STAT5002 Lab7 Question Sheet

Introduction to Statistics

STAT5002

Between 12 September and 7 November 2017 the Australian Government conducted a national postal survey in order to gauge the Australian community's opinion on changing legislation about same-sex marriage.

In the lead-up to the postal survey, various organisations conducted their own polls to infer how the final survey might turn out. One such poll returned the following results:

In favour of proposed change	Against proposed change	Don't know
1036	581	200

What can we infer from this about the final results of the official postal survey?

1 Simplifying assumptions

Since the final postal survey was to be voluntary, and we ultimately wish to estimate the proportion of participants who favour the proposed change, it is reasonable to

1. discard the "Don't know" counts;
2. model the remaining responses as being like a random sample taken with replacement from a box containing only 0s and 1s, with proportion of 1s equal to p (for some unknown $0 < p < 1$).

The parameter p is then the unknown "population" proportion of voters in favour of the change. If $p > 0.5$, the proposed change to the law will be made by the Government.

Question: Do these assumptions seem reasonable? Comment.

2 Sample size and observed sample proportion

- **What is the sample size n?**
- **What is the observed sample proportion $\bar{x}$?**

3 Confidence interval

You will need to use R for this task – please see the R markdown worksheet.

- **Determine a Wilson's 99% confidence interval for p based on the observations in the previous part. Hint:** use `binom.confint()`.
- **Perform a "sanity check" that the endpoints of the interval are such that the observed value $\bar{x}$ is "right on the edge" of a corresponding 99% prediction interval.**

4 Hypothesis test

The observed proportion is greater than 0.5, but is it significantly greater? Answer this question with an appropriate formal hypothesis test by taking the following steps.

As an exercise, you should carry out the calculation as much as possible by hand. You will need to use R for calculating the P-value.

- **Assumptions**: state the neccessary assumptions
- **Null hypothesis and alternative hypothesis**: what are the null and alternative hypotheses? Is this a one-sided or two-sided test?
- **Test statistic**: Calculate the test statistic (z-statistic). What values of z-statistic argue against the null hypothesis?
- **P-value**: What is the P-value?
- **Conclusion** Draw your conclusion based on the P-value. You can use either the P-value itself or a significance level of 0.05.

5 Results

The final outcome of the postal survey (i.e. the true population parameter p) was 61.62% support.

In light of the results of your calculations above, what can you say about the Essential poll?