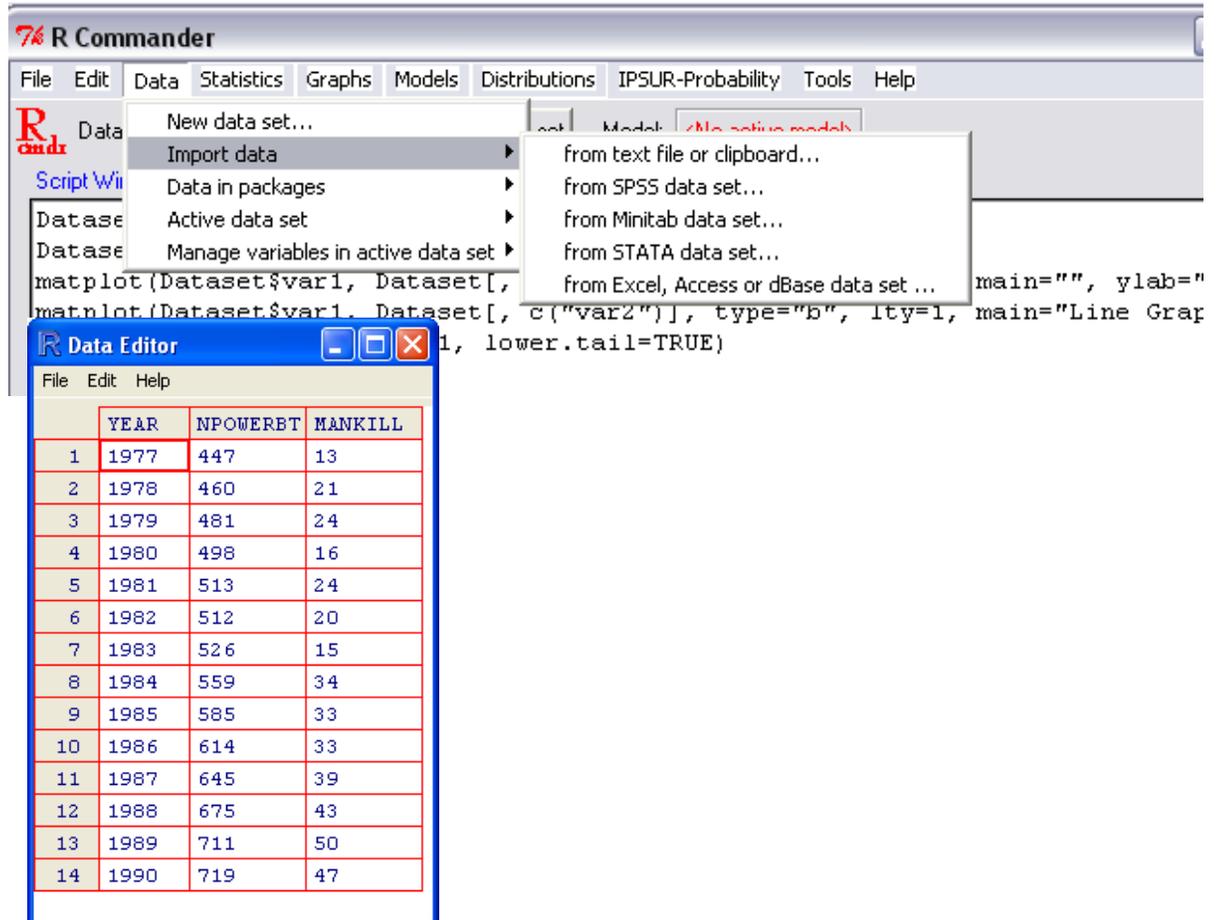


Correlation Test & Linear Regression

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

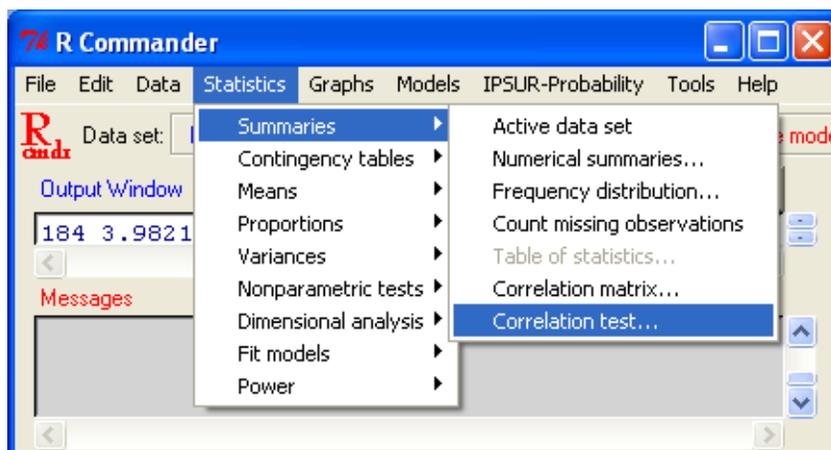
- 1) First, enter your data in R commander or upload an existing file into R by selecting **Data, New data or Import Data**.



The screenshot shows the R Commander interface. The 'Data' menu is open, showing options like 'New data set...', 'Import data', 'Data in packages', 'Active data set', and 'Manage variables in active data set'. The 'Import data' submenu is also visible, listing sources like 'from text file or clipboard...', 'from SPSS data set...', 'from Minitab data set...', 'from STATA data set...', and 'from Excel, Access or dBase data set...'. A 'Data Editor' window is open in the foreground, displaying a table with 14 rows and 4 columns: YEAR, NPOWERBT, and MANKILL. The data is as follows:

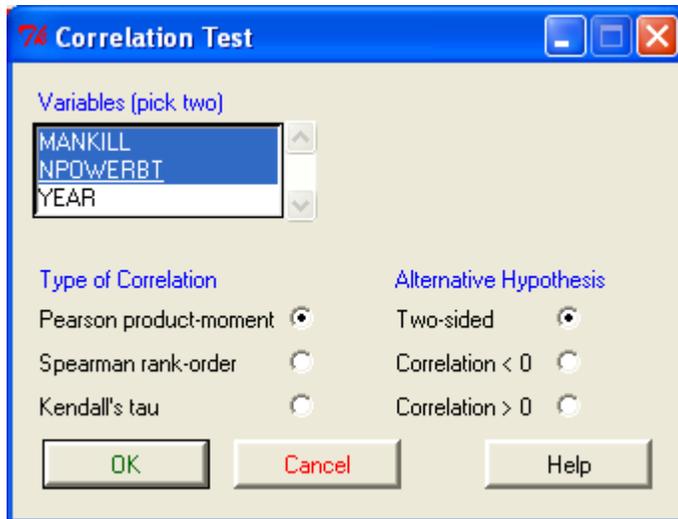
	YEAR	NPOWERBT	MANKILL
1	1977	447	13
2	1978	460	21
3	1979	481	24
4	1980	498	16
5	1981	513	24
6	1982	512	20
7	1983	526	15
8	1984	559	34
9	1985	585	33
10	1986	614	33
11	1987	645	39
12	1988	675	43
13	1989	711	50
14	1990	719	47

- 2) In R commander, select Statistics, Summaries, and select Correlation test as show in the following figure.



The screenshot shows the R Commander interface with the 'Statistics' menu open. The 'Summaries' submenu is also open, and 'Correlation test...' is highlighted. The 'Data set:' field shows '184 3.9821'. The 'Output Window' and 'Messages' panels are visible at the bottom.

- 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:

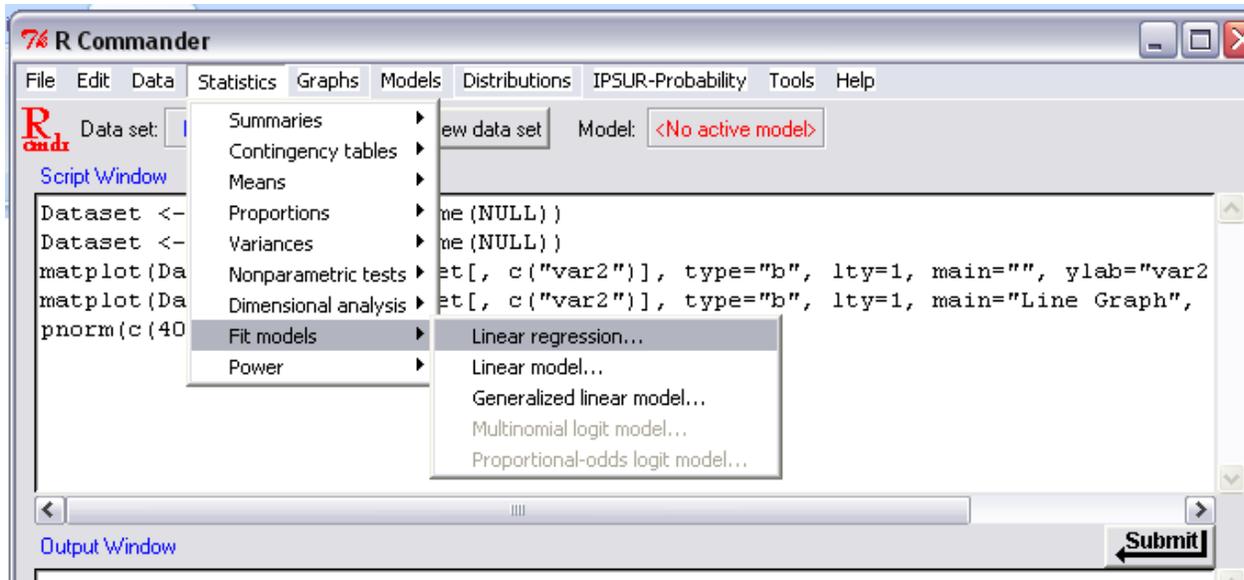
- 4) The p-value of the test is $p\text{-value} = 5.109e-07$ which is less than .05, so we can conclude that the correlation is statistically significant different from 0, at 5% level of significance. The **Pearson correlation coefficient** is **0.9414773**.

Pearson's product-moment correlation

```
data: Dataset$MANKILL and Dataset$NPOWERBT
t = 9.6755, df = 12, p-value = 5.109e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8210214 0.9816797
sample estimates:
cor
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 ~ NPOWERBT, 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  -5.589 0.000118 ***
NPOWERBT      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
F-statistic: 93.61 on 1 and 12 DF, p-value: 5.109e-07
```

Estimate **Std. Error** **t value** **Pr(>|t|)**

-41.4304 **7.4122** **-5.589** **0.000118** *******

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 28

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 α and β are highlighted above.

Equation of the regression line:

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

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

- 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.

$$\begin{aligned}\hat{y} &= -41.430439 + .124862 \cdot x \\ &= -41.430439 + .124862 \cdot 700 \\ &= 45.973\end{aligned}$$

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:

> `residuals(LinearModel.1)`

```
 1  2  3  4  5  6  7
-1.3827375  4.9940605  5.3719650 -4.7506838  1.3763908 -2.4987475 -9.2468112
 8  9 10 11 12 13 14
 5.6327530  1.3863490 -2.2346401 -0.1053526  0.1487966  2.6537757 -1.3451178
```