ADM2304 ***Assignment#2*** ***Winter 2014***

Solution: 100 marks

Question#1 (30 marks)

a. (9 marks)

S1: H0: $mu_1 - mu_2 = 0$ Ha: $mu_1 - mu_2 \neq 0$

S2: $t_{Calc} = \{(X_{Bar1} - X_{Bar2}) - 0\}/SE(X_{Bar1} - X_{Bar2}) = \{13.67 - 0\}/4.7424$

$t_{Calc} = 2.8825$

{It is assumed that the two population variances are unequal and the df = 105 is obtained from the MiniTab output. If equal population variances are assumed (it is inappropriate but..), then $SE(X_{Bar1} - X_{Bar2}) = 5.5024$ and $t_{Calc} = 2.4844$ with df = 127.}

S3: With LS = 5%, $t_{Crit} = t_{\alpha/2}(df = 105) \approx 1.984$ (more precise value 1.9828) (or 1.96)

{With df = 127, $t_{Crit} = 1.9788$}

S4: Since $\{|t_{Calc}| = 2.8825\} > \{t_{Crit} = 1.984\}$ ==> Reject H0.

With equal variance assumption, the conclusion would be the same.

{2 marks each for the 4 steps or their equivalent }

Based on the available statistical evidence, one can say that there is a difference in the population means. {1 mark}

b. (5 marks)

CI: $(X_{Bar1} - X_{Bar2}) \pm t_{\alpha/2}(df = 105) \times SE(X_{Bar1} - X_{Bar2}) = 13.67 \pm 1.9828 \times 4.7424$

CI: 13.67 ± 9.4013 which is (4.2687, 23.0713) {3 marks}

{ With df = 127, CI would be (2.78, 24.56)}

Since the CI does not contain the value '0', the hypothesized difference in H0, it is consistent with the conclusion reached in part 'a'. {2 marks}

c. (2 marks)

Here, $4.2687 < mu_1 - mu_2 < 23.0713$, or

$mu_2 + 4.2687 < mu_1 < mu_2 + 23.0713$, or {1 mark}

mu_1 is more than mu_2 by a value of anywhere from 4.2687 to 23.0713. {1 mark}

d. (9 marks)

S1: H0: $mu_1 - mu_2 = 5$ Ha: $mu_1 - mu_2 > 5$

S2: $t_{Calc} = \{(X_{Bar1} - X_{Bar2}) - 5\}/SE(X_{Bar1} - X_{Bar2}) = \{13.67 - 5\}/4.7424$

$t_{Calc} = 1.8282$

{It is assumed that the two population variances are unequal and the df = 105 is obtained from the MiniTab output. If equal population variances are assumed (it is inappropriate but..), then $SE(X_{Bar1} - X_{Bar2}) = 5.5024$ and $t_{Calc} = 1.5757$ with df = 127.}

S3: With LS = 5%, $t_{Crit} = t_{\alpha}(df = 105) \approx 1.660$ (more precise value 1.6595) (or 1.645)

{With df = 127, $t_{Crit} = 1.6569$}

S4: Since $\{|t_{Calc}| = 1.8282\} > \{t_{Crit} = 1.660\}$ ==> Reject H0.

{With equal variance assumption, the conclusion would not be the same, since $\{t|Calc| = 1.5757\} < \{tCrit = 1.6569\}$==> Do not Reject H0}.

{2 marks each for the 4 steps or their equivalent }

Based on the available statistical evidence, one can say that there is a difference of more than 5 Calories in the two population means. {1 mark}

e. (5 marks)

The appropriate asymmetrical CI for the 'Right Tail' test is the Lower Bound, LB. Here,

LB = (X_{Bar1} - X_{Bar2}) - t_{α}(df = 105) x SE(X_{Bar1} - X_{Bar2}) = 13.67 - 1.6595 x 4.7424

LB = 5.80, or CI = (5.80, ∞ (accept 13.67 -1.645*4.7424 = 5.87) {3 marks}

{ With df = 127, LB = 13.67 - 1.6569 x 5.5024 = 4.5531} accept 13.67 – 1.645 * 5.5024 = 4.62

Since the Difference of '5' is outside the LB, it would have to be rejected and it would be consistent with test Right Tail test in part 'd' where H0 is rejected. {2 marks}

{With equal variance and 'df' = 127, Difference '5' would be within the LB would be consistent with Right Tail test in part 'd' where H0 cannot be rejected.}

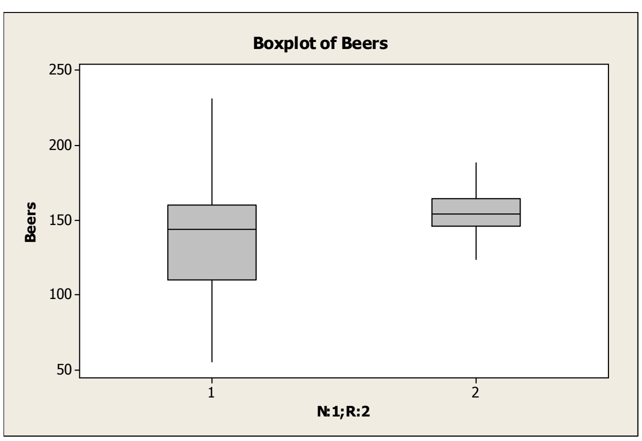
N.B.: If 'Un-Equal Variance' or 'Equal Variance' assumption is made, then all the sub-questions must have the same consistent assumption.

Question#2 (15 marks)

Mann-Whitney Test and CI: Regional Beer:Cals, National Beer:Cals

```
                          N   Median
Regional Beer:Cals   52   154.50
National Beer:Cals   77   144.00

Point estimate for ETA1-ETA2 is 15.00
95.0 Percent CI for ETA1-ETA2 is (4.99,26.00)
W = 3999.0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.0030
The test is significant at 0.0030 (adjusted for ties)
```



a. (6 marks)

S1: H0: Md_{Pop1} - Md_{Pop2} = 0 Ha: Md_{Pop1} - Md_{Pop2} ≠ 0 {3 marks}

S2, S3, S4 not required

From the MiniTab output,

S5: Since {p-Val = 0.003} < {α = 0.05} ==> Reject H0 {2 marks}

Based on the available statistical evidence, one cannot say that the two population medians are the same or conclude they are different {1 mark}

b. (3 marks)

The 95% CI for the difference of two population medians is (from the output above is: (4.99, 26.00).

Obviously the value '0' is not in it and as such the H0 has been rejected.

c. (6 marks)

The two samples are statistically independent and as such the boxplots for the two data from the two samples must be observed {1 mark}. The two boxplots do not indicate the presence of real or suspected outliers and both boxplots are reasonably symmetrical {2 marks}. This suggests that two samples and by inference, the two populations from which they are drawn are reasonably normal{2 marks}. Thus parametric test is more appropriate than the non-parametric test. Although test on the difference of the two medians is always right, in this situation, the test on the difference of the two means is more appropriate. Thus the test done in Question 1 'a' is the more appropriate one {1 mark}.

Question#3 (25 marks)

a. (4 marks)

The 30 observations in each of the two samples are for the same pair of organic food items making up the basket of food items. The samples are not independent but related. The appropriate test to be performed would on the 'Mean of the Difference', mu_d, or 'Median of the Difference', Md_{Pop-d}. If the test is on mu_d, it is the 'Paired-t' test; if it is on the Md_{Pop-d}, it is the Wilcoxon test.

b. (6 marks)

One-Sample T: Difference

```
Test of mu = 0 vs not = 0

Variable     N   Mean  StDev  SE Mean       95% CI          T      P
Difference  30  0.361  0.991    0.181  (-0.009, 0.731)  2.00  0.055
```

S1: H0: $mu_d = 0$ Ha: $mu_d \neq 0$

S2: $t_{Calc} = \{(d\text{-}_{Bar}) - 0\} / SE(d\text{-Bar})$

$t_{Calc} = \{0.361) / 0.181 = 1.9945$ (calculations not required)

S3: $t_{Crit} = t_{\alpha/2}(df = n - 1) = t_{0.025}(29)$ (or note p-value = 0.055)

$t_{Crit} = 2.045$

S4: Since $\{|t_{Calc}| = 1.9945\} < \{t_{Crit} = 2.045\}$ (or p-value not < .05) ==> Do not reject H0.

{2 marks for S2 and 1 each for S1, S3 and S4}

Based on the available statistical evidence, there is no change/difference in the mean of the difference of basket of goods at the two supermarket chains. {1 mark}

c. (5 marks)

The 95% CI is:

$d_{Bar} \pm t_{0.025}(29)\ SE(d_{Bar}) = 0.361 \pm 2.045 \times 0.181$

$= 0.361 \pm 0.3701$

$= (-0.0091, 0.7311)$ (manual calc. not req'd) {4 marks}

Please note '0' is contained in the CI above. {1 mark}

d. (5 marks)

Wilcoxon Signed Rank Test: Difference

```
Test of median = 0.000000 versus median not = 0.000000

                N for   Wilcoxon         Estimated
             N   Test  Statistic      P     Median
Difference  30     28      282.0  0.074     0.2500
```

S1: H0: $Md_{Pop\text{-}d} = 0$ Ha: $Md_{Pop\text{-}d} \neq 0$ {2 marks}

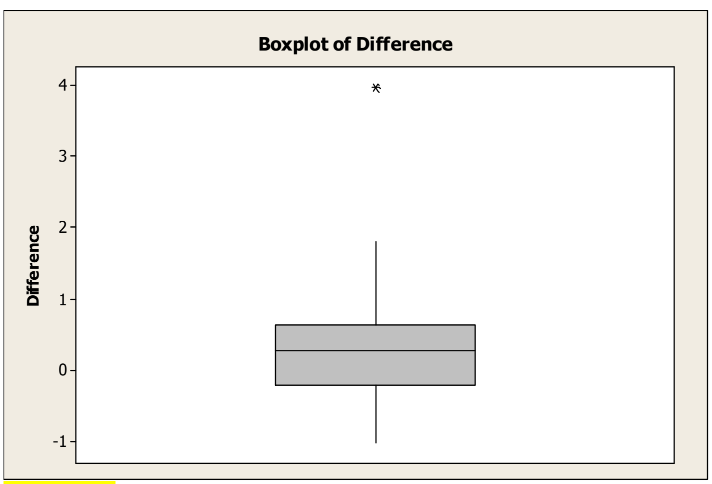
S2, S3, S4 not required

From the MiniTab output,

S5: Since {p-Val = 0.074} > {α = 0.05} ==> Do NOT Reject H0 {2 marks}

Based on the available statistical evidence, one cannot say that the median prices in the populations differ; they are the same {1 mark}

e. (5 marks)



{2 marks} for any reasonable graph

Since the data are related and paired, the boxplot need be checked only for the the 'Difference' sample. It can be seen from the diagram above that this boxplot shows the presence of real/suspected outliers, and therefore the sample and the population it is drawn from are not normally distributed. A non-parametric test is called for and as such, the Wilcoxon test performed in part 'd' above is more appropriate. {3 marks}

Question#4 **(30 marks)**

a. (8 marks)

Here 'r' = # of rows = 5 and 'c' = # of columns = 5

df = (r - 1) x (c - 1)

= (5 - 1) x (5 - 1) = 16 {2 marks}

$E_{11} = (O_{1.} \times O_{.1})/ n = (197 \times 265)/905 = 57.6851$ {1 mark}

$\chi^2_{11} = (O_{11} - E_{11})^2/E_{11} = (50 - 57.6851)^2/57.6851 = 1.0238$ {2 marks}

$E_{35} = (O_{3.} \times O_{.5})/ n = (190 \times 125)/905 = 26.2431$ {1 mark}

$\chi^2_{35} = (O_{35} - E_{35})^2/E_{35} = (25 - 26.2431)^2/26.2431 = 0.0589$ {2 marks}

b.

S1: (3 marks)

H0: The Variables 'Age' and 'Media Preference' are independent {1 mark}

Ha: The Variables 'Age' and 'Media Preference' are not independent {1 mark}

This is called the 'Contingency Test' or 'Test of Independence' based on χ^2 Analysis.

(accept 'test of homogeneity' even though it is not) {1 mark}

c. (7 marks)

S2: $\chi^2_{Calc} = \sum\chi^2_{ij} = 1.0238 + ... + 0.0589 + ... + 0.538 = 24.0405$ {See MiniTab output}

S3: $\chi^2_{Crit} = \chi^2_{\alpha}(df) = \chi^2_{0.05}(16) = 26.30$

S4: Since { $\chi^2_{Calc} = 24.0405$} < { $\chi^2_{Crit} = 26.30$}==> Do Not Reject H0. {2 marks each}

Based on the available statistical evidence you cannot conclude that the variable 'Age' and 'Media Preference' are related or associated (more simply, that they are statistically independent.} {1 mark}

d. (3 marks)

p-Val = P[$\chi^2(16) > 24.0405$] x 1 ≈ 0.075 from the table.

More precise value with MiniTab for p-Val = 0.0886 (0.089 from output) {2 marks}

Since {p-Val = 0.089} > {α = 0.05}==> Do Not Reject H0, just as before in 'c', S4. {1 mark}

e. (9 marks)

S1: H0: p1 = 0.35, p2 = 0.25, p3 = 0.20, p4 = 0.10, p5 = 0.10

Ha: Not all $p_{i's}$ are as above. {2 marks}

S2: $\chi^2_{Calc} = \sum\chi^2_i = \sum(O_i - E_i)^2/E_i$

$\chi^2_{Calc} = 8.4548 + 0.1727 + 0.0884 + 4.2017 + 13.1519 = 26.0695$ {2 marks}

{See the MiniTab output}

S3: $\chi^2_{Crit} = \chi^2_{\alpha}(df) = \chi^2_{0.05}(df = 5 - 1) = \chi^2_{0.05}(df = 4) = 9.49$ (more precisely 9.4877) {1 mark}

S4: Since { $\chi^2_{Calc} = 26.0695$} > { $\chi^2_{Crit} = 9.49$}==> Reject H0. {1 mark}

Based on the available statistical evidence, one cannot state that the 5 proportions are as given. {1 mark}

S5: p-Val = P[$\chi^2(df = 4) > \chi^2_{Calc} = 26.0695$] x 1 < 0.0001 (from Tables) {1 mark}

More precise value is from MiniTab output and is 1 - 0.999969 = 0.000031

Since {p-Val < 0.0001} < {α = 0.05}==> Reject H0. {1 mark}

Appendix: MiniTab Output

Test#1: Two-Sample T-Test and CI: Regional Beer:Cals, National Beer:Cals

```
Two-sample T for Regional Beer:Cals vs National Beer:Cals

                     N   Mean  StDev  SE Mean
Regional Beer:Cals  52  155.8   14.5      2.0
National Beer:Cals  77  142.1   37.8      4.3


Difference = mu (Regional Beer:Cals) - mu (National Beer:Cals)
Estimate for difference:  13.67
95% CI for difference:  (4.25, 23.10)
T-Test of difference = 0 (vs not =): T-Value = 2.88  P-Value = 0.005  DF = 105
```

Test#1: Two-Sample T-Test and CI: Regional Beer:Cals, National Beer:Cals

```
Two-sample T for Regional Beer:Cals vs National Beer:Cals

                     N   Mean  StDev  SE Mean
Regional Beer:Cals  52  155.8   14.5      2.0
National Beer:Cals  77  142.1   37.8      4.3

Difference = mu (Regional Beer:Cals) - mu (National Beer:Cals)
Estimate for difference:  13.67
95% CI for difference:  (2.78, 24.56)
T-Test of difference = 0 (vs not =): T-Value = 2.48  P-Value = 0.014  DF = 127
Both use Pooled StDev = 30.6552
```

Test#2: Two-Sample T-Test and CI: Regional Beer:Cals, National Beer:Cals

```
Two-sample T for Regional Beer:Cals vs National Beer:Cals

                     N   Mean  StDev  SE Mean
Regional Beer:Cals  52  155.8   14.5      2.0
National Beer:Cals  77  142.1   37.8      4.3

Difference = mu (Regional Beer:Cals) - mu (National Beer:Cals)
Estimate for difference:  13.67
95% lower bound for difference:  5.78
T-Test of difference = 5 (vs >): T-Value = 1.82  P-Value = 0.035  DF = 105
```

Test#2: Two-Sample T-Test and CI: Regional Beer:Cals, National Beer:Cals

```
Two-sample T for Regional Beer:Cals vs National Beer:Cals

                     N   Mean  StDev  SE Mean
Regional Beer:Cals  52  155.8   14.5      2.0
National Beer:Cals  77  142.1   37.8      4.3

Difference = mu (Regional Beer:Cals) - mu (National Beer:Cals)
Estimate for difference:  13.67
95% lower bound for difference:  4.55
T-Test of difference = 5 (vs >): T-Value = 1.58  P-Value = 0.059  DF = 127
Both use Pooled StDev = 30.6552
```

Question#3 (25 marks)

One-Sample T: Difference

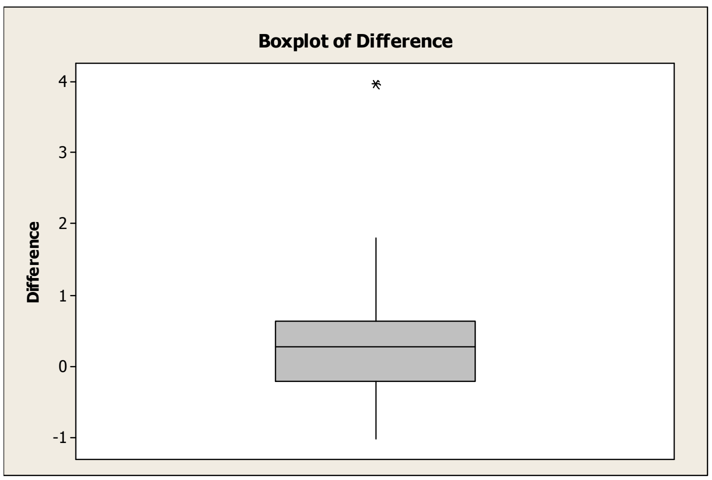
```
Test of mu = 0 vs not = 0

Variable     N   Mean  StDev  SE Mean        95% CI          T      P
Difference  30  0.361  0.991    0.181  (-0.009, 0.731)  2.00  0.055
```

Wilcoxon Signed Rank Test: Difference

```
Test of median = 0.000000 versus median not = 0.000000

                N for   Wilcoxon         Estimated
            N   Test  Statistic      P      Median
Difference  30     28      282.0  0.074      0.2500
```



Question#4 (30 marks)

Chi-Square Test: E-mail, Facebook, Twitter, Instagram, Other

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

     E-mail  Facebook  Twitter  Instagram  Other  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
```

Data Display

Row	C12	Oi	pi	Ei	(Oi-Ei)^2	Chis-i
1	E-mail:	265	0.35	316.75	2678.06	8.4548
2	Facebook:	220	0.25	226.25	39.06	0.1727
3	Twitter:	185	0.20	181.00	16.00	0.0884
4	Instagram:	110	0.10	90.50	380.25	4.2017
5	Other:	125	0.10	90.50	1190.25	13.1519

Data Display

```
K1    905.000              K2    26.0695
K3    5.00000              K4    4.00000
```

Inverse Cumulative Distribution Function

```
Chi-Square with 4 DF

P( X <= x )        x
       0.95  9.48773
```

Cumulative Distribution Function

```
Chi-Square with 4 DF

      x  P( X <= x )
26.0695     0.999969
```