

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 2006, 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

Comparing Means Amongst Two (or More) Independent Populations

John McGready

Johns Hopkins University

Lecture Topics

- ◆ CI's for mean difference between two independent populations
- ◆ Two sample t-test
- ◆ Non-parametric alternative, Mann Whitney
- ◆ Comparing means amongst more than two independent populations: ANOVA



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section A

Two Sample t-test , Part 1

Comparing Two Independent Groups

- ◆ “A Low Carbohydrate as Compared with a Low Fat Diet in Severe Obesity”¹
 - 132 severely obese subjects randomized to one of two diet groups
 - Subjects followed for six month period

Comparing Two Independent Groups

- ◆ At the End of Study Period
 - “Subjects on the low-carbohydrate diet lost more weight than those on a low fat diet (95% confidence interval for the difference in weight loss between groups, -1.6 to -6.2 kg; $p < .01$)”

Comparing Two Independent Groups

	Diet Group	
	Low-Carb	Low-Fat
Number of Subjects	64	68
Mean Weight Change (kg) Post-diet less pre-diet	-5.7	-1.8
Standard Deviation of Weight Change (kg)	8.6	3.9

- ◆ **Scientific Question**—Is Weight Loss Associated with Diet Type?

Comparing Two Independent Groups

- ◆ In statistical terms, is there a non-zero difference in the average weight loss for the subjects on the low-fat diet as compared to subjects on the low-carbohydrate diet?

Notes

- ◆ The comparison of interest is not “paired”
 - There are different subjects in each diet group
- ◆ For each subject we compute a change in weight (after diet—before weight)

Notes

- ◆ However, we are comparing the changes in weight between two independent groups!

Notes

- ◆ How do we calculate
 - Confidence interval for difference?
 - p-value to determine if the difference in two groups is “significant?”

Sampling Distribution of the Difference in Sample Means

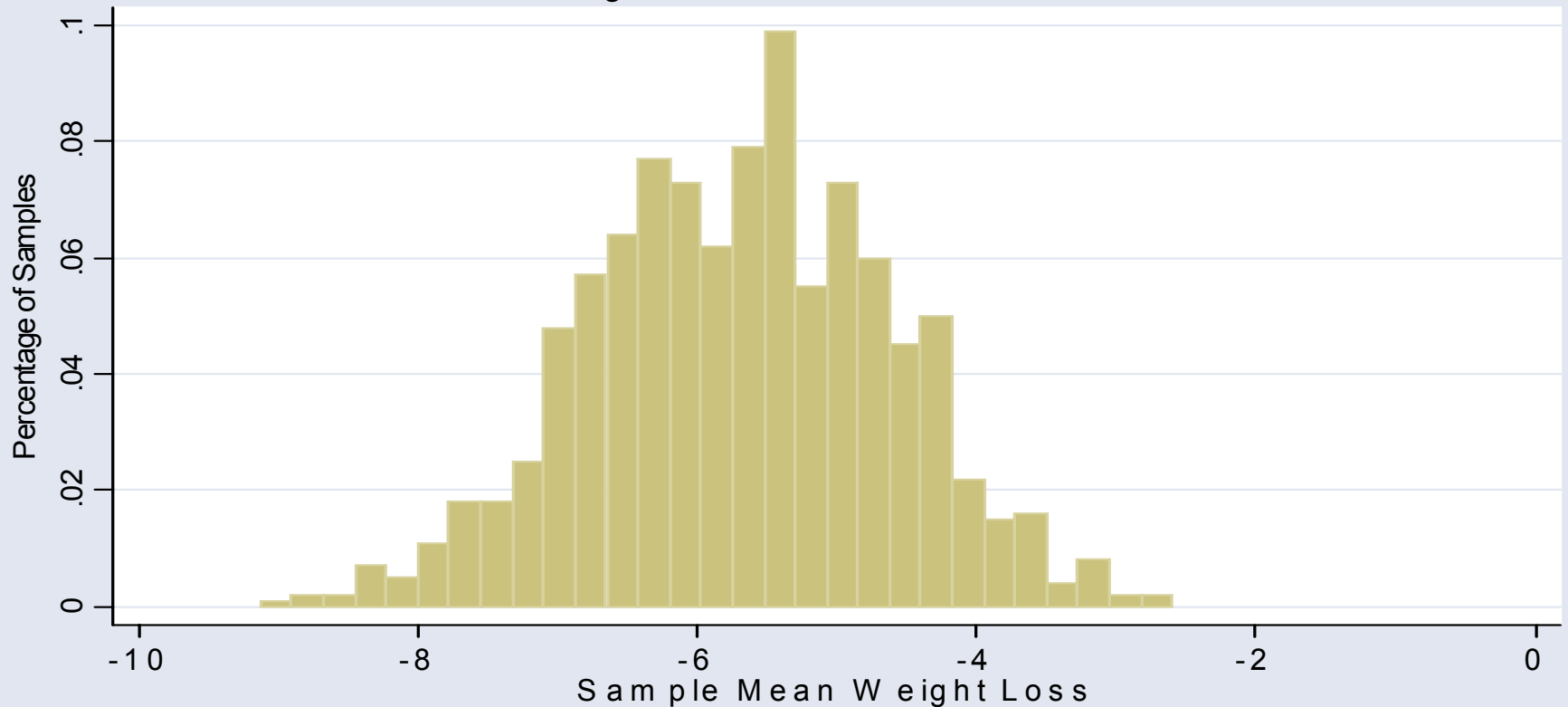
- ◆ Since we have large samples (both greater than 60) we know the sampling distributions of the sample means in both groups are approximately normal
- ◆ It turns out the difference of quantities, which are (approximately) normally distributed, are also normally distributed

Sampling Distribution of the Difference in Sample Means

- ◆ So, the big news is . . .
 - The sampling distribution of the difference of two sample means, each based on large samples, approximates a normal distribution
 - This sampling distribution is centered at the true mean difference, $\mu_1 - \mu_2$

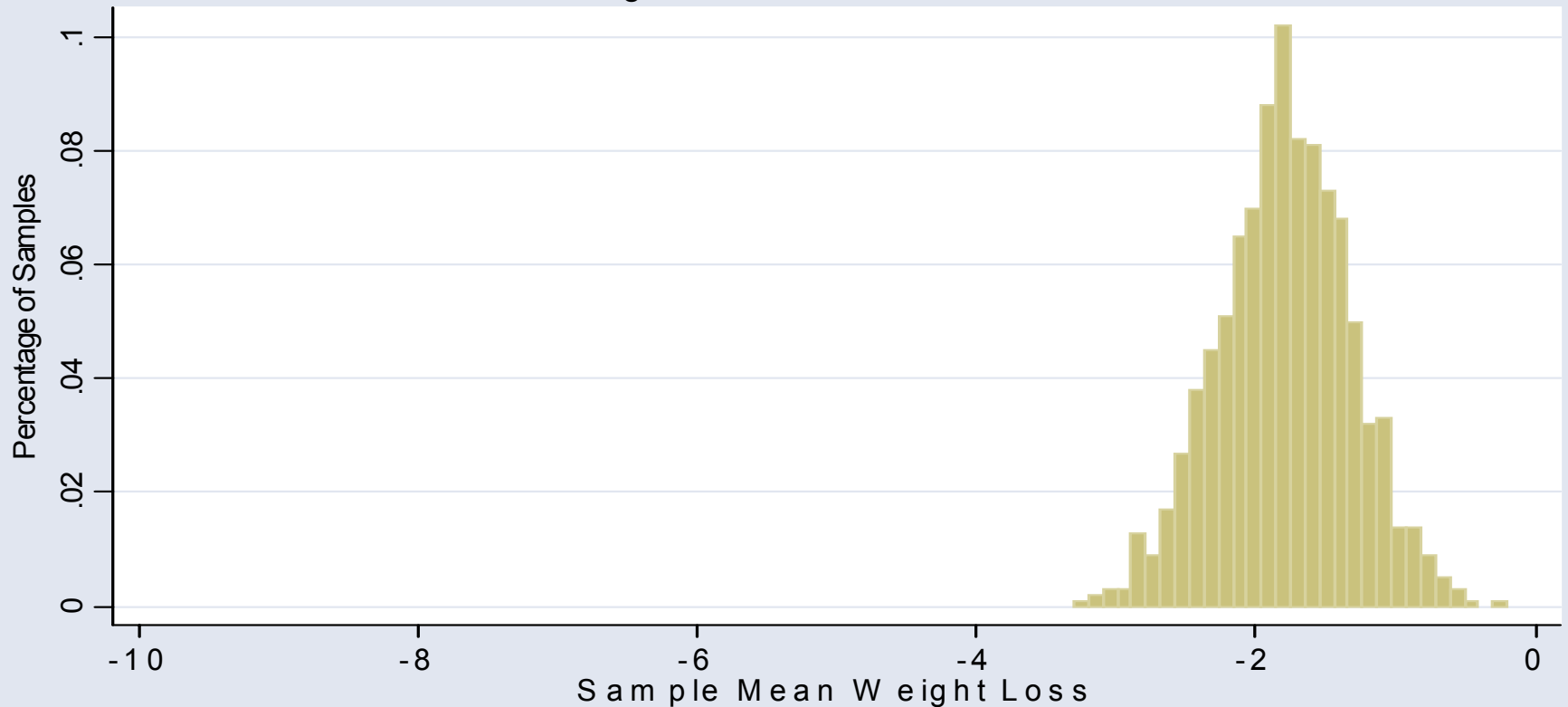
Simulated Sampling Distribution of Sample Mean Weight Loss

Simulated Sampling Distribution
Mean Weight Loss, Low Carb Diet, n = 64



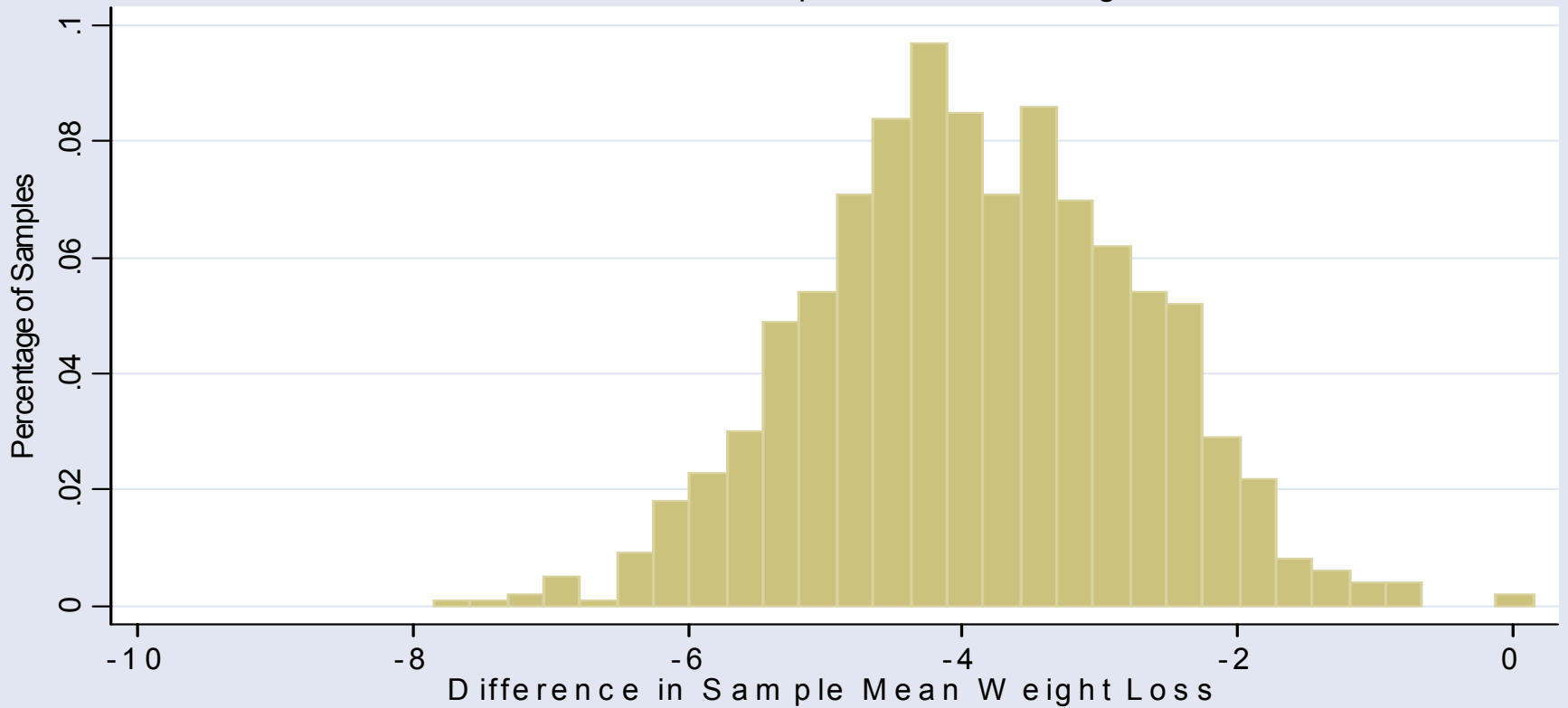
Simulated Sampling Distribution of Sample Mean Weight Loss

Simulated Sampling Distribution
Mean Weight Loss, Low Fat Diet, $n = 68$



Simulated Sampling Distribution of the Difference in Sample Means

Simulated Sampling Distribution
Difference in Sample Mean Weight Loss



95% Confidence Interval for Difference in Means

- ◆ Our most general formula:

(our best estimate) ± 2 *(SE of our best estimate)

95% Confidence Interval for Difference in Means

- ◆ Well, our best estimate for the mean difference would be . . .

$$\bar{X}_1 - \bar{X}_2$$

- ◆ Where

\bar{X}_1 = mean weight loss in low carbohydrate diet group

\bar{X}_2 = mean weight loss in low fat diet group

95% Confidence Interval for Difference in Means

- ◆ Since $\bar{X}_1 - \bar{X}_2 = -5.7 - (-1.8) = -3.9$, our formula is . . .

$$-3.9 \pm 2 \text{ SE } (\bar{X}_1 - \bar{X}_2)$$

- ◆ $\text{SE } (\bar{X}_1 - \bar{X}_2)$ = standard error of the difference of 2 sample means

Two Independent (Unpaired) Groups

- ◆ The standard error of the difference for two independent samples is calculated differently than we did for paired designs
 - With paired design we reduced data on two samples to one set differences between two groups

Two Independent (Unpaired) Groups

- ◆ Statisticians have developed formulas for the standard error of the difference
 - These formulas depend on sample sizes in both groups and standard deviations in both groups

Two Independent (Unpaired) Groups

- ◆ The SE $(\bar{X}_1 - \bar{X}_2)$ is greater than either the SE (\bar{X}_1) or SE (\bar{X}_2)
 - Why do you think this is?

Two Independent (Unpaired) Groups

- ◆ In the example

$$SE(\bar{X}_1 - \bar{X}_2) = 1.15$$

$$SE(\bar{X}_1) = 1.08$$

$$SE(\bar{X}_2) = 0.47$$

Example

- ◆ 95% confidence interval for difference in mean
 - $-3.9 \pm 2SE(\bar{X}_1 - \bar{X}_2)$
 - $-3.9 \pm 2 * 1.15$
 - -3.9 ± 2.3

 - -6.2 kg to -1.6 kg

Note

- ◆ The confidence interval does not include 0

The SE of the Difference in Sample Means

- ◆ How to compute $SE(\bar{X}_1 - \bar{X}_2)$?

Principle

- ◆ Variation from independent sources can be added

$$\text{Variance } (\bar{X}_1 - \bar{X}_2) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

- Why do you think we add?

Principle

- ◆ Variation from independent sources can be added

$$SE(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Principle

- ◆ Formula depends on n_1, n_2, s_1, s_2
- ◆ There are other slightly different equations for $SE(\bar{X}_1 - \bar{X}_2)$
 - (e.g. Altman, p.192, see “pooled variance”)
- ◆ But they all **give similar answers**

Hypothesis Test to Compare Two Independent Groups

- ◆ Two sample (unpaired) t-test
- ◆ Is the weight loss equal in the two diet groups?
 - $H_0: \mu_1 = \mu_2$
 - $H_a: \mu_1 \neq \mu_2$

Hypothesis Test to Compare Two Independent Groups

- ◆ Two sample (unpaired) t-test
- ◆ In other words, is the expected difference in weight loss zero?
 - $H_0: \mu_1 - \mu_2 = 0$
 - $H_a: \mu_1 - \mu_2 \neq 0$

Hypothesis Test to Compare Two Independent Groups

- ◆ Recall, general “recipe” for hypothesis testing . . .
 1. State null and alternative hypotheses
 2. Proceed under assumption H_0 true—calculate test statistic based on sample
 3. Compare test statistic to appropriate distribution to get p-value

Two-Sample (Unpaired) t-test

- ◆ Formula for test statistic

$$t = \frac{(\text{observed diff}) - (\text{null diff})}{\text{SE of the difference}}$$

Two-Sample (Unpaired) t-test

- ◆ But since null difference is zero, this reduces to . . .

$$t = \frac{\text{(observed diff in sample means)}}{\text{SE of the difference}}$$

Two-Sample (Unpaired) t-test

- ◆ Which is just . . .

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}(\bar{X}_1 - \bar{X}_2)}$$

Two-Sample (Unpaired) t-test

- ◆ Which is just . . .

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}(\bar{X}_1 - \bar{X}_2)}$$

$$t = \frac{-5.7 - (-1.8)}{1.15} = \frac{-3.9}{1.15} = -3.4$$

Note

- ◆ This is a 2-sample (unpaired) t-test
 - The value $t = -3.4$ is the test statistic
- ◆ We calculate a p-value which is the probability of obtaining a test statistic as extreme as we did if H_0 was true

How Are p-values Calculated?

- ◆ Is a result 3.4 standard errors below 0 unusual?
 - It depends on what kind of distribution we are dealing with

How Are p-values Calculated?

- ◆ The p-value is the probability of getting a test statistic as or more extreme than what you observed (-3.4) by chance if H_0 was true
- ◆ The p-value comes from the sampling distribution of the difference in two sample means

Sampling Distribution

- ◆ What is the sampling distribution of the difference in sample means?
 - If both groups are large (more than 60 subjects) then this distribution is approximately normal

Diet/Weight Loss Sample

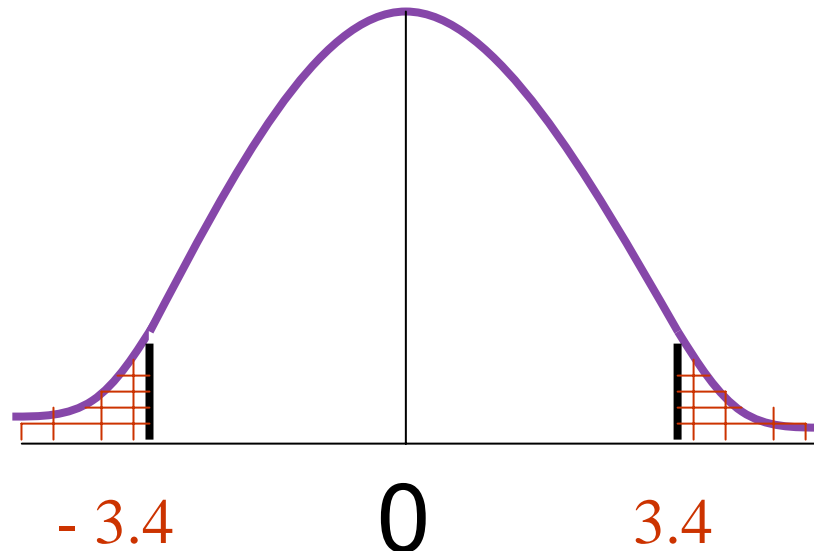
- ◆ So, since both our samples are large our sampling distribution will be approximately normal
 - This sampling distribution will be centered at true difference, $\mu_{\text{lowcarb}} - \mu_{\text{lowfat}}$
 - Under null hypothesis, this true difference is 0

Diet/Weight Loss Sample

- ◆ We observed a sample result that was 3.4 standard errors below zero
- ◆ Is a result 3.4 standard errors below 0 unusual?
 - We can compute this from sampling distribution

Diet/Weight Loss Sample

- ◆ To compute a p-value, we would need to compute the probability of being 3.4 or more standard errors away from 0 on a standard normal curve



How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

How to Use STATA to Perform a 2-Sample T-Test

- ◆ Command syntax:

- `ttesti` n_1 \bar{X}_1 s_1 n_2 \bar{X}_2 s_2

Stata Results

```
. ttesti 64 -5.7 8.6 68 -1.8 3.9
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	64	-5.7	1.075	8.6	-7.848216	-3.551784
y	68	-1.8	.4729445	3.9	-2.744001	-.8559989
combined	132	-3.690909	.5978226	6.868458	-4.873545	-2.508273
diff		-3.9	1.151038		-6.177191	-1.622809

Degrees of freedom: 130

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -3.3882
P < t = 0.0005

Ha: diff != 0
t = -3.3882
P > |t| = 0.0009

Ha: diff > 0
t = -3.3882
P > t = 0.9995

Stata Results

```
. ttesti 64 -5.7 8.6 68 -1.8 3.9
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	64	-5.7	1.075	8.6	-7.848216	-3.551784
y	68	-1.8	.4729445	3.9	-2.744001	-.8559989
combined	132	-3.690909	.5978226	6.868458	-4.873545	-2.508273
diff		-3.9	1.151038		-6.177191	-1.622809

Degrees of freedom: 130

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -3.3882
P < t = 0.0005

Ha: diff != 0
t = -3.3882
P > |t| = 0.0009

Ha: diff > 0
t = -3.3882
P > t = 0.9995

Stata Results

```
. ttesti 64 -5.7 8.6 68 -1.8 3.9
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	64	-5.7	1.075	8.6	-7.848216	-3.551784
y	68	-1.8	.4729445	3.9	-2.744001	-.8559989
combined	132	-3.690909	.5978226	6.868458	-4.873545	-2.508273
diff		-3.9	1.151038		-6.177191	-1.622809

Degrees of freedom: 130

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -3.3882
P < t = 0.0005

Ha: diff != 0
t = -3.3882
P > |t| = 0.0009

Ha: diff > 0
t = -3.3882
P > t = 0.9995

Summary: Weight Loss Example

◆ Statistical Method

- “We randomly assigned 132 severely obese patients . . . to a carbohydrate restricted (low-carbohydrate) diet or a calorie- and fat – restricted diet”

Summary: Weight Loss Example

◆ Statistical Method

- “For comparison of continuous variables between the two groups, we calculated the change from baseline to six months in each subject, and compared the mean changes in the two diet groups using an unpaired t-test”

Summary: Weight Loss Example

◆ Result

- “Subjects on the low-carbohydrate diet lost more weight than those on a low fat diet (95% confidence interval for the difference in weight loss between groups, -1.6 to -6.2 kg; $p < .01$)”



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section A

Practice Problems

Practice Problems

- ◆ In a high school in the United States, a dietary counseling program is being tested to measure the program's long-term impact on student's fat intake
- ◆ Of the three hundred students at the school, 150 are randomized to receive five one-hour sessions of dietary counseling
- ◆ The other 150 students receive no counseling

Practice Problems

- ◆ Six months after the last counseling sessions, all students are asked to keep a food diary for one week
- ◆ Each student's average fat intake in grams, is calculated at the end of this week
- ◆ The results of this exercise are as follows:

Practice Problems

Intervention Group

- ◆ $\bar{X}_1 = 54.8$ grams , $s_1 = 28.1$ grams, $n_1 = 146$

Control Group

- ◆ $\bar{X}_2 = 62.8$ grams, $s_2 = 34.7$ grams, $n_2 = 142$
- ◆ (Please note—follow up sample sizes differ slightly from initial sample size because of loss to follow up)

Practice Problems

- ◆ The public-health question of interest is whether there is a difference in mean fat intake between the two groups, six months after the intervention ended
- ◆ You are going to help answer this question

Practice Problems

1. Construct a 95% CI for the population mean difference in daily fat intake for the control group as compared to the intervention group.
2. What is the appropriate statistical test for this question? State the null and alternative hypotheses, and obtain a p-value.

Practice Problems

- ◆ You of course are free, and in fact encouraged to use Stata to help with this exercise!



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section A

Practice Problem Solutions

Practice Problems

- ◆ In a high school in the United States, a dietary counseling program is being tested to measure the program's long-term impact on student's fat intake
- ◆ Of the three hundred students at the school, 150 are randomized to receive five one-hour sessions of dietary counseling
- ◆ The other 150 students receive no counseling

Practice Problems

- ◆ Six months after the last counseling sessions, all students are asked to keep a food diary for one week
- ◆ Each student's average fat intake in grams, is calculated at the end of this week
- ◆ The results of this exercise are as follows:

Practice Problems

Intervention Group

- ◆ $\bar{X}_1 = 54.8$ grams , $s_1 = 28.1$ grams, $n_1 = 146$

Control Group

- ◆ $\bar{X}_2 = 62.8$ grams, $s_2 = 34.7$ grams, $n_2 = 142$
- ◆ (Please note—follow up sample sizes differ slightly from initial sample size because of loss to follow up)

Practice Problems

- ◆ The public-health question of interest is whether there is a difference in mean fat intake between the two groups, six months after the intervention ended
- ◆ You are going to help answer this question:

Practice Problems

1. Construct a 95% CI for the population mean difference in daily fat intake for the control group as compared to the intervention group

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

$$(\bar{X}_2 - \bar{X}_1) \pm 2 * SE (\bar{X}_2 - \bar{X}_1)$$

Business as usual!!

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

$$SE(\bar{X}_2 - \bar{X}_1) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

$$SE(\bar{X}_2 - \bar{X}_1) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= \sqrt{\frac{(28.1)^2}{146} + \frac{(34.7)^2}{142}}$$

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

$$\begin{aligned} \text{SE}(\bar{X}_2 - \bar{X}_1) &= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ &= \sqrt{\frac{(28.1)^2}{146} + \frac{(34.7)^2}{142}} = 3.7 \end{aligned}$$

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

◆ Now, $\bar{X}_2 - \bar{X}_1 = 62.8 - 54.8 = 8$

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

$$8 \pm 2^*(3.7)$$

$$8 \pm 7.4$$

$$(0.6\text{gm}, 15.4 \text{ gm})$$

Practice Problems

1. 95% CI for $\mu_2 - \mu_1$:

(0.6, 15.4)

- ◆ Notice that 0 is not included in the 95% CI- therefore we would reject the null from the t-test at the $\alpha = .05$ level.

Practice Problems

2. What is the appropriate statistical test for this question? State the null and alternative hypotheses.

t-test for two independent populations

$$H_0: \mu_2 - \mu_1 = 0$$

$$H_a: \mu_2 - \mu_1 \neq 0$$

Practice Problems

```
. ttesti 142 62.8 34.7 146 54.8 28.1
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	142	62.8	2.911959	34.7	57.04326	68.55674
y	146	54.8	2.325573	28.1	50.2036	59.3964
combined	288	58.74444	1.869474	31.72603	55.06483	62.42406
diff		8	3.715851		.6861165	15.31388

Degrees of freedom: 286

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
 t = 2.1529
 P < t = 0.9839

Ha: diff != 0
 t = 2.1529
 P > |t| = 0.0322

Ha: diff > 0
 t = 2.1529
 P > t = 0.0161

Practice Problems

```
. ttesti 146 54.8 28.1 142 62.8 34.7
```

Two-sample t test with equal variances

```
-----+-----  
          |      Obs      Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]  
-----+-----  
      x   |      146      54.8    2.325573     28.1     50.2036     59.3964  
      y   |      142      62.8    2.911959     34.7     57.04326    68.55674  
-----+-----  
combined |      288    58.74444    1.869474    31.72603    55.06483    62.42406  
-----+-----  
      diff |              -8     3.715851    -15.31388    -.6861165  
-----+-----  
Degrees of freedom: 286
```

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -2.1529
P < t = 0.0161

Ha: diff != 0
t = -2.1529
P > |t| = 0.0322

Ha: diff > 0
t = -2.1529
P > t = 0.9839



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section B

Two Sample t-test, Part 2

Sampling Distribution

- ◆ What is sampling distribution of the difference in sample means?
 - If either (or both) sample sizes are less than 60, a t distribution is used with $n_1 + n_2 - 2$ degrees of freedom—that is the degrees of freedom is the total sample size from both groups minus two

Two Sample t-test

- ◆ Example

- In a randomized design, 23 patients with hyperlipidemia were randomized to either take Treatment A or Treatment B for 12 weeks

Two Sample t-test

- ◆ Example
 - 12 patients assigned to Treatment A
 - 11 patients assigned to Treatment B

Two Sample t-test

- ◆ Example
 - LDL cholesterol levels (mmol/L) measured on each subject at baseline, and 12 weeks after start of study
 - The 12-week change in LDL cholesterol was computed for each subject

Two Sample t-test

- ◆ Summary of results:

	Treatment Group	
	A	B
Number of Subjects	12	11
Mean LDL Change (mmol/L) Post-treatment less pre-treatment	-1.41	-0.32
Standard Deviation of LDL Change (mmol/L)	0.55	0.65

Two Sample t-test

- ◆ Scientific Question
 - Is there a difference in LDL change between the two treatment groups?

Methods of Inference

- ◆ Confidence interval for the difference in mean LDL cholesterol will change between the two groups
- ◆ Statistical hypothesis test

95% Confidence Interval for Difference in Means

- ◆ Our most general formula (large samples):

(our best estimate) $\pm 2^*(\text{SE of our best estimate})$

95% Confidence Interval for Difference in Means

- ◆ Our most general formula (smaller samples):

(our best estimate) \pm t^* (SE of our best estimate)

95% Confidence Interval for Difference in Means

- ◆ Well, our best estimate for the mean difference would be:

$$\bar{X}_1 - \bar{X}_2$$

- ◆ Where . . .
- ◆ \bar{X}_1 = mean LDL change in Treatment A group
- ◆ \bar{X}_2 = mean LDL change in Treatment B group

95% Confidence Interval for Difference in Means

- ◆ Since $\bar{X}_1 - \bar{X}_2 = -1.41 - (-.32) = -1.09$, our formula is . . .

$$-1.09 \pm t^* \text{ SE } (\bar{X}_1 - \bar{X}_2)$$

- ◆ $\text{SE } (\bar{X}_1 - \bar{X}_2) = 0.25$

95% Confidence Interval for Difference in Means

- ◆ How many standard errors to add and subtract?
 - Since sample sizes small we will have to add slightly more than two standard errors
- ◆ Number we need add and subtract for 95% confidence comes from a t-distribution with $(12 + 11 - 2 =)$ 21 degrees of freedom

95% Confidence Interval for Difference in Means

- ◆ Since $\bar{X}_1 - \bar{X}_2 = -1.41 - (-.46) = -.95$, our formula is . . .

$$-1.09 \pm 2.08^* \text{ SE } (\bar{X}_1 - \bar{X}_2)$$

$$-1.09 \pm 2.08^* .25$$

$$(-1.61, -0.57)$$

Hypothesis Test to Compare Two Independent Groups

- ◆ Two-Sample (Unpaired) t-test
- ◆ Is the change in LDL cholesterol the same in the two treatment groups?
 - $H_0: \mu_1 = \mu_2$
 - $H_a: \mu_1 \neq \mu_2$

Hypothesis Test to Compare Two Independent Groups

- ◆ Two-Sample (Unpaired) t-test
- ◆ In other words, is the expected difference in LDL change zero?
 - $H_0: \mu_1 - \mu_2 = 0$
 - $H_a: \mu_1 - \mu_2 \neq 0$

Hypothesis Test to Compare Two Independent Groups

- ◆ Recall, general “recipe” for hypothesis testing:
 1. State null and alternative hypotheses
 2. Proceed under assumption H_0 is true—calculate test statistic based on sample
 3. Compare test statistic to appropriate distribution to get p-value

Two-Sample (Unpaired) t-test

- ◆ Formula for test statistic

$$t = \frac{(\text{observed diff}) - (\text{null diff})}{\text{SE of the difference}}$$

Two-Sample (Unpaired) t-test

- ◆ But since null difference is zero, this reduces to . . .

$$t = \frac{\text{(observed diff in sample means)}}{\text{SE of the difference}}$$

Two-Sample (Unpaired) t-test

- ◆ Which is just . . .

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}(\bar{X}_1 - \bar{X}_2)}$$

Two-Sample (Unpaired) t-test

- ◆ Which is just . . .

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}(\bar{X}_1 - \bar{X}_2)}$$

$$t = \frac{-1.41 - (-0.32)}{0.25} = \frac{-1.09}{0.25} = -4.4$$

Note

- ◆ This is a two sample (unpaired) t-test
 - The value $t = -4.4$ is the test statistic
- ◆ We calculate a p-value which is the probability of obtaining a test statistic as extreme as we did if H_0 was true

How Are p-values Calculated?

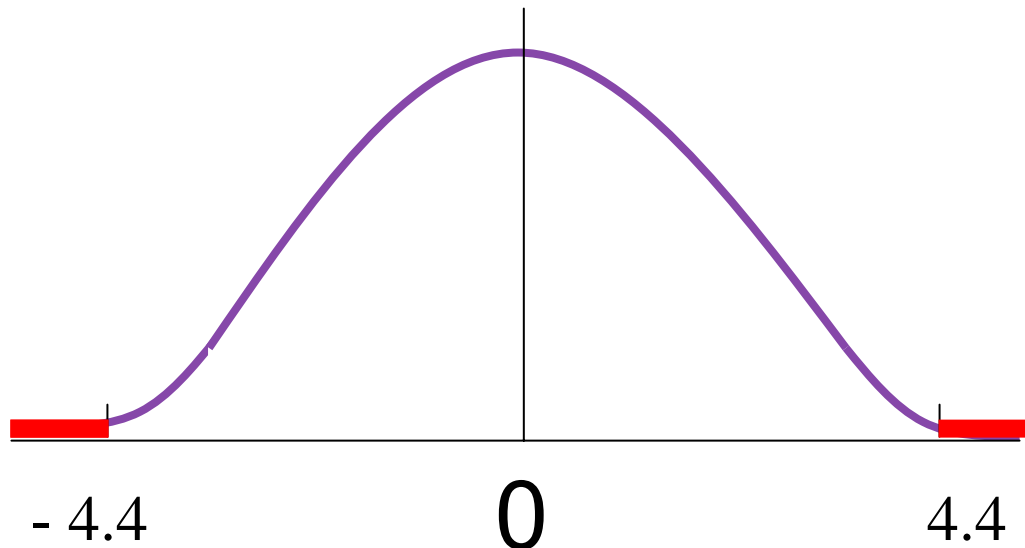
- ◆ Is a result 4.4 standard errors below 0 unusual?
 - It depends on what kind of distribution we are dealing with

How Are p-values Calculated?

- ◆ The p-value is the probability of getting a test statistic as or more extreme than what you observed (-4.4) by chance if H_0 was true
- ◆ The p-value comes from the sampling distribution of the *difference in two sample means*

Hyperlipidemia Example

- ◆ To compute a p-value, we would need to compute the probability of being 4.4 or more standard errors away from 0 on a t-distribution with 21 degrees of freedom



Using Stata

```
. ttesti 12 -1.41 .55 11 -.32 .65
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	12	-1.41	.1587713	.55	-1.759453	-1.060547
y	11	-.32	.1959824	.65	-.7566759	.1166759
combined	23	-.8886957	.1685275	.8082293	-1.2382	-.5391911
diff		-1.09	.2503299		-1.610589	-.5694105

Degrees of freedom: 21

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -4.3543
P < t = 0.0001

Ha: diff != 0
t = -4.3543
P > |t| = 0.0003

Ha: diff > 0
t = -4.3543
P > t = 0.9999

Using Stata

```
. ttesti 12 -1.41 .55 11 -.32 .65
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	12	-1.41	.1587713	.55	-1.759453	-1.060547
y	11	-.32	.1959824	.65	-.7566759	.1166759
combined	23	-.8886957	.1685275	.8082293	-1.2382	-.5391911
diff		-1.09	.2503299		-1.610589	-.5694105

Degrees of freedom: 21

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -4.3543
P < t = 0.0001

Ha: diff != 0
t = -4.3543
P > |t| = 0.0003

Ha: diff > 0
t = -4.3543
P > t = 0.9999

Using Stata

```
. ttesti 12 -1.41 .55 11 -.32 .65
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	12	-1.41	.1587713	.55	-1.759453	-1.060547
y	11	-.32	.1959824	.65	-.7566759	.1166759
combined	23	-.8886957	.1685275	.8082293	-1.2382	-.5391911
diff		-1.09	.2503299		-1.610589	-.5694105

Degrees of freedom: 21

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -4.3543
P < t = 0.0001

Ha: diff != 0
t = -4.3543
P > |t| = 0.0003

Ha: diff > 0
t = -4.3543
P > t = 0.9999

Summary: Weight Loss Example

◆ Statistical Method

- Twenty-three patients with hyperlipidemia were randomly assigned to one of two treatment groups: Treatment A or Treatment B
- 12 patients were assigned to receive Treatment A
- 11 patients were assigned to receive Treatment B

Summary: Weight Loss Example

◆ Statistical Method

- Baseline LDL cholesterol measurements were taken on each subject, and LDL was again measured after 12 weeks of treatment
- The change in LDL cholesterol was computed for each subject

Summary: Weight Loss Example

◆ Statistical Method

- The mean LDL changes in the two treatment groups were compared using an unpaired t-test and a 95% confidence interval was constructed for the difference in mean LDL changes

Summary: Weight Loss Example

◆ Result

- Patients on treatment A showed a decrease in LDL cholesterol of 1.41 mmol/L and subjects on treatment B showed a decrease of .32 mmol/L (a difference of 1.09 mmol/L, 95% CI .57 to 1.61 mmol/L)
- The difference in LDL changes was statistically significant ($p < .001$)

Using Stata

```
. ttesti 12 -1.41 .55 11 -.32 .65
```

```
Two-sample t test with equal variances
```

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	12	-1.41	.1587713	.55	-1.759453	-1.060547
y	11	-.32	.1959824	.65	-.7566759	.1166759
combined	23	-.8886957	.1685275	.8082293	-1.2382	-.5391911
diff		-1.09	.2503299		-1.610589	-.5694105

Degrees of freedom: 21

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -4.3543
P < t = 0.0001

Ha: diff != 0
t = -4.3543
P > |t| = 0.0003

Ha: diff > 0
t = -4.3543
P > t = 0.9999

FYI: Equal Variances Assumption

- ◆ The “traditional” t-test assumes equal variances in the two groups
 - This can be formally tested with another hypothesis test?!?!
- ◆ There is a slight modification to allow for unequal variances—this modification adjusts the degrees of freedom for the test, using slightly different SE computation

Equal Variances Assumption

- ◆ If you want to be truly “safe”
 - More conservative to use test that allows for unequal variance
- ◆ Makes little to no difference in large samples

Unequal Variances t-test in Stata

```
. ttesti 12 -1.41 .55 11 -.32 .65, unequal
```

Two-sample t test with unequal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	12	-1.41	.1587713	.55	-1.759453	-1.060547
y	11	-.32	.1959824	.65	-.7566759	.1166759
combined	23	-.8886957	.1685275	.8082293	-1.2382	-.5391911
diff		-1.09	.2522249		-1.616622	-.5633782

Satterthwaite's degrees of freedom: 19.714

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -4.3215
P < t = 0.0002

Ha: diff != 0
t = -4.3215
P > |t| = 0.0003

Ha: diff > 0
t = -4.3215
P > t = 0.9998.

Unequal Variances t-test in Stata

```
. ttesti 12 -1.41 .55 11 -.32 .65, unequal
```

Two-sample t test with unequal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	12	-1.41	.1587713	.55	-1.759453	-1.060547
y	11	-.32	.1959824	.65	-.7566759	.1166759
combined	23	-.8886957	.1685275	.8082293	-1.2382	-.5391911
diff		-1.09	.2522249		-1.616622	-.5633782

Satterthwaite's degrees of freedom: 19.714

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = -4.3215
P < t = 0.0002

Ha: diff != 0
t = -4.3215
P > |t| = 0.0003

Ha: diff > 0
t = -4.3215
P > t = 0.9998.

Summary: Two-Sample Unpaired t-test

- ◆ The unpaired t-test is a useful statistical tool for comparing mean differences between two independent populations based on data from two independent samples

Summary: Two-Sample Unpaired t-test

- ◆ Designate null and alternative hypotheses
- ◆ Collect data

Summary: Two-Sample Unpaired t-test

- ◆ Computations
 - Compute \bar{X}_1, \bar{X}_2 sample mean outcome for each of the two samples
 - Compute s_1, s_2 sample standard deviation in each of the two samples

Summary: Two-Sample Unpaired t-test

- ◆ Computations
 - Compute standard error of the difference in sample means

Summary: Two-Sample Unpaired t-test

- ◆ Computations
 - Test statistic

$$t = \frac{\text{Difference in sample means}}{\text{SE}(\bar{X}_1 - \bar{X}_2)}$$

Summary: Two-Sample Unpaired t-test

- ◆ Compare test statistic to appropriate distribution to get p-value

Computing 95% CI For True Mean Difference

- ◆ General recipe . . .

$$\bar{X}_1 - \bar{X}_2 \pm t^* \text{SE}(\bar{X}_1 - \bar{X}_2)$$

- $t = 2$ if $n_1, n_2 \geq 60$
- t slightly bigger than 2 if n_1 and/or $n_2 < 60$

Summary: Two-Sample Unpaired t-test

- ◆ If the samples we are comparing are very small (< 10 observations each), the t-test works well if our samples come from normally distributed populations (sometimes a big, hard to check assumption)
- ◆ Alternative—non-parametric tests



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section B

Practice Problems

Practice Problems

- ◆ A small study was performed to compare carboxyhemoglobin levels between smokers and non-smokers
- ◆ A random sample of 40 was taken, 13 who were smokers, 27 who were non-smokers

Practice Problems

- ◆ Carboxyhemoglobin measurements (measured in percentage of total hemoglobin) were taken on each of the subjects
- ◆ Summary statistics for the two groups of subjects are presented on the next slide

Practice Problems

Non-Smokers

$$\bar{X}_1 = 1.3 \text{ percent}, s_1 = 1.3 \text{ percent}, n_1 = 27$$

Smokers

$$\bar{X}_2 = 4.1 \text{ percent}, s_2 = 2.0 \text{ percent}, n_2 = 13$$

Practice Problems

1. Construct a 95% CI for the mean difference in carboxyhemoglobin levels for the smokers as compared to the non-smokers

Practice Problems

2. What is the appropriate statistical test for this question? State the null and alternative hypotheses, and obtain a p-value.

Practice Problems

- ◆ You of course are free, and in fact encouraged to use Stata to help with this exercise!



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section B

Practice Problem Solutions

Practice Problems

- ◆ A small study was performed to compare carboxyhemoglobin levels between smokers and non-smokers
- ◆ A random sample of 40 was taken, 13 who were smokers, 27 who were non-smokers

Practice Problems

- ◆ Carboxyhemoglobin measurements (measured in percentage of total hemoglobin) were taken on each of the subjects
- ◆ Summary statistics for the two groups of subjects are presented on the next slide

Practice Problems

Non-Smokers Group

- ◆ $\bar{X}_1 = 1.3$ percent, $s_1 = 1.3$ percent, $n_1 = 27$

Control Group

- ◆ $\bar{X}_2 = 4.1$ percent, $s_2 = 2.0$ percent, $n_2 = 13$

Practice Problems

1. Construct a 95% CI for the mean difference in carboxyhemoglobin levels for the smokers as compared to the non-smokers

Practice Problems

```
. ttesti 13 4.1 2 27 1.3 1.3
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	13	4.1	.5547002	2	2.891412	5.308588
y	27	1.3	.2501851	1.3	.7857371	1.814263
combined	40	2.21	.3209921	2.030132	1.560732	2.859268
diff		2.8	.5250938		1.737003	3.862997

Degrees of freedom: 38

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
 t = 5.3324
 P < t = 1.0000

Ha: diff != 0
 t = 5.3324
 P > |t| = 0.0000

Ha: diff > 0
 t = 5.3324
 P > t = 0.0000

Practice Problems

```
. ttesti 13 4.1 2 27 1.3 1.3, unequal
```

Two-sample t test with unequal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
x	13	4.1	.5547002	2	2.891412 5.308588
y	27	1.3	.2501851	1.3	.7857371 1.814263
combined	40	2.21	.3209921	2.030132	1.560732 2.859268
diff		2.8	.6085104		1.51646 4.08354

Satterthwaite's degrees of freedom: 17.0531

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = 4.6014
P < t = 0.9999

Ha: diff != 0
t = 4.6014
P > |t| = 0.0003

Ha: diff > 0
t = 4.6014
P > t = 0.0001

Practice Problems

2. What is the appropriate statistical test for this question?
 - ◆ State the null and alternative hypotheses, and obtain a p-value

Practice Problems

```
. ttesti 13 4.1 2 27 1.3 1.3
```

Two-sample t test with equal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
x	13	4.1	.5547002	2	2.891412	5.308588
y	27	1.3	.2501851	1.3	.7857371	1.814263
combined	40	2.21	.3209921	2.030132	1.560732	2.859268
diff		2.8	.5250938		1.737003	3.862997

Degrees of freedom: 38

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = 5.3324
P < t = 1.0000

Ha: diff != 0
t = 5.3324
P > |t| = 0.0000

Ha: diff > 0
t = 5.3324
P > t = 0.0000

Practice Problems

```
. ttesti 13 4.1 2 27 1.3 1.3, unequal
```

Two-sample t test with unequal variances

	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
x	13	4.1	.5547002	2	2.891412 5.308588
y	27	1.3	.2501851	1.3	.7857371 1.814263
combined	40	2.21	.3209921	2.030132	1.560732 2.859268
diff		2.8	.6085104		1.51646 4.08354

Satterthwaite's degrees of freedom: 17.0531

Ho: mean(x) - mean(y) = diff = 0

Ha: diff < 0
t = 4.6014
P < t = 0.9999

Ha: diff != 0
t = 4.6014
P > |t| = 0.0003

Ha: diff > 0
t = 4.6014
P > t = 0.0001



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section C

*Non-Parametric Analogue to the
Two Sample t -test*

Alternative to the Two Sample T-Test

- ◆ Nonparametric test for comparing two groups
 - “Non-parametric” refers to a class of tests that do not assume anything about distribution of the data

Alternative to the Two Sample T-Test

- ◆ Nonparametric test for comparing two groups
 - Mann-Whitney Rank Sum Test (Wilcoxon Rank Sum Test)
 - Also called Mann-Whitney-Wilcoxon (a mouthful)

Objective

- ◆ Tries to answer the question:
 - Are the two populations different?

Advantages

- ◆ Does not assume populations being compared are normally distributed
 - The two-sample t-test requires that assumption with small samples sizes
- ◆ Uses only ranks
- ◆ Not sensitive to outliers

Disadvantage of the Nonparametric Test

- ◆ Nonparametric methods are often less sensitive (powerful) for finding true differences because they throw away information (they use only ranks)
- ◆ Need full data set, not just summary statistics

Example:

Health Education Study

- ◆ Evaluate an intervention to educate high school students about health and lifestyle over a two-month period
- ◆ 10 students randomized to “intervention” or “control” group

Y = Post- Pretest Score is
outcome of interest

Example: Health Education Study

- ◆ Y = post- pretest score for both groups
- ◆ Intervention (I) 5 0 7 2 19
- ◆ Control (C) 6 -5 -6 1 4

Only five individuals in each sample!!!!!!

Example:

Health Education Study

- ◆ We want to compare the control and intervention groups to assess whether the “improvement” (post – pre) in scores are different

Example:

Health Education Study

- ◆ With such a small sample size, we need to be sure score improvements are normally distributed if we want to use t-test (BIG assumption)

Example: Health Education Study

- ◆ Solution:
 - Mann-Whitney-Wilcoxon non-parametric test!

Example: Health Education Study

- ◆ First step—rank the pooled data (ignore groupings)

	-6	-5	0	1	2	4	5	5	7	19
Rank	1	2	3	4	5	6	7	8	9	10

Example: Health Education Study

- ◆ Second step—“reattach” group status

	-6	-5	0	1	2	4	5	5	7	19
Rank	1	2	3	4	5	6	7	8	9	10
Group	C	C	I	C	I	C	I	C	I	I

Example:

Health Education Study

- ◆ Find the average rank in each of the two groups
- ◆ Intervention group average rank

$$\frac{3+5+6+9+10}{5} = 6.8$$

Example: Health Education Study

- ◆ Control group average rank

$$\frac{1 + 2 + 4 + 6 + 8}{5} = 4.2$$

Example:

Health Education Study

- ◆ Statisticians have developed formulas and tables to determine the probability of observing such an extreme discrepancy in ranks (6.8 vs 4.2) by chance alone
 - That's the p-value

Notes

- ◆ In the health education study, the p-value was .17
 - The interpretation is that the Mann-Whitney test did not show any significant difference in test score “improvement” between the intervention and control group ($p = .17$)

Notes

- ◆ The two-sample t-test would give a different answer ($p = .14$)
- ◆ Different statistical procedures can give different p-values

Notes

- ◆ If the largest observation, 19, was changed, the p-value based on the Mann-Whitney test would not change but the two-sample t-test would change
- ◆ For purposes of this class, test recognition is important, not a method for obtaining p-value!

Notes

- ◆ The t-test or the nonparametric test?
 - Statisticians will not always agree, but there are some guidelines

Notes

- ◆ The t-test or the nonparametric test?
 - Use nonparametric test if sample size is small and you have no reason to believe data is “well behaved” (normally distributed)
 - Only “ranks” available

Using Stata to Perform Mann-Whitney-Wilcoxon

- ◆ Data, as entered

```
. list
```

```
      +-----+
      | diff   group |
      +-----+
  1.  |     5     1 |
  2.  |     0     1 |
  3.  |     7     1 |
  4.  |     2     1 |
  5.  |    19     1 |
      +-----+
  6.  |     6     0 |
  7.  |    -5     0 |
  8.  |    -6     0 |
  9.  |     1     0 |
 10.  |     4     0 |
      +-----+
```

Using Stata to Perform Mann-Whitney-Wilcoxon

- ◆ “ranksum” command
 - Syntax:
 - ranksum *varname*, by(*group_var*)

Using Stata to Perform Mann-Whitney-Wilcoxon

```
. ranksum diff,by(group)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

group	obs	rank sum	expected
0	5	21	27.5
1	5	34	27.5
combined	10	55	55

```
unadjusted variance      22.92
```

```
adjustment for ties      0.00
```

```
-----  
adjusted variance      22.92
```

```
Ho: diff(group==0) = diff(group==1)
```

```
z = -1.358
```

```
Prob > |z| = 0.1745
```

Using Stata to Perform Mann-Whitney-Wilcoxon

```
. ranksum diff,by(group)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

group	obs	rank sum	expected
0	5	21	27.5
1	5	34	27.5
combined	10	55	55

unadjusted variance 22.92

adjustment for ties 0.00

adjusted variance 22.92

Ho: diff(group==0) = diff(group==1)

z = -1.358

Prob > |z| = 0.1745

Summary: Educational Intervention Example

◆ Statistical Methods

- 10 high school students were randomized to either receive a two-month health and lifestyle education program (or no program)
- Each student was administered a test regarding health and lifestyle issues prior to randomization (and after the two-month period)

Summary: Educational Intervention Example

◆ Statistical Methods

- Differences in the two test scores (after-before) were computed for each student
- Mean and median test score changes were computed for each of the two study groups

Summary: Educational Intervention Example

◆ Statistical Methods

- A Mann-Whitney rank sum test was used to determine if there was a statistically significant difference in test score change between the intervention and control groups at the end of the two-month study period

Summary: Educational Intervention Example

◆ Result

- Participants randomized to the educational intervention scored a median five points higher on the test given at the end of the two-month study period, as compared to the test administered prior to the intervention

Summary: Educational Intervention Example

◆ Result

- Participants randomized to receive no educational intervention scored a median one point higher on the test given at the end of the two-month study period

Summary: Educational Intervention Example

◆ Result

- The difference in test score improvements between the intervention and control groups was not statistically significant ($p = .17$)



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section C

Practice Problems

Practice Problems

- ◆ A health policy researcher is doing an ad-hoc study of gender differences in attitudes about medical confidentiality
- ◆ He spends some time at a shopping mall and polls and asks individuals to rate their degree of agreement for a statement related to confidentiality using the following five-point scale:

Practice Problems

- ◆ Strongly Disagree (1)
- ◆ Disagree (2)
- ◆ Neutral (3)
- ◆ Agree (4)
- ◆ Strongly Agree (5)

Practice Problems

- ◆ Responses are coded from 1 to 5. At the end of his tenure at the mall, this researcher has a total of 11 respondents:
 - 6 male, 5 female

- ◆ The distribution of responses:

◆ M	1	3	4	2	3	5
◆ F	3	4	4	3	5	

Practice Problems

- ◆ Despite the small sample size, the researcher is interested in testing for a difference (i.e.: obtaining a p-value) in degree of agreement for males compared to females
- ◆ How could he do this?
- ◆ Compute the average rank for each gender group



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section C

Practice Problem Solutions

Solutions

◆ The final dataset was as follows:

◆ M	1	3	4	2	3	5
◆ F	3	4	4	3	5	

Solutions

- ◆ Despite the small sample size, the researchers is interested in testing for a difference in degree of agreement for males compared to females
- ◆ How could he do this?

Solutions

1. How could he do this?
 - ◆ In this situation, the outcome of interest is an ordinal variable—it cannot be assumed to have a normal distribution
 - ◆ Additionally, the sample size is small—must appeal to non-parametric Mann-Whitney-Wilcoxon

Solutions

2. Compute the average rank for each group
 - ◆ First, we must arrange the data from lowest to highest without regard for group

1 2 3 3 3 3 4 4 4 5 5

Solutions

2. Compute the average rank for each group
 - ◆ Now, assign ranks to the data

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	3	?	?	?					

Solutions

2. Compute the average rank for each group
 - ◆ What to do about the ties?

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	3	?	?	?					

Solutions

2. Compute the average rank for each group
 - ◆ With ties, we take the “average rank” of the group with ties, and assign it to each member of the group

Solutions

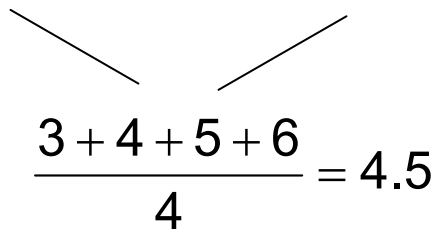
2. Compute the average rank for each group
 - ◆ For example, with the series of 3's we first assign them ascending ranks as if they were increasing

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	3	4	5	6					

Solutions

2. Compute the average rank for each group
 - ◆ Then we average these ascending ranks

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	3	4	5	6					


$$\frac{3 + 4 + 5 + 6}{4} = 4.5$$

Solutions

2. Compute the average rank for each group
 - ◆ Now reassign each "3" the average rank, 4.5

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	4.5	4.5	4.5	4.5					

Solutions

2. Compute the average rank for each group
 - ◆ Rank the rest of the values, treating ties the same way

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	4.5	4.5	4.5	4.5	7	8	9		

$$\frac{7 + 8 + 9}{3} = 8$$

Solutions

2. Compute the average rank for each group.
 - ◆ Rank the rest of the values, treating ties the same way

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	4.5	4.5	4.5	4.5	8	8	8	10.5	10.5

Solutions

2. Compute the average rank for each group
 - ◆ Reassign the group labels to the data . . .

	1	2	3	3	3	3	4	4	4	5	5
Rank	1	2	4.5	4.5	4.5	4.5	8	8	8	10.5	10.5
Group	M	M	F	M	F	M	F	F	M	F	M

Solutions

2. Compute the average rank for each group
 - ◆ Now we can compute group average ranks!
 - ◆ For males:

$$\bar{R}_{\text{males}} = \frac{1 + 2 + 4.5 + 4.5 + 8 + 10.5}{6} = 5.1$$

Solutions

2. Compute the average rank for each group
 - ◆ Now we can compute group average ranks!
 - ◆ For females:

$$\bar{R}_{\text{females}} = \frac{4.5 + 4.5 + 8 + 8 + 10.5}{5} = 7.1$$

Solutions

2. The Mann-Whitney-Wilcoxon compares the average ranks for males compared to females and tests:
 - ◆ H_0 : Distribution of Degree of Agreement is Same for M and F

Solutions

2. The p-values for this test is .29, suggesting that “degree of agreement” is not statistically different for males and females



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section D

*Comparing Means Between More than
Two Independent Populations*

Motivating Example

- ◆ Suppose you are interested in the relationship between smoking and mid-expiratory flow (FEF), a measure of pulmonary health

Motivating Example

- ◆ Suppose you recruit study subjects and classify them into one of six smoking categories

Motivating Example

- ◆ Nonsmokers (NS)
- ◆ Passive smokers (PS)
- ◆ Non-inhaling smokers (NI)
- ◆ Light smokers (LS)
- ◆ Moderate smokers (MS)
- ◆ Heavy smokers (HS)

Motivating Example

- ◆ You are interested in whether differences exist in mean FEF amongst the six groups

Motivating Example

- ◆ Main outcome variable is mid-expiratory flow (FEF) in liters per second (amount of air expelled per second)

Motivating Example

- ◆ One strategy is to perform lots of two-sample t-tests (for each possible two-group comparison)
- ◆ In this example, there would be 15 comparisons you would need to do!
 - NS to PS, NS to NI, and so on . . .

Motivating Example

- ◆ It would be nice to have one “catch-all” test
 - Something which would tell you whether there were *any* differences amongst the six groups
 - If so, you could then do group to group comparisons to look for specific group differences

Extension of the Two-Sample t-Test

- ◆ Analysis of Variance (One-Way ANOVA)
 - The t-test compares means in two populations
 - ANOVA compares means amongst more than two populations with one test

Extension of the Two-Sample t-Test

- ◆ The p-value from ANOVA helps answer the question
 - “Are there *any* differences in the means among the populations?”

Extension of the Two-Sample t-Test

- ◆ General idea behind ANOVA, comparing means for k-groups ($k > 2$):

$$H_0: \mu_1 = \mu_2 = \dots \mu_k$$

H_a : At least one mean different

Example

- ◆ Smoking and FEF¹
 - A sample of over 3,000 persons was classified into one of six smoking categorizations based on responses to smoking related questions

1 White, J.R., Froeb, H.F. Small-Airways Dysfunction in Non-Smokers Chronically Exposed to Tobacco Smoke, New England Journal of Medicine 348: 21

Example

- ◆ Nonsmokers (NS)
- ◆ Passive smokers (PS)
- ◆ Non-inhaling smokers (NI)
- ◆ Light smokers (LS)
- ◆ Moderate smokers (MS)
- ◆ Heavy smokers (HS)

Example

- ◆ Smoking and FEF
 - From each smoking group, a random sample of 200 men was drawn (except for the non-inhalers, as there were only 50 male non-inhalers in the entire sample of 3,000)

Example

- ◆ Smoking and FEF
 - FEF measurements were taken on each of the subjects

Example—Table

FEF Data for Smoking and Non-Smoking Males

Group	Mean FEF (L/s)	SD FEF (L/s)	n
NS	3.78	0.79	200
PS	3.30	0.77	200
NI	3.32	0.86	50
LS	3.23	0.78	200
MS	2.73	0.81	200
HS	2.59	0.82	200

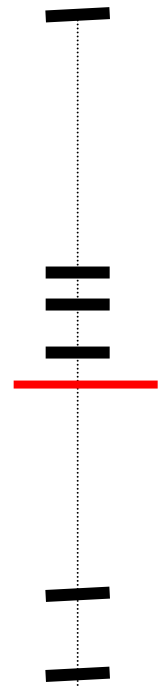
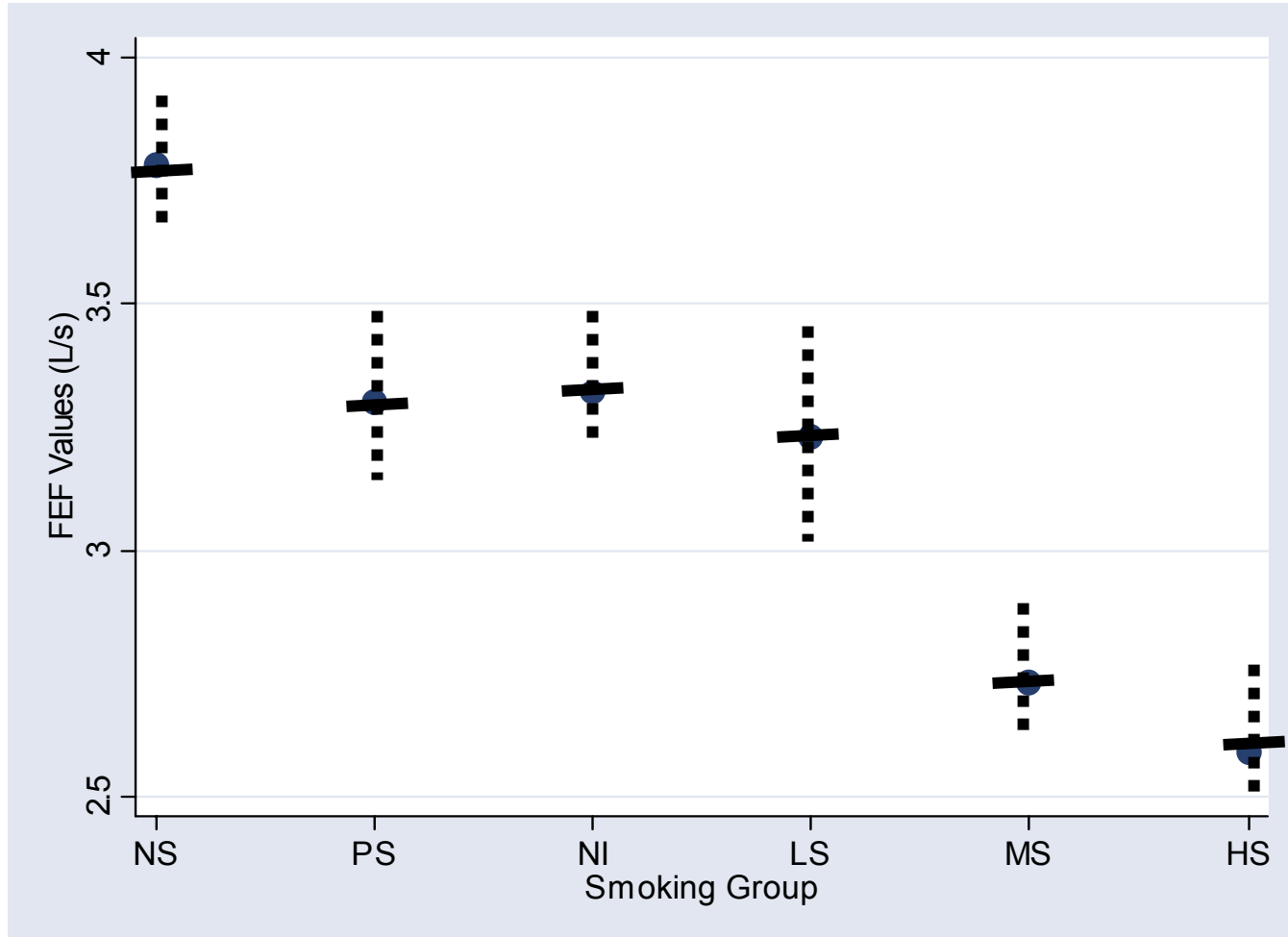
Example

- ◆ Based on a one-way analysis of variance, there are statistically significant differences in FEF levels among the six smoking groups ($p < .001$)

What's the Rationale behind Analysis of Variance?

- ◆ The variation in the sample means between groups is compared to the variation within a group
- ◆ If the between group variation is a lot bigger than the within group variation, that suggests there are some differences among the populations

Analysis of Variance



Summary: Smoking and FEF

◆ **Statistical Methods**

- 200 men were randomly selected from each of five smoking classification groups (non-smoker, passive smokers, light smokers, moderate smokers, and heavy smokers), as well as 50 men classified as non-inhaling smokers for a study designed to analyze the relationship between smoking and respiratory function

Summary: Smoking and FEF

◆ **Statistical Methods**

- Analysis of variance was used to test for any differences in FEF levels amongst the six groups of men
- Individual group comparisons were performed with a series of two sample t-tests, and 95% confidence intervals were constructed for the mean difference in FEF between each combination of groups

Summary: Smoking and FEF

◆ Results

- Analysis of variance showed statistically significant ($p < .001$) differences in FEF between the six groups of smokers
- Non-smokers had the highest mean FEF value, 3.78 L/s, and this was statistically significantly larger than the five other smoking-classification groups

Summary: Smoking and FEF

◆ Results

- The mean FEF value for non-smokers was 1.19 L/s higher than the mean FEF for heavy smokers (95% CI 1.03 – 1.35 L/s), the largest mean difference between any two smoking groups
- Confidence intervals for all smoking group FEF comparisons are in Table 1



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section D

Practice Problems

Practice Problems

1. Write a paragraph describing ANOVA . . .
 - When it is an appropriate test
 - What it is testing
 - And the basic ideas of how the test is carried out

Practice Problems

2. Suggest at least two research situations, preferably of interest to you, for which ANOVA would be a useful tool



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section D

Practice Problem Solutions

Practice Problems

1. Write a paragraph describing ANOVA . . .
 - When it is an appropriate test
 - What it is testing
 - And the basic ideas of how the test is carried out

Practice Problems

- ◆ ANOVA is a statistical method for comparing means amongst more than two independent populations
- ◆ ANOVA tests the general null hypothesis that all group means are equal
- ◆ If the p-value from ANOVA results in the rejection of the null, it means that at least one statistical difference in means between the groups being compared

Practice Problems

- ◆ More investigation is needed to determine which group mean(s) is/are statistically different, and confidence intervals need to be computed for the group comparisons of interest
- ◆ The ANOVA method compares within group variability to the variability between group means

Practice Problems

2. Suggest at least two research situations, preferably of interest to you, for which ANOVA would be a useful tool