A. Notebook Examples

1. (Example 1 Case 2)

(a) P-value $= 1 - 0.822021 = \boxed{0.177979}$

(b) 0.396569

2. (Example 1 Case 3)

(a)

$$s_1^2 = 37$$
$$s_2^2 = 30.5$$
$$s_3^2 = 62.5$$

$$s_p^2 = \text{average}(37, 30.5, 62.5) = \boxed{43.33}$$

(b)

$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{500}{43.33} = \boxed{11.54}$$

(c) The variation between sample means is 11.54 times as large as expected, if the null hypothesis is true.

(d)

$$1 - 0.998398 = \boxed{0.001602}$$

(e)
- $F = 11.54$
- P-value $= 0.002$
- Almost! For the Notes Prof. Whitten used additional MINITAB graphing options to make small changes such as “jiggling” the points horizontally so that they’re not obscured by the sample means.

3. (Example 2 Case 1)

(a)

$$\text{MSG} = \frac{\text{SSG}}{\text{DFG}} = \frac{206}{2} = 103$$

$$\text{MSE} = \frac{\text{SSE}}{\text{DFE}} = \frac{7006}{13} = 538.9230769$$

$$F = \frac{\text{MSG}}{\text{MSE}} = \boxed{0.191121896}$$

(b) P-value $= 0.828312$

(c) $F = 0.19$, P-value $= 0.829$

B. Textbook

1. (Pretest)

(a)

1. μ_1 = mean pretest score of all children taught with Basal
   $\mu_2 = \ldots\ldots$ DRTA
   $\mu_3 = \ldots\ldots$ Strat

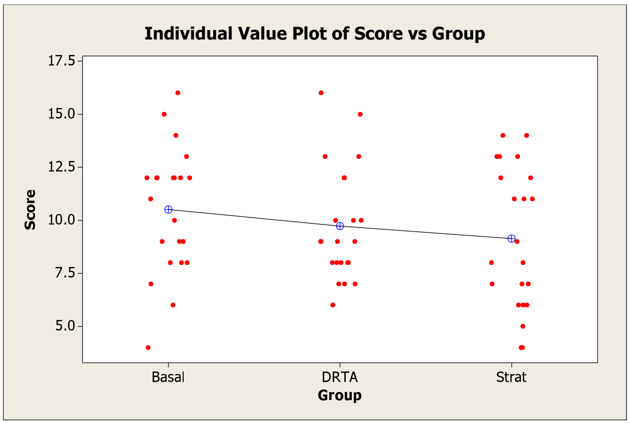
$$H_0\colon \mu_1 = \mu_2 = \mu_3$$
$$H_A\colon \text{At least one } \mu_i \text{ differs from the others}$$

2. From MINITAB: $F = \boxed{1.13}$, P-value $= \boxed{0.329}$

3. Fail to Reject H_0 since P-value $= 0.329 > .05 = \alpha$
4. There is insufficient evidence to show any difference in mean pretest scores between the three groups of children.

(b) Yes. The F statistic is 1.13 and the P-value is 0.329.

(c) Yes. The difference in sample means is small compared to within-sample variation.



(d) No. Pairwise comparisons of means are probably not necessary since the F test of the ANOVA hypothesis has already indicated no statistically-significant differences between means.

(e) Let μ = mean pretest score for all children. A 95% confidence interval for μ is

(9.05 to 10.53) points.

Combining the three samples makes sense because the F test shows no real differences between the three means. In this case μ is the common value:

$$\mu = \mu_1 = \mu_2 = \mu_3$$

2. (Test scores)

(a)

1. μ_1 = mean test score of all children taught with Basal method
   $\mu_2 = \ldots\ldots$ DRTA
   $\mu_3 = \ldots\ldots$ Strat

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3$$
$$H_A\colon\ \text{At least one } \mu_i \text{ differs from the others}$$

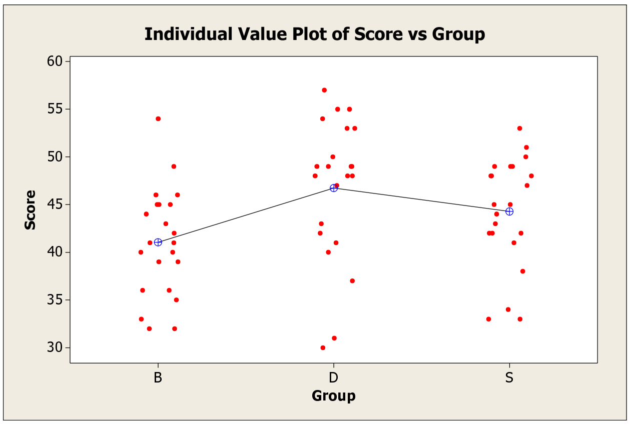
2. From MINITAB: $F = \boxed{4.48}$, P-value $= \boxed{0.015}$

3. Reject H_0 since P-value $= 0.015 < .05 = \alpha$

4. There is sufficient evidence to show that mean test scores differ between teaching methods.

(b) Yes. The F statistic is 4.48 and the P-value is 0.015.

(c) It's difficult to judge from the graph whether there are any real differences in the three population means. For that reason, we should rely on the P-value.



(d) One of the Tukey confidence intervals shows a significant difference between the Basal and DRTA methods. A 95% confidence interval for $(\mu_2 - \mu_1)$ is

(1.118 to 10.245) points

so we are 95% confident that using DRTA rather than Basal will result in improved test scores of between 1.12 and 10.25 points, on average.

Neither of the other two Tukey confidence intervals shows a statistically significant difference:

- No clear difference between the Basal and Strat methods
- No clear difference between the DRTA and Strat methods

(e) The company can claim that its DRTA method is superior to Basal but cannot make a similar claim for Strat.

3. (Bookstores)

(a)

1. μ_1 = mean age of all Bookstore 1 customers (in years)
   μ_2 = ... Bookstore 2
   μ_3 = ... Bookstore 3
   μ_4 = ... Bookstore 4
   μ_5 = ... Bookstore 5

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_A\colon \text{ At least one } \mu_i \text{ differs from the others}$$

2. $F = \boxed{472.46}$, P-value $= \boxed{0.000}$

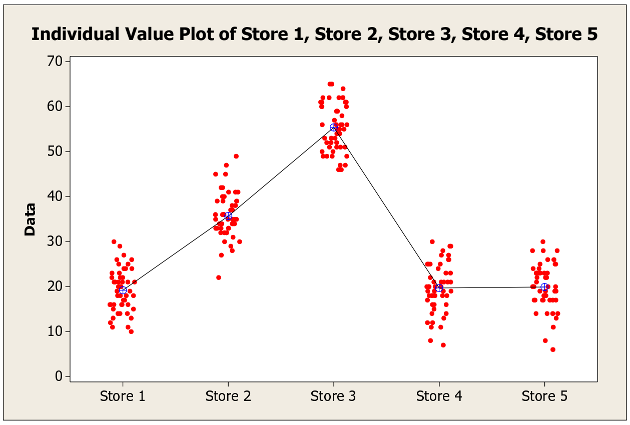
3. <u>Reject H_0</u> since P-value $= 0 < .05 = \alpha$.

4. There is sufficient evidence to show that mean customer age is not the same for all bookstores.

(b)

- Larger than Bookstore 1: Bookstores 2 and 3
- Larger than Bookstore 2: Bookstore 3
- Larger than Bookstore 3: None
- Larger than Bookstore 4: Bookstores 2 and 3
- Larger than Bookstore 5: Bookstores 2 and 3

(c) The Tukey confidence intervals in (b) show that there are <u>three</u> distinct clusters. The ANOVA graph illustrates this conclusion:



(d) Cluster 1 = Bookstores 1, 4, 5
Cluster 2 = Bookstore 2
Cluster 3 = Bookstore 3

(e) If the clusters are defined as in (d),

Cluster 1 has youngest customers.
Cluster 2 has intermediate customers.
Cluster 3 has oldest customers.

(f) Bookstore customers at the mall tend to be younger than bookstore customers downtown.

(g) Cluster 3 = Bookstore 3 has the oldest customers. A 95% confidence interval for μ_3 is

$$(53.9,\ 56.9) \text{ years.}$$

(h) Cluster 1 = Bookstores 1, 4, 5 has the youngest customers. Let μ represent the (common) mean age of customers at these three stores, all located at the mall. A 95% confidence interval for μ is

$$(18.782,\ 20.418) \text{ years.}$$

4. (More about clusters)

(a)
- Cluster 1 = Basil and Strat methods
- Cluster 2 = DRTA and Strat methods

(b) The clusters are not distinct since they contain a teaching method (Strat) in common. Therefore neither cluster is shown to be clearly superior to the other one.

5. Exercise 14.54

(a)

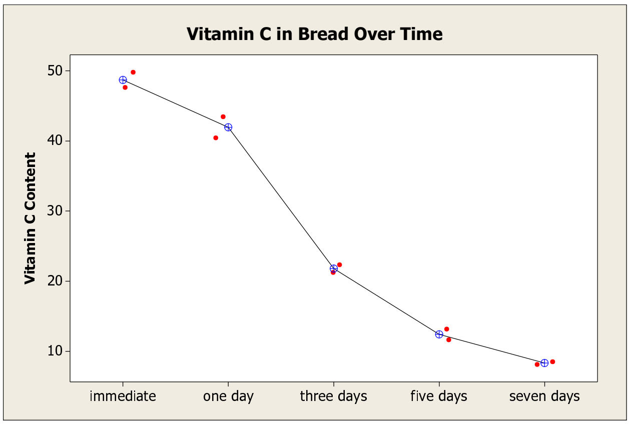
1.
μ_1 = mean Vitamin C content in bread immediately after baking
μ_2 = ... one day after baking
μ_3 = ... three days after baking
μ_4 = ... five days after baking
μ_5 = ... seven days after baking

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_A\colon \text{ At least one } \mu_i \text{ differs from the others.}$$

2. $F = \boxed{367.74}$ P-value = $\boxed{0.000}$

3. Reject H_0 since P-value $= 0 < .05 = \alpha$

4. There is sufficient evidence to show that mean Vitamin C content in bread changes over time.



(b) Tukey 95% confidence intervals for adjacent time points:

- For $(\mu_2 - \mu_1)$: $(-12.045\ ,\ -1.455)$
- For $(\mu_3 - \mu_2)$: $(-25.455\ ,\ -14.865)$
- For $(\mu_4 - \mu_3)$: $(-14.675\ ,\ -4.085)$
- For $(\mu_5 - \mu_4)$: $(-9.390\ ,\ 1.200)$

There are significant differences among the first four adjacent time points, but not between the fourth and fifth time points.

Interpret: We are 95% confident that the average Vitamin C content in bread decreases by between

- 1.455 and 12.045 mg per 100g of flour from immediately after baking to one day after baking.
- 14.865 and 25.455 mg per 100g of flour from one day after baking to three days after baking.
- 4.085 and 14.675 mg per 100g of flour from three days after baking to five days after baking.

(c) Loss in Vitamin C (on average) begins immediately after baking and continues until five days after baking, at which time it appears to taper off between five and seven days after baking.

(d) There are [4] distinct clusters of time points:

Cluster 1 = Immediately after baking
Cluster 2 = One day after baking
Cluster 3 = Three days after baking
Cluster 4 = Five days and seven days after baking

6. Exercise 14.56

Vitamin A

(a)

1.

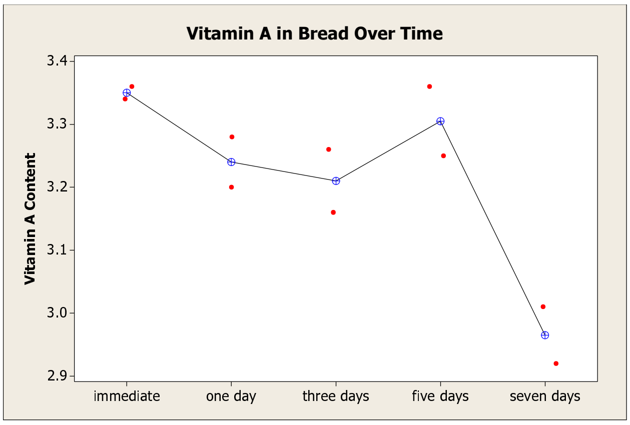
μ_1 = mean Vitamin A content in bread immediately after baking
μ_2 = ... one day after baking
μ_3 = ... three days after baking
μ_4 = ... five days after baking
μ_5 = ... seven days after baking

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_A\colon \text{ At least one } \mu_i \text{ differs from the others.}$$

2. $F = \boxed{12.09}$ P-value $= \boxed{0.009}$

3. Reject H_0 since P-value $= 0.009 < .05 = \alpha$

4. There is sufficient evidence to show that mean Vitamin A content in bread changes over time.



(b) Tukey 95% confidence intervals for adjacent time points:

- For $(\mu_2 - \mu_1)$: $(-0.354\ ,\ 0.134)$
- For $(\mu_3 - \mu_2)$: $(-0.274\ ,\ 0.214)$
- For $(\mu_4 - \mu_3)$: $(-0.149\ ,\ 0.339)$
- For $(\mu_5 - \mu_4)$: $(-0.584\ ,\ -0.096)$

Therefore there are no significant differences among the first four adjacent time points, but there is a significant difference between the fourth and fifth time points.

Interpret:

We are 95% confident that the average Vitamin A content in bread decreases by between 0.096 and 0.584 mg per 100g of flour from five days after baking to seven days after baking.

(c) There is no detectable loss in average Vitamin A content in bread until five days after baking, but there is a loss between five and seven days after baking.

(d) There are $\boxed{2}$ distinct clusters of time points:

Cluster 1 = Immediately, one day, three days, five days after baking
Cluster 2 = Seven days after baking

Vitamin E

(a)

1.

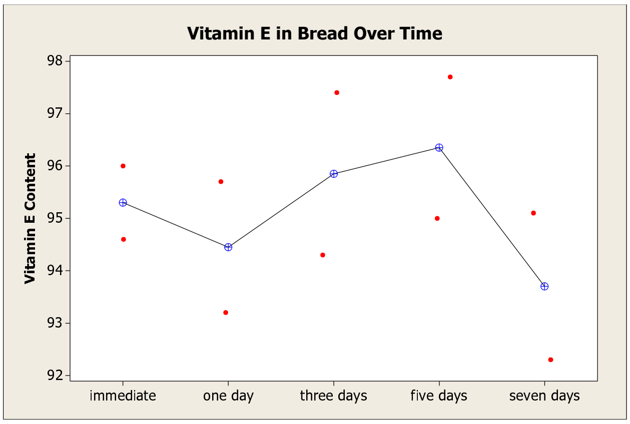
μ_1 = mean Vitamin E content in bread immediately after baking
μ_2 = ... one day after baking
μ_3 = ... three days after baking
μ_4 = ... five days after baking
μ_5 = ... seven days after baking

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$
$$H_A\text{: At least one } \mu_i \text{ differs from the others.}$$

2. $F = \boxed{0.69}$ P-value $= \boxed{0.630}$

3. Fail to Reject H_0 since P-value $= 0.630 > .05 = \alpha$.

4. There is insufficient evidence to show that mean Vitamin E content in bread changes over time.



(b) Here are the Tukey 95% confidence intervals for adjacent points in time:

- For $(\mu_2 - \mu_1)$: $(-8.128\ ,\ 6.428)$
- For $(\mu_3 - \mu_2)$: $(-5.878\ ,\ 8.678)$
- For $(\mu_4 - \mu_3)$: $(-6.778\ ,\ 7.778)$
- For $(\mu_5 - \mu_4)$: $(-9.928\ ,\ 4.628)$

Therefore there are no significant differences between any of the five time points, which is consistent with the fact that we did not reject the ANOVA null hypothesis!

(c) There is no apparent loss in Vitamin E in bread (on average) over the first seven days after baking.

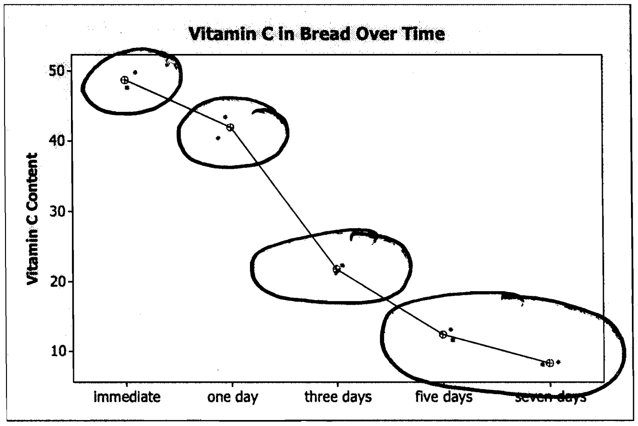
(d) There is a single cluster of time points:

Cluster 1 = Immediately, one day, three days, five days, seven days after baking.

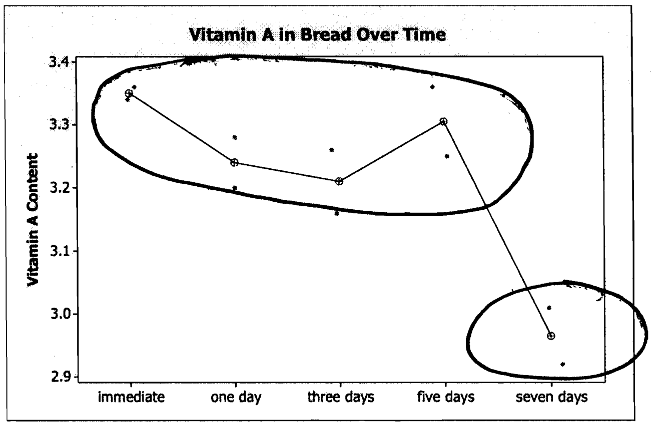
(solution for Exercise 7 on next page)

7. (a) and (b)

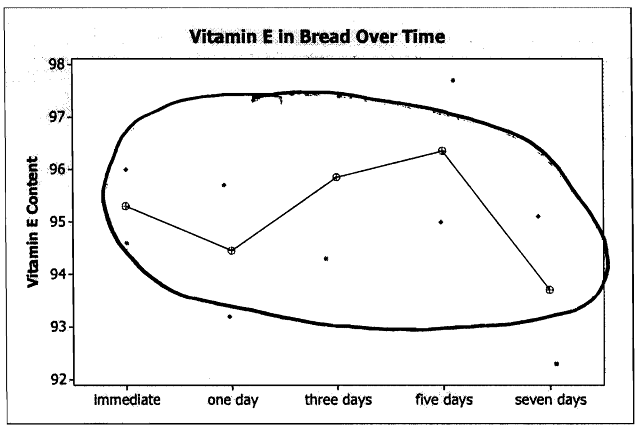
Vitamin C, A, E Loss in Bread Over Time



Vitamin C: $F =$ 367.74 P-value = 0.000 (Also circle clusters for Vitamin C on graph.)



Vitamin A: $F =$ 12.09 P-value = 0.009 (Circle clusters for Vitamin A.)



Vitamin E: $F =$ 0.69 P-value = 0.630 (Circle clusters for Vitamin E.)

(c) As the F statistics decrease, the P-values <u>increase</u> while the numbers of clusters <u>decrease</u>.