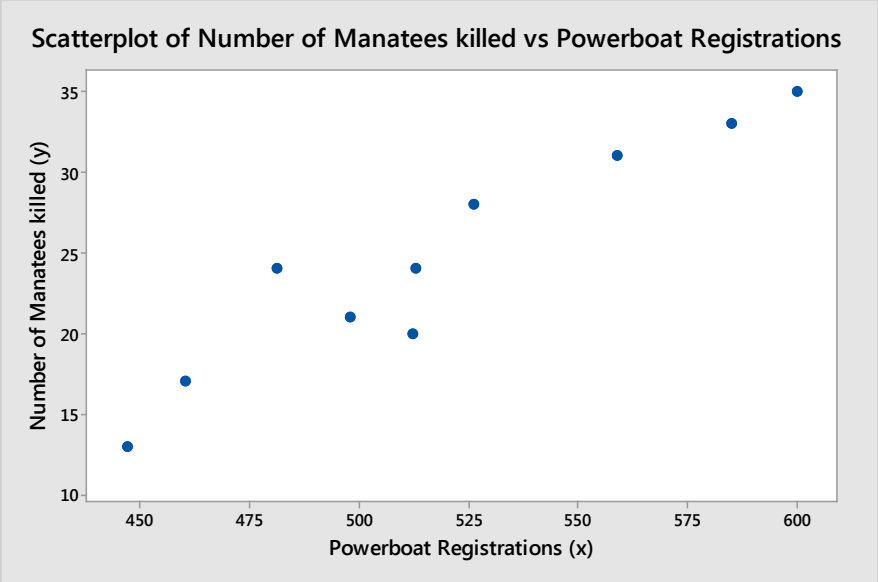
MATU9D2 : Practical Statistics Spring 2017

Weekly Assignment 8 : Minitab : Solutions

(i) **Subjective Impression** : Positive linear relationship between number of manatees killed an number of powerboat registrations



(ii) From the Minitab Output below :

Pearson's Product Moment Correlation i.e. the r value = 0.949

(iii) $H_0 : \rho = 0 \quad H_1 : \rho \neq 0$ and $p<0.001$

So we can reject H_o in favour H_1 at 1% level : correlation is significantly different to zero i.e. significant relationship between the number of manatees killed and the number of powerboat registrations.

Minitab Output for parts (ii) & (iii)

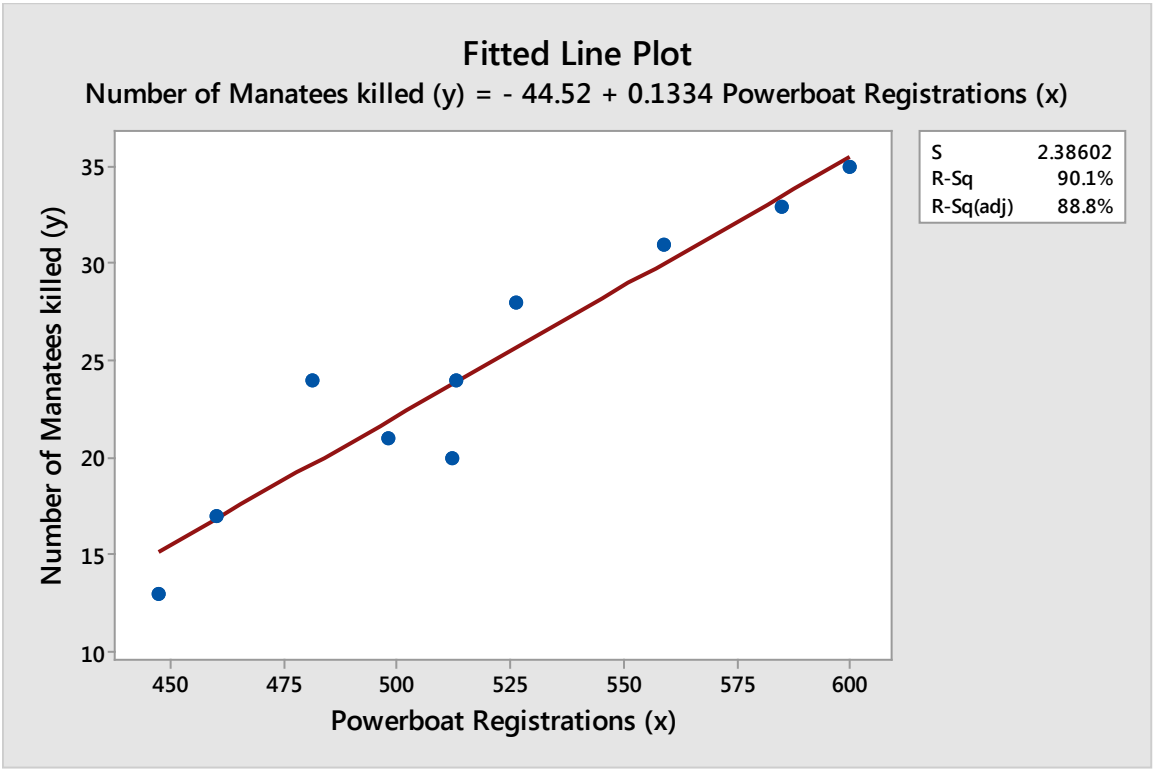
Correlation: Number of Manatees killed (y), Powerboat Registrations (x)

```
Pearson correlation of Number of Manatees killed (y) and Powerboat Registrations (x) = 0.949
P-Value = 0.000
```

(iv) From the Output on Page 4 : **Analysis only valid if assumptions are valid – Residual Plots** on Page 4

the fitted line is (Manatees Killed) = -44.52 +0.1334 (Number of Registrations)

Intercept = α = -44.52 Slope = β = 0.1334 Error Variance = σ^2 = 5.69



Interpreting Minitab Output on Pages 4 & 5

(v) $H_0 : \beta = 0 \quad H_1 : \beta \neq 0$

Observed Test Statistic = 8.52 & p<0.001

So we can reject H_o in favour H_1 at 1% level : slope is significantly different to zero
i.e. significant relationship between the number of manatees killed and the number of powerboat registrations.

(vi) 95% confidence interval for the average number of manatees killed given the number of powerboat registrations is 460,000.

Since Powerboat Registrations in thousands x = 460

95% CI for Mean Number of Manatees killed when 460,000 Powerboat Registrations is `(14.122, 19.575)`

(vii) 95% prediction interval for the number of manatees killed given the number of powerboat registrations is 460,000.

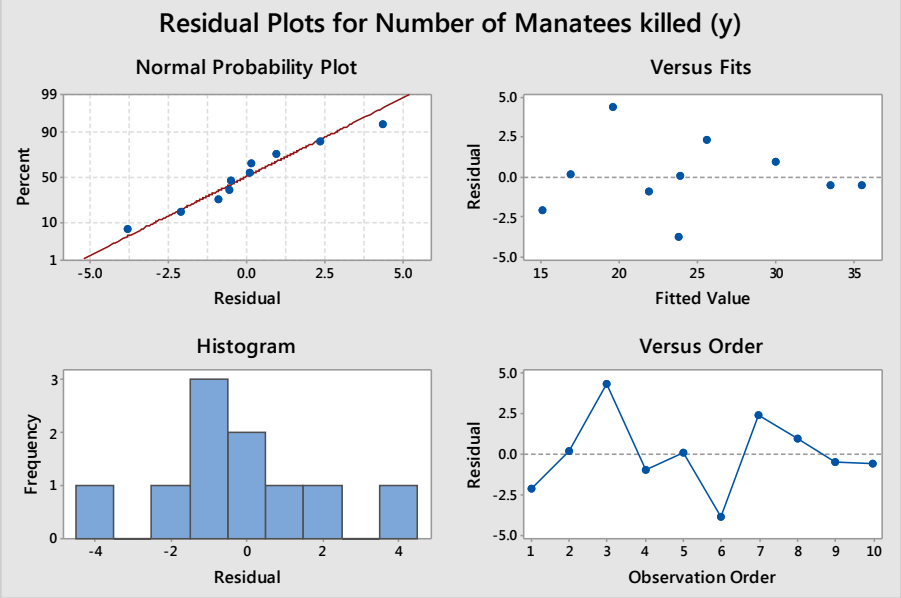
Since Powerboat Registrations in thousands x = 460

95% CI for Number of Manatees killed for single year when 460,000 Powerboat Registrations is `(10.708, 22.989)`

Practically Useful??

(vi) & (vii) Will help decide whether this statistically significant relationship is useful practically.

For example, in order that the manatee population is not under threat how many (if any) can we 'allow' to be killed by powerboats?



Validate Assumptions

Top Left : Can assume normality since graph is approximately linear.

Top Right : Slight problem with assumption of constant variance as points not evenly spread about zero i.e seem closer as fitted value larger

Regression Analysis: Number of Manatees killed (y) versus Powerboat Registrations (x)

Error Variance = σ^2 = 5.69

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	412.86	412.855	72.52	0.000
Powerboat Registrations (x)	1	412.86	412.855	72.52	0.000
Error	8	45.54	5.693		
Total	9	458.40			

R^2 = 90.1%

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.38602	90.06%	88.82%	85.89%

$H_0 : \alpha = 0 \quad H_1 : \alpha \neq 0$

Observed Test Statistic = -5.46 p=0.001

So can reject H_o in favour H_1 at 1% level : intercept is significantly different to zero

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-44.52	8.15	-5.46	0.001	
Powerboat Registrations (x)	0.1334	0.0157	8.52	0.000	1.00

Regression Equation

Number of Manatees killed (y) = -44.52 + 0.1334 Powerboat Registrations (x)

Fitted Line : y = -44.5 – 0.133 x

$H_0 : \beta = 0 \quad H_1 : \beta \neq 0$

Observed Test Statistic = 8.52 p<0.001

So can reject H_o in favour H_1 at 1% level : slope is significantly different to zero i.e. significant relationship

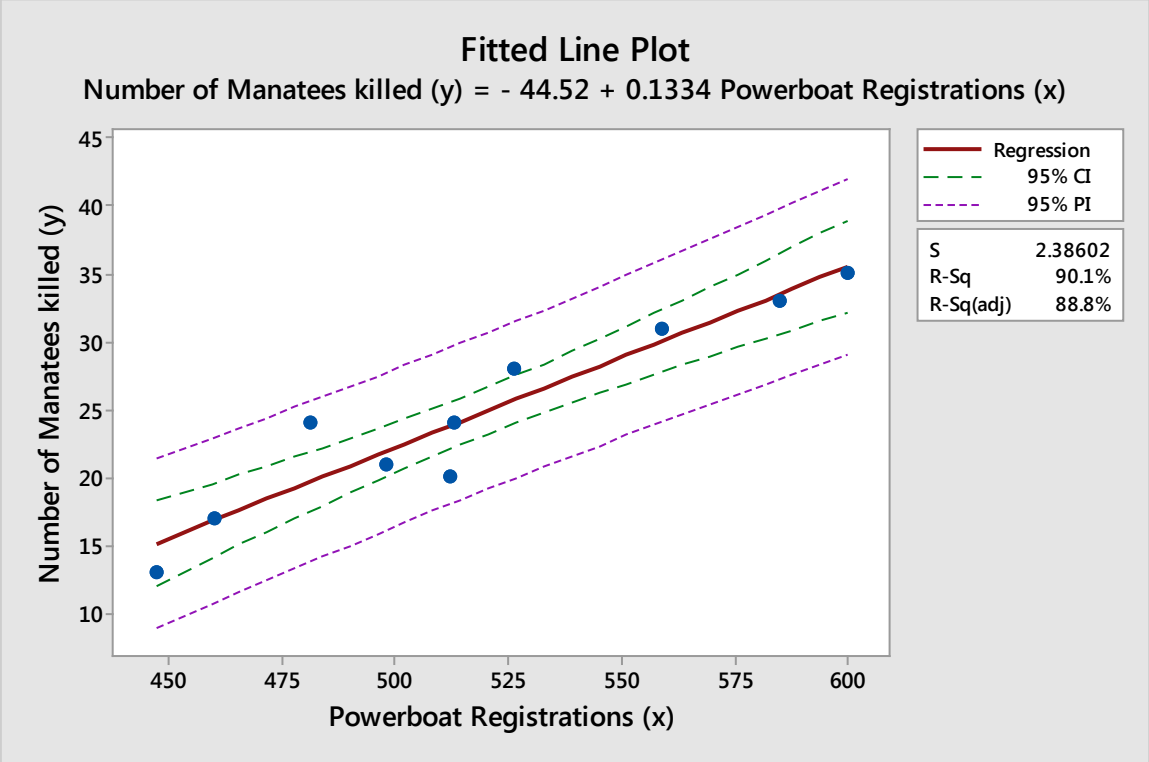
Prediction for Number of Manatees killed (y)

Regression Equation

Number of Manatees killed (y) = -44.52 + 0.1334 Powerboat Registrations (x)

Variable	Setting
Powerboat Registrations (x)	460

Fit	SE Fit	95% CI	95% PI
16.8483	1.18233	(14.1218, 19.5748)	(10.7076, 22.9889)



R^2 = 90.1% so 90.1% of the variability in y is explained by the linear relationship with x

Green lines show 95% Prediction Interval for single future numbers of registrations

Red lines show 95% Confidence Intervals for mean y values for given number of registrations values

Both are relatively narrow showing a good relationship in practice