TELFER

VOTRE LIEN AVEC CE QUI COMPTE — CONNECTS YOU TO WHAT MATTERS

Hypothesis Testing: Coincidental vs Real Change

ADM 2304
Winter 2016

© Tony Quon

Question: Is support for the Liberals higher than it was during the last election?

There are two points of view:

Null Hypothesis (H_o):
It is not higher;

Alternative Hypothesis (H_a):
It is higher.

Hypotheses using the population proportion p who support Liberals

Null Hypothesis (H_o):

$p \leq 0.3947$ (or $p = 0.3947$)

Alternative Hypothesis (H_a):

$p > 0.3947$.

The hypothesized value $p_o = 0.3947$ was the proportion who voted for the Liberals in the 2015 federal election. It is **not** a sample result.

Characterizations of Hypotheses

- The Null Hypothesis represents the **status quo**; it is “dull” and “uninteresting”;
- The Alternative Hypothesis represents the “interesting” or “**research**” hypothesis;
- As the proponent, the **burden of proof** rests on you to show that the evidence supports the alternative hypothesis and not the null hypothesis.

Start with the null hypothesis...

- We take the “status quo” null hypothesis as a **starting** assumption;
- To convince someone who is skeptical of new ideas and who prefers the status quo, we need to obtain empirical evidence which is so **“astonishing”** or **“startling”**, and clearly **inconsistent** with the **status quo.**
- What constitutes inconsistency and how do we measures the degree of inconsistency?

Do a survey

- Take a random sample of n voters;
- Calculate the sample proportion **p-hat**.
- Suppose that the most recent results were based on a sample size of 1500.

What would we expect?

- If H_o were true, then p-hat would have a sampling distribution which is *normally* distributed, with mean $p_0 = .3947$ and stdev (SD) = $\sqrt{(p_0 q_0 / n)} = 0.0126$ (for n=1500);
- Consider the (standardized) test statistic

  $$Z_{stat} = (\text{p-hat} - 0.3947) / 0.0126$$

- This measures the distance from $p_0 = .3947$, but in standard deviation (SD) units.
- We would expect Z "close" to zero.

How “close”?

- $|z| < 1$ would be “close”, but not $z > 3$;
- Is there a “critical value” above which we should “reject” the null hypothesis?
- What evidence should be considered “inconsistent” with the null hypothesis?

Critical Value

- Suppose we found a **critical value** z_α such that $P(z_{stat} > z_\alpha) = \alpha$, where α is some small probability called the "significance level".

Type I and II Errors

- Suppose we reject H_o if $z_{stat} > z_\alpha$ (this is the "rejection region")
- We would make the wrong decision with only small probability α if H_o were true (Type I Error). On the other hand, this decision rule means we would not reject H_o if $z_{stat} \leq z_\alpha$ (the non-rejection region). But if H_o were false, then we would be making a different wrong decision (Type II Error).

		Reality	
		H_o true	H_o false
Decision	Do not reject H_o	Correct decision	Type II error
	Reject H_o	Type I error	Correct decision

The conservative position

- A conservative position would worry more about a Type I error than a Type II error--better to miss out possibly on a new discovery than to tell the world now and be embarrassed later.

Analogy with jury trial

- Suppose a person is on trial for murder. We presume he is innocent until proven guilty (beyond "reasonable" shadow of doubt);
- Here the null hypothesis is one of innocence, the alternative is one of guilt;
- There is a tradeoff between the Type I error (convicting an innocent person) and the Type II error (acquitting a guilty person).

Analogy (cont'd)

- Beyond a "reasonable" shadow of doubt is quantified in the statistical test as
  Prob(Type I error) $\leq \alpha$.
- In hypothesis testing, we commonly use
  α =.05 or .01; that is,
  Pr(Reject H_o, given H_o true) $\leq$ .05 or .01.

Definition of Critical Value

- Given alpha, we need to find a critical value z_α such that $P(Z_{stat} > z_\alpha) = \alpha$.
- If α =.01, then $z_\alpha = 2.326$.
- At the .01 significance level, we reject H_o if

$$z_{stat} > 2.326, \text{ where}$$

$$z_{stat}=(\text{p-hat} - .3947) / \sqrt{(.3947*.6053/1500)}$$

(or if p-hat > .3947 + 2.326 * .0126 = .424).

One-sided confidence interval

- In testing the hypotheses, we sincerely believe the alternative ($p > 0.3947$) a priori ; therefore, the open-ended side of the interval should be in the "greater than" direction.
- The 99% one-sided confidence interval for estimating p is:

  "greater than p-hat – 2.326 √(p-hat *q-hat/n) "
- If this interval does not contain p_0= .3947, then the evidence is "inconsistent" with the null hypothesis.

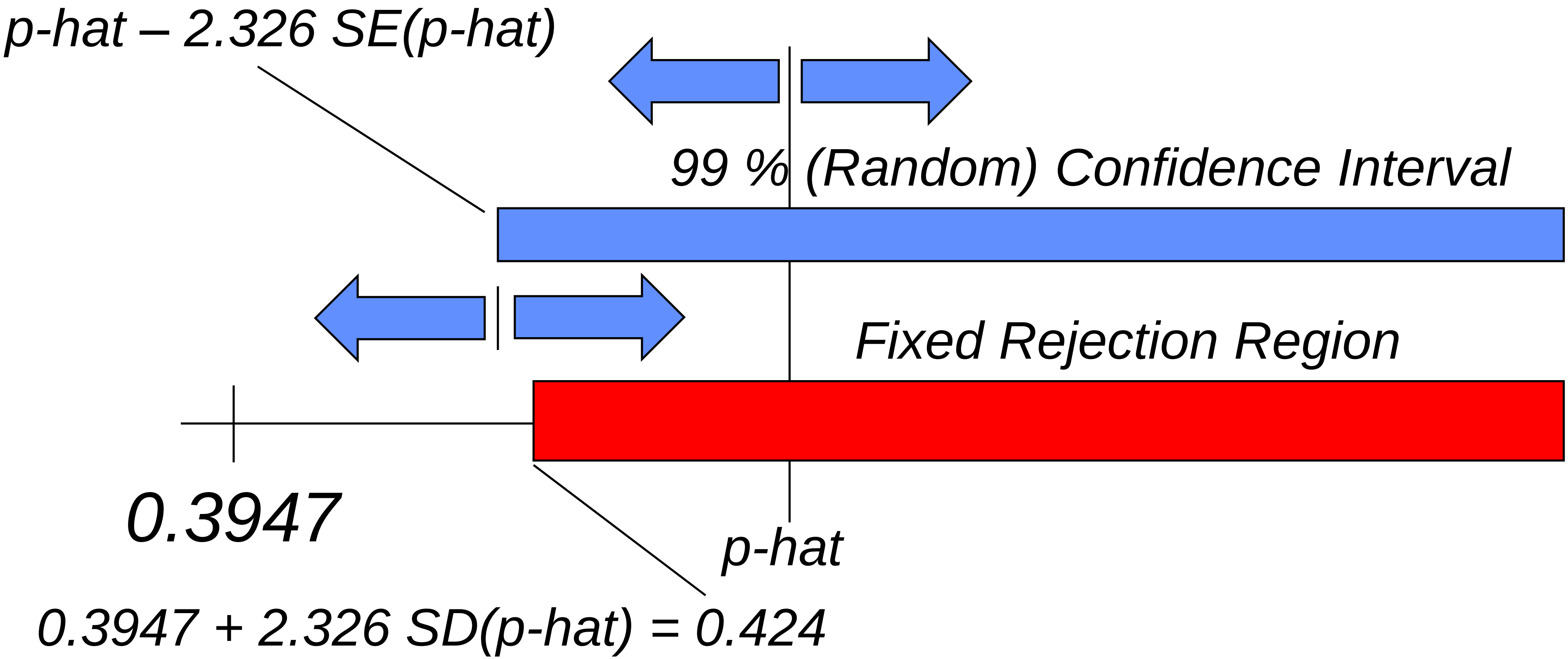
Let's look at the data

- The recent Abacus poll found p-hat=.45, based on a sample size of 1500 (http://www.threehundredeight.com)
- We reject H_o if $z_{stat} > 2.326$.
- Since $z_{stat} = (.45 - .3947) / \sqrt{(.3947*(.6053)/1500)} = 4.38$, we reject H_o at the .01 level of significance.
- The 95% 1-sided confidence interval for estimating the pop'n proportion is:

  "greater than $.45 - 2.326 \sqrt{(.45*(.55)/1500)}$" or [0.42, 1.00].

  Since the interval does not cover the value $p_0 = .3947$, we reject H_o.
- Conclude that there is sufficient evidence to show that Liberal support has increased since the federal election.

Schematic Picture



P-hat falls inside the rejection region if and only if the 1-sided interval excludes 0.3947.

Relationship between Confidence Interval and Hypothesis Test

In general, we reject $H_0 : p \leq p_0$ in favour of $H_a : p > p_0$ at the significance level α

if and only if

the one-sided confidence interval "greater than p-hat - z_α*SE(p-hat)"

does not cover the hypothesized value p_0 .

Meaning of Test

- In the population, either
  $p \leq 0.3947$ or $p > 0.3947$.
- We collect data (n=1500) and observe a p-hat value of 0.45.
- Clearly p-hat > 0.3947, but what about the value of the population proportion?

Coincidental or real difference?

- Claim p-hat = 0.45 is "**consistent**" with the population proportion being ≤ 0.3947 and the observed p-hat is higher simply due to **sampling variability** (Hypothesis of a **coincidental** difference), OR

- Claim p-hat = 0.45 is **inconsistent** with the pop'n proportion being ≤ 0.3947 and the observed p-hat shows that the value of p is **really** > 0.3947 (Hypothesis of a **real** difference). Here, we say that p-hat = 0.45 is "statistically significantly" greater than .3947 at the 0.01 level.

Practical vs Statistical Significance

- Since Z-stat = (p-hat - p_0)/$\sqrt{}$(p_0 q_0/n), a large enough sample size n can detect "statistically significant" differences, even though the observed difference may not be of any practical significance or importance.
- To avoid this confusion, we should report statistically significant differences only when they are judged to be of practical importance.

Type II Errors

- We design our tests around the Type I Error, but we cannot forget β = P(Type II Error).
- Power = 1 - P(Type II Error)
  = P(reject H_0, given H_0 is false)
- We want to have high power for values of $p > p_0$, where the effect size $(p - p_0)$ is of practical importance.
- Generally increasing the sample size will increase the power.

Probability-value

- Comparing a test statistic to a critical value results in a reject/not reject decision. Then we tend to forget the strength of the evidence against the null hypothesis.
- The confidence interval is very helpful by estimating p, along with a margin of error, for a direct comparison with p_0 .
- The p-value approach uses the test statistic to measure the strength of the evidence against the null hypothesis.

Probability-Value

- This is the probability that a *new* sample would produce a result just as extreme or even more extreme in the direction of H_a, assuming H_0 is true;
- For z_{stat} = (p-hat - .3947)/.0126 = 4.38,

  the p-value is $P(Z \geq 4.38) = 0.00001$;
- The p-value measures the strength of the evidence against the null hypothesis, with smaller p-values indicating stronger evidence for the alternative;
- We reject the null hypothesis if the p-value is less than the significance level α.

Three Ways to Test

- The critical value approach calculates
  z_{stat} = (p-hat - p_0)/SD(p-hat),
  and rejects H_0 if $z_{stat} > z_\alpha$;
- The p-value approach rejects H_0
  if the p-value = $P(Z > z_{stat}) < \alpha$;
- The confidence interval approach rejects H_0 if the interval [p-hat - z_α* SE(p-hat),1] estimating p does not cover p_0 .

Using Minitab*

- From the "Stat" Menu, select "Basic Statistics", then "1 proportion".
- Summarize the data as "number of trials" = 1500, and "number of events" = 675, since n = 1500 and X = 0.45*1500 = 675.
- Under "Options", specify the value of p_0 = .3947, the "greater than" alternative, and the use of the normal approximation if appropriate.
- We obtain the following output:

```
Test and CI for One Proportion

Test of p = 0.3947 vs p > 0.3947

                                              99%
                                            Lower
Sample      X      N  Sample p      Bound  Z-Value  P-Value
1         675   1500  0.450000   0.420118     4.38    0.000
```

* The Data Analysis Tool in MS Excel does not automate these calculations, but you can use the spreadsheet to program the basic formulae.