BMR 617

Hypothesis tests for proportions

Test for a single proportion

- Example:
- In a study of Neonatal Abstinence Syndrome, DNA samples were taken from 34 pregnant women in West Virginia with substance use disorder and sequenced.
  - For the variant rs6972158 in the gene NPSR1, 27 of the 68 alleles were found to be the variant allele G (41 were the reference allele A).
  - According to the 1000 Genomes Project, the G allele frequency in the American population is 0.235.
- We want to know if the variant allele frequency in this population is different to that of the general American population

Formulating the question as null and alternative hypotheses

- The null hypothesis is the hypothesis we would tend to assume with no other data
- The alternative hypothesis is typically the hypothesis we would like to "prove" (or provide evidence to support)
- The null hypothesis here is:
  - The G allele frequency in the study population is 0.235
- The alternative hypothesis is:
  - The G allele frequency in the study population is not 0.235

R code for testing a value of a single proportion

- In R, we can test the null hypothesis that a proportion is equal to some given value using the `prop.test()` function
- To see how to use the function, type `?prop.test` in the console, or search for `prop.test` in the help tab.
- In our case, we have 27 G alleles out of a total of 68 alleles, and the probability of a G allele under the null hypothesis is 0.235:

```
prop.test(27,68,0.235)
```

Output

```
      1-sample proportions test with continuity correction


data:  27 out of 68, null probability 0.235
X-squared = 9.053, df = 1, p-value = 0.002623
alternative hypothesis: true p is not equal to 0.235
95 percent confidence interval:
 0.2826780 0.5231249
sample estimates:
        p
0.3970588
```

Interpreting the output

- The sample estimate (i.e. the estimate of the probability of a G allele from the sample) is 0.397
  - You can check that this is just the number of G alleles in the sample divided by the total number of alleles: 27/68
- The 95% confidence interval for the proportion is [0.283, 0.523]
  - We are 95% confident that the interval[0.283, 0.523] contains the true value of the proportion of G alleles for West Virginian pregnant women with substance use disorder
- The p-value is 0.002623. This means that if the null hypothesis were true, there would be a 0.2623% chance of seeing data "this extreme"
  - i.e. if the proportion of G alleles among West Virginia pregnant women with SUD were 0.235, there would be a 0.2623% chance of seeing data this different to that in a study of 34 such women

Interpreting the result

- Since the p-value of 0.002623 is less than our predetermined threshold of 0.05, we would reject the null hypothesis and conclude that the proportion of G alleles in this population is different to 0.235
  - Since the entire 95% confidence interval is above 0.235, we'd conclude it's more than 0.235
- Possible explanations:
  - The allele frequency in West Virginians is greater than that of the general US population
  - The allele frequency among women likely to become pregnant is greater than that of the general US population
  - The allele frequency among those with SUD is greater than that of the general US population
  - Some combination of the above

Comparing proportions

- Remember the Pfizer COVID-19 vaccine trial:

		Treatment		
		Placebo	Vaccine	Total
SARS-CoV-2 status	Infected	162	8	170
	Not infected	18163	18190	36353
	Total	18325	18198	36523

- What we want to know here is if the proportion infected (the risk) in the vaccine group is different to the proportion infected (the risk) in the placebo group
  - We want to compare two proportions

Comparing proportions formulated as a hypothesis test

- The null hypothesis is that the risk in the vaccine group is the same as the risk in the placebo group
- The alternative hypothesis is that the risk is different in each group
- To calculate a p-value, we assume the risk is the same in both groups, and calculate the probability that we would get data as different as we got in the actual trial
- There are two distinct approaches
  - Fisher's exact test
  - The Chi-squared test

Fisher's exact test

- The trial has 18325 and 18198 participants in the placebo and vaccine groups, respectively
- 170 participants were infected and 36353 were not infected
- Fisher's exact test, at least conceptually, examines all possible contingency tables with these row and column totals, and calculates the probability of each one
- It then sums the probabilities of the tables in which the proportions between the groups are at least as big as the one observed

Chi-squared test

- The Chi-squared test works by calculating the "expected" value of each cell in the table if the proportions were equal in each group
- Overall, 170 out of 36523, or 0.004655 of the participants became infected
- There were 18325 participants in the placebo group, so assuming the null hypothesis, we'd expect, on average, 0.004655 x 18325 = 85.3 participants in the placebo group to become infected, and 18239.7 not to become infected
- Similarly, there were 18198 participants in the vaccine group, so we'd expect 0.004655 x 18198 = 84.7 to become infected, and 18113.3 not to become infected

Chi-squared test: observed and expected tables

Observed data	
162	8
18163	18190

Expected data	
85.3	84.7
18239.7	18113.3

How the Chi-squared test works

- In the Chi-squared test, for each cell in the table we calculate $\frac{(O-E)^2}{E}$, where O is the observed value and E is the expected value
- This is summed over all cells
- This statistic, the χ^2 statistic, is *approximately* distributed according to a distribution called the χ^2-distribution with n degrees of freedom
  - The number of degrees of freedom is (r-1)(c-1) where r and c are the number of rows and columns in the table; here there is 1 degree of freedom

Pros and Cons of the Chi-squared test and Fisher's exact test

- Fisher's exact test calculates the *exact* probability of getting results at least as extreme as those observed in the data, assuming the null hypothesis is true
- However, it is computationally intensive
  - For very large samples sizes, especially if there are more than two rows or columns, it can be prohibitive
  - Before the advent of computers, the calculation was frequently prohibitive
- The chi-squared test is an approximate test
  - The approximation starts to fail if the expected value in any cell is below 5
    - Below 10 for 2x2 tables
- However it is not computationally intensive
- Use Fisher's exact test if the computer can handle it; use the chi-squared test otherwise

Fisher's exact test in R

- The `fisher.test(…)` function will run Fisher's exact test
  - Requires a matrix

```
pfizer.data <- matrix(c(162, 18163, 8, 18190), nrow=2)
fisher.test(pfizer.data)
```

- Output:

```
        Fisher's Exact Test for Count Data

data:  pfizer.data
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 10.04364 47.79924
sample estimates:
odds ratio
  20.28092
```

Odds and the odds ratio

- Note the Fisher's exact test generates an odds ratio
- The *odds* are the number infected divided by the *number not infected*
  - Not divided by the total!
  - For "rare" events, such as infection in this case, the odds are close to the risk
- The odds for the placebo group are 162/18163, and the odds for the vaccine group are 8/18190, so the odds ratio is (162/18163)/(8/18190) = 20.28, as reported
- The 95% confidence interval for the odds ratio is [10.04, 47.8]
  - We are 95% confident the interval [10.04, 47.8] contains the true odds ratio
- Note that if the proportions in both groups are the same, the odds are the same, and so the odds ratio would be 1.

The p-value

- The p-value for this test is reported as $p < 2.2 \times 10^{-16}$
- The value 2.2×10^{-16} is essentially the smallest value the computer can represent easily
  - So this is really saying the computer can't distinguish between the p-value and zero
- Remember the p-value is always associated with a null hypothesis
  - The null hypothesis for the Fisher's exact test is that the odds ratio is 1
- If the chances of infection were the same for the vaccine group and the placebo group, the probability we would see data this extreme in the clinical trial is less than 2.2×10^{-16}, i.e. it is essentially zero.

Chi-squared test in R

- To run the Chi-squared test in R, we can use the `chisq.test()` function:

```
chisq.test(pfizer.data)
```

- Output:

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  pfizer.data
X-squared = 137.28, df = 1, p-value < 2.2e-16
```

Output from chi-squared test

- The chi-squared test also gave us a p-value of 2.2×10^{-16}.
- Remember, this is interpreted in the context of a null hypothesis
- The null hypothesis for the chi-squared test is that the response variable (infection) is independent of the explanatory variable (treatment)
  - There is no estimate of an odds ratio

Effect size

- Fisher's exact test gave an estimate of the odds ratio, along with a 95% confidence interval
- This is useful information
- The p-value merely tells us how likely these data would be if there were no different between the placebo and vaccinated groups
- But we want to know *how different* they are
- The odds ratio tells us this
- It is a measure of *effect size*: how much effect does the treatment have

Effect size when you cannot use Fisher's Exact Test

- In the (rare) cases when you cannot use Fisher's Exact Test, the Chi-Squared test gives a p-value but no effect size
- In this case we can also use the `prop.test()` with two proportions
- It needs the number infected and the *total*:

```
prop.test(x=c(162, 8), n=c(18325, 18198))
```

```
2-sample test for equality of proportions with continuity correction

data:  c(162, 8) out of c(18325, 18198)
X-squared = 137.28, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.006956920 0.009844627
sample estimates:
      prop 1       prop 2
0.0088403820 0.0004396087
```

Interpreting the output from `prop.test()`

- Note the output from prop.test() also gives a χ^2 statistic, and a p-value
  - These have exactly the same interpretation as the chi-squared test
- It also gives the proportion in each group
  - This is the number infected divided by the *total* in that group
  - I.e. the *risk*
- The difference in the risk is the *attributable risk*: 0.00884 - 0.00044 = 0.00840
- The confidence interval is the 95% confidence interval for the *attributable risk*: we are 95% confident the range [0.00696, 0.00984] contains the true attributable risk

Note on odds ratio and FDA requirements for EUA

- As a side note, the Federal Drug Administration (FDA) requires that the vaccine is at least 50% effective in order to grant emergency use authorization (EUA)
- To be 50% effective, the odds in the vaccinated group must be 50% or less than the odds in the placebo group
- So the odds ratio (placebo/vaccinated) must be greater than 2
- The 95% confidence interval for the odds ratio was [10.04, 47.8]
- So we are 95% confident the vaccine is at least 90% effective

Summary

- To perform a hypothesis test to see if a single proportion is different to some fixed value, use `prop.test(…)` specifying the value, total, and the assumed fixed proportion
- To compare two proportions, favor using Fisher's Exact Test (`fisher.test(…)` in R)
- If `fisher.test(…)` cannot run, use `chisq.test(…)`
  - In this case you can also use `prop.test(…)` with two proportions, which generates more information
- Be careful! `fisher.test(…)` and `chisq.test(…)` expect a matrix with the contingency table; `prop.test(…)` expects the value(s) and *total(s)*.