MATH9D2 : Practical Statistics
Practical 7 : Solutions

1. Quantitative Data ; Question about the standard deviation ; One Sample
Appropriate technique : Confidence Interval for the standard deviation & interpret

95% CI for σ $\left(\sqrt{\dfrac{(n-1)s^2}{\chi^2(n-1;0.025)}} , \sqrt{\dfrac{(n-1)s^2}{\chi^2(n-1;0.975)}} \right)$

$n = 16$ $\quad s = 0.2608$ (see Practical 6)

$\chi^2(15;0.025) = 27.49$

$\chi^2(15;0.975) = 6.26$

$$\left(\sqrt{\frac{15 \times 0.2608^2}{27.49}} , \sqrt{\frac{15 \times 0.2608^2}{6.26}} \right)$$

$$\left(\sqrt{0.0371} , \sqrt{0.1630} \right)$$

$$(0.193 , 0.404)$$

95% confident that the population standard deviation of crankshaft dimension lies between 0.193 mm & 0.404 mm.

$H_0 : \sigma = 0.35$ $\quad H_1 : \sigma \neq 0.35$ where σ = population standard deviation of crankshaft dimensions

0.35 is within the 95% CI for σ so we cannot reject H_0 in favour of H_1 at 5% level

ie insufficient evidence, at the 5% level, that the standard deviation of crankshaft dimensions is different to 0.35 mm

2. Quantitative Data ; Two Independent Samples
Question about the Variances ; Two Tailed since asks about difference
→ so test is F test.

[Note : this should always be done before comparing means for two independent samples]

$H_0: \sigma_1^2 = \sigma_2^2$ where σ_1^2 = population variance of cleaner A weight losses
$H_1: \sigma_1^2 \neq \sigma_2^2$ σ_2^2 = population variance of cleaner B weight losses

Significance Level 0.05

[A labelled 1 because Larger variance]

Test Statistic $F = \frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$ under H_0 ← Larger variance

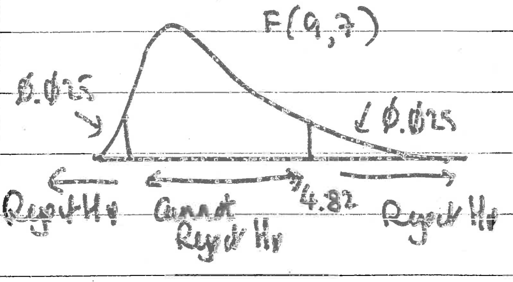
Observed Test Statistic

A $n_1 = 10$ $S_1 = 0.756$
B $n_2 = 8$ $S_2 = 0.676$

[Calculated in Practical 6]

$$F = \frac{0.756^2}{0.676^2} = 1.25$$ //

Rejection Region 2 tailed ; 0.05 ; $F(n_1-1, n_2-1) = F(9,7)$



Upper Critical Value =
$F(9,7; 0.025) = 4.82$
9 along 7 down in tables!!

Conclusion Observed Test Statistic (1.25) is not in the Rejection Region (< 4.82) so we cannot reject H_0 in favour of H_1 at 5% level. Insufficient evidence of a significant difference in the variances of weights losses. So can assume equal variances, unpaired t-test (assuming equal variance) performed in Practical 6 is appropriate

3. (a) See attached

(b) $n = 10$ $\quad \Sigma x = 190 \quad \Sigma x^2 = 3940$

$\Sigma y = 173 \quad \Sigma y^2 = 3269$

$\Sigma xy = 2988$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 2988 - \frac{190 \times 173}{10} = -299$$

negative since negative relationship

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 3940 - \frac{190^2}{10} = 330$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 3269 - \frac{173^2}{10} = 276.1$$

MUST always be negative

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{-299}{\sqrt{330 \times 276.1}} = -0.991$$

$R^2 = (-0.991)^2 = 0.982$

ie. 98.2% of the variability in y is explained by the linear relationship with x.

4. (a) See attached

(b) $n = 8$ $\quad \Sigma x = 136 \quad \Sigma x^2 = 2480$

$\Sigma y = 152 \quad \Sigma y^2 = 3048$

$\Sigma xy = 2746$

$$S_{xy} = \Sigma xy - \frac{\Sigma x \Sigma y}{n} = 2746 - \frac{136 \times 152}{8} = 162$$

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 2480 - \frac{136^2}{8} = 168$$

$$S_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 3048 - \frac{152^2}{8} = 160$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{162}{\sqrt{168 \times 160}} = \frac{162}{163.9512}$$
$$= 0.988_{//}$$

$R^2 = 0.988^2 = 0.976$

ie 97.6 % of the variability in y is explained by the linear relationship with x.

(c) $H_0 : \rho = 0$ where ρ = population correlation
$H_1 : \rho \neq 0$ between x & y

Significance Level 0.05

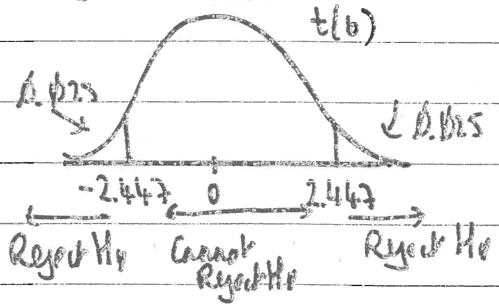
Test Statistic $t = \dfrac{r\sqrt{(n-2)}}{\sqrt{1-r^2}} \sim t(n-2)$ under H_0

Observed Test Statistic

$n = 8 \qquad r = 0.988$

$$t = \frac{0.988\sqrt{6}}{\sqrt{1-0.988^2}}$$
$$= \frac{2.420}{0.154} = 15.71_{//}$$

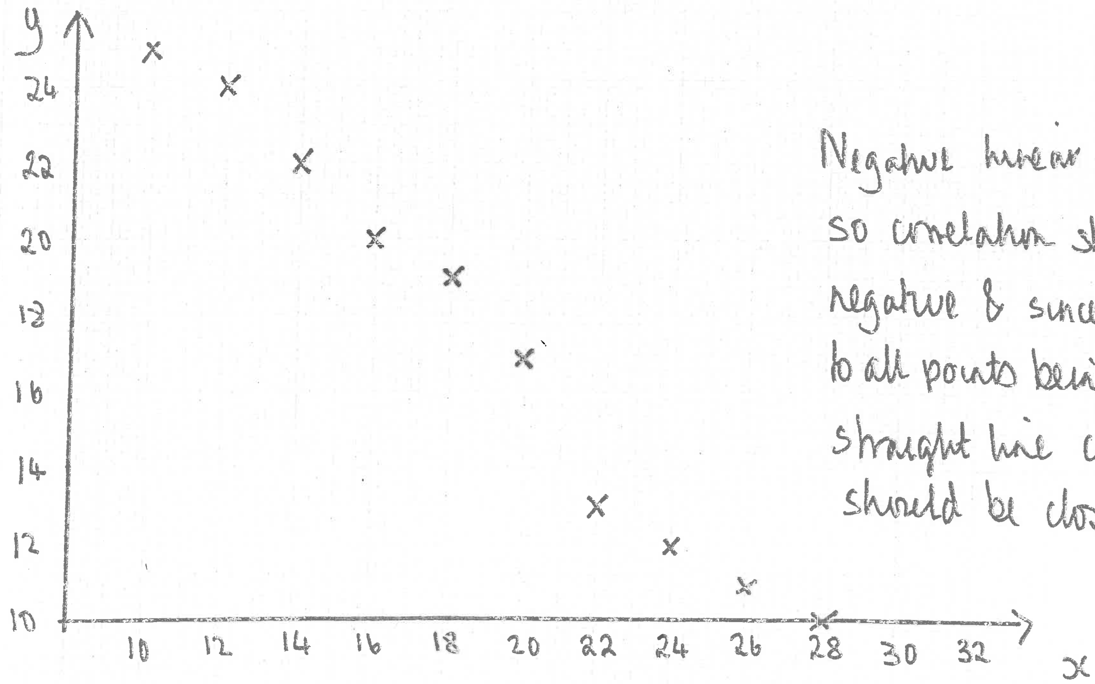
Rejection Region 0.05 ; 2 tailed ; t(6)



Critical Value = $t(6; 0.025)$
= 2.447

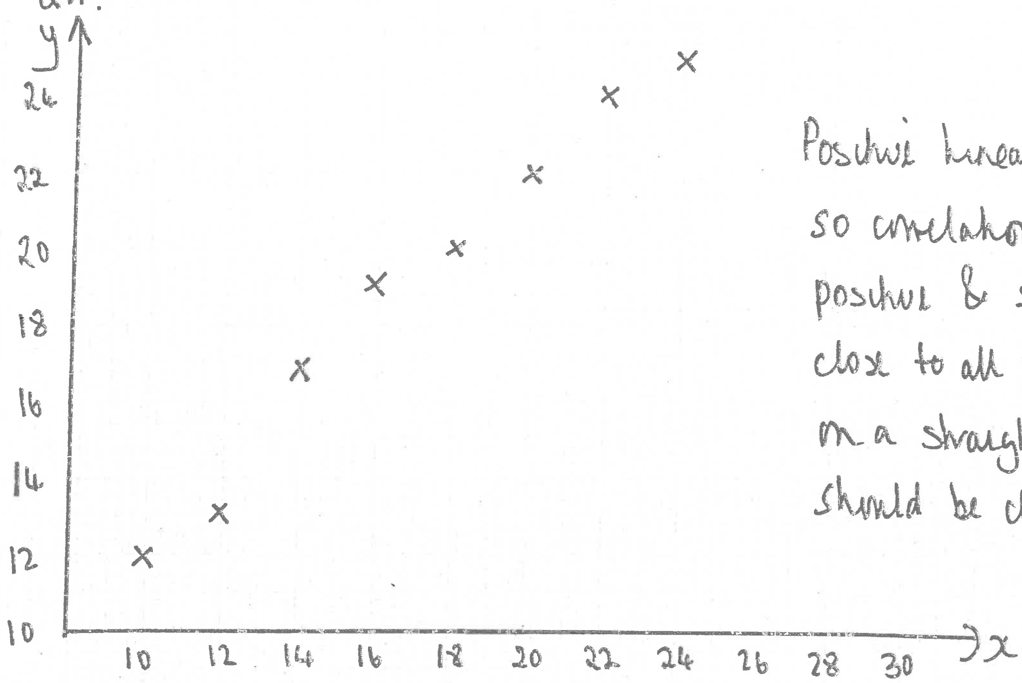
Conclusion Observed Test Statistic (15.71) is in the Rejection Region (> 2.447) so we can reject H_0 in favour of H_1 at 5% level & conclude that the correlation between x & y is significantly different to zero.

Q3.



Negative linear Relationship so correlation should be negative & since very close to all points being on a straight line correlation should be close to -1

Q4.



Positive linear Relationship so correlation should be positive & since very close to all points being on a straight line correlation should be close to +1.