

Lecture 09:

***t*-tests, and One-way ANOVA**

Outline

- Effect size based on t
- Independent-measures t test (continued)
- F-test for the equality of two variances
- Repeated-measures versus independent-measures designs
- ANOVA: comparison of means among 3 or more groups

Example of independent-measures (independent-sample) t test

Research has shown that people are more likely to show dishonest and self-interested behaviors in darkness than in a well-lit environment (Zhong, Bohns, & Gino, 2010). In one experiment, participants were given a set of 20 puzzles and were paid \$0.50 for each one solved in a 5-minute period. However, the participants reported their own performance and there was no obvious method for checking their honesty. Thus, the task provided a clear opportunity to cheat and receive undeserved money. One group of participants was tested in a room with dimmed lighting and a second group was tested in a well-lit room. The reported number of solved puzzles was recorded for each individual. The following data represent results similar to those obtained in the study.

Data

Number of Solved Puzzles			
Well-Lit Room		Dimly Lit Room	
11	6	7	9
9	7	13	11
4	12	14	15
5	10	16	11
$n = 8$		$n = 8$	
$M = 8$		$M = 12$	
$SS = 60$		$SS = 66$	

Number of Solved Puzzles			
Well-Lit Room		Dimly Lit Room	
11	6	7	9
9	7	13	11
4	12	14	15
5	10	16	11
$n = 8$		$n = 8$	
$M = 8$		$M = 12$	
$SS = 60$		$SS = 66$	

STEP 1 State the hypotheses and select the alpha level.

The prediction is that students in a dimly lit room are more likely to cheat (report higher scores) than are students in a well-lit room.

two-tailed or one-tailed?

$H_0: \mu_{\text{Dimly Lit}} \leq \mu_{\text{Well-Lit}}$ (Reported scores are not higher in the dimly lit room.)

$H_1: \mu_{\text{Dimly Lit}} > \mu_{\text{Well-Lit}}$ (Reported scores are higher in the dimly lit room.)

Suppose we use $\alpha = .01$.

STEP 2 Locate the critical region.

$$\begin{aligned}
 df &= df_1 + df_2 \\
 &= (n_1 - 1) + (n_2 - 1) \\
 &= 7 + 7 \\
 &= 14
 \end{aligned}$$

With $df = 14$, the one-tailed critical value for $\alpha = .01$ is $t = 2.624$.

STEP 3 Obtain the data and compute the test statistic.

Number of Solved Puzzles			
Well-Lit Room		Dimly Lit Room	
11	6	7	9
9	7	13	11
4	12	14	15
5	10	16	11
$n = 8$		$n = 8$	
$M = 8$		$M = 12$	
$SS = 60$		$SS = 66$	

First, find the pooled variance for the two samples:

$$\begin{aligned}s_p^2 &= \frac{SS_1 + SS_2}{df_1 + df_2} \\ &= \frac{60 + 66}{7 + 7} = \frac{126}{14} = 9\end{aligned}$$

Second, use the pooled variance to compute the estimated standard error:

$$\begin{aligned}s_{(M_1 - M_2)} &= \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{9}{8} + \frac{9}{8}} \\ &= \sqrt{2.25} \\ &= 1.50\end{aligned}$$

Third, compute the t statistic:

$$\begin{aligned}t &= \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}} = \frac{(8 - 12) - 0}{1.5} \\ &= \frac{-4}{1.5} = -2.67\end{aligned}$$

STEP 4 Make a decision.

Reported scores were significantly higher for students in the dimly lit room, $t(14) = -2.67$, $p < .01$, one-tailed. ■

The exact p value should also be reported.

Measuring Effect Size for the Independent-Measures t

- Effect size for the independent-measures t is measured in the same way that we measured effect size for the single-sample t in Chapter 8.
- Specifically, you can compute an estimate of Cohen's d or you can compute r^2 to obtain a measure of the percentage of variance accounted for by the treatment effect.

Cohen's d:

$$\text{estimated } d = \frac{\text{estimated mean difference}}{\text{estimated standard deviation}} = \frac{M_1 - M_2}{\sqrt{s_p^2}}$$

The percentage of variance accounted for:

$$r^2 = \frac{t^2}{t^2 + df}$$

Number of Solved Puzzles			
Well-Lit Room		Dimly Lit Room	
11	6	7	9
9	7	13	11
4	12	14	15
5	10	16	11
$n = 8$		$n = 8$	
$M = 8$		$M = 12$	
$SS = 60$		$SS = 66$	

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

$$= \frac{60 + 66}{7 + 7} = \frac{126}{14} = 9$$

$$d = \frac{M_1 - M_2}{\sqrt{s_p^2}} = \frac{8 - 12}{\sqrt{9}} = \frac{-4}{3} = -1.33$$

Note: Cohen's d is typically reported as a positive value; in this case $d = 1.33$.

we obtained $t = -2.67$ with $df = 14$.

$$r^2 = \frac{-2.67^2}{-2.67^2 + 14} = \frac{7.13}{7.13 + 14} = \frac{7.13}{21.13} = 0.337$$

For this study, 33.7% of the variability in the scores can be explained by the difference between the two lighting conditions.

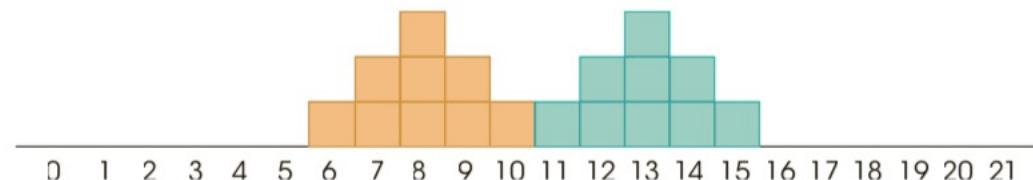
The influence of sample variance on the results of hypothesis testing

FIGURE 10.5

Two sample distributions representing two different treatments. These data show a significant difference between treatments, $t(16) = 8.62, p < .01$, and both measures of effect size indicate a very large treatment effect, $d = 4.10$ and $r^2 = 0.82$.

Treatment 1 ■ Treatment 2 ■

$n = 9$ $n = 9$
 $M = 8$ $M = 13$
 $s = 1.22$ $s = 1.22$



Treatment 1 ■ Treatment 2 ■

$n = 9$ $n = 9$
 $M = 8$ $M = 13$
 $s = 6.65$ $s = 6.65$



FIGURE 10.6

Two sample distributions representing two different treatments. These data show exactly the same mean difference as the scores in Figure 10.5, however the variance has been greatly increased. With the increased variance, there is no longer a significant difference between treatments, $t(16) = 1.59, p > .05$, and both measures of effect size are substantially reduced, $d = 0.75$ and $r^2 = 0.14$.

Recap: one-sample t tests (paired-sample t tests) and independent-sample t tests

	Sample Data	Hypothesized Population Parameter	Estimated Standard Error	Sample Variance
Single-sample t statistic	M	μ	$\sqrt{\frac{s^2}{n}}$	$s^2 = \frac{SS}{df}$
Independent-measures	$(M_1 - M_2)$	$(\mu_1 - \mu_2)$	$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$ homogeneity of variance (equal variance) assumed

(equivalent forms)

$$s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

What are the assumptions for t tests?

The Homogeneity of Variance Assumption

- Although most hypothesis tests are built on a set of underlying assumptions, the tests usually work reasonably well even if the assumptions are violated.
- The one notable exception is the assumption of **homogeneity of variance** for the independent-measures *t* test.
- The assumption requires that the two populations from which the samples are obtained have equal variances.
- This assumption is necessary in order to justify pooling the two sample variances and using the pooled variance in the calculation of the *t* statistic.

$$s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

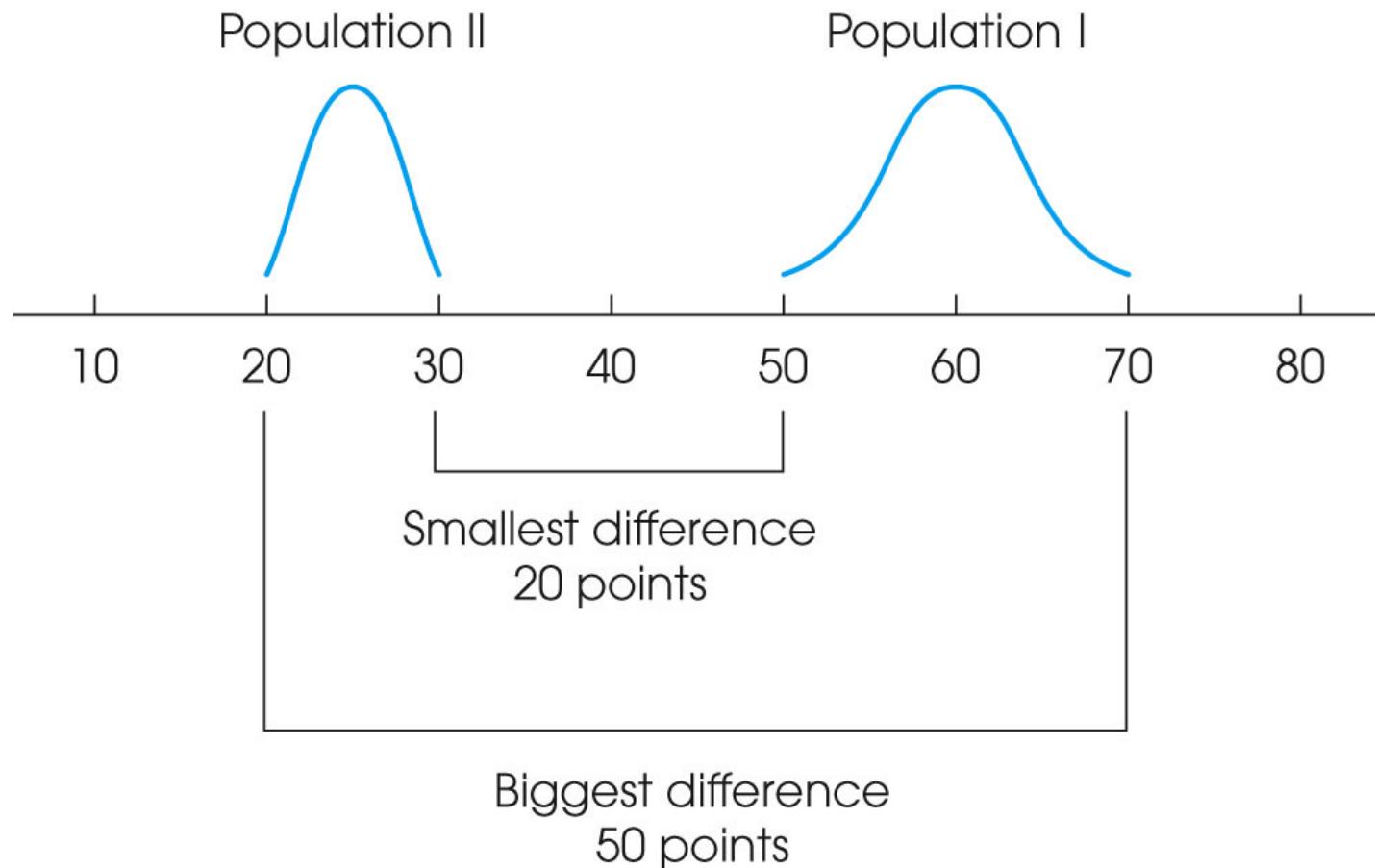
The Homogeneity of Variance Assumption (cont.)

- If the assumption is violated, then the t statistic contains two questionable values: (1) the value for the population mean difference which comes from the null hypothesis, and (2) the value for the pooled variance.
- The problem is that you cannot determine which of these two values is responsible for a t statistic that falls in the critical region.
- In particular, you cannot be certain that rejecting the null hypothesis is correct when you obtain an extreme value for t.

What if homogeneity of variance is violated?
(see Box 10.2 of the textbook)

$$s_{(M_1 - M_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(Note that the variability for the difference is the addition of the variability for the two populations.)



Summary of independent-sample tests of **unknown** variance

Welch T-test

Equal variance in two groups

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$df = n_1 + n_2 - 2$$

Unequal variance in two groups

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

$$d' = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\left(\frac{S_1^2}{n_1} \right)^2 / (n_1 - 1) + \left(\frac{S_2^2}{n_2} \right)^2 / (n_2 - 1)}$$

Satterthwaite Correction/approximation

Two-sample test of Normal variance for independent samples

- In the two-sample test of **mean** for independent samples we use different tests depending on whether we can assume the underlying variances for the two groups are equal.
 - how can we validate this assumption?
- Sometimes we simply want to know whether the variances of two Normal populations are the same.

Two-sample test of Normal variance for independent samples

Test $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_1 : \sigma_1^2 \neq \sigma_2^2$

Sample are assumed to follow

$N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively.

Test will be based on the relative magnitudes of sample variances S_1^2, S_2^2 . That is based on S_1^2 / S_2^2

Two-sample F-test of Normal variance for independent samples

(see also Hartley's F-max test in the textbook)

We need sampling distribution of S_1^2 / S_2^2 under H_0

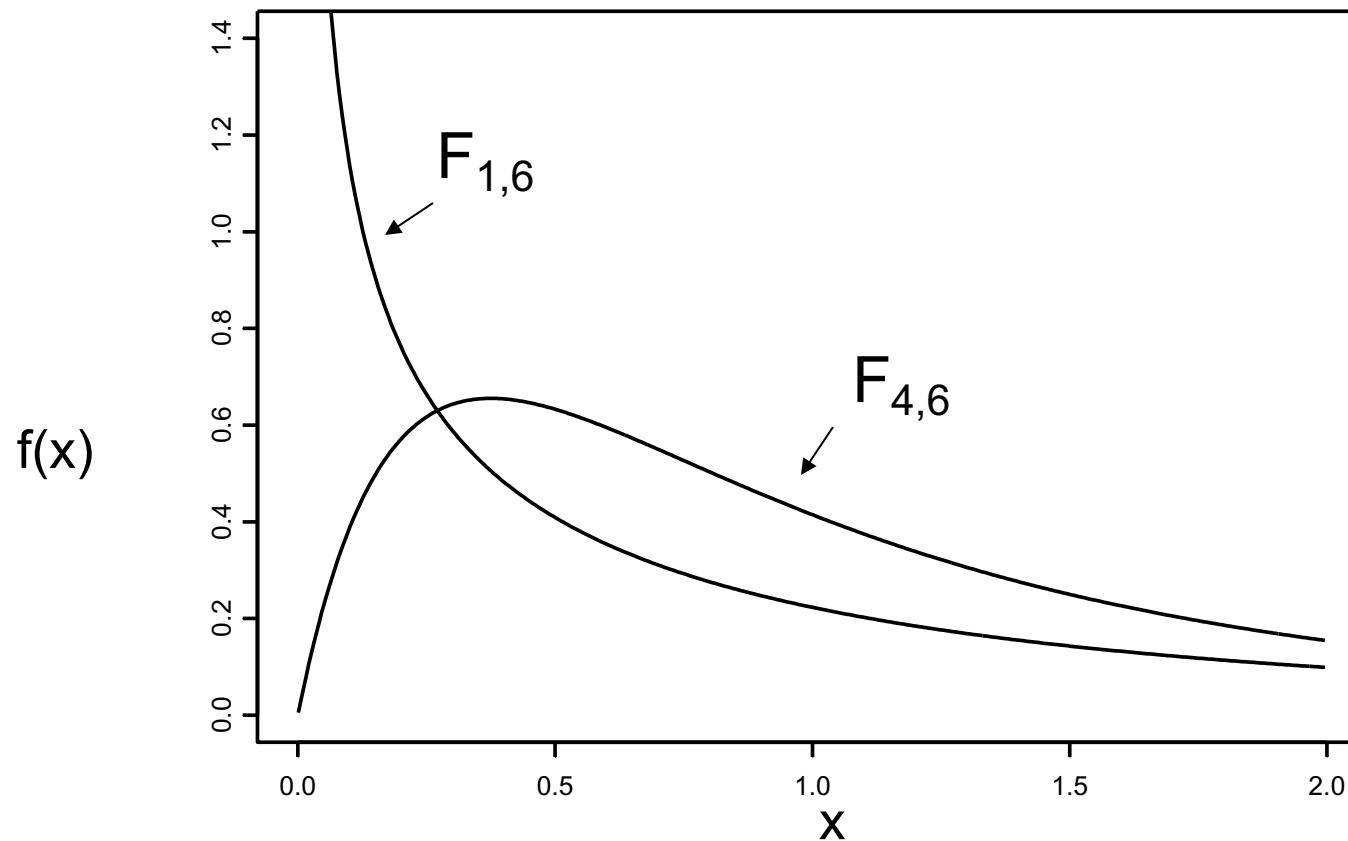
The ratio S_1^2 / S_2^2 follows an *F Distribution* .

This distribution has 2 parameters.

Numerator df, Denominator df.

$$S_1^2 / S_2^2 \sim F_{n1-1, n2-1}$$

Probability Density Function of F Distributions



The shape of the F distribution depends on df; and it is non-symmetrical.

The p th percentile of an F distribution with d_1 and d_2 dfs is denoted as:

$$F_{d_1, d_2, p} \quad \text{where} \quad \Pr(F_{d_1, d_2} \leq F_{d_1, d_2, p}) = p$$

F Distributions

The following relationship is needed when looking up values in a F table:

$$F_{d_1, d_2, p} = \frac{1}{F_{d_2, d_1, 1-p}}$$

$$\left[\Pr\left(\frac{S_1^2}{S_2^2} \leq F_{d_1, d_2, p}\right) = p, \quad \Pr\left(\frac{S_2^2}{S_1^2} \leq F_{d_2, d_1, 1-p}\right) = 1 - p \right]$$

F test for the Equality of two variances

Test the Hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_1 : \sigma_1^2 \neq \sigma_2^2$ with α level.

$$\text{If } F = \frac{S_1^2}{S_2^2} > F_{n_1-1, n_2-1, 1-\alpha/2} \quad \text{or} \quad \text{If } F < F_{n_1-1, n_2-1, \alpha/2}$$

Then H_0 is rejected.

One complete example

Cholesterol levels are assessed in 100 2-14 year old children whose father died from Heart Disease. This group has $\bar{x}_1 = 207.2$ mg%/ml, $s_1 = 35.6$.

Cholesterol levels are also assessed in 74 controls. These are children whose fathers are alive and do not have heart disease and are from the same census tract as the case children. This group has $\bar{x}_2 = 193.4$ mg%/ml, $s_2 = 17.3$.

Test at $\alpha = 0.05$

Want to compare means of two groups.

First compare variances to determine if they are equal.

$$F = \frac{s_1^2}{s_2^2} = \frac{35.6^2}{17.3^2} = 4.23$$

$H_0: \sigma_1^2 = \sigma_2^2$ rejected if $F > F_{99,73,.975} = 1.549$ or

if $F < F_{99,73,.025} = 0.655$

The two variances are significantly different.

Use unequal variances two-sample t-test.

$$t = \frac{207.3 - 193.4}{\sqrt{35.6^2/100 + 17.3^2/74}} = 3.40$$

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

$$df = d' = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)}$$

$$= \frac{35.6^2/100 + 17.3^2/74}{(35.6^2/100)^2/99 + (17.3^2/74)^2/73} = 151.4$$

$$d'' = 151$$

$$Note: n_1 + n_2 - 2 = 172$$

From t table $t=3.40 > t_{120,0.975} = 1.980 > t_{151,0.975}$

The table in the textbook is up to 120, too small....

Therefore reject H_0 . Children of fathers who died from HD have higher cholesterol levels than those of fathers without HD.

$p\text{-value} = 2(\text{area to right of } 3.40 \text{ of } t_{151})$
 $\approx 2(0.0004) < 0.001$

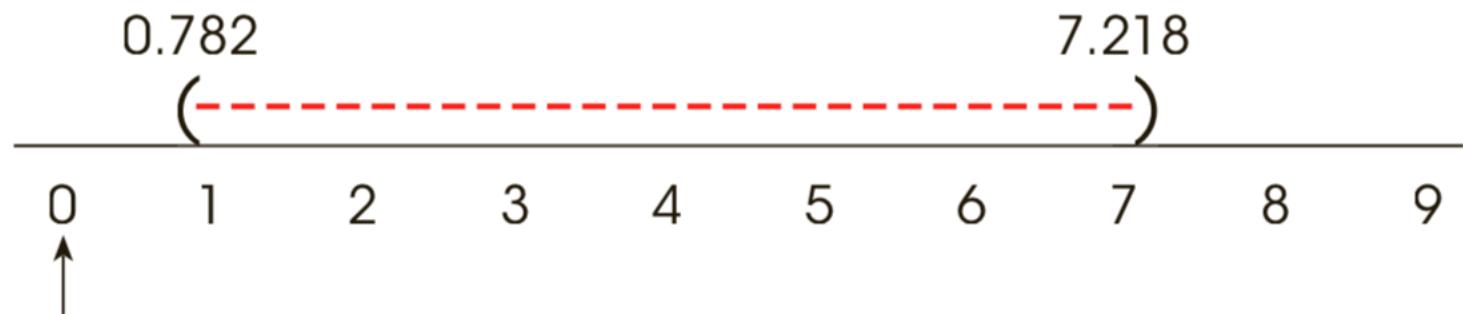
Confidence Interval for the mean difference of two normal distributions

Two sided CI $100(1-\alpha)$ for $\mu_1 - \mu_2$ When $\sigma_1^2 \neq \sigma_2^2$.

$$\left(\bar{x}_1 - \bar{x}_2 - t_{d'',1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{d'',1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

Confidence Intervals and Hypothesis Tests

95% confidence interval
estimate for $\mu_1 - \mu_2$



$\mu_1 = \mu_2$
according to H_0

Comparing Repeated-Measures and Independent-Measures Designs

- Because a repeated-measures design uses the same individuals in both treatment conditions, this type of design usually requires fewer participants than would be needed for an independent-measures design.
- In addition, the repeated-measures design is particularly well suited for examining **changes** that occur over time, such as learning or development.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

- The primary advantage of a repeated-measures design, however, is that it reduces variance and error by removing **individual differences**.
- The first step in the calculation of the repeated-measures t statistic is to find the difference score for each subject.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

This simple process has two very important consequences:

1. First, the D score for each subject provides an indication of how much difference there is between the two treatments. If all of the subjects show roughly the same D scores, then you can conclude that there appears to be a consistent, systematic difference between the two treatments. You should also note that when all the D scores are similar, the variance of the D scores will be small, which means that the standard error will be small and the t statistic is more likely to be significant.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

2. Also, you should note that the process of subtracting to obtain the D scores **removes the individual differences** from the data. That is, the initial differences in performance from one subject to another are eliminated. **Removing individual differences** also tends to reduce the variance, which creates a smaller standard error and increases the likelihood of a significant t statistic.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

The following data demonstrate these points:

Subject	X_1	X_2	D
A	9	16	7
B	25	28	3
C	31	36	5
D	58	61	3
E	72	79	7

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

- First, notice that all of the subjects show an increase of roughly 5 points when they move from treatment 1 to treatment 2.
- Because the treatment difference is very consistent, the D scores are all clustered close together will produce a very small value for s^2 .
- This means that the standard error in the bottom of the t statistic will be very small.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

- Second, notice that the original data show big differences from one subject to another. For example, subject B has scores in the 20's and subject E has scores in the 70's.
- These big **individual differences** are eliminated when the difference scores are calculated.
- Because the individual differences are removed, the D scores are usually much less variable than the original scores.
- Again, a smaller variance will produce a smaller standard error, which will increase the likelihood of a significant t statistic.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

- Finally, you should realize that there are potential **disadvantages** to using a repeated-measures design instead of independent-measures.
- Because the repeated-measures design requires that each individual participate in more than one treatment, there is always the risk that exposure to the first treatment will cause a change in the participants that influences their scores in the second treatment.

Comparing Repeated-Measures and Independent-Measures Designs (cont.)

- For example, practice in the first treatment may cause improved performance in the second treatment.
- Thus, the scores in the second treatment may show a difference, but the difference is not caused by the second treatment.
- When participation in one treatment influences the scores in another treatment, the results may be distorted by **order effects**, and this can be a serious problem in repeated-measures designs.

Think deep into these concepts

Distributions: z, t, F

How are they defined; What shapes they have; why we need them?

Equations: SD, SEM, z-score, t-value, p-value, power & sample size, degree of freedom

How do they relate to distributions; why are they defined in their particular ways;
How to relate those equations?

Estimation and hypothesis testing

How are they mapped onto distributions? How does confidence interval relate to hypothesis testing? What do those tests mean? How do they differ?

Steps in Hypothesis Testing

1. State the hypotheses (both null and alternative)
2. Specify the significance level (α)
3. Draw sample of size n , *compute the test statistic*
Use Normal (Z) when σ is known and t when you have estimated σ with S
4. Determine P-value
5. Compare P-value to the significance level α and decide whether or not to reject H_0
6. State conclusions in terms of subject matter

Significance Conclusion from p-values

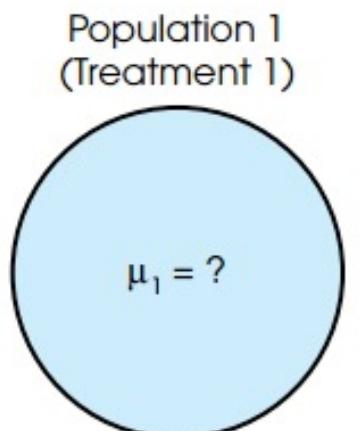
- If $0.01 \leq p < 0.05$ significant
- If $0.001 \leq p < 0.01$ highly significant
- If $p < 0.001$ very highly significant
- If $p > 0.05$ not statistically significant

Sometimes $0.045 \leq p < 0.10$ is called borderline significant or marginally significant.

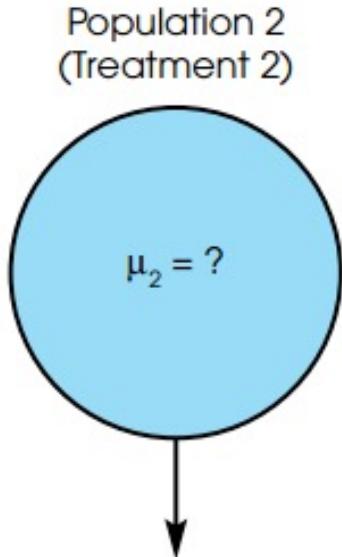
Statistical vs. Scientific Significance

- There is an important distinction between statistical significance and scientific (or clinical) significance.
- It is possible to get a statistically significant result that is not clinically significant (due to large sample size). e.g. 0.1gram difference
- The converse may also occur; statistically not significant but clinically significant, usually found in exploratory analysis. This could lead to a larger confirmatory study.

Think about the following question

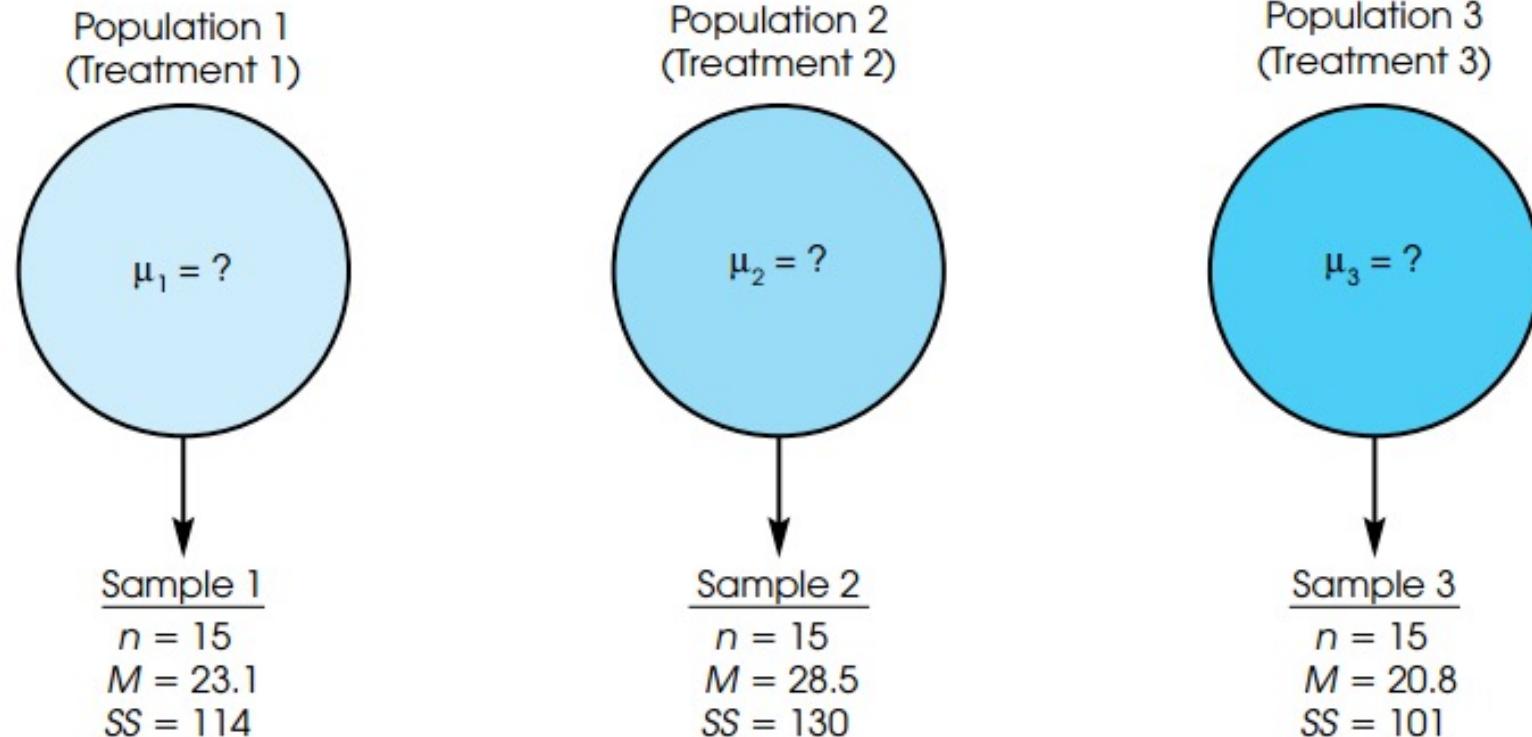


Sample 1
 $n = 15$
 $M = 23.1$
 $SS = 114$



Sample 2
 $n = 15$
 $M = 28.5$
 $SS = 130$

The question for one-way ANOVA



- Compare the means from multiple conditions (>2). Similar to t tests, which only deals with two conditions. Here we have three levels.

Driving performance example

We **randomly** select **three groups of subjects**, and compare the driving performance under three experimental conditions:

1. No phone
2. A hands-free phone
3. A hand-held phone

Question: Whether the phone use affects driven performance X

Hypothesis settings

H_0 : Phone use does not affect driving performance.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

H_1 : There is at least one mean difference among the populations.

Note: there are multiple versions of H_1 . In other words, there are multiple ways to refute H_0 .

Test statistics of ANOVA

The statistic of ANOVA is **F-ratio**, which is based on variance, instead of sample mean difference as in t-test.

$$t = \frac{\text{obtained difference between two sample means}}{\text{standard error (the difference expected with no treatment effect)}}$$

$$F = \frac{\text{variance (differences) between sample means}}{\text{variance (differences) expected with no treatment effect}}$$

Variance here is a measure of *overall differences among the sample means*.

Why not simply use t-test?

- The overall probability of a Type I error (when having multiple simultaneous t-tests) accumulates over a series of separate hypothesis tests.
- Typically, the experiment-wise alpha level is substantially greater than the test-wise alpha level used for any one of the individual tests.
- E.g., do 3 separate pair-wise t-tests (set $\alpha = 0.05$). What's the overall α ?

$$\text{Experimental-wise } \alpha = 1 - (1 - \text{test-wise } \alpha)^3$$

$$1 - 0.95^3 = 1 - 0.86 = 0.14 > 0.05$$

The logic of ANOVA

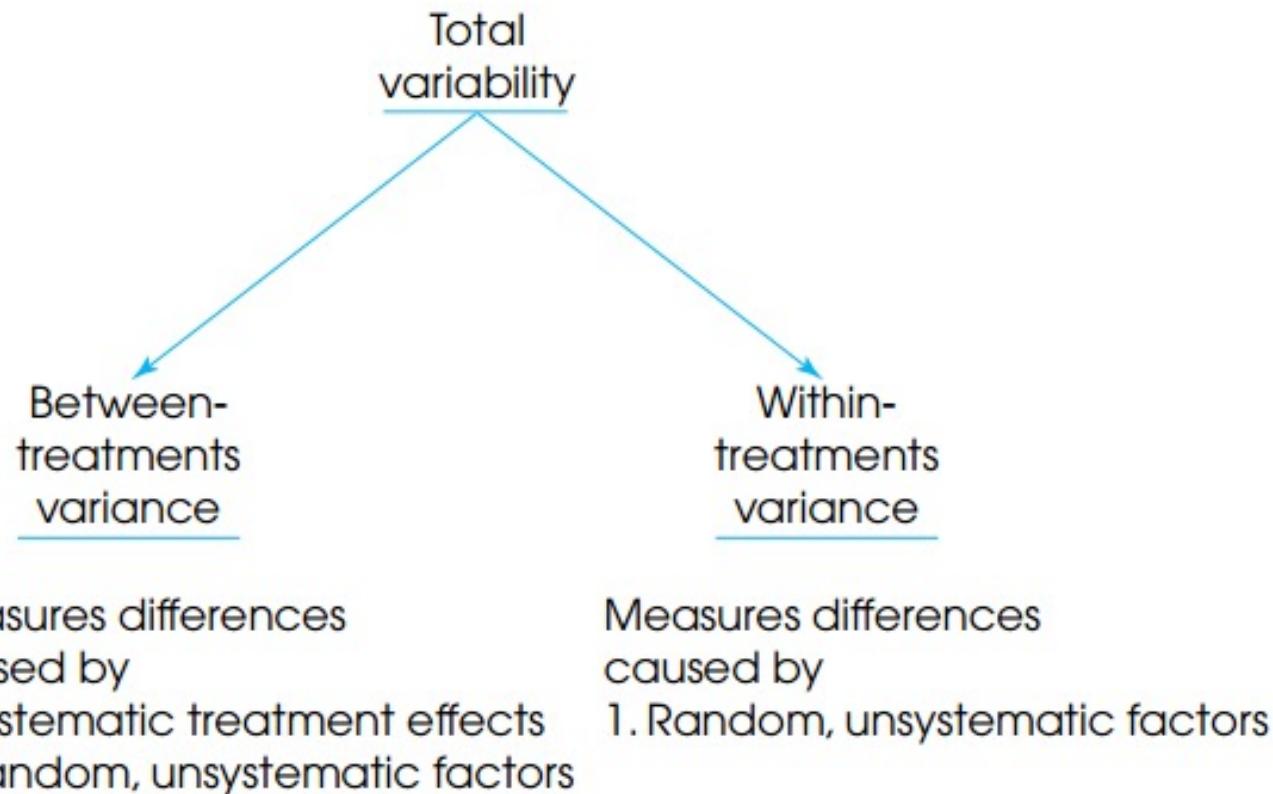
<i>Treatment 1: No Phone (Sample 1)</i>	<i>Treatment 2: Hand-Held Phone (Sample 2)</i>	<i>Treatment 3: Hands-Free Phone (Sample 3)</i>
4	0	1
3	1	2
6	3	2
3	1	0
4	0	0
$M = 4$	$M = 1$	$M = 1$

Part of the difference between scores can be attributed to the grouping of scores (treatment effect). E.g., $M = 4, 1$ and 1 here.

<i>Treatment 1: No Phone (Sample 1)</i>	<i>Treatment 2: Hand-Held Phone (Sample 2)</i>	<i>Treatment 3: Hands-Free Phone (Sample 3)</i>
4	0	1
3	1	2
6	3	2
3	1	0
4	0	0
$M = 4$	$M = 1$	$M = 1$

- The other part of the difference between scores can be attribute to the variance within each treatment.
- Total variance = Between-treatments variance + Within-treatment variance

Partitioning of variance



- Random effect is inevitable: sampling error, individual difference and so on.
- Unspecific, cannot be explained by a finite reason.

F is a ratio between variances

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}} = \frac{\text{differences including any treatment effects}}{\text{differences with no treatment effects}}$$

$$F = \frac{\text{systematic treatment effects} + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$

$$F = \frac{0 + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$

Error term

- When there is no systematic treatment effect, F will be close to 1.
- When there is a treatment effect, F should be substantially larger than 1.

How to compute

Telephone Conditions			
<i>Treatment 1</i> <i>No Phone</i> <i>(Sample 1)</i>	<i>Treatment 2</i> <i>Hand-Held Phone</i> <i>(Sample 2)</i>	<i>Treatment 3</i> <i>Hands-Free Phone</i> <i>(Sample 3)</i>	
4	0	1	$\Sigma X^2 = 106$
3	1	2	$G = 30$
6	3	2	$N = 15$
3	1	0	$k = 3$
4	0	0	

$T_1 = 20$	$T_2 = 5$	$T_3 = 5$
$SS_1 = 6$	$SS_2 = 6$	$SS_3 = 4$
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$M_1 = 4$	$M_2 = 1$	$M_3 = 1$

In general, the variance of X is computed as the sum of squared difference (SS) divided by the degree of freedom:

$$\text{sample variance} = s^2 = \frac{SS}{df} \qquad SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

If the number of X is k , then $df = k-1$

The final goal for the ANOVA is an *F*-ratio

$$F = \frac{\text{Variance between treatments}}{\text{Variance within treatments}}$$

Each variance in the *F*-ratio is computed as SS/df

$$\text{Variance between treatments} = \frac{SS \text{ between}}{df \text{ between}}$$

$$\text{Variance within treatments} = \frac{SS \text{ within}}{df \text{ within}}$$

To obtain each of the *SS* and *df* values, the total variability is analyzed into the two components

$$\begin{array}{ccc} \text{SS total} & & \\ \swarrow & \searrow & \\ \text{SS between} & & \text{SS within} \end{array}$$

$$\begin{array}{ccc} \text{df total} & & \\ \swarrow & \searrow & \\ \text{df between} & & \text{df within} \end{array}$$

The total *SS* of all scores:

$$SS_{\text{total}} = \sum X^2 - \frac{G^2}{N} = 106 - 30^2/15 = 106 - 60 = 46$$

$$df_{\text{total}} = N - 1 = 15 - 1 = 14$$

Between-treatment variance

$$SS_{\text{between}} = SS_{\text{total}} - SS_{\text{within}}$$

When n is equal across conditions:

$$SS_{\text{between}} = n(SS_{\text{means}}) = 5 (6) = 30$$

When n is NOT equal across conditions:

$$SS_{\text{between}} = \sum \frac{T^2}{n} - \frac{G^2}{N} = 20^2/5 + 5^2/5 + 5^2/5 - 30^2/5 = 80 + 5 + 5 - 60 = 30$$

where T is the treatment total, instead of treatment mean

$$df_{\text{between}} = k - 1 = 3 - 1 = 2$$

The number of treatment (X for this variance) minus 1

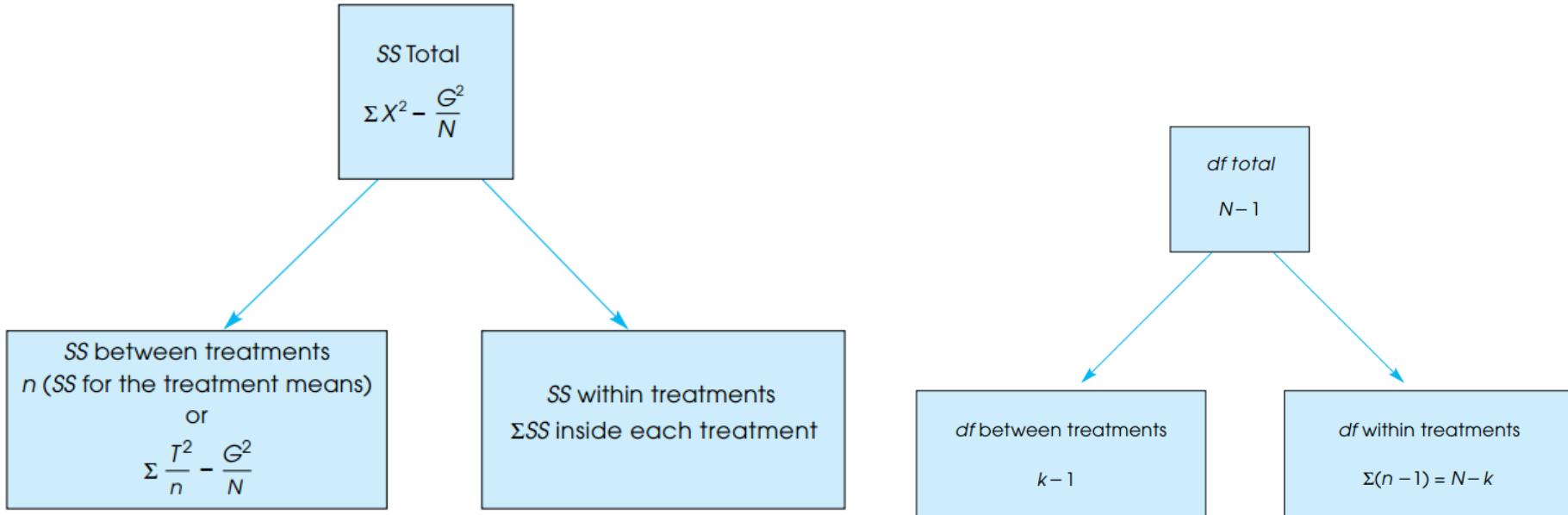
Within-treatment variance

Within-treatment SS is the sum of SS within each treatment:

$$SS_{\text{within treatments}} = \sum SS_{\text{inside each treatment}} = 6 + 6 + 4 = 16$$

Accordingly, the degree of freedom for within-treatment SS is the sum of degrees of freedom from all treatments:

$$df_{\text{within}} = \sum (n - 1) = \sum df_{\text{in each treatment}} = (5-1) + (5-1) + (5-1) = 12$$



$$MS(\text{variance}) = s^2 = \frac{SS}{df}$$

$$MS_{\text{between}} = s_{\text{between}}^2 = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{30}{2} = 15$$

$$MS_{\text{within}} = s_{\text{within}}^2 = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{16}{12} = 1.33$$

$$F = \frac{s_{\text{between}}^2}{s_{\text{within}}^2} = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

$$= 15/1.33 = 11.288$$

ANOVA table

Source	SS	df	MS	
Between treatments	30	2	15	$F = 11.28$
Within treatments	16	12	1.33	
Total	46	14		

$$F = \frac{0 + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$

Error term

Here, the error term is the within-treatment MS.

Formulas

Total sum of squares (or SS_{total} , 总和方)

$$SS_{total} = \sum (X_{i,j} - \bar{G})^2 = \sum X_{i,j}^2 - (G^2 / N)$$

Within-treatments Sum of Squares or (SS_{within} , 组内和方), the summation of the variability inside each treatment condition

$$SS_{\text{within treatments}} = \sum SS_{\text{inside each treatment}}$$

Between-treatments Sum of Squares, or ($SS_{between}$, 组间和方)

$$SS_{between} = \sum_{i=1}^k n_i (M_i - \bar{G})^2 = \sum_{i=1}^k (T_i^2 / n_i) - G^2 / N$$

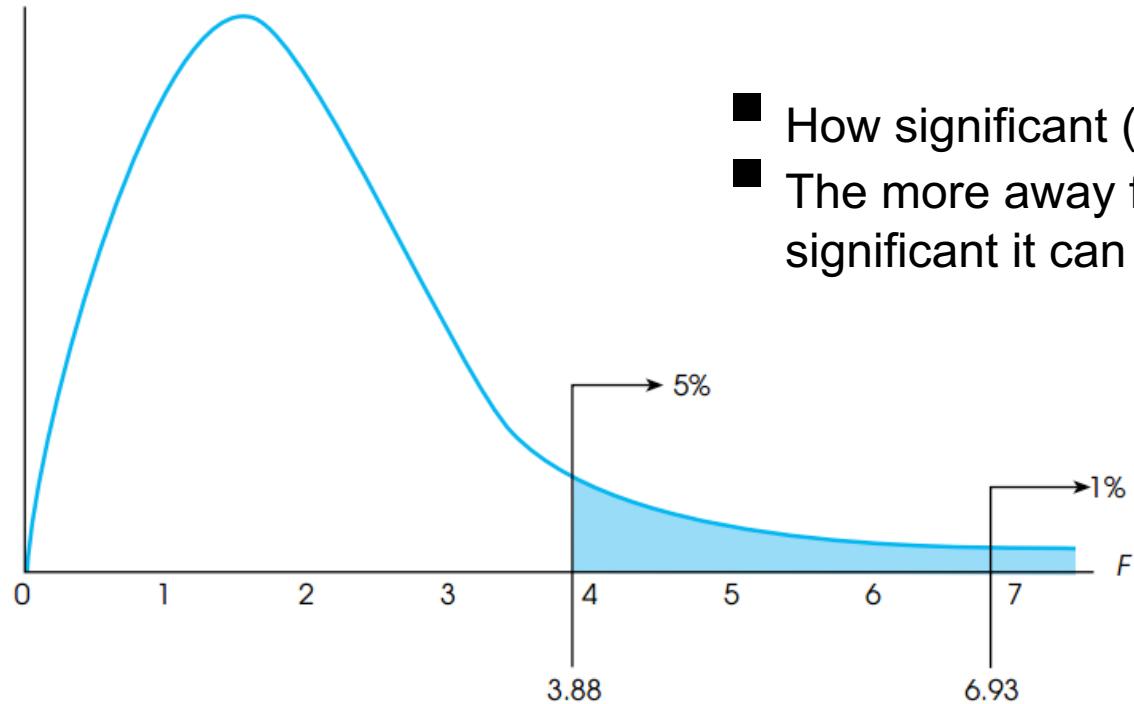
Distribution of F-ratios

Two characteristics of the F-distribution:

- F values will always be positive numbers because we are using variance (which is always positive) to compute them
- When the null hypothesis is true, the F-ratio should be near 1.00. This means that the distribution of F-ratios should pile up around 1.00

Specifically for ANOVA, a sufficiently large F value indicates a significantly difference.

F distribution



- How significant (extreme) the F is?
- The more away from 1, the more significant it can be.

Degrees of Freedom: Denominator	Degrees of Freedom: Numerator					
	1	2	3	4	5	6
10	4.96	4.10	3.71	3.48	3.33	3.22
	10.04	7.56	6.55	5.99	5.64	5.39
11	4.84	3.98	3.59	3.36	3.20	3.09
	9.65	7.20	6.22	5.67	5.32	5.07
12	4.75	3.88	3.49	3.26	3.11	3.00
	9.33	6.93	5.95	5.41	5.06	4.82
13	4.67	3.80	3.41	3.18	3.02	2.92
	9.07	6.70	5.74	5.20	4.86	4.62
14	4.60	3.74	3.34	3.11	2.96	2.85
	8.86	6.51	5.56	5.03	4.69	4.46

Another visual representation

