Solutions

STAT 651 Midterm Test (55 minutes) - 3pm - 3:55pm **March, 2010**

NAME: **Total number of Marks: /25**

There are 6 questions in this paper, do not be deterred, they are all straightforward. Read each question carefully. There are questions on both side 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 an open book exam you can use all written materials that you want, normal tables and a calculator.

Write your solutions in the question paper.

GOOD LUCK!

Spare page.

(1) Suppose you can draw from the following four populations and calculate the sample mean. Which population with sample would you choose, and explain why?

(A) The population variance is 1 the sample size is 40. $1/40$

(B) The population variance is 2 the sample size is 80. $1/40$

(C) The population variance is 0.25 the sample size is 20. $1/80$

(D) The population variance is 0.5 the sample size is 60. $1/120$

Choose (D) because the sample mean has the smallest variance.

[2]

(2) A doctor wants to investigate the probability that two siblings (brothers and/or sisters) are over 9.5 pounds. She uses the following information.

The probability a randomly selected newborn baby is over 9.5 pounds is 0.05.

The probability that a new born baby is over 9.5 pounds *given* that its older sibling birth weight is over 9.5 pounds is 0.2.

(a) Calculate the probability that both siblings are over 9.5 pounds. [2]

$$P\{\text{both siblings} > 9.5\} = P\{\text{Baby } 2 > 9.5 \mid \text{Baby } 1 > 9.5\}\, P\{\text{Baby } 1 > 9.5\}$$
$$= 0.2 \times 0.05 = 0.01$$

(a) Based on the information above do you think the birth weight of siblings are independent of each other? Give a reason for your answer. [1]

They are dependent because

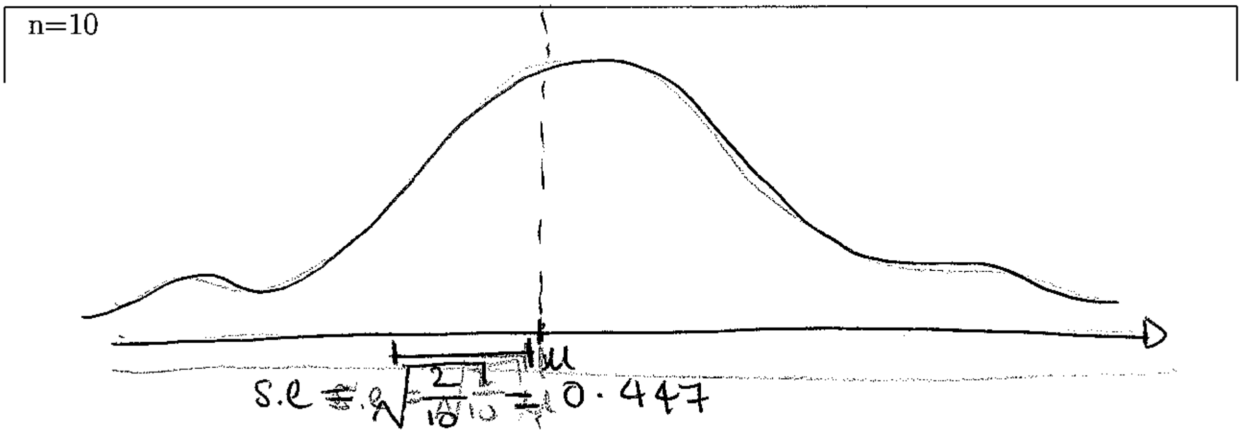
$P\{\text{Baby } 2 > 9.5 \mid \text{Baby } 1 > 9.5\} \neq P\{\text{Baby } 2 > 9.5\}$.

(3) A draw 200 random samples, each sample is of size 40 ($n = 40$). For each sample, I construct a 90% confidence interval for the mean. On average how many of the confidence intervals will *not* contain the population mean. [1]

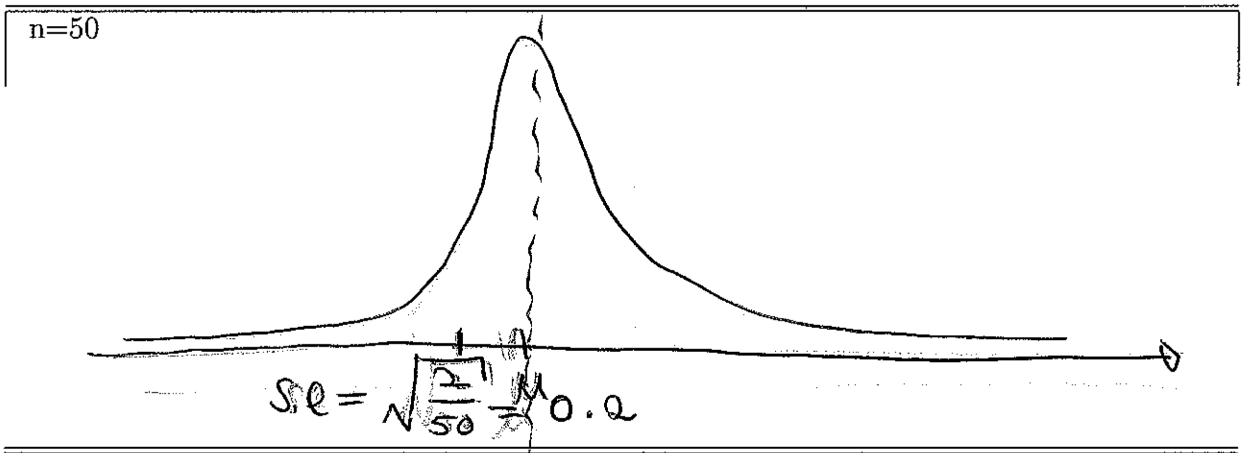
$0.1 \times 200 = 20$

(4) Consider a population which has mean μ and variance 2. Suppose we draw random samples of size $n = 10$, $n = 50$ and $n = 200$, and for each sample we calculate the sample mean $\bar{X}$. In the boxes below draw the distribution of $\bar{X}$ for $n = 10$, $n = 20$, $n = 50$ and $n = 200$. Indicate the mean and standard error of $\bar{X}$ on the plot. [3]

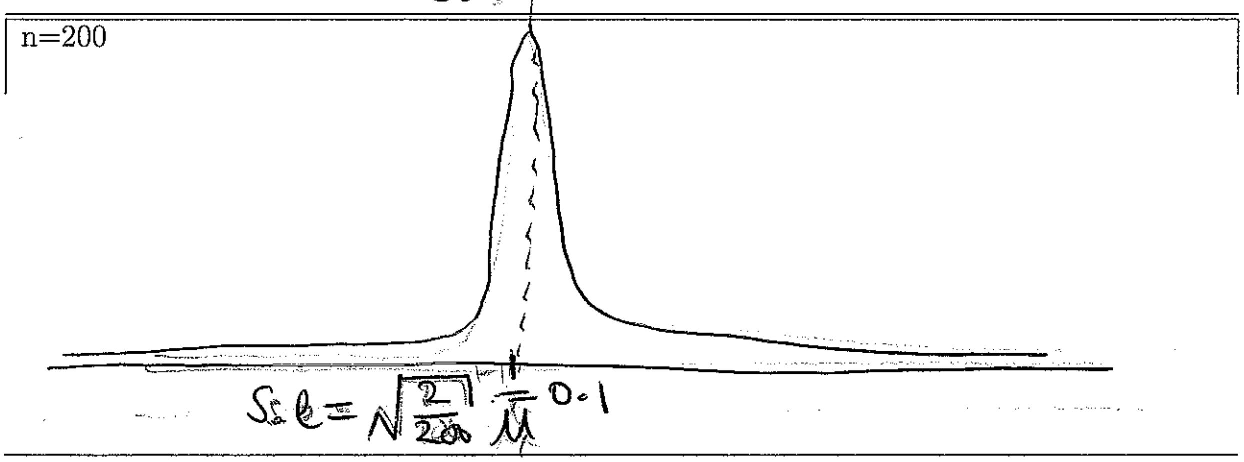
n=10



n=50

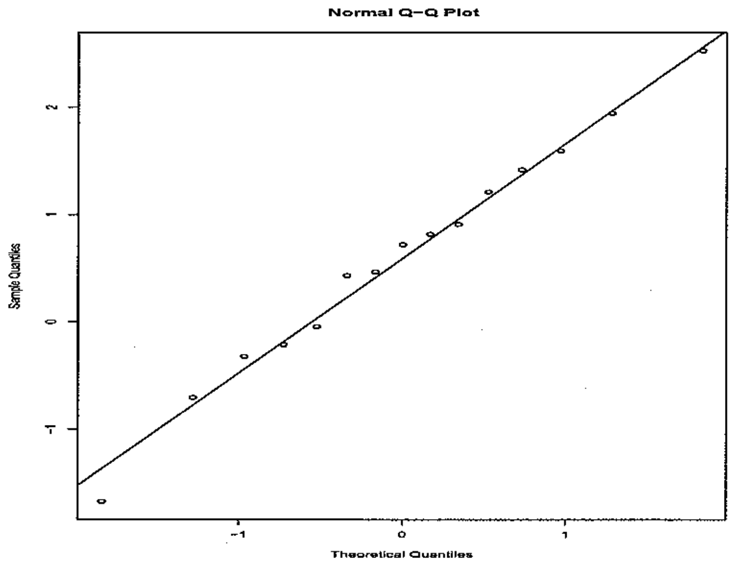


n=200



- Observe the curve becomes more 'bell shaped' as the sample size grows.
- The standard error gets smaller as sample size grows.

(5) I draw of random sample of 15. The QQplot is given below. Suppose that the sample mean is $\bar{X} = 0.606$ and the population variance is $\sigma^2 = 1$.



(a) Construct a 95% for the sample mean. [2]

$$\left[0.606 \pm 1.96 \times \sqrt{\frac{1}{15}}\right] = $$

$$= [0.01, 1.11]$$

(b) Based on the QQplot comment on whether the 95% CI for the sample mean is reliable. Give a reason for your answer. [2]

The sample size $n = 15$ is small, hence for the CI to be reliable the distribution of the population should be close to normal.

Above is a QQplot of the observations, the points tend to be on the 45° line, suggesting that the observations have come from a normal distribution. Based on this, the CI ~~is~~ seems to be reliable.

(6) Anthropologists are trying to classify what they believe are ancient handaxes (ancient knives) made by early hominid (ancient humans) found in the Olduvai Gorge, in East Africa.

It is known that axes made by Ancient Human species A has mean length 2 inches.

But Ancient Human species B had learnt to make *longer* handaxes. These hand axes had mean length *longer* than 2 inches.

It is known that the standard deviation of all handaxes from this age is about $\sigma = 2$ inches.

(i) Suppose we want to construct a 99% CI for the mean handaxe length from the Olduvai Gorge, and we want the 99% CI to have length 0.05inches, how should we choose the sample size? [3]

$$2 \times 2.56 \times \frac{2}{\sqrt{n}} = 0.05$$

hence $n = \left(\frac{2 \times 2.56 \times 2}{0.05}\right)^2 \simeq 41943$

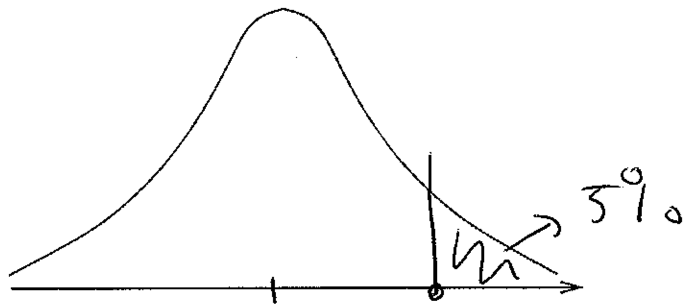
This is a huge sample size. Such a large sample size is required for this level of accuracy.

(ii) Anthropologists are collecting ancient handaxes samples from the Olduvai Gorge. They want to investigate whether the handaxes found in this particular site were made by Human species B.

State the null and alternative hypothesis that the Anthropologists want to investigate. [1]

$H_0: \mu \leq 2$ $\quad\quad$ $H_A: \mu > 2$

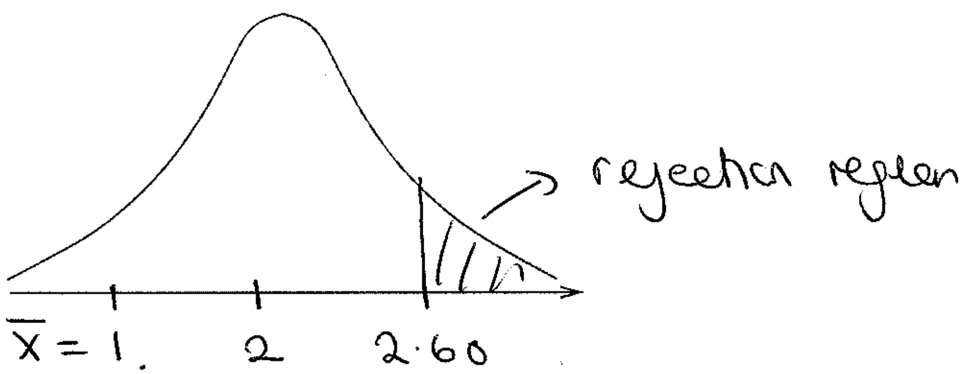
(iii) Suppose that the anthropologists collect a random sample of 30 handaxes. Based on this sample, the sample mean is $\bar{X} = 3$ inches. Test the hypthesis stated in (ii) (do the test at the 5% level). [3]



$$2 + 1.64 \times \frac{2}{\sqrt{30}} = 2.60$$

Since $3 > 2.60$, $\bar{X} = 3$ is sufficiently far from 2 for us to reject the null.

(iv) Suppose that another group of anthropologists collect a random sample of 25 handaxes. Based on this sample, the sample mean is $\bar{X} = 1$ inches. Test the hypthesis stated in (ii) (do the test at the 5% level). [2]



Hence there is not enough evidence to reject the null.

Exercise: calculate the p-value for part (iii).

Please turn over page.

(v) What is the chance of rejecting the null if the sample size is $n = 20$ and the alternative is that the mean axe length is greater than 4 inches. [3]

The above question should read,

'What is the chance of rejecting the null that $\mu \leq 2$, when the alternative that the mean axe length is greater than 4 inches is true'.

Notice there isn't any data in this question, all is given is

$n = 20$, $\sigma = 2$, $H_0: \mu \leq 2$ against $H_A: \mu \geq 4$

and the test is done at the 5% level.

$$P\{\text{reject null} \mid \text{when } \mu \geq 4\} = \text{power}$$

$$= 1 - P\left\{Z \leq 1.64 - \frac{|2-4|}{2/\sqrt{20}}\right\}$$

One-side test at 5% level

$$= 1 - P\{Z \leq -2.83\} = 0.998$$

Power = 99.8% (really large - this is because null is far from alternative).