Admir Makas

STT-6660 HW #4

In [452]:
```
import numpy as np
import scipy as sp
import sympy as sy
import warnings

%matplotlib inline
import matplotlib.pyplot as plt

sy.init_printing(use_latex='mathjax')
from IPython.display import display, Math, Latex
```

Problem 1.22

1.22. **Plastic hardness.** Refer to Problems 1.3 and 1.14. Sixteen batches of the plastic were made, and from each batch one test item was molded. Each test item was randomly assigned to one of the four predetermined time levels, and the hardness was measured after the assigned elapsed time. The results are shown below: X is the elapsed time in hours, and Y is hardness in Brinell units. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	14	15	16
X_i:	16	16	16	...	40	40	40
Y_i:	199	205	196	...	248	253	246

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here?

b. Obtain a point estimate of the mean hardness when $X = 40$ hours.

c. Obtain a point estimate of the change in mean hardness when X increases by 1 hour.

Part a

In [453]:
```
#Define the data

Y = sy.Matrix([199, 205, 196, 200, 218, 220, 215, 223, 237, 234, 235, 230,
250, 248, 253, 246]).T
X = sy.Matrix([16, 16, 16, 16, 24, 24, 24, 24, 32, 32, 32, 32, 40, 40, 40, 40]
T
```

In [454]:
```
display(X)
display(Y)
```

$$\left[\begin{array}{cccccccccccccccc}16 & 16 & 16 & 16 & 24 & 24 & 24 & 24 & 32 & 32 & 32 & 32 & 40 & 40 & 40 & 40\end{array}\right]$$

$$\left[\begin{array}{ccccccccccccccc}199 & 205 & 196 & 200 & 218 & 220 & 215 & 223 & 237 & 234 & 235 & 230 & 250 & 248 & 2\end{array}\right.$$

Calculate mean for X

In [455]:
```
Xbar = np.mean(X).evalf(5)
display(Math(r'\bar{X} = '))
display(Xbar)
```

$\bar{X} =$

28.0

Calculate mean for Y

In [456]:
```
Ybar = np.mean(Y).evalf(4)
display(Math(r'\bar{Y} = '))
display(Ybar)
```

$\bar{Y} =$

225.6

Get S_{xy}

In [457]:
```
Sxy = np.sum(np.multiply(X,Y)) - (1/16)*(np.sum(X)*np.sum(Y))
display(Math(r'S_{xy} = '))
display(Sxy)
```

$S_{xy} =$

2604.0

Get S_{xx}

In [458]:
```
Sxx = np.sum(np.multiply(X,X)) - (1/16)*(np.sum(X)*np.sum(X))
display(Math(r'S_{xx} = '))
display(Sxx)
```

$S_{xx} =$

1280.0

Get b_1

In [459]:
```
b1 = (Sxy/Sxx).evalf(5)
display(Math(r'b_{1} = '))
display(b1)
```

$b_1 =$

2.0344

Get b_0

In [460]:
```
b0 = (Ybar - b1*Xbar).evalf(4)
display(Math(r'b_{0} = '))
display(b0)
```

$b_0 =$

168.6

Regression Function is:

$$\hat{Y} = 2.0344X + 168.6$$

Next plot the data and the corresponding regression model.

From the figure below the regression model fits the data quite well.

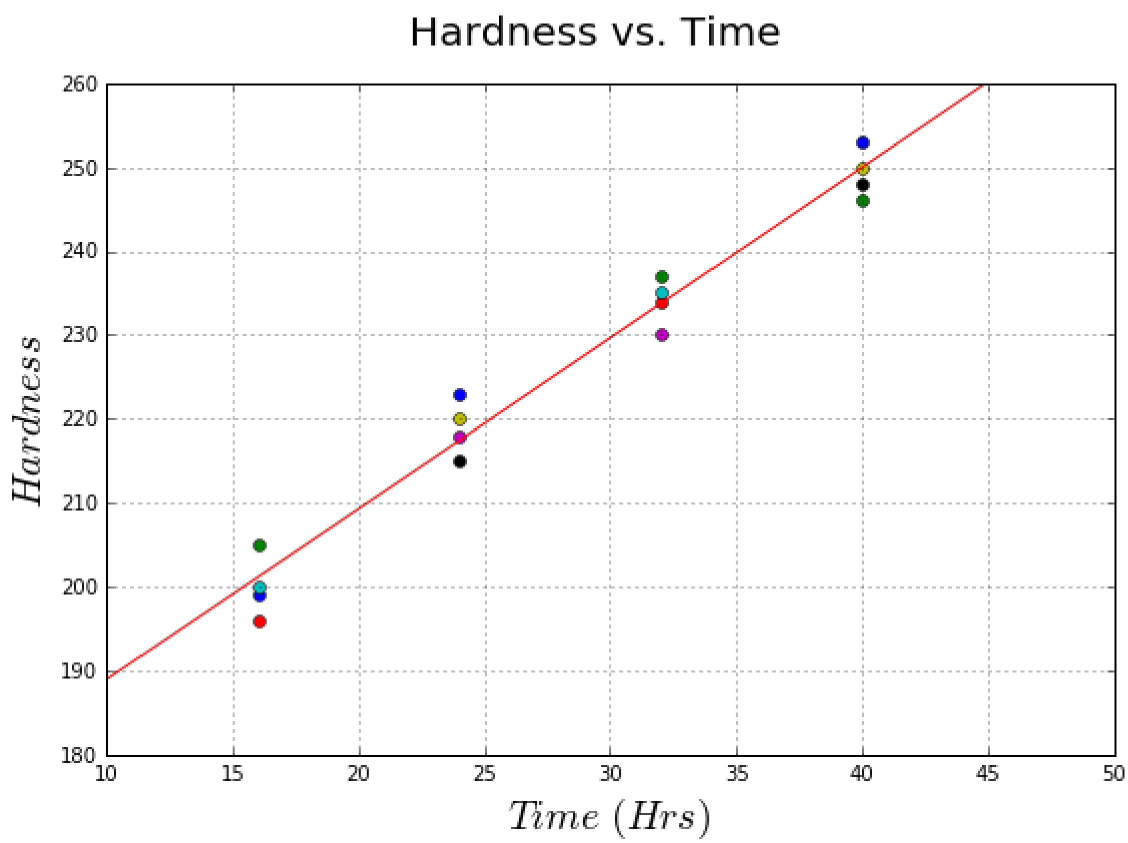
In [461]:
```
X1=np.linspace(0.0,300.0,500)

Y1 = 2.0344*X1 + 168.6;

fig=plt.figure(1, figsize=(9,6))
ax = fig.add_subplot(111)

ax.plot(X, Y, "o")
ax.plot(X1, Y1)

#Plot annotations
axis_span=[10, 50, 180, 260]
plt.axis(axis_span)
fig.suptitle('Hardness vs. Time', fontsize=20)
plt.xlabel('$Time\ (Hrs)$', fontsize=20)
plt.ylabel('$Hardness$', fontsize=20)
plt.grid()
```



Part b

In [462]:
```
y_est = (2.0344*40 + 168.6)
display(Math(r'\hat{y_1} = '))
display(y_est)
```

$\hat{y_1} =$

249.976

Part c

In [463]:
```
y2_est = (2.0344*41 + 168.6)
display(Math(r'\hat{y_2} = '))
display(y2_est)
```

$\hat{y_2} =$

252.0104

The difference simply ends up being b_1 since predictor was increased by a unit value.

In [464]:
```
diff = (y2_est) - (y_est)
display(Math(r'\hat{y_2} - \hat{y_1} = '))
diff
```

$\hat{y_2} - \hat{y_1} =$

Out[464]: 2.0344000000000005

Problem 2.25

*2.25. Refer to **Airfreight breakage** Problem 1.21.

a. Set up the ANOVA table. Which elements are additive?

b. Conduct an F test to decide whether or not there is a linear association between the number of times a carton is transferred and the number of broken ampules; control the α risk at .05. State the alternatives, decision rule, and conclusion.

c. Obtain the t^* statistic for the test in part (b) and demonstrate numerically its equivalence to the F^* statistic obtained in part (b).

d. Calculate R^2 and r. What proportion of the variation in Y is accounted for by introducing X into the regression model?

Part a

In [465]:
```
#Define the data

X = sy.Matrix([1, 0, 2, 0, 3, 1, 0, 1, 2, 0]).T
Y = sy.Matrix([16, 9, 17, 12, 22, 13, 8, 15, 19, 11]).T
```

In [466]:
```
display(X)
display(Y)
```

$$\begin{bmatrix}1 & 0 & 2 & 0 & 3 & 1 & 0 & 1 & 2 & 0\end{bmatrix}$$

$$\begin{bmatrix}16 & 9 & 17 & 12 & 22 & 13 & 8 & 15 & 19 & 11\end{bmatrix}$$

Calculate mean for Y

In [467]:
```
Ybar = np.mean(Y).evalf(4)
display(Math(r'\bar{Y} = '))
display(Ybar)
```

$\bar{Y} =$

14.2

Calculate SSTO

In [468]:
```
SSTO = (np.sum(np.power((Y-np.ones((1,10))*Ybar),2))).evalf(5)
display(Math(r'SSTO = '))
display(SSTO)
```

$SSTO =$

177.6

Calculate SSE

In [469]:
```
Y_est1 = np.multiply(np.ones((1,10))*4, X)
Y_est = np.add(Y_est1, np.ones((1,10))*10.2)
```

In [470]:
```
SSE = np.sum(np.power((Y-Y_est),2))
display(Math(r'SSE = '))
display(SSE)
```

$SSE =$

17.6

Calculate SSR

In [471]:
```
SSR = (SSTO-SSE).evalf(6)
display(Math(r'SSR = SSTO - SSE'))
display(SSR)
```

$SSR = SSTO - SSE$

160.0

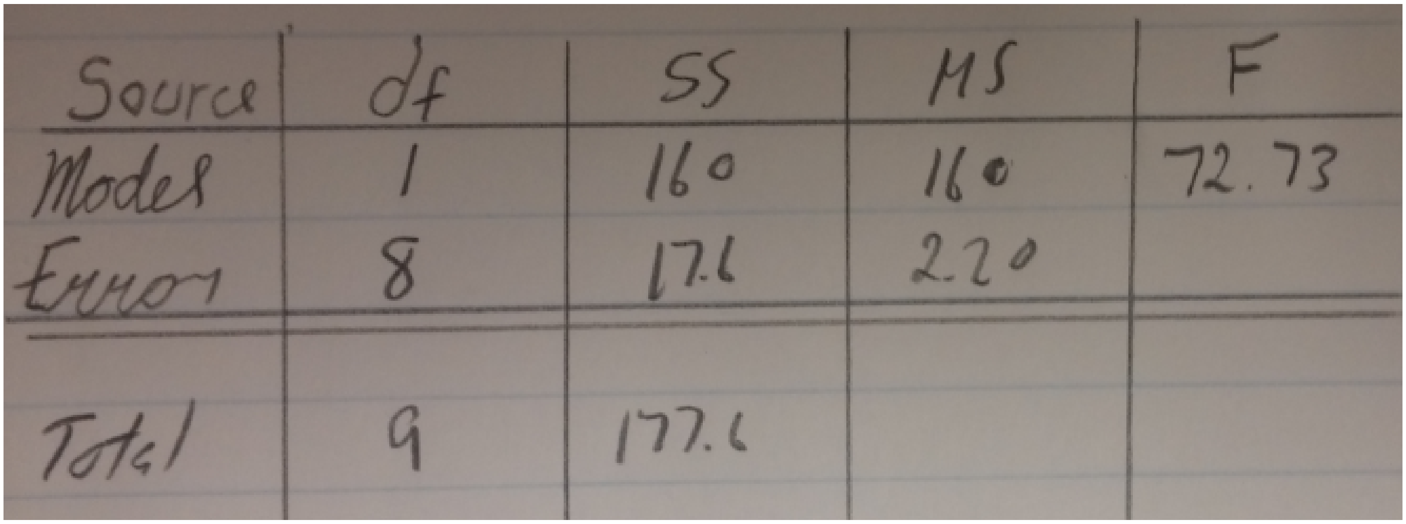
Calculate MSE

In [472]:
```
MSE = (SSE/8).evalf(5)
display(Math(r'MSE = '))
display(MSE)
```

$MSE =$

2.2

Below is the ANOVA table



Source	df	SS	MS	F
Model	1	160	160	72.73
Error	8	17.6	2.20	
Total	9	177.6		

Part b

$H_0 : \beta_1 = 0$
$H_a : \beta_1 \neq 0$

In [473]:
```
Fobs = SSR/MSE
display(Math(r'F_{obs} = '))
Fobs
```

$F_{obs} =$

Out[473]: 72.7273

$F(0.95, 1, 8) = 5.32$

Since $F_{obs} = 72.7273 \geq F = 5.32$ and is therefore in the rejection region we conclude that H_0 cannot be accepted

Part c

Get S_{xy}

In [474]:
```
Sxy = np.sum(np.multiply(X,Y)) - (1/10)*(np.sum(X)*np.sum(Y))
display(Math(r'S_{xy} = '))
display(Sxy)
```

$S_{xy} =$

40.0

Get S_{xx}

In [475]:
```
Sxx = np.sum(np.multiply(X,X)) - (1/10)*(np.sum(X)*np.sum(X))
display(Math(r'S_{xx} = '))
display(Sxx)
```

$S_{xx} =$

10.0

Get b_1

In [476]:
```
b1 = Sxy/Sxx
display(Math(r'b_{1} = '))
b1
```

$b_1 =$

Out[476]: 4.0

Get S_{yy}

In [477]:
```
Syy = np.sum(np.multiply(Y,Y)) - (1/10)*(np.sum(Y)*np.sum(Y))
display(Math(r'S_{yy} = '))
display(Syy)
```

$S_{yy} =$

177.6

Get SSE

In [478]:
```
SSE = (Syy).evalf(4) - (b1*Sxy).evalf(4)
display(Math(r'SSE = '))
SSE
```

$SSE =$

Out[478]: 17.6

Finally, get s^2

In [479]:
```
s2 = (SSE/8).evalf(4)
display(Math(r's^2 = '))
s2
```

$s^2 =$

Out[479]: 2.2

In [480]:
```
t_obs = (b1/(sy.sqrt(s2)/sy.sqrt(Sxx))).evalf(4)
display(Math(r't_{obs} = '))
t_obs
```

$t_{obs} =$

Out[480]: 8.528

$t_{0.95}(8)$ **= 2.306**
t_{obs} **= 8.5280**

Since $t_{obs} \geq t_{0.95}(8)$, the null hyppothesis cannot be accepted. This is the same result obtained by ANOVA.

$(t_{obs})^2$ will yield F as seen below, which demostrated the equivalency between the two methods.

In [481]:
```
display(Math(r'(t_{obs})^2 = F_{obs}'))
(t_obs**2).evalf(5)
```

$(t_{obs})^2 = F_{obs}$

Out[481]: 72.729

Part d

In [482]:
```
R2 = (Sxy**2/(Sxx*Syy)).evalf(6)
display(Math(r'R^2 = '))
R2
```

$R^2 =$

Out[482]: 0.900901

In [483]:
```
r = (Sxy/(sy.sqrt(Sxx*Syy))).evalf(4)
display(Math(r'r = '))
r
```

$r =$

Out[483]: 0.9492

90.09% of the variation is accounted for by introducing X into the regression model.

Problem 3.6

3.6. Refer to **Plastic hardness** Problem 1.22.

a. Obtain the residuals e_i and prepare a box plot of the residuals. What information is provided by your plot?

b. Plot the residuals e_i against the fitted values $\hat{Y}_i$ to ascertain whether any departures from regression model (2.1) are evident. State your findings.

In [484]:
```
Y = sy.Matrix([199, 205, 196, 200, 218, 220, 215, 223, 237, 234, 235, 230,
250, 248, 253, 246]).T
X = sy.Matrix([16, 16, 16, 16, 24, 24, 24, 24, 32, 32, 32, 32, 40, 40, 40, 40]
T
```

Get predicted values

In [485]:
```
Yhat = 2.0344*X + 168.6*np.ones((1,16))
```

Get residuals e_i

In [486]:
```
e = Y-Yhat
display(Math(r'e_i = '))
e.T.evalf(5)
```

$e_i =$

Out[486]:

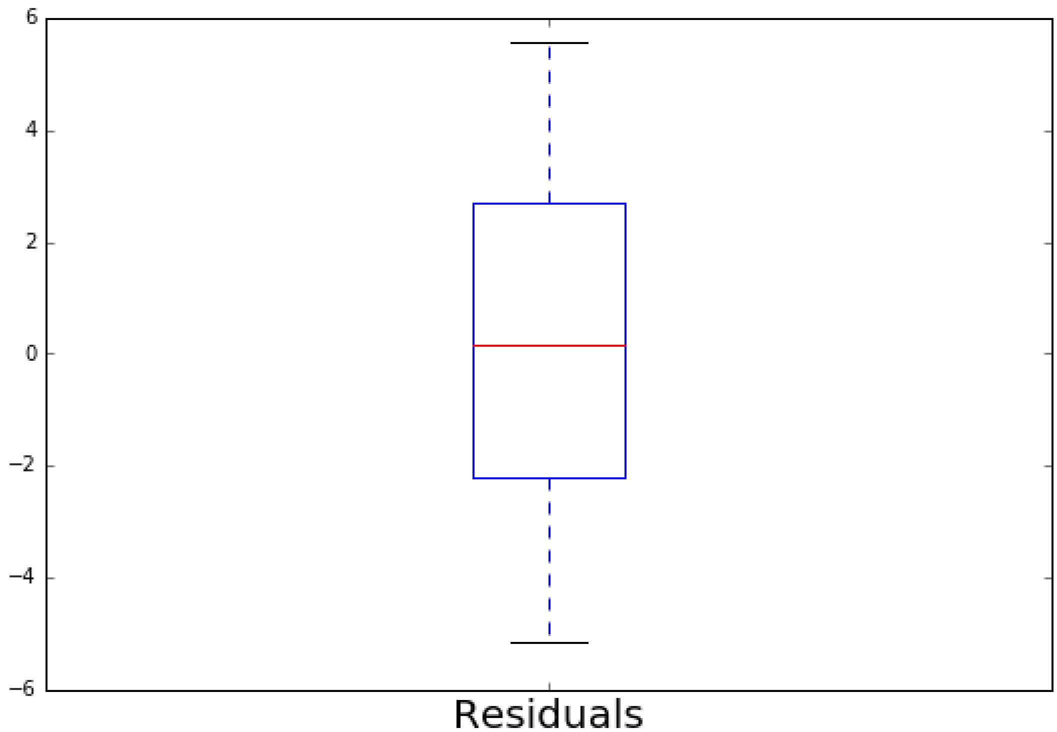
$$\begin{bmatrix} -2.1504 \\ 3.8496 \\ -5.1504 \\ -1.1504 \\ 0.5744 \\ 2.5744 \\ -2.4256 \\ 5.5744 \\ 3.2992 \\ 0.2992 \\ 1.2992 \\ -3.7008 \\ 0.024 \\ -1.976 \\ 3.024 \\ -3.976 \end{bmatrix}$$

Here is the boxplot for the residuals

In [487]:
```
e2=np.array(e[:],dtype=float)
```

In [490]:
```
fig = plt.figure(1, figsize=(9, 6))
ax = fig.add_subplot(111)
ax.boxplot(e2)
ax.set_xticklabels(['Residuals'], fontsize=20)
```

Out[490]: [<matplotlib.text.Text at 0x2e00b7700b8>]



From the boxplot above, it appears that the residuals are normally distributed

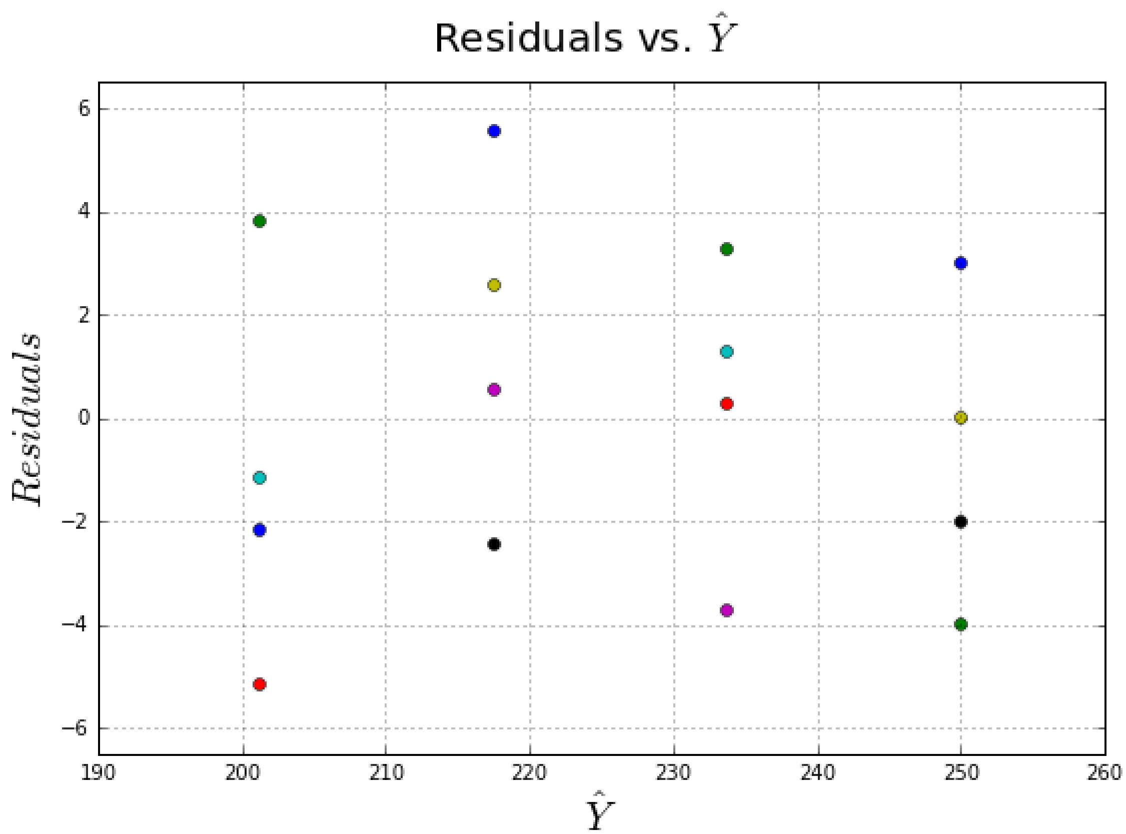
Below is plot of Residuals vs. $\hat{Y}$

In [489]:
```
fig=plt.figure(1, figsize=(9, 6))

ax = fig.add_subplot(111)
ax.plot(Yhat, e, "o")

#Plot annotations
axis_span=[190, 260, -6.5, 6.5]
plt.axis(axis_span)

fig.suptitle('Residuals vs. $\hat{Y}$', fontsize=20)
plt.xlabel('$\hat{Y}$', fontsize=20)
plt.ylabel('$Residuals$', fontsize=20)
plt.grid()
```



No unusual departures from the X-axis are noticable. The assumption of constant variance seems to apply for this data-set.