ADM2304 **Assignment #2**
Solution **100 Marks**

Due Date and Time: Friday, March 4, 2015 before 23:59 hrs.
50% Penalty for 1 day late.

Qu.#1 (20 marks)

Here the MiniTab output is given but it is only as a double-check!

Descriptive Statistics: RE_Agency1, RE_Agency2

```
Variable      N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
RE_Agency1   10   0  2887.2     65.4  206.9   2592.0  2715.0  2862.0  3060.0
RE_Agency2   12   0  2803.0     55.3  191.7   2532.0  2667.0  2775.0  2847.0

Variable    Maximum
RE_Agency1   3180.0
RE_Agency2   3180.0
```

Two-Sample T-Test and CI: RE_Agency1, RE_Agency2

```
Two-sample T for RE_Agency1 vs RE_Agency2
                                   SE
             N  Mean  StDev  Mean
RE_Agency1  10  2887    207    65
RE_Agency2  12  2803    192    55

Difference = mu (RE_Agency1) - mu (RE_Agency2)
Estimate for difference:  84.2
95% CI for difference:  (-95.8, 264.2)
T-Test of difference = 0 (vs not =): T-Value = 0.98  P-Value = 0.339  DF = 18
```

a. (9 marks) **Based on Unequal Variances**

(2 marks) S1: $H_0: \mu_1 - \mu_2 = (\Delta_0 = 0.0)$

$H_A: \mu_1 - \mu_2 \neq (\Delta_0 = 0.0)$ (Two-Tailed Test)

(2 marks) Here, $SE(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{206.9^2}{10} + \frac{191.7^2}{12}} = 85.6923$

And df = 18 (see the MiniTab output with Unequal Variances.)

(2 marks) S2: $t_{Calc} = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{SE(\overline{X}_1 - \overline{X}_2)} = \frac{(2887.20 - 2803.00) - 0}{85.6923} = 0.9826 \approx 0.98$

With $\alpha = 0.05$

(1 mark) S3: $t_{Crit} = t_{\alpha/2}(df=18) = t_{0.025}(18) = 2.101$ (≈ 2.1)

N.B. The values of df = 18 and p-Val = 0.339 are taken from the printout and the p-Val will be calculated in part 'c'.

(2 marks) S4: Since $\{|t_{Calc}| = 0.9826\} < \{t_{Crit} = 2.101\}$ ➔ **Do not Reject** H_0.

{This is a Parametric Test for testing the Difference in Means of two Populations using two Independent Random Samples with Unequal Variances.}

b. (6 Marks)

p-Val = P[t(18) > $|t_{Calc}|$ = 0.9826] x 2 --- 1 mark

= [p > 0.10] x 2 with the table only approximate value

> 0.2 can be obtained. --- 1 mark

More precise value can be found by using MiniTab

```
K1     0.830586       # P[t(18)< 0.9826]
K2     0.169414       # P[t(18)> 0.9826]
K3     0.338828       # p-Val = P[t(18)> 0.9826] x 2    --- 3 marks for
                                                                MiniTab
```

Since {p-Val = 0.3389/(> 0.2)} > {α = 0.05}==> Do not reject H0.--- 2 marks

Based on the available statistical evidence, we cannot claim that the two population means are different.

{Other things being equal, without MiniTab, maximum 4 marks; with MiniTab maximum full 6 marks}

c. (5 marks)

The Confidence Interval is given by:

$$\overline{X}_1 - \overline{X}_2 \pm \{t^*(18)\}\{SE(\overline{X}_1 - \overline{X}_2)\}$$

= (2887.20 - 2803.00) ± 2.101 x (85.6923)

= 84.20 ± 180.0395

CI: (-95.8395 to 264.2395) (3 marks)

{The MiniTab printout gives an acceptable CI as:

CI:(-95.8 to 264.2)

This CI is Consistent with the Conclusion reached in part 'a', since it **does** straddle '$\Delta_0 = 0$'. (2 mark)

Qu.#2 (25 marks)

Two-Sample T-Test and CI: RE_Agency1, RE_Agency2: Equal Variance

```
Two-sample T for RE_Agency1 vs RE_Agency2
                                    SE
             N  Mean  StDev  Mean
RE_Agency1  10  2887    207    65
RE_Agency2  12  2803    192    55

Difference = mu (RE_Agency1) - mu (RE_Agency2)
Estimate for difference:  84.2
95% CI for difference:  (-93.3, 261.7)
T-Test of difference = 0 (vs not =): T-Value = 0.99  P-Value = 0.334  DF = 20
Both use Pooled StDev = 198.6864

MTB > mann c6 c7
```

Mann-Whitney Test and CI: RE_Agency1, RE_Agency2

```
             N  Median
RE_Agency1  10  2862.0
RE_Agency2  12  2775.0

Point estimate for ETA1-ETA2 is 84.0
95.6 Percent CI for ETA1-ETA2 is (-84.0,288.1)
```

```
W = 128.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.3913
The test is significant at 0.3903 (adjusted for ties)
```

a. (9 marks) **Based on Equal Variances**

(2 marks) S1: $H_0: \mu_1 - \mu_2 = (\Delta_0 = 0.0)$

$H_A: \mu_1 - \mu_2 \neq (\Delta_0 = 0.0)$ (Two-Tailed Test)

$$s_P^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)} = \frac{9(206.90)^2 + 11(191.70)^2}{(10+12-2)} = \frac{789506.2800}{20} = 39475.3140$$

$s_P = \sqrt{39475.3140} = 198.6840$ (1 marks)

{As in the MiniTab output for Equal Variance}. Here,

$$SE(\overline{X}_1 - \overline{X}_2) = s_P\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 198.6840\sqrt{\left(\frac{1}{10} + \frac{1}{12}\right)} = 85.0714$$

And df = $n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ (1 marks)

{See MiniTab Output}

(2 marks) S2: $t_{Calc} = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{SE(\overline{X}_1 - \overline{X}_2)} = \frac{(2887.20 - 2803.00) - 0}{85.0714} = 0.9898 \approx 0.99$

With $\alpha = 0.05$

(1 mark) S3: $t_{Crit} = t_{\alpha/2}(df = 20) = t_{0.025}(20) = 2.086$

(2 marks) S4: Since $\{|t_{Calc}| = 0.9898\} < \{t_{Crit} = 2.086\}$ ➔ **Do not Reject** H_0.

b. (4 marks)

p-Val = $P[t(20) > |t_{Calc}| = 0.9898] \times 2$ (1 mark)

= [p > 0.10] x 2 with the table only approximate value

> 0.2 can be obtained. (1 mark)

More precise value can be found by using MiniTab

```
K1    0.832952      # P[t(18)< 0.9826]
K2    0.167048      # P[t(18)> 0.9826]
K3    0.334096      # p-Val = P[t(18)> 0.9826] x 2          (2 marks)
```

Since {p-Val = 0.3341} > {α = 0.05}==> Do not reject H0. (1 mark)

Based on the available statistical evidence, we cannot claim that the two population means are different.

{Other things being equal, without MiniTab, maximum 3 marks; with MiniTab maximum full 4 marks}

c. (4 marks)

The Confidence Interval is given by:

$\overline{X}_1 - \overline{X}_2 \pm \{t_{.025}(20)\}\{SE(\overline{X}_1 - \overline{X}_2)\}$ (1 mark)

= (2887.20-2803.00) ± 2.086 x (85.0714)

= 84.20 ± 177.4589 (1 mark)

CI: (-93.2589 to 261.6589) (1 mark)

{The MiniTab printout gives an acceptable CI as:(-93.3 to 261.7)}

This CI is Consistent with the Conclusion reached in part ‘a’, since it **does** straddle ‘$\Delta_0 = 0.0$’. (1mark)

d. (No marks)

Equal or Unequal Population variance is determined by:
1. Following considerations:
i. Are the populations similar or dissimilar?
ii. Are s_1 and s_2 sufficiently close to each other
2. **Rule of Thumb:** If n1 and n2 are reasonably small and the ratio (s1/s2) is less than 1.5 to 2.0, it can be considered as a situation where $\sigma_1^2 = \sigma_2^2$ or 'Population Variances are equal'.

Aside: An actual formal statistical test is there to test equality of population variances. It is based on 'F' distribution and it gives reliable results. Although we do not cover this test in ADM2304, it can readily be done. {In this case, the population variances are equal and as such Equal Population Variance test with 'Pooling' is the best choice. This comment is not required.}

e. (1 marks for the MiniTab output)
(4 marks for writing the formal test as shown below.)

S1: H_0: $Md_1 - Md_2 = 0$ {You could write: $\eta_1 - \eta_2 = 0$ or $Eta_1 - Eta_2 = 0$ etc}

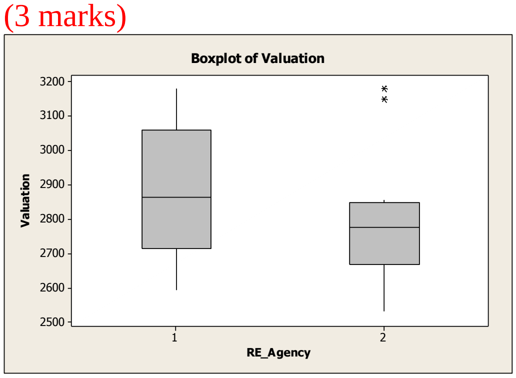
H_A: $Md_1 - Md_2 \neq 0$ (2 mark)

S2, S3 and S4 will **not** be calculated.

S5: Since {p-Val = 0.3903} > { $\alpha = 0.05$}➔ Do **not** Reject H_0. (1 mark)

Based on the available statistical data, one cannot claim that the two population medians are different.
This is a Non-Parametric Test for testing the Difference in Medians of two Populations using two Independent Random Samples. It is the Mann-Whitney Test. (1 mark)

f. (3 marks)



The boxplots of the two independent samples indicate that:

i. Although Sample1is reasonably symmetrical with equally long whiskers, Sample2 is not.

ii. Sample1 has no suspected or real outliers but Sample2 does have.

iii. The medians, or Q2 (the middle horizontal lines) is almost half-way between the first and third quartiles, Q1 and Q3 in Sample2 but not in Sample1.

This suggests that the samples and the populations they come from are **not** reasonably normal. (2 marks for reasoning)

The methodology used in part 'a' Qu.#1 (Two Independent Samples with Unequal Variance Test) and part 'a' Qu.#2 (Two Independent Samples with Equal Variance Test)is Parametric and is **not** appropriate. The methodology used in part 'd' of Qu.#2 (Mann-Whitney Test on Difference in Medians) is Non-Parametric and **is** appropriate here since one of the independent samples is non-normal.
Other possible diagrams: Histograms, NScore Plots etc. (1 mark)

Qu.#3 (25 marks)

Here the MiniTab output is given but it is only as a double-check!

Descriptive Statistics: Loss
Here "Loss" is defined as "Weight1 - Weight2" or the "Difference between the Before and After" weight. This is a "Paired-Sample" data case. The dieting programme to be "Effective", $\mu_d > 0$. It could be defined the other way around making it a left tailed test.

```
Variable   N  N*   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3  Maximum
Loss      10   0  10.08     3.21  10.15    -9.00   4.50   11.70  18.45    23.40

MTB > onet c5
```

One-Sample T: Loss

```
Test of mu = 0 vs > 0
                                            95%
                                          Lower
Variable   N   Mean  StDev  SE Mean  Bound     T      P
Loss      10  10.08  10.15     3.21   4.19  3.14  0.006
```

a. (7 marks)

(1 marks) S1: H_0: $\mu_d = 0.0$

H_A: $\mu_d > 0.0$ (Right Tailed Test)

(1 mark) Here, $SE(\bar{d}) = \frac{s(d)}{\sqrt{n}} = \frac{10.15}{\sqrt{10}} = 3.2097 \approx 3.21$

And , df = n – 1 = 10 – 1 = 9

(1 mark) S2: $t_{Calc} = \frac{(\bar{d}) - \mu_{d_0}}{SE(\bar{d})} = \frac{10.08 - 0}{3.2097} = 3.1405 \approx 3.14$

With LS= $\alpha = 0.05$,

(1 mark) S3: $t_{Crit} = t_{\alpha}(df=n-1) = t_{0.05}(9) = 1.833$

(1 mark) S4: Since $\{t_{Calc} = 3.14\} > \{t_{Crit} = 1.833\}$ ➔ **Reject** H_0.

Based on the available statistical evidence, we can claim that there is a change in the mean of population difference. (1 mark)

This is a Parametric Test for testing the Mean value of the Difference of two Populations using **two Matched Pair Related Samples**. It is the 'Matched Pair 't' Test.' (1 mark)

b. (3 marks)

The Asymmetric Confidence Interval is given by the LB, Lower Bound:

$\bar{d} - \{t_{0.05}(9)\}\{s(\bar{d})\}$ = 10.08 - 1.833 (3.2097)

= 10.08 - 5.8834

LB: (4.1966, ∞ (2 marks)

The MiniTab printout gives an acceptable LB as: (4.19, ∞

This CI is Consistent with the Conclusion reached in part 'a', since it does not contain '$\mu d = 0$'. (1 mark)

c. (2 marks)

Since LB = 4.1966, it implies that $\mu_d > 4.1966$ (1 mark)

The above numbers are in lbs.

In other words, the Mean value of the Difference in "Loss" of weight was at least 4.1966 lbs. (1 mark)

d. (4 marks)

p-Val = P[t(9) >t_{Calc} = 3.1405] x 1{One -Tailed Test}(1 mark)

= {1 – P[t(9) < 3.1405]} x 1 (1 mark)

= {1 – 0.9940} x 1 {Use 'CDF' function from MiniTab}

= 0.006 x 1

= 0.006 (as in the MiniTab output)**]** (2 mark)

{If an approximate answer of 0.01 < p-Val < 0.005 is given after defining what the p-Val is, shown in first line above, give only 2 marks}

e. (5 marks)

Wilcoxon Signed Rank Test: Loss

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

          N for   Wilcoxon           Estimated
      N   Test   Statistic       P      Median
Loss  10     10        50.0  0.025       10.80
```

(1.5 marks for MiniTab output)

(3 marks for writing the formal test as shown below and 1/2 Mark for the name.)

S1: H_0: $Md = 0$ {You could write: $\eta = 0$ or Eta = 0 etc}

H_A: $Md \neq 0$ (1 mark)

S2, S3 and S4 will not be calculated.

S5: Since {p-Val = 0.025} < { $\alpha = 0.05$}➔ **Reject** H_0. (1.5 marks)

Based on the available statistical evidence, there is a difference in the Median of the "Loss" of weight. (1/2 mark)

This is a Non-Parametric Test for testing the Median of the Difference in the two Populations using two Matched Pair Random Samples which are not Independent. It is the Wilcoxon Test. (1/2 mark)

f. (4 marks)

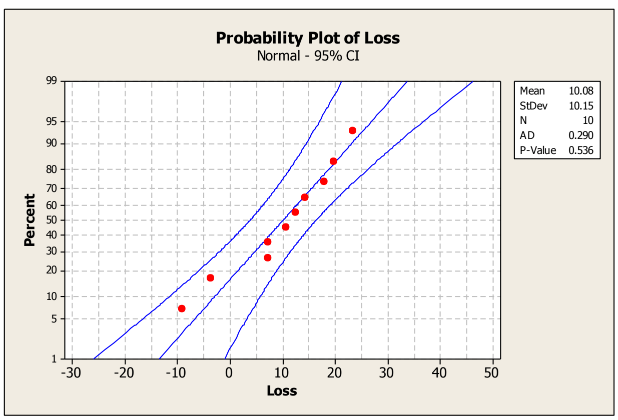
It is essential to see how normal only the **'difference'** data values are. This can be done with the help of Boxplots, Histograms superposed with Normal Curve Diagrams or Normal Score Plots, or Probability Plot of Difference. Here we will use the Probability Plot.

(1 mark)

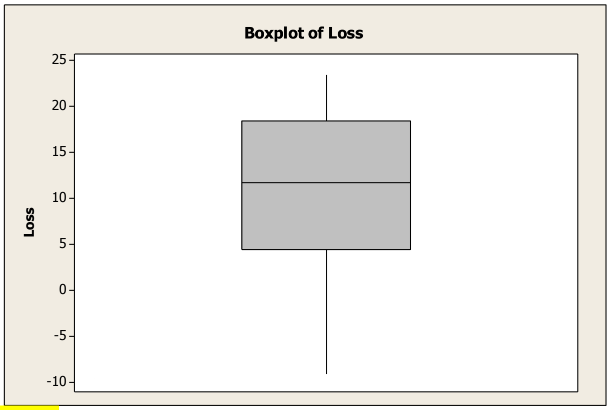
From the Probability Plot below, it will be seen that the "Differences" obtained in the Matched Pair experiment **are** reasonably Normal. This is so because the Probability Plot **does** have the various points lying approximately and symmetrically on and around the regressed straight line within the 95% CI. (2 marks for reasoning)

The methodology used in part **'a' is Parametric and is quite appropriate.** The methodology used in **part 'e' is Non-Parametric and, here it is not appropriate**. (1 mark)

Probability Plot of Difference



Or Boxplot:



Qu.# 4 (20 marks)

(MiniTab approach is, once again used only as a double-check and as a tutorial!)

Test and CI for Two Proportions

```
Sample     X    N  Sample p
1         98  200  0.490000
2        120  300  0.400000

Difference = p (1) - p (2)
Estimate for difference:  0.09
95% CI for difference:  (0.00126968, 0.178730)
Test for difference = 0 (vs not = 0):  Z = 1.99  P-Value = 0.047

Fisher's exact test: P-Value = 0.053
```

a. (7 marks)

{3 Marks for the above calculations}

(1 mark) S1: $H_0: p_1 - p_2 = (\Delta_0 = 0)$
$H_A: p_1 - p_2 \neq (\Delta_0 = 0)$ (Two Tailed Test)

(1 mark) $\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{98 + 120}{200 + 300} = 0.4360$ {You must pool}

$$SD(\hat{p}_1 - \hat{p}_2) = \sqrt{\bar{p} * \bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.436 * 0.564\left(\frac{1}{200} + \frac{1}{300}\right)} = \sqrt{0.002049} = 0.0453$$

(1 mark)

(1 mark) S2: $Z_{Calc} = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta_0}{SD(\hat{p}_1 - \hat{p}_2)} = \frac{(0.49 - 0.40) - 0.0}{0.0453} = 1.9882$

(In the Printout, it is 1.99). With {LS = α} = 0.05

(1 mark) S3: $Z_{Crit} = Z_{\alpha/2} = Z_{0.025} = 1.96$

(1 mark) S4: Since $\{|Z_{Calc}| = 1.9882\} > \{Z_{Crit} = 1.96\}$ ➔ **Reject H_0.**

{Or, since [p-Val = 0.047] < [α = 0.05] ➔ **Reject H_0.**}

Based on the available statistical evidence, we cannot claim that the two population proportions are different. (1 mark)

b. (3 marks)

$$SE(\hat{p}_1-\hat{p}_2)=\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}=\sqrt{\frac{0.49*0.51}{200}+\frac{0.40*0.60}{300}}=\sqrt{0.002050}=0.0453$$

`{No Pooling: it is wrong, although the answer, here, is the same. The above calculation must be shown.}`

(1 mark)

CI: $(\hat{p}_1-\hat{p}_2)\pm Z_{\alpha/2}SE(\hat{p}_1-\hat{p}_2)=(0.49-0.40)\pm 1.96(0.0453)=0.09\pm 0.0888$

Symmetrical CI: (0.0012, 0.1788) (1 marks)

This Symmetrical CI is consistent with the conclusion reached in the Hypothesis Test done in part 'a' since $\Delta_0 = 0$, the difference assumed in H_0 is, not straddled by the CI. (1 mark)

c. (3 marks)

p-Val = P[Z > $|Z_{Calc}|$ = 1.9882] x 2
= {1 - P[Z_{Calc} < 1.99]} x 2
= {1 - 0.9767} x 2
= 0.0466 (≈ 0.047) (2 marks)

Since {p-Val = 0.0466} < {α = 0.05}==> Reject H0. (1 mark)

d. (5 marks)

Here:

{-1 mark for pooling: must **not** use 'Pooling'}

Test and CI for Two Proportions

```
Sample      X    N  Sample p
1          98  200  0.490000
2         120  300  0.400000

Difference = p (1) - p (2)
Estimate for difference:  0.09
95% lower bound for difference:  0.0155352
Test for difference = 0.025 (vs > 0.025):  Z = 1.44  P-Value = 0.076
```

(1 mark) S1: H_0: $p_1 - p_2 = 0.025$
H_A: $p_1 - p_2 > 0.025$ (Right Tailed Test)

(1 mark) S2: $$Z_{Calc}=\frac{(\hat{p}_1-\hat{p}_2)-\Delta_0}{SE(\hat{p}_1-\hat{p}_2)}=\frac{(0.49-0.40)-0.025}{0.0453}=1.4348$$

Here $SE(\hat{p}_1-\hat{p}_2)=0.0453$ `is taken from part 'b'.`

(Same as 1.44 in the Printout). With {LS = α} = 0.05

(1 mark) S3: $Z_{Crit} = Z_{\alpha/2} = Z_{0.05} = 1.645$

(1 mark) S4: Since {$|Z_{Calc}|$ = 1.4348} < {Z_{Crit} = 1.645}➔ **Do Not Reject H_0.**

{Or, since [p-Val = 0.076] > [α = 0.05] ➔ **Do Not Reject H_0.**}

Based on the available statistical evidence, we cannot claim that the two population proportions differ by more than 2.5%. (1 mark)

e. (2 marks)

CI: $(\bar{p}_1-\bar{p}_2)-Z_{Crit}SE(\bar{p}_1-\bar{p}_2)=(0.49-0.40)-1.645(0.0453)=0.09-0.07452$

Asymmetrical CI: $p_1 - p_2 > 0.01548$ (1 1/2 marks)

This Asymmetrical CI is consistent with the conclusion reached in the Hypothesis Test done in part 'a' since Δ_0, the difference assumed in H_0 is more than the lower limit or infemum. (1/2 mark)

Qu.#5 (10 marks)

Row	Income_Class	Oi	pi	Ei	(Oi-Ei)^2	(Oi-Ei)^2/Ei
1	IC1: Poor:	25	0.10	20	25	1.250
2	IC2: Lower Middle_Class:	35	0.20	40	25	0.625
3	IC3: Middle_Class:	80	0.50	100	400	4.000
4	IC$: Upper Middle_Class:	30	0.10	20	100	5.000
5	IC5: Affluent:	20	0.08	16	16	1.000
6	IC6: Wealthy:	10	0.02	4	36	9.000

```
K1    200.000
K2    1.00000
K3    20.8750
K4    6.00000
K5    5.00000
K6    11.0705
```

a. (6 marks)

S1: H_0: $p_1 = 0.10$, $p_2 = 0.20$, $p_3 = 0.50$, $p_4 = 0.10$, $p_5 = 0.08$, $p_6 = 0.02$

H_a: Not all $p_{i's}$ are as above.

S2: $\chi^2_{Calc} = \frac{\sum_{i=1}^{i=k=6}(O_i - E_i)^2}{E_i} = 1.25 + 0.625 + 4.0 + 5.0 + 1.0 + 9.0 = 20.875$

S3: $\chi_{Crit}^2 = \chi_{0.05}^2\ (k-1 = 5) = 11.0705$

S4: Since $\{\chi_{Calc}^2 = 20.875\} > \{\chi_{Crit}^2 = 11.0705\}$ ==> Reject H_0.

Based on the available statistical evidence, the assumed population proportions cannot be supported.

{1 mark each for the 4 steps, S1 to S4. The managerial statement has 1 mark and all the other calculations shown above in the MiniTab output have 1 mark. The calculations can very easily be done manually.}

{N.B. If a student combines categories #5 and #6 to meet the minimum requirement for E_i, here is what happens. $E_5 = 20$ and $O_5 = 30$ with $\chi^2_5 = 5$ and $\chi2_{Calc} = 15.875$ leaving the conclusion unchanged.}

b. (4 marks)

$$\hat{p} = \frac{X}{n} = \frac{60}{200} = 0.30 \text{ and } s(\hat{p}) = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{0.36 * 0.64}{200}} = \sqrt{0.0012} = 0.0339$$

S1: H0: $p = (p_0 = 0.36)$, Ha: $p < (p_0 = 0.36)$ (1 mark)

S2: $Z_{Calc} = \frac{\hat{p} - p_0}{SD(\hat{p})} = \frac{0.30 - 0.36}{0.0339} = -1.7699 \approx -1.77$ (1 mark)

Here S3 and S4 not necessary but given for completeness.

S3: $Z_{Crit} = Z_{0.95} = -1.645$

{Or $Z_{Crit} = Z_{0.05} = 1.645$}
S4: Since {$Z_{Calc} = -1.7699$} < {$Z_{Crit} = Z_{0.95} = -1.645$} ==> Reject H0.
{Or Since {$|Z_{Calc}| = 1.7699$} > {$Z_{Crit} = Z_{0.05} = 1.645$} ==> Reject H0.
S5:
p-Val = $P[Z < |Z_{Calc}| = 1.7699] * 1$
= {1 - 0.9616} * 1
= 0.0384
Since {p-Val = 0.0384} < {$\alpha = 0.05$} ==> Reject H0. (1 1/2 marks)
Based on the available statistical evidence, one can state that the proportion is less than 36% and as such less than 36% will benefit from the changes in tax law.
(1/2 mark)
{If students give S3 and S4 instead of S5, they should get no credit.}