classmate

Date
Page

Regression Module-1

→ Non-deterministic: $y = \beta_0 + \beta_1 x + \epsilon$

→ Linear models & regression analysis:

$y = (x_1, x_2 \dots x_k)$

$y = f(x_1, x_2 \dots x_k, \beta_1, \beta_2, \beta_3 \dots \beta_k) + \epsilon$ → stochastic term

$x_1 \dots x_k$ ↓ independent; $\beta_1 \dots \beta_k$ ↓ parameters

$e = 0$ → mathematical model

$e \neq 0$ → statistical model

(i) $y = \beta_1 x^2 + \beta_2 x + \beta_3 \log x + \epsilon$ (linear model)

$\frac{\partial y}{\partial \beta_1} = x^2 +$ " " $\quad \frac{\partial y}{\partial \beta_2} = \beta_1 x^2 + x + \dots \quad \frac{\partial y}{\partial \beta_3} = \beta_1 x^2 + x + \log x + e$

$\therefore \frac{\partial y}{\partial \beta_1}$ is independent of β_i

→ Note: If the parameters are linear (power 1) then it is linear model

→ Func for linear model is $f(x_1, x_2 \dots x_n, \beta_1, \beta_2 \dots \beta_n) = \beta_1 x_1 + \dots + \beta_n x_n$

Regression Analysis:

→ Simple linear regression model:

Model def: $y = \beta_0 + \beta_1 x + e$

y: dependent var $\quad \beta_0$: intercept reg coff $\quad e$: error term

x: independent var $\quad \beta_1$: slope

→ $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$E(\epsilon_i) = 0 \Rightarrow$ avg of $\epsilon_i = 0$

$Var(\epsilon_i) = \sigma^2$; $cov(\epsilon_i, \epsilon_j) = 0$

→ $E(y) = E(\beta_0 + \beta_1 x + \epsilon) \qquad V(y) = V(\beta_0 + \beta_1 x + \epsilon)$

$E(y) = \beta_0 + \beta_1 x \qquad V(y) = \sigma^2$

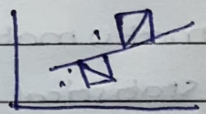
- Least square estimation:

  $\Sigma \epsilon_i^2$ is minimum (vertical)

  by which we can find β_0, β_1
- Reverse regression method:

  the sum of squares of difference (horizontal)
- Orthogonal:

  sum of squares of ⊥lar distance
- Reduced Major axis regression:



- Least absolute deviation regression:

  $\min \Sigma y_i - \beta_0 - \beta_1 x_i$

* Least Square method:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i = y_i - \beta_0 - \beta_1 x_i$$

$$S(\beta_0, \beta_1) = \Sigma \epsilon_i^2 = \Sigma (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 2 \Sigma (y_i - \beta_0 - \beta_1 x_i)(-1)$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 2 \Sigma (y_i - \beta_0 - \beta_1 x_i)(-x_i)$$

The least square estimators of β_0 and β_1 are $\hat{\beta}_0$ and $\hat{\beta}_1$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 0$$

$$2 (\Sigma y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0$$

$$\Sigma (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\Sigma y_i - \Sigma \hat{\beta}_0 - \Sigma \hat{\beta}_1 x_i = 0$$

$$n\hat{\beta}_0 = \Sigma y_i - \hat{\beta}_1 \Sigma x_i$$

$$\hat{\beta}_0 = \frac{\Sigma y_i}{n} - \hat{\beta}_1 \frac{\Sigma x_i}{n}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 0$$

$$-2 \Sigma (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

$$\Sigma (y_i x_i - \hat{\beta}_0 x_i - \hat{\beta}_1 x_i^2) = 0$$

$$\Sigma x_i y_i - \Sigma \hat{\beta}_0 x_i = \Sigma \hat{\beta}_1 x_i^2$$

$$\hat{\beta}_1 = \frac{\Sigma x_i y_i - \hat{\beta}_0 \Sigma x_i}{\Sigma x_i^2}$$

$$\Rightarrow \Sigma x_i^2 \hat{\beta}_1 = \Sigma x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \Sigma x_i$$

$$= \Sigma x_i y_i - \bar{y} \Sigma x_i + \hat{\beta}_1 \Sigma x_i \cdot \bar{x}$$

$$\Rightarrow \hat{\beta}_1 (\Sigma x_i^2 - \bar{x} \Sigma x_i) = \Sigma x_i y_i - \bar{y} \Sigma x_i$$

Date
Page

$$\hat{\beta}_1 = \frac{\Sigma x_i y_i - \bar{y}\,\Sigma x_i}{\Sigma x_i^2 - \bar{x}\,\Sigma x_i} \quad , \quad \beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}$$

$$= \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

where $S_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y})$ and $S_{xx} = \Sigma (x_i - \bar{x})^2$

- $E(y) = \beta_0 \Rightarrow$ avg value of response variable when the independent var takes the value of zero

- $\frac{d}{dx}(E(y)) = \beta_1 \Rightarrow$ rate of change in the avg value of response var when there is a unit change in the value of independent var.

Important formulas.

- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ where $S_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y}) = \Sigma y_i (x_i - \bar{x})$
  $S_{xx} = \Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - n\bar{x}^2$

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

- $e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$
- $E(\hat{\beta}_1) = \beta_1 , E(\hat{\beta}_0) = \beta_0$

- $var(\hat{\beta}_1) = \sigma^2 \Sigma c_i^2 = \frac{\sigma^2}{S_{xx}}$

  $var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$

- $\hat{\beta}_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n.\Sigma x_i^2 - (\Sigma x_i)^2}$

Gauss markov theorem:

for the linear regression model with the assumptions $E(\epsilon)=0$, $Var(\epsilon)=\sigma^2$ & uncorrelated errors, the least squares estimators are unbiased and have min variance among all unbiased estimators that are linear combinations of y_i's

Properties of residuals:

- sum of residuals is zero $\Sigma e_i = 0$
- sum of observed values equals to Predicted values

$$\Sigma y_i = \Sigma \hat{y}_i$$

- sum of residuals weighted by corresponding regressor values equals zero $\Sigma x_i e_i = 0$
- corresponding fitted values equals to 0 $\Sigma \hat{y}_i e_i = 0$

Estimation of σ^2:

$$SS_{Res} = \Sigma e_i^2 = \Sigma (y_i - \hat{y}_i)^2$$

$$SS_{Res} = \Sigma y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy}$$

where $\Sigma (y_i - \bar{y})^2 = SS_T = \Sigma y_i^2 - n\bar{y}^2$

$$S_{xy} = \Sigma (x_i - \bar{x})(y_i - \bar{y})$$

$$SS_{Res} = SS_T - \hat{\beta}_1 S_{xy}$$

$$\hat{\sigma}^2 = \frac{SS_{Res}}{(n-2)}$$ where n is no.of data values

Dof: n − no.of parameters

$$MS_{Res} = \hat{\sigma}^2 = \frac{SS_{Res}}{n-2}$$

is the residual mean square

classmate

Date
Page

→ $E(\hat{\sigma}^2) = \sigma^2$

→ $E(SS_{Res}) = (n-2)\sigma^2$

→ ~~[illegible]~~

Q:

x	50	60	70	80	90	100
y	80	55	45	35	25	22

① Fit least square ② predict for 110 ③ 4

sol:

x	y	xy	x^2	y^2
50	80	4000	2500	6400
60	55	3300	3600	3025
70	45	3150	4900	2025
80	35	2800	6400	1225
90	25	2250	8100	625
100	22	2200	10000	484
450	262	17700	35500	13784

$\bar{x} = 75 \quad \bar{y} = 43.67$

$$\hat{\beta}_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{6(17700) - (262)(450)}{6(35500) - (450)^2}$$

$$= \frac{106200 - 117900}{213000 - 202500}$$

$$= \frac{-11700}{10500} = -1.1142$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= 43.67 + (1.1142)(75)$$

$$= 127.238$$

$\hat{y} = 127.238 - 1.1142\,x_i$

for $x = 110$:

$\hat{y} = 127.238 - 122.5620$

$= 4.676$

$\hat{y}_1 = 127.~~2380~~ - 1.1142(50) = 71.5280$

$\hat{y}_2 = 60.386$

$\hat{y}_3 = 49.244$

$\hat{y}_4 = 38.102$

$\hat{y}_5 = 27.102$

$\hat{y}_6 = 15.96$

$e_1 = 8.472$

$e_2 = -5.386$

$e_3 = -4.244$

$e_4 = -3.102$

$e_5 = -2.102$

$e_6 = 6.04$

$$\hat{\sigma}_0^2 = \frac{SS_{Res}}{(n-2)} = \frac{\Sigma e_i^2}{(n-2)} = \frac{169.4025}{4} = 42.3506$$

Hypothesis testing:

slope:

- $H_0: \beta_1 = \beta_{10}$, $H_1: \beta_1 \neq \beta_{10}$

- $H_0: \beta_p = \beta_{10}$ is true

$$z_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \sim N(0,1)$$

- ~~as~~ since MS_{Res} is an unbiased estimator of σ^2

- test statistic:

$$t_{\beta_1} = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\frac{MS_{Res}}{S_{xx}}}} \sim t_{n-2}$$

Reject if:

$$|t_0| > t_{\alpha/2, n-2}$$

→ Standard error of slope $se(\hat{\beta}_1) = \sqrt{\dfrac{MS_{Res}}{S_{xx}}}$

Intercept: ~~ttwo~~

$H_0: \beta_0 = \beta_{00}$ $\quad H_1: \beta_0 \neq \beta_{00}$

→ $t_0 = \dfrac{\hat{\beta_0} - \beta_{00}}{se(\hat{\beta_0})}$ where $se(\hat{\beta_0}) = \sqrt{MS_{Res}\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}\right)}$

→ Reject H_0: (two tailed)

$$|t_0| > t_{\alpha/2, n-2}$$

→ Reject H_0: (one-tailed)

$t_0 > t_{\alpha/2}, (n-2)$ (upper tail)

$t_0 < t_{\alpha}, (n-2)$ (lower tail)

Confidence intervals for β_0 and β_1 (procedure):

→ $P(L \leq \beta_0, \leq U) = (1-\alpha) \times 100\%$ confidence interval for β_1 is given by

$$(L, U) = \left(\hat{\beta_1} - t_{\alpha/2, n-2}\sqrt{\frac{MSE}{S_{xx}}},\ \hat{\beta_1} + t_{\alpha/2, n-2}\sqrt{\frac{MSE}{S_{xx}}}\right)$$

Q: Give estimates for β_0 & β_1. Find $MSE(\hat{r}^2)$

x	y	xy	x^2
-1	-5	5	1
0	-4	0	0
2	2	4	4
-2	-7	14	4
5	6	30	25
6	9	54	36
8	13	104	64
11	21	231	121
12	20	240	144
-3 38	-9 46	27	9 408

$$\hat{\beta}_1 = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{+5342}{2636} = 2.0266$$

$$\hat{\beta_0} = 4.6 - (3.8)(2.0266)$$

$$= 3.1011$$

$$MSE = 0.973785$$

classmate

Date

Page

Q: Using the data given in above test the hypothesis $H_0: \beta_0 = -3$ vs $H_1: \beta_0 \neq -3$ at the 0.05 level of significance

$\beta_{00} = -3$

$$t_{\beta_0} = \frac{\hat{\beta_0} - \beta_{00}}{\sqrt{MS_{Res}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}$$

$$= \frac{(-3.1011) - (-3)}{\sqrt{0.923285\left(\frac{1}{10} + \frac{(3.8)^2}{263.6}\right)}} = -0.26041$$

$\alpha = 0.05$, $n = 10$

$t_{\alpha, n-2} = t_{0.05, 8} = 2.306$

$|t_{\beta_0}| < t$

$\Rightarrow$ Not in rejection area

$\therefore$ accept H_0

Q. $H_0: \beta_1 = 2$ vs $H_0: \beta_1 \neq 2$ at $\alpha = 0.05$

$\beta_{10} = 2$.

$$t_{\beta_1} = \frac{\hat{\beta_1} - \beta_0}{\sqrt{\frac{MS_{Res}}{S_{xx}}}} = \frac{2.0266 - 2}{\sqrt{\frac{0.973785}{263.6}}}$$

$$= \frac{0.0266}{0.0608} = 0.4375$$

$t_{\alpha, n-2} =$

classmate

Date ______
Page ______

Q: ① fit regression, find $\hat{r}^2$ & MS_{Res}

② $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ & $\alpha = 5\%$.

③ $H_0: \beta_0 = 3$, $H_0: \beta_0 \neq 3$, $\alpha = 2\%$.

sol:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
15.5	2158.7	240.25	4659985.69	33459.85
23.75	1678.15	564.0625	2816187.4225	39856.0625
8	2316	64	5363856	18528
17	2061.3	289	4248957.69	35042.1
5.5	2207.5	30.25	4873056.25	12141.25
19	1708.3	361	2918288.89	32457.7
24	1784.7	576	3185154.09	42832.8
2.5	2575	6.25	6630625	6437.5
7.5	2357.9	56.25	5559692.41	17684.25
11	2256.7	121	5092694.89	24823.7
13	2165.2	169	4688091.04	28147.6
3.75	2399.55	14.0625	5757840.2025	8998.3125
25	1779.8	625	3167688.04	44495
9.75	2336.75	95.0625	5460400.5625	22783.3125
22	1765.3	484	3116284.09	38836.6
18	2053.5	324	4216862.25	36963
6	2414.4	36	5829327.36	14486.4
12.5	2200.5	156.25	4842200.25	27506.25
2	2654.2	4	7044777.64	5308.4
21.5	1753.7	462.25	3075463.69	37704.55
267.25	42627.15	4677.6875	92547433.46	528492.6375

① $\bar{y} = \frac{\Sigma y}{n} = 2131.3575$ $\quad \bar{x} = 13.3625$

$$\hat{\beta}_1 = \frac{n\Sigma x_i y_i - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{10569852.75 - 11392105.84}{93553.75 - 71422.5625}$$

$$= \frac{-822253.0875}{22131.1875} = -37.154$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$= 2131.3575 + (37.154)(13.3625)$$
$$= \cancel{1630.8872}\ 2627.8278$$
$$\hat{y} = 2627.8278 - 37.154x$$

$$SS_T = \Sigma y_i^2 - n\bar{y}^2$$
$$= 1693737.604$$
$$S_{xy} = \cancel{580056.6785}\ -41108.2247$$

$$\hat{\sigma}^2 = \frac{SS_T - \hat{\beta}_1 S_{xy}}{n-2} = 9244.59$$

(2). $H_0: \beta_1 = 0$ and $H_1: \beta_1 \neq 0$

$$t_{\beta_0} = \frac{\hat{\beta}_0 - \beta_{10}}{\sqrt{\frac{MS_{res}}{S_{xx}}}} = \frac{-37.154 - 0}{\sqrt{\frac{9244.59}{1106.5544}}}$$

$$= \frac{-37.154}{2.8904} = -12.8553$$

$t_{0.05,18} = 2.101$

(3) $\alpha = 95\%$, $t_{\beta_1} = -12.8533$,

$$(L, U) = \left(\hat{\beta}_1 - t_{\alpha/2, n-2}\sqrt{\frac{MSE}{SSx}},\ \hat{\beta}_1 + t_{\alpha/2, n-2}\sqrt{\frac{MSE}{Sxx}}\right)$$

$$(L, U) = (-37.154 - 6.0727,\ -37.154 + 6.0727)$$

$$= (-43.223, -31.0813)$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad SSE = \Sigma(y_i - \hat{y})^2$$

classmate

Date

Page

Multiple linear regression:

→ General form is $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$

y: responsible variable

$x_1, x_2 \dots x_k$: Predictor variable

$\beta_0, \beta_1 \dots \beta_k$: Regression coff

ϵ: Random error term

Ex: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$ (combined effect of predictors)

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon$ (second-order)

Data table and model assumptions:

ob(i)					
1	x_{11}	x_{12}	$\dots$	x_{1k}	y_1
2	x_{21}	x_{22}	$\dots$	x_{2k}	y_2
⋮	⋮	⋮		⋮	⋮
n	x_{n1}	x_{n2}	$\dots$	x_{nk}	y_n

$E(\epsilon) = 0$

$E(\epsilon\epsilon') = \sigma^2 I_n$

model;

→ $E(\epsilon) = 0$, $var(\epsilon) = \sigma^2$, errors are uncorrelated

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji} + \epsilon_i$$

y_i : observed response for ith obser

x_{ij} : value of j^{th} regression for ith

β_j : regression coff

ϵ : error term

least square func:

$$S(\beta_0, \beta_1 \dots \beta_j) = \sum \epsilon_i^2$$

$$\epsilon_i = y_i - \beta_0 - \sum \beta_j x_{ij} = S(\beta)$$

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_j \beta_j x_{ij} \right) = 0$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_i x_{ij} \left(y_i - \beta_0 - \sum_j \beta_j x_{ij} \right) = 0$$

$$\Rightarrow \sum y_i = \beta_0 + \sum_j \beta_j x_{ij}$$

$$\sum_i x_{ij} y_j = \beta_0 \sum_i x_{ij} + \beta_1 \sum_{i=1}^{n} x_{ij} x_{ij} + \dots + \beta_k \sum_{i=1}^{n} x_{ij}{}^2$$

Regression in matrix form:

$$y = X\beta + \epsilon$$

$y : (n \times 1)$ $\quad X : (n \times (k+1))$ $\quad \beta : ((k+1) \times 1)$ $\quad \epsilon : (n \times 1)$

$$S(\beta) = \sum \epsilon_i^2 = \epsilon' \epsilon = (y - X\beta)'(y - X\beta)$$
$$= y'y - 2\beta' X y + \beta' X' X \beta$$

Note: $(\beta' x' y)' = y' x \beta$

$$\frac{\partial S}{\partial \beta} = 0 \Rightarrow -2X'y + 2X'X\beta = 0$$
$$X'X\beta = X'y$$
$$\beta = (X'X)^{-1} X'y$$

Matrix form of normal eq

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \cdots & \sum x_{ik} x_{i1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum x_{ik} & \sum x_{i1} x_{ik} & \sum x_{i1} x_{i2} & \cdots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \vdots \\ \sum x_{ik} y_i \end{bmatrix}$$

classmate

Date

Page

The vector of fitted values is

$$\hat{y} = X\beta = X(X'X)^{-1}X'y = Hy$$

where $H = X(X'X)^{-1}X'y$

→ residuals $e = y - \hat{y}$

→ $H = H'$, $H^2 = H$, $\hat{y} = Hy$

Q

x_1	x_2	y_1
1	4	1
2	5	6
3	8	8
4	2	12

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$$A = \begin{bmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 3 & 8 \\ 1 & 4 & 2 \end{bmatrix} \qquad A' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 4 & 5 & 8 & 2 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 4 & 10 & 19 \\ 10 & 30 & 46 \\ 19 & 46 & 109 \end{bmatrix}$$

$$A^T y = \begin{bmatrix} 27 \\ 85 \\ 122 \end{bmatrix}$$

$$\hat{\beta} = (A^T A)^{-1} (A^T y)$$

$$= \begin{bmatrix} 577/183 & -36/61 & -55/183 \\ -36/61 & 25/122 & 1/61 \\ -55/183 & 1/61 & 10/183 \end{bmatrix} \begin{bmatrix} 27 \\ 85 \\ 122 \end{bmatrix} = \begin{bmatrix} -1.699 \\ 3.4836 \\ -9.831 \end{bmatrix}$$

Date
Page

Q:

y	x_1	x_2
3	2	1
2	3	5
4	5	3
5	7	6
8	8	7

$$y = \begin{bmatrix} 3 \\ 2 \\ 4 \\ 5 \\ 8 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 5 \\ 1 & 5 & 3 \\ 1 & 7 & 6 \\ 1 & 8 & 7 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 7 & 8 \\ 1 & 5 & 3 & 6 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 5 \\ 1 & 5 & 3 \\ 1 & 7 & 6 \\ 1 & 8 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 25 & 22 \\ 25 & 151 & 130 \\ 22 & 130 & 120 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 305/254 & -35/254 & -9/127 \\ -35/254 & 29/254 & -25/254 \\ -9/127 & -25/254 & 65/508 \end{bmatrix}$$

$$X'y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 7 & 8 \\ 1 & 5 & 3 & 6 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 4 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 22 \\ 131 \\ 111 \end{bmatrix}$$

$$\hat{\beta} = (X'X)^{-1}(X'y)$$

$$\hat{\beta} = \begin{bmatrix} 1/2 \\ 1 \\ 1/4 \end{bmatrix}$$

$$\hat{y} = X\hat{\beta} = \hat{\beta_0} + \hat{\beta} x_1 + \hat{\beta_2} x_2$$

$$= 1/2 + x_1 - 1/4 x_2 = 0.5 + x_1 - 0.25 x_2$$

$\hat{y_1} = 0.5 + 2 - 0.25$
$= 2.25$ $\qquad e_1 = -0.75$

$\hat{y_2} = 2.25$ $\qquad e_2 = -0.25$

$\hat{y_3} = 4.75$ $\qquad e_3 = -0.75$

$\hat{y_4} = 6$ $\qquad e_4 = -1$

$\hat{y_5} = 6.75$ $\qquad e_5 = -1.25$

classmate

Date ______
Page ______

(2)

y	x_1	x_2
100	1	1
80	5	1
104	5	2
94	10	2
130	20	3

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 5 & 2 \\ 1 & 10 & 2 \\ 1 & 20 & 3 \end{bmatrix} \quad X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 5 & 5 & 10 & 20 \\ 1 & 1 & 2 & 2 & 3 \end{bmatrix} \quad y = \begin{bmatrix} 100 \\ 80 \\ 104 \\ 94 \\ 130 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 5 & 5 & 10 & 20 \\ 1 & 1 & 2 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 5 & 2 \\ 1 & 10 & 2 \\ 1 & 20 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1+1+1+1+1 & 1+5+5+10+20 & 1+1+2+2+3 \\ 1+5+5+10+20 & 1+25+25+100+400 & 1+5+10+20+60 \\ 1+1+2+2+3 & 1+5+10+20+60 & 1+1+4+4+9 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 41 & 9 \\ 41 & 551 & 96 \\ 9 & 96 & 19 \end{bmatrix}$$

$$X^Ty = \begin{bmatrix} 100+80+104+94+130 \\ 100+400+520+940+2600 \\ 100+80+208+188+390 \end{bmatrix} = \begin{bmatrix} 508 \\ 4560 \\ 966 \end{bmatrix}$$

$$(X^TX)^{-1} = \frac{Adj(X^TX)}{|X^TX|} = \begin{bmatrix} 1253 & 85 & -1023 \\ 85 & +14 & +111 \\ -1023 & -111 & 1074 \end{bmatrix} \frac{1}{543}$$

$$\beta = (X^TX)^{-1}(X^Ty) = \begin{bmatrix} 66.125 \\ -0.371 \\ 21.436 \end{bmatrix}$$

classmate

Date

Page

(3)

y	x	$x_i y_i$	x_i^2	$\hat{y}_i$	ϵ_i^2
49	40	1960	1600	36.8718	147.8218
38	40	1520	1600	36.8718	1.2728
27	40	1080	1600	36.8718	97.4524
24	55	1320	3025	30.9618	48.4667
38	55	2090	3025	30.9618	49.5363
33	55	1815	3025	30.9618	4.1543
19	70	1330	4900	25.0518	36.6243
28	70	1960	4900	25.0518	8.6916
16	70	1120	4900	25.0518	81.9351
18	90	1620	8100	17.1718	0.6859
23	90	2070	8100	17.1718	33.9679
313	675	17885	44775		501.1764

$\bar{y} = 28.4545$ $\quad \bar{x} = 61.3636$

$$\hat{\beta}_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$= \frac{11 \times 17885 - (675)(313)}{11 \times 44775 - (675)^2} = \frac{-14540}{36900} = -0.3940$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= 28.4545 + (0.3940)(61.3636) = 52.6318$$

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_i = 52.6318 - 0.394(x_i)$$

$$SSE = \Sigma e_i^2 = \Sigma(y_i - \hat{y}_i)^2 = 501.1764$$

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = 55.6863$$

$$MSE = \frac{SS_{Res}}{n-2} = \hat{\sigma}^2$$

$S_{xx} = 36900$