

STUVAC Statistical Snacks 2

MATH1062/MATH1005: Mathematics 1B/Statistical Thinking With Data Semester 1, 2024

Lecturers: J. Baine, T. Cui, J. Spreer, M. Stewart

Tests for proportions A market research company conducts regular opinion polls to get some idea of support for different political parties over time. On 13 May it released the results of a poll based on “a representative cross-section of 1,654 Australian electors from May 6-12, 2024” (see original article). It found that on a “two-party preferred” basis, 52% of those surveyed indicated support for Labor, while the remaining 48% supported the Liberal/Nationals coalition.

1. Describe a box model we could use for making an inference about the “true” proportion p of electors in Australia who would support Labor on a two-party preferred basis (that is, if asked to choose between Labor and the Liberal/National coalition) based on this data. Comment on the suitability of your model for this example.
2. Estimate the standard error of the sample proportion. If necessary, round to 3 decimal places.
3. Is the value $p = 0.50$ included in a 95% Wilson confidence interval for p based on this data? The R output below may be helpful:

```
qnorm(0.95)

## [1] 1.644854

qnorm(0.975)

## [1] 1.959964
```

4. Would the test of $H_0: p = 0.5$ versus the alternative $H_1: p \neq 0.5$ be rejected at the 5% level of significance? What about at the 10% level of significance?
5. The company produces such poll results every week. For the subsequent two weeks it produces the following results

Date	Sample Size n	Labor	Coalition
May 20	1674	50.5%	49.5%
May 27	1488	48.5%	51.5%

These results may be combined with the earlier poll and summarised in the frequency table below (percentages have been converted to counts):

Date	Labor	Coalition	Total
May 13	869	785	1654
May 20	845	829	1674
May 27	722	766	1488
Total	2436	2380	4816

It is desired to perform a statistical test to assess whether this data indicates a change in “true, underlying voting preferences” in the whole population over the 3-week period.

- This may be done using a certain type of chi-squared goodness-of-fit test. What type exactly? Explain by describing the statistical model used to model these observations.
- Are the 3 different sample proportions *significantly* different at the 5% level? What about at the 10% level? Explain with reference to the R output below.

```
0.13 = c(869, 785)
0.20 = c(845, 829)
0.27 = c(722, 766)

combined = 0.13 + 0.20 + 0.27
combined
## [1] 2436 2380

sum(combined)
## [1] 4816

combined/sum(combined)
## [1] 0.505814 0.494186

E.13 = sum(0.13)*combined/sum(combined)
E.13
## [1] 836.6163 817.3837

sum(((0.13-E.13)^2)/E.13)
## [1] 2.536511

E.20 = sum(0.20)*combined/sum(combined)
E.20
## [1] 846.7326 827.2674

sum(((0.20-E.20)^2)/E.20)
## [1] 0.007173629

E.27 = sum(0.27)*combined/sum(combined)
E.27
## [1] 752.6512 735.3488

sum(((0.27-E.27)^2)/E.27)
## [1] 2.525862

qchisq(c(0.9, 0.95), df=1)
## [1] 2.705543 3.841459
```

```
qchisq(c(0.9, 0.95), df=2)
## [1] 4.605170 5.991465
qchisq(c(0.9, 0.95), df=3)
## [1] 6.251389 7.814728
qchisq(c(0.9, 0.95), df=4)
## [1] 7.779440 9.487729
```