Tutorial #6

Question 1:

Refer to **Grocery retailer**

A large, national grocery retailer tracks productivity and costs of its facilities closely. Data below were obtained from a single distribution centre for a one-year period. Each data point for each variable represents one week of activity. The variables included are the number of cases shipped (X1) the indirect costs of the total labor hours as a percentage (X2), a qualitative predictor called holiday that is coded 1 if the week has a holiday and 0 otherwise (X3), and the total labor hours (Y).

a) Fit regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.

b) Estimate β_1, β_2 and β_3 jointly confidence interval, using a 95 percent confidence coefficient.

c) Test the significant for β_1, β_2 and β_3? using level of significance 0.05.

	C1	C2	C3	C4	C5
	y	X1	X2	X3	X0
1	4264	305657	7.17	0	1
2	4496	328476	6.20	0	1
3	4317	317164	4.61	0	1
⋮	⋮	⋮	⋮	⋮	⋮
52	4342	292087	7.77	0	1

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}_{n\times 4}\mathbf{B}_{4\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

$$\boldsymbol{E}(\boldsymbol{Y}_{n\times 1}) = \boldsymbol{X}_{n\times 4}\mathbf{B}_{4\times 1}$$

$$\mathbf{B}_{4\times 1} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

$$\boldsymbol{V}(\mathbf{B}) = MSE(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$MSE = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-(p+1)}$$

$$\hat{\beta}_k - t_{1-\frac{\alpha}{2}, n-(p+1)} S.E(\hat{\beta}_k) \le \beta_k \le \hat{\beta}_k + t_{1-\frac{\alpha}{2}, n-(p+1)} S.E(\hat{\beta}_k)$$

a)

Editor >> **Enable commands**

```
MTB > copy c5 c2 c3 c4 m1     X
MTB > tran m1 m2              X'
MTB > mult m2 m1 m3           X'X
MTB > print m3
```

Data Display

```
 Matrix M3
      52  1.57400E+07       383        6
15740042  4.92022E+12  116223168  1857680
     383  1.16223E+08      2864       46
       6  1.85768E+06        46        6
```

$$(X'X)_{4\times4} = \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \sum X_{3i} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \sum X_{1i}X_{3i} \\ \sum X_{2i} & \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{1i}X_{3i} & \sum X_{2i}X_{3i} & \sum X_{3i}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 52 & 1.57400E+07 & 383 & 6 \\ 15740042 & 4.92022E+12 & 116223168 & 1857680 \\ 383 & 1.16223E+08 & 2864 & 46 \\ 6 & 1.85768E+06 & 46 & 6 \end{bmatrix}$$

```
MTB > inver m3 m4
MTB > print m4
```

$(X'X)^{-1}$

Data Display

```
 Matrix M4

 1.86275  -0.0000017  -0.180557   0.047324
-0.00000   0.0000000  -0.000000  -0.000000
-0.18056  -0.0000000   0.025971  -0.007750
 0.04732  -0.0000000  -0.007750   0.191112
```

```
MTB > copy c1 m5
MTB > mult m2 m5 m6
MTB > print m6
```

Y

$X'_{4\times n}Y_{n\times1} = X'Y_{4\times1}$

Data Display

```
 Matrix M6
2.26878E+05
6.88202E+10
1.67289E+06
2.94990E+04
```

$$(X'Y)_{4\times1} = \begin{bmatrix} \sum Y_i \\ \sum Y_iX_{1i} \\ \sum Y_iX_{2i} \\ \sum Y_iX_{3i} \end{bmatrix} = \begin{bmatrix} 226878 \\ 68820200000 \\ 1672890 \\ 29499 \end{bmatrix}$$

```
MTB > mult m4 m6 m7
MTB > print m7
```

$(X'X)^{-1}_{4\times4}(X'Y)_{4\times1} = \mathbf{B}_{4\times1}$

Data Display

```
 Matrix M7
4149.89
   0.00
 -13.17
 623.55
```

$$\mathbf{B}_{4\times1}=\begin{bmatrix}\hat\beta_0\\ \hat\beta_1\\ \hat\beta_2\\ \hat\beta_3\end{bmatrix}=\begin{bmatrix}4149.89\\ 0.00\approx 0.000787\\ -13.17\\ 623.55\end{bmatrix}$$

$$\hat{Y}=4149.89+0.000787\,X_1-13.17\,X_2+623.55\,X_3$$

b)

```
MTB > tran m5 m8
MTB > mult m8 m5 m9
```
Y'

$Y'Y=\sum {Y_i}^2$

```
Answer = 993039576.0000

MTB > mult m1 m7 m10
MTB > copy m10 c6
```
$\hat{Y}_{n\times1}=X_{n\times4}\mathbf{B}_{4\times1}$

```
MTB > Let C7=C1-C6
```
$e=Y-\hat{Y}$

```
MTB > copy c7 m11
```
e

```
MTB > tran m11 m12
```
e'

```
MTB > mult m12 m11 m13
```
$e'e=SSE$

```
Answer = 985529.7464

let c8= 985529.7464/48
```

$MSE=\frac{985529.7464}{52-4}=20531.87$ $MSE=\frac{e'e}{n-P}$

```
MTB > mult 20531.87 m4 m14
```
$V(\mathbf{B})=MSE(X'X)^{-1}$

```
MTB > print m14
```

Data Display

```
 Matrix M14

38245.8   -0.0351482  -3707.17    971.66
   -0.0    0.0000001     -0.00     -0.00
-3707.2   -0.0006762    533.23   -159.12
  971.7   -0.0008312   -159.12   3923.89
```

$$V\begin{bmatrix}\hat\beta_0\\ \hat\beta_1\\ \hat\beta_2\\ \hat\beta_3\end{bmatrix}=\begin{bmatrix}Var(\widehat{\beta_0}) & cov(\widehat{\beta_0},\widehat{\beta_1}) & cov(\widehat{\beta_0},\widehat{\beta_2}) & cov(\widehat{\beta_0},\widehat{\beta_3})\\ cov(\widehat{\beta_0},\widehat{\beta_1}) & Var(\widehat{\beta_1}) & cov(\widehat{\beta_1},\widehat{\beta_2}) & cov(\widehat{\beta_1},\widehat{\beta_3})\\ cov(\widehat{\beta_0},\widehat{\beta_2}) & cov(\widehat{\beta_2},\widehat{\beta_1}) & Var(\widehat{\beta_2}) & cov(\widehat{\beta_2},\widehat{\beta_3})\\ cov(\widehat{\beta_0},\widehat{\beta_3}) & cov(\widehat{\beta_1},\widehat{\beta_3}) & cov(\widehat{\beta_2},\widehat{\beta_3}) & Var(\widehat{\beta_3})\end{bmatrix}_{4\times4}$$

$$=\begin{bmatrix}38245.8 & -0.0351482 & -3707.17 & 971.66\\ -0.0351482 & 0.0000001 & -0.0006762 & -0.0008312\\ -3707.2 & -0.0006762 & 533.23 & -159.12\\ 971.7 & -0.0008312 & -159.12 & 3923.89\end{bmatrix}$$

$$\hat{\beta}_k - t_{1-\frac{\alpha}{2},n-p} S.E(\hat{\beta}_k) \le \beta_k \le \hat{\beta}_k + t_{1-\frac{\alpha}{2},n-p} S.E(\hat{\beta}_k)$$

$t_{1-\frac{\alpha}{2},n-p} = t_{0.975,52-4} = 2.0106$

```
Inverse Cumulative Distribution Function
Student's t distribution with 48 DF
P( X <= x )          x
      0.975   2.01063
```

$S.E(\hat{\beta}_0) = \sqrt{38245.8} = 195.5653$

$S.E(\hat{\beta}_1) = \sqrt{0.0000001} = 0.000316$

$S.E(\hat{\beta}_2) = \sqrt{533.23} = 23.09177$

$S.E(\hat{\beta}_3) = \sqrt{3923.89} = 62.64096$

$\hat{\beta}_0 - t_{0.975,52-4} S.E(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{0.975,52-4} S.E(\hat{\beta}_0)$

$4149.89 - 2.0106 * 195.5653 \le \beta_0 \le 4149.89 + 2.0106 * 195.5653$

$3756.686 \le \beta_0 \le 4543.094$

$\hat{\beta}_1 - t_{0.975,52-4} S.E(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{0.975,52-4} S.E(\hat{\beta}_1)$

$0.000787 - 2.0106 * 0.000316 \le \beta_1 \le 0.000787 + 2.0106 * 0.000316$

$0.000151 \le \beta_1 \le 0.001423$

$\hat{\beta}_2 - t_{0.975,52-4} S.E(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{0.975,52-4} S.E(\hat{\beta}_2)$

$-13.17 - 2.0106 * 23.09177 \le \beta_2 \le -13.17 + 2.0106 * 23.09177$

$-59.5983 \le \beta_2 \le 33.25832$

$\hat{\beta}_3 - t_{0.975,52-4} S.E(\hat{\beta}_3) \le \beta_3 \le \hat{\beta}_3 + t_{0.975,52-4} S.E(\hat{\beta}_3)$

$623.55 - 2.0106 * 62.64096 \le \beta_3 \le 623.55 + 2.0106 * 62.64096$

$497.6041 \le \beta_3 \le 749.4959$

c)

1. Hypothesis

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

2. Test statistic

$$T_0 = \frac{b_k - \beta_k}{s(b_k)} = \frac{b_k}{s(b_k)}$$

3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2},n-p\right)}$

p-value=$2P\left(t_{(n-p)} > |T_0|\right)$

Reject H_0 if $\text{p} - \text{value} \le \alpha$

* Test β_1
1. Hypothesis
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$$T_0 = \frac{b_1 - \beta_1}{s(b_1)} = \frac{b_1}{s(b_1)} = \frac{0.000787}{0.000316} = 2.4887$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-p\right)} = 2.0106$
p-value = $2P\left(t_{(n-p)} > |T_0|\right) = 2P\left(t_{(48)} > |2.4887|\right) = 2(1 - 0.991829) = 0.0164$
reject H_0
* Test β_2
1. Hypothesis
$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$
2. Test statistic
$$T_0 = \frac{b_2}{s(b_2)} = \frac{-13.17}{23.091} = -0.5703$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-p\right)} = 2.0106$
p-value=$2P\left(t_{(n-p)} > |T_0|\right) = 2P\left(t_{(48)} > |-0.5703|\right) = 2(1 - 0.7144) = 0.5712$
not reject H_0
* Test β_3
1. Hypothesis
$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$
2. Test statistic
$$T_0 = \frac{b_3}{s(b_3)} = \frac{623.55}{62.6409} = 9.9543$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-p\right)} = 2.0106$
p-value=$2P\left(t_{(n-p)} > |T_0|\right) = 2P\left(t_{(48)} > |9.9543|\right) = 2(1 - 1) = 0.00$
reject H_0

By using Minitab16:
Stat→*Regression*→General *Regression*→
Response: Y
Model: X1 X2 X3
>>Result >>

√ Regression equation

√ coefficient table

√ (Display CI)

√ Analysis of Variance table

>> ok

By using Minitab17:
Stat→Regression→ Regression→FitRegressionModel

Analysis of Variance

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Regression	3	2176606	2176606	725535	35.3370	0.000000
X1	1	136366	95707	95707	4.6614	0.035877
X2	1	5726	6675	6675	0.3251	0.571227
X3	1	2034514	2034514	2034514	99.0905	0.000000
Error	48	985530	985530	20532		
Total	51	3162136				

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
143.289	68.83%	66.89%	64.78%

Coefficients

Term	Coef	SE Coef	T	P	95% CI
Constant	4149.89	195.565	21.2199	0.000	(3756.68, 4543.10)
X1	0.00	0.000	2.1590	0.036	(0.00, 0.00)
X2	-13.17	23.092	-0.5702	0.571	(-59.60, 33.26)
X3	623.55	62.641	9.9544	0.000	(497.61, 749.50)

Regression Equation
Y = 4150 + 0.000787 X1 - 13.2 X2 + 623.6 X3

Question 2:

Refer to **Grocery retailer**. Assume that regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ for three predictor variables with independent normal error terms is appropriate.

a. Test whether there is a regression relation, using level of significance *0.05*. State the alternatives, decision rule, and conclusion. What does your test result imply about β_1 β_2, and β_3? What is the P-value of the test? (ANOVA Table)

b. Calculate the coefficient of multiple determination R^2. How is this measure interpreted here?

c. Management desires simultaneous interval estimates of the total labor hours for the following typical weekly shipments: $X_{h1} = 302000$, $X_{h2} = 7.20$, $X_{h3} = 0$. Obtain the family of estimates using a 95 percent confidence coefficient.

d. Obtain a 95 percent prediction interval for the mean handling time for three new shipments $X_{h1} = 282000, X_{h2} = 7.10$, and $X_{h3} = 0$.

1- Hypotheses:

$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \quad v.s \quad H_1: at\ least\ one\ of\ \beta_k \neq 0.\ k = 1,2,3$

2- Test Statistic:

$$F = \frac{MSR}{MSE}$$

3- Decision:

Reject H_0 if $F > F_{p-1,n-p,\alpha}$ or $p-value \leq \alpha$

ANOVA Table

Source of Variation	d.f	SS	MS	F
Regression	p-1	SSR=$\sum(\hat{Y}_i - \bar{Y})^2$	$MSR = \frac{SSR}{p-1}$	$\frac{MSR}{MSE}$
Error	n-p	SSE=$\sum(Y_i - \hat{Y}_i)^2$	$MSE = \frac{SSE}{n-p}$	
Total	n-1	SSTo= $\sum(Y_i - \bar{Y})^2$		

$$SSTo = Y'Y - \left(\frac{1}{n}\right)Y'JY$$

$$SSE = Y'Y - b'X'Y$$

$$SSR = b'X'Y - \left(\frac{1}{n}\right)Y'JY$$

$$J_{n\times n} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

a)

```
MTB > copy c5 c2 c3 c4 m1    X
MTB > tran m1 m2             X'
MTB > mult m2 m1 m3          X'X
MTB > inver m3 m4   (X'X)^-1
MTB > copy c1 m5        Y
MTB > mult m2 m5 m6  X'Y
MTB > mult m4 m6 m7     (X'X)^-1_{4x4}(X'Y)_{4x1} = b_{4x1}
MTB > print m7
```

Data Display

```
 Matrix M7

4149.89
   0.00
 -13.17
 623.55
```

```
MTB > tran m5 m8            Y'
MTB > mult m8 m5 m9          Y'Y = ∑Y_i^2
```

```
Answer = 993039576.0000
```

```
MTB > copy c5 m10       I
```

```
MTB > tran m10 m11        I'
MTB > mult m10 m11 m12    II' = J
MTB > mult m8 m12 m13     Y'J
MTB > mult m13 m5 m14     Y'JY
```

```
Answer = 51473626884.0000
```

```
MTB > tran m7 m15         b'
MTB > mult m15 m2 m16     b'X'
MTB > mult m16 m5 m17     b'X'Y
```

```
Answer = 992054046.2536
```

$SSTo = Y'Y - \left(\frac{1}{n}\right)Y'JY = 993039576 - \left(\frac{1}{52}\right)51473626884 = 3162136$

$SSE = Y'Y - b'X'Y = 993039576 - 992054046.2536 = 985529.7464$

$SSR = SSTo - SSE = 3162136 - 985529.7464 = 2176606.177$

Source of Variation	d.f	SS	MS	F
Regression	3	2176606.177	725535.4	35.33703
Error	48	985529.7464	20531.87	
Total	51	3162136		

1- Hypotheses:
$H_0: \beta_1 = \beta_2 = \beta_3 = 0$
$H_1: at\ least\ one\ of\ \beta_k \neq 0.\ k = 1,2,3$
2- Test Statistic:
$F = 35.337$
3- Decision:
Reject H_0 if $F > F_{p-1,n-p,1-\alpha} = F_{3,48,0.95} = 2.79806$
Then we reject H_0. (there is a regression relation)
Or $p - value \leq \alpha = 0.05$
$p - value = P(F_{p-1,n-p} > 35.337) = 1 - 1 = 0.00 < 0.05$, Then we reject H_0

b)
$R^2 = \frac{SSR}{SSTo} = \frac{2176606.177}{3162136} = 0.68833 = 68.83\%$
Thus, when the three predictor variables, the number of cases shipped, the indirect cost of the total labor hours as percentage and the qualitative predictor which call holiday, are considered, the variation in the total labor hours is reduced by 68.83 percent.

By using Minitab:

Stat → Regression → Regression → Fit Regression Model

Regression Analysis: Y versus X1, X2, X3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2176606	725535	35.34	0.000
X1	1	95707	95707	4.66	0.036
X2	1	6675	6675	0.33	0.571
X3	1	2034514	2034514	99.09	0.000
Error	48	985530	20532		
Total	51	3162136			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
143.289	68.83%	66.89%	64.78%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4150	196	21.22	0.000	
X1	0.000787	0.000365	2.16	0.036	1.01
X2	-13.2	23.1	-0.57	0.571	1.02
X3	623.6	62.6	9.95	0.000	1.01

Regression Equation

```
Y = 4150 + 0.000787 X1 - 13.2 X2 + 623.6 X3
```

c) Management desires simultaneous interval estimates of the total labor hours for the following typical weekly shipments: $X_{h1} = 302000$, $X_{h2} = 7.20$, $X_{h3} = 0$
Obtain the family of estimates using a 95 percent confidence coefficient.

$$\hat{y}_h \pm t_{1-\frac{\alpha}{2},n-p} S.E(\hat{y}_h)$$

$$S.E(\hat{y}_h) = \sqrt{Var(\hat{y}_h)}$$

$$Var(\hat{y}_h) = X_h' \boldsymbol{V}(\hat{\boldsymbol{\beta}}) X_h$$

C6
1
302000
7.20
0

```
MTB > copy c6 m24
MTB > tran m24 m25
MTB > mult m25 m7 m26
Answer = 4292.7901
```

X_h ; X_h' ; $\hat{y}_h = X_h' \mathrm{B}_{4\times 1}$

```
MTB > mult 20531.87 m4 m19
MTB > mult m25 m19 m27
```

$V(\mathrm{B}) = MSE(X'X)^{-1}$; $X_h' V(\mathrm{B})$

```
MTB > mult m27 m24 m28
```
$X_h{}'V(\mathrm{B})X_h$

```
Answer = 456.1072
```

$Var(\hat{y}_h) = X_h' \boldsymbol{V}(\widehat{\boldsymbol{\beta}})X_h = 456.1072$

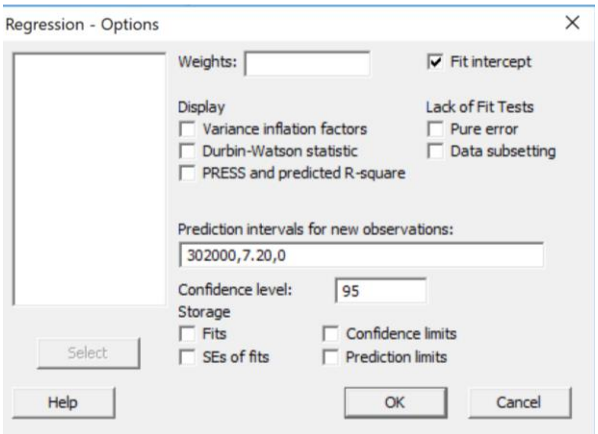
$S.E(\hat{y}_h) = \sqrt{456.1072} = 21.3567$

$t_{1-\frac{\alpha}{2},n-p} = t_{0.975,48} = 2.01063$

$4292.7901 \pm 2.01063(21.3567)$

$4249.8497 < E(Y_h) < 4335.7305$

By use Minitab17:
Stat → Regression → Regression → Predict
Or By use Minitab16:
Stat → Regression → Regression → option



Prediction for Y

```
Regression Equation
Y = 4150 + 0.000787 X1 - 13.2 X2 + 623.6 X3

Variable  Setting
X1         302000
X2            7.2
X3              0

    Fit   SE Fit          95% CI                95% PI
4292.79  21.3567  (4249.85, 4335.73)  (4001.50, 4584.08)
```

d) Obtain a 95 percent prediction interval for the mean handling time for three new shipments $X_{h1} = 282000, X_{h2} = 7.10$, and $X_{h3} = 0$.

$\hat{y}_h \pm t_{1-\frac{\alpha}{2},n-p} S.E(\hat{y}_{new})$

$S.E(\hat{y}_{new}) = \sqrt{Var(\hat{y}_{new})}$

$Var(\hat{y}_{new}) = MSE + X_h' \boldsymbol{V}(\widehat{\boldsymbol{\beta}})X_h$

C7
1
282000
7.10
0

```
MTB > copy c7 m18
MTB > mult 20531.87 m4 m19
MTB > tran m18 m20
MTB > mult m20 m19 m21
MTB > mult m21 m18 m22
```
(X_h; $V(\mathrm{B}) = MSE(X'X)^{-1}$; $X_h{}'$; $X_h{}'V(\mathrm{B})$; $X_h{}'V(\mathrm{B})X_h$)

```
Answer = 521.5551
MTB > mult m20 m7 m23
```
$\hat{y}_h$

```
Answer = 4278.3651
```

$$Var(\hat{y}_{new}) = MSE + X_h' \boldsymbol{V}(\widehat{\boldsymbol{\beta}}) X_h = 20531.87 + 521.5551 = 21053.42482$$

$$S.E(\hat{y}_{new}) = \sqrt{21053.42482} = 145.098$$

$$t_{1-\frac{\alpha}{2}, n-p} = t_{0.975,48} = 2.01063$$

$4278.3651 \pm 2.01063(145.098)$

$$\boldsymbol{3986.626748 < Y_{h(new)} < 4570.103452}$$

By use Minitab :

Stat → Regression → Regression → Predict

```
Variable  Setting
X1         282000
X2            7.1
X3              0

    Fit    SE Fit        95% CI                95% PI
4278.37   22.8376  (4232.45, 4324.28)   (3986.63, 4570.10)
```