

# Problem Set 4: Binary outcomes and comparing proportions

## Introduction to Statistical Thinking and Data Analysis

MSc in Epidemiology and MSc in Health Data Analytics, Imperial College London

31 October 2022

### A) Consolidating concepts

A1. Which of the following is true if and only if events A and B are independent?

- a)  $P(A \text{ and } B) = P(A) \times P(B|A)$
- b)  $P(A \text{ and } B) = P(A) \times P(B)$
- c)  $P(A \text{ and } B) = P(A) + P(B)$
- d)  $P(B|A) = P(B) \times P(A|B)/P(A)$

A2. Which of the following describes the relationship between odds ratios and risk ratios?

- a) Odds ratios can be used to approximate risk ratios when the data come from a randomised controlled trial.
- b) Odds ratios can be used to approximate risk ratios when the data come from a case-control study.
- c) If an odds ratio is statistically significant at the  $p < 0.05$  level, the risk ratio will also be significant at the  $p < 0.05$  level.
- d) If the risk ratio is less than 1, the odds ratio is also less than 1.
- e) The risk ratio will always be more extreme than the odds ratio (smaller if the odds ratio is less than 1; larger if the odds ratio is greater than 1).

A3. Which of the following is false about proportions from a sample:

- a) It is used for estimating the probability or risk of suffering an event (e.g. disease diagnosis, episode or death) in the population
- b) The binomial distribution describes its sampling distribution
- c) Sample proportion is defined as  $p = d/n \times h$ ; where  $p$  is proportion,  $d$  is the number of subjects who experience the event,  $n$  is the sample size and  $h$  is the number of individuals who do not experience the event
- d) None of the above

A4. For a given sample size  $n$ , when is the standard error for the sample proportion  $p = x/n$  the largest?

- a) When the probability of success  $\pi = 0.5$ .
- b) When the probability of success  $\pi = 0$  or 1.
- c) It depends on the population standard deviation in the observed number of events  $x$ .
- d) When the probability of success  $\pi = 0.33$

A5. Based on the contingency table below, what is the odds ratio of adverse events in the treatment group versus the control group?

	Adverse events = Yes	Adverse events = No	Total
Treatment	167	490	657
Control	150	193	343
Total	317	683	1000

- a) 3.057
- b) 0.327
- c) 0.439
- d) 0.183

A6. What is the formula for relative risk?

- a)  $p_1 - p_0$
- b)  $p_1/p_0$
- c)  $(p_1/d_1) - (p_0/d_0)$
- d) None of the above

A7. Which measure of effect is most appropriate to use for a statistical test when the sample size is small and only three or four cases are expected?

- a) Risk difference
- b) Risk ratio
- c) Odds ratio
- d) None of the above

A8. What does a risk ratio equal to 1 imply?

- a) That the risk of disease is the same in the exposed and unexposed groups.
- b) That the risk of disease is approximately the same across group comparisons in a case-control study.
- c) That the risk is constant across time in a population.
- d) That an event is unlikely to happen in any given individual.

A9. Contingency tables:

- a) Are useful because they can clearly display a continuous outcome with a categorical exposure.
- b) Are useful because they can help establish whether there is an association between exposure and outcome variables across categories after controlling for additional exposure variables.
- c) Typically display exposure groups in columns and outcome groups in rows
- d) Can use a  $\chi^2$  test as long as the overall total in the table is 20 or greater
- e) None of the above

## B) Practising skills

B1. The dataset `neutron.csv` contains data from a clinical trial comparing two forms of radiotherapy for cancer treatment. Cancer patients were randomly allocated to receive the standard therapy using photon particles or a new form using neutrons. Randomisation was stratified for four sites of cancer. The outcomes of interest is whether the new neutron treatment affects cancer survival compared to standard of care photon therapy. The table below describes the variables in this dataset.

Variable	Description
id	Patient ID
stime	Survival time in days
death	Death status (0 = alive; 1 = dead)
treatment	Treatment received (Neutrons, Photons)
site	Site of cancer (Bladder, Cervix, Prostate, Rectum)
phase	Randomisation phase (0 = first phase; 1 = second phase)
meta	Metastases diagnosed before death (Yes, No)
metatime	Time to metastases (days)
death1year	Death within 1 year of recruitment (Yes, No)

- Estimate the probability of death within one year of cancer diagnosis (assuming that the treatment has no effect on survival) and calculate the 95% confidence interval for the probability.
- Is the probability of death within one year below 60%? Articulate a null and alternative hypothesis to address the question. Construct an appropriate test statistic and calculate a one-sided p-value. Interpret the results of your test.
- Repeat your analysis in (b) applying the *continuity correction*. How do the results change? Which analysis do you prefer and why?
- Use the binomial distribution to calculate the probability observing the number of deaths seen in our sample or fewer, if the true probability of death within one year is 60%. (Hint: use the function `pbinom()` in R. Also think about how to get the same result using the function `dbinom()`.)

B2. In this question, use the `neutron.csv` dataset to assess whether neutron treatment compared to receiving proton treatment affects the probability of death within one year.

- Construct a 2x2 contingency table summarising the relationship between treatment assignment and death within one year. Calculate the sample proportion to estimate the probability of death and the odds of death for each treatment group.
- Estimate the risk difference, risk ratio, and odds ratio and 95% confidence intervals for each outcome. Interpret your estimates for each of these measures of difference.
- State the relevant null hypothesis for each of the measures of difference. Calculate the z-test statistic and p-value for each measure and interpret the results.
- For applied purposes, we typically choose and focus on one measure of difference and a single test statistic. This is specified during the analysis plan before conducting any data analysis. Which measure of difference would you choose and why?
- Conduct a chi-squared test of the null hypothesis that there is no difference in the risk of death in 1 year for patients receiving neutron therapy compared to proton therapy.
  - Calculate the *expected* cell counts for the contingency table constructed in part (a) if the null hypothesis is true that there is no difference in the probability of death by treatment group.

- ii. Calculate the chi-squared test statistic. Calculate the associated p-value using the `pchisq(...)` function in R.
- iii. Check your calculation with the output of the `chisq.test()` function in R. Interpret the result of your hypothesis test.
- f. Compare your chi-squared test statistic and p-value to the test statistics and p-values for the risk difference, risk ratio, and odds ratio calculated in part (b).
- g. Use `chisq.test()` to recalculate the chi-squared test with the Yates continuity correction. How has this changed the result?

B3. In this question, use the `neutron.csv` dataset to assess whether patients who experienced metastasis are more likely to die within one year.

- a. Construct a new variable `meta1year` classifying a binary outcome for patients who experienced metastasis within one year. This should use the variables `meta` and `metatime`.
- b. Construct a 2x2 contingency table summarising the relationship between metastasis within one year and death within one year. Calculate the sample proportions and odds.
- c. Calculate the risk difference, risk ratio, and odds ratio for death within one year for those who experience metastasis and those who do not.
- d. Conduct and interpret a hypothesis test for this outcome.

### C) Advanced learning: Exact tests

C1. Consider the following contingency table:

	Failure	Success
Control	28	2
Intervention	23	7

The `chisq.test()` function applied to this table gave a warning: **Chi-squared approximation may be incorrect**. In this question we will consider *exact* tests as an alternative to the chi-squared test and use simulations to explore the sample size guidance for validity of the chi-squared test.

- Why did the `chisq.test()` produce this warning for the contingency table above? Conduct an exact test for the contingency table using the R function `fisher.test()` and compare your inference to the results of `chisq.test()`.
- Create a function in R to simulate 2x2 contingency tables and conduct hypothesis tests for equal proportions using the chi-squared test, chi-squared test with continuity correction, and the exact test. Your function should do the following steps:
  - Take two arguments: `n` the number of observations per exposure group (row total) in the contingency table, and `p = c(p1, p2)` the event probability in each exposure group.
  - Simulate a 2x2 contingency table based on the input sample size and probability. (*Hint*: There are different ways one could do this in R. One approach is to use the function `rbinom(2, n, p)` to simulate the number of events (first column), then subtract the number of successes from the input row total `n` to calculate the number of failures.)
  - For the simulated contingency table, calculate the chi-squared test, chi-squared test with continuity correction, and the exact test (calculated with `fisher.test()`).
  - Return a vector with five values: (1) the chi-squared test statistic, (2) the p-value for the chi-squared test, (3) the chi-squared with the continuity correction test statistic, (4) the p-value for the chi-squared test with continuity correction, and (5) the p-value for the exact test.
- Simulations under the null hypothesis*. Under the null hypothesis, the row probabilities  $p_1 = p_2 = p$ . Use the function created in (b) to simulate a large number of replicates (5000 to 10,000) for sample sizes per group ranging from  $n = \{5, 10, 25, 50, 100, 250, 500\}$  and success probability  $p = \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ . (*Hint*: Use the `replicate()` function to repeat the simulation function.)
  - For each combination of values  $n$  and  $p$ , based on the rules of thumb for validity of the chi-squared test provided in Kirkwood and Sterne, which test (chi-squared, chi-squared with continuity correction, exact test) would you apply?
  - For some of the simulations, the chi-squared test has returned `NaN` values for the test statistic and p-value. Why does this occur? (*Hint*: Think about the formula for the chi-squared statistic.) What proportion of simulations had `NaN` values for each value of  $n$  and  $p$ ? What did the exact test return for cases where the chi-squared test returned `NaN`?
  - Under the null hypothesis, the chi-squared test statistic for a 2x2 contingency table should follow a  $\chi^2$  distribution with 1 degree of freedom. For each combination of sample size and success probability, plot a histogram or density plot of the simulated distribution of the chi-squared test statistic (without and with the continuity correction) and compare this to the chi-squared density with 1 degree of freedom.
  - For each combination of sample size and success probability, in what proportion of cases was the p-value for each test less than  $\alpha = 0.05$ , resulting in *incorrectly rejecting* the null hypothesis? This is the *Type I error rate*.

- d. *Simulations under the alternative hypothesis.* Now, consider the case where there is a true difference between the event probabilities for each exposure group. Use your function from (b) to simulate a large number of replicates assuming a risk difference of 0.1, that is  $p_1 = p_2 + 0.1$  for sample sizes  $n = \{5, 10, 25, 50, 100, 250, 500\}$  and true probability  $p_2 = \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ .

For each combination of sample size and success probability, calculate the proportion of times in which each test (chi-squared, chi-squared with continuity correction, and exact test) *correctly* rejected the null hypothesis at the  $\alpha = 0.5$  level. This proportion is referred to as the *power* of the test. The proportion of times the test incorrectly failed to reject the null hypothesis is the *Type II error rate*, often denoted  $\beta$ ; and the *power* is  $1 - \beta$ . Optionally, you may also wish to explore how the power changes as the true risk difference ( $p_1 - p_2$ ) increases.