Home Exercises 3

Your Name

7.10.2024

Write your name at the beginning of the file as "author:".

1. Return to Moodle by **9.00am, Mon 7.10.** (to section "BEFORE").
2. Watch the exercise session video available in Moodle by **10.00am, Mon 7.10.**
3. If you observe during the exercise session that your answers need some correction, return a corrected version to Moodle (to section "AFTER") by **9.00 am, Mon 14.10.**

Problem 1. Suppose that you have treated $n = 68$ patients with treatment T and for 23 the treatment has been successful.

(a) What is the point estimate of the success proportion of the treatment T?

(b) What is the 95% confidence interval around the point estimate? You can compute it using the method of your choice. Explain what is the interpretation of the 95% confidence interval.

(c) What is the two-sided P-value of the observed data under the null hypothesis that the success proportion is $p = 0.2$? Is it statistically significant at significance level 0.01? Explain what is the interpretation of the P-value.

Problem 2. The following text is adapted from August 2021 article in Lancet:

Implantable loop recorder detection of atrial fibrillation to prevent stroke (The LOOP Study): a randomised controlled trial

"We did a randomised controlled trial in four centres in Denmark. We included individuals without atrial fibrillation, aged 70–90 years, with at least one additional stroke risk factor. Participants (6004) were randomly assigned in a 1:3 ratio to ILR monitoring (treatment; 1501 individuals) or usual care (control; 4503 individuals). In the ILR group, anticoagulation was recommended if atrial fibrillation episodes lasted 6 min or longer."

(a) During the follow-up, atrial fibrillation was diagnosed in 1027 participants: 477 (31.8%) of 1501 in the ILR group versus 550 (12.2%) of 4503 in the control group. Use `binom.test()` to compute the point estimates and 95%CIs for the proportions of individuals with atrial fibrilation detected in ILR group and in control group. (NOTE: Separate applications of `binom.test()` in each group.) Given the 95%CIs, do you expect that the proportions are clearly different from each other statistically?

(b) Use `prop.test()` to compare the proportions of individuals with atrial fibrliation detected in the treatment group and in the control group. What is the 95%CI for the difference in proportions and what is the P-value?

(c) "Stroke or systemic arterial embolism occurred in 318 participants: 67 in the ILR group versus 251 in the control group." Did the different treatment between ILR group and control group lead to difference in proportion of strokes/systemic arterial embolisms between the groups? (Reminder: Total sample sizes were 1501 in ILR group and 4503 in control group.)

Problem 3. Let's study differences between Fisher's test, `prop.test()` and chi-square test as a function of sample size.

(a) Let's get small data example from the early COVID paper with 41 patients.

Huang et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395: 497-506. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30183-5/fulltext

Among the 13 ICU cases, 3 experienced Acute Kidney Injury while among the 28 non-ICU cases none experienced it. We want to study if there is a statistical difference between ICU and non-ICU patients in kidney injury status. Apply Fisher's test, `prop.test()` and Chi-square test to these data and comment which test would you trust in this case. (HINT: Only one of the tests is recommended when we observe small counts below 5.)

(b) Let's redo the analysis of Problem 2(c) with Fisher's test and chi-square tests (in addition to prop.test that was done in 2(c)). "Stroke or systemic arterial embolism occurred in 67 patients in the ILR group versus 251 in the control group. Total sample sizes were 1501 in ILR group and 4503 in control group." Are there now differences between the P-values of the three tests?

Problem 4. (From BMJ.) Over a period of 2 years a psychiatrist has classified by socioeconomic class the women aged 20-64 admitted to her unit suffering from self poisoning. At the same time she has likewise classified the women of similar age admitted to a gastroenterological unit in the same hospital.

Socioeconomic class	self-poisoning	gastroenterology	total
I	17	5	22
II	25	21	46
III	39	34	73
IV	42	49	91
V	32	25	57

Apply both `prop.test()` and `chisq.test()` to test whether there are differences between the socioeconomic classes in the proportion of these two causes of hospitalization.