ANOVA Examples **STAT 314**

1. If we define $s = \sqrt{MSE}$, then of which parameter is s an estimate?

   If we define $s = \sqrt{MSE}$, then s is an estimate of the common population standard deviation, σ, of the populations under consideration. (This presumes, of course, that the equal-standard-deviations assumption holds.)

2. Explain the reason for the word *variance* in the phrase *analysis of variance*.

   The reason for the word *variance* in the phrase *analysis of variance* is because the analysis of variance procedure for comparing means involves analyzing the ***variation*** in the sample data.

3. The null and alternative hypotheses for a one-way ANOVA test are…

$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_a: \quad \textit{Not all means are equal.}$$

   Suppose in reality that the null hypothesis is false. Does this mean that no two of the populations have the same mean? If not, what does it mean?

   If, in reality, the null hypothesis is false, this translates into "Not ***all*** of the means are the same" or equivalently "***At least two*** of the means are not the same." (Notice that "at least two" applies to 2 or 3 or … or k means.) This last statement in quotes is not equivalent to saying "***No two*** of the populations have the same mean" since this is equivalent to saying, "***All*** of the population means are different."

4. In a one-way ANOVA, identify the statistic used…
   a. as a measure of variation among the sample means.

      MSTr (or *SSTr*) is a statistic that measures the variation among the sample means for a one-way ANOVA.

   b. as a measure of variation within the samples.

      MSE (or *SSE*) is a statistic that measures the variation within the samples for a one-way ANOVA.

   c. to compare the variation among the sample means to the variation within the samples.

      The statistic that compares the variation among the sample means to the variation within the samples is $F = \dfrac{MSTr}{MSE}$.

5. The times required by three workers to perform an assembly-line task were recorded on five randomly selected occasions. Here are the times, to the nearest minute.

Hank	Joseph	Susan
8	8	10
10	9	9
9	9	10
11	8	11
10	10	9

(*Note:* $\bar{y}_1 = 9.6$, $\bar{y}_2 = 8.8$, $\bar{y}_3 = 9.8$, $s_1^2 = 1.3$, $s_2^2 = 0.7$, $s_3^2 = 0.7$, and $\bar{\bar{y}} = 9.4$.) Construct the one-way ANOVA table for the data. Compute *SSTr* and *SSE* using the defining formulas.

$$SSTr = n_1\left(\bar{y}_1 - \bar{\bar{y}}\right)^2 + n_2\left(\bar{y}_2 - \bar{\bar{y}}\right)^2 + n_3\left(\bar{y}_3 - \bar{\bar{y}}\right)^2 = 5(9.6-9.4)^2 + 5(8.8-9.4)^2 + 5(9.8-9.4)^2 = 2.8$$

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 = 4(1.3) + 4(0.7) + 4(0.7) = 10.8$$

$$SSTo = SSTr + SSE = 2.8 + 10.8 = 13.6$$

$$MSTr = \frac{SSTr}{k-1} = \frac{2.8}{3-1} = 1.4; \quad MSE = \frac{SSE}{N-k} = \frac{10.8}{15-3} = 0.9; \quad F = \frac{MSTr}{MSE} = \frac{1.4}{0.9} = 1.5556$$

Source	df	SS	MS = SS/df	F-statistic	*p*-value
Treatments	2	2.8	1.4	1.5556	*p*-value > 0.10
Error	12	10.8	0.9		
Total	14	13.6			

6. Fill in the missing entries of the partially completed one-way ANOVA table.

Source	df	SS	MS = SS/df	F-statistic
Treatments	_____	2.124	0.708	0.75
Error	20	_____	_____	
Total	_____	_____		

$$MSTr = \frac{SSTr}{df_{Tr}} \Rightarrow df_{Tr} = \frac{SSTr}{MSTr} = \frac{2.124}{0.708} = 3$$

$$F = \frac{MSTr}{MSE} \Rightarrow MSE = \frac{MSTr}{F} = \frac{0.708}{0.75} = 0.944$$

$$MSE = \frac{SSE}{df_E} \Rightarrow SSE = MSE * df_E = 0.944(20) = 18.880$$

$$SSTo = SSTr + SSE = 2.124 + 18.880 = 21.004$$

$$df_{To} = df_{Tr} + df_E = 3 + 20 = 23$$

Source	df	SS	MS = SS/df	F-statistic
Treatments	***3***	2.124	0.708	0.75
Error	20	***18.880***	***0.944***	
Total	***23***	***21.004***		

7. Data on Scholastic Aptitude Test (SAT) scores are published by the College Entrance Examination Board in *National College-Bound Senior*. SAT scores for randomly selected students from each of four high-school rank categories are displayed in the following table.

Top Tenth	Second Tenth	Second Fifth	Third Fifth
528	514	649	372
586	457	506	440
680	521	556	495
718	370	413	321
	532	470	424
			330

(*Note:* $\bar{y}_1 = 628.0$, $\bar{y}_2 = 478.8$, $\bar{y}_3 = 518.8$, $\bar{y}_4 = 397.0$, $s_1^2 = 7522.667$, $s_2^2 = 4540.700$, $s_3^2 = 8018.700$, $s_4^2 = 4614.400$, and $\bar{y}_{\bullet\bullet} = 494.1$.) Construct the one-way ANOVA table for the data. Compute *SSTr* and *SSE* using the defining formulas.

$$SSTr = n_1\left(\bar{y}_1 - \bar{\bar{y}}\right)^2 + n_2\left(\bar{y}_2 - \bar{\bar{y}}\right)^2 + n_3\left(\bar{y}_3 - \bar{\bar{y}}\right)^2 + n_4\left(\bar{y}_4 - \bar{\bar{y}}\right)^2$$

$$= 4(628.0 - 494.1)^2 + 5(478.8 - 494.1)^2 + 5(518.8 - 494.1)^2 + 6(397.0 - 494.1)^2 = 132{,}508.2$$

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 + (n_4 - 1)s_4^2$$

$$= 3(7522.667) + 4(4540.700) + 4(8018.700) + 5(4614.400) = 95{,}877.6$$

$$SSTo = SSTr + SSE = 132{,}508.2 + 95{,}877.6 = 228{,}385.8$$

$$MSTr = \frac{SSTr}{k-1} = \frac{132{,}508.2}{4-1} = 44{,}169.40; \quad MSE = \frac{SSE}{N-k} = \frac{95{,}877.6}{20-4} = 5{,}992.35;$$

$$F = \frac{MSTr}{MSE} = \frac{44{,}169.40}{5{,}992.35} = 7.37$$

Source	df	SS	MS = SS/df	F-statistic	*p*-value
Treatments	3	132,508.2	44,169.40	7.37	0.001 < *p*-value < 0.01
Error	16	95,877.6	5,992.35		
Total	19	228,385.8			

8. Four brands of flashlight batteries are to be compared by testing each brand in five flashlights. Twenty flashlights are randomly selected and divided randomly into four groups of five flashlights each. Then each group of flashlights uses a different brand of battery. The lifetimes of the batteries, to the nearest hour, are as follows.

Brand *A*	Brand *B*	Brand *C*	Brand *D*
42	28	24	20
30	36	36	32
39	31	28	38
28	32	28	28
29	27	33	25

Preliminary data analyses indicate that the independent samples come from normal populations with equal standard deviations. At the 5% significance level, does there appear to be a difference in mean lifetime among the four brands of batteries?

Let the subscripts 1, 2, 3, and 4 refer to Brand *A*, Brand *B*, Brand *C*, and Brand *D*, respectively. Each sample size is 5, and the total number of pieces of data is 20.

Step 0: **Check Assumptions**

Step 1: **Hypotheses**

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ (*The mean lifetimes are equal.*)

H_a: *Not all of the means are equal.*

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value and Rejection Region**

$$F_{\alpha,(df_1=k-1, df_2=N-k)} = F_{0.05,(df_1=4-1, df_2=20-4)} = F_{0.05,(df_1=3, df_2=16)} = 3.24$$

Reject the null hypothesis if $F \geq 3.24$ (P-value ≤ 0.05).

Step 4: **Construct the One-way ANOVA Table**

$T_1 = 168$; $T_2 = 154$; $T_3 = 149$; $T_4 = 143$

$$T = 614;\ \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = \sum_{j=1}^{n_i} y_{1j}^2 + \sum_{j=1}^{n_i} y_{2j}^2 + \sum_{j=1}^{n_i} y_{3j}^2 + \sum_{j=1}^{n_i} y_{4j}^2 = 19{,}410$$

$$SSTo = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{(T)^2}{N} = 19{,}410 - \frac{(614)^2}{20} = 560.2$$

$$SSTr = \left(\frac{(T_1)^2}{n_1} + \frac{(T_2)^2}{n_2} + \frac{(T_3)^2}{n_3} + \frac{(T_4)^2}{n_4}\right) - \frac{(T)^2}{N}$$

$$= \left(\frac{(168)^2}{5} + \frac{(154)^2}{5} + \frac{(149)^2}{5} + \frac{(143)^2}{5}\right) - \frac{(614)^2}{20} = 68.2$$

$$SSE = SSTo - SSTr = 560.2 - 68.2 = 492.0$$

Source	df	SS	MS = SS/df	F-statistic	*p*-value
Treatments	3	68.2	22.7333	0.7393	*p*-value > 0.10
Error	16	492.0	30.75		
Total	19	560.2			

Step 5: **Decision**

Since 0.7393 < 3.24 (*p*-value > 0.05), fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there is not enough evidence to conclude that the mean lifetimes of the brands of batteries differ.

9. Manufacturers of golf balls always seem to be claiming that their ball goes the farthest. A writer for a sports magazine decided to conduct an impartial test. She randomly selected 20 golf professionals and then randomly assigned four golfers to each of five brands. Each golfer drove the assigned brand of ball. The driving distances, in yards, are displayed in the following table.

Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
286	279	270	284	281
276	277	262	271	293
281	284	277	269	276
274	288	280	275	292

Preliminary data analyses indicate that the independent samples come from normal populations with equal standard deviations. Do the data provide sufficient evidence to conclude that a difference exists in mean weekly earnings among nonsupervisory workers in the five industries? Perform the required hypothesis test using $\alpha = 0.05$. (*Note:* $T_1 = 1{,}117$, $T_2 = 1{,}128$, $T_3 = 1{,}089$, $T_4 = 1{,}099$, $T_5 = 1{,}142$, and $\sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}^2 = 1{,}555{,}185$.)

Let the subscripts 1, 2, 3, 4, and 5 refer to Brand 1, Brand 2, Brand 3, Brand 4, and Brand 5 respectively. Each sample size is 4, and the total number of data pieces is 20.

Step 0: **Check Assumptions**

Step 1: **Hypotheses**

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (*The mean driving distances are equal.*)

$H_a:$ *Not all of the means are equal.*

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value and Rejection Region**

$F_{\alpha,(df_1 = k-1, df_2 = N-k)} = F_{0.05,(df_1 = 5-1, df_2 = 20-5)} = F_{0.05,(df_1 = 4, df_2 = 15)} = 3.06$

Reject the null hypothesis if $F \geq 3.06$ (P-value ≤ 0.05).

Step 4: **Construct the One-way ANOVA Table**

$$T = \sum_{i=1}^{5} T_i = 5{,}575$$

$$SSTo = \sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}^2 - \frac{(T)^2}{N} = 1{,}555{,}185 - \frac{(5{,}575)^2}{20} = 1{,}153.75$$

$$SSTr = \left(\frac{(T_1)^2}{n_1} + \frac{(T_2)^2}{n_2} + \frac{(T_3)^2}{n_3} + \frac{(T_4)^2}{n_4} + \frac{(T_5)^2}{n_5}\right) - \frac{(T)^2}{N}$$

$$= \left(\frac{(1{,}117)^2}{4} + \frac{(1{,}128)^2}{4} + \frac{(1{,}089)^2}{4} + \frac{(1{,}099)^2}{4} + \frac{(1{,}142)^2}{4}\right) - \frac{(5{,}575)^2}{20} = 458.5$$

$$SSE = SSTo - SSTr = 1{,}153.75 - 485.5 = 695.25$$

Source	df	SS	MS = SS/df	F-statistic	*p*-value
Treatments	4	458.50	114.625	2.4730	0.05 < *p*-value < 0.10
Error	15	695.25	46.35		
Total	19	1,153.75			

Step 5: **Decision**

Since 2.4730 < 3.06 (p-value > 0.05), fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there is not enough evidence to conclude that the mean driving distances of the brands of golf balls differ.

10. The U.S. Bureau of Prisons publishes data in *Statistical Report* on the times served by prisoners released from federal institutions for the first time. Independent random samples of released prisoners for five different offense categories yielded the following information on time served, in months.

Major	n_i	$\bar{y}_i$	s_i
Counterfeiting	15	14.5	4.5
Drug Laws	17	18.4	3.8
Firearms	12	18.2	4.5
Forgery	10	15.6	3.6
Fraud	11	11.5	4.7

At the 1% significance level, do the data provide sufficient evidence to conclude that a difference exists in mean time served by prisoners among the five offense groups?

Let the subscripts 1, 2, 3, 4, and 5 refer to Counterfeiting, Drug Laws, Firearms, Forgery, and Fraud, respectively. The total number of pieces of data is 65.

Step 0: Check Assumptions
Step 1: Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (*The mean times served are equal.*)

H_a: *Not all of the means are equal.*

Step 2: Significance Level

$\alpha = 0.01$

Step 3: Critical Value and Rejection Region

$F_{\alpha,(df_1=k-1, df_2=N-k)} = F_{0.01,(df_1=5-1, df_2=65-5)} = F_{0.01,(df_1=4, df_2=60)} = 3.65$

Reject the null hypothesis if $F \geq 3.65$ (P-value ≤ 0.01).

Step 4: Construct the One-way ANOVA Table

$T_1 = n_1\bar{y}_1 = 217.5$; $T_2 = n_2\bar{y}_2 = 312.8$; $T_3 = n_3\bar{y}_3 = 218.4$; $T_4 = n_4\bar{y}_4 = 156.0$;

$T_5 = n_5\bar{y}_5 = 126.5$; $T = \sum_{i=1}^{5} T_i = 1{,}031.2$

$$SSTr = \left(\frac{(T_1)^2}{n_1} + \frac{(T_2)^2}{n_2} + \frac{(T_3)^2}{n_3} + \frac{(T_4)^2}{n_4} + \frac{(T_5)^2}{n_5}\right) - \frac{(T)^2}{N}$$

$$= \left(\frac{(217.5)^2}{15} + \frac{(312.8)^2}{17} + \frac{(218.4)^2}{12} + \frac{(156.0)^2}{10} + \frac{(126.5)^2}{11}\right) - \frac{(1{,}031.2)^2}{65} = 412.9086$$

$$SSE = \sum_{i=1}^{5}(n_i - 1)s_i^2 = 14(4.5)^2 + 16(3.8)^2 + 11(4.5)^2 + 9(3.6)^2 + 10(4.7)^2 = 1{,}074.83$$

$$SSTo = SSTr + SSE = 412.909 + 1{,}074.830 = 1{,}487.739$$

Source	df	S S	MS = SS/df	F-statistic	*p*-value
Treatment	4	412.9086	103.22715	5.7656	*p*-value < 0.001
Error	60	1,074.83	17.91383		
Total	64	1,487.7386			

Step 5: Decision

Since 5.7656 ≥ 3.65 (*p*-value ≤ 0.01), reject the null hypothesis.

Step 6: State conclusion in words

At the $\alpha = 0.01$ level of significance, there exists enough evidence to conclude that the mean times served for the crime groups do differ.

11. A study of the effect of different types of anesthesia on the length of post-operative hospital stay yielded the following for cesarean patients:

Group A was given an epidural MS.
Group B was given an epidural.
Group C was given a spinal.
Group D was given general anesthesia.

The data are presented below. In general, the general anesthetic is considered to be the most dangerous, the spinal somewhat less so, and the epidural even less, with MS addition providing additional safety. Note that the data are in the form of distributions for each group.

	Length of Stay	**Number of Patients**
Group A	3	6
	4	14
Group B	4	18
	5	2
Group C	4	10
	5	9
	6	1
Group D	4	8
	5	12

a. Test for the existence of an effect due to anesthesia type. (Use $\alpha = 0.01$)

Step 0: **Check Assumptions**

Step 1: **Hypotheses**

H_0: $\mu_A = \mu_B = \mu_C = \mu_D$

H_a: *at least one inequality*

Step 2: **Significance Level**

$\alpha = 0.01$

Step 3: **Critical Value and Rejection Region**

Critical Value: $F_{\alpha,(df_1=k-1, df_2=N-k)} = F_{0.01,(df_1=3, df_2=76)} \approx F_{0.01,(df_1=3, df_2=60)} = 4.13$

Reject the null hypothesis if $F \geq 4.13$.

Step 4: **Construct the One-way ANOVA Table**

$$\sum_{i=1}^{k}\frac{(T_i)^2}{n_i} = \frac{[(3*6)+(4*14)]^2}{20} + \frac{[(4*18)+(5*2)]^2}{20} + \frac{[(4*10)+(5*9)+(6*1)]^2}{20} + \frac{[(4*8)+(5*12)]^2}{20}$$

$$= \frac{[74]^2}{20} + \frac{[82]^2}{20} + \frac{[91]^2}{20} + \frac{[92]^2}{20} = 1447.25$$

$$\frac{(T)^2}{N} = \frac{[(3*6)+(4*14)+(4*18)+(5*2)+(4*10)+(5*9)+(6*1)+(4*8)+(5*12)]^2}{80}$$

$$= \frac{[(3*6)+(4*50)+(5*23)+(6*1)]^2}{80} = \frac{[339]^2}{80} = 1436.5125$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = (3^2*6)+(4^2*14)+(4^2*18)+(5^2*2)+(4^2*10)+(5^2*9)+(6^2*1) + (4^2*8)+(5^2*12)$$

$$= 1465$$

$$SSTr = \sum_{i=1}^{k} \frac{(T_{i\cdot})^2}{n_i} - \frac{(Y_{\cdot\cdot})^2}{N} = 1447.27 - 1436.5125 = 10.7375$$

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{(T_{i\cdot})^2}{n_i} = 1465 - 1447.25 = 17.75$$

Source	*df*	*SS*	*MS = SS/df*	*F*-statistic	*p*-value
Treatments	3	10.7375	3.5792	15.3219	*p*-value < 0.001
Error	76	17.7500	0.2336		
Total	79	28.4875			

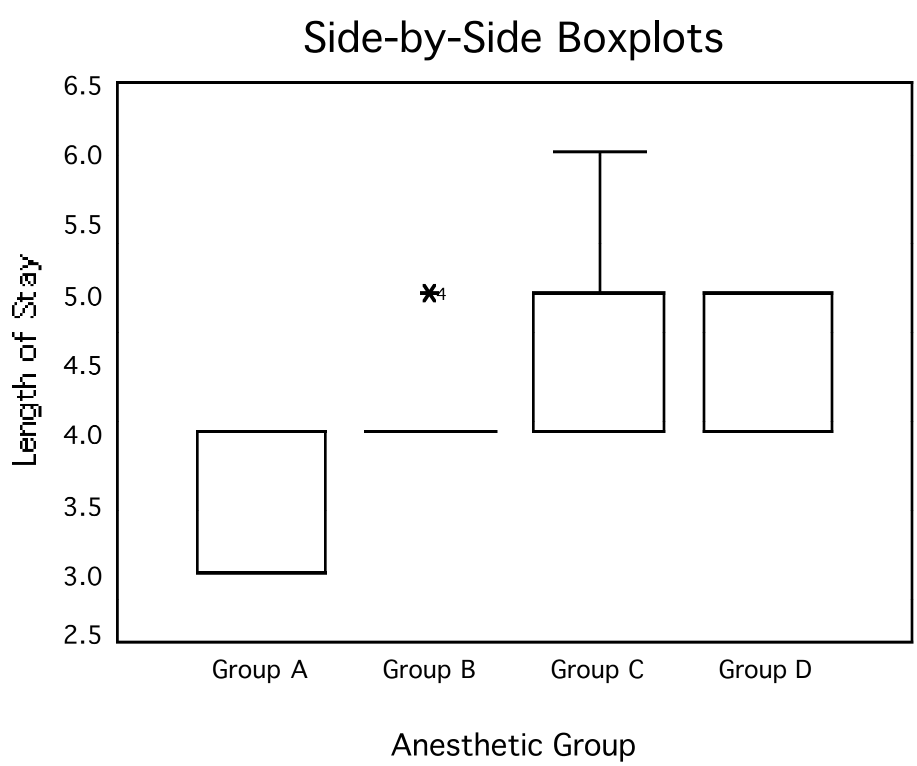
Step 5: **Decision**
Since 15.3219 ≥ 4.13 (*p*-value ≤ 0.01), we shall reject the null hypothesis.

Step 6: **State conclusion in words**
At the $\alpha = 0.01$ level of significance, there exists enough evidence to conclude that there is an effect due to anesthesia type.

b. Does it appear that the assumptions for the analysis of variance are fulfilled? Explain.

From the description of the problem, the samples seem to be independent; however, the variances appear to be a bit different.

c. Construct side-by-side boxplots to check the assumptions. Do the results support your answer in (b)?



The boxplots seem to be similar in shape (except for Group B, whose spread is close to that of the others), so our results are acceptable.

d. What specific recommendations can be made on the basis of these data?

The epidural MS appears to have the shortest average stay, and it is the safest method; therefore, I recommend use of the epidural MS for the patients.

12. Three sets of five mice were randomly selected to be placed in a standard maze but with different color doors. The response is the time required to complete the maze as seen below. Perform the appropriate analysis to test if there is an effect due to door color. (Use $\alpha = 0.01$)

Color	Time				
Red	9	11	10	9	15
Green	20	21	23	17	30
Black	6	5	8	14	7

Step 0: **Check Assumptions**

Step 1: **Hypotheses**

H_0: $\mu_{Red} = \mu_{Green} = \mu_{Black}$

H_a: *at least one inequality*

Step 2: **Significance Level**

$\alpha = 0.01$

Step 3: **Critical Value and Rejection Region**

Critical Value: $F_{\alpha,(df_1=k-1, df_2=N-k)} = F_{0.01,(df_1=2, df_2=12)} = 6.93$

Reject the null hypothesis if $F \geq 6.93$.

Step 4: **Construct the One-way ANOVA Table**

$$\sum_{i=1}^{k}\frac{(T_i)^2}{n_i} = \frac{(9+11+10+9+15)^2}{5} + \frac{(20+21+23+17+30)^2}{5} + \frac{(6+5+8+14+7)^2}{5}$$

$$= \frac{(54)^2}{5} + \frac{(111)^2}{5} + \frac{(40)^2}{5} = 3367.4$$

$$\frac{(T)^2}{N} = \frac{(9+11+10+9+15+20+21+23+17+30+6+5+8+14+7)^2}{15}$$

$$= \frac{(205)^2}{15} = 2801.6667$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = 9^2 + 11^2 + 10^2 + 9^2 + 15^2 + 20^2 + 21^2 + 23^2 + 17^2 + 30^2 + 6^2$$

$$+ 5^2 + 8^2 + 14^2 + 7^2$$

$$= 3537$$

$$SSTr = \sum_{i=1}^{k}\frac{(T_i)^2}{n_i} - \frac{(T)^2}{N} = 3367.4 - 2801.6667 = 565.7333$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k}\frac{(T_i)^2}{n_i} = 3537 - 3367.4 = 169.6$$

Source	*df*	*SS*	*MS = SS/df*	*F*-statistic	*p*-value
Treatments	2	565.7333	282.8667	20.0142	*p*-value < 0.001
Error	12	169.6000	14.1333		
Total	14	735.3333			

Step 5: **Decision**

Since $20.0142 \geq 6.93$ (p-value ≤ 0.01), we shall reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.01$ level of significance, there exists enough evidence to conclude that there is an effect due to door color.

13. The table below shows the observed pollution indexes of air samples in two areas of a city. Test the hypothesis that the mean pollution indexes are the same for the two areas. (Use $\alpha = 0.05$.) Also, verify that $t^2 = F$ for the two-sample pooled case.

Area A		Area B	
2.92	4.69	1.84	3.44
1.88	4.86	0.95	3.69
5.35	5.81	4.26	4.95
3.81	5.55	3.18	4.47

Step 0: Check Assumptions (ANOVA Approach)

Step 1: Hypotheses (ANOVA Approach)

$H_0\!: \mu_A = \mu_B$

$H_a\!: \mu_A \neq \mu_B$

Step 2: Significance Level

$\alpha = 0.05$

Step 3: Critical Value and Rejection Region

Critical Value: $F_{\alpha,(df_1 = k-1, df_2 = N-k)} = F_{0.05,(df_1 = 1, df_2 = 14)} = 4.60$

Reject the null hypothesis if $F \geq 4.60$.

Step 4: Construct the One-way ANOVA Table

$$\sum_{i=1}^{k} \frac{(T_i)^2}{n_i} = \frac{(2.92 + 1.88 + 5.35 + 3.81 + 4.69 + 4.86 + 5.81 + 5.55)^2}{8} + \frac{(1.84 + 0.95 + 4.26 + 3.18 + 3.44 + 3.69 + 4.95 + 4.47)^2}{8}$$

$$= \frac{(34.87)^2}{8} + \frac{(26.78)^2}{8} = 241.6357$$

$$\frac{(T)^2}{N} = \frac{\begin{pmatrix} 2.92 + 1.88 + 5.35 + 3.81 + 4.69 + 4.86 + 5.81 + 5.55 \\ + 1.84 + 0.95 + 4.26 + 3.18 + 3.44 + 3.69 + 4.95 + 4.47 \end{pmatrix}^2}{16}$$

$$= \frac{(61.65)^2}{16} = 237.5452$$

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} y_{ij}^2 = 2.92^2 + 1.88^2 + 5.35^2 + 3.81^2 + 4.69^2 + 4.86^2 + 5.81^2 + 5.55^2 + 1.84^2 + 0.95^2 + 4.26^2 + 3.18^2 + 3.44^2 + 3.69^2 + 4.95^2 + 4.47^2$$

$$= 267.8549$$

$$SSB = \sum_{i=1}^{t} \frac{(Y_{i\bullet})^2}{n_i} - \frac{(Y_{\bullet\bullet})^2}{N} = 241.6357 - 237.5452 = 4.0905$$

$$SSW = \sum_{i=1}^{t} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{t} \frac{(Y_{i\bullet})^2}{n_i} = 267.8549 - 241.6357 = 26.2192$$

Source	*df*	*SS*	*MS = SS/df*	*F*-statistic	*p*-value
Treatments	1	4.0905	4.0905	2.1842	p-value > 0.10
Error	14	26.2192	1.8728		
Total	15	30.3097			

Step 5: Decision

Since $2.1842 < 4.60$ (p-value > 0.05), we fail to reject the null hypothesis.

Step 6: State conclusion in words

At the $\alpha = 0.05$ level of significance, there is not enough evidence to conclude that there is a difference in the average pollution indexes for the two areas.

The table below shows the observed pollution indexes of air samples in two areas of a city. Test the hypothesis that the mean pollution indexes are the same for the two areas. (Use $\alpha = 0.05$.)

Area A		Area B	
2.92	4.69	1.84	3.44
1.88	4.86	0.95	3.69
5.35	5.81	4.26	4.95
3.81	5.55	3.18	4.47

Step 0: **Check Assumptions (Pooled-*t* Approach)**

Step 1: **Hypotheses (Pooled-*t* Approach)**

$H_0\colon \mu_A - \mu_B = 0$

$H_a\colon \mu_A - \mu_B \neq 0$

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value(s) and Rejection Region(s)**

Since we don't know the population variances (σ_A^2 and σ_B^2) but think that they are equal (air comes from areas of the same city), we'll use the pooled *t*-test.

Critical Values: $\pm t_{\alpha/2,\, df = n_A + n_B - 2} = \pm t_{0.025, df=14} = \pm 2.145$

Reject the null hypothesis if $T \leq -2.145$ or if $T \geq 2.145$.

Step 4: **Test Statistic**

$$\bar{y}_A = \frac{\sum y_A}{n_A} = \frac{34.87}{8} = 4.3588 \qquad \bar{y}_B = \frac{\sum y_B}{n_B} = \frac{26.78}{8} = 3.3475$$

$$s_p = \sqrt{\frac{\left(\sum y_A^2 - \frac{\left(\sum y_A\right)^2}{n_A}\right) + \left(\sum y_B^2 - \frac{\left(\sum y_B\right)^2}{n_B}\right)}{n_A + n_B - 2}}$$

$$= \sqrt{\frac{\left(165.3737 - \frac{(34.87)^2}{8}\right) + \left(102.4812 - \frac{(26.78)^2}{8}\right)}{8 + 8 - 2}} = \sqrt{\frac{13.3841 + 12.8352}{14}} = 1.3685$$

$$T = \frac{\left(\bar{y}_A - \bar{y}_B\right) - \delta_0}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{(4.3588 - 3.3475) - 0}{1.3685\sqrt{\frac{1}{8} + \frac{1}{8}}} = 1.4780$$

$p\text{-value} = 2 * P(t \geq 1.4780) \Rightarrow 2 * 0.05 < p\text{-value} < 2 * 0.1 \Rightarrow 0.1 < p\text{-value} < 0.2$

Step 5: **Conclusion**

Since $-2.145 \leq 1.4780 \leq 2.145$ ($0.2 > p$-value $> 0.1 > 0.05 = \alpha$), we fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there is not enough evidence to conclude that the mean pollution indexes are the same for the two areas.

***** Show $t^2 = F$:**

$t = 1.4780 \;\Rightarrow\; t^2 = 2.1845 \approx 2.1842 = F$ (off due to rounding error).

14. A study of firefighters in a large urban area centered on the physical fitness of the engineers employed by the Fire Department. To measure the fitness, a physical therapist sampled five engineers each with 5, 10, 15, and 20 years' experience with the department. She then recorded the number of pushups that each person could do in 60 seconds. The results are listed below. Perform an analysis of variance to determine if there are differences in the physical fitness of engineers by time with department group. Use $\alpha = 0.05$.

Time with Department, years	5	10	15	20
	56	64	45	42
	55	61	46	39
	62	50	45	45
	59	57	39	43
	60	55	43	41

Step 0: **Check Assumptions**

Step 1: **Hypotheses**

H_0: $\mu_5 = \mu_{10} = \mu_{15} = \mu_{20}$

H_a: *at least one inequality*

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value and Rejection Region**

Critical Value: $F_{\alpha,(df_1 = k-1, df_2 = N-k)} = F_{0.05,(df_1 = 3, df_2 = 16)} = 3.24$

Reject the null hypothesis if $F \geq 3.24$.

Step 4: **Construct the One-way ANOVA Table**

$$\sum_{i=1}^{k} \frac{(T_i)^2}{n_i} = \frac{(56+55+62+59+60)^2}{5} + \frac{(64+61+50+57+55)^2}{5} + \frac{(45+46+45+39+43)^2}{5} + \frac{(42+39+45+43+41)^2}{5}$$

$$= \frac{(292)^2}{5} + \frac{(287)^2}{5} + \frac{(218)^2}{5} + \frac{(210)^2}{5} = 51851.4$$

$$\frac{(T)^2}{N} = \frac{\left(\begin{array}{c}56+55+62+59+60+64+61+50+57+55\\ +45+46+45+39+43+42+39+45+43+41\end{array}\right)^2}{20} = \frac{(1007)^2}{20} = 50702.45$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = 56^2 + 55^2 + 62^2 + 59^2 + 60^2 + 64^2 + 61^2 + 50^2 + 57^2 + 55^2 + 45^2 + 46^2 + 45^2 + 39^2 + 43^2 + 42^2 + 39^2 + 45^2 + 43^2 + 41^2 = 52053$$

$$SSTr = \sum_{i=1}^{k} \frac{(T_i)^2}{n_i} - \frac{(T)^2}{N} = 51851.4 - 50702.45 = 1148.95$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{(T_i)^2}{n_i} = 52053 - 51851.4 = 201.6$$

Source	*df*	*SS*	*MS = SS/df*	*F*-statistic	*p*-value
Treatments	3	1148.95	382.9833	30.3955	*p*-value < 0.001
Error	16	201.6	12.600		
Total	19	1350.55			

Step 5: **Decision**

Since $30.3955 \geq 3.24$ (p-value ≤ 0.05), we shall reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there are differences in the physical fitness of engineers by time with department group.

15. A local bank has three branch offices. The bank has a liberal sick leave policy, and a vice-president was concerned with employees taking advantage of this policy. She thought that the tendency to take advantage depended on the branch at which the employee worked. To see if there were differences in the time employees took for sick leave, she asked each branch manager to sample employees randomly and record the number of days of sick leave taken during 1990. Ten employees were chosen, and the data are listed below. Do this data indicate a difference in branches? Use a level of significance of 0.05.

Branch 1	15	20	19	14
Branch 2	11	15	11	
Branch 3	18	19	23	

Step 0: **Check Assumptions**

Step 1: **Hypotheses**

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

H_a: *at least one inequality*

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value and Rejection Region**

Critical Value: $F_{\alpha,(df_1 = k-1, df_2 = N-k)} = F_{0.05,(df_1 = 2, df_2 = 7)} = 4.74$

Reject the null hypothesis if $F \geq 4.74$.

Step 4: **Construct the One-way ANOVA Table**

$$\sum_{i=1}^{k} \frac{(T_i)^2}{n_i} = \frac{(15+20+19+14)^2}{4} + \frac{(11+15+11)^2}{3} + \frac{(18+19+23)^2}{3}$$

$$= \frac{(68)^2}{4} + \frac{(37)^2}{3} + \frac{(60)^2}{3} = 2812.3333$$

$$\frac{(T)^2}{N} = \frac{(15+20+19+14+11+15+11+18+19+23)^2}{10}$$

$$= \frac{(165)^2}{10} = 2722.5$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 = 15^2 + 20^2 + 19^2 + 14^2 + 11^2 + 15^2 + 11^2 + 18^2 + 19^2 + 23^2$$

$$= 2863$$

$$SSTr = \sum_{i=1}^{k} \frac{(T_i)^2}{n_i} - \frac{(T)^2}{N} = 2812.3333 - 2722.5 = 89.8333$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{(T_i)^2}{n_i} = 2863 - 2812.3333 = 50.6667$$

Source	*df*	*SS*	*MS = SS/df*	*F*-statistic	*p*-value
Treatments	2	89.8333	44.9167	6.2056	0.01 < *p*-value < 0.05
Error	7	50.6667	7.2381		
Total	9	140.5000			

Step 5: **Decision**

Since $6.2056 \geq 4.74$ (p-value ≤ 0.05), we shall reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there are differences in the branches regarding sick leave taken during 1990.