Solutions

STAT 651 Midterm Test (60 minutes) - 10:20pm - 11:20pm October, 2019

NAME: Total number of Marks: /25

There are **5** questions in this paper, do not be deterred, they are all straightforward. Read each question carefully. There are questions on both sides of the page. The number of marks for each question are given in brackets. Be smart about how you answer. If you can't answer one question move on the to next and return to the questions you could not do after answering all the other questions!

Rubric: This exam is a closed book exam, but you can use a 4-sided cheat sheet, normal and t-tables and a calculator.

Write your solutions in the question paper.

GOOD LUCK!

Total marks

$7 + 10 + 4 + 6 = 27$ in total (but grade out of 25).

(1) The table below gives the numbers of people who survived the Titantic (a steamliner that was crossing from UK to the USA and hit an iceberg, 107 year ago) together with the class of the passengers.

	Did not Survive	Survived	Total
Crew	673	212	885
First Class	122	203	325
Second Class	167	118	285
Third Class	528	178	706
Total	1490	711	2201

(a) Calculate the proportion of passengers in first, second and third class who survived. [1.5]

let S = Survival

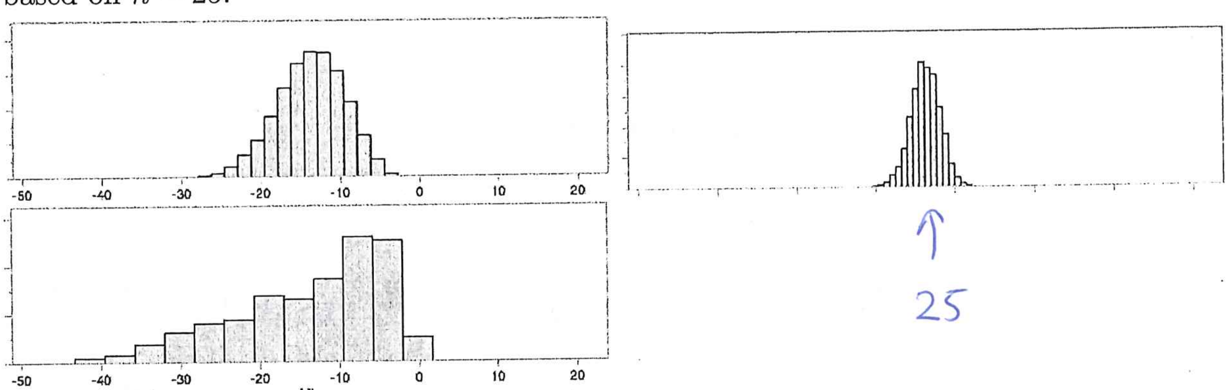
$P[S \mid FC] = \frac{203}{325}$, $P[S \mid SC] = \frac{118}{285}$, $P[S \mid TC] = \frac{178}{706}$

(b) Is there a relationship between survival rates and the class the passenger was in? Explain your answer. [1.5]

The survival rates are very different between the classes, suggesting there is a relationship between the class a passenger was in and their survival.

(2) Below are three plots. One is the distribution of the original data, one is the distribution of the sample mean based on $n = 5$ and one is the distribution of the sample mean based on $n = 25$.

5 →



↑ 25

↑ original

(a) Identify which plot should be associated with which sample size. [2]

See previous page

(b) Explain the two effects that are observed in the distribution of the sample mean as the sample size increase. [2]

(i) As sample size grows the ~~sam~~ distribution of sample mean gets closer to ~~cn~~ normality

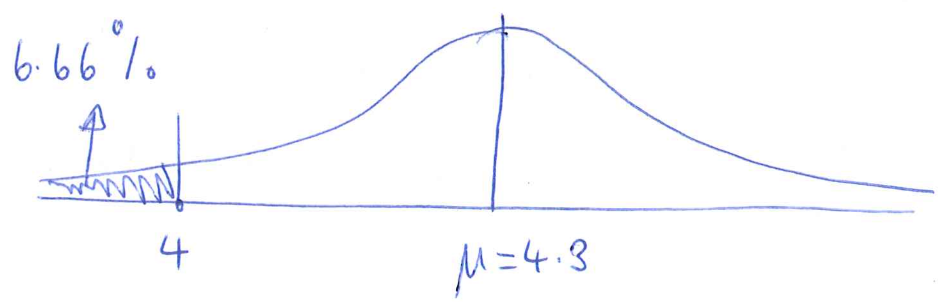
(ii) As the sample size grows the standard error (narrowness of distribution) get smaller.

(3) If the mean blood potassium level (denoted μ) of a patient is **below** 4.0mEq/dl the patient has Hypokalemia.

A few blood samples from a patient are drawn. The level of potassium in the blood follows a normal distribution with mean μ and standard deviation $\sigma = 0.2$. If the potassium level in the **sample mean,** $\bar{X}$, is **less** than 4.0, a diagnoses of low potassium is made ($\mu < 4.0$).

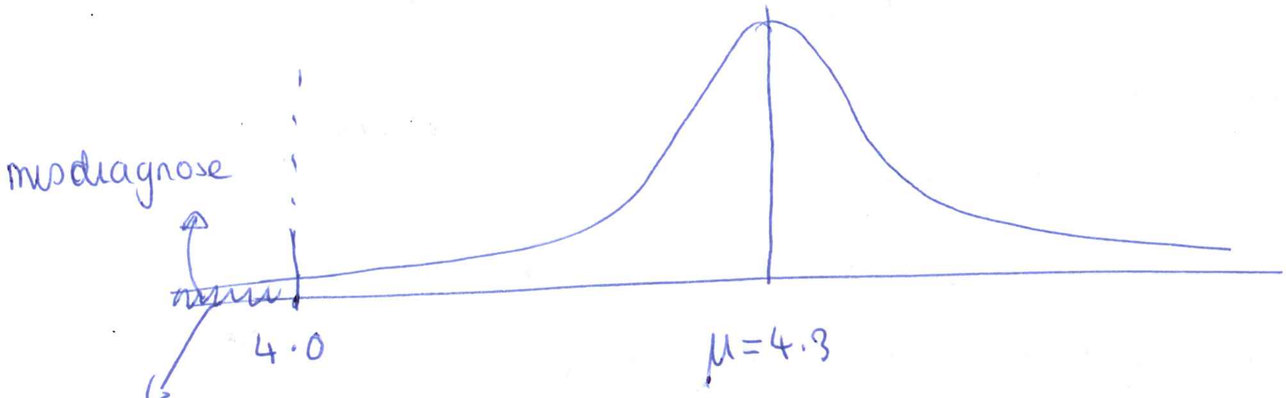
A patient goes to a clinic. Their potassium levels varies according to a Normal($\mu = 4.3, \sigma = 0.2$) distribution (this patient, by definition, does not have low potassium since $\mu = 4.3 > 4.0$).

(a) Calculate the chance the patient (with $\mu = 4.3$) is misdiagnosed with hypokalemia (low potassium). [2]



$$s.e = \frac{0.2}{\sqrt{n}} = 0.2 \qquad z = \frac{4-4.3}{0.2} = -1.5$$

(b) **5** blood samples are drawn from the healthy patient described above. If the sample mean is less than 4.0 they are (mistakenly) diagnosed with low potassium. Calculate the chance of the patient being misdiagnosed (based on a sample mean of size five). [2]



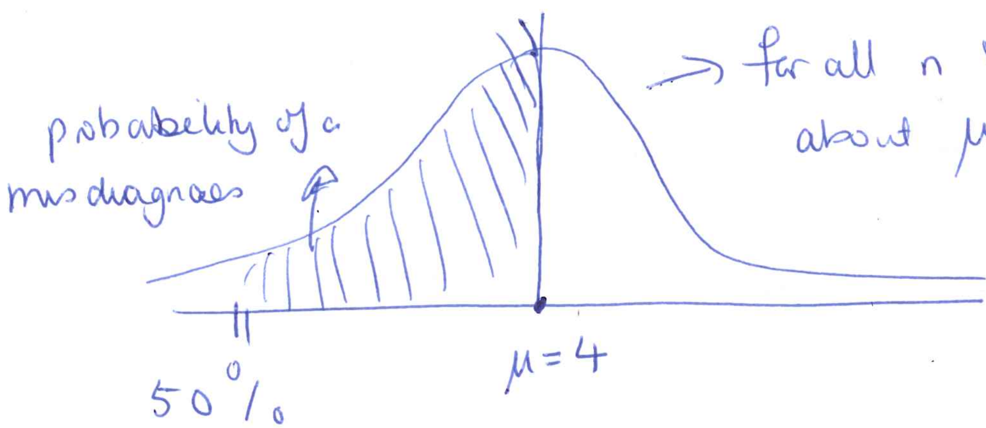
prob = 0.04%

$$s.e = \frac{0.2}{\sqrt{5}} \simeq 0.09 \qquad z = \frac{4.0-4.3}{0.09} = -3.35$$

(c) Based on the above criterion, where a diagnoses is made when the sample mean is less than 4.0. What happens to the probability of misdiagnosing a healthy patients, whose mean level is $\mu = 4.3$, as the sample size grows? Briefly explain your answer. [1]

As the sample size grows the standard error gets smaller. This leads to a ~~ms~~ narrower distribution and greater accuracy. Thus the probability of a misdiagnoses when $\mu = 4.3$ decreases.

(d) What is the probability of a misdiagnoses when a patient has mean potassium level $\mu = 4.0$ (by definition, this patient does not have low potassium) for **any** sample size. Explain your answer with a plot. [2]

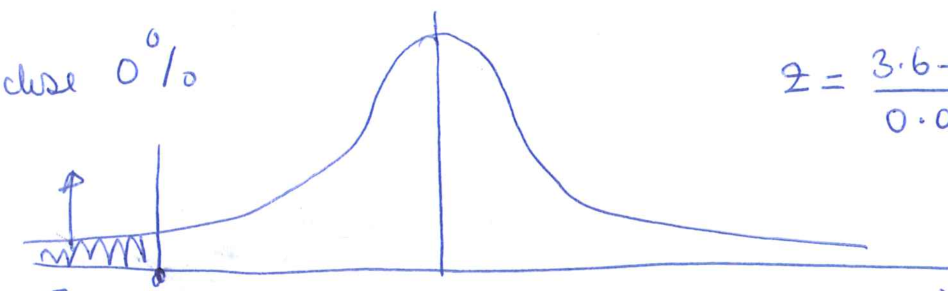


→ for all n the plot is symmetric about $\mu = 4.0$

(e) We now consider a **new** patient and and new diagnostic method.
We test the hypothesis $H_0 : \mu \geq 4.0$ against $H_A : \mu < 4.0$. **9** bloods samples are drawn from a patient and the sample mean $\bar{X} = 3.6$ is evaluated ($\sigma = 0.2$). The test is done at the 5% level. What is the conclusion of the test? [2]

very close 0%

$z = \dfrac{3.6-4}{0.066} = -4.47.$



$\bar{X} = 3.6$ $\mu = 4$ (always under the null)

$s.e = \dfrac{0.2}{\sqrt{9}} \cong 0.066...$ p-value $\simeq$ 0%. Reject null.

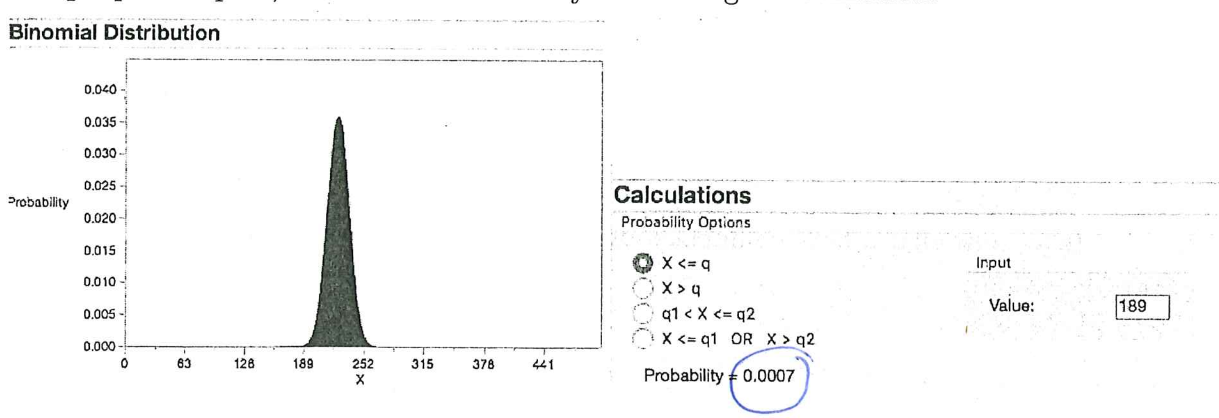
(f) Suppose we test the hypothesis $H_0 : \mu \geq 4.0$ against $H_A : \mu < 4.0$ at the 5% level. 200 healthy patients are tested. Roughly, how many patients will be misdiagnosed with low potassium? [1]

evidence of low potassium

5% of 200 $\simeq$ 10 healthy patients.
(assuming these patients have mean $\mu = 4.0$)

(4) In the 1990s the proportion of young people in the USA (those in the 20-30 age group) who had a religious affiliation was 45%.

A recent report claimed that the proportion of young people who have a religious affiliation has **decreased**. To check if there is any evidence of this, a sample of 500 young people were random sampled and asked if they had a political affiliation. Of the 500 people sampled, 189 claimed that they had a religious affiliation.



(a) What is the hypothesis of interest? $= 0.07\%$ [2]

$H_0: p \geq 0.45$ $H_A: p < 0.45$

4

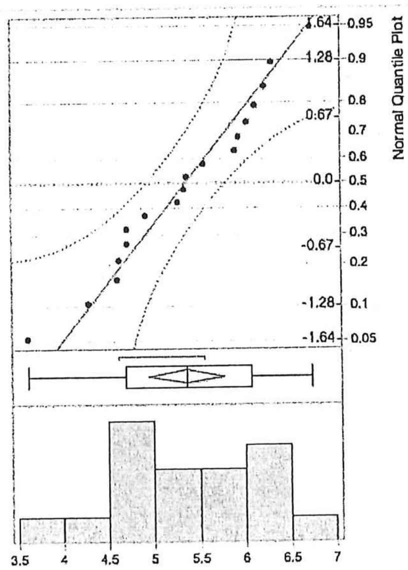
(b) The distribution calculator gives the distribution of $\text{Bin}(n = 500, p = 0.45)$. Using this, is there any evidence of a decrease in the proportion (conduct the test at the 5% level)? Give a reason for your answer. [2]

From the ghA the ~~p~~ probabili of ~~getting~~ 189 or less, claiming on affiliation when $p = 0.45$ is 0.07%. The p-value = 0.07% < 5%. Reject null. Evidence of a reduction

(5) It is well documented that the amount of CO_2 entering our atmosphere has substantially increased over the past 50 years. Scientists are studying the impact this has on the size of Arctic ice (which varies each year).

The size of Arctic ice is measured each year (in $10^6 m^2$). The mean size of the ice prior to 1999 was known to be **7** ($10^6 m^2$). Scientists hypothesis that the size of Arctic ice has **decreased** over the past few years. They measure the size of Arctic ice yearly between 1999-2016. The data is summarized below. Let μ denote the population mean. Conduct all tests at the 5% level.

Data can be found at `https://www.epa.gov/climate-indicators/`



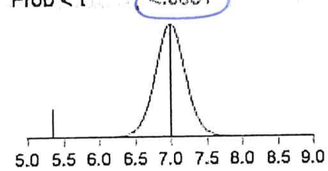
Summary Statistics

Mean	5.355
Std Dev	0.8259344
Std Err Mean	0.1946746
Upper 95% Mean	5.7657275
Lower 95% Mean	4.9442725
N	18

Test Mean

Hypothesized Value	7
Actual Estimate	5.355
DF	17
Std Dev	0.82593

	t Test
Test Statistic	-8.4500
Prob > |t|	<.0001*
Prob > t	1.0000
Prob < t	<.0001*



(a) Construct a 98% confidence interval for the mean size of ice in the Arctic. [2]

$[\,5.355 \pm 2.56 \times 0.194\,]$

$t_{17}(1\%) = 2.56$

(b) Based on the QQplox, can we reliably say we have 98% confidence the mean is in the interval? [1]

The QQplot shows that the data does not deviate too much from nomality. Thus the sample mean must be close to nomal. Yes it appears releable.

(c) The mean size of the ice prior to 1999 was known to be **7** ($10^6 m^2$). Scientists hypothesis that the size of Arctic ice has **decreased** over the past few years. What is the hypothesis of interest and is there any evidence of a **drop below 7**? [2]

$H_0: \mu \geq 7$ $H_A: \mu < 7$.

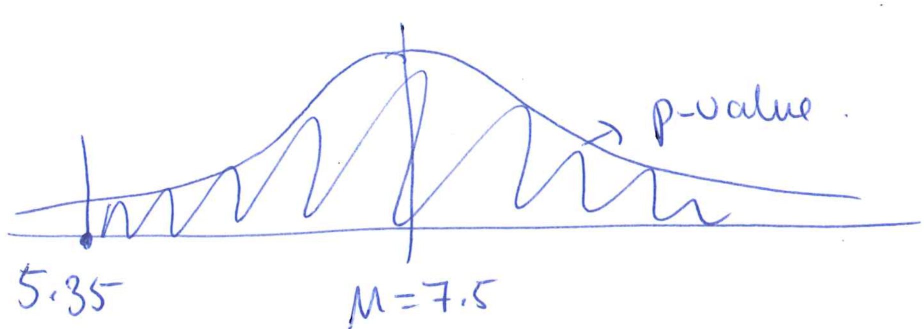
p-value $< 0.01\%$ Reject null at 5%.

Evidence the size of the ice shelf has dropped!

(considerably).

(d) A certain organisation believes that the size of ice has increased over the past few years. They claim that the ice is now **over** 7.5 ($10^6 m^2$)? What is the hypothesis of interest and is there any evidence of an **increase to over** 7.5? [1]

$H_0: \mu \leq 7.5$ $H_A: \mu > 7.5$



p-value is huge! No evidence of the mean being over 7.5!!