Solutions

STAT 651 Final Test (2 hours) **December 10th, 2013**

NAME: **Total number of Marks: /40**

Answer all the questions in the exam (questions are on both sides of the paper). There are 8 questions in this paper. Unless stated otherwise, do all tests at the 5% level.

Advice: Look at the marks allocated for each question and don't spend a disproportionately large amount of time on any one question.

When conducting a test, always state your null and alternative. Also state the distribution and/or the test that you use.

Write your solutions in the question paper.

(1) According to the National Bureau of Statistics (NBS), the mean number of pets in a household is 0.8.

A television show sets out to check this claim. They randomly sample 100 customers who go to PetCo (a pet supplies shop) and asks each person how many pets they have in their household. The sample mean (average of this sample) was 1.5. The television company test whether this difference is statistically significant $H_0 : \mu \leq 0.8$ against $H_A : \mu > 0.8$. They get a p-value of 0.01% and thus determine that the NBS is wrong and the mean number per household is greater than 0.8.

What is wrong with the data collection/methodology of the television show? Your answer should be no more than two lines. [2]

The sample will be biased since it collected for a place which caters for pets. This sample will be biased towards higher outcomes

(2) Eric wants to use statistics to 'prove' his research. As he learnt in his statistics class he stated his research hypothesis as the alternative and the oppositive of that as his null hypothesis.

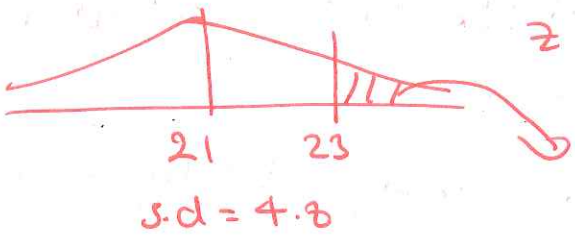
After collecting the data, and doing the test at the 5% level, he is unable to reject the null. Disappointed with this result, he collects another sample does the test at the 5% level. Again he is unable to reject the null. He does this 8 times and, finally, on the eight attempt rejects the null. He then writes an article saying that 'there is evidence to reject the null hypothesis, since his p-value was less than 5%'.

Comment on the validity of his comment. Your answer should be no more than two lines. [2]

Suppose the null is true the probability he will be able to reject the null at least once is 8 attempts is $1-(0.95)^8 = 33.6$. This is a large probability, indeed it can be viewed as a p-value; therefore there is no evidence the alternative is true.

(3) Many high school students take ACT exams. A student can get a grade from 0 to 36 (but only integer values, ie. 0, 1, 2,..., 36). This year the national mean ~~mean~~ score of an ACT test is 21 with standard deviation 4.8 (the maximum score is 36).

(i) Assume that the distribution of grades is close to normal. Calculate the probability that a student scores 23 or greater in the exam. [2]



$z = \frac{23-21}{4.8} = 0.416$

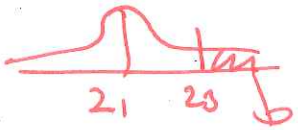
$= 33.8\%$

(ii) Using the information given about the grading scheme, explain why the normal approximation and probability calculated cannot be particularly accurate. [1]

The grades are integer value. The normal distribution is for continuous variables. The grade distribution is unlikely to be close to normal.

(iii) A high school teacher teaches a class of 20 students. He wants to know the probability that his class average will be over 23. Calculate this probability and explain why this probability will be more accurate of the truth than the probability calculated in (i). [3]

$s.e = \frac{4.8}{\sqrt{20}} = 1.07$



$z = \frac{23-21}{1.07} = 1.86$

Area to right $= 3.08\%$

The CLT means the sample mean based on 20 observations will be closer to normally distributed than the original population.

(iv) A high school teacher is using the recently introduced Flipping method to teach his class. He believes this method may lead to an improvement in the mean ACT grade. The average (sample mean) ACT grade in his class of 20 was 23 ($\bar{x} = 23$). State the null and alternative and do the test at the 5% level. [3]

$H_0: \mu \leq 21$ $H_A: \mu > 21$ the p-value $= 3.08\%$.

There is evidence to reject the null, that is, his teaching has improved average grades.

(4) A student society is staging a campaign for people to use the stairwell in Blocker. Before the campaign they randomly sample 300 people in Blocker and find that 30% of them used the stairwell. After a concerted campaign to use the Stairwell (with many posters) they interview 250 people and find that 40% now use the stairwell. The society wants to test whether this different is statistically significant. Name the TWO tests (together with and the hypotheses for each test) that they can use. [4]

1) two sample test for proportion $H_0: p_B - p_A \geq 0$ $H_A: p_B - p_A < 0$

A chi-squared test for independence

H_0: There is no dependence between campaing and stairwell use

H_A: There is a dependence

(5) Let us return to the student society health campaign to use the stairwell. The society wants to see whether the campaign has increased the frequency of stairwell use.

They random interview 10 people before the campaign and then 10 people after the campaign.

They ask each person, on average, how often in one day do they use the stairwell.

The JMP output is giveb below. Note that **1 = Before** the Campaign and **2 = After** the Campaign.

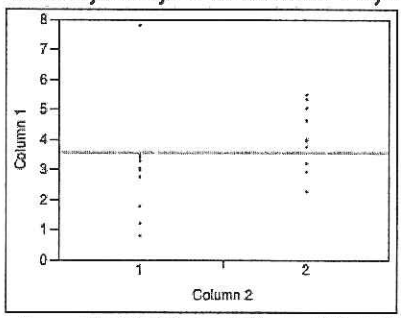
Before	1 3.277	1 3.48	1 3.036	1 1.217	1 2.753	1 3.4	1 0.79	1 7.8	1 2.981	1 1.779
After	2 3.985	2 3.94	2 5.487	2 5.328	2 2.258	2 3.207	2 2.921	2 3.752	2 5.048	2 4.634

(i) State the null and alternative hypothesis of interest. [1]

$H_0: \mu_A - \mu_B \leq 0$ $H_A: \mu_A - \mu_B > 0$

The results of the tests are given on the test page.

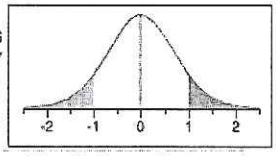
Oneway Analysis of Column 1 By Column 2



t Test

2-1
Assuming unequal variances

Difference	1.0047	t Ratio	1.446246
Std Err Dif	0.6947	DF	14.12197
Upper CL Dif	2.4935	Prob > |t|	0.1699
Lower CL Dif	-0.4841	Prob > t	0.0850
Confidence	0.95	Prob < t	0.9150



Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
1	10	79.000	105.000	7.9000	-1.928
2	10	131.000	105.000	13.1000	1.928

(ii) What are the results of the independent sample t-test at the 5% level? [2]

(6) Not enough evidence to reject null since p-value $\geq 8.5\%$.

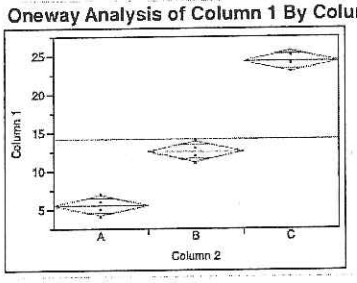
(iii) What are the results of the Wilcoxon test at the 5% level? [2]

Since it a one-sided test the point of rejection is 127. Since $131 > 127$, the p-value is less than 5% and there is evidence the campaign increased use.

(iv) Explain why there is a difference between the conclusion in the t-test and the Wilcoxon test. [2]

There is one person in the data set who used the shuttle 7.8 times (even before the campaign). This was a large outlier ~~and~~, which increased the sample mean and the s.d. The Wilcoxon test gave less weight to the outlier. ~~and~~

(6) 4 samples are taken from three populations (A,B and C) are taken.



Oneway Anova

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Column 2	2	718.16667	359.083	253.4706	<.0001*
Error	9	12.75000	1.417		
C. Total	11	730.91667			

A	A	A	A	B	B	B	B	C	C	C	C
4	5	6	7	11	12	13	14	23	24	25	25

The data and results of an ANOVA are summarised above.

(i) By studying the data, do you believe the standard assumptions are satisfied to do an ANOVA? [1]

No, the data does not appear to have a normal distribution as the observations are integers.

(ii) State the null and alternative of interest and the results from the ANOVA at the 5% significance level. [3]

$H_0: \mu_A = \mu_B = \mu_C$ H_A: At least one mean is different.

The p-value is very small $< 0.01\%$. This there is evidence to reject the null.

(iii) How does your answer in (i) influence (ii). Do you think it will have any impact on the conclusions of the test (your answer should be no more than two lines)? [2]

No, the data is so separated and the p-value is so small, that the normality assumption used to obtain the p-value won't effect it't excessly small size. The F-value = 253.4 is so large & there is overwhelming that there is a lot of evidence agan the null.

(7) The National Institute of Health would like to see whether parental smoking has an influence on their children smoking. They randomly select 5384 high school students. They ask each student whether they smoke and whether both, one of neither of their parents smoke. The data is summarized below.

Contingency table results:
Rows: Parent
Columns: None

Cell format
Count (Row percent) (Column percent) (Total percent)

	Student smokes	Does not smoke	Total
Both smoke	400 (22.36%) (39.84%) (7.429%)	1389 (77.64%) (31.71%) (25.8%)	1789 (100.00%) (33.23%) (33.23%)
One smokes	416 (18.58%) (41.43%) (7.727%)	1823 (81.42%) (41.62%) (33.86%)	2239 (100.00%) (41.59%) (41.59%)
None smoke	188 (13.86%) (18.73%) (3.492%)	1168 (86.14%) (26.67%) (21.69%)	1356 (100.00%) (25.19%) (25.19%)
Total	1004 (18.65%) (100.00%) (18.65%)	4380 (81.35%) (100.00%) (81.35%)	5384 (100.00%) (100.00%) (100.00%)

(i) What is the probability that a student smokes (regardless of whether their parents smoke)? [1]

18.65%

(ii) What is the probability that student smoke given that both their parents smoke. [1]

22.36%

(iii) Suppose there is NO dependence between a student smoking and their parent smoking. How many of the 1789 students whose (both) parents smoke would you expect to smoke? [1]

0.1865×1789

(iv) You want to test whether there is an association between student smoking and their parents smoking. What is the null and alternative hypothesis? [2]

H_0: There is no association

H_1: There is an association between parent and child smoky

(v) You do a chi-squared test and get the output:

Statistic	Value
Chi-squared	36.7

. State the results of the test (at the 5% level) giving the closest bounds for the p-value. [3]

The 5% level is 7.3

Since $36.7 > 7.3$ the p-value is less than 5%.

and we can reject the null

(8) Gestational diabetes is diagnosed in a pregnant women if the mean level of glucose in her blood is over 140. To provisionally diagnose gestational diabetes three blood samples are taken and the sample mean calculated. If the average (sample mean) is greater than $140 + 1.64 \times \sqrt{\frac{4}{3}} = 141.89$, then gestational diabetes is diagnosed.

A doctor questions why the threshold 141.89 is used to diagnose gestational diabetes and not 140. Explain in terms of type I/II errors and power, why the threshold of 140.89 is used rather than 140. Your answer should be no more than two lines. [2]

By using 140 as the threshold the power of the test would be very large however so would the type I error (the proportion of false positives). By using 141.89 you ensure the type I is 5%.