Math 1710 Class 30

2-Sample Prop., Examples

Dr. Back

Nov. 6, 2009

300 stdnts, 60% approve; 200 faclty, 65%

A significant difference? CI for difference in rates of approval?

300 stdnts, 60% approve; 200 faclty, 65%

A significant difference? CI for difference in rates of approval? Let p_1 and p_2 denote the true proportions of students and faculty that approve.

300 stdnts, 60% approve; 200 faclty, 65%

A significant difference? CI for difference in rates of approval?
2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

300 stdnts, 60% approve; 200 faclty, 65%

A significant difference? CI for difference in rates of approval?
2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

$$\mu = p_1 - p_2$$

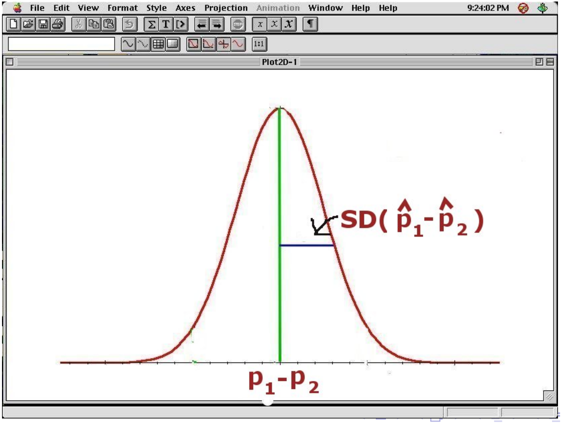
$$Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2) \quad = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

300 stdnts, 60% approve; 200 faclty, 65%

A significant difference? CI for difference in rates of approval?
2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Samp. Dist. of $\hat{p}_1 - \hat{p}_2$: $N(p_1 - p_2, SD(\hat{p}_1 - \hat{p}_2))$



300 stdnts, 60% approve; 200 faclty, 65%

2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Find a CI for $p_1 - p_2$:

300 stdnts, 60% approve; 200 faclty, 65%

2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Find a CI for $p_1 - p_2$:

Since we don't know p_1 and p_2, we can't directly compute $SD(\hat{p}_1 - \hat{p}_2)$.

So we use $SE(\hat{p}_1 - \hat{p}_2)$ instead.

300 stdnts, 60% approve; 200 faclty, 65%

2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Find a CI for $p_1 - p_2$:

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

300 stdnts, 60% approve; 200 faclty, 65%

2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Find a CI for $p_1 - p_2$:

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Same argument as in the 1-sample case gives a CI for $p_1 - p_2$ of

$$\hat{p}_1 - \hat{p}_2 \pm z^* SE(\hat{p}_1 - \hat{p}_2).$$

300 stdnts, 60% approve; 200 faclty, 65%

2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Find a CI for $p_1 - p_2$:

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Same argument as in the 1-sample case gives a CI for $p_1 - p_2$ of

$$\hat{p}_1 - \hat{p}_2 \pm z^* SE(\hat{p}_1 - \hat{p}_2).$$

Here we have

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{.6 \cdot .4}{300} + \frac{.65 \cdot .35}{200}} = .0440.$$

300 stdnts, 60% approve; 200 faclty, 65%

2-sample inference based on the sampling distribution of $\hat{p}_1 - \hat{p}_2$

Find a CI for $p_1 - p_2$:

Same argument as in the 1-sample case gives a CI for $p_1 - p_2$ of

$$\hat{p}_1 - \hat{p}_2 \pm z^* SE(\hat{p}_1 - \hat{p}_2).$$

Here we have

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{.6 \cdot .4}{300} + \frac{.65 \cdot .35}{200}} = .0440.$$

A 95% CI for $p_1 - p_2$ is:

$$(.6 - .65) \pm 1.96 \cdot .0440 = -.05 \pm .0863 = (-.1363, .0363).$$

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

Without the request for P-value, we could use the CI above.
But for the P-value we need to use "Method 1."

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

Our hypotheses are:

- H_0: $p_1 = p_2$
- H_a: $p_1 \neq p_2$

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

Our hypotheses are:

- H_0: $p_1 = p_2$
- H_a: $p_1 \neq p_2$

A twist enters. We are only interested in the reasonableness of our observed $\hat{p}_1 - \hat{p}_2$ with respect to the sampling dist if H_0 is true. There are many such distributions (since we don't know the common value of $p_1 = p_2$ to use.) In particular what we did with $SE(\hat{p}_1 - \hat{p}_2)$ above does not fit the $p_1 = p_2$ situation.

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

Our hypotheses are:

- H_0: $p_1 = p_2$
- H_a: $p_1 \neq p_2$

We resolve this conflict by making our best estimate of the common value of p_1 and p_2, namely the weighted average

$$\hat{p}_{pooled} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$$

and then

$$SE_{pooled}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\widehat{p_{pooled}}\,\widehat{q_{pooled}}}{n_1} + \frac{\widehat{p_{pooled}}\,\widehat{q_{pooled}}}{n_2}}.$$

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

Here the weighted average is

$$\hat{p}_{pooled} = \frac{300 \cdot .60 + 200 \cdot .65}{200 + 300} = .6 \cdot 300 + .4 \cdot .65 = .62$$

and then

$$SE_{pooled}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{.62 \cdot .38}{300} + \frac{.62 \cdot .38}{200}} = .0443.$$

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

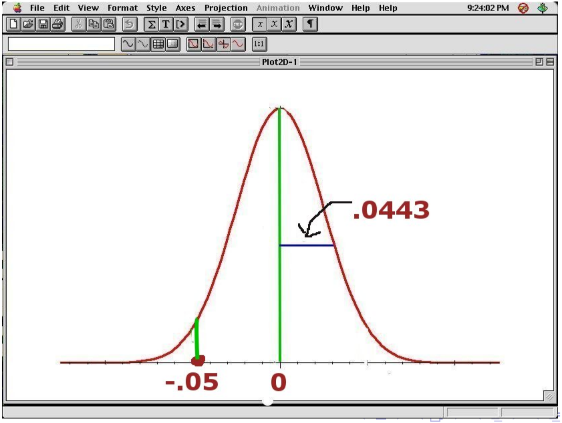
Our z-statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{pooled}(\hat{p}_1 - \hat{p}_2)} = \frac{-.05}{.0443} = -1.12.$$

300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

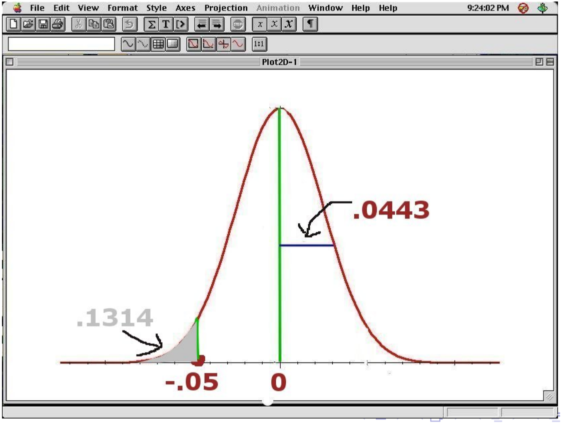
Approx Samp. Dist. of $\hat{p}_1 - \hat{p}_2$: $N(0, .0443)$ if H_0 is true.



300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

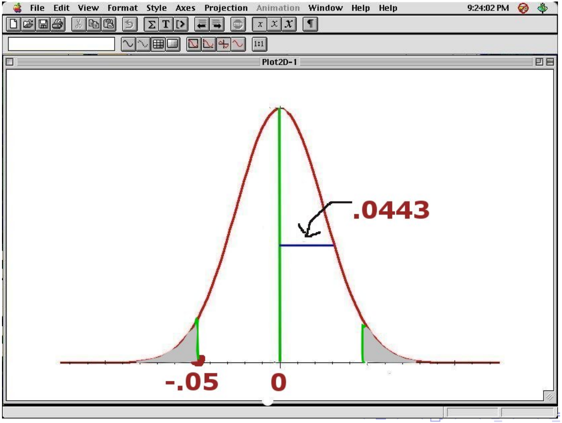
Tail Prob. is $P(Z < -1.12) = .1314$.



300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

P-value=2(Tail Prob.) $= 2(.1314) = .2628$



300 stdnts, 60% approve; 200 faclty, 65%

Carry out an HT at a sig level of $\alpha = .05$ of whether faculty and student approval rates are different. Calculate the P-value as well.

Our P-value is larger than $\alpha = .05$, so we retain H_0.

Three Kinds of 1-Sample H_a's

H_a for a prop. HT can be any of the three possibilities

Two-Sided $p \neq p_0$

One-Sided $p > p_0$

One-Sided $p < p_0$

Three Kinds of 1-Sample $H_a's$

H_a for a prop. HT can be any of the three possibilities

Two-Sided $p \neq p_0$

One-Sided $p > p_0$

One-Sided $p < p_0$

One chooses among these based on the question being studied.

Three Kinds of 1-Sample $H_a's$

H_a for a prop. HT can be any of the three possibilities

Two-Sided $p \neq p_0$

One-Sided $p > p_0$

One-Sided $p < p_0$

One chooses among these based on the question being studied.

A question like "Is there strong evidence that p has *changed* . . ." would point to 2-sided.

Three Kinds of 1-Sample H_a's

H_a for a prop. HT can be any of the three possibilities

Two-Sided $p \neq p_0$

One-Sided $p > p_0$

One-Sided $p < p_0$

One chooses among these based on the question being studied.

A question like "Is there strong evidence that p has *increased* . . ." would point to 1-sided.

Three Kinds of 1-Sample $H_a's$

H_a for a prop. HT can be any of the three possibilities

Two-Sided $p \neq p_0$

One-Sided $p > p_0$

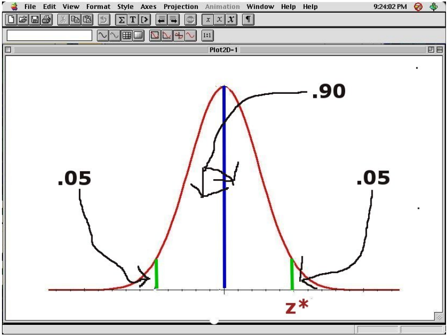
One-Sided $p < p_0$

One chooses among these based on the question being studied.

The value of $\hat{p}$ never plays a role in formulating hypotheses!

Three Kinds of 1-Sample H_a's

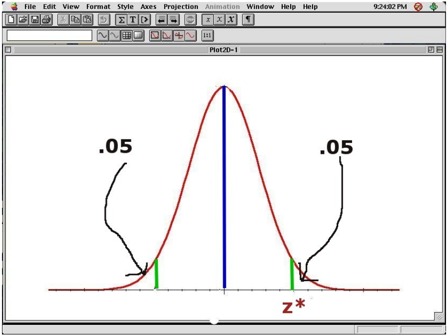
N(0,1) and the critical value z^*



Three Kinds of 1-Sample H_a's

N(0,1)

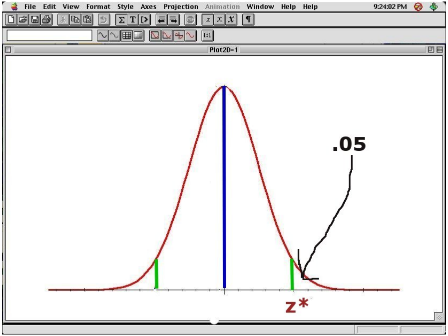
$H_a: \; p \neq p_0$ (2-sided), $\alpha = .10$



Three Kinds of 1-Sample H_a's

N(0,1)

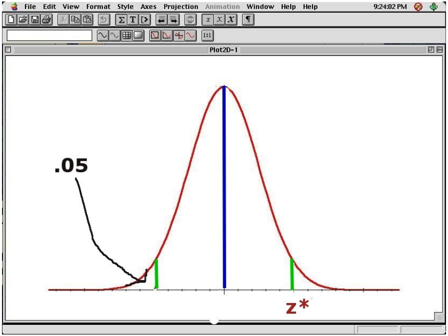
$H_a: \; p > p_0$ (1-sided), $\alpha = .05$



Three Kinds of 1-Sample H_a's

N(0,1)

$H_a: \ p < p_0$ (1-sided), $\alpha = .05$



Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

random sampling -
could be critical;
might be ok if "representative"
representative hard/impossible to define

Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

plausible independence -
could be critical
sometimes just a working hypothesis

Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

<u>What Happens if Not Satisfied:</u>

10% condition -
results in overestimation of samp. dist. st dev
gradual breakdown in formulas, not method

Conditions for Prop Tests

- plausible independence
- random sampling
- 10% condition
- success/failure

What Happens if Not Satisfied:

success/failure -
progressive reduction of accuracy
accuracy varies regardless for smaller values of n

Conditions for Prop Tests

For 2-sample inference, we add the

independence groups assumption.

The chance of an individual in one of the groups assuming a certain value should be independent of the values assumed by any of the individuals in the other group.

Conditions for Prop Tests

For 2-sample inference, we add the

independence groups assumption.

The chance of an individual in one of the groups assuming a certain value should be independent of the values assumed by any of the individuals in the other group.

In 2-sample hypothesis testing, best to use $\hat{p}_{pooled}$ in success/failure.

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

Notation: Let p denote the proportion of Americans in 1995 who had never smoked.

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

Hypotheses:

- H_0: $p = .44$
- H_a: $p \neq .44$

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

Plausible Independence: Hopefully.

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

Random Sampling: Stated.

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

10% Condition: Much less than the national population.

Smoking *2nd Edition* Ch. 20 # 15

National data in the 1960s showed that about 44% of the adult population had never smoked cigarettes. In 1995 a national health survey interviewed a random sample of 881 adults and found that 52% had never been smokers.

- (a) Create a 95% CI for the proportion of adults (in 1995) who had never been smokers.
- (b) Does this provide evidence of a change in behavior among Americans? Using your CI, test an appropriate hypothesis and state your conclusion.

Success/Failure:

- $881 \cdot .52 \geq 10$
- $881 \cdot .48 \geq 10$

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

Notation: Let p denote the proportion of executives *(in companies like this one, perhaps)* who are women.

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

Hypotheses: If potentially proving the company is wrong:

- H_0: $p = .4$ $(or\ p \geq .4)$
- H_a: $p < .4$

If potentially proving the company is right:

- H_0: $p = .4$ $(or\ p \leq .4)$
- H_a: $p > .4$

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

Plausible Independence: Hopefully.

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

Random Sampling: Hopefully representative of a much larger population.

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

10% Condition: Depends on definition of the population. Hopefully much less than 10% of population.

Women Executives Ch. 20 #25

A company is criticized because only 13 of 43 people in executive-level positions are women. The company explains that although this proportion is lower than it might wish, it's not surprising given that only 40% of all its employees are women. What do you think? Test an appropriate hypothesis and state your conclusion.

Success/Failure:

- $43\left(\frac{13}{43}\right) \geq 10$
- $43\left(\frac{30}{43}\right) \geq 10$

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

Notation: Let p denote the proportion of students in districts like this one who drop out.

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

Hypotheses:

- H_0: $p = 10.3$ *(or* $p \leq .103)$
- H_a: $p > .103$

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

Plausible Independence: Hopefully.

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

Random Sampling: Hopefully representative of a much larger population.

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

10% Condition: Depends on definition of the population.
Hopefully much less than 10% of population.
Certainly much less than the national population.

Dropouts Ch. 20 #27

Some people are concerned that new tougher standards and high-stakes tests adopted in many states have driven up the high school dropout rate. The National Center for Education Statistics reported that the high school dropout rate for the year 2004 was 10.3%. One school district whose dropout rate has always been very close to the national average reports that 210 of their 1782 high school students dropped out last year. Is this evidence that their dropout rate may be increasing? Explain.

Success/Failure:

- $1782\left(\frac{210}{1782}\right) \geq 10$
- $1782\left(\frac{1572}{1782}\right) \geq 10$

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

Notation: Let p denote the proportion of lost luggage that is returned within 24 hours.

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

Hypotheses:

- H_0: $p = .9$ (or $p \geq .9$)
- H_a: $p < .9$

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

Plausible Independence: Hopefully.

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

Random Sampling: Hopefully.

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

10% Condition: Depends on definition of the population.
Hopefully much less than 10% of population.
Certainly much less than total volume of luggage.

Lost Luggage Ch. 20 #29

An airline's public relations department says that the airline rarely loses passengers' luggage. It further claims that on those occasions when luggage is lost, 90% is recovered and delivered to its owner within 24 hours. A consumer group that surveyed a large number of air travelers found that only 103 of 122 people who lost luggage on that airline were reunited with the missing items by the next day. Does this cast doubt on the airline's claim? Explain.

Success/Failure:

- $122\left(\frac{103}{122}\right) \geq 10$
- $122\left(\frac{19}{122}\right) \geq 10$

Chicken Contamination

Suppose:

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Questions:

1. Purdue safer than Store Brand?
2. Tyson safer than Store Brand?
3. Tyson different in safety than Store Brand?
4. Confidence interval for difference in safety between Store Brand and Tyson?

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Purdue safer than Store Brand?

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Purdue safer than Store Brand?

Notation: Let p_1 denote the proportion of Purdue which are contaminated and p_2 the proportion for Store Brand.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Purdue safer than Store Brand?

Notation: Let p_1 denote the proportion of Purdue which are contaminated and p_2 the proportion for Store Brand.

Hypotheses:

- H_0: $p_1 = p_2$ (or $p_1 \geq p_2$)
- H_a: $p_1 < p_2$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Hypotheses:

- H_0: $p_1 = p_2$ $(or\ p_1 \geq p_2)$
- H_a: $p_1 < p_2$

$$\hat{p}_{pooled} = \frac{.33 \cdot 75 + .45 \cdot 75}{75 + 75} = .39$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Hypotheses:

- H_0: $p_1 = p_2$ (or $p_1 \geq p_2$)
- H_a: $p_1 < p_2$

$$\hat{p}_{pooled} = \frac{.33 \cdot 75 + .45 \cdot 75}{75 + 75} = .39$$

$$SE_{pooled}(\hat{p}_1 - \hat{p}_2) = \sqrt{.39 \cdot .61 \left(\frac{1}{75} + \frac{1}{75} \right)} = .0796.$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Hypotheses:

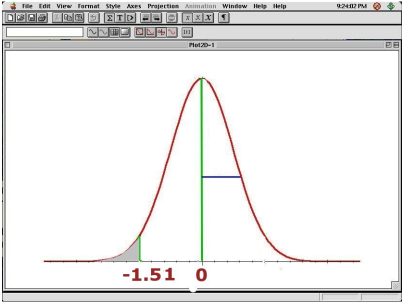
- H_0: $p_1 = p_2$ $(or\ p_1 \geq p_2)$
- H_a: $p_1 < p_2$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{pooled}} = \frac{-.12}{.0796} = -1.51.$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{pooled}} = \frac{-.12}{.0796} = -1.51.$$



N(0,1)

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Hypotheses:

- H_0: $p_1 = p_2$ (or $p_1 \geq p_2$)
- H_a: $p_1 < p_2$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{pooled}} = \frac{-.12}{.0796} = -1.51.$$

P-value = tail probability = $P(Z < -1.51) = .0655$.
At a level of $\alpha = .05$, we'd retain H_0.
Purdue might not be safer.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson safer than Store Brand?

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson safer than Store Brand?

Notation: Let p_2 denote the proportion of Store Brand which are contaminated and p_3 the proportion for Tyson.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson safer than Store Brand?

Notation: Let p_2 denote the proportion of Store Brand which are contaminated and p_3 the proportion for Tyson.

Hypotheses:

- H_0: $p_3 = p_2$ $(or\ p_3 \geq p_2)$
- H_a: $p_3 < p_2$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

$$\hat{p}_{pooled} = \frac{.45 \cdot 75 + .56 \cdot 75}{75 + 75} = .505$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

$$\hat{p}_{pooled} = \frac{.45 \cdot 75 + .56 \cdot 75}{75 + 75} = .505$$

$$SE_{pooled}(\hat{p}_2 - \hat{p}_3) = \sqrt{.505 \cdot .495 \left(\frac{1}{75} + \frac{1}{75} \right)} = .0816.$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

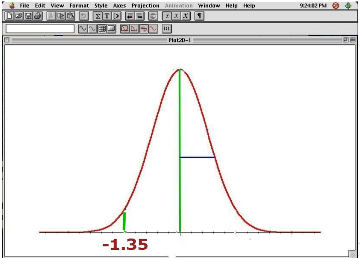
$$z = \frac{\hat{p}_2 - \hat{p}_3}{SE_{pooled}} = \frac{-.11}{.0816} = -1.35.$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

$$z = \frac{\hat{p}_2 - \hat{p}_3}{SE_{pooled}} = \frac{-.11}{.0816} = -1.35.$$

Which side provides as much or more support for H_a of $p_3 < p_2$?

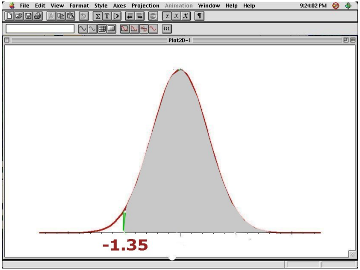


Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

$$z = \frac{\hat{p}_2 - \hat{p}_3}{SE_{pooled}} = \frac{-.11}{.0816} = -1.35.$$

Which side provides as much or more support for $p_3 < p_2$?



Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Our statistic provides no support for H_a so we immediately retain H_0.
It is a matter of convention whether we'd view the P-value as .5 or even larger.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson different in safety than Store Brand?

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson different in safety than Store Brand?
Notation: Let p_2 denote the proportion of Store Brand which are contaminated and p_3 the proportion for Yson.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson different in safety than Store Brand?

Notation: Let p_2 denote the proportion of Store Brand which are contaminated and p_3 the proportion for Yson.

Hypotheses:

- H_0: $p_2 = p_3$
- H_a: $p_2 \neq p_3$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

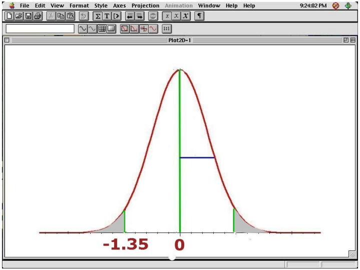
Question: Tyson different in safety than Store Brand?

Still $\hat{p}_{pooled} = .505$, $SE_{pooled}(\hat{p}_2 - \hat{p}_3) = .0816$, $z = -1.35$.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Question: Tyson different in safety than Store Brand?



Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

tail probability $= P(Z < -1.35) = .0885$.
P-value $=$ 2(tail probability)=2(.0885)=.177
At a level of $\alpha = .05$, we'd retain H_0.
Tyson might not have a different level of safety than Store Brand.

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Confidence interval for difference in safety between Store Brand and Tyson?

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Confidence interval for difference in safety between Store Brand and Tyson?

$$SE_{pooled} = \sqrt{\frac{.45 \cdot .55}{75} + \frac{.56 \cdot .44}{75 + 75}} = .0812$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Confidence interval for difference in safety between Store Brand and Tyson?

$$SE_{pooled} = \sqrt{\frac{.45 \cdot .55}{75} + \frac{.56 \cdot .44}{75+75}} = .0812$$

A 95% CI for $p_2 - p_3$ would be

$$-.11 \pm 1.96 \cdot .0812 = -.11 \pm .159 = (-.269, .049)$$

Chicken Contamination

1. 33% of 75 Perdue chickens contaminated.
2. 45% of 75 Store Brand chickens contaminated.
3. 56% of 75 Tyson chickens contaminated.

Confidence interval for difference in safety between Store Brand and Tyson?

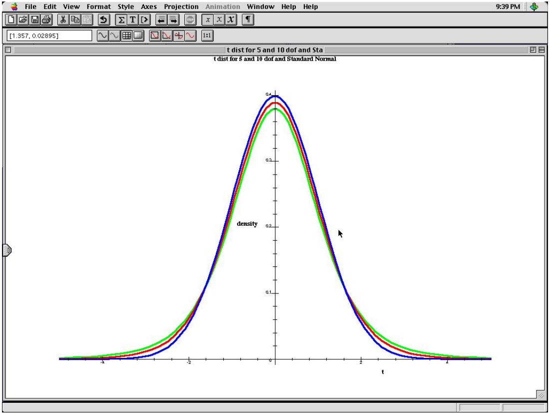
$$SE_{pooled} = \sqrt{\frac{.45 \cdot .55}{75} + \frac{.56 \cdot .44}{75 + 75}} = .0812$$

A 95% CI for $p_2 - p_3$ would be

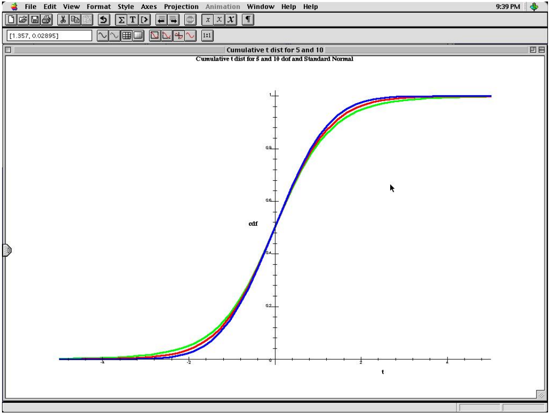
$$-.11 \pm 1.96 \cdot .0812 = -.11 \pm .159 = (-.269, .049)$$

The fact that this CI contains 0 is another way of doing the last 2 HT's.

t-Distributions



t-Distributions



t-Distributions

t distribution critical value

two tail	0.20	0.10	0.05	0.02	0.01
one tail	0.10	0.05	0.025	0.01	0.005
df					
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
35	1.306	1.690	2.030	2.438	2.725
40	1.303	1.684	2.021	2.423	2.704
45	1.301	1.679	2.014	2.412	2.690
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
75	1.293	1.665	1.992	2.377	2.643
100	1.290	1.660	1.984	2.364	2.626
120	1.289	1.658	1.980	2.358	2.617
140	1.288	1.656	1.977	2.353	2.611
180	1.286	1.653	1.973	2.347	2.603
250	1.285	1.651	1.969	2.341	2.596
400	1.284	1.649	1.966	2.336	2.588
1000	1.282	1.646	1.962	2.330	2.581
∞	1.282	1.645	1.960	2.326	2.576
evels / conf. levels	80%	90%	95%	98%	99%