Correlation Test & Linear Regression

Test of Correlation (Testing for \( \rho = 0 \))

1) First, enter your data in R commander or upload an existing file into R by selecting Data, New data or Import Data.

2) In R commander, select Statistics, Summaries, and select Correlation test as show in the following figure.
3) In the correlation dialog box, click and drag mouse to select the two variables, MANKILL and NPOWERBT for computing the correlation, and have the Pearson product-moment bullet checked, and click OK.

Interpret:

4) The p-value of the test is \( \text{p-value} = 5.109 \times 10^{-7} \) which is less than .05, so we can conclude that the correlation is statistically significant different from 0, at 5% level of significance. The Pearson correlation coefficient is 0.9414773.
Linear Regression: (Using the same data.)

5) From IPSUR, select Statistics, Fit Models, Linear Regression

6) The following is the R Output and the necessary numbers needed from it.

```
R Output

Call:
  lm(formula = MANKILL ~ NPWERT, data = Dataset)

Residuals:
     Min      1Q  Median      3Q     Max
-9.24681 -2.02166  0.02172  2.33692  5.63275

Coefficients:  estimate Std. Error t value Pr(>|t|)
(Intercept)     -41.4304     7.4122  -5.689  0.000118 ***
NPWERT           0.1249      0.0129   9.675   5.11e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.276 on 12 degrees of freedom
Multiple R-Squared: 0.8864,  Adjusted R-squared: 0.8769

Coefficient of determination = $R^2$
```

F-statistic: 93.61 on 1 and 12 DF,  p-value: 5.109e-07
Interpret:

7) The Intercept (p-value = .000118) and the beta coefficient (p-value = 5.11e-07) are both significantly different from zero. The following equation is used to find the equation of the regression line, the variables \( \alpha \) and \( \beta \) are highlighted above.

\[
\hat{y} = \hat{\alpha} + \hat{\beta} \cdot x;
\]

\[
\hat{y} = -41.4304 + .1249 \cdot x
\]

8) Example: If at a certain year the number of power boats registered is 700, estimate how many manatee on average would be killed.

To solve this example simply plug 700 into the equation that was found above.

\[
\hat{y} = -41.430439 + .124862 \cdot x
\]

\[
= -41.430439 + .124862 \cdot 700
\]

\[
= 45.973
\]

The average response at \( x = 700 \) is 45.973.

9) The residual for the data point for year 1978 at \( x = 460 \) (see data table in the beginning) would be the actual observed \( y = 21 \) (from year 1978) at \( x = 460 \) minus the fitted value of \( y \) at \( x = 460 \) that is \( \hat{y} = 16.006 \).

Residual at for data point from year 1978 (at \( x = 460 \)) = 21 − 16.006 = 4.994.

The residuals for all data points can be obtained using the following R command:

\[
> \text{residuals(LinearModel.1)}
\]

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
9 & 10 & 11 & 12 & 13 & 14 \\
& & & & & & & -0.1053526 & 0.1487966 & 2.6537757 & -1.3451178
\end{array}
\]