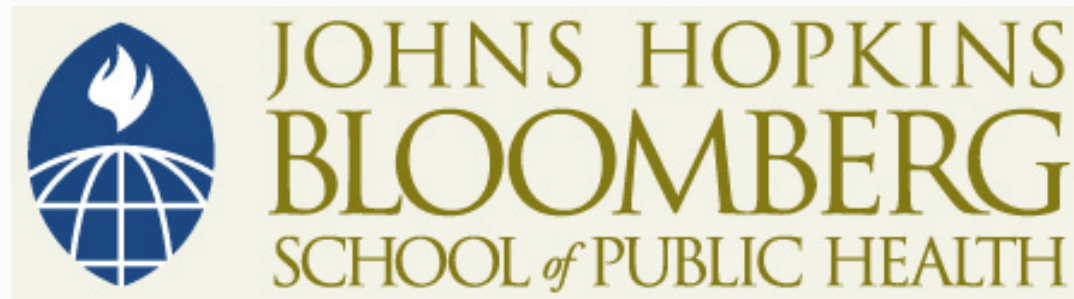


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2009, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section D

The p-Value *in Even More Detail!*

p-Values

- p-values are probabilities (numbers between 0 and 1)
- Small p-values mean that the sample results are unlikely when the null is true
- The p-value is the probability of obtaining a result as extreme or more extreme than you did by chance alone assuming the null hypothesis H_0 is true
 - How likely your sample result (and other result less likely) are if null is true

p-Values

- The p-value is not the probability that the null hypothesis is true!
- The p-value alone imparts no information about scientific/ substantive content in result of a study
- Example: from Example 3, the researchers found a statistically significant ($p=0.005!$) difference in average LDL cholesterol levels in men who had been on a diet including corn flakes versus the same men on a diet including oat bran cereal
 - Which diet showed lower average LDL levels?
 - How much was the difference; does it mean anything nutritionally?

p-Values

- If the p-value is small either a very rare event occurred and
 - H_0 is true
 - or
 - H_0 is false
- Type I error
 - Claim H_A is true when in fact H_0 is true
 - The probability of making a Type I error is called the *alpha-level (α -level)* or *significance level*

Note on the p-Value and the Alpha-Level

- If the p-value is less than some pre-determined cutoff (e.g., .05), the result is called *statistically significant*
- This cutoff is the *α -level*
 - α -level is the probability of a type I error
 - It is the probability of falsely rejecting H_0 when H_0 true
- Idea: to keep the chance of “making a mistake” when the H_0 is true low and only reject if the sample result is “unlikely”
 - Unlikeliness threshold is determined by *α -level*

Note on the p-Value and the Alpha-Level

- Truth versus decision made by hypothesis testing

		TRUTH	
		H_0	H_A
Reject H_0	Type I Error alpha-level	Power 1-beta	
Not Reject H_0		Type II Error beta	

Note on the p-Value and the Alpha-Level

- Truth versus decision made by hypothesis testing

		TRUTH	
		H_0	H_A
Decision	Reject H_0	Type I Error alpha-level	Power 1-beta
	Not Reject H_0		Type II Error beta

Note on the p-Value and the Alpha-Level

- Truth versus decision made by hypothesis testing

		TRUTH	
		H_0	H_A
Reject H_0	Type I Error alpha-level	Power 1-beta	
Not Reject H_0		Type II Error beta	

Note on the p-Value and the Alpha-Level

- Truth versus decision made by hypothesis testing

		TRUTH	
		H_0	H_A
Reject H_0	Type I Error alpha-level	Power 1-beta	
Not Reject H_0		Type II Error beta	

More on p-Value: One-Sided vs. Two-Sided Controversy

- Two-sided p-value (BP/OC: $p = .009$)
 - Probability of a result as or more extreme than observed (either positive or negative)
- One-sided p-value
 - Probability of a more extreme positive result than observed or a more extreme negative result: only considers extremes in one direction of null when evaluation how likely your sample result is (and results less likely)
 - If the direction of the alternative hypothesis in the one-sided test is the same as the direction of the sample result in terms of above/below the null, then the one-sided p-value will be half the two-sided p-value

Stata Output

- One-sided alternative: true mean difference >0
 - Sample mean difference was greater than 0

```
. ttesti 10 4.8 4.6 0
```

```
One-sample t test
```

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	10	4.8	1.454648	4.6	1.509358	8.090642

```
mean = mean(x)                                t = 3.2998
Ho: mean = 0                                  degrees of freedom = 9
```

```
Ha: mean < 0
Pr(T < t) = 0.9954
```

```
Ha: mean != 0
Pr(|T| > |t|) = 0.0092
```

```
Ha: mean > 0
Pr(T > t) = 0.0046
```

Stata Output

- One-sided alternative: true mean difference < 0
 - Sample mean difference was greater than 0

```
. ttesti 10 4.8 4.6 0
```

```
One-sample t test
```

```
-----  
      |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]  
-----+-----  
      x |      10      4.8     1.454648     4.6     1.509358     8.090642  
-----  
      mean = mean(x)                                t =      3.2998  
Ho: mean = 0                                       degrees of freedom =      9
```

```
Ha: mean < 0  
Pr(T < t) = 0.9954
```

```
Ha: mean != 0  
Pr(|T| > |t|) = 0.0092
```

```
Ha: mean > 0  
Pr(T > t) = 0.0046
```

More on the p-Value

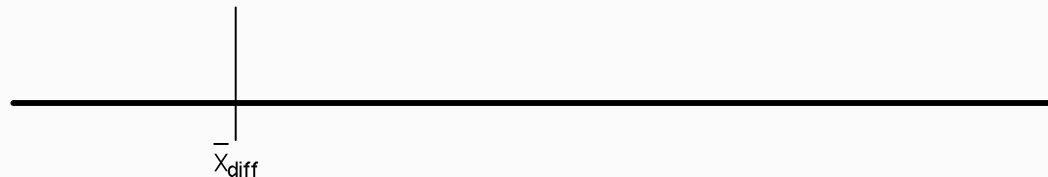
- In some cases, a one-sided alternative may not make scientific sense
 - In the absence of pre-existing information, in evaluating the BP/OC relationship, wouldn't either result be interesting and useful? (i.e., negative or positive association?)
- In some cases, a one-sided alternative often makes scientific sense
 - For example: not really interested if new treatment is worse than old treatment—only care whether it's better
- However: because of “culture of p-value” and sanctity of “.05,” one-sided p-values are viewed with suspicion
- In this course, we will use two-sided p-values exclusively

Connection: Hypothesis Testing and CIs

- The confidence interval gives plausible values for the population parameter
 - “Data take me to the truth”
- Hypothesis testing postulates two choice for the population parameter
 - “Here are two possibilities for the truth; data help me choose one”

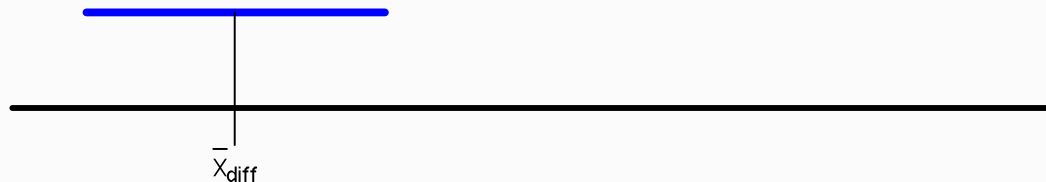
95% Confidence Interval

- If 0 is not in the 95% CI, then we would reject H_0 that $\mu = 0$ at level $\alpha = .05$ (the p-value $< .05$)
- Why?
- With confidence interval we start at sample mean difference and go two standard errors in either direction (or slightly more in small samples)



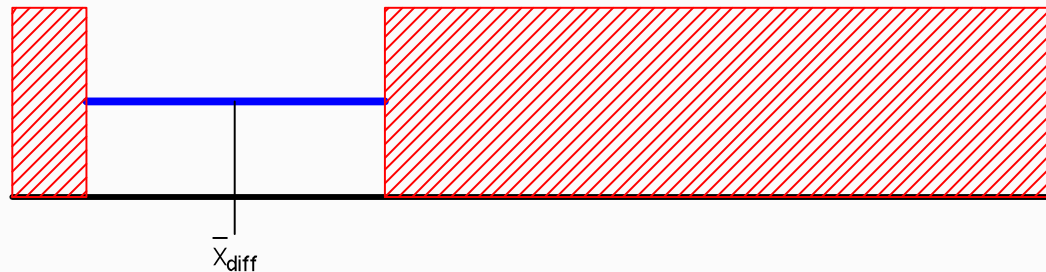
95% Confidence Interval

- If 0 is not in the 95% CI, then we would reject H_0 that $\mu = 0$ at level $\alpha = .05$ (the p-value $< .05$)
- Why?
- With confidence interval we start at sample mean difference and go two standard errors in either direction (or slightly more in small samples)



95% Confidence Interval

- If 0 is not in the 95% CI, then this must mean \bar{x} is > 2 standard errors away from 0 (either above or below)
- Hence, the distance (t) will be > 2 or < -2 : and the resulting p-value $< .05$



95% Confidence Interval and p-Value

- In the BP/OC example, the 95% confidence interval tells us that the p-value is less than .05, but it doesn't tell us that it is $p = .009$
- The confidence interval and the p-value are complementary
- However, you can't get the exact p-value from just looking at a confidence interval, and you can't get a sense of the scientific/substantive significance of your study results by looking at a p-value

More on the p-Value

- Statistical significance does not imply/prove causation
- For example: in the blood pressure/oral contraceptives example, there could be other factors that could explain the change in blood pressure
- A significant p-value is only ruling out random sampling (chance) as the explanation
- Need a comparison group to better establish causality
 - Self-selected (may be okay)
 - Randomized (better)

More on the p-Value

- *Statistical significance* is not the same as *scientific significance*
- Hypothetical example: blood pressure and oral contraceptives:
 - Suppose:
 - ▶ $n = 100,000$; $\bar{x}_{diff} = .03$ mmHg; $s = 4.6$ mmHg
 - ▶ p-value = .04
- Big n can sometimes produce a small p-value, even though the magnitude of the effect is very small (not scientifically/substantively significant)
- Very important
 - Always report a confidence interval
 - 95% CI: 0.002-0.058 mmHg

More on the p-Value

- *Lack of statistical significance* is not the same as *lack of scientific significance*
 - Must evaluate in context of study, sample size
- Small n can sometimes produce a non-significant even though the magnitude of the association at the population level is real and important (our study just can't detect it)
- *Low power* in small sample studies makes not rejecting hard to interpret
- Sometimes small studies are designed without power in mind just to generate preliminary data