Brock Francom
A02052161

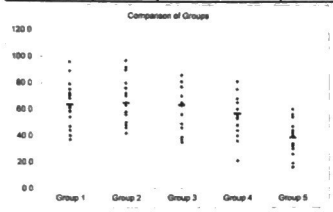
Homework ANOVA

Instructions: Assume the conditions are met unless you are told otherwise.

1. We want to know if the reproduction level is related to the life expectancy of male fruit flies. Independent random samples were taken for male fruit flies in different levels of reproduction. Conduct an ANOVA test with $\alpha = .05$ to determine if the population means are significantly different for the different levels of reproduction.

Mean	n	Std. Dev	
63.6	25	16.45	Group 1
64.8	25	15.65	Group 2
63.4	25	14.54	Group 3
56.8	25	14.93	Group 4
38.7	25	12.10	Group 5
57.4	125	17.56	Total

Source	SS	df	MS	F	p-value
Treatment	11,939.28	4	2,984.820	13.61	3.52E-09
Error	26,313.52	120	219.279		
Total	38,252.80	124			



(a) Conditions:

i. I found the normal curve plots for the groups to make sure that they are normal enough to do ANOVA and they are.

ii. Are all of the standard deviations close enough to each other to use ANOVA?

yes 12.1 - 17.56 less than double.

iii. Are the samples independent?

yes

(b) What are the null and alternative hypotheses?

H_0: all pop. μ's are equal

H_A: at least one pop μ different

(c) What is the test statistic?

$F = 13.61$

(d) What is the p-value?

$Pval = 3.52 E^{-09}$

(e) Should you reject, or fail to reject the null hypothesis?

small pval, reject null

(f) What is your conclusion?

We have strong evidence that at least one population mean is different from the others.

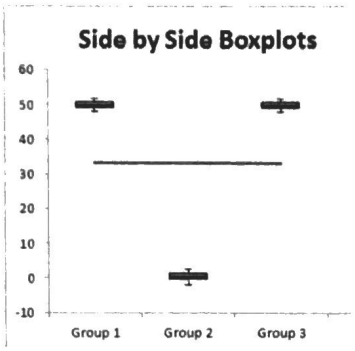
(g) Find and interpret R^2.

$R^2 = \frac{11939.28}{38252.8} = .312$

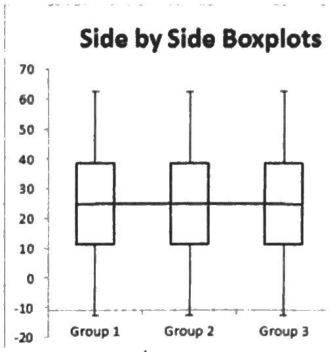
31.2% of total variation can be explained by which group they are in.

2. For each set of boxplots, rank the three boxplots from most likely to least likely to have Equal population means.

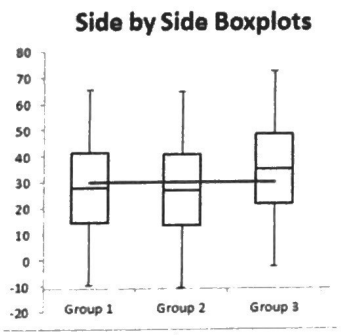
(a)



least

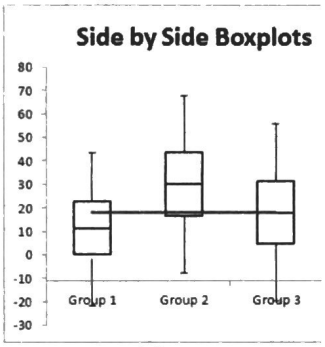


most

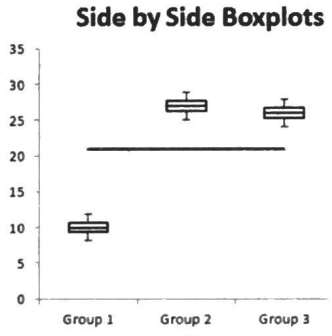


middle.

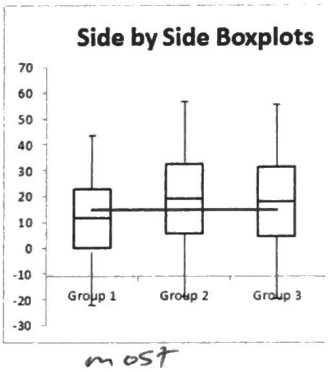
(b)



middle

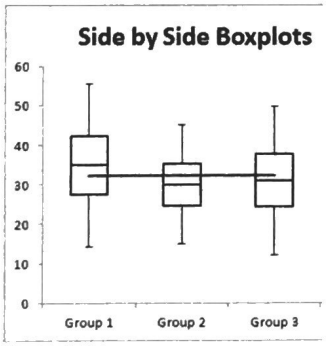


least

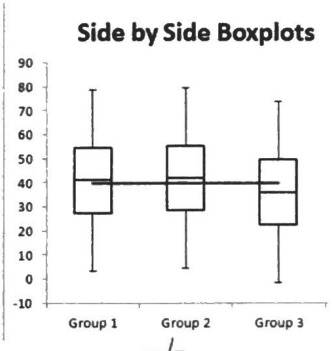


most

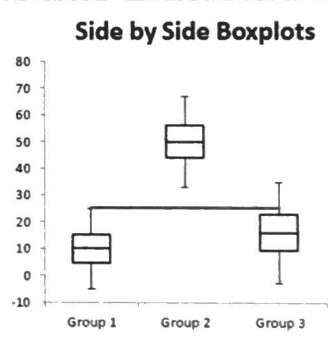
(c)



middle



most



least.

3. We want to compare the mean monthly sales when using the bottom, middle, and top shelf display heights.

One factor ANOVA

Mean	n	Std. Dev	
55.80	6	2.477	Bottom
77.20	6	3.103	Middle
51.50	6	1.648	Top
61.50	18	11.798	Total

ANOVA table

Source	SS	df	MS	F	p-value
Treatment	2,273.880	2	1,136.9400	184.57	2.74E-11
Error	92.400	15	6.1600		
Total	2,366.280	17			

(a) What is the sample mean for the bottom shelves?

$\bar{x}_{Bot} = 55.8$

(b) What is the sample size for the Top shelves?

$n_{top} = 6$

(c) What is the overall or total sample mean for all the shelves combined?

$\bar{x} = 61.5$

(d) What is the Sum of Squares for the Error?

92.4

(e) What is the Mean Square for the Treatment or Group?

1136.94

(f) What is the Total degrees of freedom?

17

(g) What is the F statistic?

$F = 184.57$

(h) Test the null hypothesis that the true mean monthly sales are the same for all display heights. Use $\alpha = .05$. What are the appropriate null and alternative hypotheses?

cond. met
$\alpha = .05$
H_0: All means same
H_A: at least one different
$F = 184.57$,

→ Pval = $2.74E^{-11}$
Small, reject null.
we found very strong evidence that at least one population mean is different from the others.

(i) State your conclusion in practical terms (as if you were explaining your conclusion to a boss who had never taken a statistics course.)

we found evidence that there are differences between shelf averages

(j) Find and interpret R^2.

$R^2 = \frac{2273.88}{2366.28} = .961$

96.1% of variation can be explained by which group they were in.

4. We want to see if there is a difference in the mean durability of four different brands of golf balls. Conduct an ANOVA test with $\alpha = .05$.

One factor ANOVA

Mean	n	Std. Dev	
253.6	5	24.68	Alpha
306.4	5	27.21	Best
241.8	5	21.67	Century
336.6	5	24.60	Divot
284.6	20	45.63	Total

ANOVA table

Source	SS	df	MS	F	p-value
Treatment	29,860.40	3	9,953.467	16.42	3.85E-05
Error	9,698.40	16	606.150		
Total	39,558.80	19			

(a) What is the degrees of freedom for the error?

16

(b) What is the Mean Square for Treatment or Groups?

9953.467

(c) What is the F statistic?

16.42

(d) What would be the appropriate null and alternative hypotheses for this ANOVA test?

H_0: all means same

H_A: at least 1 different

(e) Conduct the ANOVA test. What is your conclusion?

Conditions met ✓ → pval = 3.85E-05

$\alpha = .05$

Small, reject null

F = 16.42

We found strong evidence that at least 1 population mean is different from the others.

(f) Find and interpret R^2.

$R^2 = \frac{29860.4}{39558.8} = .7548$

75.48% of variation can be explained by which group they were in.

5. **NOTE:** In all of our hypothesis tests we first calculate a test statistic and then find the area in the tail as the p-value. So the point of a test statistic whether it is Z, T, or F (depending on which test you do) is to let you find the p-value.

- For ANOVA, we need to compare the *variation between the groups* to the *variation within the groups*.
- So we use the F test statistic:

$$F = \frac{\text{mean square group}}{\text{mean square error}}$$

- If the population means are different, what kind of F values will you expect?
- You'd expect to see *more variation between* the groups if the population means are different.
- So you should get big values on the top.
- So bigger values of F give more evidence that the population means are different.
- So the p-value is always the area to the right of the F test statistic.

6. A college has 5 different professors that teach statistics. The dean wants to know if the different professors courses have the same true population mean final score, or if there is a difference in the population means across their courses. He takes a sample from each professors courses.

(a) Which statistical test would be appropriate?

ANOVA!

One factor ANOVA

Mean	n	Std. Dev	
80.59367	44	11.638343	Course 1
80.63797	42	11.568092	Course 2
76.84809	40	10.894762	Course 3
83.01148	44	11.987494	Course 4
79.75520	43	11.492241	Course 5
80.22920	213	11.591860	Total

ANOVA table

Source	SS	df	MS	F	p-value
Treatment	820.406098	4	205.1015245	1.54	.1913
Error	27,666.290873	208	133.0110138		
Total	28,486.696971	212			

(b) What are the appropriate null and alternative hypotheses?

H_0: all means same

H_A: at least 1 different

(c) What is the F statistic?

$F = 1.54$

(d) If I don't tell you a significance level to use, what would be an acceptable choice?

$\alpha = .05$

(e) What is the p-value?

.1913

(f) Should you reject or fail to reject the null hypothesis.

fail to reject.

(g) State your conclusion in practical terms.

We don't have evidence that there are differences between courses mean grades.

(h) Find and interpret R^2.

$R^2 = \frac{820.4}{28486.7} = .02879$

2.879% of variation can be explained by which group they are in.

7. We want to compare the pricing for the internet, television, and telephone bundles for different types of providers. The summary information is

Group	n	$\overline{x}$	s
DSL	19	104.49	26.09
Cable	20	119.98	40.39
Fiber	8	83.87	31.78

The ANOVA table is

	Degrees of Freedom	Sum of Squares	Mean Squares	F	P-value
Group	2	7753.92	3876.96	3.39	0.0427
Error	44	50317.85	1143.59		
Total	46	58071.78			

(a) What is the degrees of freedom for the error?

44

(b) What is the Sum of Squares (Total)?

58071.78

(c) What is the Mean Square Group?

3876.96

(d) What are the null and alternative hypotheses?

H_0: all means are same

H_A: At least 1 different

(e) What is the test statistic F?

$F = 3.39$

(f) What is the p-value?

.0427

(g) Use $\alpha = .05$. Should you reject or fail to reject the null hypothesis?

reject

(h) What is your conclusion?

We found evidence that at least 1 population mean is different from the others.

8. We want to compare the effects of different music genre on how much people are willing to spend on restaurants. Someone took a three independent random samples of diners at a expensive restaurant in England. The summary information is

Music Type	n	$\overline{x}$	s
Classical	44	24.13	2.243
Pop	49	21.912	2.627
None	48	21.697	3.332

The ANOVA table is

	Degrees of Freedom	Sum of Squares	Mean Squares	F	P-value
Group	2	164.66	82.33	10.62	5.11×10^{-5}
Error	138	1069.39	7.75		
Total	140	1234.05			

(a) What is the degrees of freedom for the group?

2

(b) What is the Sum of Squares for the error?

1069.39

(c) What is the Mean Square Group?

82.33

(d) What are the null and alternative hypotheses?

H_0: all means equal

H_A: at least one different.

(e) What is the test statistic F?

$F = 10.62$

(f) What is the p-value?

5.11×10^{-5}

(g) Use $\alpha = .05$. Should you reject or fail to reject the null hypothesis?

reject

(h) What is your conclusion?

We found strong evidence that, at least 1 population mean is different from the others.