STUDENT'S t – DISTRIBUTION INTRODUCTION

Question 1 (**)

The weights, in grams, of ten bags of popcorn are shown below

$$91,\ 101,\ 98,\ 98,\ 103,\ 97,\ 102,\ 105,\ 94,\ 90.$$

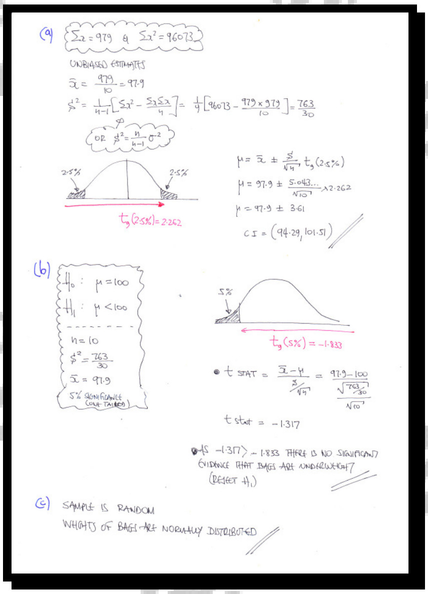
a) Find a 95% confidence interval for the mean weight of a bag of popcorn.

The seller of the popcorn claims that the mean weight of the bags is 100 grams.

b) Test at the 5% level of significance whether there is evidence that the bags of popcorn bought from this particular seller are underweight.

c) State clearly any assumptions made.

$\boxed{(94.29, 101.51)}$, $\boxed{\text{not significant, } \ -1.317 > -1.833}$


(a) $\sum x = 979$ & $\sum x^2 = 96073$

UNBIASED ESTIMATES

$\bar{x} = \frac{979}{10} = 97.9$

$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right] = \frac{1}{9}\left[96073 - \frac{979 \times 979}{10}\right] = \frac{763}{30}$

OR $s^2 = \frac{n}{n-1}\sigma^2$

2.5%   2.5%

$t_9(2.5\%) = 2.262$

$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_9(2.5\%)$

$\mu = 97.9 \pm \frac{5.043\ldots}{\sqrt{10}} \times 2.262$

$\mu = 97.9 \pm 3.61$

$CI = (94.29, 101.51)$

(b) $H_0 : \mu = 100$

$H_1 : \mu < 100$

$n = 10$

$s^2 = \frac{763}{30}$

$\bar{x} = 97.9$

5% SIGNIFICANCE (ONE TAILED)

5%

$t_9(5\%) = -1.833$

• t STAT $= \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{97.9 - 100}{\sqrt{763/30}/\sqrt{10}}$

t stat $= -1.317$

• AS $-1.317 > -1.833$ THERE IS NO SIGNIFICANT EVIDENCE THAT BAGS ARE UNDERWEIGHT (REJECT H_1)

(c) SAMPLE IS RANDOM

WEIGHTS OF BAGS ARE NORMALLY DISTRIBUTED

Question 2 (**)

It is claimed by a school that the typical waiting time to see a teacher in a parents' evening is 5 minutes.

In a recent parents' evening one of the parents recorded the waiting time, in minutes, for each of his son's 9 teachers. His results are shown below

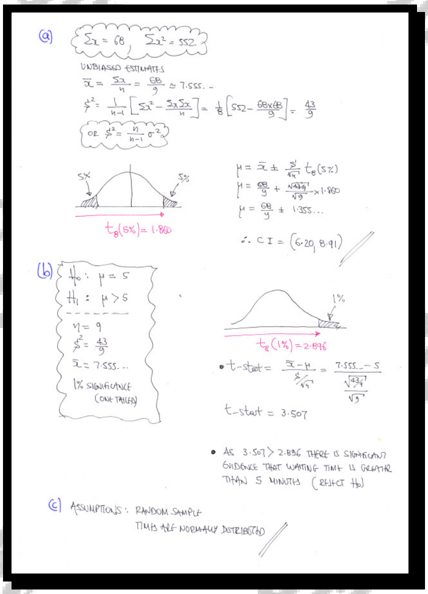
$$4,\ 12,\ 9,\ 7,\ 7,\ 6,\ 8,\ 7,\ 8.$$

a) Find a 90% confidence interval for the mean waiting time in a parents evening in that school.

b) Test at the 1% level of significance whether there is evidence that the mean waiting time in a parents evening in that school is higher than 5 minutes.

c) State clearly any assumptions made.

$\boxed{(6.20, 8.91)}$, $\boxed{\text{significant, } 3.507 > 2.896}$



(a) $\sum x = 68$, $\sum x^2 = 552$

UNBIASED ESTIMATES

$\bar{x} = \frac{\sum x}{n} = \frac{68}{9} \simeq 7.555\ldots$

$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right] = \frac{1}{8}\left[552 - \frac{68 \times 68}{9}\right] = \frac{43}{9}$

OR $s^2 = \frac{n}{n-1}\sigma^2$

5% 5%

$t_8(5\%) = 1.860$

$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_8(5\%)$

$\mu = \frac{68}{9} \pm \frac{\sqrt{43/9}}{\sqrt{9}} \times 1.860$

$\mu = \frac{68}{9} \pm 1.355\ldots$

$\therefore$ CI $= (6.20, 8.91)$

(b) $H_0: \mu = 5$

$H_1: \mu > 5$

$n = 9$

$s^2 = \frac{43}{9}$

$\bar{x} = 7.555\ldots$

1% SIGNIFICANCE (ONE TAILED)

1%

$t_8(1\%) = 2.896$

• t-stat $= \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{7.555\ldots - 5}{\frac{\sqrt{43/9}}{\sqrt{9}}}$

t-stat $= 3.507$

• AS $3.507 > 2.896$ THERE IS SIGNIFICANT EVIDENCE THAT WAITING TIME IS GREATER THAN 5 MINUTES (REJECT H_0)

(c) ASSUMPTIONS: RANDOM SAMPLE
TIMES ARE NORMALLY DISTRIBUTED

Question 3 (**)

Mark is a shot putter.

The distances, in metres, for 8 of his throws are shown below

$$15.82\,,\ 16.07\,,\ 15.37\,,\ 19.01\,,\ 17.52\,,\ 14.98\,,\ 15.64\,,\ 16.28\,.$$

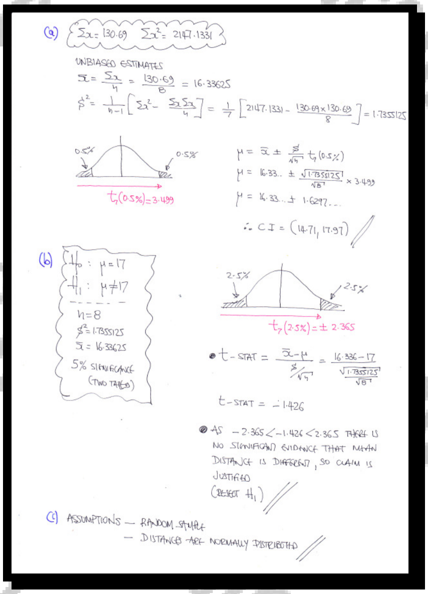
a) Find a 99% confidence interval for the mean distance thrown.

Mark claims that his mean throwing distance is 17 metres.

b) Test at the 5% level of significance whether Mark's claim is justified.

c) State clearly any assumptions made.

$\boxed{(14.71, 17.97)}$, $\boxed{\text{not significant,}\quad -2.365 < -1.426 < 2.365}$



(a) $\sum x = 130.69 \quad \sum x^2 = 2147.1331$

UNBIASED ESTIMATES

$\bar{x} = \frac{\sum x}{n} = \frac{130.69}{8} = 16.33625$

$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right] = \frac{1}{7}\left[2147.1331 - \frac{130.69 \times 130.69}{8}\right] = 1.7355125$

0.5%, 0.5%

$t_7(0.5\%) = 3.499$

$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_7(0.5\%)$

$\mu = 16.33.. \pm \frac{\sqrt{1.7355125}}{\sqrt{8}} \times 3.499$

$\mu = 16.33... \pm 1.6297...$

$\therefore CI = (14.71, 17.97)$

(b) $H_0 : \mu = 17$

$H_1 : \mu \neq 17$

$n = 8$

$s^2 = 1.7355125$

$\bar{x} = 16.33625$

5% SIGNIFICANCE (TWO TAILED)

2.5%, 2.5%

$t_7(2.5\%) = \pm 2.365$

- $t\text{-STAT} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{16.336 - 17}{\frac{\sqrt{1.7355125}}{\sqrt{8}}}$

$t\text{-STAT} = -1.426$

- AS $-2.365 < -1.426 < 2.365$ THERE IS NO SIGNIFICANT EVIDENCE THAT MEAN DISTANCE IS DIFFERENT, SO CLAIM IS JUSTIFIED (REJECT H_1)

(c) ASSUMPTIONS — RANDOM SAMPLE
— DISTANCES ARE NORMALLY DISTRIBUTED

Question 4 (**)

Joe disagrees with the statement made on the label of a packet of crisps, which states that the net content is 40 grams.

He weighs the content, x grams, of 10 randomly chosen packets of these crisps.

His results are summarized below.

$$\sum_{10} x = 387.5 \quad \text{and} \quad \sum_{10} x^2 = 15096.25$$

Test, at the 5% level of significance, whether there is evidence to support Joe's belief.

$\boxed{}$, $\boxed{\text{not significant, } -2.262 < -1.3206... < 2.262}$

SETTING HYPOTHESES

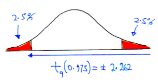
$H_0: \mu = 40$

$H_1: \mu \neq 40$ WHERE μ IS THE MEAN AMOUNT OF ALL CRISP PACKETS (POPULATION MEAN)

OBTAINING SAMPLE STATISTICS

- $\bar{x} = \frac{\sum x}{n} = \frac{387.5}{10} = 38.75$
- $s = \sqrt{\frac{1}{n-1}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right]} = \sqrt{\frac{1}{9}\left[15096.25 - \frac{387.5^2}{10}\right]} = 2.99304...$

LOOKING AT A t-DISTRIBUTION DIAGRAM



$$t\text{-stat} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{38.75 - 40}{\frac{2.99304...}{\sqrt{10}}} = -1.3206...$$

AS $-2.262 < -1.3206 < 2.262$, THERE IS NO SIGNIFICANT EVIDENCE THAT THE MEAN WEIGHT IS NOT 40

WE DO NOT HAVE SUFFICIENT EVIDENCE TO REJECT H_0 //

Question 5 (**+)

An industrial wood shredder must be rested for a minimum period of 20 minutes after a set usage time. The times of these rest brakes are thought to be modelled by a Normal variable T, with mean μ and standard deviation σ.

16 random of values of T are summarized below.

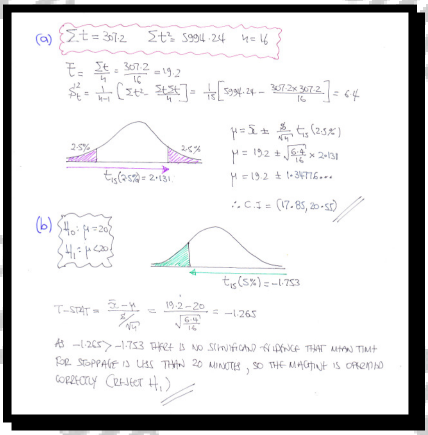
$$\sum t = 307.2, \quad \sum t^2 = 5994.24.$$

a) Calculate a 95% confidence interval for μ.

For the shredder to be operated correctly, μ should not be less than 20 minutes.

b) Test at the 5% level of significance whether the shredder is operated correctly.

$\boxed{(17.85, 20.55)}$, $\boxed{\text{not significant, } -1.265 > -1.753, \therefore \text{correctly operated}}$


(a) $\sum t = 307.2 \quad \sum t^2 = 5994.24 \quad n = 16$

$\bar{t} = \frac{\sum t}{n} = \frac{307.2}{16} = 19.2$

$s_t^2 = \frac{1}{n-1}\left[\sum t^2 - \frac{\sum t \sum t}{n}\right] = \frac{1}{15}\left[5994.24 - \frac{307.2 \times 307.2}{16}\right] = 6.4$

2.5% 2.5%

$t_{15}(2.5\%) = 2.131$

$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} t_{15}(2.5\%)$

$\mu = 19.2 \pm \sqrt{\frac{6.4}{16}} \times 2.131$

$\mu = 19.2 \pm 1.34776\ldots$

$\therefore$ C.I $= (17.85, 20.55)$

(b) $H_0: \mu = 20$

$H_1: \mu < 20$

$t_{15}(5\%) = -1.753$

T-STAT $= \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{19.2 - 20}{\sqrt{\frac{6.4}{16}}} = -1.265$

AS $-1.265 > -1.753$ THERE IS NO SIGNIFICANT EVIDENCE THAT MEAN TIME FOR STOPPAGE IS LESS THAN 20 MINUTES, SO THE MACHINE IS OPERATING CORRECTLY (REJECT H_1)

Question 6 (+)**

In the main road, passing through the village of Cockfosters, the speed limit is 30 mph, however a speed measuring device indicated that the mean driving speed through the village was 37.2 mph.

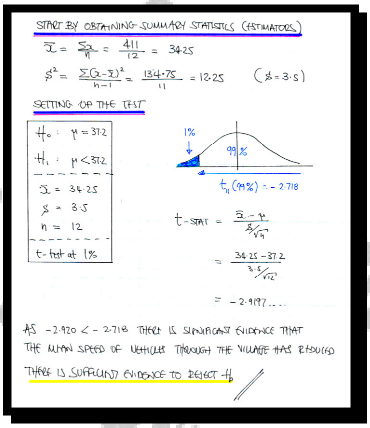
In an attempt to reduce these excessive driving speeds a carton real size model of a police officer was placed by the side of the road, in the approach to the village.

The speeds, X mph, of a random sample of 12 vehicles were subsequently measured yielding the following summary statistics.

$$\sum_{i=1}^{12} x_i = 411 \qquad \sum_{i=1}^{12} (x_i - \bar{x})^2 = 134.75$$

Stating your hypotheses clearly, carry out a suitable test, at the 1% level of significance, to determine whether there has been a reduction in the mean speed of the vehicles passing through the village, after the placement of the police officer model.

$\boxed{\quad}$, $\boxed{\text{significant, } -2.920 < -2.718}$



Question 7 (***)

A machine packs peas into bags with mean weight μ grams.

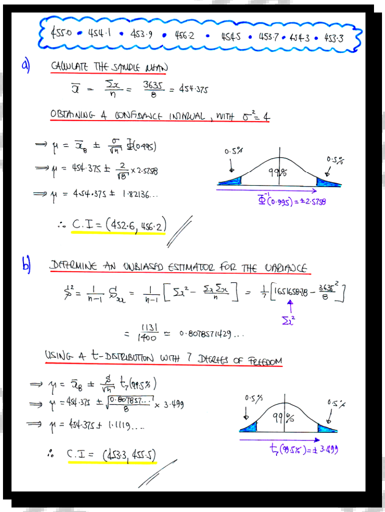
The weights, in grams, of a random sample of 8 bags is shown below.

$$455.0,\ 454.1,\ 453.9,\ 456.2,\ 454.5,\ 453.7,\ 454.3,\ 453.3.$$

a) Assuming the standard deviation of the weights of all bags packed by the machine is 2 grams, find a 99% confidence interval for μ.

b) Given instead that the standard deviation of the weights of all bags packed by the machine is unknown, find another 99% confidence interval for μ.

$\boxed{}$, $\boxed{(452.6, 456.2)}$, $\boxed{(453.3, 455.5)}$


455.0 • 454.1 • 453.9 • 456.2 • 454.5 • 453.7 • 454.3 • 453.3

a) CALCULATE THE SAMPLE MEAN

$$\bar{x} = \frac{\Sigma x}{n} = \frac{3635}{8} = 454.375$$

OBTAINING A CONFIDENCE INTERVAL, WITH $\sigma^2 = 4$

$$\Rightarrow \mu = \bar{x}_8 \pm \frac{\sigma}{\sqrt{n}} \Phi(0.995)$$

$$\Rightarrow \mu = 454.375 \pm \frac{2}{\sqrt{8}} \times 2.5758$$

$$\Rightarrow \mu = 454.375 \pm 1.82136\ldots$$

$$\therefore \text{C.I} = (452.6, 456.2)$$

b) DETERMINE AN UNBIASED ESTIMATOR FOR THE VARIANCE

$$s^2 = \frac{1}{n-1} S_{xx} = \frac{1}{n-1}\left[\Sigma x^2 - \frac{\Sigma x \Sigma x}{n}\right] = \frac{1}{7}\left[\underbrace{1651658.78}_{\Sigma x^2} - \frac{3635^2}{8}\right]$$

$$= \frac{1131}{1400} \simeq 0.8078571429\ldots$$

USING A t-DISTRIBUTION WITH 7 DEGREES OF FREEDOM

$$\Rightarrow \mu = \bar{x}_8 \pm \frac{s}{\sqrt{n}} t_7(99.5\%)$$

$$\Rightarrow \mu = 454.375 \pm \sqrt{\frac{0.807857\ldots}{8}} \times 3.499$$

$$\Rightarrow \mu = 454.375 \pm 1.1119\ldots$$

$$\therefore \text{C.I} = (453.3, 455.5)$$

Question 8 (***)

A coffee machine, placed in the waiting room of a garage, dispenses coffee into cups. The volume of the coffee in a cup is Normally distributed with mean 250 ml.

The manager of the garage claims that the mean volume of coffee in a cup is no longer 250 ml due to the age of the machine.

He records the volume, x ml, of 10 randomly selected cups, producing the following summary statistics.

$$\sum_{i=1}^{10} x_i = 2390 \quad \text{and} \quad \sum_{i=1}^{10} x_i^2 = 574495 .$$

a) Test, at the 5% significance level, the garage manager's claim.

b) Carry another test, at the 5% significance level, if instead the garage manager claimed that the mean amount of coffee dispensed by the machine in every cup is less than 250 ml.

$\boxed{\text{not significant evidence, } -1.821 > -2.262}$, $\boxed{\text{not significant evidence, } -1.821 > -1.833}$

a) STARTING BY OBTAINING THE SAMPLE MEAN & STANDARD DEVIATION

$$\bar{x} = \frac{\sum x}{n} = \frac{2390}{10} = 239$$

$$s = \sqrt{\frac{1}{9}\left[\sum x^2 - \frac{\sum x \sum x}{n}\right]} = \sqrt{\frac{1}{9}\left[574495 - \frac{2390 \times 2390}{10}\right]} = \sqrt{365}$$

SETTING THE HYPOTHESES USING THE t DISTRIBUTION

$H_0 : \mu = 250$

$H_1 : \mu \neq 250$

$\bar{x} = 239$

$n = 10$

$s = \sqrt{365}$

TWO TAILED t TEST AT 5%

$t_9(0.975) = -2.262$

$$t\text{-STAT} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{239 - 250}{\frac{\sqrt{365}}{\sqrt{10}}} = -1.821$$

AS $-1.821 > -2.262$ THERE IS NO SIGNIFICANT EVIDENCE TO SUPPORT THE MANAGER'S CLAIM – WE DO NOT HAVE SUFFICIENT EVIDENCE TO REJECT H_0

b) ON THIS OCCASION THE HYPOTHESES WOULD BE

$H_0 : \mu = 250$

$H_1 : \mu < 250$

- THE CRITICAL VALUE NOW WILL BE -1.833.
- THE t-STAT WILL BE UNCHANGED AT -1.821

AS $-1.821 > -1.833$ THERE IS STILL NO SIGNIFICANT EVIDENCE TO SUPPORT THE MANAGER'S CLAIM – REJECT H_1

STUDENT'S t – DISTRIBUTION

TWO SAMPLE t–TEST

Question 1 (**+)

A Mathematics exam was given to the Year 8 pupils of a certain school.

The percentage marks for some of the boys and some of the girls that sat this exam are given below.

Boys: 67 , 72 , 56, 91, 55, 68, 55, 45 , 80 .

Girls: 58, 97 , 65, 69 , 57 , 57 , 77 , 69 .

The Head of Maths thinks that the Year 8 boys have a different mean mark than the Year 8 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.

$\boxed{}$, $\boxed{\text{not significant, } -2.131 < -0.470 < 2.131}$

OBTAIN SUMMARY STATISTICS FROM THE CALCULATOR

- $\sum x_B = 589$ • $\sum x_B^2 = 40189$ • $n_B = 9$
- $\sum x_G = 549$ • $\sum x_G^2 = 38947$ • $n_G = 8$

SETTING HYPOTHESES

$H_0 : \mu_B = \mu_G$

$H_1 : \mu_B = \mu_G$

WHERE μ REPRESENTS THE POPULATION MEANS (TWO TAILED AT 5% SIGNIFICANCE, IE TESTING AT 2.5% IN EACH TAIL)

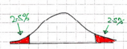
CALCULATING AUXILIARIES

- $\bar{x}_B = \dfrac{\sum x_B}{n_B} = \dfrac{589}{9} = 65.444\ldots$
- $\bar{x}_G = \dfrac{\sum x_G}{n_G} = \dfrac{549}{8} = 68.625$
- $s_B^2 = \dfrac{1}{n_B - 1}\left[\sum x_B^2 - \dfrac{\sum x_B \sum x_B}{n_B}\right] = \dfrac{1}{8}\left[40189 - \dfrac{589 \times 589}{9}\right] = \dfrac{3695}{18}$
- $s_G^2 = \dfrac{1}{n_G - 1}\left[\sum x_G^2 - \dfrac{\sum x_G \sum x_G}{n}\right] = \dfrac{1}{7}\left[38947 - \dfrac{549 \times 549}{8}\right] = \dfrac{10175}{56}$

NEXT THE POOLED ESTIMATE OF THE VARIANCE

$$s_p^2 = \frac{(n_B - 1)s_B^2 + (n_G - 1)s_G^2}{n_B + n_G - 2} = \frac{8 \times \frac{3695}{18} + 7 \times \frac{10175}{56}}{9 + 8 - 2} = \frac{41963}{216}$$

USING A t DISTRIBUTION WITH 15 DEGREES OF FREEDOM



(5% TWO TAILED)

$t_{15}(2.5\%) = \pm 2.131$

$$t\text{-STATISTIC} = \frac{(\bar{x}_B - \bar{x}_G) - (\mu_B - \mu_G)}{s_p\sqrt{\frac{1}{n_B} + \frac{1}{n_G}}}$$

$$= \frac{(65.444\ldots - 68.625) - (0)}{\sqrt{\frac{41963}{216}}\sqrt{\frac{1}{9} + \frac{1}{8}}}$$

$$= -0.46961\ldots$$

AS $-2.131 < -0.470 < 2.131$ THERE IS NO SIGNIFICANT EVIDENCE THAT THE POPULATION MEAN OF THE BOYS IS DIFFERENT TO THAT OF THE GIRLS — INSUFFICIENT EVIDENCE TO REJECT H_0 //

ASSUMPTIONS MADE

- SAMPLES ARE RANDOM
- SAMPLES COME FROM NORMAL DISTRIBUTIONS
- POPULATION VARIANCES ARE THE SAME FOR BOYS & GIRLS

Question 2 (+)**

A decathlete feels that his throwing distances have a higher mean this season than the previous one.

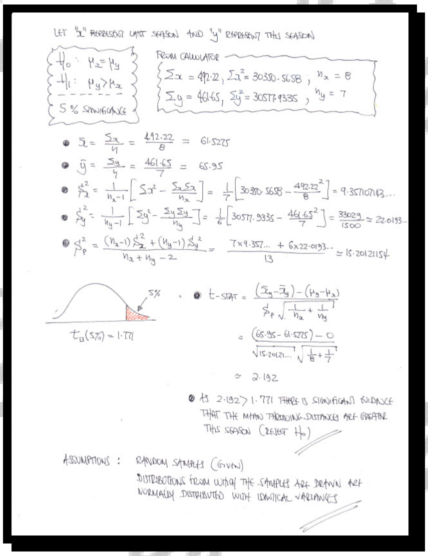
Two samples for his throwing distances, in m, for the last two seasons are shown in the table below.

Last season: 61.37, 57.29, 66.56, 60.91, 59.95, 61.10, 59.59, 65.45.

This season: 58.99, 66.97, 70.55, 69.02, 59.54, 68.22, 68.36.

Assuming the throwing distances considered for each season are random, test at the 5% level of significance, whether the mean throwing distances of the decathlete have improved this season. State your hypotheses clearly, stating any additional assumptions made.

$\boxed{\text{significant, } 2.192 > 1.771}$


LET "x" REPRESENT LAST SEASON AND "y" REPRESENT THIS SEASON

$H_0 : \mu_x = \mu_y$
$H_1 : \mu_y > \mu_x$
5% SIGNIFICANCE

FROM CALCULATOR
$\Sigma x = 492.22,\ \Sigma x^2 = 30350.5658,\ n_x = 8$
$\Sigma y = 461.65,\ \Sigma y^2 = 30577.9335,\ n_y = 7$

- $\bar{x} = \frac{\Sigma x}{n} = \frac{492.22}{8} = 61.5275$
- $\bar{y} = \frac{\Sigma y}{n} = \frac{461.65}{7} = 65.95$
- $s_x^2 = \frac{1}{n_x - 1}\left[\Sigma x^2 - \frac{\Sigma x \Sigma x}{n_x}\right] = \frac{1}{7}\left[30350.5658 - \frac{492.22^2}{8}\right] = 9.357107143\ldots$
- $s_y^2 = \frac{1}{n_y - 1}\left[\Sigma y^2 - \frac{\Sigma y \Sigma y}{n_y}\right] = \frac{1}{6}\left[30577.9335 - \frac{461.65^2}{7}\right] = \frac{33029}{1500} \simeq 22.0193\ldots$
- $s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{7 \times 9.357\ldots + 6 \times 22.0193\ldots}{13} \simeq 15.20121154$

5%

$t_{13}(5\%) = 1.771$

- $t\text{-STAT} = \frac{(\bar{x}_y - \bar{x}_y) - (\mu_y - \mu_x)}{s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} = \frac{(65.95 - 61.5275) - 0}{\sqrt{15.20121\ldots}\sqrt{\frac{1}{8} + \frac{1}{7}}} \simeq 2.192$
- AS $2.192 > 1.771$ THERE IS SIGNIFICANT EVIDENCE THAT THE MEAN THROWING DISTANCES ARE GREATER THIS SEASON (REJECT H_0)

ASSUMPTIONS : RANDOM SAMPLES (GIVEN)
DISTRIBUTIONS FROM WHICH THE SAMPLES ARE DRAWN ARE NORMALLY DISTRIBUTED WITH IDENTICAL VARIANCES

Question 3 (***)

A Mathematics exam was given to the Year 11 pupils of a certain school.

The marks for some of the boys and some of the girls that sat this exam are given in the table below.

Boys: 67 , 62 , 56, 91, 55, 68, 55, 45 , 80 .

Girls: 68, 87 , 55, 69 , 79 , 70 , 77 , 69 .

The Head of Maths thinks that the mean of the Year 11 girls is more than 5 marks higher than the mean of the Year 11 boys.

Test, at the 10% level of significance, whether the claim of the Head of Maths is justified. State your hypotheses clearly, stating any additional assumptions made.

[illegible] , $\boxed{\text{not significant, } 0.408 < 1.341}$

OBTAIN SUMMARY STATISTICS USING A CALCULATOR

- $\sum x_B = 579$ • $\sum x_B^2 = 38849$ • $n_B = 9$
- $\sum x_G = 574$ • $\sum x_G^2 = 41810$ • $n_G = 8$

SETTING HYPOTHESES

$H_0: \mu_G = \mu_B + 5$

$H_1: \mu_G > \mu_B + 5$, WHERE μ DENOTES THE POPULATION MEANS

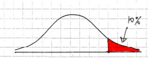
CALCULATING AUXILIARIES

- $\bar{x}_B = \frac{\sum x_B}{n_B} = \frac{579}{9} = 64.333\ldots$
- $\bar{x}_G = \frac{\sum x_G}{n_G} = \frac{574}{8} = 71.75$
- $s_B^2 = \frac{1}{n_B - 1}\left[\sum x_B^2 - \frac{\sum x_B \sum x_B}{n_B}\right] = \frac{1}{8}\left[38849 - \frac{579 \times 579}{9}\right] = 200$
- $s_G^2 = \frac{1}{n_G - 1}\left[\sum x_G^2 - \frac{\sum x_G \sum x_B}{n_G}\right] = \frac{1}{7}\left[41810 - \frac{574 \times 574}{8}\right] = \frac{1251}{14}$

NEXT THE POOLED ESTIMATE OF THE VARIANCE

$$s_p^2 = \frac{(n_B-1)s_B^2 + (n_G-1)s_G^2}{n_B + n_G - 2} = \frac{8 \times 200 + 7 \times \frac{1251}{14}}{9 + 8 - 2} = \frac{4451}{30}$$

USING A t DISTRIBUTION WITH 15 DEGREES OF FREEDOM



(10% ONE TAILED)

$t_{15}(10\%) = 1.341$

$$t\text{-STATISTIC} = \frac{(\bar{x}_G - \bar{x}_B) - (\mu_G - \mu_B)}{s_p\sqrt{\frac{1}{n_B} + \frac{1}{n_G}}} = \frac{(71.75 - 64.333\ldots) - (5)}{\sqrt{\frac{4451}{30}}\sqrt{\frac{1}{9} + \frac{1}{8}}} = 0.408\ldots$$

AS $0.408 < 1.341$, THERE IS NO SIGNIFICANT EVIDENCE THAT THE POPULATION MEAN MARK OF THE GIRLS EXCEED THAT OF THE BOYS BY 5 MARKS – INSUFFICIENT EVIDENCE TO REJECT H_0

ASSUMPTIONS MADE

THE SAMPLES ARE RANDOM, TAKEN FROM NORMAL DISTRIBUTIONS WITH IDENTICAL VARIANCES

Question 4 (***)

A factory produces batteries which are claimed to have a mean lifetime of 40 hours. The mean lifetime of a battery, in hours, is assumed to be a Normal variable, and it is denoted by X. The factory uses two identical machines for the battery manufacture.

The manager of the factory tests a random sample of 10 batteries produced by machine A, and the results are summarized below.

$$\sum x_A = 378, \quad \sum x^2{}_A = 14598.$$

a) Carry out a one tail test, at the 10% level of significance, to check whether the claim of the factory is justified.

The manager of the factory next tests a random sample of 12 batteries produced by machine B, and these results are summarized below.

$$\sum x_B = 495, \quad \sum x^2{}_B = 20988.$$

b) Test, at the 10% level of significance, whether the mean battery lifetime differs between the two machines, stating any additional assumptions made.

c) Make a criticism of one of the assumptions made in part **(b)**

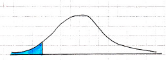
$\boxed{}$, $\boxed{\text{not significant, } -1.186 > -1.383}$, $\boxed{\text{not significant, } -1.725 < 1.216 < 1.725}$

a) OBTAIN ESTIMATORS FROM THE SAMPLE

- $\bar{x}_A = \frac{\sum x_A}{n_A} = \frac{378}{10} = 37.8$
- $s_A^2 = \frac{1}{n_A - 1}\left[\sum x_A^2 - \frac{\sum x_A \sum x_A}{n_A}\right] = \frac{1}{9}\left[14598 - \frac{378 \times 378}{10}\right] = \frac{172}{5} = 34.4$

SETTING HYPOTHESES

$H_0 : \mu_A = 40$
$H_1 : \mu_A < 40$
$\bar{x}_A = 37.8$
$s_A^2 = 34.4$
$n_A = 10$
10% ONE TAILED



$t_9(10\%) = -1.383$

t-STATISTIC $= \frac{\bar{x}_A - \mu_A}{s_A / \sqrt{n_A}} = \frac{37.8 - 40}{\sqrt{\frac{34.4}{10}}} = -1.186$

AS $-1.186 > -1.383$ THERE IS NO SIGNIFICANT THAT THE MEAN LIFETIME (FOR THE POPULATION) IS LESS THAN 40 HOURS — INSUFFICIENT EVIDENCE TO REJECT H_0

b) ASSUMING THAT THE POPULATION VARIANCES ARE THE SAME

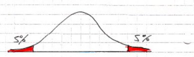
- $\bar{x}_B = \frac{\sum x_B}{n_B} = \frac{495}{12} = 41.25$
- $s_B^2 = \frac{1}{n_B - 1}\left[\sum x_B^2 - \frac{\sum x_B \sum x_B}{n_B}\right] = \frac{1}{11}\left[20988 - \frac{495 \times 495}{12}\right] = 51.75$

ESTIMATE THE COMMON VARIANCE (POOLED ESTIMATE s_p^2)

$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = \frac{(9 \times 34.4) + (11 \times 51.75)}{10 + 12 - 2} = 43.9425$

SETTING HYPOTHESES

$H_0 : \mu_A = \mu_B$
$H_1 : \mu_A \neq \mu_B$
$\bar{x}_A = 37.8$, $\bar{x}_B = 41.25$
$n_A = 10$, $n_B = 12$
$s_p^2 = 43.9425$



$t_{20}(5\%) = \pm 1.725$

- t-STATISTIC $= \frac{(\bar{x}_B - \bar{x}_A) - \cancel{(\mu_B - \mu_A)}}{\sqrt{s_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$
- t-STATISTIC $= \frac{41.25 - 37.8}{\sqrt{43.9425\left(\frac{1}{10} + \frac{1}{12}\right)}}$
- t-STAT $= 1.21550\ldots$

AS $-1.725 < 1.216 < 1.725$, THERE IS NO SIGNIFICANT EVIDENCE THAT THE MEAN LIFETIMES ARE DIFFERENT IN THE TWO MACHINES INSUFFICIENT EVIDENCE TO REJECT H_0

c) VARIANCES ARE UNLIKELY TO BE IDENTICAL, AS REQUIRED BY THIS TYPE OF TESTING, AS THE VARIANCE ESTIMATORS ARE VERY DIFFERENT, I.E. $s_A^2 = 34.4$ & $s_B^2 = 51.75$

STUDENT'S t – DISTRIBUTION

Difference of Means Confidence Interval

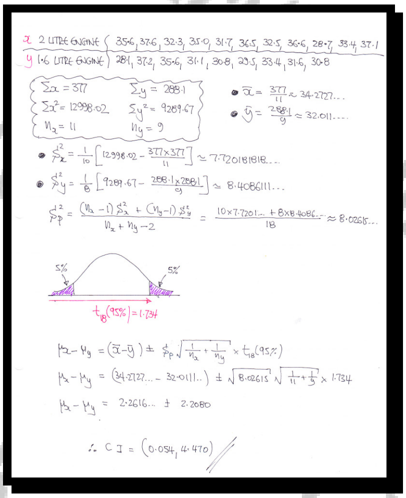
Question 1 (+)**

The table below shows the fuel consumption in mpg for a random sample of 11 cars with a 2 litre engine and a random sample of 9 cars with a 1.6 litre engine.

2 litre engine	35.6	37.6	32.3	35.0	31.7	36.5	32.5	36.6	28.7	33.4	37.1
1.6 litre engine	28.1	37.2	35.6	31.1	30.8	29.5	33.4	31.6	30.8		

Assuming that these samples come from populations with equal variance, find a 90% confidence interval for the difference in mean fuel consumption between cars with a 2 litre engine and cars with a 1.6 litre engine.

$\boxed{(0.054, 4.470)}$



x 2 LITRE ENGINE: 35.6, 37.6, 32.3, 35.0, 31.7, 36.5, 32.5, 36.6, 28.7, 33.4, 37.1

y 1.6 LITRE ENGINE: 28.1, 37.2, 35.6, 31.1, 30.8, 29.5, 33.4, 31.6, 30.8

$\sum x = 377 \qquad \sum y = 288.1$

$\sum x^2 = 12998.02 \qquad \sum y^2 = 9289.67$

$n_x = 11 \qquad n_y = 9$

- $\bar{x} = \frac{377}{11} \approx 34.2727...$
- $\bar{y} = \frac{288.1}{9} \approx 32.011...$

- $s_x^2 = \frac{1}{10}\left[12998.02 - \frac{377 \times 377}{11}\right] \simeq 7.720181818...$
- $s_y^2 = \frac{1}{8}\left[9289.67 - \frac{288.1 \times 288.1}{9}\right] \simeq 8.4086111...$

$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{10 \times 7.7201... + 8 \times 8.4086...}{18} \approx 8.02615...$

$\mu_x - \mu_y = (\bar{x} - \bar{y}) \pm s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}} \times t_{18}(95\%)$

$\mu_x - \mu_y = (34.2727... - 32.0111..) \pm \sqrt{8.02615} \sqrt{\frac{1}{11} + \frac{1}{9}} \times 1.734$

$\mu_x - \mu_y = 2.2616... \pm 2.2080$

$\therefore \ CI = (0.054, 4.470)$

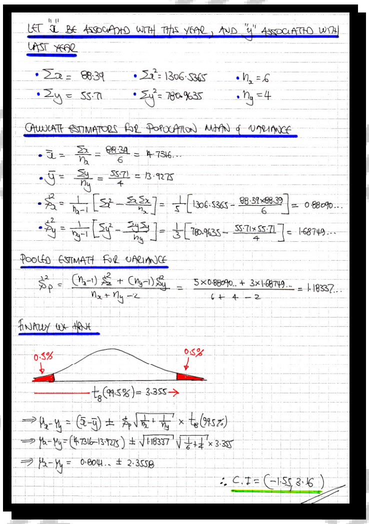
Question 2 (**+)

The table below shows the lengths jumped by a triple jumper for a random sample of 6 attempts this year and the lengths jumped by the same triple jumper for a random sample of 4 attempts last year.

Jumps this Year (m)	14.48	15.08	13.88	13.50	15.68	15.77
Jumps last Year (m)	15.45	12.32	13.69	14.25		

Assuming that the variances in the lengths of the triple jumper in both years are identical, find a 99% confidence interval for the difference in the mean length of the jumps of the athlete between this year and last year.

$\boxed{\text{S2J}}$, $\boxed{(-1.55, 3.16)}$



Question 3 (***)

The effectiveness of two teaching systems for learning Spanish, A and B, is to be compared using 9 randomly chosen suitable adults.

5 of these adults are allocated system A and 4 are allocated system B.

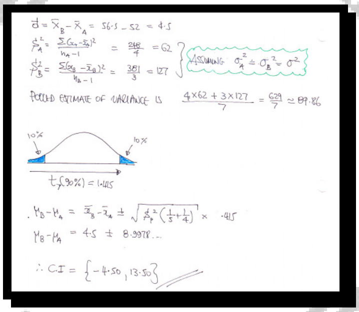
After a suitable testing period, all 9 are given the same test and information about their scores, x_A and x_B, is given below.

$$\bar{x}_A = 52, \quad \bar{x}_B = 56.5, \quad \sum (x_A - \bar{x}_A)^2 = 248.0, \quad \sum (x_B - \bar{x}_B)^2 = 381.0.$$

It is assumed that X_A and X_B follow normal distributions with identical variance.

Determine an 80% confidence interval for the difference in the population scores using the two methods.

$\boxed{(-4.50, 13.50) \text{ or } (-13.50, 4.50)}$



$\bar{d} = \bar{x}_B - \bar{x}_A = 56.5 - 52 = 4.5$

$s_A^2 = \frac{\sum(x_A - \bar{x}_A)^2}{n_A - 1} = \frac{248}{4} = 62$

$s_B^2 = \frac{\sum(x_B - \bar{x}_B)^2}{n_B - 1} = \frac{381}{3} = 127$

ASSUMING $\sigma_A^2 = \sigma_B^2 = \sigma^2$

POOLED ESTIMATE OF VARIANCE IS $\frac{4 \times 62 + 3 \times 127}{7} = \frac{629}{7} \simeq 89.86$

10%

10%

$t_7(90\%) = 1.415$

$\mu_B - \mu_A = \bar{x}_B - \bar{x}_A \pm \sqrt{s_p^2\left(\frac{1}{5} + \frac{1}{4}\right)} \times 1.415$

$\mu_B - \mu_A = 4.5 \pm 8.9978\ldots$

$\therefore \text{C.I} = \{-4.50, 13.50\}$

STUDENT'S t – DISTRIBUTION

PAIRED t–TEST

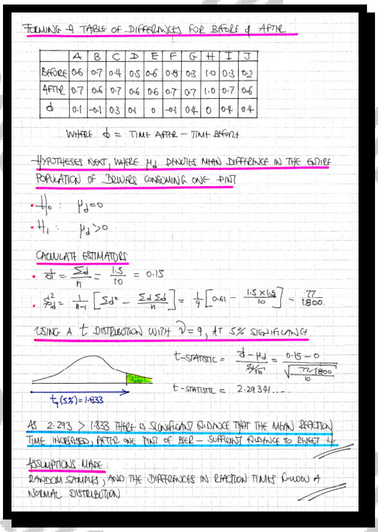
Question 1 (**+)

Research is conducted to investigate the effect of consumption of small amounts of alcohol in driving, by measuring the reaction times of 10 drivers before and after consuming one pint of beer. The results are summarised below.

Driver	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Reaction Time Before (sec)	0.6	0.7	0.4	0.5	0.6	0.8	0.3	1.0	0.3	0.2
Reaction Time After (sec)	0.7	0.6	0.7	0.6	0.6	0.7	0.7	1.0	0.7	0.6

Test, at the 5% level of significance, whether the consumption of one pint of beer appears to increase the reaction time of drivers, stating any assumptions used.

$\boxed{}$, $\boxed{\text{significant, } 2.293 > 1.833}$


FORMING A TABLE OF DIFFERENCES FOR BEFORE & AFTER

	A	B	C	D	E	F	G	H	I	J
BEFORE	0.6	0.7	0.4	0.5	0.6	0.8	0.3	1.0	0.3	0.2
AFTER	0.7	0.6	0.7	0.6	0.6	0.7	0.7	1.0	0.7	0.6
d	0.1	-0.1	0.3	0.1	0	-0.1	0.4	0	0.4	0.4

WHERE d = TIME AFTER − TIME BEFORE

HYPOTHESES NEXT, WHERE μ_d DENOTES MEAN DIFFERENCE IN THE ENTIRE POPULATION OF DRIVERS CONSUMING ONE PINT

- $H_0: \mu_d = 0$
- $H_1: \mu_d > 0$

CALCULATE ESTIMATORS

- $\bar{d} = \frac{\sum d}{n} = \frac{1.5}{10} = 0.15$
- $s_d^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{\sum d \sum d}{n}\right] = \frac{1}{9}\left[0.61 - \frac{1.5 \times 1.5}{10}\right] = \frac{77}{1800}$

USING A t DISTRIBUTION WITH $\nu = 9$, AT 5% SIGNIFICANCE

$t_9(5\%) = 1.833$

$t\text{-STATISTIC} = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{0.15 - 0}{\sqrt{\frac{77/1800}{10}}}$

$t\text{-STATISTIC} = 2.29341\ldots$

AS $2.293 > 1.833$ THERE IS SIGNIFICANT EVIDENCE THAT THE MEAN REACTION TIME INCREASED, AFTER ONE PINT OF BEER – SUFFICIENT EVIDENCE TO REJECT H_0

ASSUMPTIONS MADE

RANDOM SAMPLES, AND THE DIFFERENCES IN REACTION TIMES FOLLOW A NORMAL DISTRIBUTION

Question 2 (**+)

An Examining Board claims that their 2 Mathematics papers had identical grade boundaries for achieving the top grade.

The head of Mathematics of a large school decides to test this claim by looking at a random sample of 10 students from this school, whose marks were in the region of the top grade.

The percentage marks in each of the 2 papers for these 10 students are shown below.

Student	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Percentage mark in paper A	91	85	81	90	78	82	71	88	75	94
Percentage mark in paper B	85	86	82	80	80	83	72	84	70	90

Test, at the 10% level of significance, the Examining Board's claim

$\boxed{}$, $\boxed{\text{not significant, } -1.833 < 1.793 < 1.833}$

CANDIDATE	A	B	C	D	E	F	G	H	I	J
PAPER 1	91	85	81	90	78	82	71	88	75	94
PAPER 2	85	86	82	80	80	83	72	84	70	90

SETTING UP A PAIRED TWO TAILED TEST

$\boxed{\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array}}$ OR $\boxed{\begin{array}{l} H_0 : \mu_{diff} = 0 \\ H_0 : \mu_{diff} \neq 0 \end{array}}$

PREPARING ALL THE "AUXILIARIES" – LET "$d = 1 - 2$"

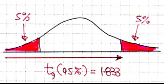
- $\sum \text{differences } (d) = 6 - 1 - 1 + 10 - 2 - 1 - 1 + 4 + 5 + 4$
- $\sum d = 23$
- $\sum d^2 = 201$

FINDING THE ESTIMATORS

$$\bar{d} = \frac{\sum d}{n} = \frac{23}{10}$$

$$s_d^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{\sum d \sum d}{n}\right] = \frac{1}{9}\left[201 - \frac{23 \times 23}{10}\right] = 16.4555\ldots$$

USING THE t-DISTRIBUTION, WITH 9 DEGREES OF FREEDOM, AT 10%, TWO TAILED



OBTAIN THE TEST STATISTIC TO COMPARE WITH THE CRITICAL VALUE OF 1.833

$$t\text{-stat} = \frac{\bar{d} - 0}{s/\sqrt{n}} = \frac{2.3}{\sqrt{\frac{16.555\ldots}{10}}} = 1.7929$$

AS $-1.8333 < 1.7929 < 1.8333$, THERE IS NO SIGNIFICANT EVIDENCE (AT 10%) THAT THE PAPERS WERE OF DIFFERENT DIFFICULTY

NO SUFFICIENT EVIDENCE TO REJECT H_0

Question 3 (**+)

A nutritional researcher is investigating the effect of dieting literature in promoting weight loss in overweight individuals.

The weights of random sample of 9 subjects were recorded, then they were given the dieting literature, and their weights were recorded again 9 weeks later.

The results are shown below.

Subject	A	B	C	D	E	F	G	H	I
Weight Before (kg)	95.2	96.0	100.2	88.2	91.7	85.0	74.3	83.7	87.0
Weight After (kg)	93.1	95.1	98.1	90.7	90.6	87.2	71.3	80.1	89.1

Test, at the 10% level of significance, whether there is evidence that the dieting literature has some effect in promoting weight loss.

, not significant, $0.8499... < 1.397$

SUBJECT	A	B	C	D	E	F	G	H	I
INITIAL WEIGHT	95.2	96.0	100.2	88.2	91.7	85.0	74.3	83.7	87.0
WEIGHT AFTER	93.1	95.1	98.1	90.7	90.6	87.2	71.3	80.1	89.1

SETTING UP A PAIRED VALUES t-TEST – LET d = "BEFORE – AFTER"

SUBJECT	A	B	C	D	E	F	G	H	I
DIFFERENCE (d)	2.1	0.9	2.1	-2.5	1.1	-2.2	3.0	3.6	-2.1

- $\sum d = 6.0$ • $\sum d^2 = 48.3$ • $n = 9$

$H_0 : \mu_d = 0$
$H_1 : \mu_d > 0$

OR

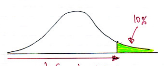
$H_0 : \mu_{BEF} = \mu_{AFTER}$
$H_1 : \mu_{BEF} > \mu_{AFTER}$

OBTAIN SUMMARY STATISTICS FOR THE SAMPLE DIFFERENCES

$$\bar{d} = \frac{\sum d}{n} = \frac{6}{9} = 0.666...$$

$$s_d^2 = \sqrt{\frac{1}{n-1}\left[\sum d^2 - \frac{\sum d \sum d}{n}\right]} = \sqrt{\frac{1}{8}\left[48.3 - \frac{6^2}{9}\right]} = 2.353189...$$

LOOKING FOR A CRITICAL VALUE AT 10%



$t_8(90\%) = 1.397$

FINDING THE t-STAT

$$t \text{ stat} = \frac{\bar{d} - 0}{\frac{s}{\sqrt{n}}} = \frac{0.6666...}{\frac{2.3531...}{\sqrt{9}}} = 0.8499$$

AS $0.8499 < 1.397$ THERE IS NO SIGNIFICANT EVIDENCE THAT THE LITERATURE IS EFFECTIVE

NO SUFFICIENT EVIDENCE TO REJECT H_0.

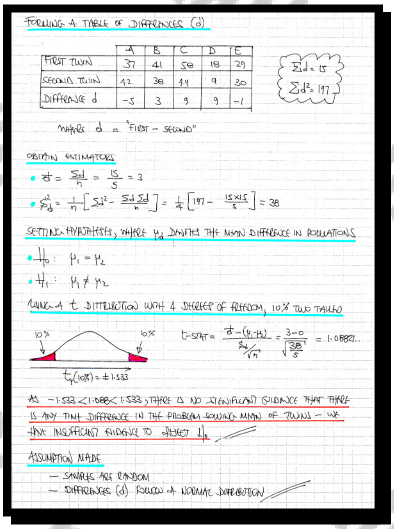
Question 4 (**+)

Five randomly chosen pairs of identical twins were each given the same puzzle to solve, and the times taken, in seconds, are summarized in the table below.

Twins	A	B	C	D	E
Time for first twin (sec)	37	41	58	18	29
Time for second twin (sec)	42	38	49	9	30

Test, at the 20% level of significance, whether there is a difference in the mean time taken to solve a puzzle, among identical twins, stating clearly any assumptions made.

FS2-L, $\boxed{\text{not significant, } -1.533 < 1.088 < 1.533}$



FORMING A TABLE OF DIFFERENCES (d)

	A	B	C	D	E
FIRST TWIN	37	41	58	18	29
SECOND TWIN	42	38	49	9	30
DIFFERENCE d	−5	3	9	9	−1

$\sum d = 15$

$\sum d^2 = 197$

WHERE d = "FIRST − SECOND"

OBTAIN ESTIMATORS

- $\bar{d} = \frac{\sum d}{n} = \frac{15}{5} = 3$
- $s_d^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{\sum d \sum d}{n}\right] = \frac{1}{4}\left[197 - \frac{15 \times 15}{5}\right] = 38$

SETTING HYPOTHESES, WHERE μ_d DENOTES THE MEAN DIFFERENCE IN POPULATIONS

- $H_0: \mu_1 = \mu_2$
- $H_1: \mu_1 \neq \mu_2$

USING A t DISTRIBUTION WITH 4 DEGREES OF FREEDOM, 10% TWO TAILED

10% 10%

$t_4(10\%) = \pm 1.533$

$$t\text{-STAT} = \frac{\bar{d} - (\mu_1 - \mu_2)}{s_d/\sqrt{n}} = \frac{3-0}{\sqrt{\frac{38}{5}}} = 1.08821\ldots$$

AS $-1.533 < 1.088 < 1.533$, THERE IS NO SIGNIFICANT EVIDENCE THAT THERE IS ANY TIME DIFFERENCE IN THE PROBLEM SOLVING MEAN OF TWINS – WE HAVE INSUFFICIENT EVIDENCE TO REJECT H_0

ASSUMPTION MADE

— SAMPLES ARE RANDOM

— DIFFERENCES (d) FOLLOW A NORMAL DISTRIBUTION

Question 5 (**+)

Two juice extracting machines are tested by a consumer magazine.

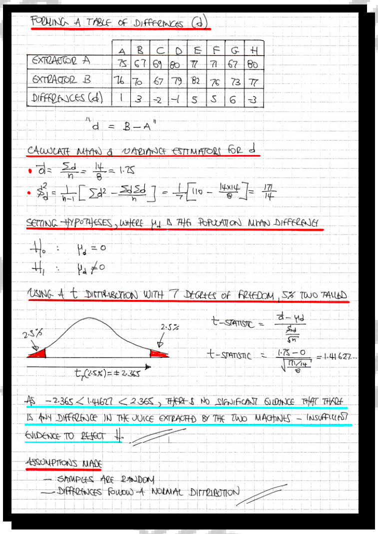
Eight oranges are cut in half and one half has its juice extracted by machine 1 and the other half has its juice extracted by machine 2.

The amounts of juice extracted, in ml, are summarised in the table below.

Orange	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Machine 1	75	67	69	80	77	71	67	80
Machine 2	76	70	67	79	82	76	73	77

Test, at the 5% level of significance, whether there is a difference in the amount of juice extracted between the two machines, stating any assumptions made.

$\boxed{}$, $\boxed{\text{not significant, } -2.365 < 1.416 < 2.365}$



FORMING A TABLE OF DIFFERENCES (d)

	A	B	C	D	E	F	G	H
EXTRACTOR A	75	67	69	80	77	71	67	80
EXTRACTOR B	76	70	67	79	82	76	73	77
DIFFERENCES (d)	1	3	-2	-1	5	5	6	-3

"$d = B - A$"

CALCULATE MEAN & VARIANCE ESTIMATORS FOR d

- $\bar{d} = \frac{\sum d}{n} = \frac{14}{8} = 1.75$
- $s_d^2 = \frac{1}{n-1}\left[\sum d^2 - \frac{\sum d \sum d}{n}\right] = \frac{1}{7}\left[110 - \frac{14 \times 14}{8}\right] = \frac{171}{14}$

SETTING HYPOTHESES, WHERE μ_d IS THE POPULATION MEAN DIFFERENCE

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

USING A t DISTRIBUTION WITH 7 DEGREES OF FREEDOM, 5% TWO TAILED

$t\text{-STATISTIC} = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$

$t\text{-STATISTIC} = \frac{1.75 - 0}{\sqrt{\frac{171/14}{8}}} = 1.41627\ldots$

AS $-2.365 < 1.41627 < 2.365$, THERE IS NO SIGNIFICANT EVIDENCE THAT THERE IS ANY DIFFERENCE IN THE JUICE EXTRACTED BY THE TWO MACHINES – INSUFFICIENT EVIDENCE TO REJECT H_0

ASSUMPTIONS MADE

- SAMPLES ARE RANDOM
- DIFFERENCES FOLLOW A NORMAL DISTRIBUTION

Question 6 (**+)

A new track surface is developed which is claimed to decrease the times of 100 m sprinters, compared with the old track surfaces currently used. To investigate this claim, the performances of 9 randomly chosen sprinters are measured on both surfaces. The results are summarised below.

Sprinter	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Time on Old Surface (sec)	10.7	11.2	11.5	10.9	11.8	12.0	10.6	13.1	12.1
Time on New Surface (sec)	10.8	11.0	11.4	11.1	11.4	11.6	10.7	12.5	12.5

Test, at the 2.5% level of significance, the claim about the new track surface.

, $\boxed{\text{not significant, } -2.306 < 0.915 < 2.306}$

FORMING A TABLE OF DIFFERENCES (d)

ATHLETE	A	B	C	D	E	F	G	H	I
TIME ON OLD TRACK	10.7	11.2	11.5	10.9	11.8	12.0	10.6	13.1	12.1
TIME ON NEW TRACK	10.8	11.0	11.4	11.1	11.4	11.6	10.7	12.5	12.5
DIFFERENCES (d)	0.1	-0.2	-0.1	0.2	-0.4	-0.4	0.1	-0.6	0.4

"d = NEW − OLD"

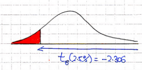
CALCULATE UNBIASED ESTIMATORS FOR THE MEAN & VARIANCE OF d

- $\bar{d} = \dfrac{\Sigma d}{n} = \dfrac{-0.9}{9} = -0.1$
- $s_d^2 = \dfrac{1}{n-1}\left[\Sigma d^2 - \dfrac{\Sigma d \Sigma d}{n}\right] = \dfrac{1}{8}\left[0.95 - \dfrac{(-0.9)(-0.9)}{9}\right] = \dfrac{43}{400}$

SETTING HYPOTHESES, WHERE μ_d IS THE POPULATION MEAN DIFFERENCE

- $H_0: \mu_d = 0$ OR $H_0: \mu_{BEF} = \mu_{AFTER}$
- $H_1: \mu_d < 0$ OR $H_1: \mu_{BEF} > \mu_{AFTER}$

USING A t DISTRIBUTION, WITH $\nu = 8$, AT 2.5% SIGNIFICANCE



$t\text{-STATISTIC} = \dfrac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$

$t\text{-STATISTIC} = \dfrac{-0.1 - 0}{\sqrt{\frac{43/400}{9}}}$

$t\text{-STATISTIC} = -0.91499\ldots$

AS $-2.306 < -0.915 < 2.306$, THERE IS NO SIGNIFICANT EVIDENCE TO SUPPORT THE CLAIM − INSUFFICIENT EVIDENCE TO REJECT H_0

Question 7 (**+)

Research is conducted to investigate the effect of consumption of large amounts of caffeine in problem solving ability.

Having avoided consumption of caffeine for a number of days, ten subjects were given a puzzle to solve and their times, in seconds, were recorded.

The same individuals returned and this time before attempting similar puzzles, they consumed large amounts of caffeine.

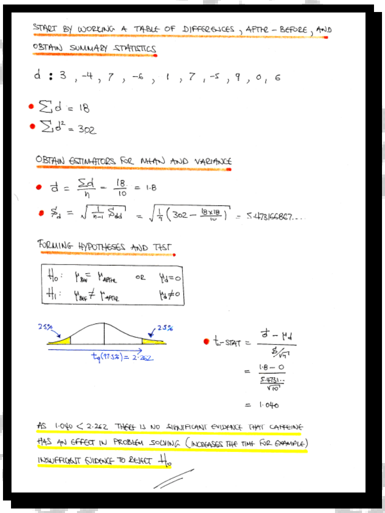
Their times were again recorded.

This information is shown in the table below.

Subject	A	B	C	D	E	F	G	H	I	J
Time without Caffeine (sec)	50	50	45	66	51	72	95	48	73	56
Time with Caffeine (sec)	53	46	52	60	52	79	90	57	73	62

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

$\boxed{}$, $\boxed{\text{not significant, } 1.040 < 2.262}$


START BY WORKING A TABLE OF DIFFERENCES, AFTER − BEFORE, AND OBTAIN SUMMARY STATISTICS

d : 3, −4, 7, −6, 1, 7, −5, 9, 0, 6

- $\sum d = 18$
- $\sum d^2 = 302$

OBTAIN ESTIMATORS FOR MEAN AND VARIANCE

- $\bar{d} = \frac{\sum d}{n} = \frac{18}{10} = 1.8$
- $s_d = \sqrt{\frac{1}{n-1} S_{dd}} = \sqrt{\frac{1}{9}\left(302 - \frac{18 \times 18}{10}\right)} = 5.473166867...$

FORMING HYPOTHESES AND TEST

$$\boxed{\begin{array}{llll} H_0: & \mu_{BEF} = \mu_{AFTER} & \text{OR} & \mu_d = 0 \\ H_1: & \mu_{BEF} \neq \mu_{AFTER} & & \mu_d \neq 0 \end{array}}$$

2.5% 2.5%

$t_9(97.5\%) = 2.262$

- $t\text{-STAT} = \frac{\bar{d} - \mu_d}{s/\sqrt{n}} = \frac{1.8 - 0}{\frac{5.4731...}{\sqrt{10}}} = 1.040$

AS $1.040 < 2.262$ THERE IS NO SIGNIFICANT EVIDENCE THAT CAFFEINE HAS AN EFFECT IN PROBLEM SOLVING (INCREASES THE TIME FOR EXAMPLE)

INSUFFICIENT EVIDENCE TO REJECT H_0

Question 8 (**+)

Research is conducted to investigate the effect of some recently implemented traffic measures to improve cycling routes into a city centre.

Before the measures were implemented, the times of 10 randomly chosen cyclists into the city centre were recorded.

The times of the same 10 randomly chosen cyclists were also recorded after the measures were implemented.

Their times were again recorded and the results are summarized in the table below.

Cyclist	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Time before measures (min)	21	18	25	17	22	18	31	25	24	17
Time after measures (min)	20	20	20	19	21	13	26	27	19	15

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

$\boxed{}$, $\boxed{\text{significant, } -1.846 < -1.833}$

START BY WORKING A TABLE OF DIFFERENCES, d = AFTER − BEFORE, AND OBTAIN SUMMARY STATISTICS

d: −1, 2, −5, 2, −1, −5, −5, 2, −5, −2

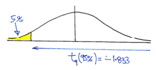
- $\sum d = -18$
- $\sum d^2 = 118$

OBTAIN ESTIMATORS FOR THE MEAN AND VARIANCE OF THE DIFFERENCES FOR THE POPULATION

- $\bar{d} = \frac{\sum d}{n} = \frac{-18}{10} = -1.8$
- $s_d = \sqrt{\frac{1}{n-1} S_{dd}} = \sqrt{\frac{1}{9}\left(118 - \frac{(-18)^2}{10}\right)} = 3.084008935$

FORMING HYPOTHESES AND CARRY OUT THE TEST

H_0: $\mu_{BEF} = \mu_{AFTER}$		OR	$\mu_d = 0$
H_1: $\mu_{BEF} > \mu_{AFTER}$			$\mu_d < 0$



$$t_{STAT} = \frac{\bar{d} - \mu_d}{s/\sqrt{n}} = \frac{-1.8 - 0}{\frac{3.084..}{\sqrt{10}}} = -1.846$$

AS $-1.846 < -1.833$ THERE IS SIGNIFICANT EVIDENCE THAT THE RECENTLY IMPLEMENT MEASURES HAD THE DESIRED EFFECT, HOWEVER AS THE TEST VALUE IS VERY CLOSE TO THE CRITICAL VALUE MORE TESTING MIGHT BE MORE APPROPRIATE — INSUFFICIENT EVIDENCE TO REJECT H_1