Read →

12.11

In[52]:=
```
mortarair =
  {5.7, 6.8, 9.6, 10.0, 10.7, 12.6, 14.4, 15.0, 15.3, 16.2, 17.8, 18.7, 19.7, 20.6, 25.0};
dryden = {119, 121.3, 118.2, 124.0, 112.3, 114.1, 112.2,
   115.1, 111.3, 107.2, 108.9, 107.8, 111.0, 106.2, 105.0};
Length[
 mortarair]
```

These are the X-Values

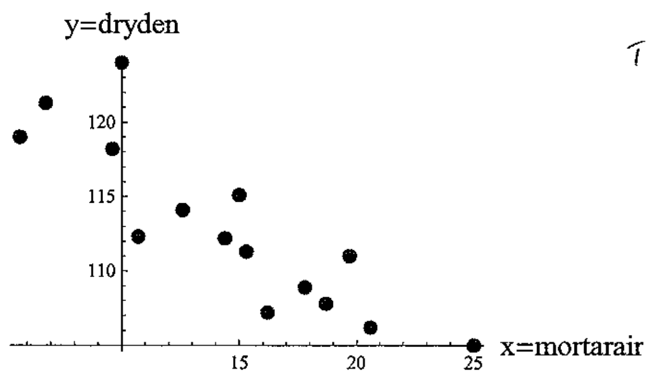
dryden are the y values

length gives the number of values within mortarair

In[63]:=
```
15
ListPlot[Table[{mortarair[[i]], dryden[[i]]}, {i, 1, 15}],
 AxesLabel → {"x=mortarair", "y=dryden"}, PlotStyle -> PointSize[0.03]]
```

Out[63]= 15

Out[64]=



This plots the x and y data

In[66]:=
```
xmortarair = Table[{1, mortarair[[i]]}, {i, 1, 15}];
MatrixForm[xmortarair]
```

- This shows mortarair (x values) with the 1^{st} column all ones in matrix form. This is known as the design Matrix. It is presented in this fashion to take the dot product with dryden to find $\hat{\beta}_0$ and $\hat{\beta}_1$.

This probability model stated in matrix form is $y = x.\beta + \epsilon$

Out[67]//MatrixForm=

$$\begin{pmatrix} 1 & 5.7 \\ 1 & 6.8 \\ 1 & 9.6 \\ 1 & 10. \\ 1 & 10.7 \\ 1 & 12.6 \\ 1 & 14.4 \\ 1 & 15. \\ 1 & 15.3 \\ 1 & 16.2 \\ 1 & 17.8 \\ 1 & 18.7 \\ 1 & 19.7 \\ 1 & 20.6 \\ 1 & 25. \end{pmatrix}$$

$(X^{tr}X)^{-1}X^{tr}$ design Matrix y

In[104]:=
```
betahatmortar = PseudoInverse[xmortarair].dryden
```

Out[104]= {126.249, -0.917622}

$\hat{\beta}_0$ $\hat{\beta}_1$

PseudoInverse is the least squares solver applicable to systems of linear equations. It produces the unique solution of simultaneous linear equations in several variables. The least squares solver uses Pseudo Inverse to find a least squares solution $\hat{\beta}_0$ and $\hat{\beta}_1$ by taking the dot product of xmortarair and dryden.

$\hat{\beta}_0$ and $\hat{\beta}_1$ represent the slope and the intercepts

In[107]:=
```
Beta0hat = 126.249;
Beta1hat = -.917622
```

$\hat{\beta}_1 = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$ $\sqrt{\overline{y^2} - \bar{y}^2}$ $\sqrt{\overline{x^2} - \bar{x}^2}$

$\bar{x}, \bar{y}$ is on L.S line if $\dfrac{y - \bar{y}}{x - \bar{x}} = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$

Out[108]= -0.917622

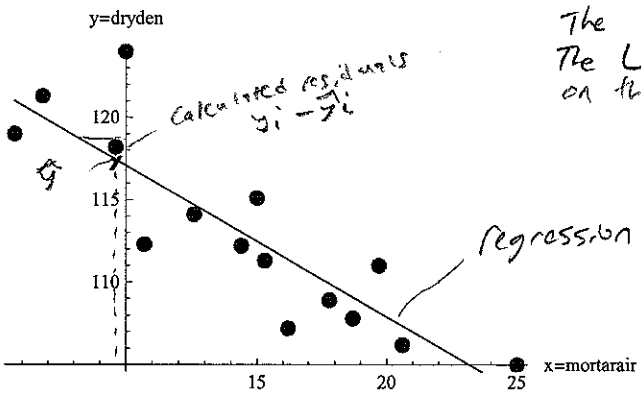
In[69]:=
```
drydenhat = xmortarair.betahatmortar
```

From $\hat{\beta}$ mortar, The slope and y intercept are used to find $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Out[69]= {121.018, 120.009, 117.44, 117.073, 116.43, 114.687, 113.035, 112.485, 112.209, 111.383, 109.915, 109.089, 108.172, 107.346, 103.308}

In[70]:=
```
Show[ListPlot[Table[{mortarair[[i]], dryden[[i]]}, {i, 1, 15}],
  AxesLabel → {"x=mortarair", "y=dryden"}, PlotStyle -> PointSize[0.03]],
 Graphics[Line[Table[{mortarair[[i]], drydenhat[[i]]}, {i, 1, 15}]]]]
```

Out[70]=



The Line joins the points $(x_i, \hat{y}_i)$. The LS produces fitted values on the line.

$\hat{\beta}_0$ is indeed the y-intercept at 126.249

$\hat{\beta}_1$ is the slope at -.917622.

The heights of the regression line are the previously calculated fitted values. Also, the residuals are the signed vertical gaps between the points of the plot and regression line

In[71]:=
```
drydenresid = dryden - drydenhat
```

Out[71]= {-2.01844, 1.29094, 0.760281, 6.92733, -4.13033, -0.586853, -0.835134, 2.61544, -0.909274, -4.18341, -1.01522, -1.28936, 2.82826, -1.14588, 1.69166}

This pulls off the residuals $y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ left by least squares

In[87]:= Inverse[Transpose[xmortarair].xmortarair] $\dfrac{15-1}{15-2}$ (s[drydenresid])^2

Out[87]= {{5.0824, -0.309887}, {-0.309887, 0.0213128}}

The least squares estimators of intercept and slope have variances and covariances that are estimated by entries of the matrix.

In[90]:=

sigmahatsquared = $\dfrac{15-1}{15-2}$ (s[drydenresid])^2

$\hat{\sigma}^2 = \dfrac{n-1}{n-d} s^2_{\text{Residuals}}$ d = # of columns of the design matrix (d = 2)

Out[90]= 8.64948

In[79]:=
```
MatrixForm[%]
```

Out[79]//MatrixForm=

$$\begin{array}{c} \hat{\beta}_0 \\ \begin{pmatrix} 5.0824 & -0.309887 \\ -0.309887 & 0.0213128 \end{pmatrix} \end{array}$$

Covarience of $\hat{\beta}_0$ and $\hat{\beta}_1$

Var $\hat{\beta}_1$

In[124]:=
```
variencebetahat1 = 0.02131275243772429
```

Out[124]= 0.0213128

In[125]:=
```
standarderror = Sqrt[variencebetahat1]
```

Out[125]= 0.145989

$s_{\hat{\beta}_1} = \dfrac{s}{\sqrt{S_{xx}}} = \sqrt{\text{var}\,\hat{\beta}_1}$. This is used in the calculation of confidence interval degree Freedom $n-2 = 15-2 = 13$

In[128]:= **tscore = t[13, .95]**

Out[128]= 2.16037

95% CI. $\alpha = .025$

This use of t results in an exact CI provided the measurements (x, y) are from a process under statistical control.

In[129]:= **ci = {Beta1hat - tscore * standarderror, Beta1hat + tscore * standarderror}**

Out[129]= {-1.23301, -0.602232}

$ci = \hat{\beta}_1 \pm (t)\cdot(S_{\hat{\beta}_1})$

for 95% CI, $\alpha = .025$

This CI agrees with the one reported on page 472.

In[80]:= **mean[(dryden - xmortarair.{126.248889, -0.917622})^2]**

Out[80]= 7.49622

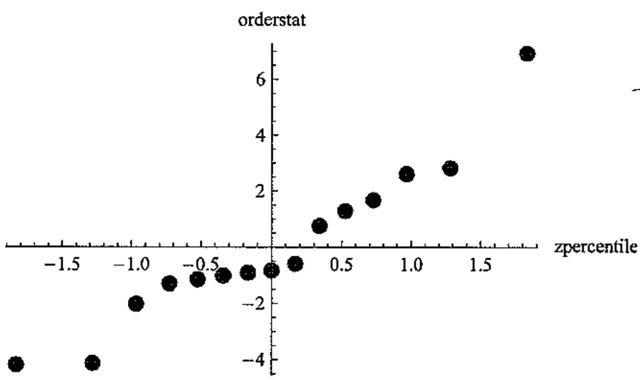
In[81]:= **mean[(dryden - xmortarair.{126.32776, -0.913876})^2]**

Out[81]= 7.51438

There is a minor disagreement between my mean and the book mean. Since the Mathematica residuals' mean of squares are less than SAS residuals' SAS must be in error (least squares not achieved by SAS fit)

In[82]:= **normalprobabilityplot [drydenresid, 0.03]**

Out[82]=



This is the normal probability plot for the residuals to give a partial check on the normal errors assumption of the probability model.

In[122]:= **r[mortarair, dryden]**

Out[122]= -0.867421

The correlation between the independent variable mortarair and the dependent variable dryden.

In[123]:= **%^2**

Out[123]= 0.75242

Squaring it gives the coefficient of determination or "the fraction of var Y accounted for by regression on x".

Want to be close to 1. It is interpereted as fraction of $\overline{y_i - \bar{y}}^2$ Accounted for by Regression in x

12.4

In[49]:= `mortarair = {99, 101.1, 102.7, 103.0, 105.4, 107.0, 108.7, 110.8, 112.1, 112.4, 113.6, 113.8, 115.1, 115.4, 120.0};` *X values*

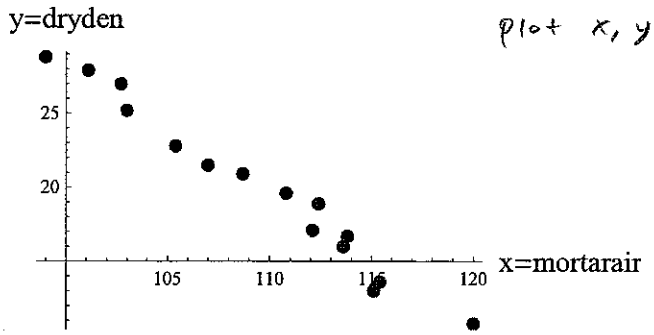
In[52]:= `dryden = {28.8, 27.9, 27, 25.2, 22.8, 21.5, 20.9, 19.6, 17.1, 18.9, 16.0, 16.7, 13, 13.6, 10.8};` *Y values*
`Length[mortarair]`

Out[53]= 15 *Length of terms*

In[55]:= `ListPlot[Table[{mortarair[[i]], dryden[[i]]}, {i, 1, 15}], AxesLabel → {"x=mortarair", "y=dryden"}, PlotStyle -> PointSize[0.03]]`

plot x, y of mortarair, dryden

Out[55]=



In[57]:= `xmortarair = Table[{1, mortarair[[i]]}, {i, 1, 15}];`

In[58]:= `MatrixForm[xmortarair]`

Out[58]//MatrixForm=

$$\begin{pmatrix} 1 & 99 \\ 1 & 101.1 \\ 1 & 102.7 \\ 1 & 103. \\ 1 & 105.4 \\ 1 & 107. \\ 1 & 108.7 \\ 1 & 110.8 \\ 1 & 112.1 \\ 1 & 112.4 \\ 1 & 113.6 \\ 1 & 113.8 \\ 1 & 115.1 \\ 1 & 115.4 \\ 1 & 120. \end{pmatrix}$$

design matrix with 1st column 1's in order to take the dot product with dryden to find $\hat{\beta}_0$ and $\hat{\beta}_1$.

design matrix

In[59]:= `betahatmortar = PseudoInverse[xmortarair].dryden`

Out[59]= {118.91, -0.904731}
$\hat{\beta}_0$ $\hat{\beta}_1$

least squares solver application that takes the dot product of the design matrix and dryden

In[60]:= `Beta0hat = 118.91`

Out[60]= 118.91

In[61]:= `Beta1hat = -.904731`

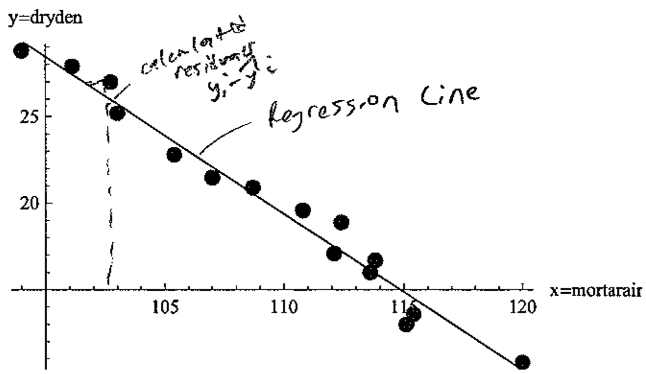
Out[61]= -0.904731

In[62]:= **drydenhat = xmortarair.betahatmortar** $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Out[62]= {29.3416, 27.4416, 25.9941, 25.7227, 23.5513, 22.1037, 20.5657, 18.6658, 17.4896, 17.2182, 16.1325, 15.9516, 14.7754, 14.504, 10.3422}

In[63]:=
```
Show[ListPlot[Table[{mortarair[[i]], dryden[[i]]}, {i, 1, 15}],
  AxesLabel → {"x=mortarair", "y=dryden"}, PlotStyle -> PointSize[0.03]],
 Graphics[Line[Table[{mortarair[[i]], drydenhat[[i]]}, {i, 1, 15}]]]]
```

Out[63]=



$\hat{\beta}_0$ is the y-intercept
$\hat{\beta}_1$ is the slope
The heights of the regression line are the previously calculated fitted values residuals are signed vertical gaps between the points of the plot and the regression line.

In[64]:= **drydenresid = dryden - drydenhat**   => pulls out the residuals $y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Out[64]= {-0.541582, 0.458353, 1.00592, -0.522659, -0.751305, -0.603736, 0.334306, 0.93424, -0.38961, 1.68181, -0.132514, 0.748432, -1.77542, -0.903999, 0.457762}

In[68]:= **Inverse[Transpose[xmortarair].xmortarair]** $\dfrac{15-1}{15-2}$ **(s[drydenresid])**^{2}

Out[68]= {{20.2421, -0.184593}, {-0.184593, 0.00168825}}

In[69]:= **MatrixForm[%]**

Out[69]//MatrixForm= $\hat{\beta}_0$ Covarience of $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\begin{pmatrix} 20.2421 & -0.184593 \\ -0.184593 & 0.00168825 \end{pmatrix}$$

Var $\hat{\beta}_1$

In[66]:= **sigmahatsquared =** $\dfrac{15-1}{15-2}$ **(s[drydenresid])**^{2}

$\hat{\sigma}^2 = \dfrac{n-1}{n-d} s^2_{residuals}$

Out[66]= 0.87991

In[70]:= **variencebetahat1 = 0.00168825**

Out[70]= 0.00168825

In[71]:= **standarderror = Sqrt[variencebetahat1]** $s_{\beta_1} = \dfrac{s}{\sqrt{s_{xx}}} = \sqrt{\mathrm{var}\,\hat{\beta}}$ => used in calculation of Ci

Out[71]= 0.0410883

degrees Freedom $(n-2) = 15-2 = 13$

In[72]:= **tscore = t[13, .95]**

Out[72]= 2.16037

In[73]:= **ci = {Beta1hat - tscore * standarderror, Beta1hat + tscore * standarderror}**

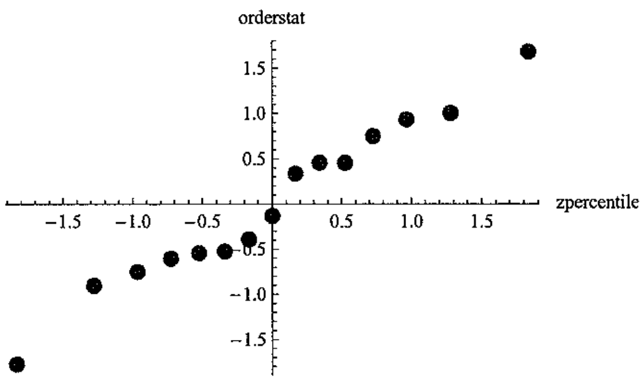
Out[73]= {-0.993497, -0.815965}

$C_i = \hat{\beta}_1 + (t)\cdot(s\beta_1)$

In[76]:= `mean[(dryden - xmortarair.{118.91, -0.9047306579482643})^2]`

Out[76]= 0.762589

In[77]:= `normalprobabilityplot[drydenresid, 0.03]`

Out[77]=


In[78]:= `r[mortarair, dryden]` => the regular correlation between the y's observed and the fitted values $y_i, \hat{y}_i$.

Out[78]= -0.986857

In[79]:= `%^2`

Out[79]= 0.973887 $0 < r^2 < 1$

↳ closer to one, is a better correlation

R is interpreted as a fraction of $y^2 - \bar{y}^2$

accounted for by the regression x