TELFER

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

ADM 2304
APPLIED STATISTICAL METHODS IN BUSINESS

15 October 2016, 9:00-11:00

Name (print) ______________________________ Student Number ____________________

SECTION REGISTERED (Circle one): A (day) B (evening)

Instructions

Length of Exam: 5 pages, plus 3 pages of Minitab output (please return all).
Please show all your work and explain your answers briefly where required. Make sure your tests of hypotheses are completed with all the necessary steps.
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.
Three pages of statistical tables (normal, t and binomial) are provided in Appendix C.

Marks:									
	________	+	________	+	________	+	________	=	____________
	10		10		10		10		40

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 headsets), 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]

Ohio is a key swing state in the U. S. presidential election. No Republican nominee has ever won without winning Ohio's electoral votes. On October 3, a Quinnipiac poll based on 497 likely voters showed Trump with 47% and Clinton with 42% of the sample in Ohio. On October 5, a Monmouth poll based on 405 likely voters showed Trump with 42% and Clinton with 44% of the sample in Ohio.

a) Keeping the probability of a Type I error at 0.05, perform a test to determine whether there is a statistically significant difference in the support for Trump between the two polls. Use the critical value approach.

[5]

```
          Sample    X    N  Sample p
          1       234  497  0.470825        Trump1
          2       170  405  0.419753        Trump2

Estimate for difference:  0.0510719
95% CI for difference:  (-0.0140122, 0.116156)
Test for difference = 0 (vs not = 0):  Z = 1.53  P-Value = 0.125

Sample    X    N  Sample p
1       209  497  0.420523        Clinton1
2       178  405  0.439506        Clinton2

Difference = p (1) - p (2)
Estimate for difference:  -0.0189830
95% CI for difference:  (-0.0839450, 0.0459790)
Test for difference = 0 (vs not = 0):  Z = -0.57  P-Value = 0.567
```

-1 for hypotheses, 1 for pooled proportion of 0.448 (for Trump), 1 for test statistic, 1 for rejection region |z| > 1.96, 1 for decision (not to reject Ho) and conclusion of "no difference".

b) Find the p-value for the test statistic in (a).

[1] P(|z|>1.53) = 2*0.063 = 0.126
P(|z| > 0.57) = 2*0.2843=0.569

c) Now calculate the appropriate 95% confidence interval for the difference in popular support for Trump between the two surveys.

[2] 0.05 +/- 1.96*sqrt(.47*.53/497 + .42*.58/405) = 0.05 +/- 1.96*0.0332 = (-0.015, 0.115)
(for Trump), or (for Clinton):
0.02 +/- 1.96*sqrt(.42*.58/497 + .44*.56/405) = 0.02 +/- 1.96*0.0331 = (-0.045, 0.085)

d) Explain whether and how the three approaches above are or are not consistent.

[2] p-value not < 0.05 and CI covers zero, therefore do not reject Ho; therefore all three approaches are consistent. (Many stated incorrectly that the CI should cover α or the p-value.)

Question 2. [10 marks]

Appendix A shows, for a sample of Ottawa-Nepean homes, both the 2015 and the 2016 assessed values in thousands of dollars. Suppose you are interested in determining whether the typical increase in assessed values is more than $18,000.

a) Explain whether you should do an independent samples test or a paired samples test.

[1] There is only one sample of homes. Must explain why the samples are dependent (because the assessed values are for the same homes). Paired samples test is appropriate.

b) Given your answer to (a), identify the most appropriate test. Explain briefly, with reference to specific boxplots.

[2] For paired sample test, only the differences matter. The differences boxplot suggests it is reasonable to assume a normal distribution of differences. Therefore, the parametric matched t-test is appropriate.

If samples are considered independent in (a), then look at the two samples. Since it is not reasonable to assume these samples come from normally distributed populations, the non-parametric Mann-Whitney is appropriate.

c) Notwithstanding your answer to (b), but being consistent to your answer to (a),

i. Perform the appropriate parametric test to determine whether there is sufficient evidence to conclude that the mean increase is more than $18,000. Use the 0.05 level of significance.

[4] Ho: $\mu_d = 18$; Ha: $\mu_d > 18$
$T = (20.73 - 18) / (4.45/\sqrt{15}) = 2.73/1.15 = 2.37$
Rejection region is one-sided: $t > 1.76$ (prob of 0.05 and 14 df), not $|t| > 1.76$.
We reject the null H and conclude the mean increase exceeds $18,000.

ii. Now perform an appropriate non-parametric test to determine whether there is sufficient evidence to conclude that the median increase is more than $18,000. Use the 0.05 level of significance.

[3] Ho: pop med diff =18; Ha: pop med diff > 18
p-value = 0.021 < 0.05
reject Ho and conclude the median increase exceeds $18,000

Question 3. [10 marks]

Appendix B shows some graphs and summaries of median household incomes for two samples of neighbourhoods, one from Calgary (HhldInc_Cal) and the other from Edmonton (HhldInc_Edm).

a) Test at the 0.05 level of significance to determine if there is evidence of a difference in average household income between the two cities. Use a parametric test and do not assume equal population variances.

[5] -Ho: $\mu 1 - \mu 2 = 0$; Ha: $\mu 1 - \mu 2 \neq 0$
-t = (90716 – 80444) / SE, where SE = sqrt(26715^2/59 + 20688^2/60) = 4385.166, so
t = 10273 / 4385 = 2.34 (2 marks)
-rejection region is |t| > 1.96
-Reject Ho, conclude there is evidence of a difference.

b) Find the p-value for the result above.

[1] p-value is P (|t| > 2.34) ≅ P(|z| > 2.34) = 2*P(z < -2.34) = 2 * 0.0096 = 0.0192.
Using t-dist with 109 (or 100) df, p-value very close to 2*0.01 = 0.02,
since t(0.01) = 2.36 for 100 df.

c) Now calculate an appropriate 95% confidence interval for the difference in average household income between the two cities.

[2] 10273 ± 1.96 * 4385 = 10273 ± 8595 = (1678, 18868)
More accurately, 10273 ± 1.98 * 4385 = 10273 ± 8682 = (1591, 18955)

d) What assumptions regarding the population data are *required* to justify your calculations above? Looking at the boxplots of the two samples of median incomes, explain whether these assumptions are reasonable.

Necessary assumption(s): since both samples are large (> 30), we only need to assume the populations are not extremely skewed (we require the data to be normally distributed only if the samples are small).

[2] Clearly some skewness exists particularly in the Calgary sample, but it is reasonable to assume the populations are not extremely skewed.

Some solutions stated "boxplots not too skewed" as an assumption. This is an observation, not an assumption regarding the population data. The observation makes the assumption reasonable.

Question 4. [10 marks]

A CPA with a large accounting firm wants to select a random sample of transactions to determine if the population deviation rate (the proportion with material errors) is below the tolerable rate of 7%.

a) What hypotheses should he test to be able to conclude that the population deviation rate is below 7%? (This means that the control procedures are operating effectively.)

[1] Ho: p = 0.07; Ha: p < 0.07

b) Based on his previous auditing experience, he expects to find a deviation rate of 4%. What sample size should he select if he wants to calculate the appropriate 95% confidence interval that will allow him to reject the null hypothesis?

[3]	0.04*0.96*(1.645/0.03)^2 = 116	0.07*0.93*(1.645/0.03)^2 = 196
	0.04*0.96*(1.96/0.03)^2 = 164	0.07*0.93*(1.96/0.03)^2 = 278

-1 mark for identifying ME of 0.03 (0.07 – 0.04)
(To reject Ho, the CI must have an UB of 0.04 + ME < 0.07. Thus, ME < 0.07-0.04.
-1 mark for identifying 1.645 for a 1-sided CI, or 0.5 for using 1.96 for a 2-sided CI
-1 mark for using either 0.04 or 0.07 as the value of the proportion.

c) In your calculations above, you are assuming that the
__sample proportion__ has a __normal__ distribution.
(A binomial population or a binomial sample cannot have a normal distribution, and a population proportion is a fixed value and cannot have a distribution.)
Assuming the null hypothesis is true, this is reasonable since
np = 196*.07 > 10, np = 278 * 0.07 > 10,
or not reasonable since np = 115*.04 < 10, np = 164 * 0.04 < 10.

d) Suppose he ends up taking a sample size of 200 and finds only 6 transactions with errors. Find the exact p-value of this result and state your final conclusion for the test in (a).

[2] the exact p-value is P(X ≤ 6) = 0.0119 < 0.05, based on n=200 and p=0.07.
Since the p-value < 0.05 , we reject Ho. The control procedures are operating effectively.

Since np = 200*0.07 > 10, the approximate p-value is based on
z = (0.03 – 0.07) / sqrt(0.07*0.93/200) = -.04 /0.018 = -2.22. Here,
p-value = P(z < -2.22) = 0.0132. This leads to the same conclusion as above.
This approximate solution is only worth 1 mark.