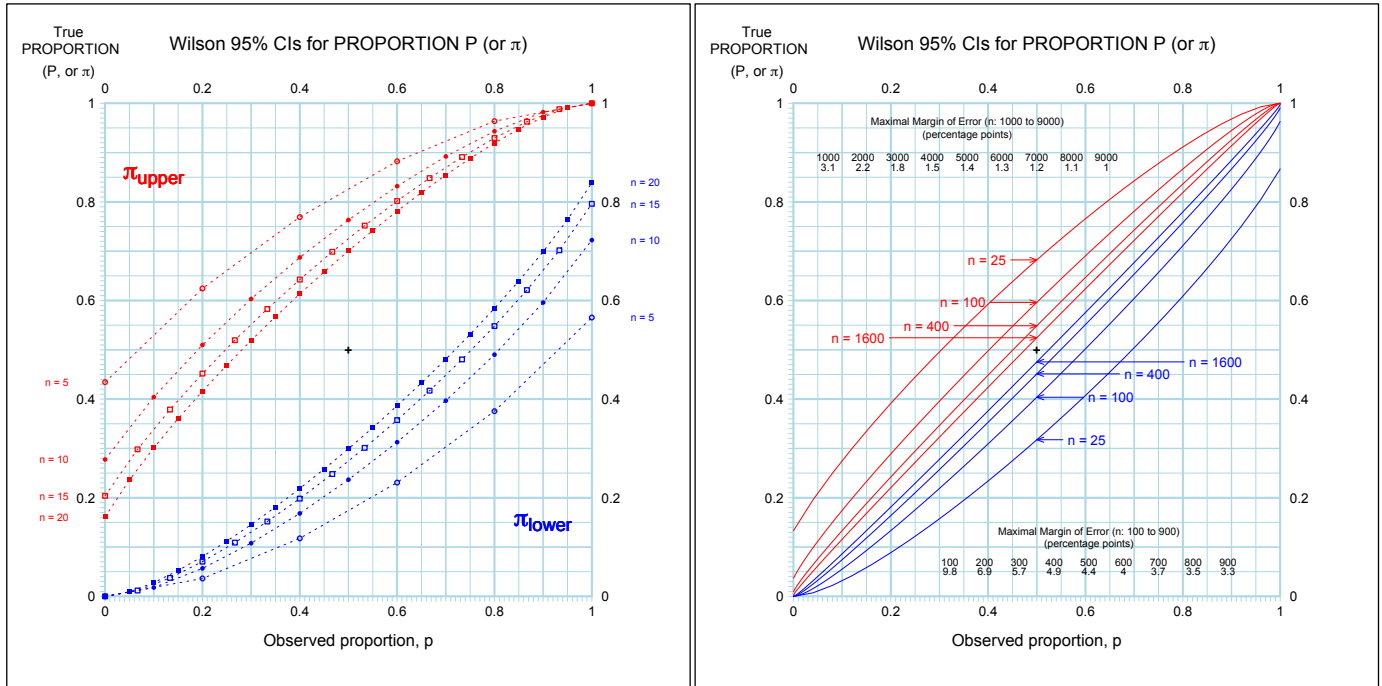

Fig. 1. Wilson 95% CIs for proportion π

1. Drunken cyclists

In the United States approximately 900 people die in bicycle accidents each year. One study examined the records of 1711 bicyclists aged 15 or older who were fatally injured in bicycle accidents between 1987 and 1991 and were tested for alcohol. Of these, 542 tested positive for alcohol (blood alcohol concentration of 0.01% or higher).

    a. To do statistical inference for these data, we think in terms of a model where p is parameter that represents the probability that a tested bicycle rider is positive for alcohol. Find a 99% confidence interval for p.

    b. Can you conclude from your analysis of this study that alcohol causes fatal bicycle accidents? Explain

    c. In this study 386 bicyclists had blood alcohol levels above 0.10%, a level defining legally drunk in many states at the time. Give a 99% confidence interval for the proportion who were legally drunk according to this criterion.

2. Handling contact lenses

Failure to follow recommended contact lens wear and care practices can lead to serious eye infection. A survey of a random sample of 281 Americans who wear contact lenses regularly asked about contact lens practices. The

survey found that 139 respondents do not consistently wash their hands before handing their contact lenses.

    a. Obtain a plus four 99% confidence interval for the proportion p of all American contact lens wearers who do not consistently wash their hands before handing their lenses. Verify that the conditions for your confidence interval are met.

    b. Obtain a large sample 99% confidence interval for the proportion p of all American contact lens wearers who do not consistently wash their hands before handing their lenses. How does this compare to the interval you calculated in part (a)?

    c. The researchers indicated that this survey had a substantial nonresponse rate. How does this information affect your interpretation of the confidence interval in context?

    d. Survey participants simply answered a questionnaire, and no attempt was made to verify the answers. How does this information affect your interpretation of the confidence interval in context?

3. Cancer-detecting dogs

A study was designed to determine whether dogs can be trained to identify urine specimens from individuals with bladder cancer. Dogs were first trained to discriminate between urine specimens from patients with bladder cancer and urine specimens from patients with other conditions. After the training was completed, the dogs had to pick one of seven new urine specimens. Each time, only one of the seven urine specimens came from a patient with bladder cancer. Out of 54 trials, the dogs identified the correct urine specimen 22 times.

    a. If the dogs were simply picking a urine specimen at random, we would expect them to be correct, on average, 1 out of 7 times. The experiment was designed to test whether dogs can perform better than chance. State the null and alternative hypotheses for this test.

    b. Obtain the test statistic and the P-value. What do you conclude?

4. (FALL 2018 Midterm) The US presidential campaign is a very costly affair. One pundit/comedian has suggested that much time and trouble could be saved, if the candidates simply had their height measured. In 20 of the 25 elections where the data are known, the taller candidate has won. Tall people are credited with qualities expected of capable people

    a. In order to test this claim about tall people, we need to test it against a null hypothesis. State the null hypothesis in plain English.

    b. Translate the null and the alternative hypotheses into statistical notation.

    c. Write down the steps, including the R code, to carry out the statistical test. Include a sentence on how you would report your findings.

    d. You don't have access to software to run the code you have suggested in part (c), but can you anticipate what your findings will be? Explain.