Lecture 20: Chapter 11, Section 3 Categorical & Quantitative Variable Inference in Several-Sample Design

- Compare and Contrast Several- and 2-sample
- Variation Among Means or Within Groups
- *F* Statistic as Ratio of Variation
- Role of Sample Size
- Formulating Hypotheses Correctly
- Questions about Practice Midterm 2?

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-3)
  - Displaying and Summarizing (Lectures 3-8)
  - Probability (discussed in Lectures 9-14)
  - Statistical Inference
    - 1 categorical (discussed in Lectures 14-16)
    - 1 quantitative (discussed in Lectures 16-18)
    - cat and quan: paired, 2-sample, several-sample
    - 2 categorical
    - 2 quantitative

Inference Methods for C→Q *(Review)*

- Paired: reduces to 1-sample *t*
  - Focused on mean of differences
- Two-Sample: 2-sample *t* (similar to 1-sample *t*)
  - Focused on difference between means
- Several-Sample: need new distribution (*F*)
  - Focus on difference among means

Display & Summary, Several Samples *(Review)*

- **Display: Side-by-side boxplots:**
  - One boxplot for each categorical group
  - All share same quantitative scale
- **Summarize:** Compare
  - Five Number Summaries (looking at boxplots)
  - Means and Standard Deviations

Looking Ahead: *Inference for population relationship focuses on* ***means and standard deviations****.*

Notation

	Sizes	Means	s.d.s
Sample	I =no. of groups compared $n_1, n_2, \cdots, n_I$ sum to N	$\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_I$ (overall $\bar{x}$)	$s_1, s_2, \cdots, s_I$
Population		$\mu_1, \mu_2, \cdots, \mu_I$	$\sigma_1, \sigma_2, \cdots, \sigma_I$

Two- vs. Several-Sample Inference

- **Similar:** test statistic standardizes difference among sample means, taking sample sizes and standard deviations into account
- **Different:** several-sample test statistic (***F***) focuses on
  - Squared differences of means in numerator
  - Squared standard deviations **(variances)** in denominator

Procedure called **ANOVA** (ANalysis Of VAriance)

Two- vs. Several-Sample Inference

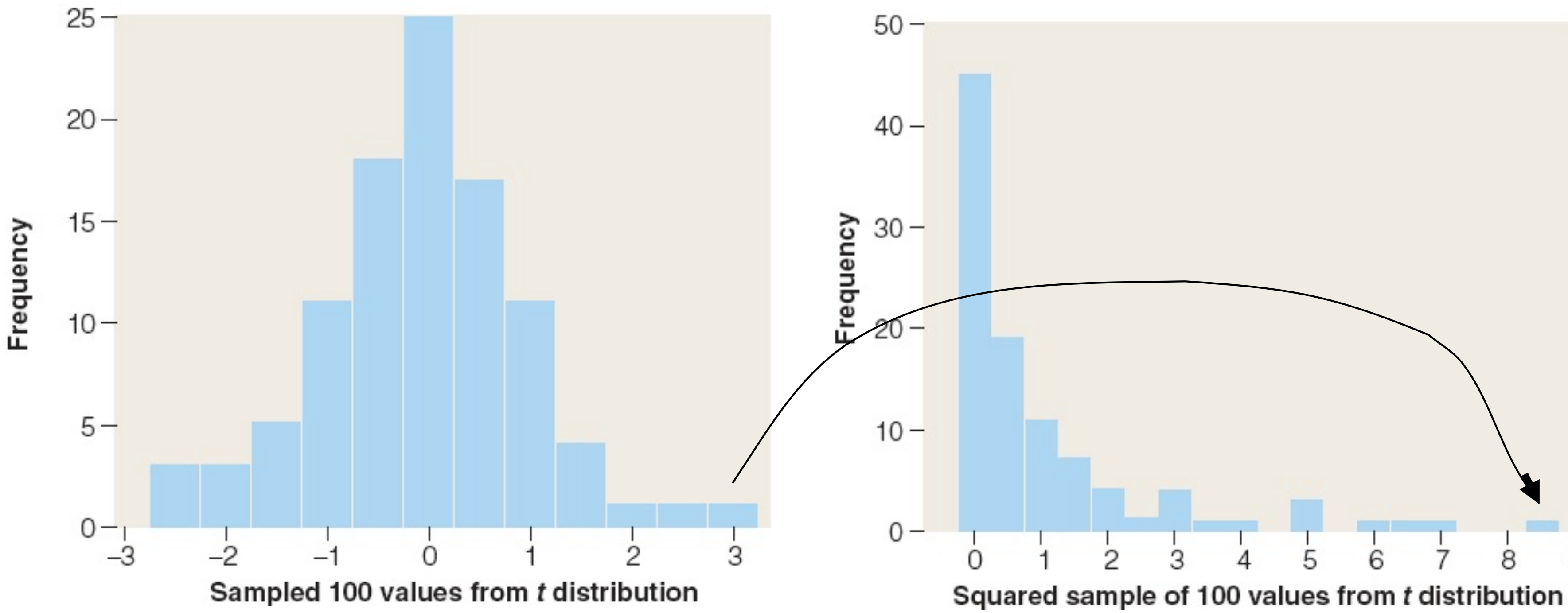
- **Similar:** test statistic standardizes difference among sample means, taking sample sizes and standard deviations into account.

For 2 groups of equal sizes and $\sigma_1 = \sigma_2, F = t^2$ and conclusions (including *P*-value) are the same.

t and *F* Distributions

- Left: sampled 100 values from a *t* distribution
- Right: squared the 100 values from *t* distribution

Squaring makes *F* non-negative, right-skewed (makes extreme values even more extreme; for example, 3 becomes 9)



Two- vs. Several-Sample Statistics

- **Similar:** test statistic standardizes how different sample means are, taking sample sizes and standard deviations into account

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$F = \frac{\left[n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2\right] / (I-1)}{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2\right] / (N - I)}$$

Two- vs. Several-Sample Statistics

- How different are sample means?
- How spread out are the distributions?
- How large are the samples? (As far as contributing to the size of F is concerned, the individual group sizes "cancel out": the main contributor is N.)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$F = \frac{\left[n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2\right] / (I-1)}{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2\right] / (N - I)}$$

What Makes *t* or *F* Statistics Large

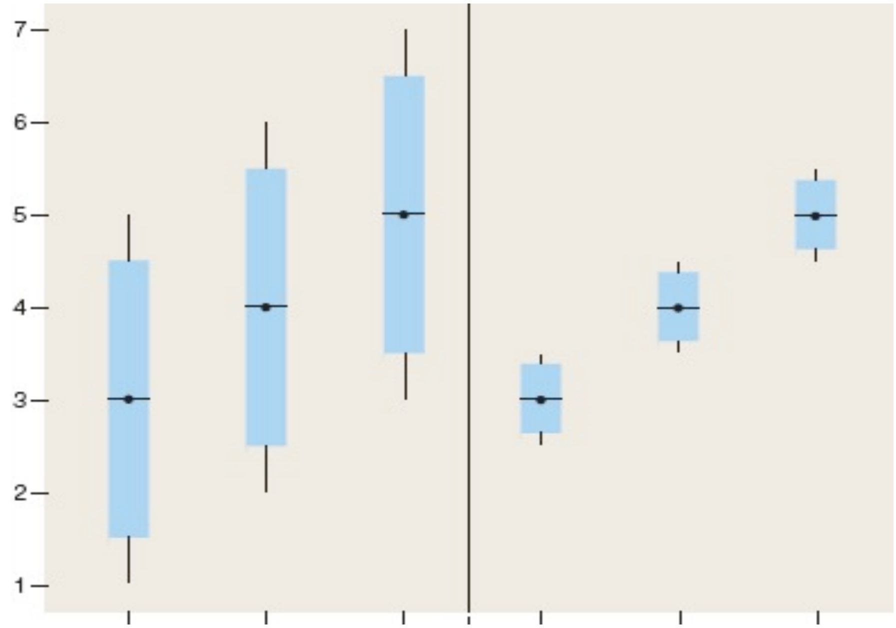
- Large diff among sample means (in numerator)
- Small spreads (in denominator)
- Large sample sizes (denominator of denominator)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$F = \frac{\left[n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2\right] / (I - 1)}{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2\right] / (N - I)}$$

Example: *Sample S.D.s' Effect on P-Value*

- **Background**: Boxplots with $\bar{x}_1 = 3, \bar{x}_2 = 4, \bar{x}_3 = 5$ could appear as on left or right, depending on s.d.s.

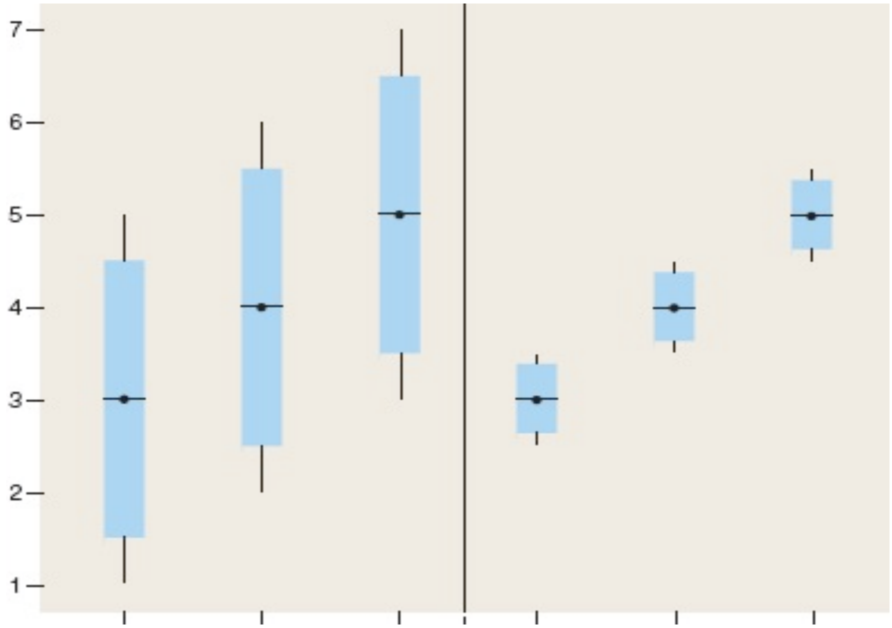


Context: *sample mean monthly pay (in $1000s) for 3 racial/ethnic groups.*

- **Question:** For which scenario does the difference among means appear more significant?
- **Response:** Difference among means appears more significant on ______________________________

Example: *Sample S.D.s' Effect on P-Value*

- **Background**: Boxplots with $\bar{x}_1 = 3, \bar{x}_2 = 4, \bar{x}_3 = 5$ could appear as on left or right, depending on s.d.s.



***Context:** sample mean monthly pay (in \$1000s) for 3 racial/ethnic groups.*

- **Question:** For which scenario are we more likely to reject hypothesis of equal population means?
- **Response:** Scenario on ______: smaller s.d.s→larger *F* stat→smaller *P*-val→likelier to reject H_0, conclude ________________

Practice: 11.35b p.563

Measuring Variation Among and Within

$$F = \frac{\left[n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2\right] / (I-1)}{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2\right] / (N-I)}$$

- **Numerator:** variation among groups
  - How different are $\bar{x}_1, \cdots, \bar{x}_I$ from one another?
- **Denominator:** variation within groups
  - How spread out are samples? (sds $s_1, \cdots, s_I$)

Numerator of F (Difference Among Means)

- **SSG:** Sum of Squared diffs among Groups

$$SSG = 5(3-4)^2 + 5(4-4)^2 + 5(5-4)^2 = \boxed{10}$$

- **DFG:** Degrees of Freedom for Groups

$$DFG = I - 1 = 3 - 1 = \boxed{2}$$

- **MSG:** Mean Squared diffs among Groups

$$MSG = \frac{SSG}{DFG} = \frac{10}{2} = \boxed{5}$$

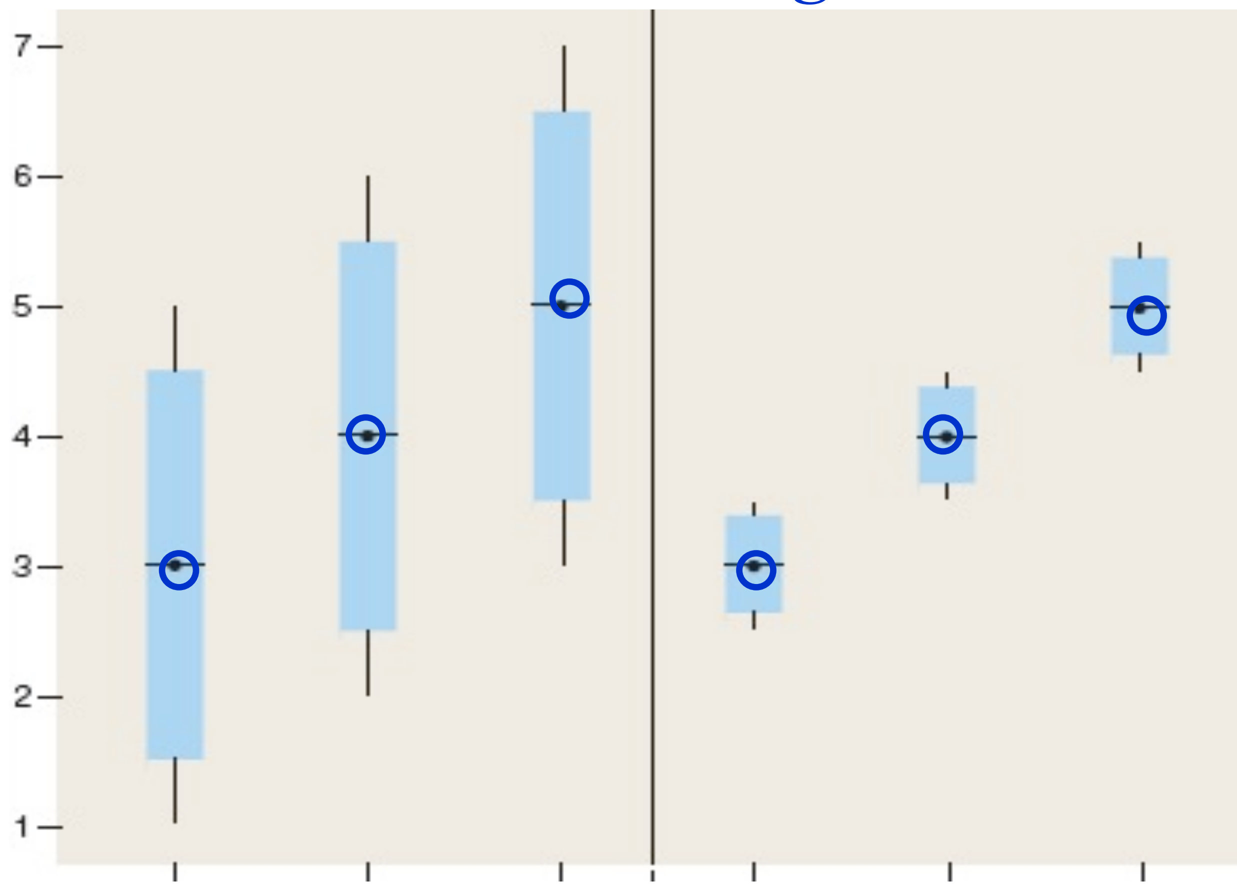
$I = 3$

$n_1 = 5$	$\bar{x}_1 = 3$	$s_1 = 1.58$
$n_2 = 5$	$\bar{x}_2 = 4$	$s_2 = 1.58$
$n_3 = 5$	$\bar{x}_3 = 5$	$s_3 = 1.58$
$N = 15$	$\bar{x} = 4$	

monthly earnings (in \$1000s) for 3 racial/ethnic groups (hypothetical)

Numerator of *F* (Difference Among Means)

Note: numerator of *F* is the same for both scenarios because the difference *among* means is the same.



Denominator of F (Spread Within Groups)

- **SSE:** Sum of Squared Error within Groups

$$SSE = (5-1)1.58^2 + (5-1)1.58^2 + (5-1)1.58^2 = \boxed{30}$$

- **DFE:** Degrees of Freedom for Error

$$DFE = N - I = 15 - 3 = \boxed{12}$$

- **MSE:** Mean Squared Error within Groups

$$MSE = \frac{SSE}{DFE} = \frac{30}{12} = \boxed{2.5}$$

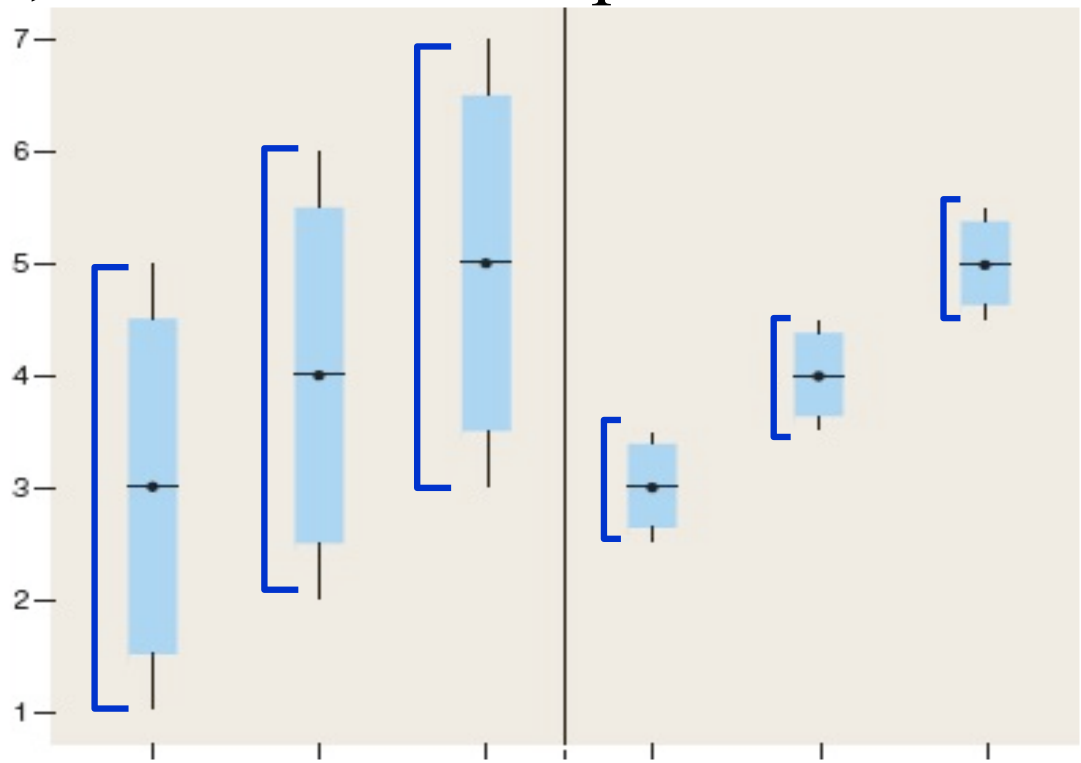
I* = *3

$n_1 = 5$	$\bar{x}_1 = 3$	$s_1 = 1.58$
$n_2 = 5$	$\bar{x}_2 = 4$	$s_2 = 1.58$
$n_3 = 5$	$\bar{x}_3 = 5$	$s_3 = 1.58$
$N = 15$	$\bar{x} = 4$	

monthly earnings (in $1000s) for 3 racial/ethnic groups (hypothetical)

Denominator of *F* (Spread Within Groups)

- Note: denominator of *F* is smaller for the scenario on the right, because of less spread.



- Because the numerators are the same, *F* (the quotient) is considerably larger on the right.

The *F* Statistic

$$F = \frac{\left[n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2\right]/(I-1)}{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2\right]/(N-I)}$$

$$= \frac{MSG}{MSE} = \frac{5}{2.5} = 2$$ ***Is 2 large???***

measures difference among sample means
(relative to spreads and sample sizes)

If *F* is large, reject $H_o : \mu_1 = \mu_2 = \mu_3$

Conclude population means differ.

Example: *Size of Standardized Statistics*

- **Background:** Say standardized statistic is 2.
- **Question:** Is 2 large…
  - For z?
  - For t?
  - For F?
- **Response:**
  - z=2 large? ____________ (combined tail probs 0.05)
  - t=2 large? depends on ___________________________
  - F=2 large?

depends on ____
(based on total sample size N **and** number of groups I)

F and its Degrees of Freedom

Family of *F* curves all non-neg, right-skewed.

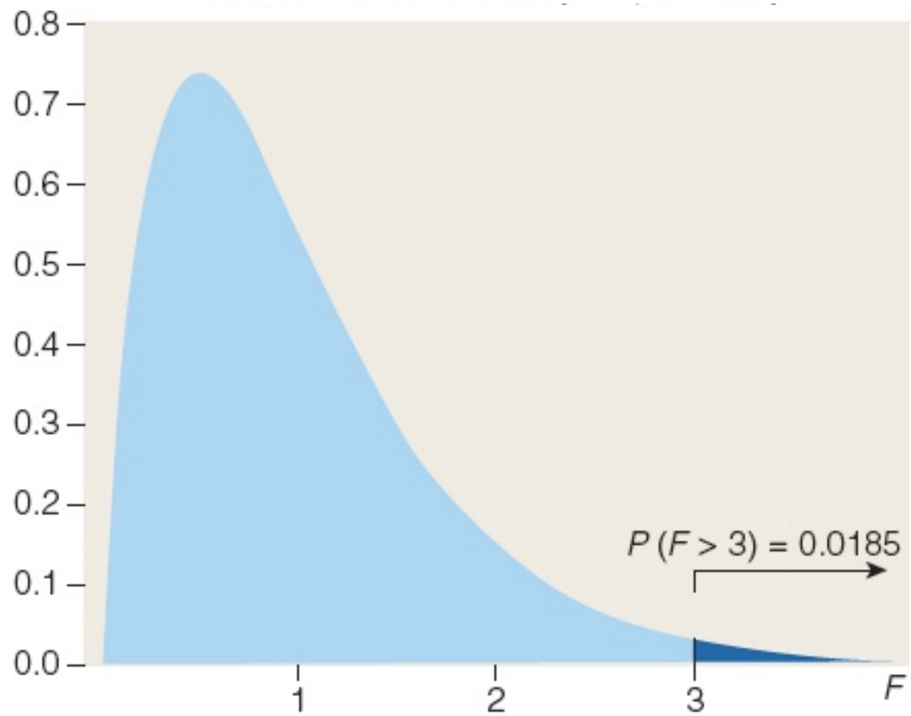
Spreads vary, depending on DFG = I - 1 in numerator, DFE = N - I in denominator.

Example: *Degrees of Freedom for F*

- **Background**: Consider these *F* distributions
  - *F* with *I*=5, *N*=390
  - *F* with DFG=2, DFE=12 [written *F*(2, 12)]
- **Questions:**
  - What are degrees of freedom if *I*=5, *N*=390?
  - What are *I* and *N* if DFG=2, DFE=12?
- **Responses:**
  - I = 5, N = 390→
    DFG = ______________, DFE = ___________________
  - DFG = 2, DFE = 12→

Example: *Assessing Size of F Statistic*

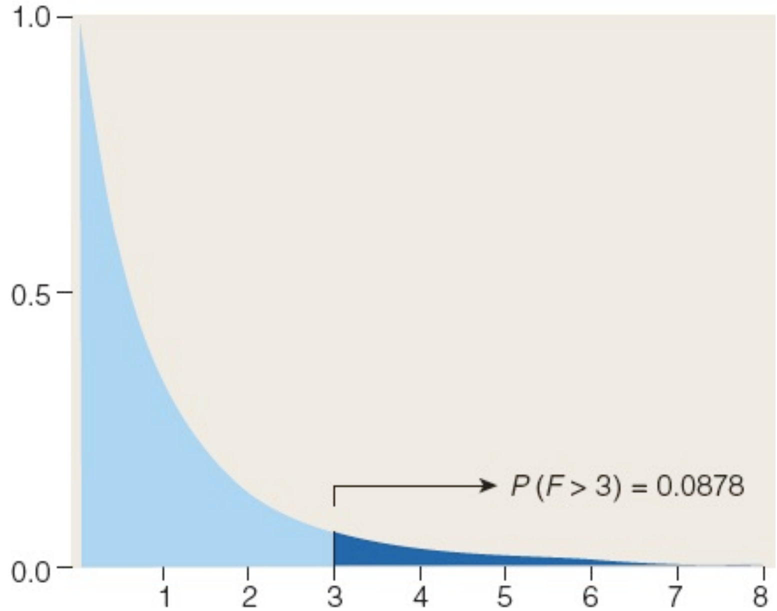
- **Background**: *F*=3 for DFG=4, DFE=385:



- **Questions:** Is *F*=3 large? Will we reject a claim that the 5 population means are equal?
- **Responses:** *P*-val= 0.0185 →Very little area past *F*=3 → *F* is ______Reject claim that 5 population means are equal?____

Example: *Assessing F for Different DF*

- **Background**: F=3 for DFG=2, DFE=12



- **Questions:** Is F=3 large?
  What would we conclude if F=2 for DFG=2, DFE=12?
- **Responses:** P-val=0.0878→F=3 is ____________________
  P-val for F=2 must be ________________ Reject H_0? _____ Conclude population means may be equal? _____

Practice: 11.29h p.561

The *F* Statistic *(Review)*

$$F = \frac{\left[n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_I(\bar{x}_I - \bar{x})^2\right]/(I-1)}{\left[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2\right]/(N-I)}$$

$$= \frac{MSG}{MSE} = \frac{5}{2.5} = 2$$

Is 2 large for DFG=2, DFE=12?

NO

measures difference among sample means
(relative to spreads and sample sizes)

If *F* is large, reject $H_o : \mu_1 = \mu_2 = \mu_3$

Conclude population means differ.

Example: *Drawing Conclusions Based on F*

- **Background**: Earnings for 5 sampled individuals from three racial/ethnic groups had means 3, 4, 5 (in thousands of dollars). ANOVA procedure resulted in $F=2$, which in this case is **not** large.
- **Question:** What do we conclude about mean earnings for populations in the three racial/ethnic groups?
- **Response:** Since *F* is not large, sample means ________ differ significantly from one another. Conclude population mean earnings ______________

Practice: 11.24j-k p.560

Example: *Role of n in ANOVA Test*

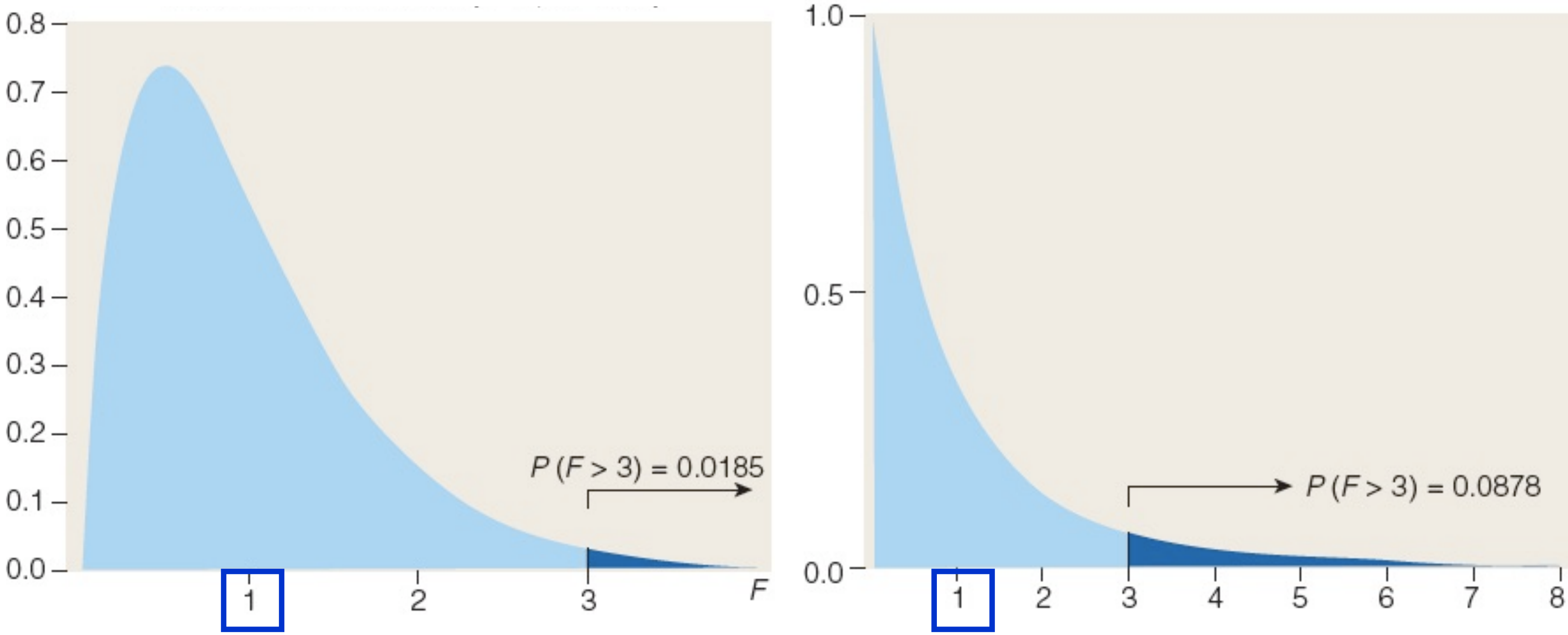
- **Background**: Earnings for **12** (instead of 5) sampled individuals from three racial/ethnic groups had means 3, 4, 5 (in thousands of dollars). ANOVA procedure resulted in ***F*=4.8**, and a *P*-value of 0.015.
- **Question:** What do we conclude about mean earnings for populations in the three racial/ethnic groups?
- **Response:** Conclude population mean earnings for the three groups are ____________

______samples help provide more evidence against Ho.

Practice: 11.29i p.561

Mean of *F*

Since *t* has s.d.= typical distance of values from 0 = approximately 1, and *F* is similar to squaring *t* distribution, mean of *F* is approximately 1.



Example: *Test Relationship/Parameters (Review)*

- **Background**: Research question: "For all students at a university, are Math SATs related to what year they're in?"
- **Question:** How can the question be reformulated in terms of relevant parameters (means) instead of in terms of whether or not the variables are related?
- **Response:**

Example: *Testing Relationship or Parameters*

- **Background**: Research question: "Do mean earnings differ significantly for three racial/ethnic groups?"
- **Question:** How can the question be reformulated in terms of relevant variables instead of in terms of whether or not the means are equal?
- **Response:**

Practice: 11.30j p.562

Inference Methods for C→Q *(Review)*

- Paired: reduces to 1-sample *t*
  - Focused on mean of differences
- Two-Sample: 2-sample *t* (similar to 1-sample *t*)
  - Focused on difference between means
- Several-Sample: *F* distribution
  - Focus on difference among means

Inference for Relationship *(Review)*

- H_0 and H_a about variables: not related or related
  - Applies to all three C→Q, C→C, Q→Q
- H_0 and H_a about parameters: equality or not
  - C→Q: pop means equal? (or mean=0? for paired)
  - C→C: pop proportions equal?
  - Q→Q: pop slope equals zero?

ANOVA Null and Alternative Hypotheses

$H_0:$ explanatory C & response Q *not* related

- Equivalently, $H_o: \mu_1 = \mu_2 = \cdots = \mu_I$

(difference among sample means just chance)

$H_a:$ explanatory C & response Q *are* related

- Equivalently, H_a: not all the μ_i are equal

(difference too extreme to be due to chance)

Depending on formulation, the word "not" appears in Ho or Ha.

Example: *How to Refute a Claim about "All"*

- **Background**: Reader asked medical advice columnist: "Dear Doctor, does everyone with Parkinson's disease shake?" and doctor replied: *All patients with Parkinson's disease do not shake.*
- **Question:** Is this what the doctor meant to say?
- **Response:**

 Practice: 11.37e-f p.564

Example: *ANOVA Alternative Hypothesis*

- **Background**: Null hypothesis to test for relationship between race (3 groups) and earnings:

$$H_o : \mu_1 = \mu_2 = \mu_3$$

- **Question:** Is this the correct alternative?

$$H_a : \mu_1 \neq \mu_2 \neq \mu_3$$

- **Response:**

Words are better: say "____".

Practice: 11.37b p.564

Questions about Practice Midterm 2?

Lecture Summary
(Inference for Cat & Quan: ANOVA)

- Several-sample vs. 2-sample design
  - Notation
  - Compare and contrast *t* and *F* statistics
  - What makes *t* or *F* large?
- Variation among means or within groups; *F* as ratio of variations
- How large is "large" *F*?
  - *F* degrees of freedom
  - *F* distribution
- Role of sample size
- Formulating hypotheses correctly