Example 1.1

Graph > Stem-and-Leaf

Stem-and-leaf of fundrais  N = 60
Leaf Unit = 1.0

19  0  01111222233333344
(17)  0  555666666778888
24  1  0001222244
14  1  55666789
  6  2  01
  4  2  6
  3  3  4
  2  4
  2  4  8
  1  5
  1  5
  1  6
  1  6
  1  7
  1  7
  1  8  3

Fundraising expenses (% of total expenses) for organized charities have a unimodal distribution, skewed to the right (+) with two unusual observations at 48 and 83 percent.

Graph > Histogram
Example 1.16

Stat > Basic Statistics > Display Descriptive Statistics (select C1 % Copper)

Results for: exp01-16.mtw

Descriptive Statistics: % Copper

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Copper</td>
<td>26</td>
<td>0</td>
<td>3.654</td>
<td>0.303</td>
<td>1.547</td>
<td>2.000</td>
<td>2.70</td>
<td>3.35</td>
<td>3.875</td>
<td>10.100</td>
</tr>
</tbody>
</table>

Conclusion: The average (or mean) copper content is 3.65%. Half of the Bidri samples had copper content of less than 3.35% and half had a greater copper content.
Example 1.18 (p. 41)

Stat > Basic Statistics > Display Descriptive Statistics (select C1)

Results for: exp1-18.mtw

Descriptive Statistics: C1

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>N*</th>
<th>Mean</th>
<th>SE Mean</th>
<th>StDev</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>19</td>
<td>0</td>
<td>86.32</td>
<td>5.35</td>
<td>23.32</td>
<td>40.00</td>
<td>70.00</td>
<td>90.00</td>
<td>98.00</td>
<td>125.00</td>
</tr>
</tbody>
</table>

Conclusion: A typical pit depth is 86 thousandths of an inch, with a typical spread of 23 thousandths of an inch either way.

Graph > Boxplot (Simple, select C1)

Conclusion: The distribution of pit depths is positively skewed with no apparent outliers.
Example 3.32

Calc > Probability Distributions > Binomial (n=15, p=0.2, CDF, x=8)

**Cumulative Distribution Function**

Binomial with \( n = 15 \) and \( p = 0.2 \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P( X &lt;= x ) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.999215</td>
</tr>
</tbody>
</table>

Conclusion: Probability that at most 8 fail the test is .9992

Calc > Probability Distributions > Binomial (n=15, p=0.2, pmf, x=8)

**Probability Density Function**

Binomial with \( n = 15 \) and \( p = 0.2 \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P( X = x ) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.0034548</td>
</tr>
</tbody>
</table>

Conclusion: Probability that exactly 8 fail the test is .003

Calc > Probability Distributions > Binomial (n=15, p=0.2, inverse pmf, prob=.5)

**Inverse Cumulative Distribution Function**

Binomial with \( n = 15 \) and \( p = 0.2 \)

<table>
<thead>
<tr>
<th>( x )</th>
<th>( P( X &lt;= x ) )</th>
<th>( x )</th>
<th>( P( X &lt;= x ) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.398023</td>
<td>3</td>
<td>0.648162</td>
</tr>
</tbody>
</table>

Conclusion: Median is 3. There is at least a 50% chance that <=3 fail the test, and there is at least a 50% chance that >=3 fail the test.
Example 3.40

Calc > Probability Distributions > Poisson (Probability, mean=2, input constant 1)

**Probability Density Function**

Poisson with mean = 2

\[\begin{array}{c|c}
x & P( X = x ) \\
1 & 0.270671 \\
\end{array} \]

Conclusion: Probability that exactly one error is found would be 27%

Calc > Probability Distributions > Poisson (Cumulative probability, mean=2, input constant 3)

**Cumulative Distribution Function**

Poisson with mean = 2

\[\begin{array}{c|c}
x & P( X \leq x ) \\
3 & 0.857123 \\
\end{array} \]

Conclusion: Probability that at most three errors are found would be 85.7%

Calc > Probability Distributions > Poisson (Inverse cumulative probability, mean=2, input constant 0.75)

**Inverse Cumulative Distribution Function**

Poisson with mean = 2

\[\begin{array}{c|c|c}
x & P( X \leq x ) & x & P( X \leq x ) \\
2 & 0.676676 & 3 & 0.857123 \\
\end{array} \]

Conclusion: Third quantile Q3 is equal to 3. There is at least a 75% chance that \( \leq 3 \) errors are found, and there is at least a 25% chance that \( \geq 3 \) errors are found.
Example 4.16

Calc > Probability Distributions > Normal (Cumulative probability, mean 1.25, standard deviation 0.46, input constant 1.00)

Cumulative Distribution Function

Normal with mean = 1.25 and standard deviation = 0.46

<table>
<thead>
<tr>
<th>x</th>
<th>P( X &lt;= x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.293400</td>
</tr>
</tbody>
</table>

Calc > Probability Distributions > Normal (Cumulative probability, mean 1.25, standard deviation 0.46, input constant 1.75)

Cumulative Distribution Function

Normal with mean = 1.25 and standard deviation = 0.46

<table>
<thead>
<tr>
<th>x</th>
<th>P( X &lt;= x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.75</td>
<td>0.861472</td>
</tr>
</tbody>
</table>

Conclusion: Then \( P(1.00 < X < 1.75) = 0.861 - 0.293 = 0.568 \) so the probability that reaction time is between 1.00 and 1.75 seconds is 56.8%.

Calc > Probability Distributions > Normal (Cumulative probability, mean 1.25, standard deviation 0.46, input constant 2.00)

Cumulative Distribution Function

Normal with mean = 1.25 and standard deviation = 0.46

<table>
<thead>
<tr>
<th>x</th>
<th>P( X &lt;= x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.948495</td>
</tr>
</tbody>
</table>

Conclusion: Then \( P(X > 2) = 1 - 0.948 = 0.052 \) so the probability that reaction time exceeds 2.0 seconds is 5.2%.

Calc > Probability Distributions > Normal (Inverse cumulative probability, mean 1.25, standard deviation 0.46, input constant 0.99)

Inverse Cumulative Distribution Function

Normal with mean = 1.25 and standard deviation = 0.46

<table>
<thead>
<tr>
<th>P( X &lt;= x )</th>
<th>x</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>2.32012</td>
</tr>
</tbody>
</table>

Conclusion: 99th percentile is 2.32. There is a 99% chance that reaction time is less than 2.32 seconds.
Example 4.24

Calc > Probability Distributions > Gamma (Cumulative probability, shape=8, scale=15, input constant 60)

Cumulative Distribution Function

Gamma with shape = 8 and scale = 15

\[ x \quad P( X \leq x ) \]

60 \hspace{1cm} 0.0511336

Calc > Probability Distributions > Gamma (Cumulative probability, shape=8, scale=15, input constant 120)

Cumulative Distribution Function

Gamma with shape = 8 and scale = 15

\[ x \quad P( X \leq x ) \]

120 \hspace{1cm} 0.547039

Conclusion: Then \( P(60<X<120) = .547 - .051 = .496 \) so the probability that a mouse survives between 60 and 120 weeks is 49.6%.

Calc > Probability Distributions > Gamma (Cumulative probability, shape=8, scale=15, input constant 30)

Cumulative Distribution Function

Gamma with shape = 8 and scale = 15

\[ x \quad P( X \leq x ) \]

30 \hspace{1cm} 0.0010967

Conclusion: Then \( P(X>30) = 1-.001 = .999 \) so the probability that a mouse survives at least 30 weeks is 99.9%.

Calc > Probability Distributions > Gamma (Inverse cumulative probability, shape=8, scale=15, input constant 0.25)

Inverse Cumulative Distribution Function

Gamma with shape = 8 and scale = 15

\[ P( X \leq x ) \quad x \]

0.25 \hspace{1cm} 89.3416

Conclusion: First quantile \( Q_1 \) is equal to 89.3. There is a 25% chance that a mouse will survive less than 89.3 weeks.
Example 4.25

Calc > Probability Distributions > Weibull (Cumulative probability, shape=2, scale=10, input constant 10)

**Cumulative Distribution Function**

Weibull with shape = 2 and scale = 10

\[
\begin{array}{c|c}
  x & P( X \leq x ) \\
  10 & 0.632121 \\
\end{array}
\]

Conclusion: There is a 63% chance that nitrous oxide emissions are less than 10.

Calc > Probability Distributions > Weibull (Cumulative probability, shape=2, scale=10, input constant 25)

**Cumulative Distribution Function**

Weibull with shape = 2 and scale = 10

\[
\begin{array}{c|c}
  x & P( X \leq x ) \\
  25 & 0.998070 \\
\end{array}
\]

Conclusion: There is a 99.8% chance that nitrous oxide emissions are less than 25.

Calc > Probability Distributions > Weibull (Inverse cumulative probability, shape=2, scale=10, input constant 0.95)

**Inverse Cumulative Distribution Function**

Weibull with shape = 2 and scale = 10

\[
\begin{array}{c|c}
  P( X \leq x ) & x \\
  0.95 & 17.3082 \\
\end{array}
\]

Conclusion: 95% of NOx emissions are less than 17.3.
Example 4.30

Graph > Probability Plot (Single, select data)

Results for: exp4-30.mtw

Probability Plot of X(1):

Conclusion: The distribution of dialectric breakdown voltage data appears to fit a normal distribution with mean 27.79 and standard deviation 1.462.
Example 7.11

Stat > Basic Statistics > One sample-t (select column C1)

Results for: EXP07-11.MTW

One-Sample T: rupture

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>rupture</td>
<td>30</td>
<td>7203.2</td>
<td>543.5</td>
<td>99.2</td>
<td>(7000.2, 7406.2)</td>
</tr>
</tbody>
</table>

Conclusion: 95% sure that average strength is between 7000 and 7406 psi.

Graph > Probability Plot (Single, select data)

Conclusion: OK, data appears normal, t-test procedure is valid.
Example 8.9

Stat > Basic Statistics > One sample-t > Options (Alternative: not equal)

One-Sample T: conc

Test of mu = 4 vs not = 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95% CI</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>conc</td>
<td>5</td>
<td>3.814</td>
<td>0.718</td>
<td>0.321</td>
<td>(2.922, 4.706)</td>
<td>-0.58</td>
<td>0.594</td>
</tr>
</tbody>
</table>

Calculation: Reject $H_0$ at the 95% level if $T > t_{0.025,4} = 2.776$. Since $T = -0.58$, do not reject. [Or, do not reject because the P-value is $P = 0.594 > .05$.]

Conclusion: Insufficient evidence to be 95% sure that average concentration differs from 4 mg/mL.

Graph > Probability Plot (Single, select data)

Conclusion: OK, data appears normal, t-test procedure is valid.
Example 8.11

Stat > Basic Statistics > 1 Proportion > (Number of events: 16, Number of trials: 91, Perform hypothesis test, Hypothesized proportion: 0.15, Options: Confidence level: 90, Alternative: greater than, Use test and interval based on normal distribution)

Test and CI for One Proportion

Test of p = 0.15 vs p > 0.15

<table>
<thead>
<tr>
<th>Sample</th>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>Bound</th>
<th>Z-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>91</td>
<td>0.175824</td>
<td>0.124684</td>
<td>0.69</td>
<td>0.245</td>
</tr>
</tbody>
</table>

Using the normal approximation.

Since the test statistic z = 0.69 is less than the 90% cut off value of 1.282, we do not reject the null hypothesis.

Conclusion: Insufficient evidence to be 90% sure that more than 15% of corks are bad.
Example 8.18

Stat > Basic Statistics > One sample-t > Options (Alternative: not equal)

One-Sample T: conc

Test of mu = 4 vs not = 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95% CI</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>conc</td>
<td>5</td>
<td>3.814</td>
<td>0.718</td>
<td>0.321</td>
<td>(2.922, 4.706)</td>
<td>-0.58</td>
<td>0.594</td>
</tr>
</tbody>
</table>

Conclusion: We are 41.6% sure that average concentration differs from 4 mg/mL. [There is insufficient evidence to conclude that average concentration differs from 4 mg/mL.]

Graph > Probability Plot (Single, select data)

Conclusion: OK, data appears normal, t-test procedure is valid.
Example 9.9

Stat > Basic Statistics > One sample-t > Options (level, alternative)

One-Sample T: Differen

Test of μ = 0 vs not = 0

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95% CI</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differen</td>
<td>16</td>
<td>6.7500</td>
<td>8.2340</td>
<td>2.0585</td>
<td>(2.36237, 11.13763)</td>
<td>3.28</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Conclusion: We are 99.5% sure that the average proportion of time at which arm angle is less than 30 degrees has changed after work conditions were changed (since p=.005). Also, we are 95% sure that the average proportion of time at which arm angle is less than 30 degrees, after the change in working conditions, is between 2% and 11% less than it was before (based on the 95% CI).

Graph > Probability Plot (Single, select data)

Conclusion: OK, data appears normal, t-test procedure is valid.
Alternative procedure based on two-sample t-test

Stat > Basic Statistics > 2 sample-t > Options (level, alternative)

**Two-Sample T-Test and CI: Before:, After:**

Two-sample T for Before: vs After:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before:</td>
<td>16</td>
<td>78.63</td>
<td>9.89</td>
<td>2.5</td>
</tr>
<tr>
<td>After:</td>
<td>16</td>
<td>71.9</td>
<td>10.9</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Difference = mu (Before:) - mu (After:)

Estimate for difference: 6.75000

95% CI for difference: (-0.75711, 14.25711)

T-Test of difference = 0 (vs not =): T-Value = 1.84  P-Value = 0.076  DF = 29

Conclusion: We are 92.4% sure that the average proportion of time at which arm angle is less than 30 degrees has changed after work conditions were changed (since p=.076). Also, we are 95% sure that the average amount of this change is between -0.7% and +14.3% (based on the 95% CI). Note that the paired t-test gives more definitive results!

Graph > Probability Plot (Single, select data)

Conclusion: OK, each data set appears normal, 2-sample t-test procedure is valid.
Example 9.11

Stat > Basic Statistics > 2 Proportions (Options: Alternative: greater than, Use pooled estimate for p test)

Test and CI for Two Proportions

Sample  X   N  Sample p  
1     81  549  0.147541  
2    141  730  0.193151  

Difference = p (1) - p (2)  
Estimate for difference:  -0.0456097  
95% upper bound for difference:  -0.0110060  
Test for difference = 0 (vs < 0):  Z = -2.13  P-Value = 0.017  
Fisher’s exact test: P-Value = 0.019

Since p-value is 0.017 we reject the null hypothesis at level 0.01, but not at level 0.01.

Conclusion: We are 98.3% sure that aspirin use increases the 15 year survival rate for colorectal cancer victims.
Examples 12.1 and 12.2

Graph > Scatterplot

Conclusion: There appears to be a strong positive linear relation between \( y = \text{OSA} \) and \( x = \text{palprebal fissure width} \).

Conclusion: There appears to be a weak negative linear relation between \( y = \text{mean crown dieback} \) and \( x = \text{pH} \).
Example 12.4

Stat > Regression > Regression (Storage: Residuals)

Results for: EXP12-04.MTW

Regression Analysis: y versus x

The regression equation is
\[ y = 75.2 - 0.209 \, x \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>75.212</td>
<td>2.984</td>
<td>25.21</td>
<td>0.000</td>
</tr>
<tr>
<td>x</td>
<td>-0.20939</td>
<td>0.03109</td>
<td>-6.73</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 2.56450 \quad R-Sq = 79.1\% \quad R-Sq(adj) = 77.3\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>298.25</td>
<td>298.25</td>
<td>45.35</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>12</td>
<td>78.92</td>
<td>6.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>377.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion: The true mean cetane number \( y \) for diesel fuel with iodine value \( x \) is estimated to be
\[ y = 75.2 - 0.209 \, x \], with a typical spread of 2.6. This regression model explains 77.3\% of the variations in cetane number in terms of variations in iodine value.
Stat > Regression > Fitted Line Plot

Fitted Line Plot

\[ y = 75.21 - 0.2094 \times \]

\[ S = 2.56450 \]

\[ R-Sq = 79.1\% \]

\[ R-Sq(adj) = 77.3\% \]

Conclusion: The regression line provides a reasonably good fit to the data, indicating that \( x = \) iodine value can give a useful model to predict \( y = \) cetane number.

Probability Plot of RESI1

Scatterplot of RESI1 vs \( x \)

Conclusion: Probability plot (left) indicates that the residuals fit a normal distribution. Scatterplot (right) indicates the residual seem to be independent and identically distributed. Hence the basic assumptions of the regression model are satisfied.
Example 12.11

Stat > Regression > Regression (Storage: Residuals)

Results for: exp12-11.mtw

Regression Analysis: y: versus x:

The regression equation is
\[
y: = 126 - 0.918 \times:
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>126.249</td>
<td>2.254</td>
<td>56.00</td>
<td>0.000</td>
</tr>
<tr>
<td>x:</td>
<td>-0.9176</td>
<td>0.1460</td>
<td>-6.29</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ S = 2.94100 \quad R-Sq = 75.2\% \quad R-Sq(adj) = 73.3\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>341.73</td>
<td>341.73</td>
<td>39.51</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>13</td>
<td>112.44</td>
<td>8.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>454.17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>x:</th>
<th>y:</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>10.0</td>
<td>124.000</td>
<td>117.07</td>
<td>1.008</td>
<td>6.927</td>
<td>2.51R</td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual.

Conclusion: The true mean density for mortar with an air content of x % is estimated to be \( \mu = 126 - 0.918 \times \) lb/cu-ft, with a typical spread of 3 lb/cu-ft. This regression model explains 75.2% of the variations in density in terms of variations in air content.

The t statistic is -6.29 with a p-value of 0.000, so we are virtually certain that the regression model provides a useful predictor of mortar density in terms of air content.
Conclusion: The regression line provides a reasonably good fit to the data, indicating that $x$=air content can give a useful model to predict $y$=density.

Conclusion: Probability plot (left) indicates that the residuals fit a normal distribution. Scatterplot (right) indicates the residual seem to be independent and identically distributed. Hence the basic assumptions of the regression model are satisfied. Note: The p-value of 0.275 in the probability plot indicates that there is insufficient evidence to reject the null hypothesis of a normal fit.
Example 12.13

Stat > Regression > Regression (Storage: Residuals, Options: Prediction intervals for new observation: 45)

Results for: exp12-13.mtw

Regression Analysis: y: versus x:

The regression equation is

\[ y: = 27.2 - 0.298 \times \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>27.183</td>
<td>1.651</td>
<td>16.46</td>
<td>0.000</td>
</tr>
<tr>
<td>x:</td>
<td>-0.29756</td>
<td>0.04116</td>
<td>-7.23</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S = 2.86403  R-Sq = 76.6%  R-Sq(adj) = 75.1%

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>428.62</td>
<td>428.62</td>
<td>52.25</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>16</td>
<td>131.24</td>
<td>8.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>559.86</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Predicted Values for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>Fit</th>
<th>SE Fit</th>
<th>95% CI</th>
<th>95% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.793</td>
<td>0.758</td>
<td>(12.185, 15.400)</td>
<td>(7.512, 20.073)</td>
</tr>
</tbody>
</table>

Values of Predictors for New Observations

<table>
<thead>
<tr>
<th>New Obs</th>
<th>x:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>45.0</td>
</tr>
</tbody>
</table>

Conclusion: For cement samples with a carbonation depth of 45mm, we are 95% sure that the average strength is between 12.2 MPa and 15.4 MPa. For one sample of cement with a carbonation depth of 45mm, we are 95% sure that the strength of this individual sample is between 7.5 MPa and 20.1 MPa.

The regression model \( y = 27.2 - 0.298 \times \) gives the estimated average strength \( y \) in MPa for concrete samples with a given carbonation depth \( x \) in mm. The t-value of -7.23 and the corresponding p-value of 0.000 indicates strong evidence that there is a positive linear relation between these two variables.
Conclusion: The regression line provides a reasonably good fit to the data, indicating that \( x = \text{carbonation depth content} \) can give a useful model to predict \( y = \text{strength} \) with a typical error of 2.86 MPa. 77% of the variations in strength can be attributed to variations in carbonation depth.

Conclusion: Probability plot (left) indicates that the residuals fit a normal distribution. Scatterplot (right) indicates the residuals seem to be independent and identically distributed. Hence the basic assumptions of the regression model are satisfied. Note: The p-value of 0.242 in the probability plot indicates that there is insufficient evidence to reject the null hypothesis of a normal fit.
Example 12.16

Stat > Regression > Regression (Storage: Residuals)

Regression Analysis: y versus x

The regression equation is
\[ y = 1.00 + 93.4 \times x \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.998</td>
<td>2.703</td>
<td>0.37</td>
<td>0.717</td>
</tr>
<tr>
<td>x</td>
<td>93.38</td>
<td>24.36</td>
<td>3.83</td>
<td>0.002</td>
</tr>
</tbody>
</table>

\[ S = 3.89192 \quad R-Sq = 51.2\% \quad R-Sq(adj) = 47.7\% \]

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>222.48</td>
<td>222.48</td>
<td>14.69</td>
<td>0.002</td>
</tr>
<tr>
<td>Residual Error</td>
<td>14</td>
<td>212.06</td>
<td>15.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>434.54</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>x</th>
<th>y</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.074</td>
<td>16.600</td>
<td>7.908</td>
<td>1.210</td>
<td>8.692</td>
<td>2.35R</td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual.

Stat > Basic Statistics > Correlation

Correlations: x, y

Pearson correlation of x and y = 0.716
P-Value = 0.002

Conclusion: The correlation of .716 indicates a strong positive relation between ozone and carbon concentrations. The p-value of 0.002 indicates we are 99.8% sure there is a linear relation between ozone and carbon concentrations. Note that the p-values for the slope and the correlation are identical, this is always the case!

The regression model \[ y = 1.00 + 93.4 \times x \] gives the estimated average carbon concentration in \( \mu g/mm^3 \) for air samples with a given ozone concentration \( x \) ppm. Variations in ozone concentration account for 51% of the variations in carbon concentration.
Conclusion: The regression line provides a reasonably good fit to the data, indicating that $x=\text{ozone concentration}$ can give a useful model to predict $y=\text{carbon concentration}$ with a typical error of $3.89 \mu g/mm^3$.

Conclusion: Probability plot (left) indicates that the residuals fit a normal distribution. Scatterplot (right) indicates the residuals seem to be independent and identically distributed. Hence the basic assumptions of the regression model are satisfied.