Math 140
Introductory Statistics

Extra hours at the tutoring center

Fri Dec 3rd 10-4pm, Sat Dec 4 11-2 pm

Final Dec 14th 5:30 - 7:30pm CH 5122

Left handed people

113 out of 1067 males are left handed
92 out of 1170 females are left handed

Perform a significance test to see whether males
Are more likely to be left handed than women.

Null hypothesis?
Alternate hypothesis?

Left handed people

113 out of 1067 males are left handed
92 out of 1170 females are left handed

Perform a significance test to see whether males
Are more likely to be left handed than women.

p_1 male, p_2 female

Null hypothesis: $p_1 = p_2$
Alternate hypothesis? $p_1 > p_2$

Left handed people

You try: check applicability of normal distribution,
Check for the 95% confidence interval.

$$\hat{p}_1 = \frac{113}{1067} = 0.106;$$

$$\hat{p}_2 = \frac{92}{1170} = 0.079;$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \ldots.$$

Let’s do it a bit differently

The null hypothesis says that $p_1 = p_2$

We are going to create a POOLED estimate

Total number of left handed people = 113+ 92 = 205

Total number of people = 1067+1170 = 2237

Then, according to the null hypothesis,

p1, p2 should both be approximately 205/2237 = 0.091

Let's do it a bit differently

So we ESTIMATE the error of the difference in the following way:

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}$$

$$= \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

We also introduce a new test statistic z

Let's do it a bit differently

$$z = \frac{statistic - parameter}{estimated \quad standard \quad error} =$$

$$= \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 2.23$$

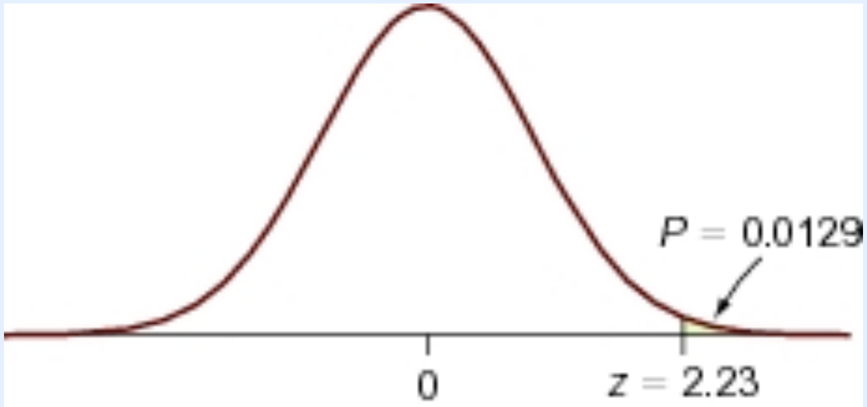
The alternative hypothesis is that $p_1 - p_2 > 0$

Let's do it a bit differently

$$z = \frac{statistic - parameter}{estimated \quad standard \quad error} =$$

$$= \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 2.23$$

Our P value is 0.0129



Overweight in America

2004: 1000 adults surveyed, 30% were overweight
2009: 1000 adults surveyed, 32% were overweight

Does this represent a real population change at the 10% significance level?

Of course, the difference is 0.32 - 0.30 = 0.02 = 2%
But is this value statistically significant?

Overweight in America

Null hypothesis, the real percentages of the Overweight population

p1 (2009) = p2 (2004)

Alternative hypothesis

p1 (2009) different than p2 (2004)

Overweight in America

Check applicability:

0.30*1000 = 300;
0.32*1000 = 320;
(1-0.30)*1000 = 700;
(1-0.32)*1000 = 680;

OK all are greater than 5

Overweight in America

Find the POOLED value of proportion
of overweight people

Total number of overweight people = 300+320

Total number of people polled = 1000+1000

The pooled estimate is

$$\hat{p} = \frac{620}{2000} = 0.31$$

The test statistic z

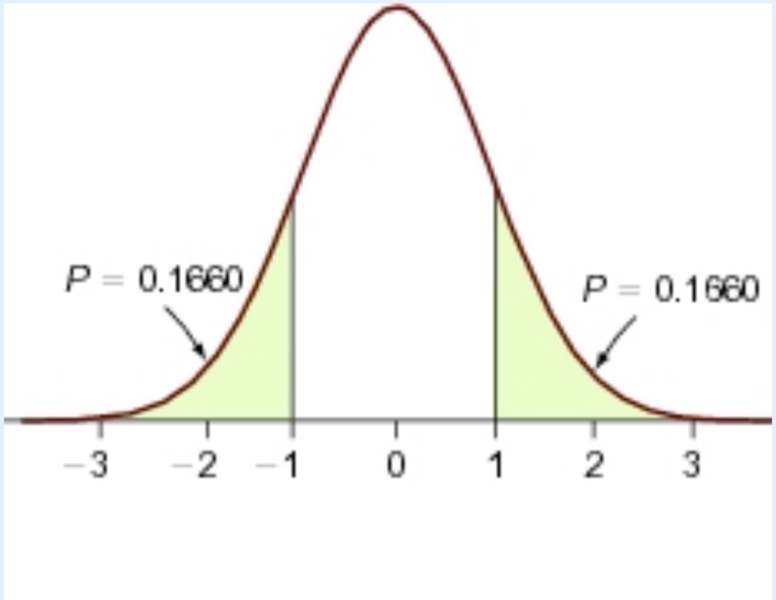
$$\hat{p} = \frac{620}{2000} = 0.31$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -0.97$$

Now, recall, we want the two proportions to be statistically DIFFERENT

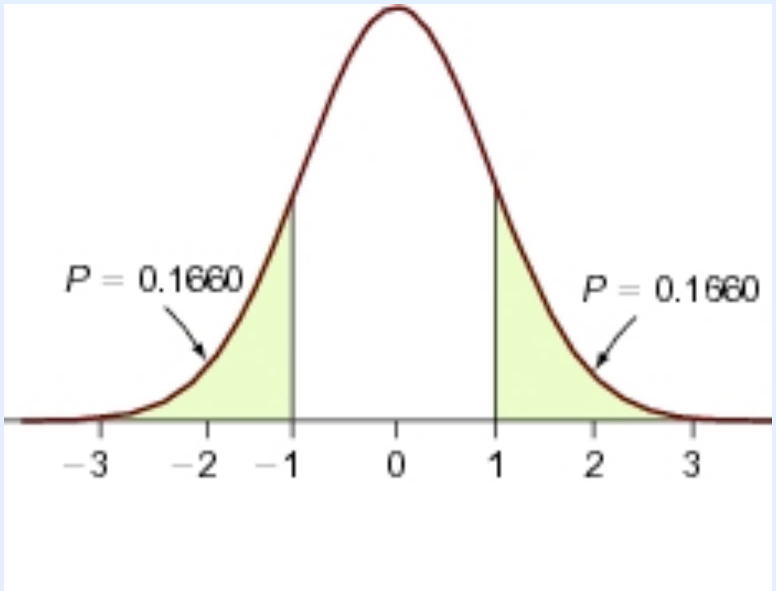
The test statistic z

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -0.97$$



So this is a double sided test,
we need to ADD the two P values

The test statistic z



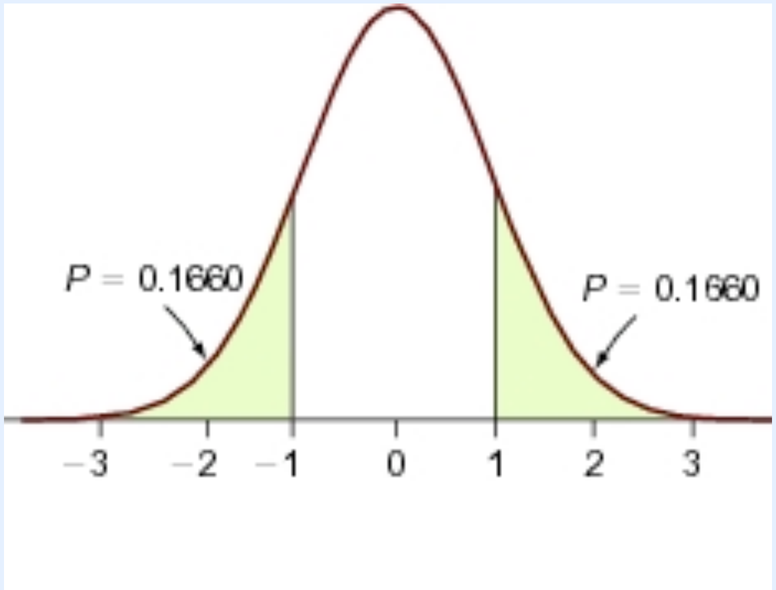
P = 0.332

To discard the null hypothesis we need

$$P < \alpha$$

For us the level of significance was 10% on either side, so we need to look at $\alpha = 0.2$

The test statistic z



P = 0.332

To discard the null hypothesis we need

Either: 0.332 < 0.2
or: 0.166 < 0.1

This is false, so we don't discard the null hypothesis

You try it in the "old way"

Construct the 90% confidence interval
For the difference of overweight people
between the years 2004 and 2009

p1 (2009) = 320 out of 1000
p2(2004) = 300 out of 1000

You should see that 0 is plausible.

AIDS: AZT or AZT+ACV?

Treated with

Survived?		AZT	AZT + ACV	Total
	No	28	13	41
	Yes	41	49	90
	Total	69	62	131

What is the null hypothesis?

What is the alternative hypothesis?

n1, n2, $\hat{p}_1, \hat{p}_2$

Hypotheses:

Null: the treatments
(AZT alone, or the cocktail AZT + ACV)
give the same survival rate

(AZT) $p1 = p2$ (ACV)

Alternative: the AZT + ACV is better

$p1 < p2$

Find:

$n_1 = 69$	sample size of patients treated with AZT
$n_2 = 62$	sample size of patients treated with AZT+ACV
$\hat{p}_1 = \frac{41}{69}$	proportion of patients in the sample treated with AZT that survived
$\hat{p}_2 = \frac{49}{62}$	proportion of patients in the sample treated with AZT+ACV that survived
p_1	True proportion of survival if all patients were treated with AZT (unknown)
p_2	True proportion of survival if all patients were treated with AZT+ACV (unknown)

Check all requirements are satisfied

The z statistic

$n_1 = 69$	sample size of patients treated with AZT
$n_2 = 62$	sample size of patients treated with AZT+ACV
$\hat{p}_1 = \frac{41}{69}$	proportion of patients in the sample treated with AZT that survived
$\hat{p}_2 = \frac{49}{62}$	proportion of patients in the sample treated with AZT+ACV that survived
p_1	True proportion of survival if all patients were treated with AZT (unknown)
p_2	True proportion of survival if all patients were treated with AZT+ACV (unknown)

Check all Conditions

Find the z statistic using

$$\hat{p} = \frac{survived}{total} = \frac{41 + 49}{69 + 62} = 0.687$$

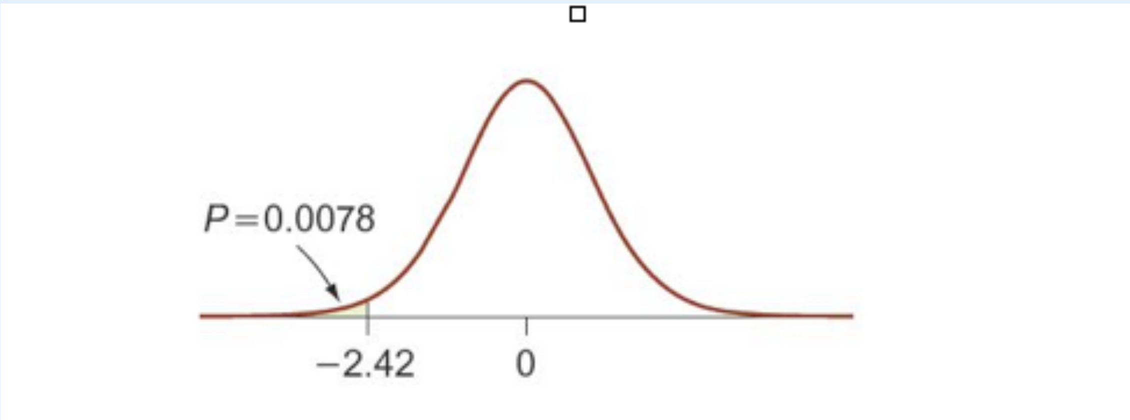
The z statistic

Our alternative hypothesis is that p1 (AZT) is worse than p2 (AZT+ACV),

$p1 < p2$

Where do we look on Table 1, page 759?

The z statistic



Check all Conditions

P value 0.0078

$$z = \frac{(0.594 - 0.790) - 0}{\sqrt{0.687(1 - 0.687)\left(\frac{1}{69} + \frac{1}{62}\right)}} = -2.415$$

The z statistic

We reject the null hypothesis, because the P value is extremely small.

$P = 0.0078$

AZT+ACV is better

The Flu

		Treatment		
		Surgical Mask	N95 Respirator	Total
Influenza?	Yes	50	48	98
	No	175	173	348
	Total	225	221	446

Two anti-flu masks esist

Surgical are cheaper, but don’t always fit right
N95 are more expensive, but fit better

Experiments to check their effectiveness

The Flu

		Treatment		
		Surgical Mask	N95 Respirator	Total
Influenza?	Yes	50	48	98
	No	175	173	348
	Total	225	221	446

Null hypothesis: p1 = p2

Alternate hypothesis p1 different than p2

The Flu

		Treatment		
		Surgical Mask	N95 Respirator	Total
Influenza?	Yes	50	48	98
	No	175	173	348
	Total	225	221	446

Check applicability using

p1 = 50/225, p2 = 48/221

The Flu

		Treatment		
		Surgical Mask	N95 Respirator	Total
Influenza?	Yes	50	48	98
	No	175	173	348
	Total	225	221	446

Test statistic

$$z = \frac{(50/225 - 48/221) - 0}{\sqrt{0.219(1-0.219)\left(\frac{1}{225} + \frac{1}{221}\right)}} = 0.128$$

The Flu

		Treatment		
		Surgical Mask	N95 Respirator	Total
Influenza?	Yes	50	48	98
	No	175	173	348
	Total	225	221	446

P value = 2*(0.4483) = 0.8966

We need to see whether they are different,
so this is a two sided test

LARGE! The two masks perform more in the same way.

The Flu

This means:

If the difference in flu infectivity
was really 0 between the two mask types

then the probability we would get a difference
of p1-p2 = 50/225 - 48/221 = 0.005
between our measured data is 0.8990

Large probability, most likely our assumption is correct.

Helmets

Location Type	% Helmet Use		Sample Size	
	1999	2002	1999	2002
Local streets	16	19	1116	848
Collector streets	25	30	592	513
Greenways	42	55	404	369
Mountain biking trails	84	90	336	219
Total			2448	1949

A law imposing helmet use was passed in early 2002.
Do these numbers support or disprove the fact that that the law was effective in having people wear their helmet?

Helmets

	% Helmet Use		Sample Size	
Location Type	1999	2002	1999	2002
Local streets	16	19	1116	848
Collector streets	25	30	592	513
Greenways	42	55	404	369
Mountain biking trails	84	90	336	219
Total			2448	1949

1999 = 0.16*1116 =178.56
people were wearing their helmets

2002 = 0.19*848 = 161.2
people were wearing their helmets

Helmets

Location Type	% Helmet Use 1999	% Helmet Use 2002	Sample Size 1999	Sample Size 2002
Local streets	16	19	1116	848
Collector streets	25	30	592	513
Greenways	42	55	404	369
Mountain biking trails	84	90	336	219
Total			2448	1949

Pooled p = (178.56+161.2)/(1116+848) = 0.173

Helmets

$$z = \frac{(0.16 - 0.19) - 0}{\sqrt{0.173(1 - 0.173)\left(\frac{1}{1116} + \frac{1}{848}\right)}} = -1.741$$

We are asking if the law changed things so that
$p1 < p2$

The P value is 0.04

Very low. Certainly there is a difference,
But maybe not necessarily due to the law!

Homework

Page 455 E35 - E45

Page 441 E17-E27