Eco311 Optional Reading: Hypothesis Testing

(Jing Li, Miami University)

1. This note provides one example of hypothesis testing, a key tool for *statistical thinking*.

2. President Trump is unhappy with the election result, and he suspects someone may cheat in the process of voting. We wonder if statistics can be used to resolve this issue.

3. Focus on Georgia. One obvious solution is recounting all ballots in that state. But this *population* may be contaminated—ballots might be fabricated, or stolen, or be cast by persons who are not eligible for voting. Statistics can help!

4. Using statistics entails collecting a sample. It is a bad idea to use a non-iid or biased sample such as doing a survey in downtown Atlanta. Ideally, all Georgian voters should have an *equal chance* to be selected. For instance, an iid or *random sample* may include all voters with 7 as last digit in their social security number. This sample is more *representative* than people walking at downtown Atlanta since Trump or Biden supporters have equal chance to have 7 as the last digit of SSN.

5. Suppose the size of this random sample is $n = 100$, and 45 of those people recalled that they voted for Trump. Does that indicate Trump had lost in Georgia? Not necessarily. The main concern is that we may use *different* random sample (e.g., people with 9 as last digit of SSN) and get different result. The key issue of statistics is accounting for the *uncertainty* associated with using samples. We use *standard error* as the measurement of sampling uncertainty.

6. Let's do math. The sample average or sample proportion of Trump votes is

$$\bar{y} = \frac{45}{100} = 0.45$$

The standard deviation is (see Q2 of HW1)

$$\sigma = \sqrt{variance} = \sqrt{0.45(1 - 0.45)} = 0.4974937$$

The standard error is

$$\frac{\sigma}{\sqrt{n}} = \frac{0.4974937}{\sqrt{100}} = 0.04974937$$

Now we know the left eye and mouth of that panda (t statistic).

7. The right eye of panda is determined by the null hypothesis

$$H_0 : \mu = 0.5 \quad (\texttt{there is a tie in election})$$

The t statistic equals

$$\texttt{one sample t test} = \texttt{panda} = \frac{0.45 - 0.5}{0.04974937} = -1.005038$$

We *cannot* reject the null hypothesis since $1.005038 < 1.96$. The R codes and results are

```
> y = c(rep(1,45),rep(0,55))
> t.test(y, mu = 0.5, alternative = "two.sided")

One Sample t-test
data:  y
t = -1, df = 99, p-value = 0.3197
alternative hypothesis: true mean is not equal to 0.5
95 percent confidence interval:
0.3507892 0.5492108
sample estimates:
mean of x
0.45
```

We see the p-value 0.3197 exceeds 0.05, and the 95 confidence intervals $(0.3507892, 0.5492108)$ include the hypothesized value 0.5. Both imply no rejection.

8. So based on the result from this relatively *small* imaginary sample, our conclusion is that there is no statistical evidence strong enough to suggest that Trump had lost in Georgia. In other words, we cannot rule out that there was a tie in the election.

9. What if the sample gets larger, and have 1000 voters? Suppose 450 of them voted for Trump. The sample average is still 0.45, but now standard error becomes smaller, implying less uncertainty

$$\frac{\sigma}{\sqrt{n}} = \frac{0.4974937}{\sqrt{1000}} = 0.01573213$$

The new t test becomes
$$\frac{0.45 - 0.5}{0.01573213} = -3.178209.$$
Because 3.178209 is greater than 1.96, we *reject* the hypothesis that there is a tie. Since $0.45 < 0.5$, we conclude from this imaginary sample that Trump had lost in Georgia if the *significance level* (denoted by α) is 0.05.

10. Lesson 1: big sample is *more informative* than small sample.

11. Recall that there are two inherent errors associated with hypothesis testing. We may reject a correct null hypothesis (type I error), or we may fail to reject an incorrect null hypothesis (type II error). Those two errors are like a judge sending an innocent person to jail, or a judge setting a criminal free. By using 1.96 as the critical value, we let the probability of type I error be 0.05. People may not like this, and they may prefer a direct measurement of the *strength of the evidence* against the null hypothesis. P value is such a measurement. In this case, the p-value is

$$\texttt{p value} = \texttt{two shaded tails} = 2\text{pnorm}(-3.178209) = 0.001481879.$$

This p value is very close to zero. So here we can largely ignore the type I error. In other words, the evidence that suggests the victory of Biden is *very strong*. The new R results are

```
> ybig = c(rep(1,450),rep(0,550))
> t.test(ybig, mu = 0.5, alternative = "two.sided")

One Sample t-test
data:  ybig
t = -3.1766, df = 999, p-value = 0.001536
alternative hypothesis: true mean is not equal to 0.5
95 percent confidence interval:
0.4191127 0.4808873
sample estimates:
mean of x
0.45
```

12. Lesson 2: The p-value approach is *better* than the critical value approach