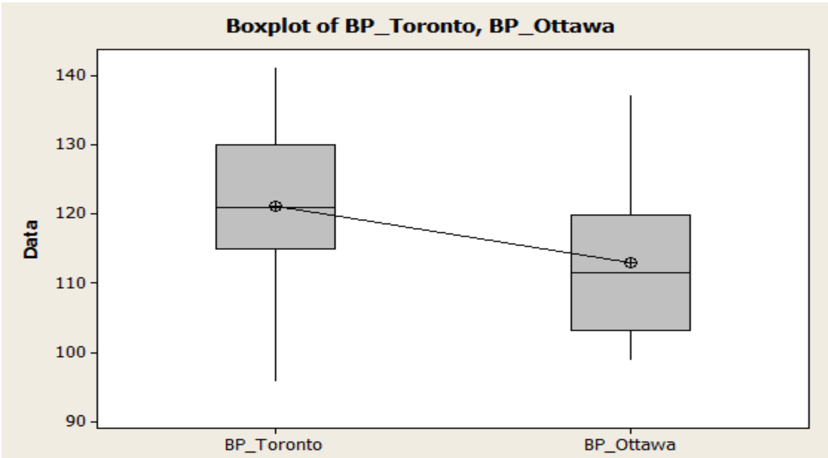
Assignment 2

Robyn Ackerman
ADM 2304 N
7710097

Question 1

PART A



It is best to use a parametric test because the data, although slightly skewed, has no extreme outliers. Therefore, it is safe to assume, for testing purposes, that the data follows a normal distribution.

PART B

2-Sample T-Test

Two-sample T for BP_Toronto vs BP_Ottawa

	N	Mean	StDev	SE Mean
BP_Toronto	35	121.1	11.4	1.9
BP_Ottawa	20	112.9	10.6	2.4

H_o: $\mu_1 - \mu_2 = 2.5$
H_A: $\mu_1 - \mu_2 > 2.5$

Level of Significance = 0.05

$S_p = \sqrt{(34)(11.6)^2 + (19)(10.6)^2 / 53}$
$S_p = 11.1198$

$T_{STAT} = \sqrt{(121.1 - 112.9) - 2.5 \div 11.1198(\sqrt{\frac{1}{35} + \frac{1}{20}})}$

$T_{STAT} = 1.828$

Df = 35 + 20 – 2 = 53

Since we have a large sample size (>30), we can use the z-table to find our p-value.
Therefore, p = 1 – 0.9664

P-value = 0.336

Since $p < 0.05$ (level of significance) we reject the null hypothesis. There is sufficient evidence to believe that the BP of Toronto males is more than 2.5 higher than the BP of Ottawa males.

PART C

$T_{CRIT} = 1.6741$

UB = infinity

$$LB = (121.1 - 112.9) - 1.6741 * (11.1198)\left(\sqrt{\frac{1}{35} + \frac{1}{20}}\right)$$

LB = 2.578

CI = (2.578, infinity)

Since the value 2.5 is not in the confidence interval, we reject the null hypothesis. This is consistent with my conclusion in part b.

PART D
The mean BP for Toronto = 126 and the mean BP for Ottawa = 124. This creates a difference of 2, which is the smallest possible difference in the entire sample.

Question 2
PART A

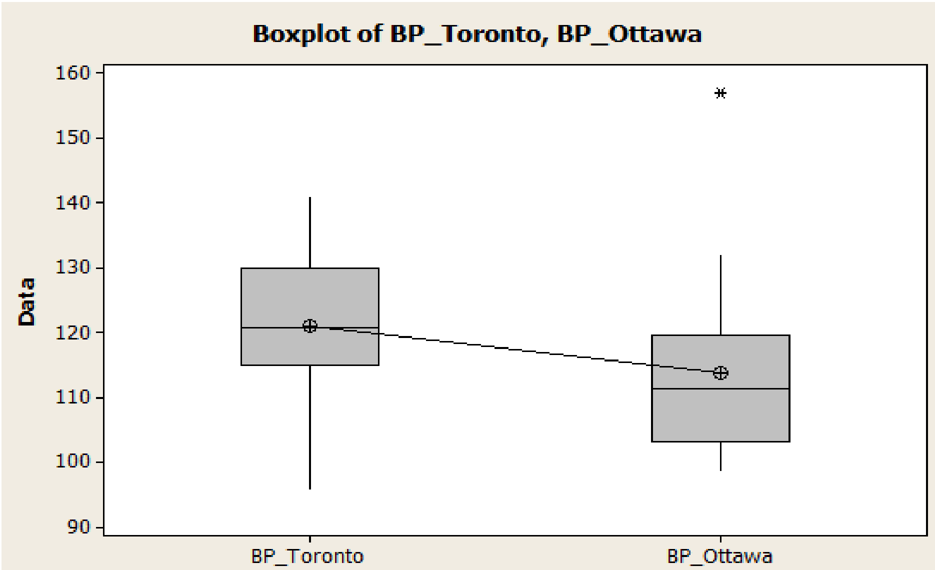
H_o: $M_d = 2.5$
H_A: $M_d > 2.5$
Level of Significance = 0.05

Mann-Whitney Test and CI: BP_Toronto, BP_Ottawa

```
             N  Median
BP_Toronto  35  121.00
BP_Ottawa   20  111.50


Point estimate for ETA1-ETA2 is 9.00
95.1 Percent CI for ETA1-ETA2 is (2.00,16.00)
W = 1125.5
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0056
The test is significant at 0.0056 (adjusted for ties)
```

PART B



Since there is an extreme outlier (in Ottawa) and the data is skewed, the data are not normal. Therefore, it would be best to use a non-parametric test in this case.

Question 3

PART A

Correlations: Wt_Week0, Wt_Week4

```
Pearson correlation of Wt_Week0 and Wt_Week4 = 0.995
P-Value = 0.000
```

I would use a paired approach because the data of week 1 is related to the data of week 4. They are dependent on each other. The correlation coefficient shows that there is a 0.995 correlation between the two sets of data. This is a very positive and strong correlation, which means that they are in fact dependent and paired. This supports my answer.

PART B

Paired T-Test

Paired T for Wt_Week0 - Wt_Week4

	N	Mean	StDev	SE Mean
Wt_Week0	15	143.667	12.971	3.349
Wt_Week4	15	137.400	10.907	2.816
Difference	15	6.26667	2.40436	0.62080

H_o: $\mu_d = 5$
H_A: $\mu_d > 5$
Level of Significance = 0.05
Df = n – 1 = 14

$T_{STAT} = \frac{6.26667 - 5}{2.40436/\sqrt{15}}$

$T_{STAT} = 2.04$

$T_{CRIT} = 1.761$
P-Value = 0.030

Since p < level of significance we reject the null hypothesis. We have sufficient evidence to believe that the claim that a person will lose more than 5lbs by the end of the programme is true.

PACT C

UB = Infinity

LB = 6.26667 – 1.761 * 2.40436/ $\sqrt{15}$
LB = 5.173436

CI = (5.173436, infinity)

Therefore, the confidence interval does not contain the μ_d value of 5, so we reject the null hypothesis.

PART D
H_o: $M_d = 5$
H_A: $M_d > 5$
level of significance = 0.05

Wilcoxon Signed Rank Test: Wt_Week0, Wt_Week4

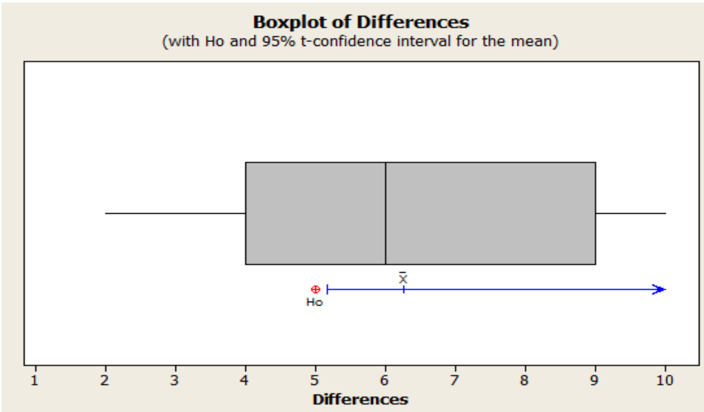
Test of median = 5.000 versus median > 5.000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Wt_Week0	15	15	120.0	0.000	142.0
Wt_Week4	15	15	120.0	0.000	135.8

$T_{CRIT} = 30$
P-value = 0

Since the p value is less than the level of significance, we reject the null hypothesis

PART E



The boxplot of differences is not normally distributed, which means it would be more appropriate to use the non-parametric approach to solving this problem. We would also use this method when calculating the median, and a parametric would be more appropriate for testing the mean, which is seen in part b of this question.

Question 4

PART A

Since we are comparing two categorical variables to determine their relationship, it is best to use the chi-square test for independence.

H_o: age and social media preference are independent
H_A: age and social media preference are not independent

PART B

Chi-Square Test: M1, M2, M3, M4, M5

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

          M1     M2     M3     M4     M5  Total
    1     50     45     50     25     27    197
       57.69  47.89  40.27  23.94  27.21
       1.024  0.174  2.351  0.047  0.002

    2     60     55     50     30     30    225
       65.88  54.70  45.99  27.35  31.08
       0.525  0.002  0.349  0.257  0.037

    3     65     45     30     25     25    190
       55.64  46.19  38.84  23.09  26.24
       1.576  0.031  2.012  0.157  0.059

    4     55     30     20     15     26    146
       42.75  35.49  29.85  17.75  20.17
       3.509  0.850  3.248  0.425  1.688

    5     35     45     35     15     17    147
       43.04  35.73  30.05  17.87  20.30
       1.503  2.402  0.815  0.460  0.538

Total    265    220    185    110    125    905

Chi-Sq = 24.041, DF = 16, P-Value = 0.089
```

Df = (5 – 1)(5 – 1)
Df = 16

E_{11} = (197 x 265)/905 = 57.6851

$X^2_{11} = (50 - 57.6851)^2/57.6851 = 1.0238$

E_{35} = (190 x 125)/905 = 26.2431

$X^2_{35} = (25 - 26.2431)^2/26.2431 = 0.0589$

Test statistic = 24.041
Given that the degrees of freedom is 16, we find a critical value of 26.296

In this case, the test statistic is not greater than the critical value, so we fail to reject the null hypothesis.

PART C

P-value 0.089 from the table above. Since the level of significance is 0.05 (95% confidence interval), P > 0.05, so we fail to reject the null hypothesis which confirms our conclusion in part b.

PART D

H_o: $P_1 = 0.25$ $P_2 = 0.25$ $P_3 = 0.1$ $P_4 = 0.2$ $P_5 = 0.3$
H_A: at least one proportion is different from the specified probability

	M1	M2	M3	M4	M5	Total
1	50	45	50	25	27	197

Observed	Percentage	Expected	Expected (counts)	Difference	$(O - E)^2 / E$
50	25.38%	25%	49.25	0.75	0.01142
45	22.84%	25%	49.25	-4.25	0.36675
50	25.38%	30%	59.10	-9.10	1.40118
25	12.69%	10%	19.70	5.30	1.42580
27	13.71%	10%	19.70	7.30	2.70500

add up the x^2 values to obtain a value of 5.91
Critical value = 9.4877
Since the test statistic (5.91) is less than the critical value (9.4877) we fail to reject the null hypothesis.