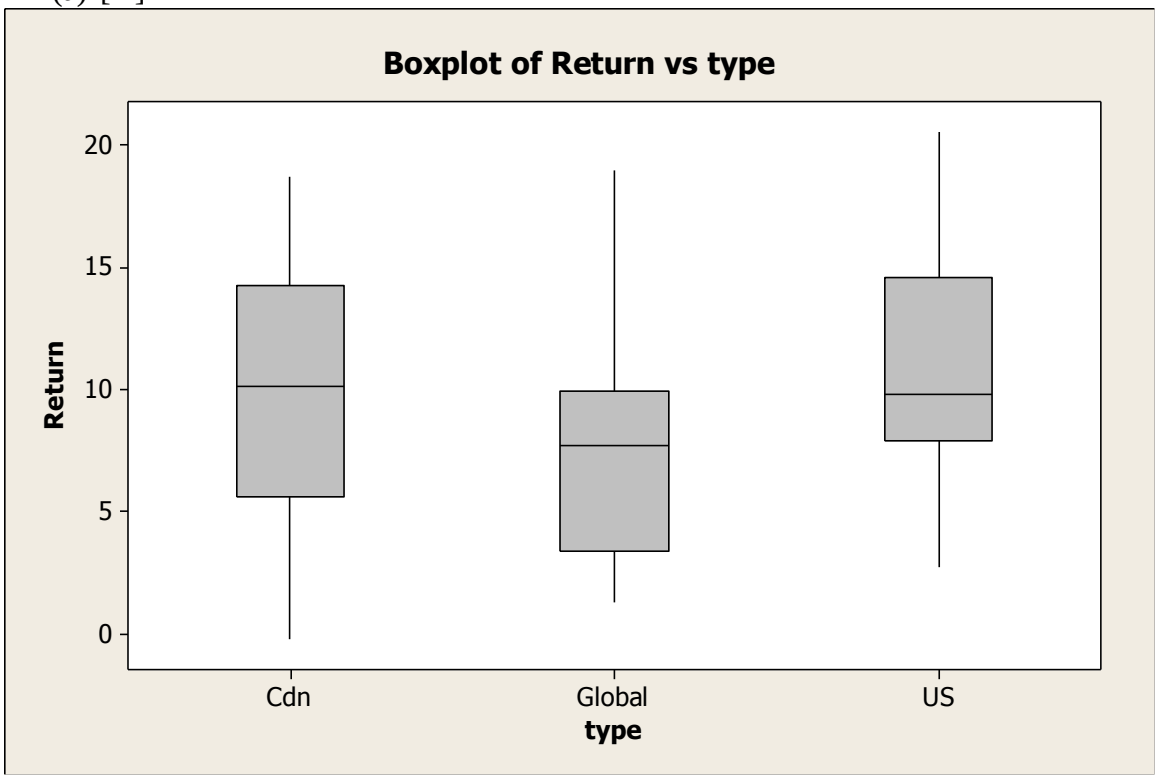
ADM 2304
Assignment 3. ***[total: 35 marks]***

Question 1. [15 marks]

(a) [2]



There is more skewness with the global sample, suggesting possibly a problem with the assumption that the data come from a normal distribution; however, given the small sample sizes, any comment that it is reasonable to assume normality is acceptable. However, the constant variance assumption looks reasonable (similar range, similar IQR for small sample sizes).
-1 mark for each comment above.

(b) [2]

Descriptive Statistics: Return

Variable	type	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Return	Cdn	16	0	9.78	1.44	5.78	-0.220	5.63	10.11	14.30
	Global	16	0	7.51	1.16	4.65	1.28	3.40	7.74	9.97
	US	13	0	11.03	1.36	4.91	2.70	7.91	9.82	14.57

Variable	type	Maximum
Return	Cdn	18.73
	Global	19.01
	US	20.56

The three sample stdevs are 5.78, 4.65 and 4.91.
The pooled variance is (15*5.78^2 + 15*4.65^2 + 12*4.91^2) /(15+15+12) = 1114.76 / 42 = 26.54
The value of 26.54 agrees with the MSE in the table below.

-1 mark for the three sample stdevs, 1 for the calculation of 26.5 (accuracy of +/- 0.5)

(c) [3]

One-way ANOVA: Return versus type

```
Source  DF      SS    MS     F      P
type     2    93.7  46.8  1.76  0.184
Error   42  1114.4  26.5
Total   44  1208.1

S = 5.151   R-Sq = 7.75%   R-Sq(adj) = 3.36%
```

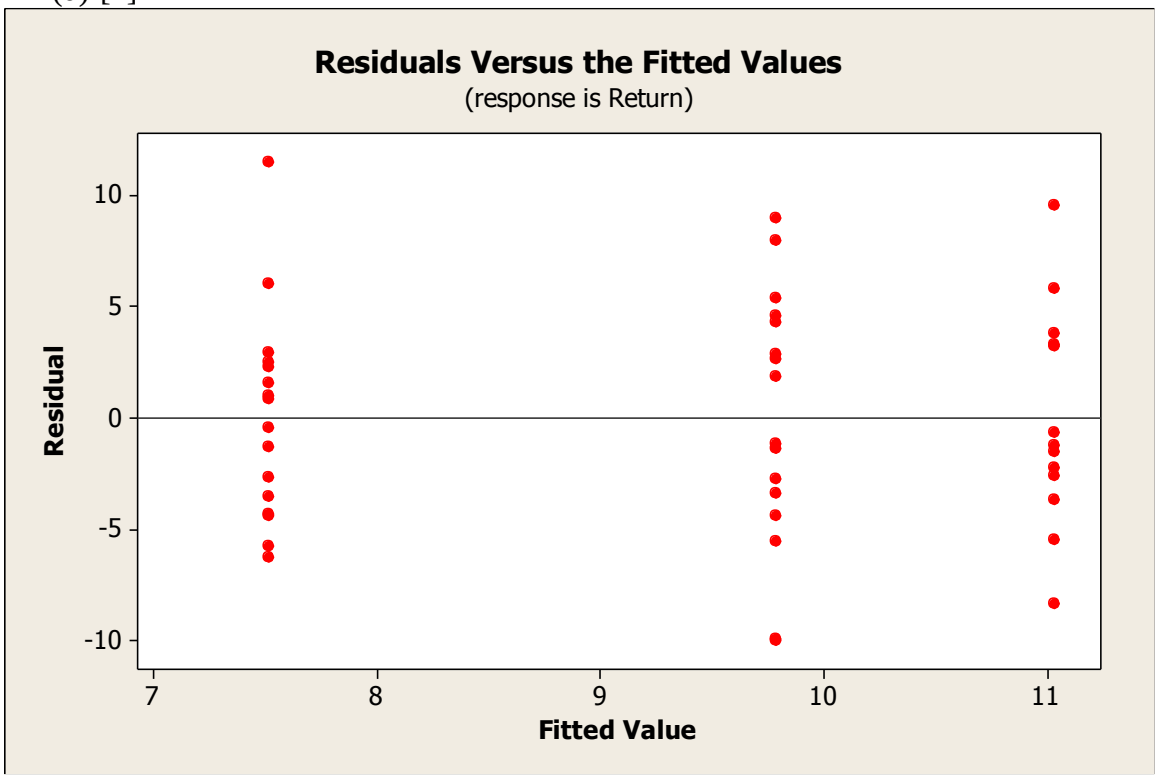
Ho: average returns are the same, Ha: average returns are different

F=1.76 with p-value of 0.184, critical value is 3.22 to 3.23

We reject the null H at the 0.05 level, and conclude the average returns differ by type.

-1 for hypotheses, 1 for F-statistic and p-value or critical value, 1 for decision/conclusion

(d) [2]



-Given the standard error of 5.15, all the residuals are within two standard errors of zero. This agrees with the assumption of normally distributed errors.
-Each group of residuals has approximately the same vertical spread. It is reasonable to assume the errors have the same variance.
-1 mark for each comment above.

(e) [3]
K =3 and J = 3.
The critical value of t (with 42 df) is +/-2.49 based on a tail probability of 0.05/(2*3).
The sample means are Cdn 9.78, Global 7.51, and U.S. 11.03.

Inverse Cumulative Distribution Function

```
Student's t distribution with 42 DF

P( X <= x )          x
   0.008333  -2.49368
```

To compare Canadian with Global, the ME is 2.49*5.15*sqrt(2/16) = 4.53.
The CI is (9.78-7.51) ± 4.53 = 2.26 ± 4.53, we conclude the two are not statistically significantly different from zero.

To compare Global with U.S. or with U.S. with Canadian, the ME is 2.49*5.15*sqrt(1/16 + 1/13) = 4.79. The CI is either (11.03 – 7.51) ± 4.79 = 3.52 ± 4.79 or (11.03 – 9.78) = 1.25 ± 4.79, respectively. Since neither CI covers zero, we do not reject the null H of no difference.

These conclusions agree with the F-test.

-1 mark for margin of error calculations, 1 mark for the three Cis, 1 mark for the conclusion of no difference.

(f) [3]

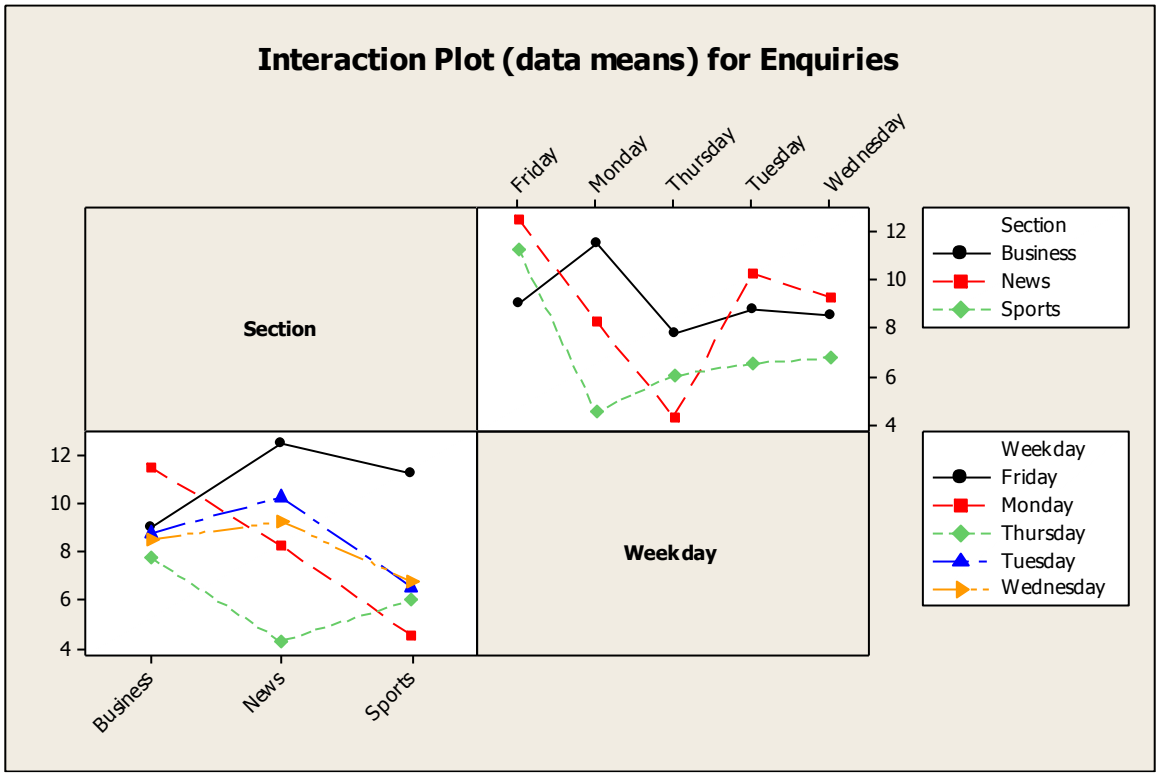
```
Kruskal-Wallis Test on Return

type      N  Median  Ave Rank      Z
Cdn      16  10.105      24.2   0.45
Global   16   7.740      18.3  -1.78
US       13   9.820      27.3   1.40
Overall  45              23.0

H = 3.57  DF = 2  P = 0.168
```

-Ho: median returns are the same, Ha: medians are different.
-p-value = 0.168 < 0.05
-Do not reject null H, conclude no difference in median returns.

-1 mark for each of the above.

Question 2. ***[20 marks]***

(a) [3 marks]



There is interaction between the section and the weekday factors *[1 mark]*
There appears to be main effects due to the section and due to the weekday *[1 mark for mentioning either main effect]*
There does not appear to be a section that is best for all weekdays nor does there appear to be a weekday that is best for all sections *[1 mark for either comment]*

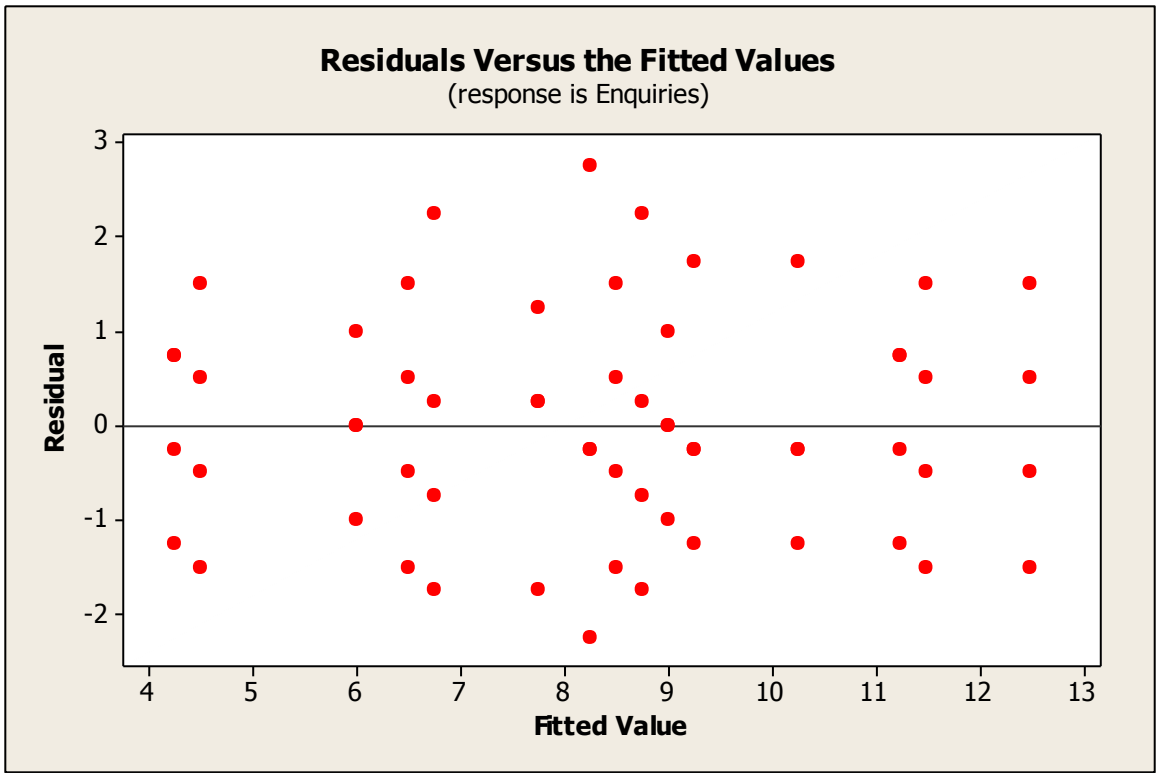
(b) [2 marks]

	Friday	Monday	Thursday	Tuesday	Wednesday	All
Business	0.816	1.291	1.258	1.708	1.291	1.744
	4	4	4	4	4	20
News	1.291	2.062	0.957	1.258	1.258	3.059
	4	4	4	4	4	20
Sports	0.957	1.291	0.816	1.291	1.708	2.575
	4	4	4	4	4	20
All	1.782	3.315	1.758	2.067	1.697	2.653
	12	12	12	12	12	60

[3 * 0.816^2 + 3 * 2.062^2 + … + 3 * 1.697^2] / 45 = 1.7556
[1 mark for showing this set of calculations]

There are 45 d.f. since each sample variance contributes 3 d.f. and there are 15 sample variances *[1 mark for this comment]*

(c) [2 marks]



With a standard error of 1.325, there are relatively few (4 out of 60 or about 6%) residuals beyond 2 standard errors of zero and the spread of the residuals is fairly constant from group to group. Therefore it is reasonable to think that the errors are normally distributed and have constant variance.

[0.5 mark for each of the two assumptions and 0.5 mark for each of the two comments]

Source	DF	SS	MS	F	P
Section	2	53.733	26.8667	15.30	0.000
Weekday	4	146.833	36.7083	20.91	0.000
Interaction	8	135.767	16.9708	9.67	0.000
Error	45	79.000	1.7556		
Total	59	415.333			

S = 1.325 R-Sq = 80.98% R-Sq(adj) = 75.06%

For d, e, f, [1 for hypotheses, 1 for F-stat and p-value or critical value, 1 for decision/ conclusion]

(d) [3 marks]

Ho: no interaction; Ha: some interaction

F = 9.67 with p-value of 0.000, critical value between 2.13 and 2.18.

Since p-value < 0.05, we reject the null hypothesis and conclude there is interaction.

(e) [3 marks]

Ho: no main effect due to day of week; Ha: some day of week effect

F = 20.91 with p-value of 0.000, critical value between 2.56 and 2.61.

Since p-value < 0.05, we reject the null hypothesis and conclude there is a main effect due to day of week.

(f) [3 marks]

Ho: no main effect due to section; Ha: some section effect

F = 15.30 with p-value of 0.000, critical value between 3.18 and 3.23.

Since p-value < 0.05, we reject the null hypothesis and conclude there is a main effect due to section.

It could be argued that with interaction, the main effects may not be meaningful. However, it can also be argued that if one wanted to advertise only one day a week or only in one section, then the main effects are meaningful.

(g) [4 marks]

There are 15 treatment means and therefore J = 15*14/2 = 105 pairwise comparisons.

The margin of error for each comparison is

± t(0.05/210) * 1.325 * sqrt(2/4) = ± t(0.000238) * 0.937 = ± 3.77 * .937 = ± 3.53

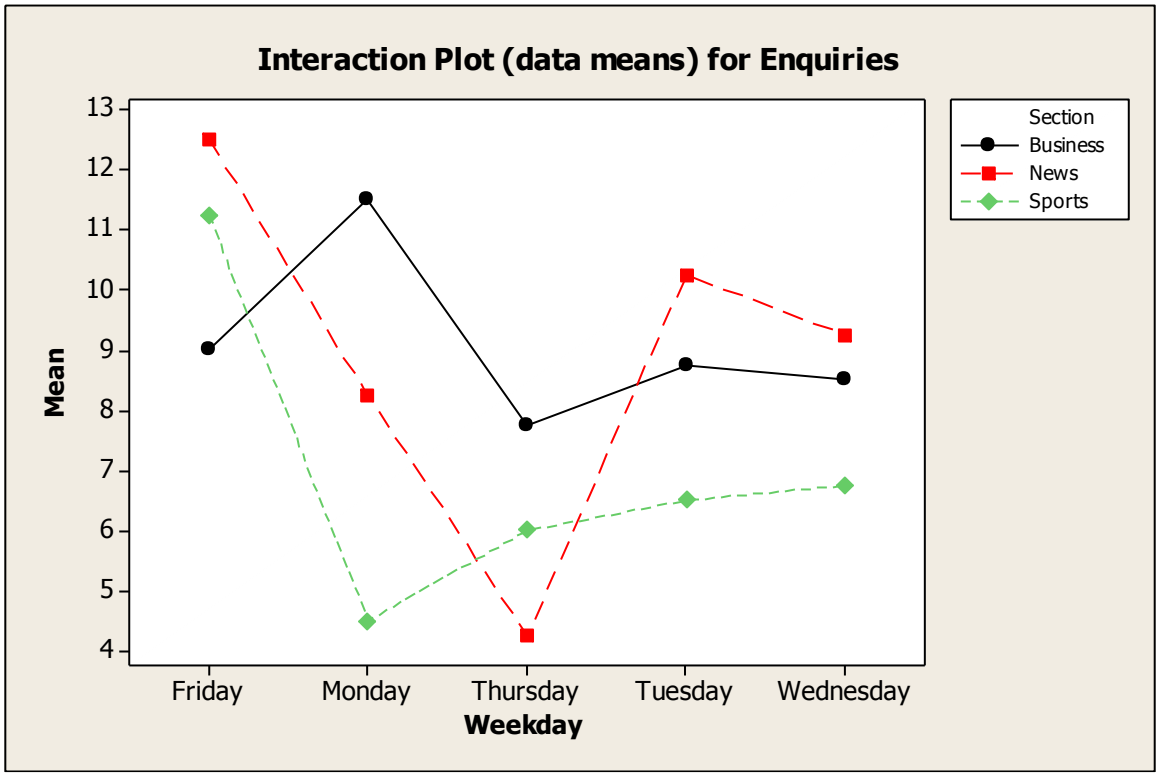
```
Student's t distribution with 45 DF

P( X <= x )          x
   0.000238  -3.76811
```

[1 mark for the t critical value, 1 for the standard error sqrt(2/4), 1 for ME 3.53]

The fifteen treatment means are in bold font below:

	Friday	Monday	Thursday	Tuesday	Wednesday	All
Business	**9.000**	**11.500**	**7.750**	**8.750**	**8.500**	9.100
	4	4	4	4	4	20
News	**12.500**	**8.250**	**4.250**	**10.250**	**9.250**	8.900
	4	4	4	4	4	20
Sports	**11.250**	**4.500**	**6.000**	**6.500**	**6.750**	7.000
	4	4	4	4	4	20
All	10.917	8.083	6.000	8.500	8.167	8.333
	12	12	12	12	12	60

Cell Contents: Enquiries : Mean
Count



There are too many pairwise comparisons that are statistically significantly different to mention all of them.
If we look at the different days of the week, we can see:
-for Tuesday, news is better than sports;
-for Wednesday, there are no differences;
-for Thursday, there are no differences;
-for Friday, there are no differences.

For Mondays, the means are Sports 4.5, news 8.25, business 11.5

The two CIs are:

News – sports is 3.75 ± 3.53,
Business – sports is 7 ± 3.53.

Since both CIs do not cover zero, we can conclude that for Monday, the sports section is worse than the other two sections.

-1 mark for this conclusion, provided the two CIs are shown.