Class 25

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

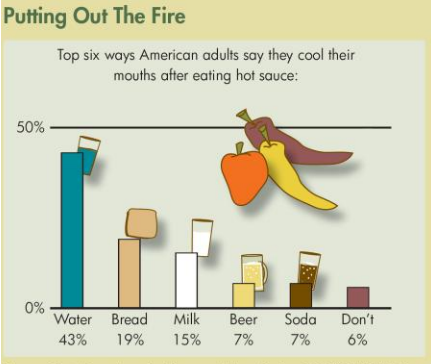
Recap Chapter 11.1-11.3

Lecture Chapter 12.1

Recap Chapter 11.1-11.3

11: Applications of Chi-Square

11.1 Chi-Square Statistic Cooling a Great Hot Taste



Quite often we have qualitative data in categories.

Example: Cooling mouth after hot spicy food.

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

11: Applications of Chi-Square

11.1 Chi-Square Statistic Data Setup

Example: Cooling mouth after hot spicy food.

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

Data set up: k cells $C_1,\ldots,C_k$ that n observations sorted into
Observed frequencies in each cell $O_1,\ldots,O_k$. $O_1+\ldots+O_k=n$
Expected frequencies in each cell $E_1,\ldots,E_k$. $E_1+\ldots+E_k=n$

Cell	C_1	C_2	.	.	.	C_k
Observed	O_1	O_2	.	.	.	O_k
Expected	E_1	E_2	.	.	.	E_k

11: Applications of Chi-Square

11.1 Chi-Square Statistic Data Setup

Example: Cooling mouth after hot spicy food.

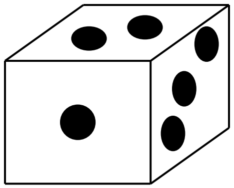
Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

Data set up: k cells $C_1,\ldots,C_k$ that n observations sorted into
Observed frequencies in each cell $O_1,\ldots,O_k$. $O_1+\ldots+O_k=n$
Expected frequencies in each cell $E_1,\ldots,E_k$. $E_1+\ldots+E_k=n$

Cell	C_1	C_2	.	.	.	C_k
Observed	O_1	O_2	.	.	.	O_k
Expected	E_1	E_2	.	.	.	E_k

11: Applications of Chi-Square

11.2 Inferences Concerning Multinomial Experiments



Example: We work for Las Vegas Gaming Commission. We have received many complaints that dice at particular casino are loaded (weighted, not fair). We confiscate all dice and test one. We roll it n=60 times. We get following data.

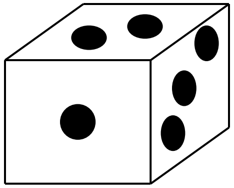
Cell, i	1	2	3	4	5	6
Observed, O_i	7	12	10	12	8	11
Expected, E_i	10	10	10	10	10	10

Expected Value for Multinomial Experiment:

$$E_i = np_i \tag{11.3}$$

11: Applications of Chi-Square

11.2 Inferences Concerning Multinomial Experiments



Example: We work for Las Vegas Gaming Commission. We have received many complaints that dice at particular casino are loaded (weighted, not fair). We confiscate all dice and test one. We roll it n=60 times. We get following data.

Cell, i	1	2	3	4	5	6
Observed, O_i	7	12	10	12	8	11
Expected, E_i	10	10	10	10	10	10

Expected Value for Multinomial Experiment:

$$E_i = np_i \qquad (11.3)$$

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

Is "Preference for math-science, social science, or humanities" … "independent of the gender of a college student?"

Sample Results for Gender and Subject Preference

Gender	Favorite Subject Area: Math–Science (MS)	Social Science (SS)	Humanities (H)	Total
Male (M)	37	41	44	122
Female (F)	35	72	71	178
Total	72	113	115	300

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

Is "Preference for math-science, social science, or humanities" … "independent of the gender of a college student?"

There is a Hypothesis test (of independence) to determine this.

Is Favorite Subject independent of Gender.

Similar to Example with the Die, now have rows i and columns j.

$$\chi^2{*} = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Observed values, O_{ij}'s.

	Favorite Subject Area			
Gender	Math–Science (MS)	Social Science (SS)	Humanities (H)	Total
Male (M)	37	41	44	122
Female (F)	35	72	71	178
Total	72	113	115	300

Figure from Johnson & Kuby, 2012.

What are E_{ij}'s?

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

$$\chi^{2*} = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

D of F for Contingency Tables:

$$df = (r-1)(c-1) \qquad (11.4)$$

$r>1, c>1$

Expected Frequencies for Contingency Tables

$$E_{ij} = \frac{row\ total \times column\ total}{grand\ total} = \frac{R_i C_j}{n} \qquad (11.5)$$

Where does this formula for E_{ij}'s come from?

rows i and columns j

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

$$E_{ij} = \frac{R_i C_j}{n}$$

Where does this formula for E_{ij}'s come from?

	Favorite Subject Area			
Gender	MS	SS	H	Total
Male	37 (29.28)	41 (45.95)	44 (46.77)	122
Female	35 (42.72)	72 (67.05)	71 (68.23)	178
Total	72	113	115	300

r=2
c=3

If Favorite Subject is independent of Gender, then

$$\chi^2* = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} < \chi^2(2, 0.05)$$

α=0.05

$$df = (r-1)(c-1) = (2-1)(3-1)$$

$$\chi^2* = 4.604 < \chi^2(2, 0.05) = 5.99$$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

$$E_{ij} = \frac{R_i C_j}{n}$$

Expected Frequencies for an $r \times c$ Contingency Table

Row	Column 1	2	. . .	jth column	. . .	c	Total
1	$\frac{R_1 \times C_1}{n}$	$\frac{R_1 \times C_2}{n}$	. . .	$\frac{R_1 \times C_j}{n}$	. . .	$\frac{R_1 \times C_c}{n}$	R_1
2	$\frac{R_2 \times C_1}{n}$						R_2
⋮	⋮			⋮			⋮
ith row	$\frac{R_i \times C_1}{n}$			$\frac{R_i \times C_j}{n}$	. . .		R_i
⋮	⋮			⋮			⋮
r	$\frac{R_r \times C_1}{n}$						
Total	C_1	C_2	. . .	C_j	. . .	. . .	n

$$\chi^2{*} = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \overset{?}{<} \chi^2((r-1)(c-1), \alpha)$$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Homogeneity

$$E_{ij} = \frac{R_i C_j}{n}$$

Is the distribution within all rows the same for all rows?

Residence	Governor's Proposal: Favor	Governor's Proposal: Oppose	Total
Urban	143	57	200
Suburban	98	102	200
Rural	13	87	100
Total	254	246	500

r=3
c=2

α=0.05

$$\chi^2{*} = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \overset{?}{<} \chi^2((r-1)(c-1), \alpha)$$

$$df = (r-1)(c-1) = (3-1)(2-1)$$

Chapter 11: Applications of Chi-Square

Questions?

Homework: Read Chapter 11

WebAssign

Chapter 11 # 3, 5, 11, 15, 21, 49, 53

Chapter 12: Analysis of Variance

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical

Lecture Chapter 12.1

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

Previously we learned about hypothesis testing for:
One Population: μ, p, and σ^2.
Two Populations: $\mu_d=\mu_1-\mu_2$, $\mu_1-\mu_2$, p_1-p_2, and σ_1^2/σ_2^2.

(We also learned about hypothesis testing for contingency tables.)

Now we are going to study hypothesis testing for three or more populations.

Three Populations: at least two of $\mu_1,\mu_2,\mu_3,\ldots$ different.

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

If we are testing for differences in means,
…why are we analyzing variance?

As it turns out, we calculate two variances and take the ratio.

If all the means are truly the same, the two variances will be the same and the ratio will be 1.

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

Example: Hypothesis Test for Three Means, μ_1, μ_2, and μ_3.
It is believed that manufacturing plant temperature affects production rate.

	Temperature Levels		
	Sample from 68°F (i = 1)	Sample from 72°F (i = 2)	Sample from 76°F (i = 3)
	10	7	3
	12	6	3
	10	7	5
	9	8	4
		7	
Column totals	$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
	$k_1 = 4$	$k_2 = 5$	$k_3 = 4$
	μ_1	μ_2	μ_3

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F (i = 1)	Sample from 72°F (i = 2)	Sample from 76°F (i = 3)
10	7	3
12	6	3
10	7	5
9	8	4
	7	
$C_1 = 41$	$C_2 = 35$	$C_3 = 15$
$\bar{x}_1 = 10.25$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 3.75$

We could test all pairs:

$H_0\!:\mu_1=\mu_2$ vs. $H_a\!:\mu_1\neq\mu_2$

$H_0\!:\mu_1=\mu_3$ vs. $H_a\!:\mu_1\neq\mu_3$

$H_0\!:\mu_2=\mu_3$ vs. $H_a\!:\mu_2\neq\mu_3$

But each test increases our Type I error rate, $\alpha=\sum\alpha_i$.

So we would like to perform one single hypothesis test

$H_0\!:\mu_1=\mu_2=\mu_3$ vs. H_a: *at least two μ's are different*

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i=1$)	Sample from 72°F ($i=2$)	Sample from 76°F ($i=3$)
10	7	3
12	6	3
10	7	5
9	8	4
	7	
$C_1 = 41$	$C_2 = 35$	$C_3 = 15$
$\bar{x}_1 = 10.25$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 3.75$

Example: μ_1, μ_2, and μ_3
Let's go through the hypothesis test procedure

Step 1: a) Parameter of Interest
Mean production at 3 temperatures
b) Statement of Hypotheses

$H_0: \mu_1 = \mu_2 = \mu_3$
vs.
H_a: *at least two μ's are different*
(it could be $\mu_1 \neq \mu_2$, or $\mu_1 \neq \mu_3$, or $\mu_2 \neq \mu_3$, or $\mu_1 \neq \mu_2 \neq \mu_3$)

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
10	7	3
12	6	3
10	7	5
9	8	4
	7	
$C_1 = 41$	$C_2 = 35$	$C_3 = 15$
$\bar{x}_1 = 10.25$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 3.75$

Example: μ_1, μ_2, and μ_3
Let's go through the hypothesis test procedure

Step 2: a) Assumptions

Normal independent populations, independent observations from each population, equal variances.

b) Test Statistic

Reject or do not reject, F-distribution and F-statistic.

c) Level of Significance

α=0.05

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F (i = 1)	Sample from 72°F (i = 2)	Sample from 76°F (i = 3)
10	7	3
12	6	3
10	7	5
9	8	4
	7	
$C_1 = 41$	$C_2 = 35$	$C_3 = 15$
$\bar{x}_1 = 10.25$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 3.75$

Example: μ_1, μ_2, and μ_3
Let's go through the hypothesis test procedure

Step 3: a) Sample Information

b) Calculate Test Statistic
Next page

c) Level of Significance
α=0.05

12: Analysis of Variance
12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

$$SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \quad \textbf{(2.8)}$$

Shortcut for Total Sum of Squares

$$SS(\text{total}) = \sum(x^2) - \frac{(\sum x)^2}{n} \quad \textbf{(12.2)}$$

$$\sum x_i = 10+12+10+9+7+6+7+8+7+3+3+5+4 = 91$$

$$\sum x_i^2 = 10^2+12^2+10^2+9^2+7^2+6^2+7^2+8^2+7^2+3^2 +3^2+5^2+4^2$$
$$= 731$$

$$SS(total) = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = 731 - \frac{(91)^2}{13} = 94$$

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

The Total Sum of Squares can be partitioned

$$SS(total) = SS(error) + SS(factor)$$

temperature

Sum of Squares Due to Factor

$$SS(\text{factor}) = \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \frac{C_3^2}{k_3} + \cdots\right) - \frac{(\sum x)^2}{n} \quad \textbf{(12.3)}$$

C_i=column total

k_i=number of replicates

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

The Total Sum of Squares can be partitioned

$$SS(total) = SS(error) + SS(factor)$$

temperature

Sum of Squares Due to Factor

$$SS(\text{factor}) = \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \frac{C_3^2}{k_3} + \cdots\right) - \frac{(\sum x)^2}{n} \qquad \textbf{(12.3)}$$

$$SS(temperature) = \left(\frac{41^2}{4} + \frac{35^2}{5} + \frac{15^2}{4}\right) - \frac{(91)^2}{13} = 84.5$$

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

The Total Sum of Squares can be partitioned

$$SS(total) = SS(error) + SS(factor)$$

temperature

Sum of Squares Due to Error

$$SS(\text{error}) = \sum(x^2) - \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \frac{C_3^2}{k_3} + \cdots\right) \quad \textbf{(12.4)}$$

$$SS(error) = 731.0 - \left(\frac{41^2}{4} + \frac{35^2}{5} + \frac{15^2}{4}\right) = 9.5$$

12: Analysis of Variance
12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

We will write these Sums of Squares in a table

Source	df	SS	MS
Factor		84.5	
Error		9.5	
Total		94.0	

(temperature)

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: $F*$

Need to calculate 3 Sums of Squares

Need the degrees of freedom for the sum of squares

Degrees of Freedom for Factor

$$\text{df(factor)} = c - 1 \quad \textbf{(12.5)}$$

$$df\,(temperature) = 3 - 1 = 2$$

Degrees of Freedom for Total

$$\text{df(total)} = n - 1 \quad \textbf{(12.6)}$$

$$df\,(total) = 13 - 1 = 12$$

Degrees of Freedom for Error

$$\text{df(error)} = n - c \quad \textbf{(12.7)}$$

$$df\,(error) = 13 - 3 = 10$$

12: Analysis of Variance
12.1 Intro to ANOVA

	Temperature Levels	
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: $F*$

Need to calculate 3 Sums of Squares

The degrees of freedom and SS must add up

$$SS(\text{factor}) + SS(\text{error}) = SS(\text{total}) \qquad \textbf{(12.8)}$$

and

$$\text{df}(\text{factor}) + \text{df}(\text{error}) = \text{df}(\text{total}) \qquad \textbf{(12.9)}$$

$$SS: 84.5 + 9.5 = 94.0$$

$$df: 10 - 2 = 12$$

$$SS(temperature) = 84.5$$
$$SS(error) = 9.5$$
$$SS(total) = 94.0$$

$$df(temperature) = 2$$
$$df(error) = 10$$
$$df(total) = 12$$

12: Analysis of Variance
12.1 Intro to ANOVA

	Temperature Levels	
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: $F*$

Need to calculate 3 Sums of Squares

Need to divide SS's by df's to get mean squares

Mean Square for Factor

$$MS(\text{factor}) = \frac{SS(\text{factor})}{df(\text{factor})} \tag{12.10}$$

$$SS(temperature) = 84.5$$

$$df(temperature) = 2$$

Mean Square for Error

$$MS(\text{error}) = \frac{SS(\text{error})}{df(\text{error})} \tag{12.11}$$

$$SS(error) = 9.5 \qquad SS(total) = 94.0$$

$$df(error) = 10 \qquad df(total) = 12$$

12: Analysis of Variance
12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F (i = 1)	Sample from 72°F (i = 2)	Sample from 76°F (i = 3)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

We then complete the ANOVA table

Source	df	*SS*	MS
Temperature	2	84.5	42.25
Error	10	9.5	0.95
Total	12	94.0	

$$\text{MS(factor)} = \frac{SS\text{(factor)}}{\text{df(factor)}}$$

$$\text{MS(error)} = \frac{SS\text{(error)}}{\text{df(error)}}$$

$$\text{df(factor)} = c - 1$$

$$\text{df(error)} = n - c$$

$$\text{df(total)} = n - 1$$

12: Analysis of Variance

12.1 Intro to ANOVA

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$
$k_1 = 4$	$k_2 = 5$	$k_3 = 4$

Step 3: b) Calculate Test Statistic: F^*

Need to calculate 3 Sums of Squares

We now can calculate the F statistic

Test Statistic for ANOVA

$$F\star = \frac{\text{MS(factor)}}{\text{MS(error)}} \quad \textbf{(12.12)}$$

$$\text{MS(factor)} = \frac{SS\text{(factor)}}{\text{df(factor)}}$$

$$\text{MS(error)} = \frac{SS\text{(error)}}{\text{df(error)}}$$

$$F^* = \frac{MS(temperature)}{MS(error)} = \frac{42.25}{0.95} = 44.47$$

$\text{df(factor)} = c - 1$

$\text{df(error)} = n - c$

$\text{df(total)} = n - 1$

So is 44.47 “large?”

12: Analysis of Variance
12.1 Intro to ANOVA

$$F\star = \frac{MS(\text{factor})}{MS(\text{error})} \quad (12.12)$$

Step 4: Probability Distribution
We need to determine if the F^* statistic is "large" or equivalently if the area to its right is "small."

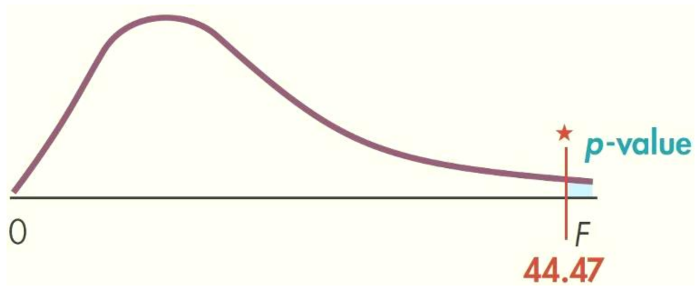
Source	df	SS	MS
Temperature	2	84.5	42.25
Error	10	9.5	0.95
Total	12	94.0	$F^* = 44.47$

12: Analysis of Variance
12.1 Intro to ANOVA

Source	df	SS	MS
Temperature	2	84.5	42.25
Error	10	9.5	0.95
Total	12	94.0	$F^* = 44.47$

Step 4: Probability Distribution: p-value approach α=0.05

$p\text{-value}=P(F^*>44.7|df_n=2,df_d=10)$



$\alpha = 0.01$

Degrees of Freedom for Numerator

Degrees of Freedom for Denominator	1	2	3	4	5	6	7	8	9	10
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6022.	6056.
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30

From Table 9C we see that

$P(F^*>7.56|df_n=2,df_d=10)=0.01$

so $P(F^*>44.7|df_n=2,df_d=10)<0.01$

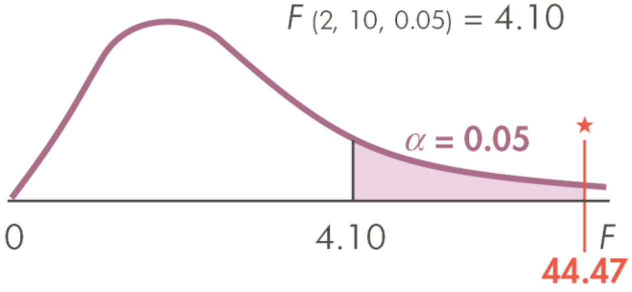
12: Analysis of Variance

12.1 Intro to ANOVA

Source	df	SS	MS
Temperature	2	84.5	42.25
Error	10	9.5	0.95
Total	12	94.0	$F^* = 44.47$

Step 4: Probability Distribution: Classical approach α=0.05

We need to find the F value that has an area of α=0.05 larger than it when df_n=2,df_d=10.



$\alpha = 0.05$

Degrees of Freedom for Numerator (columns); Degrees of Freedom for Denominator (rows)

	1	2	3	4	5	6	7	8	9	10
1	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75

From Table 9A we see that the F critical value is 4.10.

12: Analysis of Variance
12.1 Intro to ANOVA

Source	df	SS	MS
Temperature	2	84.5	42.25
Error	10	9.5	0.95
Total	12	94.0	$F^* = 44.47$

$\alpha=0.05$

Step 5: a) Decision: Reject H_0
p-value approach:
From Table 9C $P(F^*>44.7|df_n=2,df_d=10)<0.01$
But $P(F^*>44.7|df_n=2,df_d=10)>0$.
Classical approach:
From Table 9A we see that $F(2,10,0.05)=4.10$.

b) At least two of the means are statistically different.
At least one of the room temperatures does have a significant effect on the production rate.

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

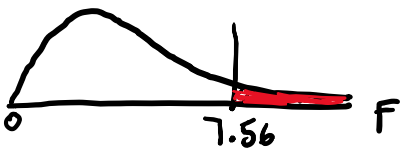
What if we had a different F^*?

Step 3

$$F^* = \frac{4.225}{0.95} = 4.447$$

Step 4

p-value approach

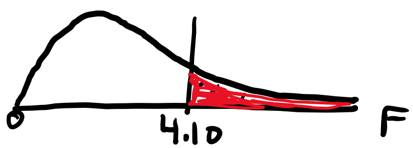


$\alpha = 0.01$

Degrees of Freedom for Numerator

Degrees of Freedom for Denominator	1	2	3	4	5	6	7	8	9	10
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6022.	6056.
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30

p-value approach



$\alpha = 0.05$

Degrees of Freedom for Numerator

Degrees of Freedom for Denominator	1	2	3	4	5	6	7	8	9	10
1	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75

$$.01 < p - value < .05$$

Step 5

Decision: Reject H_0

$$p - value < \alpha$$

$$.01 < p - value < .05$$

$$\alpha = 0.05$$

$$F^* > F_{crit}$$

$$F^* = 4.447$$

Chapter 12: Analysis of Variance

Questions?

Homework: Read Chapter 12

WebAssign

Chapter 12 # 2, 9, 17, 18, 19, 32