NORMAL DISTRIBUTION HYPOTHESIS TESTING

&

RELATED TOPICS

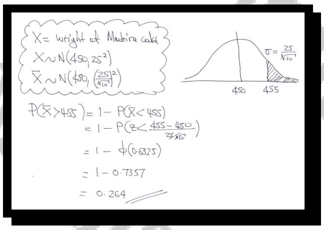
THE SAMPLING DISTRIBUTION OF THE MEAN

Question 1 (**)

The weights of Madeira cakes are Normally distributed with a mean of 450 grams and a standard deviation of 25.

Find the probability that the sample mean of 10 randomly selected Madeira cakes will exceed 455 grams.

$\boxed{0.264}$

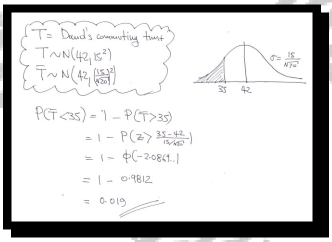


Question 2 (**)

David's commuting times are Normally distributed with a mean of 42 minutes and a standard deviation of 15.

Find the probability that 20 of David's commuting times will have a sample mean of less than 35 minutes.

$\boxed{0.019}$

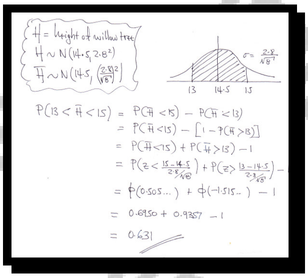


Question 3 (**)

The heights of willow trees are Normally distributed with a mean of 14.5 metres and a standard deviation of 2.8 .

Find the probability that the mean height of a random sample 8 willow trees will be between 13 and 15 metres.

$\boxed{0.631}$



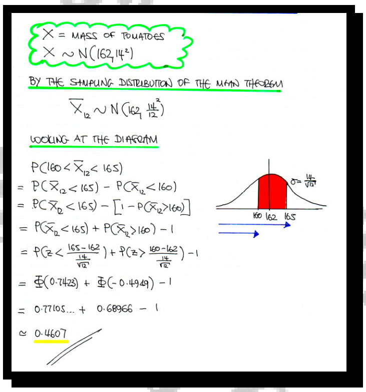
Question 4 (**)

The masses of a particular variety of tomatoes, in grams, are assumed to be Normally distributed with mean 162 and standard deviation 14.

A random sample of 12 tomatoes of this variety is selected.

Determine the probability that the mean mass of this sample will be between 160 and 165 grams.

$\boxed{}$, $\boxed{0.4607}$

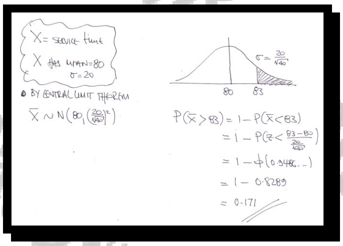


Question 5 (**)

Car servicing times have a mean of 80 minutes and a standard deviation of 20.

Find the probability that the mean servicing times of 40 cars will exceed 83 minutes.

$\boxed{0.171}$

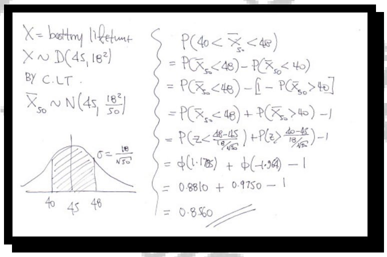


Question 6 (**)

The lifetimes of a certain brand of battery have a mean of 45 hours and a standard deviation of 18.

Find the probability that the mean lifetime of random sample of such 50 batteries will be between 40 and 48 hours.

$\boxed{0.856}$

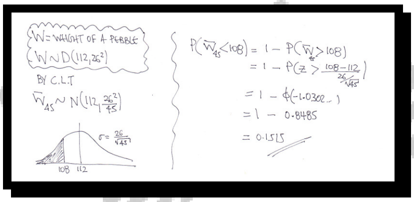


Question 7 (**)

The weights of pebbles have a mean of 112 grams and a standard deviation of 26.

Find the probability that the mean weight of a random sample of 45 pebbles will be less than 108 grams.

$\boxed{0.1515}$

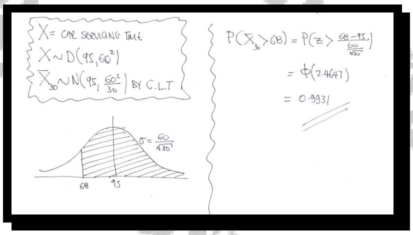


Question 8 (**)

The times taken to service a family size car have a mean of 95 minutes and a standard deviation of 60.

Find the probability that the mean servicing time of 30 family size cars will exceed 68 minutes.

$\boxed{0.9931}$



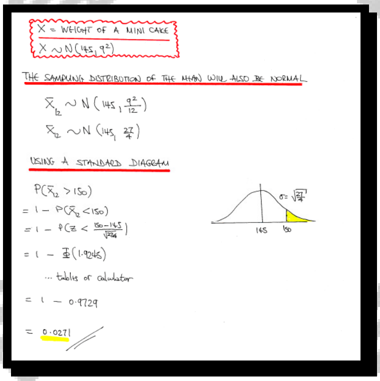
Question 9 (**+)

Mini-cakes having a mean weight of 145 grams and standard deviation 9 grams are packed in boxes of 12.

A box of these mini-cakes is selected at random.

If the weights of these mini-cakes are Normally distributed, determine the probability that the mean weight of the mini-cakes in the box will be greater than 150 grams.

$\boxed{}$, $\boxed{0.0271}$



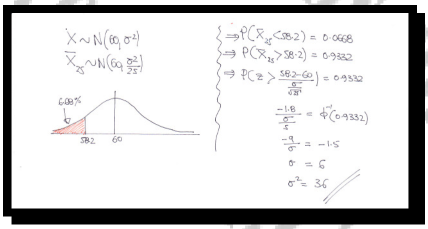
Question 10 (**+)

The mean of a normal random variable X is 60.

The mean of 25 random observations of X is denoted by $\bar{X}$.

Given that $P(\bar{X} < 58.2) = 0.0668$, determine the variance of X.

$\boxed{\sigma^2 = 36}$



Question 11 (**+)

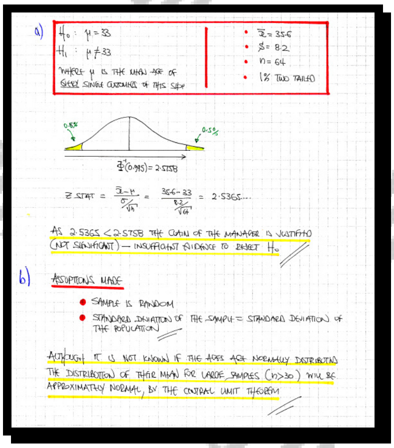
The manager of a shop claims that the mean age of his customers is 33 years.

A sample of 64 customers is taken and this sample produces a mean of 35.6 years and a standard deviation of 8.2 years.

a) Stating your hypotheses and using a 1% level of significance, test whether or not the manager's claim is supported by the data.

b) State two assumptions made in carrying this test, further explaining why this test is still valid even if the ages of the customers are not Normally distributed.

$\boxed{\quad}$, $\boxed{\text{not significant, } 2.5365 < 2.5758}$



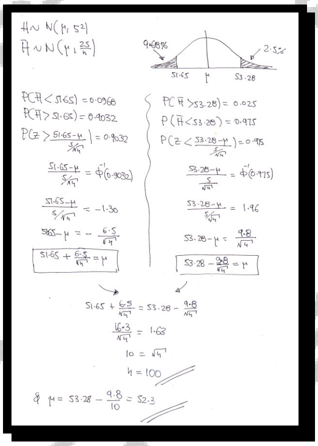
Question 12 **(***)**

The random variable H is Normally distributed with mean μ and variance 25.

The mean of a random sample of n observations of H is denoted by $\bar{H}$.

Given that $\mathrm{P}\left(\bar{H} > 53.28\right) = 0.0250$ and $\mathrm{P}\left(\bar{H} < 51.65\right) = 0.0968$, determine the value of μ and the value of n.

$\boxed{n = 100}$, $\boxed{\mu = 52.3}$



$H \sim N(\mu, 5^2)$

$\bar{H} \sim N\left(\mu, \frac{25}{n}\right)$

$\mathrm{P}(\bar{H} < 51.65) = 0.0968$

$\mathrm{P}(\bar{H} > 51.65) = 0.9032$

$\mathrm{P}\left(Z > \frac{51.65-\mu}{5/\sqrt{n}}\right) = 0.9032$

$\frac{51.65-\mu}{5/\sqrt{n}} = \Phi^{-1}(0.9032)$

$\frac{51.65-\mu}{5/\sqrt{n}} = -1.30$

$51.65-\mu = -\frac{6.5}{\sqrt{n}}$

$\boxed{51.65 + \frac{6.5}{\sqrt{n}} = \mu}$

$\mathrm{P}(\bar{H} > 53.28) = 0.025$

$\mathrm{P}(\bar{H} < 53.28) = 0.975$

$\mathrm{P}\left(Z < \frac{53.28-\mu}{5/\sqrt{n}}\right) = 0.975$

$\frac{53.28-\mu}{5/\sqrt{n}} = \Phi^{-1}(0.975)$

$\frac{53.28-\mu}{5/\sqrt{n}} = 1.96$

$53.28-\mu = \frac{9.8}{\sqrt{n}}$

$\boxed{53.28 - \frac{9.8}{\sqrt{n}} = \mu}$

$51.65 + \frac{6.5}{\sqrt{n}} = 53.28 - \frac{9.8}{\sqrt{n}}$

$\frac{16.3}{\sqrt{n}} = 1.63$

$10 = \sqrt{n}$

$n = 100$

& $\mu = 53.28 - \frac{9.8}{10} = 52.3$

Question 13 (**+)

The weights, in grams, of a random sample of ten eggs, produced by the chickens in a farm, are shown below.

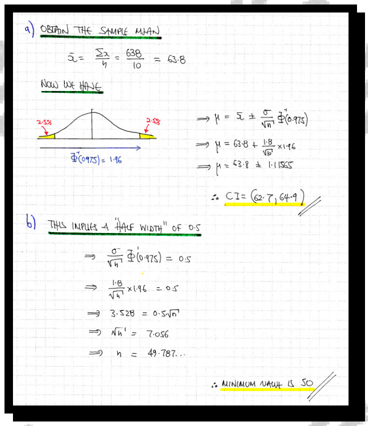
$$61.0,\ 64.6,\ 62.8,\ 67.2,\ 63.1,\ 64.8,\ 66.0,\ 63.5,\ 63.2,\ 61.8.$$

You may assume that this is a random sample coming from a population which is follows a Normal distribution with standard deviation 1.8.

a) Determine a symmetrical 95% confidence interval for the mean weight of all the eggs produced in this farm.

b) If the symmetrical 95% confidence interval obtained in part **(a)** is to have a width of at most 1 gram, determine the minimum sample size needed.

$\boxed{}$, $\boxed{(62.7, 64.9)}$, $\boxed{n_{\min} = 50}$



Question 14 (***)

A standard fair dice with faces numbered 1, 2, 3, 4, 5, 6 is rolled 80 times.

Determine the approximate probability that the mean score of the 80 scores obtained will exceed 3.8.

$\boxed{\approx 5.8\%}$

- A FAIR DICE FOLLOWS A DISCRETE UNIFORM DISTRIBUTION WITH $n=6$

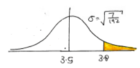
$$E(X) = \frac{n+1}{2} = \frac{6+1}{2} = 3.5$$
$$\mathrm{Var}(X) = \frac{n^2-1}{12} = \frac{35}{12}$$

- BY THE CENTRAL LIMIT THEOREM, THE MEAN OF 80 OBSERVATIONS, WILL HAVE AN APPROXIMATE DISTRIBUTION

$$\bar{X}_{80} \sim N\left(3.5, \frac{35/12}{80}\right)$$

SIMPLIFIES TO $\frac{7}{192}$

- HENCE WE NOW HAVE



$$\Rightarrow P(\bar{X}_{80} > 3.8)$$
$$= 1 - P(\bar{X}_{80} < 3.8)$$
$$= 1 - P\left(z < \frac{3.8 - 3.5}{\sqrt{7/192}}\right)$$
$$= 1 - \Phi(1.5712)$$
$$= 1 - 0.9419$$
$$= 0.0581$$

IE 5.81%

Question 15 (***+)

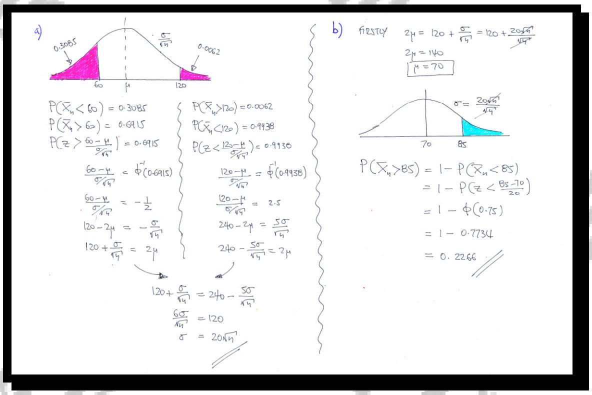
The random variable X is Normally distributed with mean μ and variance σ^2.

The mean of a random sample of n observations of X is denoted by $\bar{X}_n$.

a) Given that $\mathrm{P}\left(\bar{X}_n < 60\right) = 0.3085$ and $\mathrm{P}\left(\bar{X}_n > 120\right) = 0.0062$, express σ in terms of n.

b) Calculate the value of $\mathrm{P}\left(\bar{X}_n > 85\right)$.

$\boxed{\sigma = 20\sqrt{n}}$, $\boxed{\approx 0.2266}$



a)

0.3085, 0.0062, $\frac{\sigma}{\sqrt{n}}$, 60, μ, 120

$\mathrm{P}(\bar{X}_n < 60) = 0.3085$

$\mathrm{P}(\bar{X}_n > 60) = 0.6915$

$\mathrm{P}\left(z > \frac{60-\mu}{\sigma/\sqrt{n}}\right) = 0.6915$

$\frac{60-\mu}{\sigma/\sqrt{n}} = \Phi^{-1}(0.6915)$

$\frac{60-\mu}{\sigma/\sqrt{n}} = -\frac{1}{2}$

$120 - 2\mu = -\frac{\sigma}{\sqrt{n}}$

$120 + \frac{\sigma}{\sqrt{n}} = 2\mu$

$\mathrm{P}(\bar{X}_n > 120) = 0.0062$

$\mathrm{P}(\bar{X}_n < 120) = 0.9938$

$\mathrm{P}\left(z < \frac{120-\mu}{\sigma/\sqrt{n}}\right) = 0.9938$

$\frac{120-\mu}{\sigma/\sqrt{n}} = \Phi^{-1}(0.9938)$

$\frac{120-\mu}{\sigma/\sqrt{n}} = 2.5$

$240 - 2\mu = \frac{5\sigma}{\sqrt{n}}$

$240 - \frac{5\sigma}{\sqrt{n}} = 2\mu$

$120 + \frac{\sigma}{\sqrt{n}} = 240 - \frac{5\sigma}{\sqrt{n}}$

$\frac{6\sigma}{\sqrt{n}} = 120$

$\sigma = 20\sqrt{n}$

b) FIRSTLY $2\mu = 120 + \frac{\sigma}{\sqrt{n}} = 120 + \frac{20\cancel{\sqrt{n}}}{\cancel{\sqrt{n}}}$

$2\mu = 140$

$\boxed{\mu = 70}$

$\sigma = \frac{20\cancel{\sqrt{n}}}{\cancel{\sqrt{n}}}$, 70, 85

$\mathrm{P}(\bar{X}_n > 85) = 1 - \mathrm{P}(\bar{X}_n < 85)$

$= 1 - \mathrm{P}\left(z < \frac{85-70}{20}\right)$

$= 1 - \Phi(0.75)$

$= 1 - 0.7734$

$= 0.2266$

Question 16 (****)

One thousand pieces of positive numerical data, all of which contain a decimal part, were rounded to the nearest integer.

a) Determine the probability that the sum of the rounded data exceeds the total of the original data by at least 10.

Next n pieces of positive numerical data, all of which contain a decimal part, were rounded to the nearest integer.

b) Find the greatest value of n so that the probability that the sum of the rounded data and the total of the original data differs by 10, is greater than 0.95.

$\boxed{\approx 0.137}$, $\boxed{n = 312}$

a) X = error in rounding

$X \sim U(-0.5, 0.5)$ ← THIS DISTRIBUTION HAS MEAN $\frac{a+b}{2} = 0$

VARIANCE $\frac{(b-a)^2}{12} = \frac{1}{12}$

BY THE CENTRAL LIMIT THEOREM AS n IS LARGE

$$\bar{X}_{1000} \sim N\left(0, \frac{1/12}{1000}\right)$$

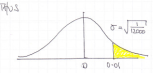
$$\bar{X}_{1000} \sim N\left(0, \frac{1}{12000}\right)$$

IF THE ERROR OF 1000 OBSERVATIONS ADDED IS GREATER THAN $+10$,

THEN $\bar{X}_{1000} > \frac{10}{1000}$ ← TOTAL ERROR ← FROM 1000 OBSERVATIONS

$$\bar{X}_{1000} > 0.01$$

THUS



$$P(\bar{X}_{1000} > 0.01) = 1 - P(\bar{X}_{1000} < 0.01)$$

$$= 1 - P\left(z < \frac{0.01 - 0}{\sqrt{\frac{1}{12000}}}\right)$$

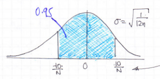
$$= 1 - \Phi(1.09544\ldots)$$

$$= 1 - 0.86334\ldots$$

$$= 0.137$$

b) NOW n IS UNKNOWN (ERROR 10 OR -10)

$\bar{X}_n \sim N\left(0, \frac{1/12}{n}\right)$ AS BEFORE



IF THE ERROR OF n OBSERVATIONS IS TO BE ± 10

$$P\left(|\bar{X}_n| < \frac{10}{n}\right) > 0.95$$

$$P\left(z < \frac{\frac{10}{n} - 0}{\sqrt{1/12n}}\right) > 0.975$$

INVERTING

$$\frac{\frac{10}{n}}{\sqrt{1/12n}} > \Phi^{-1}(0.975) \implies \frac{10}{n}\sqrt{12n} > 1.96$$

$$\implies \frac{100}{n^2}(12n) > 1.96^2$$

$$\implies \frac{1200}{n} > 1.96^2$$

$$\implies \frac{n}{1200} < \frac{1}{1.96^2}$$

$$\implies n < \frac{1200}{1.96^2}$$

$$\implies n < 312.3698\ldots$$

$$\implies n = 312$$

Question 17 (*****)

A six sided dice is labelled 1, 2, 3, 4, 5 and 6.

It is required to check if the dice, when rolled, is fair in obtaining the number 6.

It is decided that **seventy** different people are to roll this dice, 36 times each, and individually record the number of sixes obtained by each person.

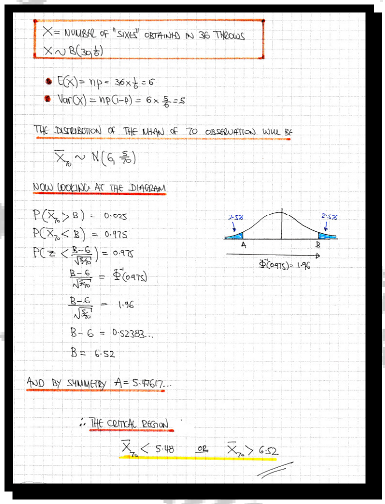
Let $\bar{X}$ be **the mean number of sixes** obtained by these seventy people and the hypotheses

$$\text{H}_0 : p = \frac{1}{6}, \quad \text{H}_1 : p \neq \frac{1}{6},$$

where p is the probability of obtaining a six.

Find, in terms of $\bar{X}$, the critical region for this test, at 5% level of significance.

$\boxed{\bar{X}_{70} < 5.48 \ \cup \ \bar{X}_{70} > 6.52}$



X = NUMBER OF "SIXES" OBTAINED IN 36 THROWS
$X \sim B(36, \frac{1}{6})$

- $E(X) = np = 36 \times \frac{1}{6} = 6$
- $Var(X) = np(1-p) = 6 \times \frac{5}{6} = 5$

THE DISTRIBUTION OF THE MEAN OF 70 OBSERVATION WILL BE

$\bar{X}_{70} \sim N(6, \frac{5}{70})$

NOW LOOKING AT THE DIAGRAM

$P(\bar{X}_{70} > B) = 0.025$
$P(\bar{X}_{70} < B) = 0.975$
$P(z < \frac{B-6}{\sqrt{5/70}}) = 0.975$
$\frac{B-6}{\sqrt{5/70}} = \Phi^{-1}(0.975)$
$\frac{B-6}{\sqrt{5/70}} = 1.96$
$B - 6 = 0.52383...$
$B = 6.52$

AND BY SYMMETRY $A = 5.47617...$

$\therefore$ THE CRITICAL REGION

$\bar{X}_{70} < 5.48$ OR $\bar{X}_{70} > 6.52$

CONFIDENCE INTERVALS

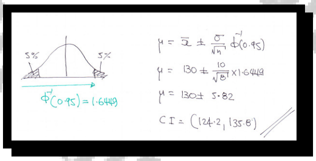
Question 1 (**)

The weights of bags of popcorn are Normally distributed with unknown mean μ and standard deviation of 10 grams.

The sample mean $\bar{x}$ of 8 such bags was found to be 130 grams.

Find a 90% confidence interval for the mean weight of a bag of popcorn.

$\boxed{(124.2,135.8)}$



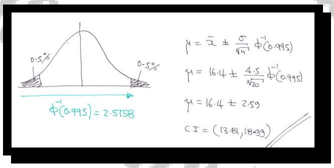
Question 2 (**)

The heights of palm trees are Normally distributed with unknown mean μ and standard deviation of 4.5 metres.

The sample mean $\bar{x}$ of 20 palm trees was measured at 16.40 metres.

Find a 99% confidence interval for the mean height of a palm tree.

$\boxed{(13.81,18.99)}$



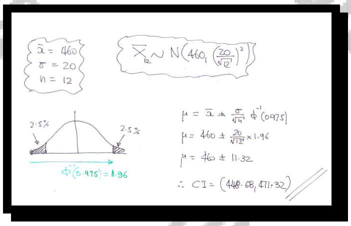
Question 3 (**)

The weights of walnut cakes are Normally distributed with unknown mean μ and standard deviation of 20 grams.

The sample mean $\bar{x}$ of 12 walnut cakes was 460 grams.

Find a 95% confidence interval for the mean weight of a walnut cake.

$\boxed{(448.68, 471.32)}$



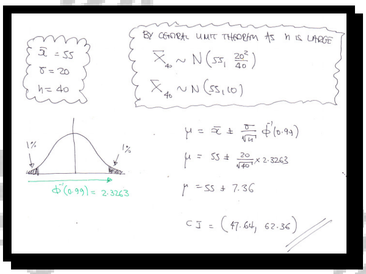
Question 4 (**)

Mark's commuting times to London have unknown mean μ and standard deviation of 20 minutes.

A sample of 40 journeys produced a sample mean $\bar{x}$ of 55 minutes.

Find an approximate 98% confidence interval for Mark's mean commuting time.

$\boxed{(47.64, 62.36)}$



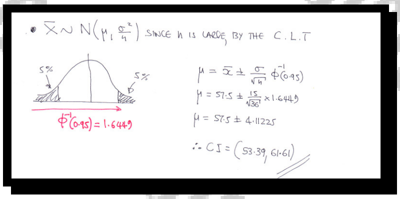
Question 5 (**)

Jeff's javelin throwing distances have a mean μ and standard deviation of 15 metres.

A sample of 36 javelin throws yielded a mean $\bar{x}$ of 57.50 metres.

Find an approximate 90% confidence interval for Jeff's mean throwing distance.

$\boxed{(53.39, 61.61)}$



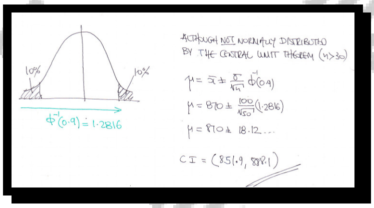
Question 6 (**)

The lifetimes of light bulbs have standard deviation of 100 hours.

50 bulbs were tested and the average lifetime was 870 hours.

Find an approximate 80% confidence interval for the mean lifetime of these bulbs.

$\boxed{(851.9, 888.1)}$

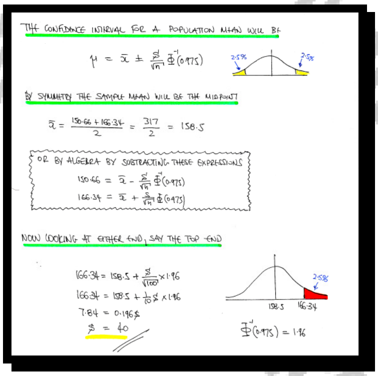


Question 7 (**)

A symmetrical 95% confidence interval for the population mean of a Normal variable, based on a sample of 100 observations, is $(150.66, 166.34)$.

Determine the mean and the standard deviation of the sample.

$\boxed{}$, $\boxed{\bar{x} = 158.5}$, $\boxed{\sigma = 40}$



THE CONFIDENCE INTERVAL FOR A POPULATION MEAN WILL BE

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}}\Phi^{-1}(0.975)$$

BY SYMMETRY THE SAMPLE MEAN WILL BE THE MIDPOINT

$$\bar{x} = \frac{150.66 + 166.34}{2} = \frac{317}{2} = 158.5$$

OR BY ALGEBRA BY SUBTRACTING THESE EXPRESSIONS

$$150.66 = \bar{x} - \frac{s}{\sqrt{n}}\Phi^{-1}(0.975)$$
$$166.34 = \bar{x} + \frac{s}{\sqrt{n}}\Phi^{-1}(0.975)$$

NOW LOOKING AT EITHER END, SAY THE TOP END

$$166.34 = 158.5 + \frac{s}{\sqrt{100}} \times 1.96$$
$$166.34 = 158.5 + \frac{1}{10}s \times 1.96$$
$$7.84 = 0.196s$$
$$s = 40$$

$$\Phi^{-1}(0.975) = 1.96$$

Question 8 (+)**

The time T, taken by a mechanic to service a certain model of car, has mean μ minutes and standard deviation 20 minutes.

A random sample of 100 servicing times showed a mean time of 45 minutes.

a) Explain, with full justification, why T is unlikely to be Normally distributed but the sample mean of the random 100 servicing times $\overline{T}$ can be modelled by a Normal distribution.

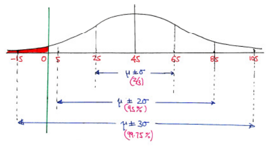
b) Construct a 99% confidence interval for μ and hence comment on the claim that the typical servicing time could be about one hour.

$\boxed{(39.85,\ 50.16)}$

a) T = TIME TO SERVICE CAR

AS $\bar{x} = 45$ BASED ON 100 OBSERVATIONS IT IS REASONABLE TO ASSUME THAT μ WILL BE CLOSE TO THAT FIGURE

SUPPOSE THAT $T \sim N(45, 20^2)$



THEN A SUBSTANTIAL PERCENTAGE OF DATA WOULD BE NEGATIVE, IT WILL NOT BE THE 3 STANDARD DEVIATION THRESHOLD

$\overline{T}_{100} \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$ (APPROXIMATELY BY THE C.L.T)

$\overline{T}_{100} \sim N\left(45, \frac{20^2}{100}\right)$

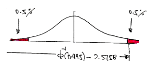
$\overline{T}_{100} \sim N(45, 2^2)$

HENCE $\overline{T}_{100}$ WILL HAVE AN APPROXIMATELY NORMAL DISTRIBUTION WITHOUT $\overline{T}_{100}$ GOING NEGATIVE (σ IS VERY SMALL NOW)

b) $\mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(0.995)$

$\mu = 45 \pm \frac{20}{\sqrt{100}} \times 2.5758$

$\mu = 45 \pm 5.1516$



$\therefore$ CI $= (39.85, 50.16)$

CLAIM NOT JUSTIFIED AS 60 IS "WAY ABOVE" THE UPPER BOUND OF SUCH "HIGH CONFIDENCE" (99%) INTERVAL, BASED ON A LARGE n (100)

Question 9 (**+)

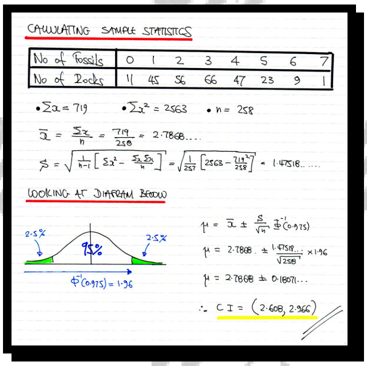
A geologist is investigating the mean number of fossils found in standard size rock samples collected from a certain area.

His data is summarized in the table below.

Number of Fossils	0	1	2	3	4	5	6	7
Number of Rocks	11	45	56	66	47	23	9	1

Find a 95% confidence interval for the mean number of fossils per rock, based on the samples collected from that area.

$\boxed{}$, $\boxed{2.608 < \mu < 2.966}$



CALCULATING SAMPLE STATISTICS

No of Fossils	0	1	2	3	4	5	6	7
No of Rocks	11	45	56	66	47	23	9	1

• $\sum x = 719$ • $\sum x^2 = 2563$ • $n = 258$

$\bar{x} = \frac{\sum x}{n} = \frac{719}{258} = 2.7868\ldots$

$s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right]} = \sqrt{\frac{1}{257}\left[2563 - \frac{719^2}{258}\right]} = 1.47518\ldots$

LOOKING AT DIAGRAM BELOW

$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} \Phi^{-1}(0.975)$

$\mu = 2.7868 \pm \frac{1.47518\ldots}{\sqrt{258}} \times 1.96$

$\mu = 2.7868 \pm 0.18071\ldots$

$\therefore \text{ CI} = (2.608, 2.966)$

Question 10 (***)

The heights of male students in a college are thought to be Normally distributed with mean μ cm and standard deviation 7.5.

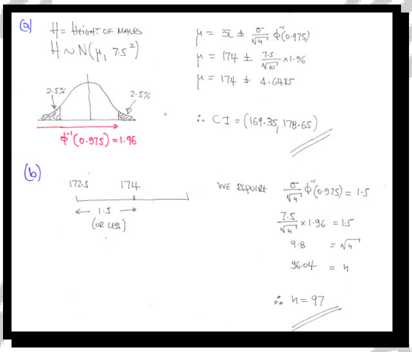
The heights of 10 male students from this college were measured and the sample mean was 174 cm.

a) Find a 95% confidence interval for the mean height of the male students in this particular college.

The value of μ is claimed to be 172.5 by the college.

b) Determine the smallest number of male students whose sample mean height is still 174 cm that must be considered, so that the claimed value of μ is **not** in the 95% confidence interval.

$\boxed{(169.35, 178.65)}$, $\boxed{n = 97}$



Question 11 (***+)

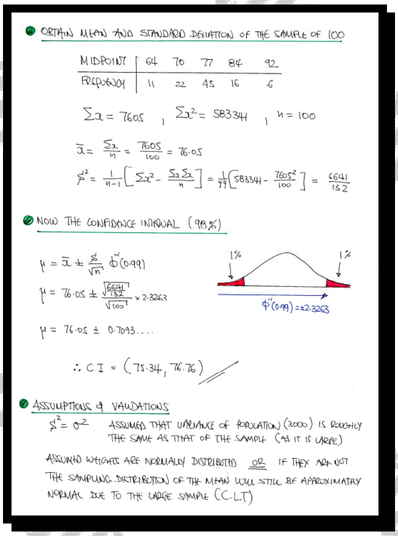
From a regiment of 3000 soldiers a random sample is selected and their weights are recorded to the nearest kilogram.

A summary of these weights is shown in the table below

Weight (nearest Kg)	62 - 66	67 - 73	74 - 80	81 - 87	88 - 96
Number of Soldiers	11	22	45	16	6

Estimate a 98% confidence interval for the mean weight of the soldiers in this regiment, stating clearly any assumptions and validations made.

, $(75.34, 76.76)$



OBTAIN MEAN AND STANDARD DEVIATION OF THE SAMPLE OF 100

MIDPOINT	64	70	77	84	92
FREQUENCY	11	22	45	16	6

$\sum x = 7605$, $\sum x^2 = 58334$, $n = 100$

$\bar{x} = \frac{\sum x}{n} = \frac{7605}{100} = 76.05$

$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right] = \frac{1}{99}\left[58334 - \frac{7605^2}{100}\right] = \frac{6641}{132}$

NOW THE CONFIDENCE INTERVAL (98%)

$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} \Phi^{-1}(0.99)$

$\mu = 76.05 \pm \frac{\sqrt{\frac{6641}{132}}}{\sqrt{100}} \times 2.3263$

$\mu = 76.05 \pm 0.7093\ldots$

$\therefore CI = (75.34, 76.76)$

ASSUMPTIONS & VALIDATIONS

$s^2 = \sigma^2$ ASSUMED THAT VARIANCE OF POPULATION (3000) IS ROUGHLY THE SAME AS THAT OF THE SAMPLE (AS IT IS LARGE)

ASSUMED WEIGHTS ARE NORMALLY DISTRIBUTED OR IF THEY ARE NOT THE SAMPLING DISTRIBUTION OF THE MEAN WILL STILL BE APPROXIMATELY NORMAL DUE TO THE LARGE SAMPLE (C.L.T)

Question 12 (***+)

The weights of baclava portions produced by a bakery are Normally distributed with a standard deviation of 5 grams and unknown mean.

The weights of seven randomly selected portions are shown below.

$$77, \quad 80, \quad 74, \quad 72, \quad 83, \quad 85, \quad 75$$

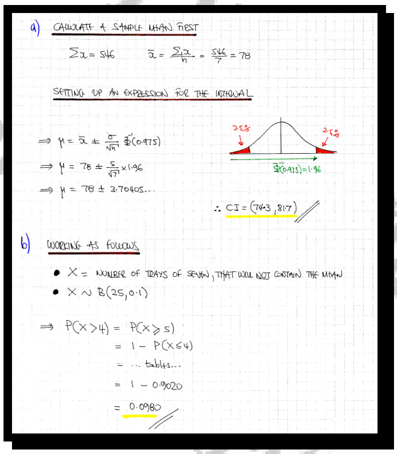
a) Find a 95% confidence interval for the mean weight of the baclava portions produced by this bakery.

The chef produces 175 baclava portions and randomly places them in 25 trays of 7 portions in each tray.

He calculates 25 separate 90% confidence intervals for the mean weight of a portion, from each tray.

b) Determine the probability that more than 4 of these intervals **will not** contain the mean.

$\boxed{}$, $\boxed{(74.3, 81.7)}$, $\boxed{0.0980}$


a) CALCULATE A SAMPLE MEAN FIRST

$\sum x = 546 \qquad \bar{x} = \frac{\sum x}{n} = \frac{546}{7} = 78$

SETTING UP AN EXPRESSION FOR THE INTERVAL

$\Rightarrow \mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(0.975)$

$\Rightarrow \mu = 78 \pm \frac{5}{\sqrt{7}} \times 1.96$

$\Rightarrow \mu = 78 \pm 3.70405\ldots$

2.5%, 2.5%, $\Phi^{-1}(0.975) = 1.96$

$\therefore \text{CI} = (74.3, 81.7)$

b) WORKING AS FOLLOWS

- X = NUMBER OF TRAYS OF SEVEN, THAT WILL NOT CONTAIN THE MEAN
- $X \sim B(25, 0.1)$

$\Rightarrow P(X > 4) = P(X \geqslant 5)$
$= 1 - P(X \leqslant 4)$
$= \ldots \text{tables} \ldots$
$= 1 - 0.9020$
$= 0.0980$

Question 13 (****)

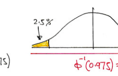
A saw mill cut planks of wood whose lengths X m are such so that $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

A random sample of these planks were chosen and a 95% confidence interval for μ was calculated to be $(5.85, 6.34)$.

Find the standard error for the mean and hence construct a 90% confidence interval for μ, based on the same sample.

$\boxed{\text{standard error} = \dfrac{\sigma}{\sqrt{n}} = 0.125}$, $\boxed{(5.89, 6.30)}$

$X \sim N(\mu, \sigma^2)$



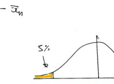
- $\dfrac{6.34 - 5.85}{2} = \dfrac{\sigma}{\sqrt{n}} \Phi^{-1}(0.975)$

$0.245 = \dfrac{\sigma}{\sqrt{n}} \times 1.96$

$\dfrac{\sigma}{\sqrt{n}} = 0.125$ (STANDARD ERROR)

- $\dfrac{6.34 + 5.85}{2} = 6.095 \leftarrow \bar{x}_n$

HENCE WE HAVE



$\mu = \bar{x}_n \pm \dfrac{\sigma}{\sqrt{n}} \Phi^{-1}(0.95)$

$\mu = 6.095 \pm (0.125)(1.6449)$

$\mu = 6.095 \pm 0.2056\ldots$

$\therefore$ C.I $= (5.89, 6.30)$

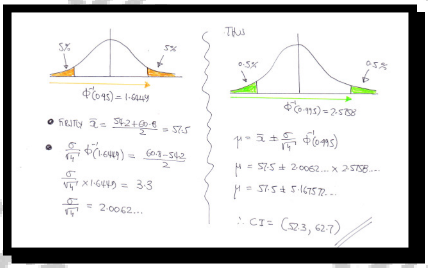
Question 14 (****)

A continuous variable has distribution $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

A random sample of n observations of X produced a 90% confidence interval for μ, given to 1 decimal place as $(54.2, 60.8)$.

Find a 99% confidence interval for μ.

$$\boxed{(52.3, 62.7)}$$



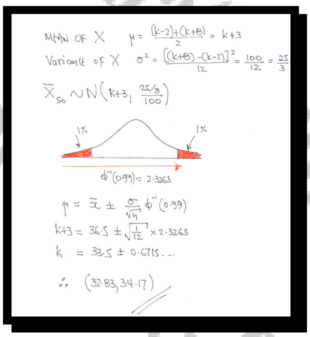
Question 15 (****)

The continuous random variable X has uniform distribution over the real interval $[k-2, k+8]$, where k is a positive constant.

One hundred random observations of X produced a sample mean of 36.5.

Find a 98% confidence interval for k.

$$\boxed{(47.2, 67.2)}$$



Question 16 (****+)

A continuous variable has distribution $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

A random sample of 100 observations of X produced a 95% confidence interval for μ, as $(60.08, 67.92)$.

a) Determine the mean and standard deviation of the sample.

b) Calculate the percentage confidence for μ of an interval of width 6, using the same statistics obtained from the same 100 observations.

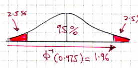
c) If instead a 95% confidence interval for μ of width 6 is required, determine the minimum number of observations of X needed to accomplish this task.

$\boxed{}$, $\boxed{(\mu, \hat{\sigma}) = (64, 20)}$, $\boxed{86.64\%}$, $\boxed{n_{\min} = 171}$

$\boxed{X \sim N(\mu, \sigma^2)}$

a) $n = 100$ CI $= (60.08, 67.92)$ AT 95%

- SAMPLE MEAN $\bar{x} = \frac{60.08 + 67.92}{2} = 64$
- HALF THE WIDTH OF THE INTERVAL $= \Phi^{-1}(0.975) \times \frac{\sigma}{\sqrt{n}}$



$\Rightarrow \frac{67.92 - 60.08}{2} = 1.96 \times \frac{\sigma}{\sqrt{100}}$

$\Rightarrow 3.92 = 0.196\sigma$

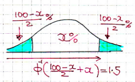
$\Rightarrow \sigma = 20$

b) $n = 100$, WIDTH $= 6$, $\sigma = 20$

$\Rightarrow \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{100-x}{2} + x\right) = 3$

$\Rightarrow \frac{20}{\sqrt{100}} \Phi^{-1}\left(\frac{100-x}{2} + x\right) = 3$

$\Rightarrow \Phi^{-1}\left(\frac{100-x}{2} + x\right) = 1.5$



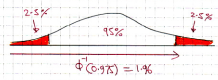
$\Rightarrow \frac{100-x}{2} + x = 93.32$

$\Rightarrow 100 - x + 2x = 186.64$

$\Rightarrow x = 86.64$

IE 86.64% CONFIDENCE INTERVAL

c) WIDTH $= 6$, 95% CONFIDENCE, $\sigma = 20$



$\Phi^{-1}(0.975) \times \frac{\sigma}{\sqrt{n}} = 3$

$1.96 \times \frac{20}{\sqrt{n}} = 3$

$\sqrt{n} = 13.0666\ldots$

$n = 170.7377\ldots$

$\therefore$ REQUIRED SAMPLE SIZE IS 171

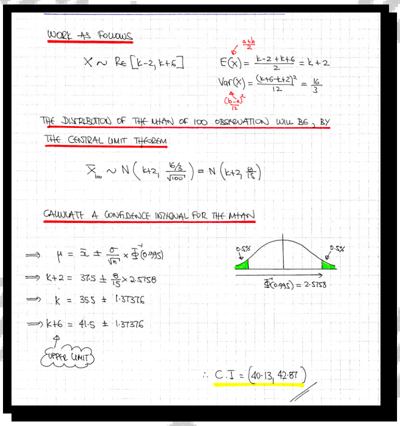
Question 27 (****+)

A continuous random variable X is uniformly distributed in the interval $[k-2, k+6]$, where k is a constant.

A random sample of 100 observations of X produced a sample mean of 37.5.

Determine a 99% confidence interval for the **upper bound** of X.

$\boxed{}$, $\boxed{(40.13, 42.87)}$



WORK AS FOLLOWS

$X \sim Re[k-2, k+6]$

$E(X) = \frac{k-2+k+6}{2} = k+2$ $\quad \frac{a+b}{2}$

$\mathrm{Var}(X) = \frac{(k+6-k+2)^2}{12} = \frac{16}{3}$ $\quad \frac{(b-a)^2}{12}$

THE DISTRIBUTION OF THE MEAN OF 100 OBSERVATION WILL BE, BY THE CENTRAL LIMIT THEOREM

$\bar{X}_{100} \sim N\left(k+2, \frac{16/3}{\sqrt{100}}\right) = N\left(k+2, \frac{8}{15}\right)$

CALCULATE A CONFIDENCE INTERVAL FOR THE MEAN

$\Rightarrow \mu = \bar{x} \pm \frac{\sigma}{\sqrt{n}} \times \Phi^{-1}(0.995)$

$\Rightarrow k+2 = 37.5 \pm \frac{8}{15} \times 2.5758$

$\Rightarrow k = 35.5 \pm 1.37376$

$\Rightarrow k+6 = 41.5 \pm 1.37376$

UPPER LIMIT

0.5% 0.5%

$\Phi^{-1}(0.995) = 2.5758$

$\therefore$ C.I $= (40.13, 42.87)$

Question 18 (****+)

The continuous random variables X and Y are independent and have respective distributions $\mathrm{N}\left(\mu,4^2\right)$ and $\mathrm{N}\left(\mu,2^2\right)$.

A random sample of 4 observation of X is taken and $\bar{X}$ denotes the sample mean.

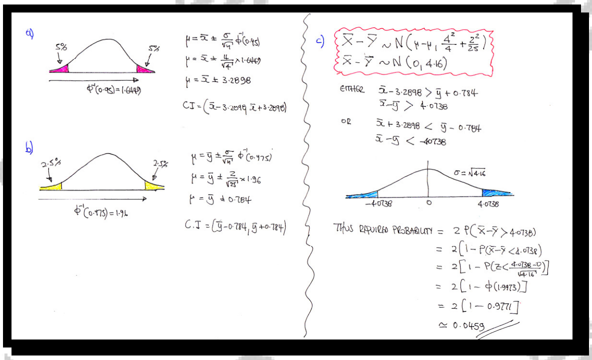
a) Find a 90% confidence interval for μ, in terms of $\bar{X}$.

A random sample of 25 observation of Y is taken and $\bar{Y}$ denotes the sample mean.

b) Find a 95% confidence interval for μ, in terms of $\bar{Y}$.

c) By considering the distribution of $\bar{X}-\bar{Y}$, calculate the probability that the two confidence intervals of part (a) and part (b) **do not** overlap.

$\boxed{\left(\bar{X}-3.2898, \bar{X}+3.2898\right)}$, $\boxed{\left(\bar{Y}-0.784, \bar{Y}+0.784\right)}$, $\boxed{0.0459}$



NORMAL DISTRIBUTION HYPOTHESIS TESTING

Question 1 **(**)**

The continuous random variable X is Normally distributed with mean μ and standard deviation 10.

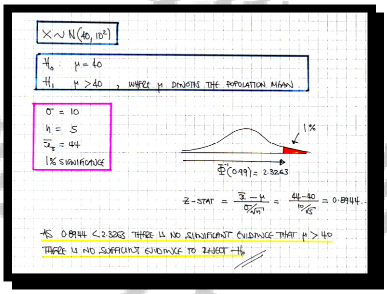
A test is to be carried out for the hypotheses

$$H_0 : \mu = 40 \text{ versus } H_1 : \mu > 40 .$$

A random sample of 5 observations of X produced a sample mean of 44.

Carry out the test at the 1% significance level.

$\boxed{\text{not significant}}$



Question 2 (**)

The continuous random variable X is Normally distributed with mean μ and standard deviation 16.

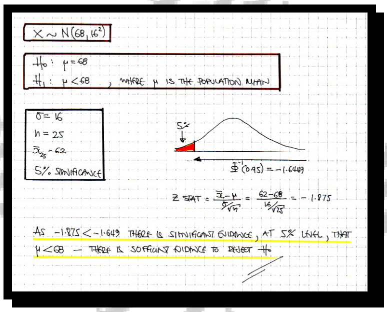
A test is to be carried out for the hypotheses

$$H_0 : \mu = 68 \text{ versus } H_1 : \mu < 68 .$$

A random sample of 25 observations of X produced a sample mean of 62.

Carry out the test at the 5% significance level.

$\boxed{\text{significant}}$



Question 3 **(**)**

The continuous random variable X is Normally distributed with mean μ and standard deviation 45.

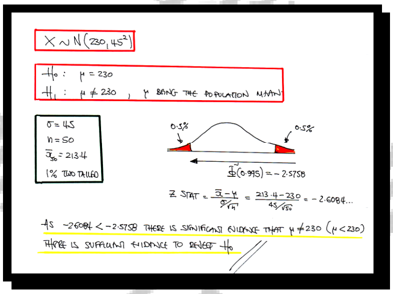
A test is to be carried out for the hypotheses

$$H_0 : \mu = 230 \text{ versus } H_1 : \mu \neq 230 .$$

A random sample of 50 observations of X produced a sample mean of 213.4.

Carry out the test at the 1% significance level.

$\boxed{\text{significant}}$



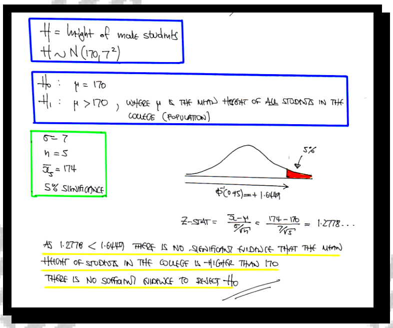
Question 4 (**)

The heights of male students in a college are thought to be Normally distributed with mean 170 cm and standard deviation 7.

The heights of 5 male students from this college are measured and the sample mean was 174 cm.

Determine, at the 5% level of significance, whether there is evidence that the mean height of the male students of this college is higher than 170 cm.

$\boxed{\text{not significant}}$



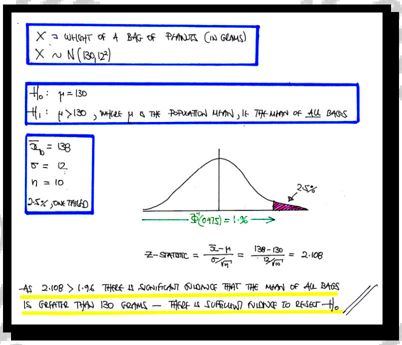
Question 5 (**)

The weights of bags of peanuts filled by a vendor in a sports venue are thought to be Normally distributed with mean 130 g and standard deviation 12.

Ten such bags are sampled and their mean weight is found to be 138 g.

Determine, at the 2.5% level of significance, whether there is evidence that the mean weight of the bags of peanuts, sold by this vendor, is over 130 g.

$\boxed{\text{significant}}$



Question 6 (**)

The lifetime of the car tyres produced by a company is known to be Normally distributed with mean 8500 miles and standard deviation 2000 miles.

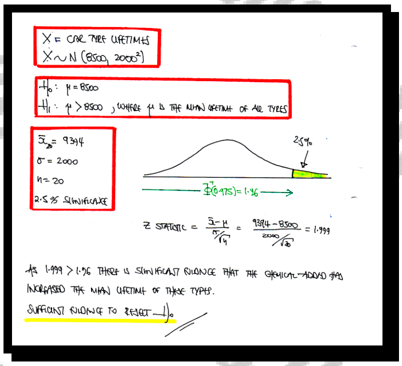
A chemical is added to the rubber compound, which is thought to increase the lifetime of these car tyres.

Twenty new tyres with the chemical added to the rubber are now tested.

Their mean lifetime was 9394 miles.

Determine, at the 2.5% level of significance, whether there is evidence that the mean lifetime of the car tyres is now higher than 8500 miles.

significant



Question 7 (+)**

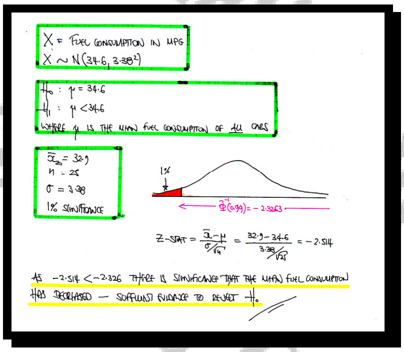
The fuel consumption of a certain make of car is modelled by a Normal distribution with a mean of 34.6 mpg and standard deviation of 3.38 mpg.

A refinement is made to the engine which is thought to improve performance without affecting the fuel economy.

The fuel consumption of 25 cars, with the refined engine, now produced a mean fuel consumption of 32.9 mpg.

Determine, at the 1% level of significance, whether there is evidence that the mean fuel consumption figure of the car has decreased after the engine refinement.

$\boxed{\text{significant}}$



Question 8 (**+)

The volume of juice in cups filled in a canteen is assumed to have a mean of 250 ml and standard deviation of 25.

The volume of forty such cups was measured and the mean was 246 ml.

Determine, at the 10% level of significance, whether the assumption that the mean is 250 ml is justified.

$\boxed{\text{not significant}}$

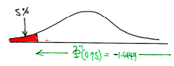
X = VOLUME OF JUICE DISPENSED IN A CUP
$\bar{X}_{40} \sim N(250, 25 \quad)$ BY C.L.T

$H_0 : \mu = 250$
$H_1 : \mu \neq 250$, WHERE μ IS THE MEAN VOLUME OF ALL CUPS

$\bar{x}_{40} = 246$
$\sigma = 25$
$n = 40$
10% SIGNIFICANCE

"TWO TAILED, i.e. 5% EACH TAIL"

5%



$\Phi^{-1}(0.95) = -1.6449$

z STAT $= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{246-250}{25/\sqrt{40}} = -1.0119$

AS $-1.0119 > -1.6449$ THE CLAIM IS JUSTIFIED
INSUFFICIENT EVIDENCE TO REJECT H_0

Question 9 (**+)

The battery lifetime of a certain make of laptop is claimed to have a mean of 6.6 hours and variance of 3.9 hours^2.

The battery lifetimes of forty such laptops were measured and the mean was 6.1 hours.

Determine, at the 10% level of significance, whether the claim on the battery lifetime of the laptop is justified.

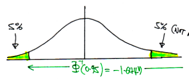
$\boxed{\text{not significant}}$

X = BATTERY LIFETIME OF A LAPTOP IN HOURS
$\bar{X}_{40} \sim N\left(6.6, \frac{3.9}{40}\right) \leftarrow$ BY THE CENTRAL LIMIT THEOREM

$H_0 : \mu = 6.6$
$H_1 : \mu \neq 6.6$, WHERE μ IS THE MEAN LIFETIME OF THE BATTERIES OF ALL LAPTOPS

$\bar{x}_{40} = 6.1$
$\sigma^2 = 3.9$
$n = 40$
10%, TWO TAILED

5% 5% (NOT NEEDED)



$\Phi^{-1}(0.95) = -1.6449$

z STAT $= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{6.1-6.6}{\sqrt{3.9}/\sqrt{40}} = -1.601$

AS $-1.601 > -1.6449$ THE CLAIM IS JUSTIFIED – INSUFFICIENT EVIDENCE TO REJECT H_0

Question 10 (***)

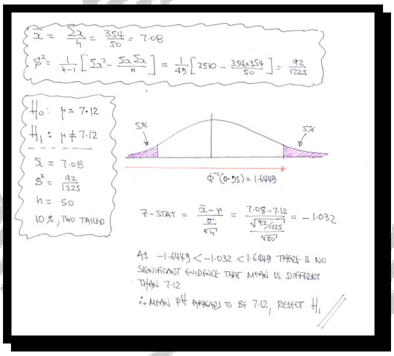
The acidity X, measured in pH, of limestone collected from different areas is thought to be Normally distributed.

The pH of 50 random samples produced the following summary statistics.

$$\sum x = 354, \quad \sum x^2 = 2510.$$

Test, at the 10% level of significance, whether the mean pH of limestone is 7.12.

$\boxed{\text{not significant}}$



Question 11 (***)

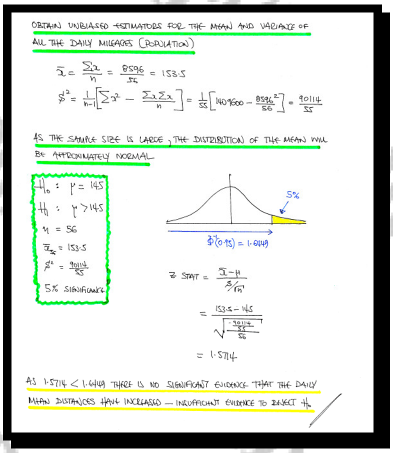
A minicab driver feels that his daily mileage figures have recently increased compared to those of last year. He knows that his daily mean mileage last year was 145.

He records the next current 56 daily mileage figures, x miles, and he obtains

$$\sum x = 8596 \qquad \text{and} \qquad \sum x = 1409600$$

Explain briefly what conclusions can be drawn from a suitable test, at the 5% level of significance.

$\boxed{}$, $\boxed{\text{not significant, } 1.5714 < 1.6449}$



OBTAIN UNBIASED ESTIMATORS FOR THE MEAN AND VARIANCE OF ALL THE DAILY MILEAGES (POPULATION)

$$\bar{x} = \frac{\sum x}{n} = \frac{8596}{56} = 153.5$$

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right] = \frac{1}{55}\left[1409600 - \frac{8596^2}{56}\right] = \frac{90114}{55}$$

AS THE SAMPLE SIZE IS LARGE, THE DISTRIBUTION OF THE MEAN WILL BE APPROXIMATELY NORMAL

$H_0 : \mu = 145$
$H_1 : \mu > 145$
$n = 56$
$\bar{x}_{56} = 153.5$
$s^2 = \frac{90114}{55}$
5% SIGNIFICANCE

$$z \text{ STAT} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{153.5 - 145}{\sqrt{\frac{\frac{90114}{55}}{56}}} = 1.5714$$

AS $1.5714 < 1.6449$ THERE IS NO SIGNIFICANT EVIDENCE THAT THE DAILY MEAN DISTANCES HAVE INCREASED — INSUFFICIENT EVIDENCE TO REJECT H_0

Question 12 (***+)

The time, in minutes, taken by a large group of students to complete an Economics exam are thought to be Normally distributed with mean μ and standard deviation σ.

15% of the students finished the exam in under 74 minutes while 20% used in excess of 115 minutes.

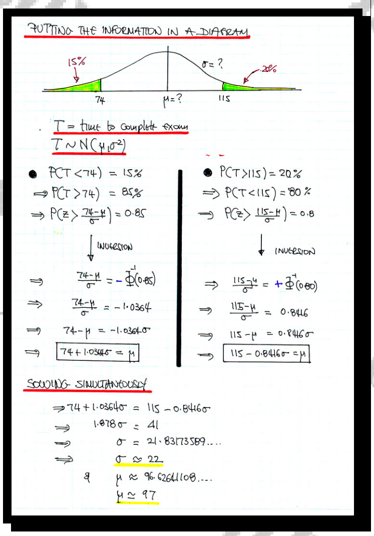
a) Find, correct to the nearest minute, the value of μ and the value of σ.

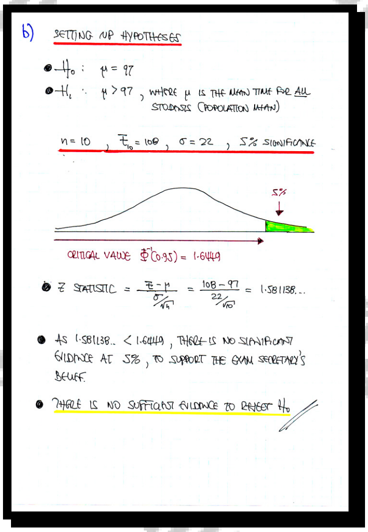
The school exam secretary believes that the value of μ is much higher than the one found in part **(a)**, based on a random sample of 10 students whose mean time to complete the exam was 108 minutes.

b) Using the value of σ found in part **(a)**, conduct a hypothesis test at the 5% level of significance to investigate the school exam secretary's belief.

State your hypotheses clearly.

$\boxed{\mu = 97}$, $\boxed{\sigma = 22}$, $\boxed{\text{not significant, } 1.5811 < 1.6449}$





Question 13 (***+)

In a clothing factory, the time taken by machines to manufacture a certain type of shirt, are assumed to be Normally distributed with a mean of 44 minutes and a standard deviation of 4 minutes.

a) Determine the value of t, if 10.56% of the shirts take more than t minutes to be manufactured.

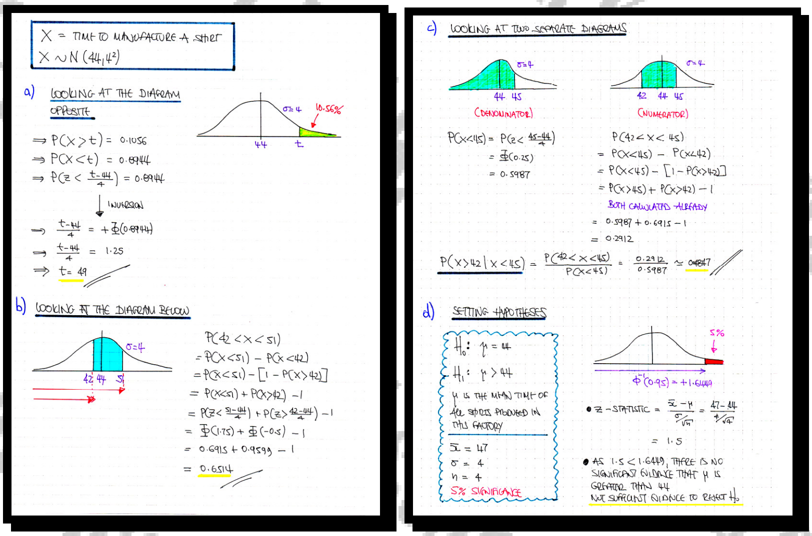
b) Find the probability that a shirt picked at random, took between 42 and 51 minutes to manufacture.

c) If a shirt took less than 45 minutes to be made, calculate the probability that it in fact took more than 42 minutes to make.

The owner of the factory believes that the mean time is greater than 44 minutes, due to the aging machinery. He finds that the mean manufacturing time of a random sample of 4 shirts to be 47 minutes.

d) By clearly stating suitable hypotheses, test at the 5% level of significance the owner's belief.

$\boxed{t=49}$, $\boxed{0.6514}$, $\boxed{0.4847}$, $\boxed{\text{not significant}}$



Question 14 (***+)

The weekly mileages of a sales rep are thought to be Normally distributed with mean μ and standard deviation σ.

5% of his weekly mileages are less than 850 miles and 1% exceed 960 miles.

a) Find, correct to the nearest mile, the value of μ and the value of σ.

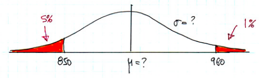
The rep believes that the value of μ is much lower than the one found in part **(a)**, based on a random sample of 4 weeks whose mean mileage was 863.

b) Using the value of σ found in part **(a)**, conduct a hypothesis test at the 1% level of significance to investigate the rep's belief.

State your hypotheses clearly.

$\boxed{\mu = 896}$, $\boxed{\sigma = 28}$, $\boxed{\text{significant, } -2.3571 < -2.3263}$

a) PUTTING THE INFORMATION IN A DIAGRAM



$\boxed{\begin{array}{l} X = \text{weekly mileages} \\ X \sim N(\mu, \sigma^2) \end{array}}$

- $P(X<850) = 5\%$
$\Rightarrow P(X>850) = 95\%$
$\Rightarrow P\left(Z > \frac{850-\mu}{\sigma}\right) = 0.95$
$\downarrow$ INVERSION
$\Rightarrow \frac{850-\mu}{\sigma} = -\Phi^{-1}(0.95)$
$\Rightarrow \frac{850-\mu}{\sigma} = -1.6449$
$\Rightarrow 850-\mu = -1.6449\sigma$
$\Rightarrow \boxed{850 + 1.6449\sigma = \mu}$

- $P(X>960) = 1\%$
$\Rightarrow P(X<960) = 99\%$
$\Rightarrow P\left(Z < \frac{960-\mu}{\sigma}\right) = 0.99$
$\downarrow$ INVERSION
$\Rightarrow \frac{960-\mu}{\sigma} = +\Phi^{-1}(0.99)$
$\Rightarrow \frac{960-\mu}{\sigma} = 2.3263$
$\Rightarrow 960-\mu = 2.3263\sigma$
$\Rightarrow \boxed{960 - 2.3263\sigma = \mu}$

SOLVING SIMULTANEOUSLY

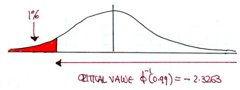
$\Rightarrow 850 + 1.6449\sigma = 960 - 2.3263\sigma$
$\Rightarrow 3.9712\sigma = 110$
$\Rightarrow \sigma = 27.69943594\ldots$
$\Rightarrow \sigma \approx 28$

$\&\ \mu = 895.5628022\ldots$
$\mu \simeq 896$

b) SETTING UP HYPOTHESES

- $H_0: \mu = 896$
- $H_1: \mu < 896$, WHERE μ IS THE MEAN OF ALL WEEKLY MILEAGES (POPULATION MEAN)

$n = 4$, $\bar{x}_4 = 863$, $\sigma = 28$, 1% SIGNIFICANCE



CRITICAL VALUE $\Phi^{-1}(0.99) = -2.3263$

- $Z\text{-STATISTIC} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{863-896}{28/\sqrt{4}} = -2.357142\ldots$
- AS $-2.357142\ldots < -2.3263$ THERE IS SIGNIFICANT EVIDENCE, AT 1% LEVEL, TO SUPPORT THE SALES REP'S BELIEF
- THERE IS SUFFICIENT EVIDENCE TO REJECT H_0

Question 15 (***+)

The manager of a cinema believes that the weights, X grams, of popcorn bags sold in his cinema are Normally distributed with mean of 340 and standard deviation of 10.

a) Taking $\mu = 340$ and $\sigma = 10$ ...

i. ... find, to the nearest gram, the weight x_0 exceeded by 5% of these popcorn bags.

ii. ... $\mathrm{P}\left(X > \mu \mid X < x_0\right)$.

A new manager which takes over asks his staff to investigate the weights of these popcorn bags and is told that it was subsequently found that $\mu = 320$ and $\sigma = 10$.

The new manager claims that μ has to be higher than 320, as the mean of a random sample of 5 bags was found to be 327.

b) Using $\sigma = 10$, conduct a hypothesis test at the 5% level of significance to investigate the new manager's claim.

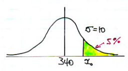
State your hypotheses clearly.

$\boxed{}$, $\boxed{356}$, $\boxed{\frac{9}{19} \approx 0.4737}$, $\boxed{\text{not significant, } 1.5652 < 1.6449}$

a) I) POTTING THE INFORMATION INTO A DIAGRAM

X = weight of pop-corn bags
$X \sim N(340, 10^2)$

$\Rightarrow P(X > x_0) = 5\%$
$\Rightarrow P(X < x_0) = 95\%$
$\Rightarrow P\left(Z < \frac{x_0 - 340}{10}\right) = 0.95$



↓ INVERSION

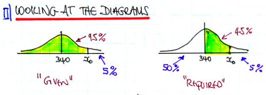
$\frac{x_0 - 340}{10} = + \Phi^{-1}(0.95)$

$\frac{x_0 - 340}{10} = 1.6449$

$x_0 = 356.449$

$\therefore x_0$ IS 356 grams

II) LOOKING AT THE DIAGRAMS

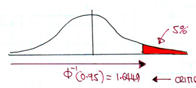


$P(X > \mu \mid X < x_0) = \frac{\text{"}45\%\text{"}}{\text{"}95\%\text{"}} = \frac{45}{95} = \frac{9}{19} \approx 0.4737$

b) SETTING UP THE HYPOTHESES

- $H_0 : \mu = 320$
- $H_1 : \mu > 320$, WHERE μ IS THE MEAN WEIGHT OF ALL BAGS OF POPCORNS

$n = 5$, $\bar{x}_5 = 327$, $\sigma = 10$, 5% SIGNIFICANCE



$\Phi^{-1}(0.95) = 1.6449$ ← CRITICAL VALUE

- Z STATISTIC $= \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{327 - 320}{10/\sqrt{5}} = 1.5652\ldots$
- AS $1.5652 < 1.6449$ THERE IS NO SIGNIFICANT EVIDENCE (AT 5%) THAT THE MEAN WEIGHT OF BAGS IS OVER 320
- THERE IS NO SUFFICIENT EVIDENCE TO REJECT H_0

Question 16 (***+)

A small factory produces $2-$litre bottles of mineral water. The mean volume of water is each bottle was known to be 2020 ml.

After the bottling machine was replaced by a new machine, the volume of water x ml was recorded in a random sample of 100 bottles.

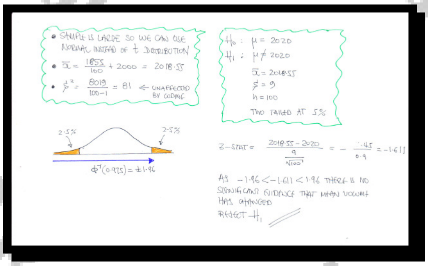
The following data was obtained

$$\sum y = 1855 \quad \text{and} \quad \sum (y-\bar{y})^2 = 8019,$$

where $y = x - 2000$.

Carry out a hypothesis test, at the 5% level of significance, to determine whether the mean volume in a $2-$litre bottle of mineral water is different when bottled by the new machine.

$\boxed{\text{not significant}}$



Question 17 (***+)

The continuous random variable X has a Normal distribution with mean of 425 and a standard deviation of 20.

a) Determine the value of ...

i. ... $\mathrm{P}(X > 455)$

ii. ... $\mathrm{P}(395 < X < 455)$

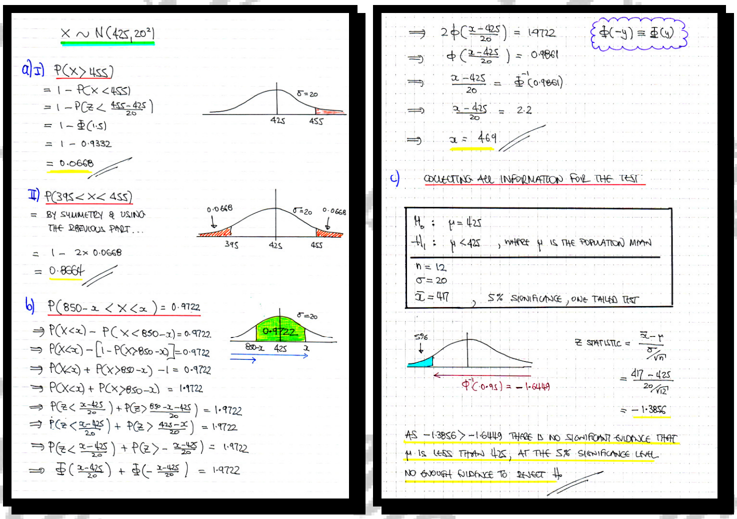
b) Find the value of x, given further that

$$\mathrm{P}(850 - x < X < x) = 0.9722 .$$

It is believed that the mean of X could be less than 425, as the mean of a random sample of 12 independent observations of X was 417.

c) Test the validity of this belief, at 5% level of significance, stating clearly all the relevant quantities and hypotheses.

$\boxed{}$, $\boxed{0.0868}$, $\boxed{0.8664}$, $\boxed{x = 469}$, $\boxed{\text{not significant, z-stat} = -1.3856 > -1.6449}$



Question 18 (***+)

Tim's tennis serve has a mean speed of 125 miles per hour.

Tim buys a new racket and wishes to know whether or not using this racket has changed the mean speed of his serve. He decides to measure the speed of a random sample of 10 serves with his new racket.

The speeds, in miles per hour, are shown below.

127.0, 124.6, 122.8, 127.2, 123.1, 124.8, 126.0, 123.5, 123.2, 121.8.

You may assume that this random sample comes from a Normal distribution with standard deviation 1.1.

Determine the p-value for these results and state the conclusion in context at the 5% level of significance.

$\boxed{p\text{-value } = 0.0846, \text{ not significant, } 8.46\% > 5\%}$

- START BY FINDING THE SAMPLE MEAN

$$\bar{x}_{10} = \frac{127.0 + 124.6 + 122.8 + \dots + 121.8}{10} = \frac{1244}{10} = 124.4$$

- SET THE HYPOTHESES

$H_0: \mu = 125$

$H_1: \mu \neq 125$

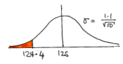
$\bar{x}_{10} = 124.4$

$\sigma = 1.1$

$n = 10$

- THUS WE NOW HAVE

$$P(\bar{x}_{10} < 124.4)$$
$$= 1 - P(\bar{x}_{10} > 124.4)$$
$$= 1 - P\left(z > \frac{124.4 - 125}{\frac{1.1}{\sqrt{10}}}\right)$$
$$= 1 - \Phi(-1.724878\ldots)$$
$$= 1 - 0.9577$$
$$= 0.0423$$
$$= 4.23\%$$



- AS THE TEST IS TWO TAILED THE p-VALUE IS $2 \times 4.23\% = 8.46\%$

COMPARING WITH 5%, THERE IS NO SIGNIFICANT EVIDENCE THAT THE MEAN SERVING SPEED HAS CHANGED – NOT SUFFICIENT EVIDENCE TO REJECT H_0

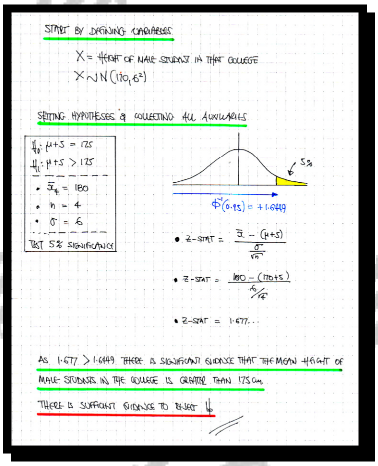
Question 19 (***+)

The heights of male students in a college are thought to be Normally distributed with mean 170 cm and standard deviation 6.

The heights of 4 male students from this college are measured and the sample mean was 180 cm.

Determine, at the 5% level of significance, whether there is evidence that the mean height of the male students of this college is greater than **175** cm.

$\boxed{}$, $\boxed{\text{significant}}$



START BY DEFINING VARIABLES

X = HEIGHT OF MALE STUDENT IN THAT COLLEGE

$X \sim N(170, 6^2)$

SETTING HYPOTHESES & COLLECTING ALL AUXILIARIES

$H_0: \mu + 5 = 175$

$H_1: \mu + 5 > 175$

- $\bar{x}_4 = 180$
- $n = 4$
- $\sigma = 6$

TEST 5% SIGNIFICANCE

5%

$\Phi^{-1}(0.95) = +1.6449$

- $z\text{-STAT} = \dfrac{\bar{x} - (\mu+5)}{\frac{\sigma}{\sqrt{n}}}$
- $z\text{-STAT} = \dfrac{180 - (170+5)}{6/\sqrt{4}}$
- $z\text{-STAT} = 1.677\ldots$

AS $1.677 > 1.6449$ THERE IS SIGNIFICANT EVIDENCE THAT THE MEAN HEIGHT OF MALE STUDENTS IN THE COLLEGE IS GREATER THAN 175 cm.

THERE IS SUFFICIENT EVIDENCE TO REJECT H_0

Question 20 (****)

The times, in minutes, taken by Year 6 students to complete the SATS Science test are assumed to be Normally distributed with mean of 48 and standard deviation of 5.

8% of the students finished the exam in less than t minutes.

a) Find the value of t, correct to the nearest minute.

b) Determine the probability that a randomly chosen student took more than 57 minutes to complete the test.

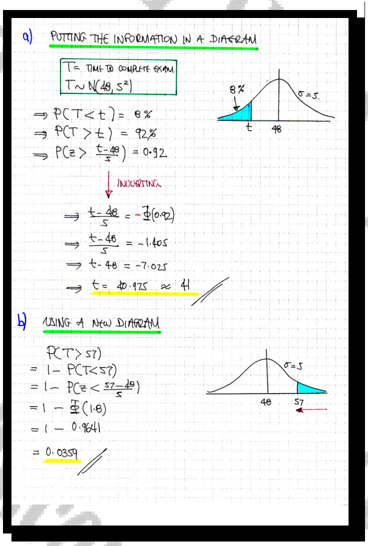
20 students that sat the SATS Science test are selected at random.

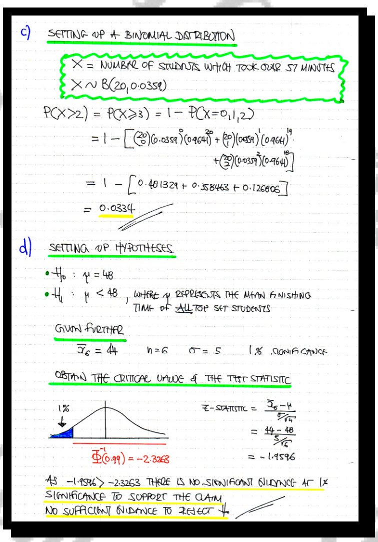
c) Calculate, correct to 3 significant figures, the probability that more than 2 of these 20 students took more than 57 minutes to complete the test.

It is claimed that the "top set students" take less time to complete the exam as the mean finishing time of a random sample of 6 "top set students" was 44 minutes.

d) Test this claim at the 1% level of significance.

$\boxed{}$, $\boxed{t \approx 41}$, $\boxed{0.0359}$, $\boxed{0.0334}$, $\boxed{\text{not significant, } -1.9596 > -2.3263}$





Question 21 (****)

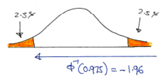
A pharmaceutical company spokesman claims that a certain pill contains 250 mg of active ingredient. Tests carried out on 120 tablets resulted in a sample mean of 249 mg, with a standard deviation of s mg.

If the pharmaceutical company's spokesman claim was just rejected at the 5% level of significance, find the largest possible value of s, correct to 1 decimal place.

$\boxed{}$, $\boxed{s \approx \sigma \approx 5.58}$

- FIRSTLY THE SAMPLE IS LARGE $n=120$, SO
  - THE SAMPLING DISTRIBUTION OF THE MEAN WILL BE APPROX NORMAL
  - $s \approx \sigma$
  - WE CAN USE NORMAL DISTRIBUTION INSTEAD OF t-DISTRIBUTION
- SETTING THE HYPOTHESES

$$\boxed{\begin{array}{l} H_0: \ \mu = 250 \\ H_1: \ \mu \neq 250 \\ \hline n = 120 \\ \bar{x}_{120} = 249 \\ s \approx \sigma = ? \\ \hline 5\% \text{ TWO TAILED} \end{array}}$$



$\Phi^{-1}(0.975) = -1.96$

- Z-STAT $= \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
- HERE $\dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < -1.96$

$$\Rightarrow \frac{249 - 250}{\frac{\sigma}{\sqrt{120}}} < -1.96$$

$$\Rightarrow \frac{-\sqrt{120}}{\sigma} < -1.96$$

$$\Rightarrow -1.96\sigma > -\sqrt{120}$$

$$\Rightarrow \sigma < 5.589$$

$$\therefore \ \sigma \approx s \approx 5.58$$

Created by T. Madas

Question 22 (****+)

The volume of coffee, X ml, poured into a cup by a drink dispenser is thought to be a Normal variable with mean of 252 ml and a standard deviation of σ ml.

a) Find the value of σ, given further that

$$\mathrm{P}(X < a) = 0.82\% \quad \text{and} \quad \mathrm{P}(X > a + 10) = 5.48\% ,$$

where a is a positive constant.

b) Determine the value of

$$\mathrm{P}\left(X - 2a - 14 + \frac{64000}{X} < 0 \right).$$

It is believed that the mean volume of coffee pored into each cup could be more than 252 ml, as the mean of a random sample of 5 such cups was 255 ml.

c) Test the validity of this belief, at 1% level of significance, stating clearly all the relevant quantities and hypotheses.

$\boxed{\sigma = 2.5}$, $\boxed{0.7333}$, $\boxed{\text{significant, z-stat} = 2.6833 > 2.3263}$

a) PUTTING INFORMATION INTO A DIAGRAM

X = VOLUME OF COFFEE DISPENSED (ml)
$X \sim N(252, \sigma^2)$

$\Rightarrow \mathrm{P}(X<a) = 0.0082$
$\Rightarrow \mathrm{P}(X>a) = 0.9918$
$\Rightarrow \mathrm{P}\left(Z > \frac{a-252}{\sigma}\right) = 0.9918$
$\Rightarrow \Phi\left(\frac{a-252}{\sigma}\right) = 0.9918$
$\Rightarrow \frac{a-252}{\sigma} = -\Phi^{-1}(0.9918)$
$\Rightarrow \frac{a-252}{\sigma} = -2.40$
$\Rightarrow a - 252 = -2.40\sigma$
$\Rightarrow a = 252 - 2.40\sigma$

$\Rightarrow \mathrm{P}(X>a+10) = 0.0548$
$\Rightarrow \mathrm{P}(X<a+10) = 0.9452$
$\Rightarrow \mathrm{P}\left(Z < \frac{a+10-252}{\sigma}\right) = 0.9452$
$\Rightarrow \Phi\left(\frac{a-242}{\sigma}\right) = 0.9452$
$\Rightarrow \frac{a-242}{\sigma} = \Phi^{-1}(0.9452)$
$\Rightarrow \frac{a-242}{\sigma} = 1.60$
$\Rightarrow a - 242 = 1.60\sigma$
$\Rightarrow a = 242 + 1.60\sigma$

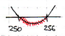
$252 - 2.4\sigma = 242 + 1.6\sigma$
$10 = 4\sigma$
$\sigma = 2.5$
($a = 246$ FOR PART (b))

b) PROCEED AS FOLLOWS

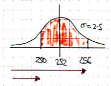
$\mathrm{P}\left(X - 2a - 14 + \frac{64000}{X} < 0\right)$ AS $X > 0$
$= \mathrm{P}(X(X-2a-14) + 64000 < 0)$
$= \mathrm{P}(X(X-506) + 64000 < 0)$
$= \mathrm{P}(X^2 - 506X + 64000 < 0)$

FACTORIZE BY THE QUADRATIC FORMULA

$= \mathrm{P}[(X-250)(X-256) < 0]$



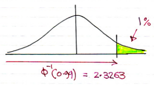
$= \mathrm{P}[250 < X < 256]$
$= \mathrm{P}(X<256) - \mathrm{P}(X<250)$
$= \mathrm{P}(X<256) - [1 - \mathrm{P}(X>250)]$
$= \mathrm{P}(X<256) + \mathrm{P}(X>250) - 1$



$= \mathrm{P}\left(Z < \frac{256-252}{2.5}\right) + \mathrm{P}\left(Z > \frac{250-252}{2.5}\right) - 1$
$= \Phi(1.6) + \Phi(-0.8) - 1$
$= 0.9452 + 0.7881 - 1$
$= 0.7333$

c) SUMMARIZING ALL INFORMATION FOR THIS TEST

$H_0: \mu = 252$
$H_1: \mu > 252$, WHERE μ IS THE POPULATION MEAN
$n = 5$
$\sigma = 2.5$
$\bar{x} = 255$, 1% SIGNIFICANCE, ONE TAILED TEST



$Z\text{-STATISTIC} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{255-252}{2.5/\sqrt{5}} = 2.6833\ldots$

AS $2.6833 > 2.3263$, THERE IS SIGNIFICANT EVIDENCE THAT μ IS GREATER THAN 252
THERE IS SUFFICIENT EVIDENCE TO REJECT H_0

NORMAL DISTRIBUTION CRITICAL REGIONS

Question 1 (***)

The random variable X is Normally distributed with mean μ and variance 25.

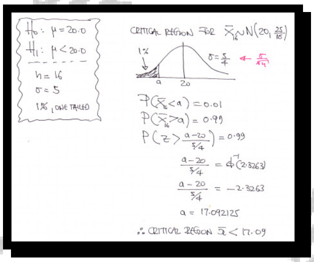
A test is to be carried out for the hypotheses

$$H_0 : \mu = 20 \text{ against } H_1 : \mu < 20.$$

A random sample of 16 observations of X produced a sample mean of $\overline{x}$.

Find, to 2 decimal places, the critical region for $\overline{x}$, at the 1% level of significance.

$\boxed{\overline{x} < 17.09}$



Question 2 (***)

It has been established over time that the car servicing times in a garage are Normally distributed with mean 62 minutes.

The new manager of the garage wants to check the validity of this claim by considering the mean of a random sample of 50 recent car servicing times.

a) Given that the times come from a Normal distribution with standard deviation of 14, find the critical region the sample mean of the 50 recent car servicing times, using a 5% level of significance.

b) Explain whether the Central Limit Theorem was used in this question.

$$\boxed{\bar{x} < 58.12 \ \cup \ \bar{x} > 65.88}$$

(a) $H_0 : \mu = 62$

$H_1 : \mu \neq 62$

$\sigma = 14$

$n = 50$

CRITICAL REGION AT 5% (TWO TAILED)

$\Phi^{-1}(0.975) = \pm 1.96$

$\bar{x} = \mu \pm \frac{\sigma}{\sqrt{n}} \Phi^{-1}(0.975)$

$\bar{x} = 62 \pm \frac{14}{\sqrt{50}} \times 1.96$

$\bar{x} = 62 \pm 3.88$

$\therefore$ CRITICAL REGION $\bar{x} < 58.12$ OR $\bar{x} > 65.88$

(b) NO NEED TO USE THE C.L.T, AS DATA COMES FROM A NORMAL DISTRIBUTION

Question 3 (***)

The running times, for a fixed distance used by the members of a jogging club, are normally distributed with a standard deviation of 12 minutes.

A two tailed hypothesis test on the mean μ of these times produced a critical region of

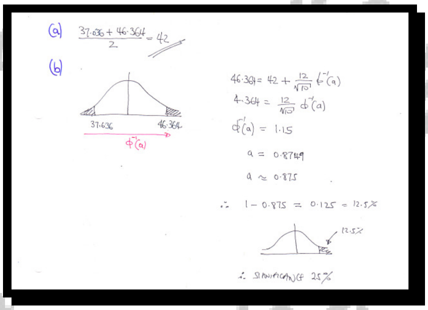
$$\bar{x}_{10} < 37.636 \ \cup \ \bar{x}_{10} > 46.364 ,$$

where $\bar{x}_{10}$ is the sample mean of 10 random running times.

a) State the value of μ.

b) Determine the significance of the region/test.

$\boxed{\mu = 42}$, $\boxed{25\%}$



NORMAL DISTRIBUTION HYPOTHESIS TESTING

Difference of Means

Question 1 (***)

The continuous random variables X and Y satisfy

$$X \sim \mathrm{N}\left(\mu_x, 4^2\right) \quad \text{and} \quad Y \sim \mathrm{N}\left(\mu_y, 6^2\right).$$

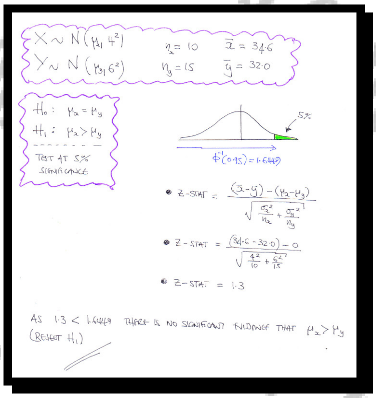
A random sample of 10 observations of X produced a sample mean of 34.6.

A random sample of 15 observations of Y produced a sample mean of 32.0.

A test is to be carried out to determine whether of not there is evidence that the mean of X is greater than the mean of Y.

Carry out the test at the 5% significance level.

$\boxed{\text{not significant}}$



Question 2 (***)

The continuous random variables X_1 and X_2 satisfy

$$X_1 \sim \mathrm{N}\left(\mu_1, 20^2\right) \quad \text{and} \quad X_2 \sim \mathrm{N}\left(\mu_2, 40^2\right).$$

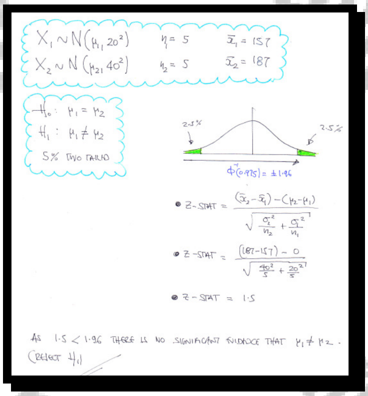
A random sample of 5 observations of X_1 produced a sample mean of 157.

A random sample of 5 observations of X_2 produced a sample mean of 187.

A test is to be carried out to determine whether of not there is evidence that the mean of X_1 is different from the mean of X_2.

Carry out the test at the 5% significance level.

$\boxed{\text{not significant}}$



Question 3 (***)

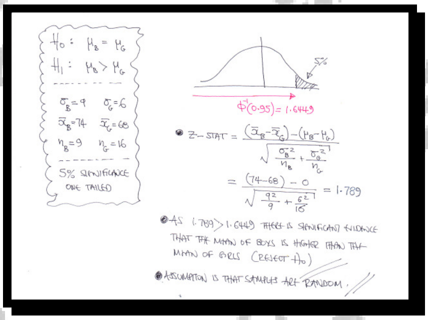
The same Mathematics mock exam is given to the Year 11 students of a certain school year after year. It has been established over time that the standard deviation of the marks is 9 for the boys and 6 for the girls. The marks for both boys and girls are thought to be normally distributed.

The marks of 9 boys in this year's mock exam had a mean of 74 while the marks of 16 girls in the same exam had a mean of 68.

The Head of Maths thinks in this academic year the Year 11 boys have a higher mean mark than that of the Year 11 girls.

Test, at the 5% level of significance, whether the claim of the Head of Maths is justified. State your hypotheses clearly, stating any additional assumptions made.

$\boxed{\text{significant}}$


$H_0 : \mu_B = \mu_G$
$H_1 : \mu_B > \mu_G$

$\sigma_B = 9 \quad \sigma_G = 6$
$\bar{x}_B = 74 \quad \bar{x}_G = 68$
$n_B = 9 \quad n_G = 16$

5% SIGNIFICANCE
ONE TAILED

5%

$\Phi^{-1}(0.95) = 1.6449$

• $Z\text{-STAT} = \dfrac{(\bar{x}_B - \bar{x}_G) - (\mu_B - \mu_G)}{\sqrt{\dfrac{\sigma_B^2}{n_B} + \dfrac{\sigma_G^2}{n_G}}} = \dfrac{(74-68) - 0}{\sqrt{\dfrac{9^2}{9} + \dfrac{6^2}{16}}} = 1.789$

• AS $1.789 > 1.6449$ THERE IS SIGNIFICANT EVIDENCE THAT THE MEAN OF BOYS IS HIGHER THAN THE MEAN OF GIRLS (REJECT H_0)

• ASSUMPTION IS THAT SAMPLES ARE RANDOM.

Question 4 (***)

It has been established over time that the javelin throwing distances of an athlete are normally distributed with a standard deviation of 8 m.

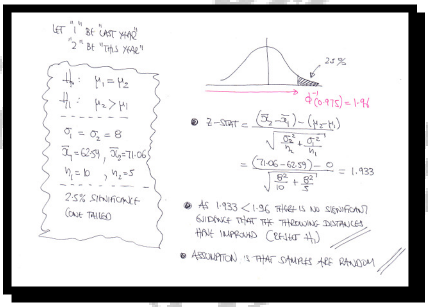
Ten throws of this athlete in the previous year had a mean of 62.59 m.

Five throws of this athlete this year had a mean of 71.06 m.

Test at the 2.5% level of significance, whether the mean throwing distances of the athlete have improved this year.

State your hypotheses clearly, stating any additional assumptions made.

$\boxed{\text{not significant}}$



Question 5 (***)

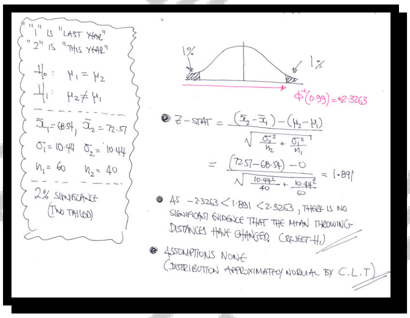
It has been established over time that the javelin throwing distances of an athlete have standard deviation of 10.44 m.

Sixty throws of this athlete in the previous year had a mean distance 68.54 m.

Forty throws of the same athlete during this year produced a mean distance 72.57 m.

Assuming the throws considered in each year are random, test at the 2% level of significance, whether the mean throwing distances of the athlete have changed since last year. State your hypotheses clearly, stating any additional assumptions made.

$\boxed{\text{not significant}}$



"1" IS "LAST YEAR"
"2" IS "THIS YEAR"

$H_0: \mu_1 = \mu_2$

$H_1: \mu_2 \neq \mu_1$

$\bar{x}_1 = 68.54$, $\bar{x}_2 = 72.57$

$\sigma_1 = 10.44$ $\sigma_2 = 10.44$

$n_1 = 60$ $n_2 = 40$

2% SIGNIFICANCE (TWO TAILED)

$\Phi^{-1}(0.99) = \pm 2.3263$

- $Z\text{-STAT} = \dfrac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\dfrac{\sigma_2^2}{n_2} + \dfrac{\sigma_1^2}{n_1}}} = \dfrac{(72.57 - 68.54) - 0}{\sqrt{\dfrac{10.44^2}{40} + \dfrac{10.44^2}{60}}} = 1.891$
- AS $-2.3263 < 1.891 < 2.3263$, THERE IS NO SIGNIFICANT EVIDENCE THAT THE MEAN THROWING DISTANCES HAVE CHANGED. (REJECT H_1)
- ASSUMPTIONS NONE (DISTRIBUTION APPROXIMATELY NORMAL BY C.L.T)

Question 6 (***)

The hardness of a certain type of rock is a Normal variable with standard deviation of 35, measured in suitable units.

Eight rock samples were collected from a certain area A and a further six rock samples were collected from a different area B.

These samples were randomly collected, and their hardness was accurately tested, recorded (in suitable units) and summarized below.

Area A: 1156, 1280, 1199, 1220, 1175, 1204, 1246, 1168.

Area B: 1175, 1143, 1159, 1142, 1224, 1147.

Test, at 5% level of significance, whether there is a difference in the mean hardness of this type of rock in the two area from where these samples were collected.

$\boxed{}$, $\boxed{\text{significant, } 2.169 > 1.96}$

OBTAIN SUMMARY STATISTICS

$\sum x_A = 9648$ $\quad \bar{x}_A = \frac{9648}{8} = 1206$

$\sum x_B = 6990$ $\quad \bar{x}_B = \frac{6990}{6} = 1165$

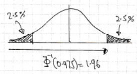
SETTING HYPOTHESES & "ESTIMATORS"

$H_0 : \mu_A = \mu_B$

$H_1 : \mu_A \neq \mu_B$

- MEAN IS 0 (AS $\mu_A - \mu_B$)
- VARIANCE $= \frac{35^2}{8} + \frac{35^2}{6} = \frac{8575}{24}$
- $\bar{x}_A - \bar{x}_B = 1206 - 1165 = 41$

USING THE Z-STATISTIC



$$Z \text{ STAT} = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{41 - 0}{\sqrt{\frac{8575}{24}}} = 2.169\ldots$$

AS $2.169 > 1.96$ THERE SIGNIFICANT EVIDENCE IN THE MEAN HARDNESS IN THE ROCKS IN THE TWO AREAS – SUFFICIENT EVIDENCE TO REJECT H_0

Question 7 (***+)

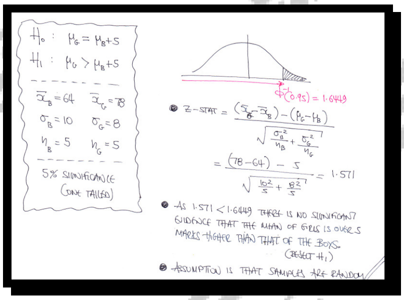
The same Mathematics mock exam is given to the Year 11 students of a certain school year after year. It has been established over time that the standard deviation of the marks is 10 for the boys and 8 for the girls. The marks for both boys and girls are thought to be normally distributed.

The marks of 5 boys in this year's mock exam had a mean of 64 while the marks of 5 girls in the same test had a mean of 78.

The Head of Maths thinks that in this academic year the mean mark of the Year 11 girls is **at least** 5 marks higher than that of the Year 11 boys.

Test, at the 5% level of significance, whether the claim of the Head of Maths is justified. State your hypotheses clearly, stating any additional assumptions made.

$\boxed{\text{not significant}}$



Question 8 (***+)

The continuous random variables X_1 and X_2 satisfy

$$X_1 \sim \mathrm{N}\left(\mu_1, 50^2\right) \quad \text{and} \quad X_2 \sim \mathrm{N}\left(\mu_2, 20^2\right).$$

A random sample of 40 observations of X_1 produced a sample mean of 1752.

A random sample of 50 observations of X_2 produced a sample mean of 1598.

A test is to be carried out to determine whether of not there is evidence that the mean of X_1 is greater than the mean of X_2 by **more than** 140.

Carry out the test at the 10% significance level.

$\boxed{\text{significant}}$

$X_1 \sim N(\mu_1, 50^2)$ $n_1 = 40$ $\bar{x}_1 = 1752$

$X_2 \sim N(\mu_2, 20^2)$ $n_2 = 50$ $\bar{x}_2 = 1598$

$H_0 : \mu_1 = \mu_2 + 140$

$H_1 : \mu_1 > \mu_2 + 140$

TEST AT 10% SIGNIFICANCE

10%

$\Phi^{-1}(0.9) = 1.2816$

- $\text{Z-STAT} = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$
- $\text{Z-STAT} = \dfrac{(1752 - 1598) - 140}{\sqrt{\dfrac{50^2}{40} + \dfrac{20^2}{50}}}$
- $\text{Z-STAT} = 1.667$

AS $1.667 > 1.2816$ THERE IS SIGNIFICANT EVIDENCE THAT μ_1 IS EXCEEDING μ_2 BY MORE THAN 140. (REJECT H_0)

Question 9 (***+)

A group of 1200 soldiers completed an assault course, early in the morning.

A random sample of 60 soldiers was selected from the group of 1200. The time taken by each of these 60 soldiers to complete the assault course, x minutes, was recorded and the following information is known.

$$\sum_{i=1}^{60} x_i = 1350 \quad \text{and} \quad \sum_{i=1}^{60} x_i^{\,2} = 30685$$

a) Find unbiased estimates for the mean and variance of the time taken by the 1200 soldiers who completed the course early in the morning.

A group of 1500 soldiers completed the same assault course, late in the afternoon.

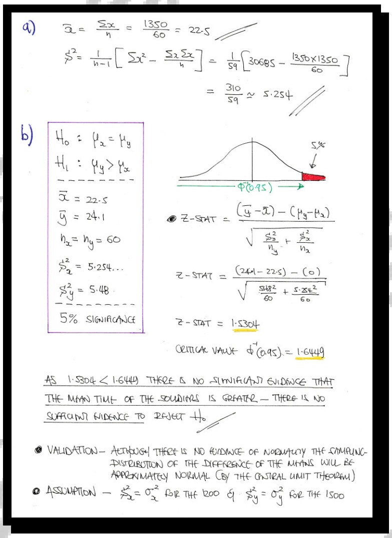
A random sample of 60 soldiers was selected from the group of 1500. The time taken by each of these 60 soldiers to complete the assault course, y minutes, was recorded and the following information is known.

$$\bar{y} = 24.1 \quad \text{and} \quad s_y^2 = 5.48$$

b) Test, at the 5% significance level, whether or not the mean time of the 1500 soldiers which completed the assault course in the afternoon is greater than that of the 1200 soldiers which completed the same course in the morning.

State the hypotheses clearly and any assumptions and validations made.

$\boxed{\bar{x} = 22.5}$, $\boxed{s_y^2 = \frac{310}{59} \approx 5.254}$, $\boxed{\text{not significant evidence, } 1.5304 < 1.6449}$



a) $\bar{x} = \frac{\Sigma x}{n} = \frac{1350}{60} = 22.5$

$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{\Sigma x \Sigma x}{n}\right] = \frac{1}{59}\left[30685 - \frac{1350 \times 1350}{60}\right]$

$= \frac{310}{59} \simeq 5.254$

b) $H_0 : \mu_x = \mu_y$

$H_1 : \mu_y > \mu_x$

$\bar{x} = 22.5$

$\bar{y} = 24.1$

$n_x = n_y = 60$

$s_x^2 = 5.254\ldots$

$s_y^2 = 5.48$

5% SIGNIFICANCE

5%

$\Phi^{-1}(0.95)$

• Z-STAT $= \dfrac{(\bar{y} - \bar{x}) - (\mu_y - \mu_x)}{\sqrt{\frac{s_y^2}{n_y} + \frac{s_x^2}{n_x}}}$

Z-STAT $= \dfrac{(24.1 - 22.5) - (0)}{\sqrt{\frac{5.48^2}{60} + \frac{5.254\ldots^2}{60}}}$

Z-STAT $= 1.5304$

CRITICAL VALUE $\Phi^{-1}(0.95) = 1.6449$

AS $1.5304 < 1.6449$ THERE IS NO SIGNIFICANT EVIDENCE THAT THE MEAN TIME OF THE SOLDIERS IS GREATER – THERE IS NO SUFFICIENT EVIDENCE TO REJECT H_0

• VALIDATION – ALTHOUGH THERE IS NO EVIDENCE OF NORMALITY THE SAMPLING DISTRIBUTION OF THE DIFFERENCE OF THE MEANS WILL BE APPROXIMATELY NORMAL (BY THE CENTRAL LIMIT THEOREM)

• ASSUMPTION – $s_x^2 = \sigma_x^2$ FOR THE 1200 & $s_y^2 = \sigma_y^2$ FOR THE 1500

Question 10 (****)

The continuous random variables X and Y are defined as

$$X \sim \mathrm{N}\left(\mu_x, 2^2\right) \quad \text{and} \quad Y \sim \mathrm{N}\left(\mu_y, 3^2\right).$$

The mean of a random sample of 7 observations from the population of X is denoted by $\bar{x}$ and the mean of a random sample of 9 observations from the population of Y is denoted by $\bar{y}$.

A test on the difference of the population means, at the 5% significance level, is to be carried out.

a) Stating your hypotheses clearly, determine the critical region for this test.

Give the answer in the form $|\bar{x}-\bar{y}| > k$, where k is a constant.

b) Determine the probability of a Type II error if $\mu_x - \mu_y = 0.9$

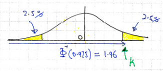
$\boxed{}$, $\boxed{k = \frac{7\sqrt{77}}{25} \approx 2.457}$, $\boxed{\approx 0.8892}$

a) IT IS GIVEN THAT

• $X \sim N(\mu_x, 2^2)$ • $Y \sim N(\mu_y, 3^2)$

$H_0 : \mu_x = \mu_y$
$H_1 : \mu_x \neq \mu_y$
5% SIGNIFICANCE
$\sigma_x = 2$ $\sigma_y = 3$
$n_x = 7$ $n_y = 9$

$$Z \text{ STAT} = \frac{\cancel{(\mu_x - \mu_y)} - (\bar{x} - \bar{y})}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$



LOOKING AT THE "TOP END" ONLY

EXPECTATION MEAN = 0
VARIANCE $= \frac{4}{7} + \frac{9}{9} = \frac{11}{7}$

$\Rightarrow P(\bar{x}-\bar{y} > k) = 0.025$
$\Rightarrow P(\bar{x}-\bar{y} < k) = 0.975$
$\Rightarrow P\left(z < \frac{k-0}{\sqrt{\frac{11}{7}}}\right) = 0.975$
$\Rightarrow \frac{k}{\sqrt{\frac{11}{7}}} = \Phi^{-1}(0.975)$
$\Rightarrow \frac{k}{\sqrt{\frac{11}{7}}} = 1.96$
$\Rightarrow k = \frac{7}{25}\sqrt{77}$
$\Rightarrow k \simeq 2.45699\ldots$
$\Rightarrow k \simeq 2.457$

$\therefore$ IF THE DIFFERENCE IN THE SAMPLE MEANS IS GREATER THAN 2.457 THE RESULT WILL BE SIGNIFICANT

$\therefore |\bar{x}-\bar{y}| > 2.457$

NOW THE PROBABILITY OF A TYPE II ERROR

NOW $\mu_x - \mu_y = 0.9$, VARIANCE IS $\frac{11}{7}$, AND WE WOULD NOT REJECT BETWEEN -2.457 & 2.457.

(TYPE II ERROR $\Rightarrow$ DO NOT REJECT H_0 WHEN H_0 IS FALSE.)

$$\begin{aligned}
P(|\bar{x}-\bar{y}| < k) &= P(-2.457 < \bar{x}-\bar{y} < 2.457) \\
&= P(\bar{x}-\bar{y} < 2.457) - P(\bar{x}-\bar{y} < -2.457) \\
&= P(\bar{x}-\bar{y} < 2.457) - \left[1 - P(\bar{x}-\bar{y} > -2.457)\right] \\
&= P(\bar{x}-\bar{y} < 2.457) + P(\bar{x}-\bar{y} > -2.457) - 1 \\
&= P\left(z < \frac{2.457-0.9}{\sqrt{11/7}}\right) + P\left(z > \frac{-2.457-0.9}{\sqrt{\frac{11}{7}}}\right) - 1 \\
&= \Phi(1.2421) + \Phi(-2.6780) - 1 \\
&= 0.8929 + 0.9963 - 1 \\
&= 0.8892
\end{aligned}$$

NORMAL DISTRIBUTION HYPOTHESIS TESTING

Difference of Means

Confidence Intervals

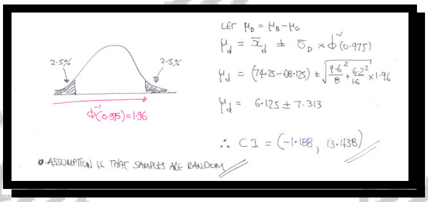
Question 1 (***)

The same Mathematics mock exam is given to the Year 11 students of a certain school year after year. It has been established over time that the standard deviation of the percentage marks is 9.6 for the boys and 6.2 for the girls. The percentage marks for both boys and girls are thought to be normally distributed.

The percentage marks of 8 boys in this year's mock exam had a mean of 74.25 while the percentage marks of 16 girls in the same test had a mean of 68.125.

Find a 95% confidence interval for the difference in the mean percentage mark between the boys and the girls, stating clearly any assumption made.

$$\boxed{\mu_B - \mu_G = (-1.188, 13.438)}$$



Question 2 (***)

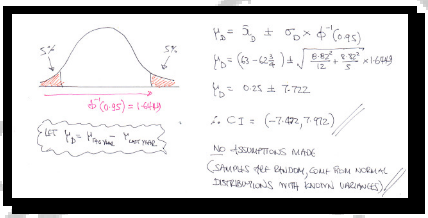
It has been established over time that the javelin throwing distances of an athlete are normally distributed with a standard deviation of 8.82 m.

Twelve throws of this athlete in the previous year had a mean of $62\frac{3}{4}$ m.

Five throws of the same athlete this year had a mean of 63 m.

Assuming the throwing distances considered for each year are random, find a 90% confidence interval for the difference in the mean throwing distances between the two years, stating clearly any assumption made.

$$\boxed{\mu_{\text{this year}} - \mu_{\text{last year}} = (-7.472, 7.722)}$$



Question 3 (***)

It has been established over time that the javelin throwing distances of an athlete have a standard deviation of 4.15 m.

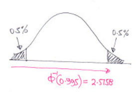
Forty throws of this athlete in the previous year had a mean of 60.14 m.

Thirty five throws of the same athlete this year had a mean of 59.95 m.

Assuming the throwing distances considered for each year are random, find a 99% confidence interval for the difference in the mean throwing distances between the two years, stating clearly any assumption made.

$$\boxed{\mu_{\text{this year}} - \mu_{\text{last year}} = (-2.771, 2.392)}$$

LET $\mu_D = \mu_{\text{THIS YEAR}} - \mu_{\text{LAST YEAR}}$



$\mu_D = \bar{x}_D \pm \sigma_{\bar{D}} \times \Phi^{-1}(0.995)$

$\mu_D = (59.95 - 60.14) \pm \sqrt{\frac{4.15^2}{40} + \frac{4.15^2}{30}} \times 2.5758$

$\mu_D = -0.19 \pm 2.582\ldots$

$\therefore$ C.I $= (-2.771, 2.392)$

- ASSUMPTIONS, RANDOM SAMPLES
- CONFIDENCE INTERVAL IS APPROXIMATE BY THE CENTRAL LIMIT THEOREM

Question 4 (***)

The continuous random variables X_1 and X_2 are assumed to have respective standard deviations

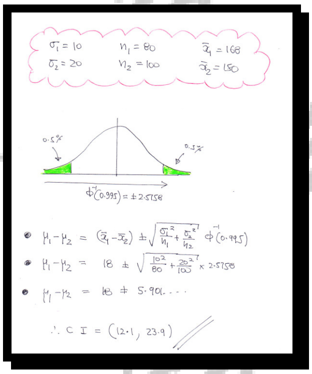
$$\sigma_1 = 10 \quad \text{and} \quad \sigma_2 = 20 .$$

A random sample of 80 observations of X_1 produced a sample mean of 168.

A random sample of 100 observations of X_2 produced a sample mean of 150.

Find a 99% confidence interval for the difference in the means between X_1 and X_2 .

$$\boxed{\mu_1 - \mu_2 = (12.1, 23.9)}$$



Question 5 (***+)

When the fat content of a 100 gram slice of cheesecake is measured, using a certain machine, the reading obtained in grams is a Normally distributed variable with mean the actual fat content and standard deviation of 1.1 grams.

The fat content of 16 slices, 8 from each of two varieties of cheesecake are recorded. All 16 slices have a mass of 100 grams.

The fat content of these slices is shown below.

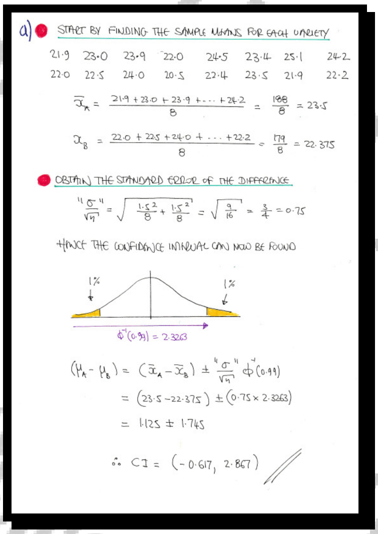
Variety A : 21.9, 23.0, 23.9, 22.0, 24.5, 23.4, 25.1, 24.2.

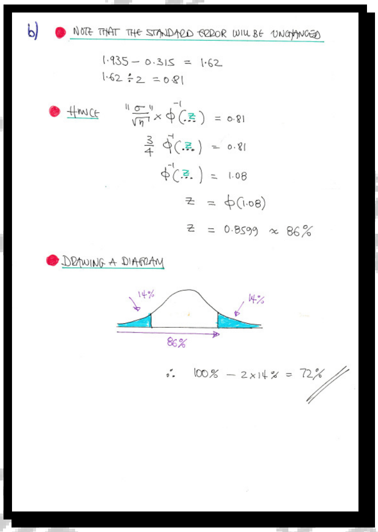
Variety B : 22.0, 22.5, 24.0, 20.5, 22.4, 23.5, 21.9, 22.2.

a) Calculate a 98% confidence interval for the difference between the mean fat content of a 100 gram cheesecake slice of variety A and variety B.

b) Determine the percentage confidence level if the confidence interval for the difference between the mean fat content of the two varieties is $[0.315, 1.935]$.

$\boxed{}$, $\boxed{(\mp 0.617, \pm 2.867)}$, $\boxed{72\%}$





Question 6 (***+)

A car manufacturer wants to compare the fuel consumption of two similar models of car they produce, model **A** and model **B**.

It is assumed that the fuel consumption for both models of these cars is a Normal variable with respective population mean in miles, μ_A and μ_B.

It is further assumed that the standard deviation for the fuel consumption of both models is 5 miles.

Sixteen cars were picked at random, **eight** from each model.

All sixteen cars were filled with exactly 3 gallons of fuel and were driven under lab conditions until they ran out of fuel.

The mileages achieved are summarized below.

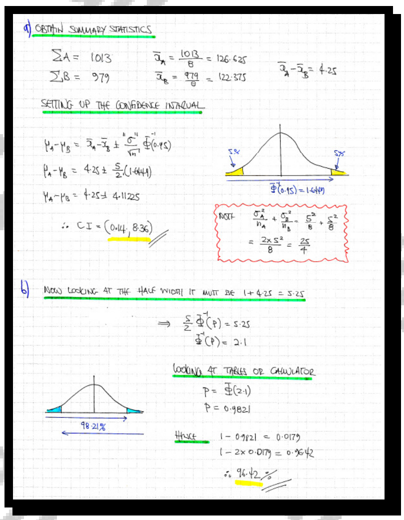
Model A : 126.0, 125.8, 128.2, 126.1, 123.9, 127.0, 124.6, 131.4

Model B : 122.3, 122.7, 118.9, 124.1, 122.6, 122.5, 124.0, 121.9

a) Determine a symmetrical 90% confidence interval for $\mu_A - \mu_B$.

b) Calculate the percentage significance of a symmetrical confidence interval for $\mu_A - \mu_B$ must have, if it is to contain -1.

$\boxed{}$, $\boxed{(0.14, 8.36)}$, $\boxed{96.42\%}$



a) OBTAIN SUMMARY STATISTICS

$\sum A = 1013$ $\quad \bar{x}_A = \frac{1013}{8} = 126.625$ $\quad \bar{x}_A - \bar{x}_B = 4.25$

$\sum B = 979$ $\quad \bar{x}_B = \frac{979}{8} = 122.375$

SETTING UP THE CONFIDENCE INTERVAL

$\mu_A - \mu_B = \bar{x}_A - \bar{x}_B \pm \frac{"\sigma"}{\sqrt{n}} \Phi^{-1}(0.95)$

$\mu_A - \mu_B = 4.25 \pm \frac{5}{2}(1.6449)$

$\mu_A - \mu_B = 4.25 \pm 4.11225$

$\therefore \text{CI} = (0.14, 8.36)$

5% 5%

$\Phi^{-1}(0.95) = 1.6449$

NOTE $\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{5^2}{8} + \frac{5^2}{8} = \frac{2 \times 5^2}{8} = \frac{25}{4}$

b) NOW LOOKING AT THE HALF WIDTH IT MUST BE $1 + 4.25 = 5.25$

$\Rightarrow \frac{5}{2}\Phi^{-1}(p) = 5.25$

$\Phi^{-1}(p) = 2.1$

LOOKING AT TABLES OR CALCULATOR

$p = \Phi(2.1)$

$p = 0.9821$

98.21%

HENCE $1 - 0.9821 = 0.0179$

$1 - 2 \times 0.0179 = 0.9642$

$\therefore 96.42\%$