

Solutions

Tuesday 8am

STAT 651 Final Test (2 hours)

May, 2008

NAME:

Total number of Marks: /40

Answer all the questions in the exam (questions are on both sides of the paper). There are 0+8 questions in this paper.

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.

(0) What is the name of your favourite cuddly bear?

[3]

(1) Scientists are monitoring the growth of the human population over the past 50 years. Each year they estimate the number of humans that year (in billions) and do a linear regression. They estimate $\hat{\beta}_0 = -139.31$ and $\hat{\beta}_1 = 0.073$, which gives the predictive equation $\hat{Y} = -139.31 + 0.0727 \times \text{year}$.

(a) Based on the above predict the population in 1900 and in 2050. [1]

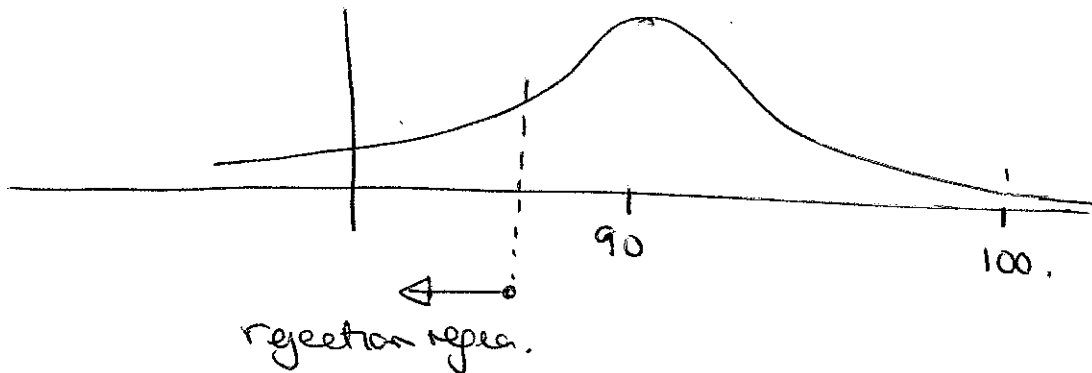
$$\underline{1900} : \hat{Y}(1900) = \textcircled{-139.31} - 139.31 + 0.0727 \times 1900 = -1.18$$

$$\underline{2050} : \hat{Y}(2050) = -139.31 + 0.0727 \times 2050 = 9.725$$

(b) State the main flaw in the above predictions. [1]

Extrapolation, we have no idea of ~~what~~ what's happening before 1950 or after 2008.

(2) I have a sample of size 50. The sample mean is 100 and the sample variance is 30. Suppose I want to test whether the sample mean μ satisfies $H_0 : \mu \geq 90$ against the alternative $H_A : \mu < 90$. I do the test with $\alpha = 1\%$. What are the conclusions of the test? [2]



not enough evidence to reject the null.

(3) You are asked to design an experiment comparing people's attitudes to immigration in city A and city B. You have money to interview a total of 500 people (in total from both cities).

(a) How many people in city A and city B would you interview? Give a reason for your answer. [2]

A : 250 B : 250

The variance of the difference $\bar{x} - \bar{y}$ is smallest.

(b) Suppose you interview 300 people in city A and 200 people in city B. 160 people in city A were pro-immigration and 140 were anti-immigration. Whereas in city B 120 were pro-immigration and 80 were anti-immigration. Test the hypothesis that there is a difference in opinion between the cities. Remember to precisely state the null and alternative and briefly mention any assumptions that you have made. Do the test at the 5% level. [3]

Proportion
 $\pi_1 =$ no. of people who are pro-immigration in city A
 $\pi_2 =$ proportion of people " " " B

$$\hat{\pi}_1 = \frac{160}{300} = 0.533 \quad \hat{\pi}_2 = \frac{120}{200} = 0.6$$

~~H_0~~ $H_0: \pi_1 - \pi_2 = 0 \quad H_A: \pi_1 - \pi_2 \neq 0$

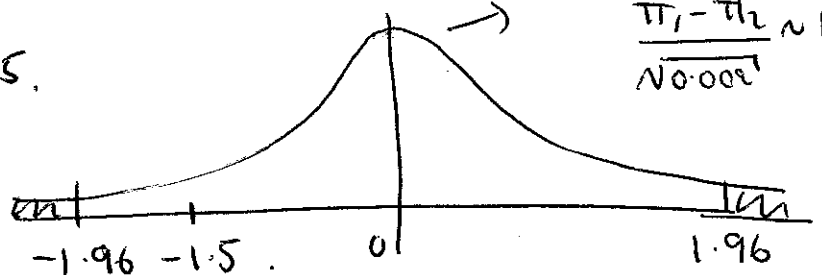
Under null $\hat{\pi}_1 - \hat{\pi}_2$ has variance (approx)

$$= \left(\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{300} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{200} \right)$$

$$\approx 0.0008 + 0.0012$$

$$\approx 0.002$$

Under null $\frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{0.002}} \sim N(0,1)$



p-value
 $= 0.066$

$$\frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{0.002}} = \frac{-0.0667}{\sqrt{0.002}} \approx -1.5$$

not enough evidence
to reject null

(4) You are given the heights of 25 people. The sample mean is $\bar{X} = 170\text{cm}$ and the sample variance is $s^2 = 40$.

(a) Construct a 95% confidence interval for the population mean. [2]

$$\left[170 \pm \overbrace{t_{24}(0.025)}^{\approx 2.064} \times \sqrt{\frac{40}{25}} \right]$$

$$= [167.40, 172.6]$$

(b) You are asked to predict the height of a randomly selected person. Predict that person's height and construct a 95% confidence interval for that person's height. [2]

• It will be the sample mean $\bar{X} = 170\text{cm}$.

The 95% CI is $\left[170 \pm 2.064 \times \sqrt{\frac{40}{25} + 40} \right]$

(c) Previous studies have shown that the mean height is greater than 168cm. State precisely the null and alternative hypothesis and conduct the test at the 5% level. [3]

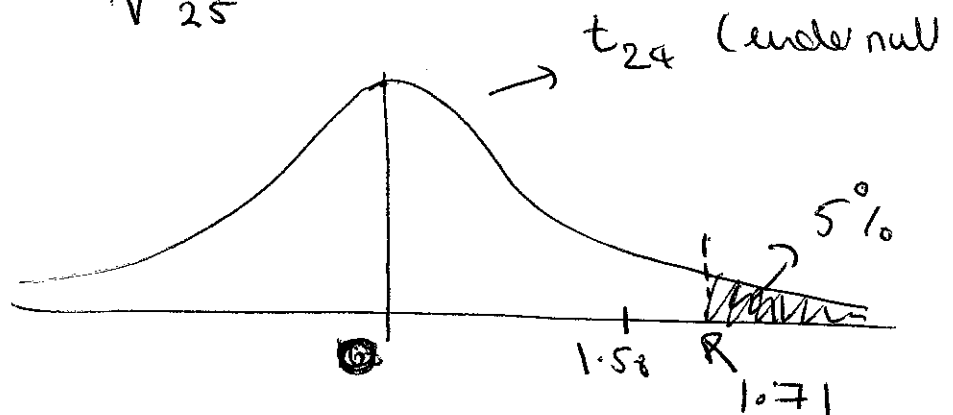
$$H_0: \mu \leq 168\text{cm}$$

$$H_A: \mu > 168$$

Under the null

$$\frac{(\bar{X}) - 168}{\sqrt{\frac{40}{25}}} \sim t_{24}$$

$$\frac{170 - 168}{\sqrt{\frac{40}{25}}} = 1.58$$



not enough evidence to reject the null.

(5) Scientists want to know whether it is nature or nurture that determines a person's preference for sugar. To do the experiment they analysed 30 sets of identical twins who were separated at birth, where one twin grew up in an affluent family while the other grew up in a working class family. They measured the amount of sugar all the twins had consumed on a certain designated day. The sample average of sugar consumed by the twins which were brought up in an affluent family was 20grams and the sample average of sugar consumed by the twins which grew up in a working class family was 19grams.

(a) State precisely (in terms of means) the null and alternative hypothesis that should be investigated. [1]

$\mu_1 = \text{mean amount of sugar working class twin consumes}$

$\mu_2 = \text{mean amount of sugar affluent twin consumes}$

$H_0: \mu_1 - \mu_2 = 0 \quad H_A: \mu_1 - \mu_2 \neq 0$

(b) What test would you recommend the scientists do? [1]

A paired t-test.

(c) The test is done using SPSS and part of the output is given below.

Mean Difference	Std. Deviation	Std. Error Mean
1	5	0.912

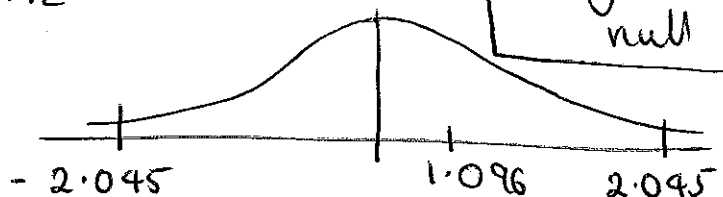
What are the conclusions of the test (state precisely the distribution you use and do the test at the 5% level). [2]

Under the null

$\frac{\bar{D}}{0.912} \sim t_{29}$

not enough evidence to reject null

$\frac{\bar{D}}{0.912} = 1.096$



(d) Use the information above to construct a 95% confidence interval for the mean difference. [2]

$$[1 \pm t_{29}(0.025) \times 0.912]$$

↓
2.045

$$= [-0.865, 2.865]$$

	Sum of Squares	df	Mean square	F	Sig.
Between Groups	248.2	6			
Within Groups	1859	143			
Total					

(6) Suppose the mean height of different varieties of oak trees are compared. An ANOVA is conducted and the analysis of variance table is given above.

(a) Using the table above, how many different varieties of trees were involved in the study? [1]

$$6 + 1 = 7$$

(b) State precisely the null and alternative hypotheses. [1]

$$H_0: \mu_1 = \dots = \mu_7 = 0$$

H_A : not all the means are the same.

(c) Fill in the rest of the table (except for the p-value). [2]

$$SSB = 248.2$$

$$SSW = 1859$$

$$\frac{SSB}{6} = \frac{248.2}{6}$$

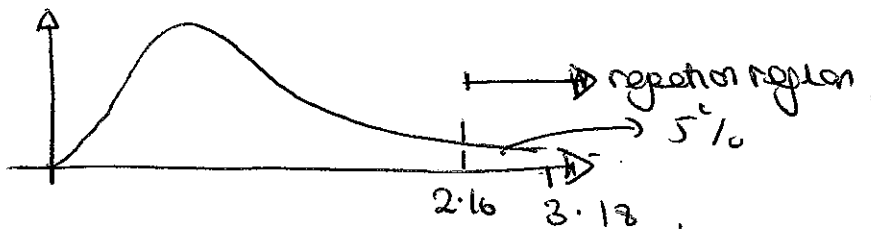
$$= 41.36$$

$$\frac{SSW}{143} = \frac{1859}{143}$$

$$= 13$$

$$F = \frac{SSB/6}{SSW/143} = 3.18$$

(iii) What are the conclusions of the test (do the test at the 5% level)? [1]



Useful information: $F_{6,143}(0.05) = 2.16$, $F_{143,6}(0.05) = 3.70$, $F_{7,144}(0.05) = 2.073$ and $F_{144,7}(0.05) = 3.26$.

~~The~~ Since $3.18 > 2.16$ there is enough evidence to reject the null.

Weight island A	Weight island B
6.7	4.2
7.2	4.3
8.6	5.3
9.4	6.2
10.3	6.8
11.4	12

$$T = 28$$

(7) A group of ornithologists want to investigate whether location had an influence on the weight of birds. They collected the data of 12 birds from same species but different islands.

(a) State the null and alternative that the ornithologists want to investigate. [1]

H_0 : The distribution of bird weight on island A and B are same

H_1 : The distribution of birds weights on island A and B are a shift of each other.

(b) State the test you would do and why? [1]

Wilcoxon sum rank test

(c) Do the test at the 5% level. [3]

looking up Wilcoxon sum rank tables (Table 5).

Non rejection region [26, 52]

since $T = 28$ lies in this region not enough evidence to reject the null.

Island	average	sample variance
A	8.9	3.25
B	6.4	8.40

] - pooled 5.83

- (8) Three instructors ~~are~~ gave STAT 651 exams last week. The number of As and Bs are recorded below. Students want to know whether the instructor influences the grade a student obtains.

	Instructor 1	Instructor 2	Instructor 3	
A	20	40	20	80
B	30	30	30	90
	50	70	50	170

- (a) State the null and alternative hypothesis to investigate. [1]

H_0 : There is no dependence between instructor and grade

H_A : There is a dependence between instructor and grade.

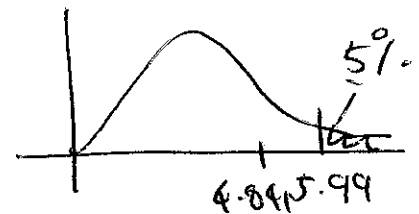
- (b) State the test you would do and why? [1]

χ^2 test for independence

- (c) Do the test at the 5% level, and report your findings. [3]

Expected Value

	1	2	3
A	$\frac{50 \times 80}{170} = 23.53$	$\frac{70 \times 80}{170} = 32.94$	$\frac{50 \times 80}{170} = 23.5$
B	$\frac{50 \times 90}{170} = 26.47$	$\frac{70 \times 90}{170} = 37.06$	$\frac{50 \times 90}{170} = 26.5$



Under the null
 χ^2 - 2 degrees
of freedom

$$\frac{(20 - 23.53)^2}{23.53} + \frac{(40 - 32.94)^2}{32.94} + \frac{(23.5 - 20)^2}{23.5} + \frac{(30 - 26.47)^2}{26.47} + \frac{(30 - 37.06)^2}{37.06} + \frac{(30 - 26.5)^2}{26.5} =$$

$$0.53 + 1.513 + 0.521 + 0.47 + 1.345 + 0.462 = 4.841$$

5.99 > 4.84
not enough
evidence to
reject null