2010 exam -

(i) $n = 10 \qquad \bar{x} = 12.2 \qquad s = 6.8$

From Box plot we get 5 number summary (2.8, 8, 12, 16.8, 23) approx.

(ii) from boxplot the data is

- approximately symmetric.
- centred on median of about 12. (mean $\bar{x} = 12.2$)
- values from 2.8 to 23 approx.
- no outliers.
- from normal quantile plot - points close to a line so the data is plausibly from a normal dist.

(iii) A 99% confidence interval for the mean PSI for Atlanta

$$\bar{X} \pm t^* \frac{s}{\sqrt{n}}$$

t^* from $t_{n-1} = t_9$ for 99% CI $t^* = 3.250$

$$12.2 \pm 3.250 \times \frac{6.8}{\sqrt{10}}$$

$$12.2 \pm 6.99\ldots$$

i.e interval (5.2, 19.2).

(iv) Assumptions:

- need to assume that PSI at Atlanta is at least approximately normal from the normal quantile plot this is plausible
- need to assume a random sample or at least data values independent.
  - can't check from these plots, this depends how the days/data were selected.

(b) PSI values for Houston in general highest; Atlanta a little higher than Chicago. Houston also more spread out than ~~Ch~~ Atlanta and Chicago. None show outliers. All fairly symmetric. All Houston values bigger than all Chicago values, largest Atlanta value is less than first quartile of Houston

c) (i) $n = 96$ tested.

sample propotion is $\hat{p} = \frac{12}{96}$.

98% Confidence interval for the true proportion

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

$$\frac{12}{96} \pm 2.33\sqrt{\frac{12/96\,(1-12/96)}{96}}$$

$$0.125 \pm 0.079.$$

Confidence interval $(0.05, 0.24)$.

i.e 5% to 24%.

(ii) Assumptions.

- random sample / independent data
- need to assume n large enough so that C.L.T. justifies using z^{*}.

Q5, $A = \dfrac{75081}{3} = 25027$

$B = \dfrac{235424}{16} = 14714$

(check $A/B = \dfrac{25027}{14714} \doteq 1.7$ good!)

(ii) $k=4$ groups.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: not all $\mu_1, \mu_2, \mu_3, \mu_4$ are equal.

Let X_1 be driving km till failure for plug of type 1.

Similarly X_2, X_3, X_4.

To do ANOVA we need assume that

$X_1 \sim N(\mu_1, \sigma)$

$X_2 \sim N(\mu_2, \sigma)$

$X_3 \sim N(\mu_3, \sigma)$

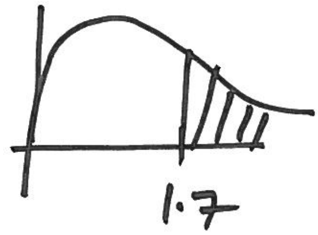
$X_4 \sim N(\mu_4, \sigma)$

Also need independent random samples from each.

$\alpha = .01$

Test statistic is F

if H_0 is true $F \sim F_{3,16}$



observed $F = 1.7$

p-value $= P(F > 1.7)$ $\quad F \sim F_{3,16}$

> 0.05 from tables

$> \alpha = 0.01$

[tables don't have $F_{3,16}$. $F_{3,15}$ and $F_{3,20}$ but for either of these we get $> .05$]

p-value is not small.

Data consistent with H_0

Accept H_0.

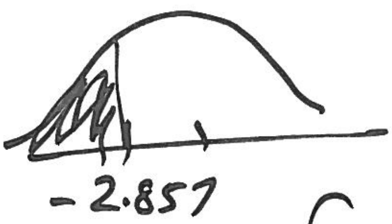
Accept 4 types of spark plugs have same average performance.

(b) X opening altitude of random parachute (m)

$X \sim N(200, 35)$ $\qquad Z = \frac{X - 200}{35} \sim N(0,1)$

$\underline{P(X < 100)} = P\left(Z < \frac{100-200}{35}\right)$

$= P(Z < \cancel{\;} -2.857$



$P(X<100) = 0.0021$

$[P(X \geq 100) = 1 - .0021 = .9979]$

5 parachutes dropped independently.

$P(\text{At least one opens} < 100\ m)$

$= 1 - P(\text{none open} < 100\ m)$

$= 1 - P(\text{All open at} \geq 100\ m)$.

$= 1 - (0.9979)^5 \doteq 0.01$.

Q5.

c) $X \sim \text{Poisson}(2)$

for $x = 0, 1, 2, \ldots$ $\quad P(X=x) = \dfrac{e^{-2} 2^x}{x!}$

$Y \sim \text{Exp } 1$. density $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0. \end{cases}$

(i) $P(X=1) = \dfrac{e^{-2} 2^1}{1!} = 2e^{-2} = 0.2707$ to 4dp

(ii) $P(Y<1) = \displaystyle\int_0^1 e^{-x}\,dx$

$= 1 - e^{-1} = 0.6321$ to 4dp.

(iii) $P[(X=1) \text{ or } (Y<1)] = P(X=1) + P(Y<1) - P(X=1 \text{ and } Y<1)$

$= P(X=1) + P(Y<1) - P(X=1)P(Y<1)$

by independence

$= 0.2707 + 0.6321 - 0.2707 \times .6321$

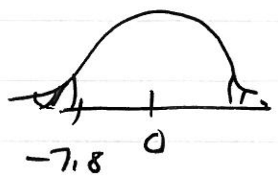
$= 0.732$ to 3dp

Q6. (a) (i) $H_0: \beta_1 = 0 \qquad H_a: \beta_1 \neq 0$

(ii) test statistic is $T = \dfrac{b_1}{se(b_1)}$. $\qquad \alpha = 0.01$

if H_0 is true $T \sim t_{n-2} = t_{28}$.

~~observed~~ $t = -7.80$



P-value $= 2\,P(T > 7.80)_{T \sim t_{28}}$

$< 2 \times .0005$ from table
$= 0.001$

p-value $< .001 < \alpha = 0.01$

Very small p-value, so reject H_0 and accept Ha that $\beta_1 \neq 0$

We conclude that ppv is significant in the regression.

(b) $R^2 = 0.685$ and the relationship is negative, so $r = -\sqrt{0.685} = -0.837647$.
Correlation is $r = -0.84$ to 2dp.

(c) $E(\text{Ratio} \mid ppv = x) = \beta_0 + \beta_1 x$

We want a 95% C.I for β_1.
$b_1 \pm t^* s_e(b_1)$ $\qquad$ t^* for 95% and t_{28} is $t^* = 2.048$

$-0.000015 \pm 2.048 \times .0000019\underline{0}$

$-0.000015 \pm 0.0000038912$.

i.e $(-0.000019, -0.000011)$.

a) (i) The 95% C.I is a 95% C.I for the mean value of the ratio when ppv = 750 mm/sec

(ii) The 95% PI is a 95% prediction interval for the value of ratio when ppv = 750 mm/sec. This is wider than the C.I because it includes the uncertainty in the estimate of the mean value, and the fact that ratio is a random variable.

(iii) Each interval is the point estimate ± margin of error.

point estimate is

$$\widehat{\text{ratio}} = 1.00007 - 0.0000\cancel{15}^{1484} \times 750$$
$$= \cancel{1.00007} - \cancel{0.01125}\ \cancel{0.0111}3.$$
$$= 0.98894$$

(e) The $\varepsilon \sim N(0, \sigma)$ ~~with~~

- We need independent errors, i.e independent data - this can't be checked from data plots

- The normality assumption can be checked by plotting the residuals - the normal probability plot of the standardized residuals; shows

points close to a line. This confirms normality is plausible
- the histogram of residuals is approximately bell shape, no extreme outliers - plausibly normal.

- The constant standard deviation σ (constant variance σ^2) assumption can be checked by plotting (standardized) residuals against fitted value. The random looking scatter in the plot shows this is plausible.

- The plot of residuals against order also shows no pattern - confirming constant σ, (~~and also reassum~~