A413

HW8 Key

1. You will be expected to be able to solve this type of problem without assistance on an exam. A sequence of number pairs (x_i, y_i), $i \le 100$ has

$\overline{x} = 8$ $\quad \overline{x^2} = 80$

$\overline{y} = 10$ $\quad \overline{y^2} = 136$ $\quad \overline{xy} = 96$ [redacted]

Express in formula, evaluated formula, and reduced answer:

a. S_x (n-1 divisor sample sd of x scores)

$S_x = \sqrt{\frac{n}{n-1}} \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{\frac{100}{99}} \sqrt{80-64} = 4\sqrt{\frac{100}{99}}$

b. $\hat{\sigma}_x$ (n-divisor sample sd of x scores)

$\hat{\sigma}_x = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{80-64} = \boxed{4}$

c. $\hat{\sigma}_y$

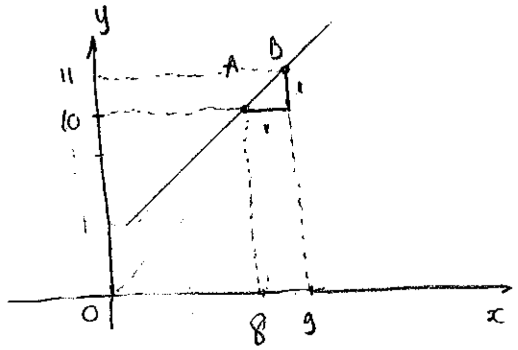
$\hat{\sigma}_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{136-100} = \boxed{6}$

d. $r = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\hat{\sigma}_x * \hat{\sigma}_y} = \dfrac{96-80}{24} = \dfrac{16}{24} = \boxed{.667}$

e. Slope $\hat{\beta}_1$ of estimated line of regression of y on x.

$\hat{\beta}_1 = \dfrac{r\hat{\sigma}_y}{\hat{\sigma}_x} = \dfrac{2}{3}\dfrac{6}{4} = \boxed{1}$

f. Sketch the line of regression identifying two points on that line with their formula and numerical values.



A (8, 10)
B (9, 11)

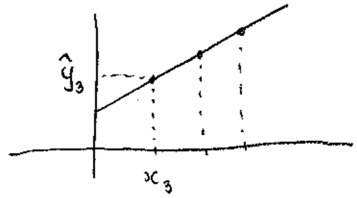
$\boxed{y = x + 2}$ $(\hat{\beta}_1 = 1 \ \hat{\beta}_0 = 2)$

g. From (d), determine the fraction of $\overline{y^2}-\overline{y}^2$ explained by regression of y on x.

fraction of $\overline{y^2} - \bar{y}^2 = r^2 = \boxed{\frac{4}{9}}$

h. Express the multiple correlation r_{MLR} mathematically in terms of y_i values and fitted values $\hat{y}_i$, Determine its value from r.

$r_{MLR} = r_{y,\hat{y}} = |r| = \boxed{.667}$

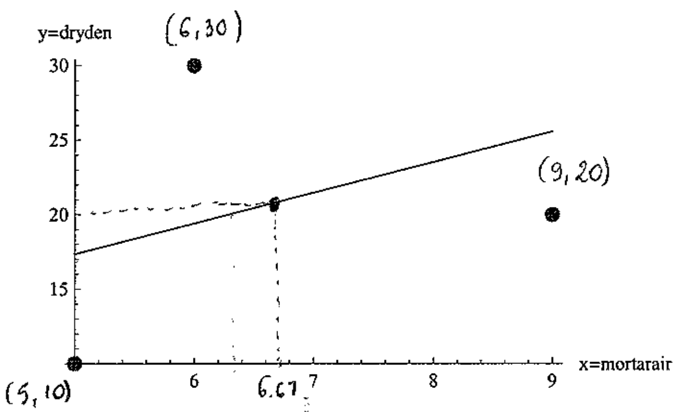


$\hat{y}_1, \hat{y}_2, \cdots \hat{y}_n$ fitted values

$\underbrace{r[y, \hat{y}]}_{r_{MLR}} = r[x, y]$

$(r_{\hat{y}, y} = r_{\hat{\beta}_0 + \hat{\beta}_1 x, y})$

2. For the scatterplot and its regression of y on x, given below, determine the requested quantities. You will need a calculator.



i	y_i	x_i
1	10	5
2	30	6
3	20	9

n = 3

a. Determine $(\overline{x}, \overline{y})$ and identify it on the regression line.

$\bar{x} = (5 + 6 + 9)/3 = 6.67$

$\bar{y} = (10 + 20 + 30)/3 = 20$

b. Determine the slope of the regression line from formula calculation and observe that the picture is correct on that.

$\overline{xy} = \frac{410}{3}, \quad \overline{x^2} = \frac{142}{3}, \quad \bar{x}^2 = \frac{400}{9}, \quad \overline{y^2} = \frac{1400}{3}, \quad \bar{y}^2 = 400$

$$\hat{\beta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\frac{410}{3} - \frac{20}{3} * 20}{\frac{142}{3} - \frac{400}{9}} = \frac{\frac{10}{3}}{\frac{26}{9}} = \frac{10}{3} * \frac{9}{26} = \boxed{\frac{15}{13}}$$

c. Determine the fitted value for the second point from the left.

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 20 - \frac{15}{13} * \frac{20}{3} = \frac{160}{13}$

$\hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1 x_2 = \frac{160}{13} + \frac{15}{13} * 6 = \boxed{\frac{250}{13}}$

d. Determine the correlation r between x and y scores.

$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2} * \sqrt{\overline{y^2} - \bar{y}^2}} = \frac{10/3}{\sqrt{26/9} * \sqrt{\frac{1400}{3} - \frac{1200}{3}}} = \frac{10/3}{\frac{\sqrt{26}}{3} * \sqrt{\frac{200}{3}}} = \boxed{.24}$

e. Determine the correlation r_{MLR} between fitted scores and y scores and verify that indeed it is equal to | r | where r is the correlation from (d).

$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 x_1 = \frac{160}{13} + \frac{15}{13} * 5 = \frac{235}{13}$; $\hat{y}_3 = \hat{\beta}_0 + \hat{\beta}_1 x_3 = \frac{160}{13} + \frac{15}{13} * 9 = \frac{295}{13}$

$\Rightarrow \bar{\hat{y}} = \frac{780}{39} = 20 \quad \overline{\hat{y}^2} = \frac{204750}{507} \Rightarrow \sigma_{\hat{y}} = \sqrt{\overline{\hat{y}^2} - \bar{\hat{y}}^2} = 1.96$, $r_{MLR} = \frac{\overline{y\hat{y}} - \bar{y}\bar{\hat{y}}}{\sigma_y * \sigma_{\hat{y}}} = \frac{403.8 - 20 * 20}{\sqrt{\frac{200}{3}} * 1.96} = \boxed{0.24} = |r|$

d. Determine the slope of regression.

I still dont understand what you are asking here

Is this the same as the $\hat{\beta}_1$ that we found in b) which equals $\boxed{\frac{15}{13}}$???

e. Give the 3 by 2 design matrix xmortarair for a matrix formulation of this LR.

$\begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 1 & 9 \end{bmatrix}$

designx = {{1,5}, {1,6}, {1,9}}

MatrixForm [%]

f. What regression quantities are produced by the *Mathematica* codes below?

PseudoInverse[xmortarair].dryden

$\hat{\beta}_0 = \frac{160}{13}$

$\hat{\beta}_1 = \frac{15}{13}$

Inverse[Transpose[xmortarair].xmortarair] $\frac{3-1}{3-2} s[\text{drydenresid}]^2$

Variance of $\hat{\beta}_0$

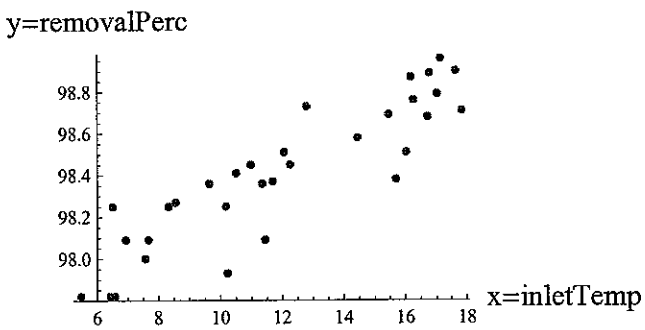
Variance of $\hat{\beta}_1$

Covariance of $\hat{\beta}_0$ with $\hat{\beta}_1$

Problem 84

```
In[32]:= inletTemp = { 7.68, 6.51, 6.43, 5.48, 6.57, 10.22, 15.69, 16.77, 17.13,
    17.63, 16.72, 15.45, 12.06, 11.44, 10.17, 9.64, 8.55, 7.57, 6.94, 8.32, 10.50,
    16.02, 17.83, 17.03, 16.18, 16.26, 14.44, 12.78, 12.25, 11.69, 11.34, 10.97 };

In[33]:= removalPerc = { 98.09, 98.25, 97.82, 97.82, 97.82, 97.93, 98.38, 98.89, 98.96, 98.90,
    98.68, 98.69, 98.51, 98.09, 98.25, 98.36, 98.27, 98.00, 98.09, 98.25, 98.41,
    98.51, 98.71, 98.79, 98.87, 98.76, 98.58, 98.73, 98.45, 98.37, 98.36, 98.45 };

In[34]:= Length[inletTemp]

Out[34]= 32

In[35]:= ListPlot[Table[{inletTemp[[i]], removalPerc[[i]]}, {i, 1, 32}],
    AxesLabel → {"x=inletTemp", "y=removalPerc"}, PlotStyle -> PointSize[0.02]]
```

Out[35]=




```
In[36]:= xinletTemp = Table[{1, inletTemp[[i]]}, {i, 1, 32}];
```

```
MatrixForm[xinletTemp]
```

1	7.68
1	6.51
1	6.43
1	5.48
1	6.57
1	10.22
1	15.69
1	16.77
1	17.13
1	17.63
1	16.72
1	15.45
1	12.06
1	11.44
1	10.17
1	9.64
1	8.55
1	7.57
1	6.94
1	8.32
1	10.5
1	16.02
1	17.83
1	17.03
1	16.18
1	16.26
1	14.44
1	12.78
1	12.25
1	11.69
1	11.34
1	10.97

PseudoInverse is a least squares solver applicable to systems of linear equations. It produces the unique solution of simultaneous linear equations in several variables (such as the normal equations of Least Squares) if there is one. If not, it produces a particular choice of a least squares solution known as the Moore - Penrose Inverse. In the present example, the matrix formulation of the equations of our linear model is

$$\begin{pmatrix} 98.09 \\ 98.25 \\ 97.82 \\ 97.82 \\ 97.82 \\ 97.93 \\ 98.38 \\ 98.89 \\ 98.96 \\ 98.9 \\ 98.68 \\ 98.69 \\ 98.51 \\ 98.09 \\ 98.25 \\ 98.36 \\ 98.27 \\ 98. \\ 98.09 \\ 98.25 \\ 98.41 \\ 98.51 \\ 98.71 \\ 98.79 \\ 98.87 \\ 98.76 \\ 98.58 \\ 98.73 \\ 98.45 \\ 98.37 \\ 98.36 \\ 98.45 \end{pmatrix} \sim \begin{pmatrix} 1 & 7.68 \\ 1 & 6.51 \\ 1 & 6.43 \\ 1 & 5.48 \\ 1 & 6.57 \\ 1 & 10.22 \\ 1 & 15.69 \\ 1 & 16.77 \\ 1 & 17.13 \\ 1 & 17.63 \\ 1 & 16.72 \\ 1 & 15.45 \\ 1 & 12.06 \\ 1 & 11.44 \\ 1 & 10.17 \\ 1 & 9.64 \\ 1 & 8.55 \\ 1 & 7.57 \\ 1 & 6.94 \\ 1 & 8.32 \\ 1 & 10.5 \\ 1 & 16.02 \\ 1 & 17.83 \\ 1 & 17.03 \\ 1 & 16.18 \\ 1 & 16.26 \\ 1 & 14.44 \\ 1 & 12.78 \\ 1 & 12.25 \\ 1 & 11.69 \\ 1 & 11.34 \\ 1 & 10.97 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Because the points (x, y) do not fall on a line (see the plot above) there is no exact solution of these 3 equations in only 2 unknowns. The least squares solver uses a Pseudo-Inverse to find a Least Squares solution

$$\begin{pmatrix} 1 & 7.68 \\ 1 & 6.51 \\ 1 & 6.43 \\ 1 & 5.48 \\ 1 & 6.57 \\ 1 & 10.22 \\ 1 & 15.69 \\ 1 & 16.77 \\ 1 & 17.13 \\ 1 & 17.63 \\ 1 & 16.72 \\ 1 & 15.45 \\ 1 & 12.06 \\ 1 & 11.44 \\ 1 & 10.17 \\ 1 & 9.64 \\ 1 & 8.55 \\ 1 & 7.57 \\ 1 & 6.94 \\ 1 & 8.32 \\ 1 & 10.5 \\ 1 & 16.02 \\ 1 & 17.83 \\ 1 & 17.03 \\ 1 & 16.18 \\ 1 & 16.26 \\ 1 & 14.44 \\ 1 & 12.78 \\ 1 & 12.25 \\ 1 & 11.69 \\ 1 & 11.34 \\ 1 & 10.97 \end{pmatrix}^{-1} \begin{pmatrix} 98.09 \\ 98.25 \\ 97.82 \\ 97.82 \\ 97.82 \\ 97.93 \\ 98.38 \\ 98.89 \\ 98.96 \\ 98.9 \\ 98.68 \\ 98.69 \\ 98.51 \\ 98.09 \\ 98.25 \\ 98.36 \\ 98.27 \\ 98. \\ 98.09 \\ 98.25 \\ 98.41 \\ 98.51 \\ 98.71 \\ 98.79 \\ 98.87 \\ 98.76 \\ 98.58 \\ 98.73 \\ 98.45 \\ 98.37 \\ 98.36 \\ 98.45 \end{pmatrix} \sim \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

(Careful! It is not a genuine inverse, only LS.)

Observe the little dot in the code below. It denotes matrix product and is very important!

```
In[37]:= betahatinletTemp = PseudoInverse[xinletTemp].removalPerc

Out[37]= {97.4986, 0.0756914}
```

Therefore, we have $\hat{\beta}_0 = 97.4986$ and $\hat{\beta}_1 = 0.0756914$

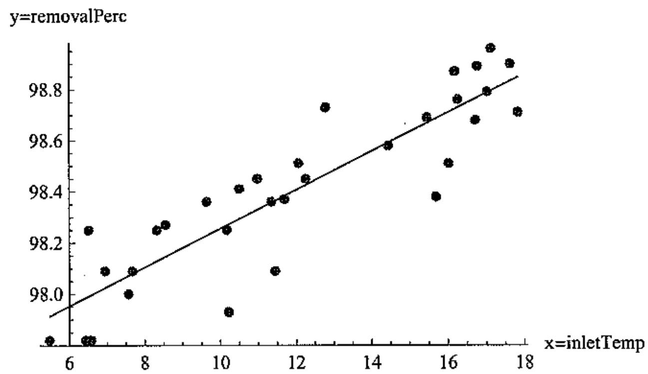
In[38]:=
```
removalPerchat = xinletTemp.betahatinletTemp
```

Out[38]=
```
{98.0799, 97.9913, 97.9853, 97.9134, 97.9959, 98.2722, 98.6862, 98.7679,
 98.7952, 98.833, 98.7641, 98.668, 98.4114, 98.3645, 98.2684, 98.2283,
 98.1457, 98.0716, 98.0239, 98.1283, 98.2933, 98.7112, 98.8482, 98.7876,
 98.7233, 98.7293, 98.5916, 98.4659, 98.4258, 98.3834, 98.3569, 98.3289}
```

Next, overlay the LS line on the plot of (x, y). Notice that the code below asks that *Mathematica* join the line of LS fitted values, i.e. the line joins points $(x_i, \hat{y}_i)$. See that LS produces fitted values falling perfectly on a line (otherwise *Mathematica* would have plotted a broken zig-zag line).

In[39]:=
```
Show[ListPlot[Table[{inletTemp[[i]], removalPerc[[i]]}, {i, 1, 32}],
  AxesLabel -> {"x=inletTemp", "y=removalPerc"}, PlotStyle -> PointSize[0.02]],
 Graphics[Line[Table[{inletTemp[[i]], removalPerchat[[i]]}, {i, 1, 32}]]]]
```

Out[39]=



In[40]:=
```
removalPercresid = removalPerc - removalPerchat
```

Out[40]=
```
{0.0101018, 0.258661, -0.165284, -0.0933772, -0.175881, -0.342154, -0.306186, 0.122067,
 0.164818, 0.0669725, -0.0841483, 0.0219797, 0.0985735, -0.274498, -0.0183698, 0.131747,
 0.12425, -0.0715722, 0.0661134, 0.121659, 0.116652, -0.201164, -0.138166, 0.00238735,
 0.146725, 0.0306697, -0.011572, 0.264076, 0.0241921, -0.0134207, 0.0030713, 0.121077}
```

In[41]:=
$$\text{Inverse[Transpose[xinletTemp].xinletTemp]}\ \frac{32-1}{32-2}\ (\text{s[removalPercresid]})^2$$

Out[41]=
```
{{0.00791118, -0.000596158}, {-0.000596158, 0.0000496462}}
```

In[42]:=
```
MatrixForm[%]
```

Out[42]//MatrixForm=
$$\begin{pmatrix} 0.00791118 & -0.000596158 \\ -0.000596158 & 0.0000496462 \end{pmatrix}$$

From the above, we estimate the variance of $\hat{\beta}_0$ to be 0.00791118, the variance of $\hat{\beta}_1$ to be 0.0000496462, and the covariance of $\hat{\beta}_0$ with $\hat{\beta}_1$ (same as cov of $\hat{\beta}_1$ with $\hat{\beta}_0$) to be -0.000596158.

In[43]:=
$$97.4985 + 2.042\sqrt{0.00791118}$$

Out[43]= 97.6801

In[44]:= `97.4985 - 2.042 √0.00791118`

Out[44]= 97.3169

So a 95% CI for β_0 (the true intercept absent errors of observation) is (97.3169, 97.6801)

In[45]:= `0.07569 + 2.042 √0.0000496462`

Out[45]= 0.0900779

In[46]:= `0.07569 - 2.042 √0.0000496462`

Out[46]= 0.0613021

So a 95% CI for β_1 (the true intercept absent errors of observation) is (0.0613021, 0.0900779)

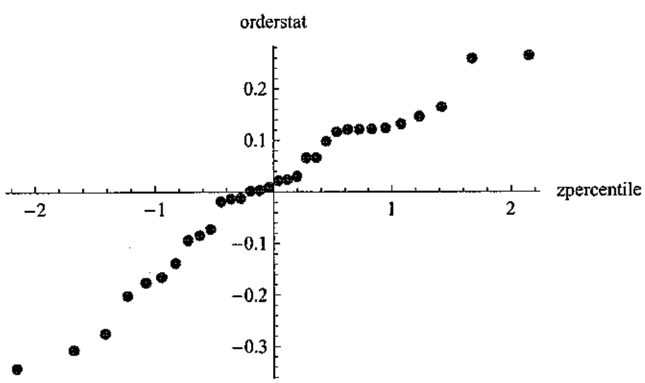
The residuals of our *Mathematica* solution have the following mean of squares

In[47]:= `mean[(removalPerc - xinletTemp.{97.49858845327189`, 0.07569137952245805`})^2]`

Out[47]= 0.0225735

For a ***partial*** check on the ***normal errors assumption*** of the probability model it is customary to perform a ***normal probability plot for the residuals*** to see if it departs very much from a straight line.

In[48]:= `normalprobabilityplot[removalPercresid, 0.02]`

Out[48]=


Here is the correlation between the independent variable inletTemp and the dependent variable removalPerc. Squaring it gives the coefficient of determination which is "the fraction of var y accounted for by regression on x."

In[49]:= `r[inletTemp, removalPerc]`

Out[49]= 0.890883

In[50]:= `%^2`

Out[50]= 0.793673

Example 13.12

```
In[32]:= force = { 30, 40, 30, 40, 30, 40, 30, 40, 30, 40, 30, 40,
            30, 40, 30, 40, 25, 45, 35, 35, 35, 35, 35, 35, 35, 35, 35, 35, 35, 35 };

In[33]:= power = { 60, 60, 90, 90, 60, 60, 90, 90, 60, 60, 90, 90,
            60, 60, 90, 90, 75, 75, 45, 105, 75, 75, 75, 75, 75, 75, 75, 75, 75, 75 };

In[34]:= temp = { 175, 175, 175, 175, 225, 225, 225, 225, 175, 175, 175, 175, 225, 225,
            225, 225, 200, 200, 200, 200, 150, 250, 200, 200, 200, 200, 200, 200, 200, 200 };

In[35]:= time = { 15, 15, 15, 15, 15, 15, 15, 15, 25, 25, 25, 25,
            25, 25, 25, 25, 20, 20, 20, 20, 20, 20, 10, 30, 20, 20, 20, 20, 20, 20 };

In[36]:= yyStrength =
          { 26.2, 26.3, 39.8, 39.7, 38.6, 35.5, 48.8, 37.8, 26.6, 23.4, 38.6, 52.1, 39.5, 32.3, 43.0,
            56.0, 35.2, 46.9, 22.7, 58.7, 34.5, 44.0, 35.7, 41.8, 36.5, 37.6, 40.3, 46.0, 27.8, 40.3 };

In[37]:= Length[force]

Out[37]= 30

In[38]:= xxStats = Table[ {1, force[[i]], power[[i]], temp[[i]], time[[i]]}, {i, 1, 30}];
```

```
MatrixForm[xxStats]
```

1	30	60	175	15
1	40	60	175	15
1	30	90	175	15
1	40	90	175	15
1	30	60	225	15
1	40	60	225	15
1	30	90	225	15
1	40	90	225	15
1	30	60	175	25
1	40	60	175	25
1	30	90	175	25
1	40	90	175	25
1	30	60	225	25
1	40	60	225	25
1	30	90	225	25
1	40	90	225	25
1	25	75	200	20
1	45	75	200	20
1	35	45	200	20
1	35	105	200	20
1	35	75	150	20
1	35	75	250	20
1	35	75	200	10
1	35	75	200	30
1	35	75	200	20
1	35	75	200	20
1	35	75	200	20
1	35	75	200	20
1	35	75	200	20
1	35	75	200	20

```
In[39]:= betahatxxStats = PseudoInverse[xxStats].yyStrength

Out[39]= {-37.4767, 0.211667, 0.498333, 0.129667, 0.258333}
```

As the above, we got β_0 = -37.48 β_1 = 0.211667 β_2 = 0.498333 β_3 = 0.1297 and β_4 = 0.2583 which matched the results in the book page 534.

```
In[40]:= yyStrengthhat = xxStats.betahatxxStats

Out[40]= {25.34, 27.4567, 40.29, 42.4067, 31.8233, 33.94, 46.7733, 48.89, 27.9233, 30.04,
  42.8733, 44.99, 34.4067, 36.5233, 49.3567, 51.4733, 36.29, 40.5233, 23.4567, 53.3567,
  31.9233, 44.89, 35.8233, 40.99, 38.4067, 38.4067, 38.4067, 38.4067, 38.4067, 38.4067}

In[41]:= yyStrengthresid = yyStrength - yyStrengthhat

Out[41]= {0.86, -1.15667, -0.49, -2.70667, 6.77667, 1.56, 2.02667, -11.09, -1.32333, -6.64, -4.27333,
  7.11, 5.09333, -4.22333, -6.35667, 4.52667, -1.09, 6.37667, -0.756667, 5.34333, 2.57667,
  -0.89, -0.123333, 0.81, -1.90667, -0.806667, 1.89333, 7.59333, -10.6067, 1.89333}
```

```
In[42]:= Inverse[Transpose[xxStats].xxStats] (30 - 1)/(30 - 5) (s[yyStrengthresid])^2

Out[42]= {{171.601, -1.55194, -0.36951, -0.35473, -0.886825},
  {-1.55194, 0.0443412, 0, 0, 0}, {-0.36951, 0, 0.0049268, 0, 0},
  {-0.35473, 0, 0, 0.00177365, 0}, {-0.886825, 0, 0, 0, 0.0443412}}

MatrixForm[%]
```

$$\begin{pmatrix} 171.601 & -1.55194 & -0.36951 & -0.35473 & -0.886825 \\ -1.55194 & 0.0443412 & 0 & 0 & 0 \\ -0.36951 & 0 & 0.0049268 & 0 & 0 \\ -0.35473 & 0 & 0 & 0.00177365 & 0 \\ -0.886825 & 0 & 0 & 0 & 0.0443412 \end{pmatrix}$$

$-37.48 + 2.060 \sqrt{171.601}$

-10.4947

$-37.48 - 2.060 \sqrt{171.601}$

-64.4653

So a 95% CI for β_0 (the true intercept absent errors of observation) is (-64.4653, -10.4947)

$0.2117 + 2.06 \sqrt{0.0443414}$

0.645482

$0.2117 - 2.06 \sqrt{0.0443414}$

-0.222082

So a 95% CI for β_1 (the true intercept absent errors of observation) is (-0.222082, 0.645482)

$0.49833 + 2.06 \sqrt{0.00492}$

0.642824

$0.49833 - 2.06 \sqrt{0.00492}$

0.353836

So a 95% CI for β_2 (the true intercept absent errors of observation) is (0.353836, 0.642824)

In[43]:= $0.12967 + 2.06 \sqrt{0.00177}$

Out[43]= 0.216337

In[45]:= $0.12967 - 2.06 \sqrt{0.00177}$

Out[45]= 0.043003

So a 95% CI for β_3 (the true intercept absent errors of observation) is (0.043003, 0.216337)

In[46]:= `0.2583 + 2.06 √0.0443412`

Out[46]= 0.692081

In[47]:= `0.2583 - 2.06 √0.0443412`

Out[47]= -0.175481

So a 95% CI for β_4 (the true intercept absent errors of observation) is (-0.175481, 0.692081)

The residuals of our *Mathematica* solution have the following mean of squares

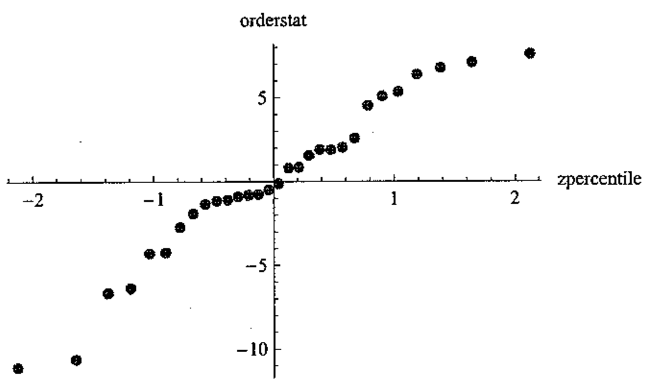
```
mean[(yyStrength - xxStats.{-37.47666666666667`, 0.2116666666666666`,
    0.49833333333333335`, 0.12966666666666665`, 0.25833333333333325`})^2]
```

22.1706

For a ***partial*** check on the ***normal errors assumption*** of the probability model it is customary to perform a ***normal probability plot for the residuals*** to see if it departs very much from a straight line.

In[50]:= `normalprobabilityplot[yyStrengthresid, 0.02]`

Out[50]=



In[48]:= `r[yyStrengthhat, yyStrength]`

Out[48]= 0.844961

In[49]:= `%^2`

Out[49]= 0.713959

Problem 47

```
In[82]:= plastics = { 18.69, 19.43, 19.24, 22.64, 16.54, 21.44, 19.53, 23.97,
        21.45, 20.34, 17.03, 21.03, 20.49, 20.45, 18.81, 18.28, 21.41, 25.11, 21.04,
        17.99, 18.73, 18.49, 22.08, 14.28, 17.74, 20.54, 18.25, 19.09, 21.25, 21.62 };

In[83]:= paper  = { 15.65, 23.51, 24.23, 22.20, 23.56, 23.65, 24.45, 19.39,
        23.84, 26.50, 23.46, 26.99, 19.87, 23.03, 22.62, 21.87, 20.47, 22.59, 26.27,
        28.22, 29.39, 26.58, 24.88, 26.27, 23.61, 26.58, 13.77, 25.62, 20.63, 22.71 };

In[84]:= garbage =  { 45.01, 39.69, 43.16, 35.76, 41.20, 35.56, 40.18, 44.11,
        35.41, 34.21, 32.45, 38.19, 41.35, 43.59, 42.20, 41.50, 41.20, 37.02, 38.66,
        44.18, 34.77, 37.55, 37.07, 35.80, 37.36, 35.40, 51.32, 39.54, 40.72, 36.22 };

In[85]:= water = { 58.21, 46.31, 46.63, 45.85, 55.14, 54.24, 47.20, 43.82,
        51.01, 49.06, 53.23, 51.78, 46.69, 53.57, 52.98, 47.44, 54.68, 48.74, 53.22,
        53.37, 51.06, 50.66, 50.72, 48.24, 49.92, 53.58, 51.38, 50.13, 48.67, 48.19 };

In[86]:= xxWaste = Table[ {1, plastics[[i]], paper[[i]], garbage[[i]], water[[i]]}, {i, 1, 30}];

In[87]:= yyEnergyContent =
       { 947, 1407, 1452, 1553, 989, 1162, 1466, 1656, 1254, 1336, 1097, 1266, 1401, 1223, 1216,
        1334, 1155, 1453, 1278, 1153, 1225, 1237, 1327, 1229, 1205, 1221, 1138, 1295, 1391, 1372 };

In[88]:= Length[xxWaste]

Out[88]= 30
```

In[89]:= **MatrixForm[xxWaste]**

Out[89]//MatrixForm=

1	18.69	15.65	45.01	58.21
1	19.43	23.51	39.69	46.31
1	19.24	24.23	43.16	46.63
1	22.64	22.2	35.76	45.85
1	16.54	23.56	41.2	55.14
1	21.44	23.65	35.56	54.24
1	19.53	24.45	40.18	47.2
1	23.97	19.39	44.11	43.82
1	21.45	23.84	35.41	51.01
1	20.34	26.5	34.21	49.06
1	17.03	23.46	32.45	53.23
1	21.03	26.99	38.19	51.78
1	20.49	19.87	41.35	46.69
1	20.45	23.03	43.59	53.57
1	18.81	22.62	42.2	52.98
1	18.28	21.87	41.5	47.44
1	21.41	20.47	41.2	54.68
1	25.11	22.59	37.02	48.74
1	21.04	26.27	38.66	53.22
1	17.99	28.22	44.18	53.37
1	18.73	29.39	34.77	51.06
1	18.49	26.58	37.55	50.66
1	22.08	24.88	37.07	50.72
1	14.28	26.27	35.8	48.24
1	17.74	23.61	37.36	49.92
1	20.54	26.58	35.4	53.58
1	18.25	13.77	51.32	51.38
1	19.09	25.62	39.54	50.13
1	21.25	20.63	40.72	48.67
1	21.62	22.71	36.22	48.19

In[90]:= **betahatxxWaste = PseudoInverse[xxWaste].yyEnergyContent**

Out[90]= {2244.92, 28.925, 7.64361, 4.29664, -37.3538}

Therefore we have $\hat{\beta}_0$= 2244.91, $\hat{\beta}_1$ = 28.925, $\hat{\beta}_2$= 7.64361, $\hat{\beta}_3$= 4.29664, $\hat{\beta}_4$= -37.3538

In[91]:= **yyEnergyContenthat = xxWaste.betahatxxWaste**

Out[91]= {924.179, 1427.31, 1430.28, 1510.45, 1020.76, 1172.56, 1406.25, 1639.14, 1294.31, 1350.22, 1067.92, 1289.42, 1423.09, 1198.72, 1164.21, 1347.08, 1155.19, 1482.33, 1232.44, 1177.24, 1253.44, 1251.91, 1338.45, 1210.64, 1234.34, 1192.89, 1179.32, 1290.27, 1374.22, 1399.41}

In[92]:= **yyEnergyContentresid = yyEnergyContent - yyEnergyContenthat**

Out[92]= {22.8211, -20.3146, 21.7216, 42.5523, -31.7572, -10.563, 59.7474, 16.8554, -40.3129, -14.2221, 29.0838, -23.424, -22.0903, 24.2827, 51.7872, -13.0824, -0.185904, -29.3347, 45.5603, -24.2379, -28.4414, -14.907, -11.45, 18.3596, -29.3372, 28.1077, -41.3203, 4.72798, 16.785, -27.4109}

In[93]:= **Inverse [Transpose [xxWaste] . xxWaste]** $\frac{30-1}{30-5}$ **(s [yyEnergyContentresid])**^{2}

```
Out[93]= {{31 648.9, -312.105, -277.179, -231.862, -193.821},
  {-312.105, 7.97316, 1.71481, 1.21959, 1.29263},
  {-277.179, 1.71481, 5.35382, 2.89513, 0.0750264},
  {-231.862, 1.21959, 2.89513, 3.67204, -0.0924896},
  {-193.821, 1.29263, 0.0750264, -0.0924896, 3.36437}}
```

In[94]:= **MatrixForm [%]**

Out[94]//MatrixForm=

$$\begin{pmatrix} 31648.9 & -312.105 & -277.179 & -231.862 & -193.821 \\ -312.105 & 7.97316 & 1.71481 & 1.21959 & 1.29263 \\ -277.179 & 1.71481 & 5.35382 & 2.89513 & 0.0750264 \\ -231.862 & 1.21959 & 2.89513 & 3.67204 & -0.0924896 \\ -193.821 & 1.29263 & 0.0750264 & -0.0924896 & 3.36437 \end{pmatrix}$$

From the above, we estimate the variance of $\hat{\beta}_0$ to be 31648.9, the variance of $\hat{\beta}_1$ to be 7.97316, the variance of $\hat{\beta}_2$ to be 5.35382, the variance of $\hat{\beta}_3$ to be 3.67204, the variance of $\hat{\beta}_4$ to be 3.36437 and the covariance of $\hat{\beta}_0$ with $\hat{\beta}_1$ (same as cov of $\hat{\beta}_1$ with $\hat{\beta}_0$) to be -312.105, the covariance of $\hat{\beta}_1$ with $\hat{\beta}_2$ to be -277.179, the covariance of $\hat{\beta}_2$ with $\hat{\beta}_3$to be -231.862, the covariance of $\hat{\beta}_3$ with $\hat{\beta}_4$ to be -193.821 .

2244.92 + 2.06 $\sqrt{\mathbf{31\,648.9}}$

2611.4

2244.92 - 2.06 $\sqrt{\mathbf{31\,648.9}}$

1878.44

So a 95% CI for β_0 (the true intercept absent errors of observation) is (1878.44, 2611.4)

28.925 + 2.060 $\sqrt{\mathbf{7.79316}}$

34.6757

28.925 - 2.060 $\sqrt{\mathbf{7.79316}}$

23.1743

So a 95% CI for β_1 (the true intercept absent errors of observation) is (23.1743, 34.6757)

7.64361 + 2.060 $\sqrt{\mathbf{5.35382}}$

12.4101

7.64361 - 2.060 $\sqrt{\mathbf{5.35382}}$

2.87712

So a 95% CI for β_2 (the true intercept absent errors of observation) is (2.87712, 12.4101)

$4.29664 + 2.060\sqrt{3.67204}$

8.24413

$4.29664 - 2.060\sqrt{3.67204}$

0.349151

So a 95% CI for β_3 (the true intercept absent errors of observation) is (0.349151, 8.24413)

$-37.3538 + 2.060\sqrt{3.36437}$

-33.5753

$-37.3538 - 2.060\sqrt{3.36437}$

-41.1323

So a 95% CI for β_4 (the true intercept absent errors of observation) is (-41.1323, -33.5753)

The residuals of our *Mathematica* solution have the following mean of squares

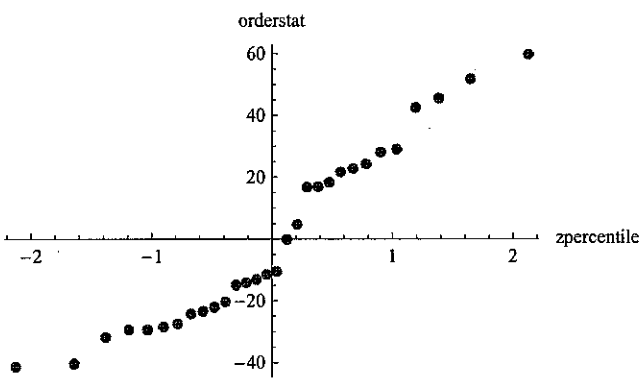
In[95]:= `mean[(yyEnergyContent - xxWaste.{2244.92, 28.925, 7.64361, 4.29664, -37.3538})^2]`

Out[95]= 825.974

For a ***partial*** check on the ***normal errors assumption*** of the probability model it is customary to perform a ***normal probability plot for the residuals*** to see if it departs very much from a straight line.

In[96]:= `normalprobabilityplot[yyEnergyContentresid, 0.02]`

Out[96]=



Here is the correlation between the xxWaste and the yyEnergyContent. Squaring it gives the coefficient of determination which is "the fraction of var y accounted for by regression on x."

In[97]:= `r[yyEnergyContenthat, yyEnergyContent]`

Out[97]= 0.981872

In[98]:= `%^2`

Out[98]= 0.964073

Problem 52

```
In[32]:= linoleic = { 30, 30, 30, 40, 30, 13.18, 20, 20, 40, 30, 30, 40, 40, 30, 30, 30, 30, 20, 20, 46.82 };

In[33]:= kerosene = { 30, 30, 30, 40, 30, 30, 40, 40, 20, 30, 30, 20, 40, 30, 46.82, 30, 13.18, 20, 20, 30 };

In[34]:= antiox =   { 10, 10, 18.41, 5, 10, 10, 5, 15, 5, 10, 1.59, 15, 15, 10, 10, 10, 10, 5, 15, 10 };

In[35]:= xxPredictors = Table[ {1, linoleic[[i]], kerosene[[i]], antiox[[i]]}, {i, 1, 20}];

In[36]:= yyBetacaro = { .7, .63, .013, .049, .7, .1, .04, .0065,
            .2020, .63, .04, .132, .150, .7, .346, .63, .397, .269, .0054, .064 };

In[37]:= Length[xxPredictors]

Out[37]= 20

In[38]:= MatrixForm[xxPredictors]

Out[38]//MatrixForm=
```

1	30	30	10
1	30	30	10
1	30	30	18.41
1	40	40	5
1	30	30	10
1	13.18	30	10
1	20	40	5
1	20	40	15
1	40	20	5
1	30	30	10
1	30	30	1.59
1	40	20	15
1	40	40	15
1	30	30	10
1	30	46.82	10
1	30	30	10
1	30	13.18	10
1	20	20	5
1	20	20	15
1	46.82	30	10

```
In[39]:= betahatxxPredictors = PseudoInverse[xxPredictors].yyBetacaro

Out[39]= {0.401075, 0.00110957, -0.00328506, -0.00456155}
```

Therefore we have $\hat{\beta}_0$= 0.401075, $\hat{\beta}_1$ = 0.00110957, $\hat{\beta}_2$= -0.00328506, $\hat{\beta}_3$= -0.00456155

```
In[40]:= yyBetacarohat = xxPredictors.betahatxxPredictors

Out[40]= {0.290195, 0.290195, 0.251832, 0.291248, 0.290195, 0.271532,
          0.269056, 0.223441, 0.356949, 0.290195, 0.328558, 0.311334, 0.245632,
          0.290195, 0.23494, 0.290195, 0.34545, 0.334758, 0.289142, 0.308858}
```

In[41]:= yyBetacaroresid = yyBetacaro - yyBetacarohat

Out[41]= {0.409805, 0.339805, -0.238832, -0.242248, 0.409805, -0.171532, -0.229056, -0.216941, -0.154949, 0.339805, -0.288558, -0.179334, -0.0956323, 0.409805, 0.11106, 0.339805, 0.0515502, -0.0657577, -0.283742, -0.244858}

In[42]:= Inverse[Transpose[xxPredictors].xxPredictors] $\frac{20-1}{20-4}$ (s[yyBetacaroresid])^2

Out[42]= {{0.145654, -0.00192639, -0.00192639, -0.00256853}, {-0.00192639, 0.0000642132, 4.38439×10^{-20}, 4.12113×10^{-19}}, {-0.00192639, 6.94222×10^{-20}, 0.0000642132, -1.2798×10^{-19}}, {-0.00256853, 3.97479×10^{-19}, -2.15821×10^{-19}, 0.000256853}}

In[43]:= MatrixForm[%]

Out[43]//MatrixForm=

$$\begin{pmatrix} 0.145654 & -0.00192639 & -0.00192639 & -0.00256853 \\ -0.00192639 & 0.0000642132 & 4.38439\times10^{-20} & 4.12113\times10^{-19} \\ -0.00192639 & 6.94222\times10^{-20} & 0.0000642132 & -1.2798\times10^{-19} \\ -0.00256853 & 3.97479\times10^{-19} & -2.15821\times10^{-19} & 0.000256853 \end{pmatrix}$$

From the above, we estimate the variance of $\hat{\beta}_0$ to be 0.155364, the variance of $\hat{\beta}_1$ to be 0.000068494, the variance of $\hat{\beta}_2$ to be 0.000068494, the variance of $\hat{\beta}_3$ to be 0.000273976 and the covariance of $\hat{\beta}_0$ with $\hat{\beta}_1$ (same as cov of $\hat{\beta}_1$ with $\hat{\beta}_0$) to be -0.00205482, the covariance of $\hat{\beta}_1$ with $\hat{\beta}_2$ to be -0.00205482, the covariance of $\hat{\beta}_2$ with $\hat{\beta}_3$to be -0.00273976.

$0.401 + 2.12\sqrt{0.155}$

1.23564

$0.401 - 2.12\sqrt{0.155}$

-0.433645

So a 95% CI for β_0 (the true intercept absent errors of observation) is (-0.433645, 1.23564)

$0.0011 + 2.12\sqrt{0.000068494}$

0.0186454

$0.0011 - 2.12\sqrt{0.000068494}$

-0.0164454

So a 95% CI for β_1 (the true intercept absent errors of observation) is (-0.0164454, 0.0186454)

$-0.003285 + 2.12\sqrt{0.000068494}$

0.0142604

```
-0.003285 - 2.12 √0.000068494
-0.0208304
```

So a 95% CI for β_2 (the true intercept absent errors of observation) is (-0.0208304, 0.0142604)

```
-0.00456 + 2.12 √0.000273976
0.0305307
```

```
-0.00456 - 2.12 √0.000273976
-0.0396507
```

So a 95% CI for β_3 (the true intercept absent errors of observation) is (-0.0396507, 0.0305307)

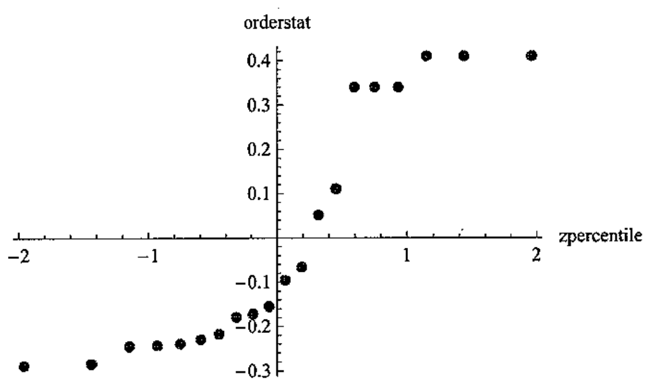
The residuals of our *Mathematica* solution have the following mean of squares

```
mean[(yyBetacaro - xxPredictors.{0.40107525345564266`,
     0.0011095713007991735`, -0.003285062622965892`, -0.004561551379064221`})^2]
0.0701631
```

For a ***partial*** check on the ***normal errors assumption*** of the probability model it is customary to perform a ***normal probability plot for the residuals*** to see if it departs very much from a straight line.

```
In[44]:= normalprobabilityplot[yyBetacaroresid, 0.02]
```

Out[44]=



Here is the correlation between xxPredictors and yyBetacaro. Squaring it gives the coefficient of determination which is "the fraction of var y accounted for by regression on x."

```
In[45]:= r[yyBetacarohat, yyBetacaro]
Out[45]= 0.128409
In[46]:= %^2
Out[46]= 0.0164887
```