PLSC 502 – Autumn 2016
Hypothesis Testing

October 18, 2016

Hypothesis Testing: Concepts

- A *null hypothesis*, H_0
- an *alternative hypothesis* H_a
- a *test statistic* $\theta = f(\mathbf{X})$
- a *rejection region* in the range of θ

Example

Quinnipiac poll:

- Clinton $=$ 47 percent
- Trump $=$ 41 percent
- $N = 660$ likely voters.

Hypothesis:

$$H_a : \pi > 0.5$$

Corresponding null:

$$H_0 : \pi = 0.5$$

Test statistic: $\hat{\pi} = 0.47$

- **Type I error** = "false positive."
- **Type II error** = "false negative."

	Reailty / Population	
Test Statistic / Sample	H_a	H_0
H_a	Correct	Type I error
H_0	Type II Error	Correct

Alphas and Significance

By convention:

$$\text{Pr(Type I Error)} = \alpha \text{ (“significance level”)}$$

and

$$1 - \alpha = \text{ “specificity”}$$

While

$$\text{Pr(Type II Error)} = \beta$$

and

$$1 - \beta = \text{ “sensitivity.”}$$

A New Table

	Reailty / Population		
Sample Result	Positive	Negative	Frequency
Positive	True Positive (N_{TP})	Type I error (False Positive) (N_{FP})	$N_P = N_{TP} + N_{FP}$
Negative	Type II Error (False Negative) (N_{FN})	True Negative (N_{TN})	$N_N = N_{TN} + N_{FN}$
Frequency	$N_{(+)} = N_{TP} + N_{FN}$	$N_{(-)} = N_{TN} + N_{FP}$	N

Components...

- *False positive / significance rate* $\alpha = N_{FP}/N_{(-)}$,
- *False negative rate* $\beta = N_{FN}/N_{(+)}$,
- *False discovery rate* $= N_{FP}/N_P$,
- *False omission rate* $= N_{FN}/N_N$,
- *Accuracy* $= (N_{TP} + N_{FP})/N$

In the Clinton / PA example, we know:

$$\hat{\pi} \sim \mathcal{N}(\pi, \sigma^2_{\hat{\pi}}).$$

and

$$\begin{aligned} \hat{\sigma}^2 &= 0.470(1 - 0.470) \\ &= 0.249 \end{aligned}$$

and

$$\begin{aligned} \hat{\sigma}^2_{\hat{\pi}} &= \frac{0.249}{660} \\ &= 0.00038. \end{aligned}$$

That means:

$$\hat{\pi} \sim \mathcal{N}(0.5, 0.00038).$$

More Hypothesis Testing

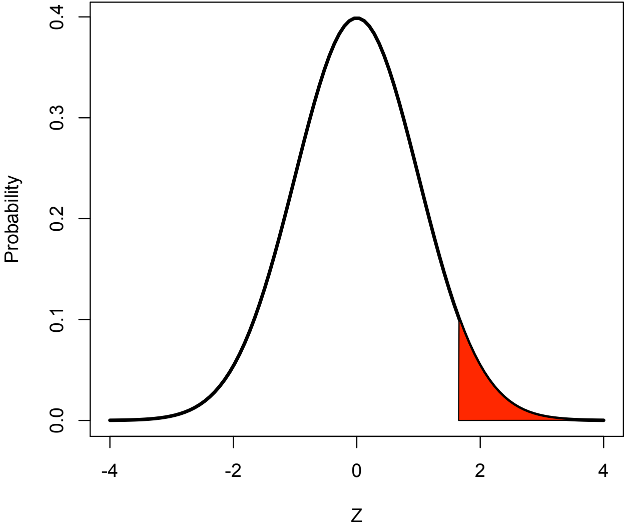
Converting $\hat{\pi}$ to z:

$$\frac{\hat{\pi} - \pi}{\sigma_{\hat{\pi}}} = Z \sim \mathcal{N}(0, 1).$$

Decision rule:

$$\text{Reject } H_0 \text{ if } Z \geq z_\alpha.$$

$$\alpha = 0.05 \rightarrow Z \geq 1.65$$



Example, Continued

Here,

$$Z = \frac{0.470 - 0.50}{0.0195} = -1.54$$

so we fail to reject the null.

Another alternative $H_0 : \pi = 0.40$ yields

$$\begin{aligned} Z &= \frac{0.47 - 0.40}{0.0195} \\ &= 3.59 \end{aligned}$$

meaning we would reject that null.

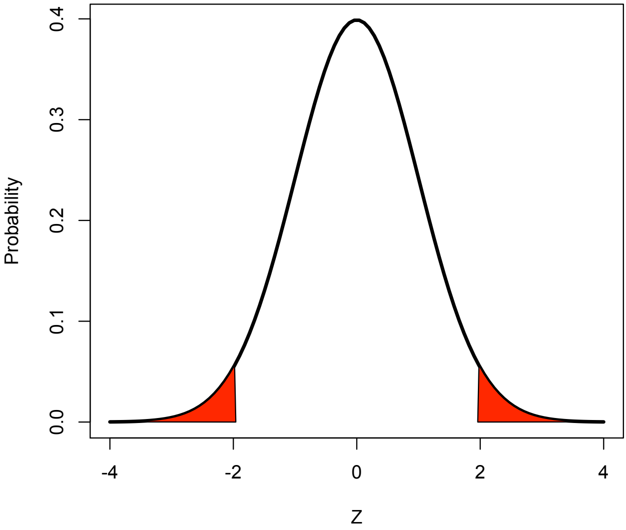
An alternative formulation:

$$H_a : \pi \neq 0.5.$$

Alternative decision rule:

$$\text{Reject } H_0 \text{ if } |Z| \geq z_{\alpha/2}$$

$$\alpha = 0.05 \rightarrow Z \geq 1.96$$



P-Values Versus Significance Tests

P-value (or "attained significance level"): *The smallest level of significance* α *for which the observed data indicate that the null hypothesis should be rejected.*

Why we like them:

- **Avoid arbitrary "cutoffs."**
- **Provide more information.**

Significance Tests and Confidence Intervals

C.I.:

$$\text{c.i.}_\alpha = \hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

vs. test:

$$|Z| \equiv \left|\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}\right| \geq z_{\alpha/2}.$$

Significance Tests and Confidence Intervals

"Acceptance region":

$$-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2} = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}},$$

I.e., "Do not reject H_0 at $\mathrm{P} = \alpha$ if θ lies within a $(1 - \alpha) \times 100$-percent confidence interval around $\hat{\theta}$, and reject H_0 if it does not."

Important Things, I

1. P-**values are not "the probability that the null hypothesis is false."**

> *"The test statistic allows us to reject the null hypothesis at $P < 0.01$, indicating that there is a less than one in 100 chance that the null hypothesis is true."*

> *"The test statistic allows us to reject the null hypothesis at $P < 0.01$, which is strong evidence that the observed result is not due to chance."*

Important Things, II

2. **One does not, in general, "accept" the null hypothesis.**

> *"The P-value for the regression coefficient on Female is 0.56, indicating that there is no relationship between gender and support for immigrants' rights."*

> *"The P-value for the regression coefficient on Female is 0.56, indicating the data do not support the hypothesized relationship between gender and support for immigrants' rights."*

Important Things, III

3. *P*-**values are not the long-run frequency of a "statistically significant" test statistic.**

"The P-value of 0.01 means that 99 out of 100 hypothetical replications would reject the null hypothesis."

Important Things, IV

4. **Statistical significance does not equate to substantive significance.**

```
> popmean <- with(data, prop.table(table(1-Active)))[1]
> popmean
     0
0.8956

> with(data[data$Sign=="Scorpio",],
+      prop.test(table(1-Active),p=popmean,
+                correct=FALSE))

1-sample proportions test without continuity correction

data:  table(1 - Active), null probability popmean
X-squared = 2, df = 1, p-value = 0.2
alternative hypothesis: true p is not equal to 0.8956
95 percent confidence interval:
 0.8832 0.8975
sample estimates:
     p
0.8905
```

Important Things, V

5. **A statistic can never be "significant in the wrong direction."**

> *"Our estimate of the effect of trade liberalization on the probability of a civil war – which we expected to be negative – is in fact positive, and statistically significant at* $P = 0.02$*."*

Important Things, VI and VII

6. **Identical** P**-values are not "better" or "more reliable" if they are based on a larger sample.**

7. **Failing to reject the null hypothesis in a larger sample is a bigger deal than failing to do so in a small one.**