TELFER

VOTRE LIEN AVEC CE QUI COMPTE — CONNECTS YOU TO WHAT MATTERS

ADM 2304
APPLIED STATISTICAL METHODS IN BUSINESS

17 October 2015
9:00 – 11:00

NAME (please print): ______Solutions______________________

Student Number: __________________

SECTION REGISTERED (Circle one):

A (Wednesday/Friday)   B (Monday evening)

Instructions

Length of Exam: 5 pages, plus 2 pages of Minitab output (please return).
Please show all your work and explain your answers briefly where required.
You are encouraged to use the Minitab output as much as possible.
You are permitted to have a non-programmable calculator and a sheet (8.5 x 11 inch) of notes.
Statistical tables (normal and t) are provided separately (please keep).

Marks:								
	________	+	________	+	________	+	________	= ____________
	10		9		8		8	35

Statement of Academic Integrity

The School of Management does not condone academic fraud, an act by a student that may result in a false academic evaluation of that student or of another student. Without limiting the generality of this definition, academic fraud occurs when a student commits any of the following offences: plagiarism or cheating of any kind, use of books, notes, mathematical tables, dictionaries or other study aid unless an explicit written note to the contrary appears on the exam, to have in his/her possession cameras, radios (radios with head sets), tape recorders, pagers, cell phones, or any other communication device which has not been previously authorized in writing.

I have read the text on academic integrity and I pledge not to have committed or attempted to commit academic fraud in this examination.

Signed: ____________________________________

Question 1. [10 marks]

The city wanted to evaluate the effect of new turning restrictions on the reduction of automobile traffic through a residential neighbourhood. It collected data on the number of cars during the peak evening rush hour on a sample of days before and on a sample of days after the restrictions were implemented. Appendix A shows some graphs of the data and some basic summaries.

a. To evaluate the effectiveness of the turning restrictions, should you do an independent samples test or a paired sample test? Explain briefly.

[1] independent samples, since each day before the change is not linked to a day after the change.

b. Test (using p-value) at the 0.05 level of significance whether the data support the city's claim that there was an average reduction of more than 50 cars per minute.

[5]

```
Two-sample T for before vs after

         N   Mean  StDev  SE Mean
before  60  104.0   11.4      1.5
after   60  50.55   7.16     0.92

Difference = μ (before) - μ (after)
Estimate for difference:  53.45
95% lower bound for difference:  50.56
T-Test of difference = 50 (vs >): T-Value = 1.98  P-Value = 0.025  DF = 99
```

-Ho: mu1 – mu2 = 50, Ha: diff > 50
-SE is sqrt(11.4^2/60 + 7.16^2/60)=1.74
-t = (53.45 - 50) / = 3.45/1.74 = 1.98, regardless of equal/unequal variance;
-p-value = Prob(t >1.98) = P(Z > 1.98) = 0.025
-Reject the null H since p-value < 0.05; conclude average reduction more than 50 *(0.5 each)*
1 mark each above; accept pooled stdev = 9.537, with same t-statistic

c. Calculate an appropriate 1-sided confidence interval for the reduction in the average number of cars during the peak evening rush hour.

[2] at least 53.45 – 1.645 * sqrt(11.4^2/60 + 7.16^2/60) or 53.45 – 1.645 * 1.74
or 53.45 – 2.86 = 50.59; interval is (50.6, infinity)

d. What key distributional assumption(s) is(are) absolutely necessary to support your statistical test?

[1] Need to assume the population distributions are not extremely skewed.

e. Is(are) this(these) assumption(s) warranted? Explain with respect to specific boxplots.

[1] The first boxplot is only slightly skewed and the second is symmetric.

Question 2. [7 marks]

Newly developed 'Application Specific Integrated Circuits' or ASICs generally have low 'yield ratios' meaning that only 60% of these chips were useful. In the past, the yield ratio was only 60%. To determine if there has been an improvement, a new sample of 200 ASICs was taken and tested. It was found that 130 of them were useful.

a. Test, using the p-value approach, whether the yield ratio has improved. Assume a level of significance of 5%.

[4] -Ho: p = 0.60, Ha: p > 0.6
-Z = (0.65 – 0.6)/sqrt((0.6*0.4/200)= 0.05 / 0.03464 = 1.44
-p-value is 0.075
-Do not reject Ho, conclude no improvement

1 mark each above.

```
Test of p = 0.6 vs p > 0.6

                                       95%
                                     Lower
Sample    X    N  Sample p       Bound  Z-Value  P-Value
1       130  200  0.650000  0.594524     1.44    0.074
```

b. Calculate the appropriate asymmetric or one-sided confidence interval and explain why it confirms the conclusions you reached in either part 'a' above.

[3] -at least 0.65 – 1.645 * 0.0346 or at least 0.65 – 0.057 = 0.593 *(2 marks)*

-We do not reject Ho since the interval covers the hypothesized value of 0.60 (*1 mark)*

Question 3. [2 marks]

In the current election campaign, all the polls indicate that the major political parties are hovering around 30% of popular vote. Suppose you wanted to estimate a 95% confidence interval for the percentage support for each of the major parties, using a margin of error of plus-or-minus 1%. What sample size would be required?

N = 0.3*0.7 * (1.96/0.01)^2 = 8067   Deduct 1 mark if p=q=0.5 used as this is unrealistic.

Question 4. [8 marks]

Campaign managers are always looking at their own internal polls to figure out if their campaigns are achieving their goals. Here is a brief summary of the various polling numbers a campaign manager obtained 'This Week' and 'Four Weeks Back' for the candidate running in the current election in a certain riding.

When?	Four weeks back	This week
# In favour of candidate	110	163
Sample size	200	250

In all the following sub-questions, use 5% as the level of significance.

a. Test if the improvement in the level of support exceeds 2%.

[4] -Ho: p1-p2 = 0.02, Ha: p1-p2 > 0.02 (here p1 is the more recent population proportion and p1-p2 is the improvement)
-SE is sqrt(11/20*9/20/200 + 163/250*87/250/250) = 0.0463 (pooled proportion unnecessary) *0.5 marks*
-Z = {(163/250 - 11/20) – 0.02}/0.0463 = (0.102 – 0.02) / 0.0463 = 1.77
-Reject Ho if Z > 1.645 *0.5 marks*
-Decide to reject Ho, conclude improvement exceeds 2%

A common mistake was to calculate
Z = {(0.55 – 0.652) – 0.02}/0.0463 = -0.122/0.0463 = -2.51
This requires Ha: p2-p1 < - 0.02 (here p2-p1 must be a "negative" drop in support), where the appropriate Z is {(0.55 – 0.652) - (- 0.02)}/0.0463 = -0.082/0.0463 = -1.77
with rejection region Z < -1.645.

Then the CI in (b) to estimate the drop would be an UB of -0.102 + 0.076 = -0.026
or [-1, -0.026], which does not cover the hypothesized value of -0.02.
It is not a LB of -0.026 and a CI of [-0.026, 1], which of course covers 0.02.
Clearly it is much more straightforward to think of a (positive) improvement.

Test and CI for Two Proportions

```
Sample    X    N  Sample p
1       163  250  0.652000
2       110  200  0.550000

95% lower bound for difference:  0.0258185
Test for difference = 0.02 (vs > 0.02):  Z = 1.77  P-Value = 0.038
```

b. Calculate an asymmetric or one-sided 95% confidence interval to estimate the improvement in the level of support.

[2] at least 0.102 – 1.645 * 0.0463 or at least 0.102 – 0.076 or 0.026

c. What is the minimum value of this improvement? Does your interval support your conclusion in part b? Explain.

[2] Since the minimum is 2.6%, the confidence interval does not cover the hypothesized value of 2%; therefore, we reject the null hypothesis.

Question 5. [8 marks]

Appendix B shows comparison data on the return on equity (ROE) in 2007 and 2008 for a sample of financial institutions. You are interested in seeing if the data support a claim that the average ROE declined from 2007 to 2008.

a. Explain whether an independent samples or a paired sample test is more appropriate.

[1] Paired test since we have ROEs in two years for the same financial institutions (note mention of a sample of ….)

b. Given your answer in part a, test at the 0.05 level of significance whether the data support your claim.

[5]

```
Paired T for ROE07 - ROE08

              N     Mean    StDev  SE Mean
ROE07        27  12.9035  17.6623   3.3991
ROE08        27   9.6128  15.0869   2.9035
Difference   27  3.29067  9.31277  1.79224

95% lower bound for mean difference: 0.23378
T-Test of mean difference = 0 (vs > 0): T-Value = 1.84  P-Value = 0.039
```

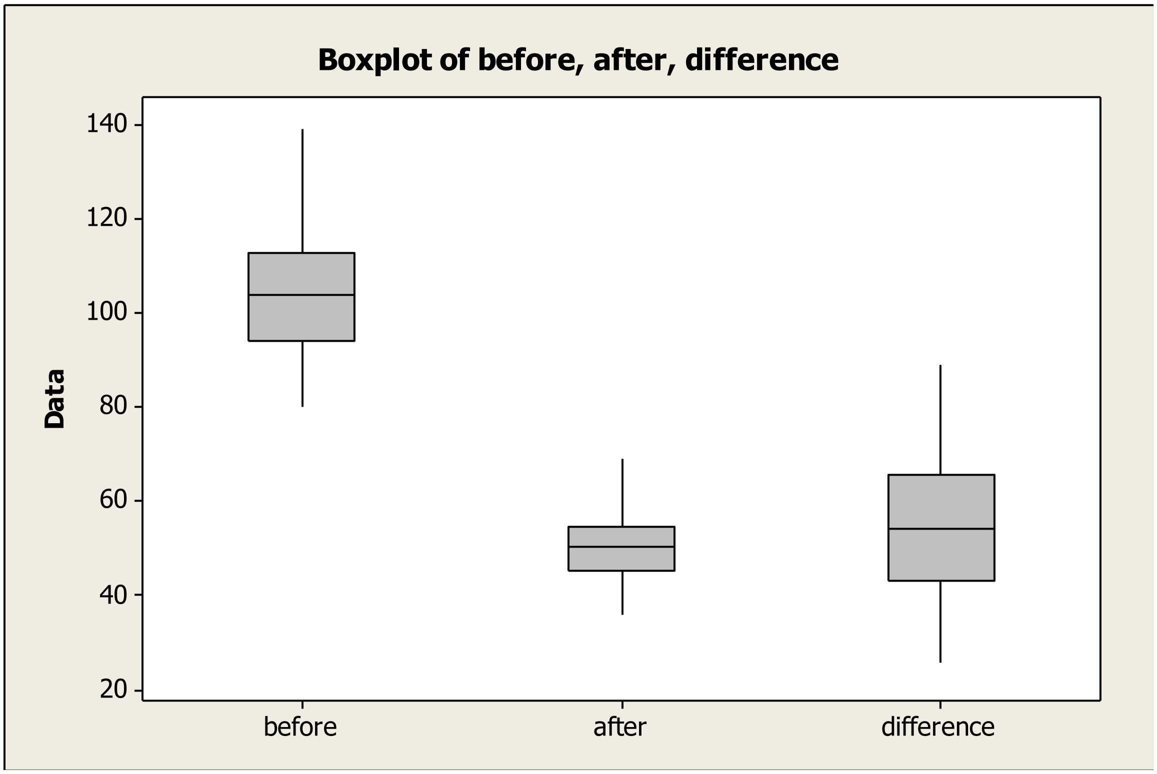
-Ho: mu(diff) = 0, Ha: mu(diff) > 0
-T = 3.29/1.79 = 1.84 (2-sample t=0.74, with $s_p = 15.56$, SE=4.47 and 50 df)
-Reject Ho if T > 1.71 (1.68 for 2-sample t) or since p-value between 0.025 and 0.05 < alpha
-Reject Ho and conclude ROE has declined (*0.5 mark each point)*

1 mark each above.

c. Given your test in part b, what assumptions do you need to make regarding the underlying population distribution(s)? Explain with reference to specific boxplot(s) whether these assumptions are warranted.

[2] -Since the sample of differences is small, we need to assume that the population of changes is normally distributed.
-Since this is a paired test, we only look at the boxplot of differences (*must mention specific boxplot)*
-The assumption is warranted since the boxplot is relatively symmetric and there are no outliers.

Appendix A



Paired T-Test and CI: before, after

```
Paired T for before - after

              N      Mean    StDev  SE Mean
before       60   104.000   11.431    1.476
after        60    50.550    7.158    0.924
Difference   60   53.4500  13.3714   1.7262


T-Test of mean difference = 50 (_________): T-Value = _____  P-Value = _____
```

Two-Sample T-Test and CI: before, after

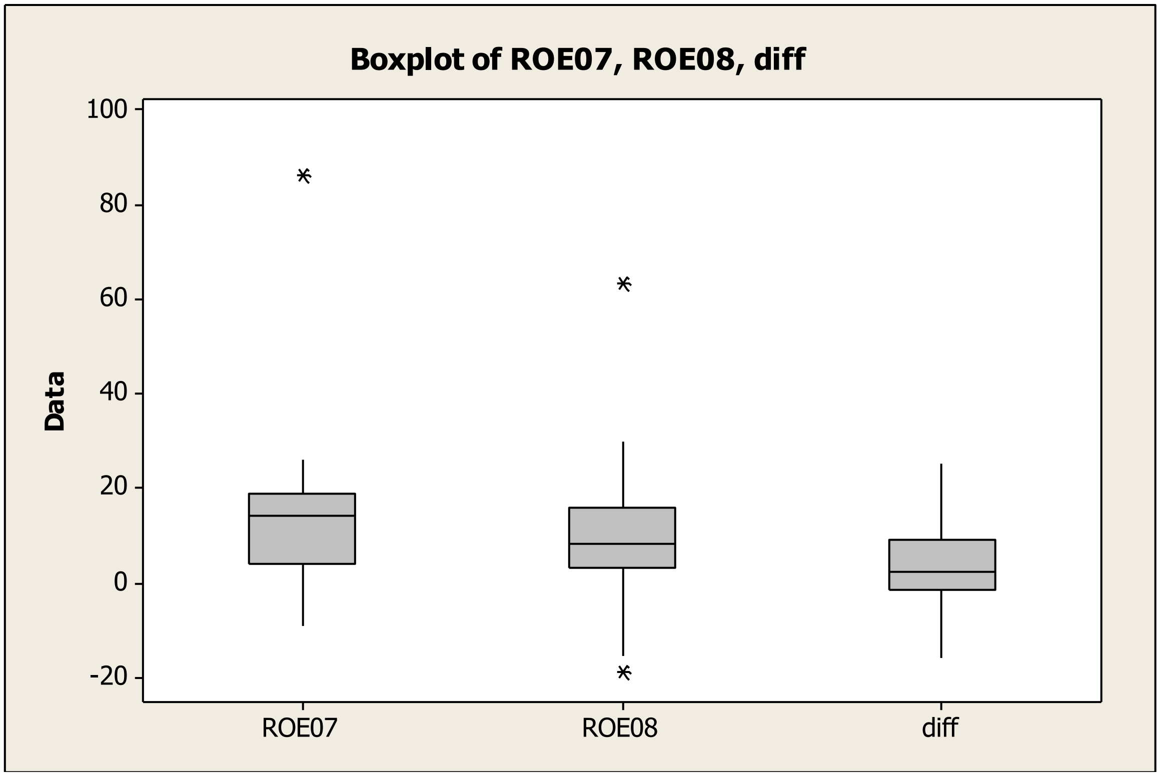
```
Two-sample T for before vs after

         N   Mean  StDev  SE Mean
before  60  104.0   11.4      1.5
after   60  50.55   7.16     0.92


Difference = mu (before) - mu (after)
Estimate for difference:  53.4500

T-Test of difference = 50 (     ): T-Value = _____  P-Value = _____  DF = 99
```

Appendix B



Two-Sample T-Test and CI: ROE07, ROE08

```
Two-sample T for ROE07 vs ROE08

        N  Mean  StDev  SE Mean
ROE07  31  13.9   16.7      3.0
ROE08  31  10.2   14.3      2.6


Difference = mu (ROE07) - mu (ROE08)
Estimate for difference:  3.71487

T-Test of difference = 0 (_____): T-Value = ______  P-Value = ______  DF = 58
```

Paired T-Test and CI: ROE07, ROE08

```
Paired T for ROE07 - ROE08

              N     Mean    StDev  SE Mean
ROE07        31  13.8862  16.7199   3.0030
ROE08        31  10.1713  14.3046   2.5692
Difference   31  3.71487  8.79247  1.57917


T-Test of mean difference = 0 (     ): T-Value = _____  P-Value = ______
```