ECO220Y

Hypothesis Testing: Type I and Type II Errors and Power

Readings: Chapter 12, 12.7-12.9

Winter 2012

Lecture 15

Linking Two Approaches to Hypothesis Testing

A magazine is considering the launch of an online edition. The magazine plans to go ahead only if it's convinced that more than 25% of current readers would subscribe. The magazine contacts a random sample of 500 currents subscribers and 137 of those surveyed expressed interest. Should the magazine go ahead?

$$H_0 : p = ?$$

$$H_A : p - ?$$

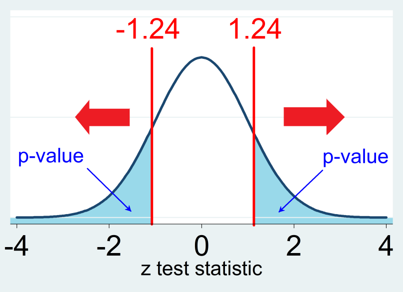
- $\hat{p} = \frac{137}{500}$
- $n = 500$

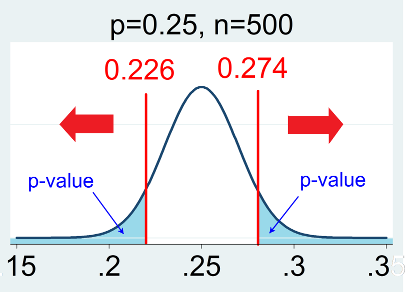
P-value approach

- Set significance level: $\alpha = 0.05$
- Calculate test-statistic:

$$\text{t-statistic} = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} =$$

- Find $P(z > \textit{and} < \text{test-statistic}) - ?$
- Compare p-value and significance level
- Conclusion?

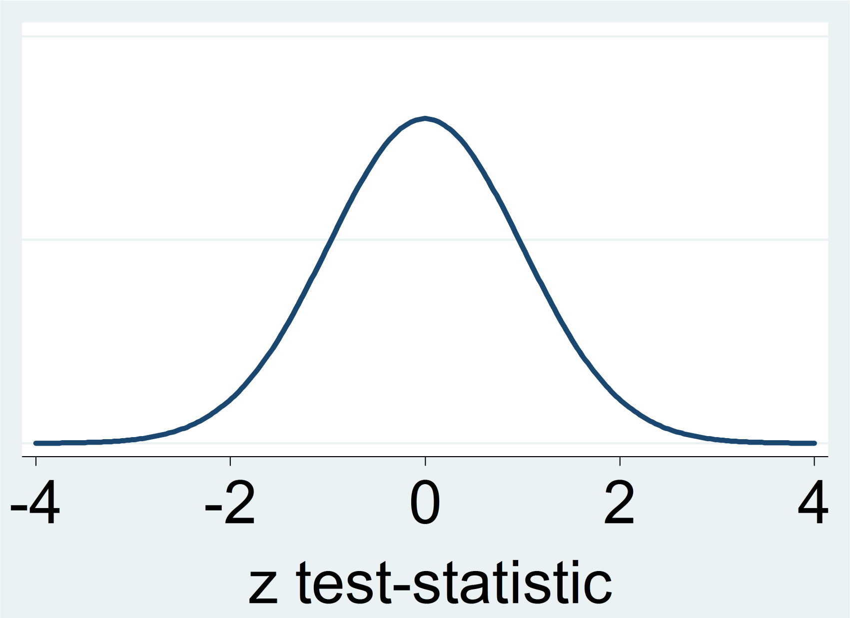

-1.24
1.24
p-value
p-value
-4
-2
0
2
4
z test statistic

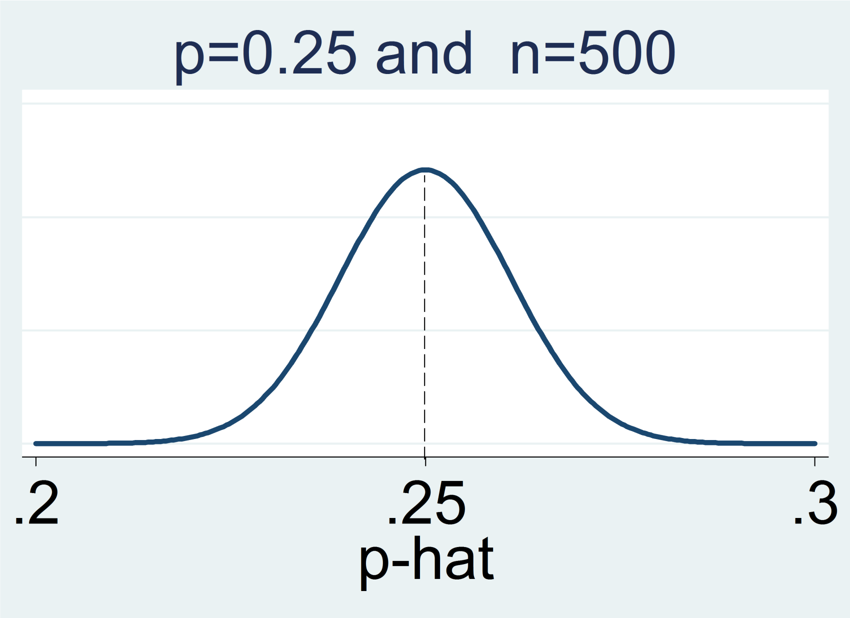

p=0.25, n=500
0.226
0.274
p-value
p-value
.15
.2
.25
.3
.35

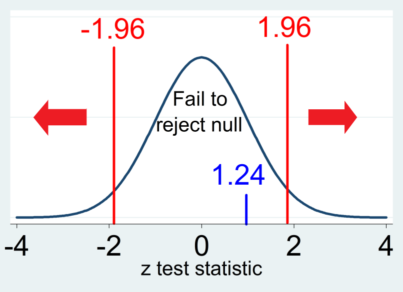
Rejection/Critical Region Approach

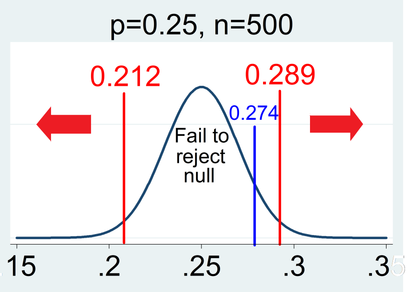
- Standardized test statistic vs Unstandardized test statistic
- Critical values from table for $\alpha = 0.05$ and two-sided test
- Draw a graph with rejection/acceptance regions
- Conclusion?
- Unstandardized critical value:

$$z_{\alpha/2} * \sqrt{\frac{p_0(1-p_0)}{n}} + p_0$$


-4
-2
0
2
4
z test-statistic


p=0.25 and n=500
.2
.25
.3
p-hat


-1.96
1.96
Fail to
reject null
1.24
-4
-2
0
2
4
z test statistic

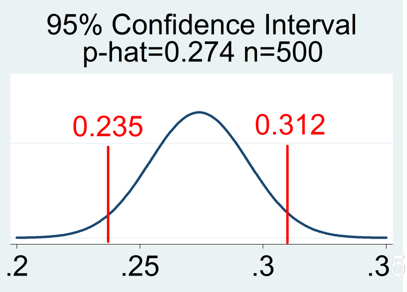

p=0.25, n=500
0.212
0.289
0.274
Fail to
reject
null
.15
.2
.25
.3
.35

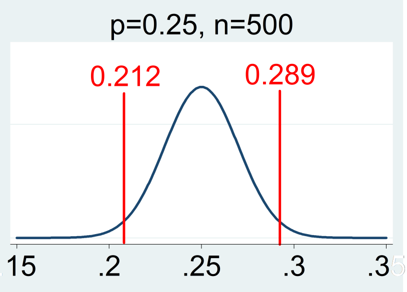
Link CI Estimators and Two-Tailed Tests

- Using a confidence interval estimator, can approximate a two-tailed hypothesis test with same α.
  - If parameter specified under H_0 is in CI, then fail to reject H_0
  - If parameter specified under H_0 is not in CI, then reject H_0 and infer H_A is true
- Recall that confidence interval is always in original units; corresponds to unstandardized version of rejection region approach

95% CI: Online Edition

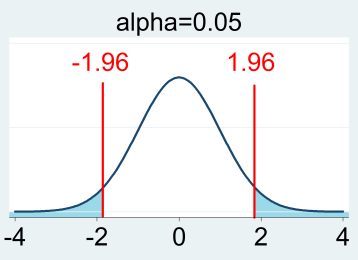
- Assume two-tailed test, i.e. $H_A : p \neq 0.25$
- Critical values from table:
- CI: $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- LCL:
- UCL:
- Important: sampling distribution when we compute CI is centered at ? , while sampling distribution in hypothesis testing is centered at ?

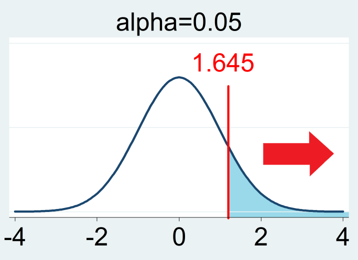

95% Confidence Interval
p-hat=0.274 n=500
0.235
0.312
.2
.25
.3
.35


p=0.25, n=500
0.212
0.289
.15
.2
.25
.3
.35

Two-Tailed vs One-Tailed Test

- Economists often use two-tailed tests even if one-tailed seems to be more reasonable
- This is because two-tailed test is more conservative - less chance of getting statistically significant results





Type I and Type II Errors

	H_0 is true state of the world (Innocent)	H_A is true state of the world (Guilty)
Fail to reject H_0 (Acquit)	No error	Type II Error
Reject H_0 (Convict)	Type I Error	No error

Type I error: reject a true null hypothesis
Type II error: fail to reject a false null hypothesis

Significance level and Type I error

- Significance level is the maximum probability of Type I error a researcher is willing to tolerate
- P-value is the actual probability of Type I error
- What can be done to reduce Type I error?

Type II error and Power

- P(Type II Error)=β=P(Fail to reject null| Null is false)
  - Trade off α and β
  - Decreasing α increases β
- Power of a test: the probability of rejecting the null hypothesis when it is false
  - Power=1-β
  - A statistical test with more power is always preferred
- P(Type II error) depends on:
  - Parameter value under H_0 and direction of H_A
  - Significance level α
  - Sample size n
  - True parameter value (p)

Finding Type II error

P(Type II error)=P(Fail to reject $H_0|H_0$ is false, H_A is true)

$$\beta = P(\hat{p} < \text{critical value}|p_A, n, \alpha)$$

$\uparrow$

opposite direction of H_A

$$\beta = P\left(\frac{\hat{p}-p_A}{\sqrt{\frac{p_A(1-p_A)}{n}}} > \frac{p_{critical}-p_A}{\sqrt{\frac{p_A(1-p_A)}{n}}}\right)$$

Example

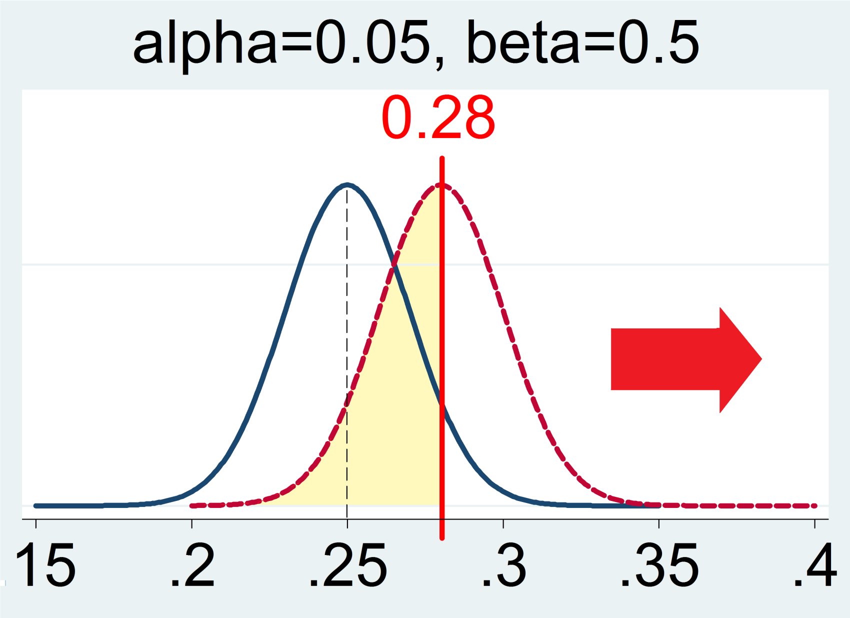
Online edition example:

- $H_0 : p = 0.25$ vs $H_A : p > 0.25$
- Significance level α=0.05
- Sample size, n=500
- True parameter value, p=0.28

$$\hat{p}_{critical} = 1.645 * \sqrt{\frac{p_0(1-p_0)}{n}} + p_0 = 0.28$$

$$\beta = P\left(z < \frac{0.28 - 0.28}{\sqrt{\frac{0.28*0.72}{500}}}\right) = P(z < 0) = 0.5$$

$$1 - \beta = 1 - 0.5 = 0.5$$


alpha=0.05, beta=0.5
0.28
15
.2
.25
.3
.35
.4

Effect of significance level, α, on P(Type II error)

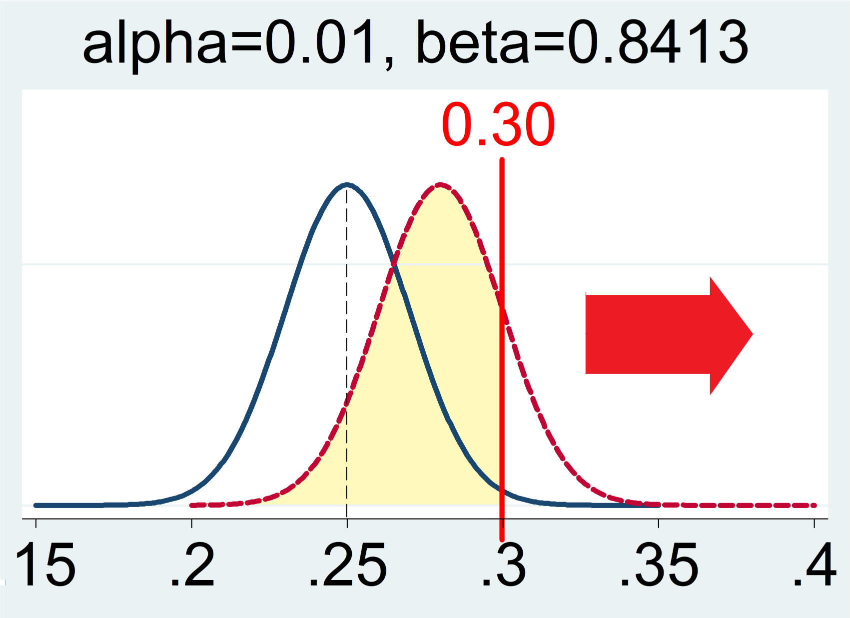
Online edition example:

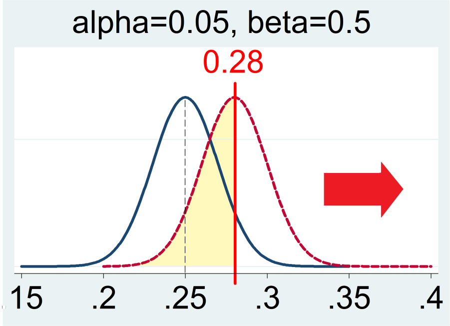
- $H_0 : p = 0.25$ vs $H_A : p > 0.25$
- Significance level α=0.01
- Sample size, n=500
- True parameter value, p=0.28

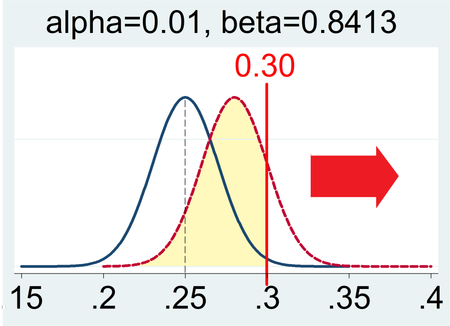
$$\hat{p}_{critical} = 2.33 * \sqrt{\frac{p_0(1-p_0)}{n}} + p_0 = 0.30$$

$$\beta = P\left(z < \frac{0.30 - 0.28}{\sqrt{\frac{0.28*0.72}{500}}}\right) = P(z < 1) = 0.8413$$

$$1 - \beta = 1 - 0.8413 = 0.1587$$


alpha=0.01, beta=0.8413
0.30
15
.2
.25
.3
.35
.4


alpha=0.05, beta=0.5
0.28
.15
.2
.25
.3
.35
.4


alpha=0.01, beta=0.8413
0.30
.15
.2
.25
.3
.35
.4

Effect of sample size on P(Type II error)

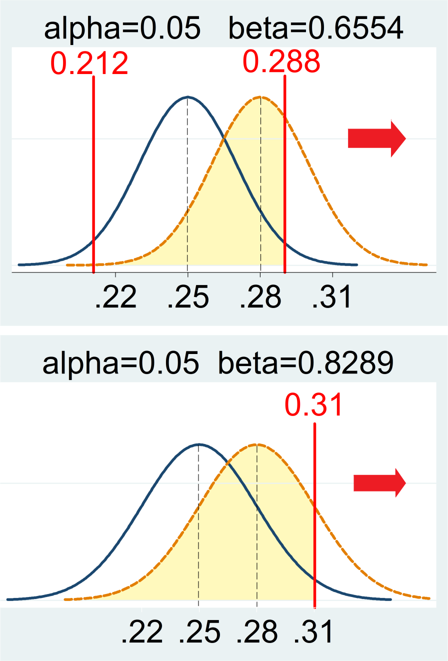
Online edition example:

- $H_0 : p = 0.25$ vs $H_A : p \neq 0.25$
- Significance level α=0.05
- Sample size,n=200
- True parameter value, p=0.28

$$\hat{p}_{critical} = 1.96 * \sqrt{\frac{0.25(1-0.25)}{200}} + 0.25 = 0.31$$

$$\beta = P\left(z < \frac{0.31-0.28}{\sqrt{\frac{0.28*0.72}{200}}}\right) = P(z < 0.95) = 0.8289$$

$$1 - \beta = 1 - 0.8289 = 0.1711$$


alpha=0.05 beta=0.6554
0.212
0.288
.22 .25 .28 .31
alpha=0.05 beta=0.8289
0.31
.22 .25 .28 .31

Effect of parameter value under H_A on P(Type II error)

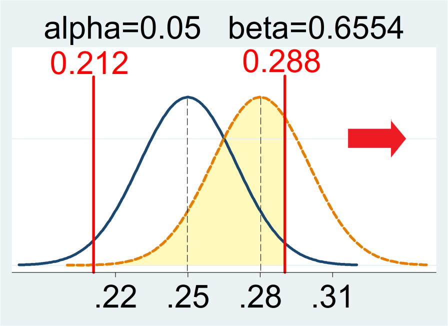
Online edition example:

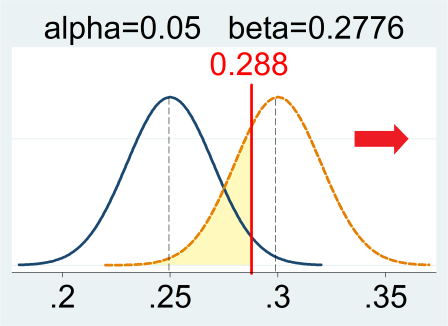
- $H_0 : p = 0.25$ vs $H_A : p \neq 0.25$
- Significance level α=0.05
- Sample size, n=500
- True parameter value, p=0.30

$$\hat{p}_{critical} = 1.96 * \sqrt{\frac{0.25(1-0.25)}{500}} + 0.25 = 0.288$$

$$\beta = P\left(z < \frac{0.288 - 0.30}{\sqrt{\frac{0.30*0.70}{500}}}\right) = P(z < -0.59) = 0.2776$$

$$1 - \beta = 1 - 0.2776 = 0.7224$$


alpha=0.05 beta=0.6554
0.212
0.288
.22
.25
.28
.31


alpha=0.05 beta=0.2776
0.288
.2
.25
.3
.35