

STATS 3N03/8J04

2004-11-29

32-1

SIMPLE LINEAR REGRESSION
ANOVA

x	40	155	260	325	420
y	2.5	3.1	3.6	3.9	5.0

x = PRESSURE

y = EXTRACTION TIME

$$n = 5$$

$$\sum x = 1200$$

$$\sum y = 18.1$$

$$\sum x^2 = 375250$$

$$\sum xy = 4884 \quad \sum y^2 = 69.03$$

MODEL:

$$E(Y|x) = \beta_0 + \beta_1 x$$

$$\text{OR} \quad a + b x$$

$$\text{OR} \quad A + B x$$

CALCULATOR:

$$\hat{\beta}_0 = 2.13461$$

$$\hat{\beta}_1 = 0.0061891$$

32-2

THEN GET:

$$SS_T = (n-1) \hat{\sigma}_y^2 = 3.508$$

$$S_{xx} = (n-1) \hat{\sigma}_x^2 = 87250$$

$$SS_\beta = \hat{\beta}_1^2 S_{xx} = 3.34212$$

$$SSE = SS_T - SS_\beta = 0.16588$$

SV	SS	DF	MS	F	P
SLOPE	3.34212	1	3.34212	60.44	0.004
RESIDUAL	0.16588	3	0.05529		
TOTAL	3.508	4			

CONCLUSIONS?

32-3

NOTATION:

FITTED LINE AT i^{th} OBS

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

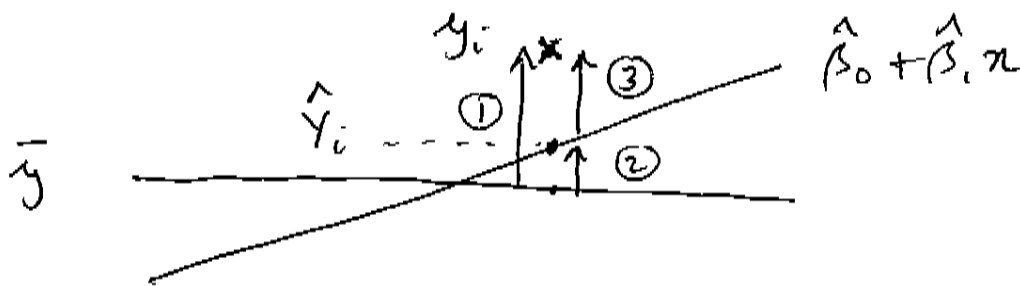
① ② ③

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

$$(n-1) = 1 + (n-2)$$

$$SS_T = SS_B + SSE$$

INDIVIDUAL OBS. FROM GRAND MEAN
FITTED LINE FROM GRAND MEAN (FLAT LINE)
INDIVIDUAL OBS. FROM FITTED LINE.



32-4

REPEATED x -VALUES

- LACK OF FIT TEST
- MEASURE OF PURE ERROR
- TEST FOR NON-LINEARITY

DATA CAN BE ANALYSED AS
SIMPLE LINEAR REGRESSION
OR AS SINGLE-FACTOR

- DO BOTH

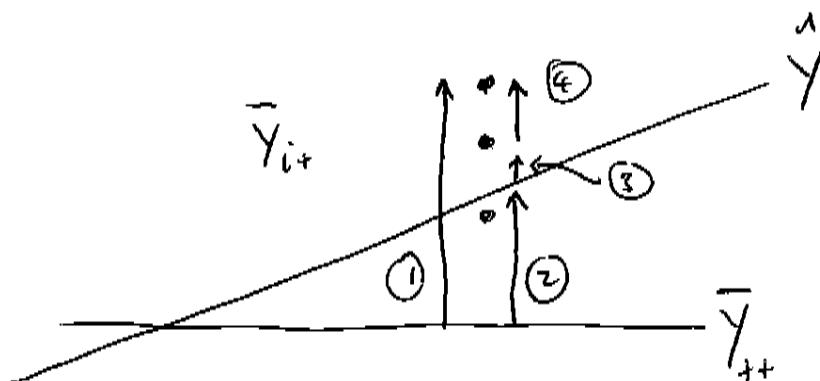
- DIFFERENCE BETWEEN
RESIDUAL FROM SLR
AND PURE ERROR FROM
SINGLE-FACTOR MEASURES
LACK OF FIT.

32-5

EX STRENGTH OF PARTICLE BOARD AS A FUNCTION OF % HARD WOOD

CONC:	5%	10%	15%	20%
	7	12	14	19
	8	17	18	25
	15	13	19	22
	11	18	17	23

$$\underbrace{(Y_{ij} - \bar{Y}_{++})}_{(1)} = \underbrace{(\hat{Y}_i - \bar{Y}_{++})}_{(2)} + \underbrace{(\bar{Y}_{it} - \hat{Y}_i)}_{(3)} + \underbrace{(Y_{ij} - \bar{Y}_{it})}_{(4)}$$



- ① OBS. FROM GRAND MEAN "TOTAL"
- ② FITTED LINE FROM GRAND MEAN "SLOPE"
- ③ GROUP MEAN FROM FITTED LINE "NON-LIN."
- ④ OBS. FROM GROUP MEAN "PURE ERROR"

32-6

```
> board
  stren conc
1      7    5
2      8    5
3     15    5
4     11    5
5     12   10
6     17   10
7     13   10
8     18   10
9     14   15
10    18   15
11    19   15
12    17   15
13    19   20
14    25   20
15    22   20
16    23   20
```

VALID ONLY IF RELATIONSHIP
IS A STRAIGHT LINE
"SIMPLE LINEAR
REGRESSION"

```
> anova(lm(stren~conc, board))
Analysis of Variance Table
```

Response: stren

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
conc	1	288.800	288.800	38.525	2.287e-05 ***
Residuals	14	104.950	7.496		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> anova(lm(stren~as.factor(conc), board))
Analysis of Variance Table
```

"SINGLE-FACTOR"

Response: stren

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(conc)	3	296.250	98.750	12.154	0.0005999 ***
Residuals	12	97.500	8.125		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

32-7

```
> anova(lm(stren~conc+as.factor(conc), board))
```

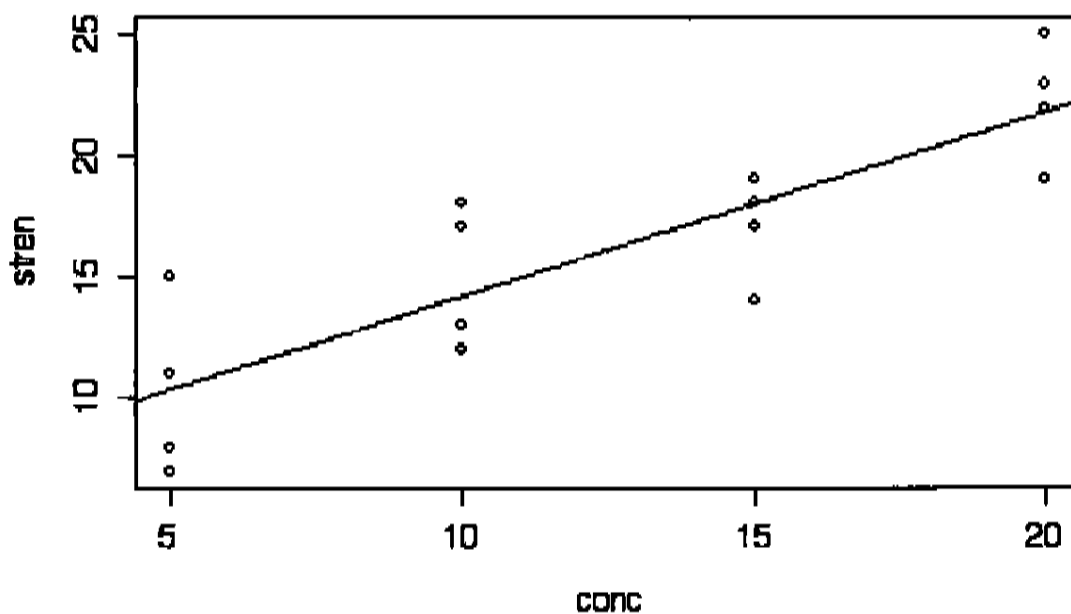
Analysis of Variance Table**Response: stren**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
conc	1	288.800	288.800	35.5446	6.591e-05 ***
as.factor(conc)	2	7.450	3.725	0.4585	0.6429
Residuals	12	97.500	8.125		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> plot(stren~conc, board)
```

```
> abline(lm(stren~conc, board))
```



32-8

RESIDUAL :

WHAT VARIATION IS LEFT
OVER AFTER FITTING THE
MODEL.

PURE ERROR:

RESIDUAL WHEN THE MODEL
ACCOUNTS FOR ALL POSSIBLE
SOURCES OF SYSTEMATIC
VARIATION

RESIDUAL OR ERROR MEAN SQUARE
IS ALWAYS THE BOTTOM ONE
ON THE TABLE.

TEST HYPOTHESES FROM BOTTOM
UP, STOP WHEN A LINE IS
SIGNIFICANT.

32-9

SV	H_0
FACTOR A	"FACTOR A HAS NO EFFECT"
FACTOR B	"FACTOR B HAS NO EFFECT"
INTERACTION AxB	"NO INTERACTION"
ERROR	

SV	H_0
REGRESSION	"SLOPE IS 0"
NON LINEARITY (PURE) ERROR	"STRAIGHT LINE IS ADEQUATE (OVER THIS RANGE)"