One-Way (Single Factor) ANOVA Table

■ **TABLE 3.3**

The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
			SS/doF	
Between treatments	$SS_{\text{Treatments}} = n\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y}_{..})^2$	$k - 1$	$MS_{\text{Treatments}}$	$F_0 = \dfrac{MS_{\text{Treatments}}}{MS_E}$
Error (within treatments)	$SS_E = SS_T - SS_{\text{Treatments}}$	$N - k$	MS_E	
Total	$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$N-1$		

compare w/t

$F_{crit} = F_{1-\alpha, k-1, N-k}$

$F_0 > F_{crit} \rightarrow H_1$ ✓ other H_0 ✓

Example

$k =$ 1, 2, 3, 4

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	$\boxed{63}$	71	71	60
4	59	64	$\boxed{67}$	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

$n_1 = 4$ $n_2 = 6$ $n_3 = 6$ $n_4 = 8$

$$y_{ij}$$

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

y_{ij} $\quad i \to 1 - k \quad k = 4$

$\quad j \to 1 - n_i$

e.g. $y_{13} = 63$, $y_{34} = 67$

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$ (effects model)

$\tau_i \to 0$

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_i - \bar{\bar{y}}) + (y_{ij} - \bar{y}_i)$$

$$N = \sum^k n_i = 4+6+6+8 = 24$$

$$\sum\sum y_{ij}^2 = N\bar{\bar{y}}^2 + \sum^k n_i(\bar{y}_i - \bar{\bar{y}})^2 + \sum\sum(y_{ij} - \bar{y}_i)^2$$

SS_T $\quad SS_m \quad SS_{treatment} \quad SS_{error}$

j	SS	CS	NA	HSS
1	65	63	67	50
2	65	63	67	50
3	65	63	67	50
4	65	63	67	50

$\varepsilon_{ij} = 0$

NOTE: You do NOT have permission to share this file or any of its contents with anyone else, and/or upload it on internet or any of the platforms where it can be accessed by others.

Example

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

$\bar{\bar{y}} = 64$ $N = 24$

y_{ij}

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

$\bar{y}_1 = 61$, $\bar{y}_2 = 66$, $\bar{y}_3 = 68$, $\bar{y}_4 = 61$

$n_1 = 4$, $n_2 = 6$, $n_3 = 6$, $n_4 = 8$

$$SS_T = \sum\sum y_{ij}^2 = \underbrace{62^2 + 60^2 + \dots + 63^2 + \dots + 59^2}_{24 \text{ terms}}$$

$$= \underline{98644} \quad (dof = 24)$$

$$SS_{mean} = N\bar{\bar{y}}^2 = 24 \times 64^2 = \underline{98304} \quad (dof = 1)$$

$$SS_{treatment} = \sum^{k} n_i(\bar{y}_i - \bar{\bar{y}})$$

$$= 4(61-64)^2 + 6(66-64)^2 + 6(68-64)^2 + 8(61-68)^2$$

$$= \underline{228} \quad (dof = k-1 = 3)$$

$$SS_{error} = \sum\sum(y_{ij} - \bar{y}_i) = \underbrace{(62-61)^2 + \dots\dots + }_{24 \text{ terms}} =$$

OR by subtra

$$SS_{err} = S_T - S_m - S_{treat} = 112$$

Example

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

$$y_{ij}$$

j^{th} observation for i^{th} metal

Here, i = 1, 2, 3, 4

(4 levels of variable 'metal')

Example

j	Stainless Steel	Carbon Steel	Nickel Alloy	High Speed Steel
1	62	63	68	56
2	60	67	66	62
3	63	71	71	60
4	59	64	67	61
5	NaN	65	68	63
6	NaN	66	68	64
7	NaN	NaN	NaN	63
8	NaN	NaN	NaN	59

CEP2022_Notebook (2.2)

■ **TABLE 3.3**

The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	228	3	$\frac{228}{3} = 76.$	$\frac{MS_{tr}}{MS_{er}} = \frac{76}{5.6} =$ 13.57
Error (within treatments)	112	$N-k = 20$	$\frac{112}{20} = 5.6$	
Total	98644	24.	~	
Mean	98304	1.	~	

compare with

$F_{1-\alpha, 3, 20}$

$F_{0.95, 3, 20}$

Summary: Tests on Variances

■ **TABLE 2.8**

Tests on Variances of Normal Distributions

Hypothesis	Test Statistic	Fixed Significance Level Criteria for Rejection
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$		$\chi_0^2 > \chi^2_{\alpha/2,n-1}$ or $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$\chi_0^2 < \chi^2_{1-\alpha,n-1}$
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$		$\chi_0^2 > \chi^2_{\alpha,n-1}$
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$F_0 = \dfrac{S_1^2}{S_2^2}$	$F_0 > F_{\alpha/2,n_1-1,n_2-1}$ or $F_0 < F_{1-\alpha/2,n_1-1,n_2-1}$
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$	$F_0 = \dfrac{S_2^2}{S_1^2}$	$F_0 > F_{\alpha,n_2-1,n_1-1}$
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$F_0 = \dfrac{S_1^2}{S_2^2}$	$F_0 > \cancel{F_{\alpha,n_1-1,n_2-1}}$ $F_{1-\alpha,n_1-1,n_2-1}$

Example

The engineer is interested in determining if the RF power setting affects the etch rate, and she has run a completely randomized experiment with four levels of RF power and five replicates (see Table1).

We will use the analysis of variance to test, $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

against the alternative, H_1: Some means are different (OR at least one mean is different)

	RF Power (W)	Observed Etch Rate (Å/min) 1	2	3	4	5
1	160	575	542	530	539	570
2	180	565	593	590	579	610
3	200	600	651	610	637	629
4	220	725	700	715	685	710

$k=4, \quad N=20, n_i=5$

$SS_T = \sum\sum y_{ij}^2 = 7704511$

$SS_m = N\bar{\bar{y}}^2 = 20 \times 617.75^2 =$

$SS_{treatment} = \sum^{k} n_i(\bar{y}_i - \bar{\bar{y}})^2$

$= 5\times(551.2-617.75)^2 +$
$5\times(587.4-617.75)^2 +$
$5\times(625.4-617.75)^2 + 5\times(707-617.75)^2$

$\bar{y}_1 = 551.2$
$\bar{y}_2 = 587.4$
$\bar{y}_3 = 625.4$
$\bar{y}_4 = 707$

RF Power (W)	Observed Etch Rate (Å/min)				
	1	2	3	4	5
160	575	542	530	539	570
180	565	593	590	579	610
200	600	651	610	637	629
220	725	700	715	685	710

$\bar{\bar{y}} = 617.75$

$SS_{error} = (575-551.2)^2 + (542-551.2)^2 + \cdots$
$+ (565-587.4)^2 + (593-587.4)^2 + \cdots$

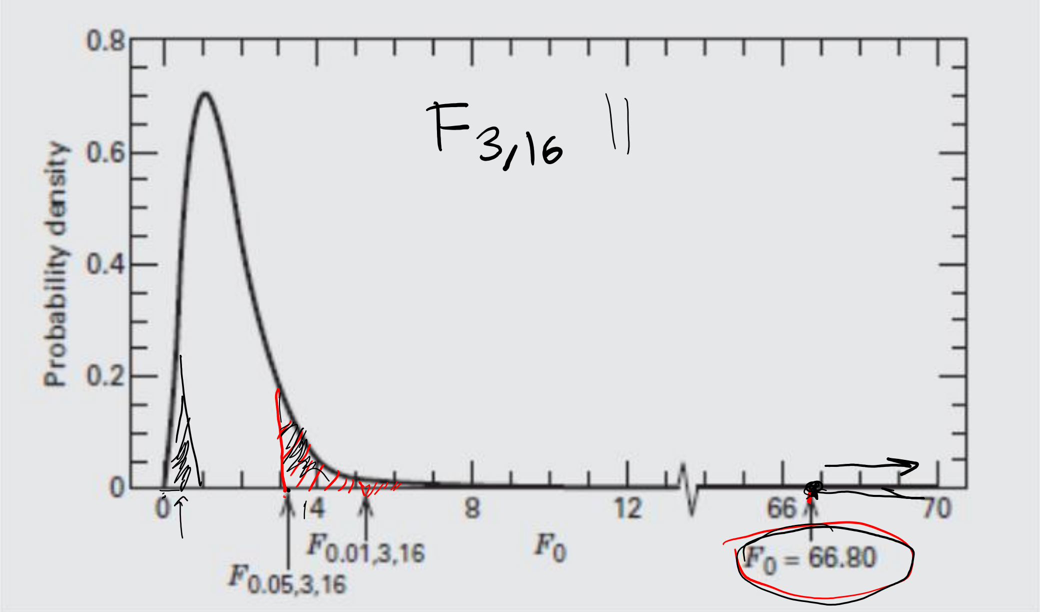
$= SS_T - SS_m - SS_{treatment}$

	SS	DOF	MS (SS/DOF)	F_0
SS_T	7704511	20	✓	
SS_m	7632301.25	1	✓	
SS_{treat}	66870.55	$k-1 = 3$	22290.18	$\frac{MS_{treat}}{MS_{error}} = 66.80$
SS_{error}	5339.20	16	333.08	

RF Power (W)	Observed Etch Rate (Å/min) 1	2	3	4	5
160	575	542	530	539	570
180	565	593	590	579	610
200	600	651	610	637	629
220	725	700	715	685	710

6

$66.80 > 3.16$



p value = ?

ANOVA: Residuals

- Residuals are the difference between what is ACTUALLY observed (Experiment) vs. what is PREDICTED from a model that is used to adequately describe the data

$$\epsilon_{ij} = y_{ij} - \widetilde{y_{ij}}$$

- In One-way ANOVA, what is the model?

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

μ = grand mean

τ_j = treatment mean

ε_{ij} = error

- What is the prediction?

$$\widetilde{y_{ij}} = \mu + \tau_i$$ **"Effects Model"**

- **Remember,** we had assumed that the residuals (or errors) are random and **normally distributed.**

So is that assumption valid IF we use the particular model? -> Model Adequacy Check!

$$\varepsilon_{ij} = y_{ij} - \bar{y}_i$$

$$\varepsilon_{ij} = y_{ij} - \underbrace{\overset{\bar{y}}{\mu} - \overset{(\bar{y}_i - \bar{y})}{\tau_i}}_{} = \bar{y}_i$$

ANOVA: Model Adequacy Checking

e_1 e_2 e_3

$\frac{i-0.5}{20}$

Normality Assumption can be checked using several methods

- A dot diagram
- Histogram of residuals
- Normal probability plot

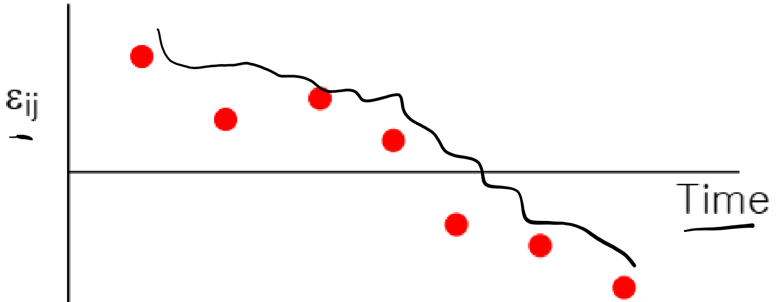
Etch Rate Data and Residuals from Example 3.1[a]

Power (w)	Observations (j) 1	2	3	4	5	$\hat{y}_{ij} = \bar{y}_{i\cdot}$	
160	$\boxed{23.8}$ 575 (13)	$\boxed{-9.2}$ 542 (14)	$\boxed{-21.2}$ 530 (8)	$\boxed{-12.2}$ 539 (5)	$\boxed{18.8}$ 570 (4)	551.2	$\bar{y}_1$
180	$\boxed{-22.4}$ 565 (18)	$\boxed{5.6}$ 593 (9)	$\boxed{2.6}$ 590 (6)	$\boxed{-8.4}$ 579 (16)	$\boxed{22.6}$ 610 (17)	587.4	$\bar{y}_2$
200	$\boxed{-25.4}$ 600 (7)	$\boxed{25.6}$ 651 (19)	$\boxed{-15.4}$ 610 (10)	$\boxed{11.6}$ 637 (20)	$\boxed{3.6}$ 629 (1)	625.4	$\bar{y}_3$
220	$\boxed{18.0}$ 725 (2)	$\boxed{-7.0}$ 700 (3)	$\boxed{8.0}$ 715 (15)	$\boxed{-22.0}$ 685 (11)	$\boxed{3.0}$ 710 (12)	707.0	$\bar{y}_4$

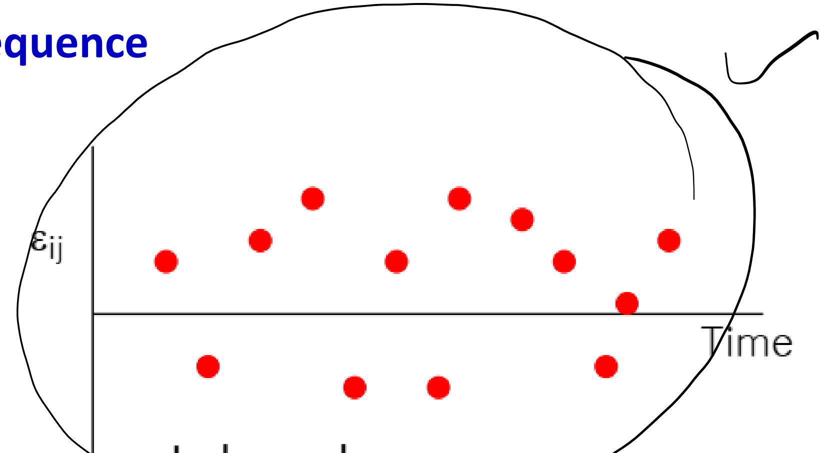
[a]The residuals are shown in the box in each cell. The numbers in parentheses indicate the order in which each experimental run was made.

Model Adequacy Checking

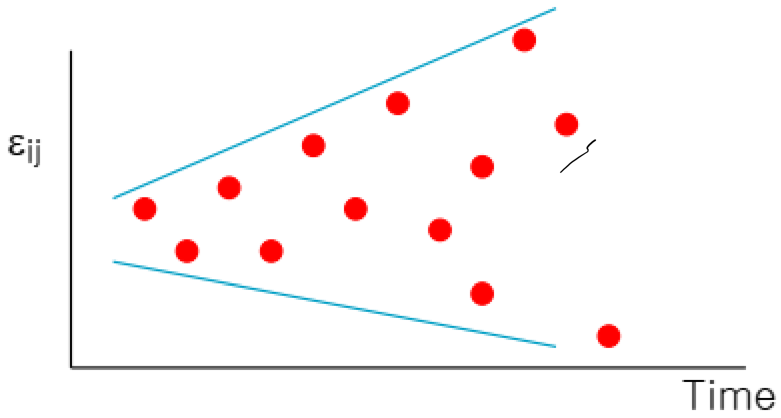
Dot Diagram of Residuals in Time Sequence



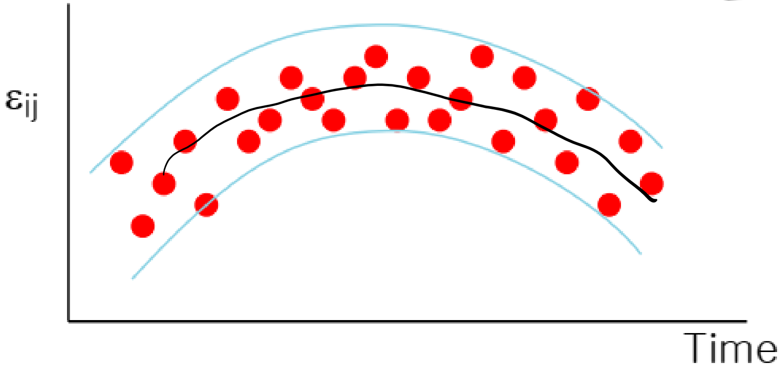
positive correlation



Independence



The variance is not constant, but increases with time.

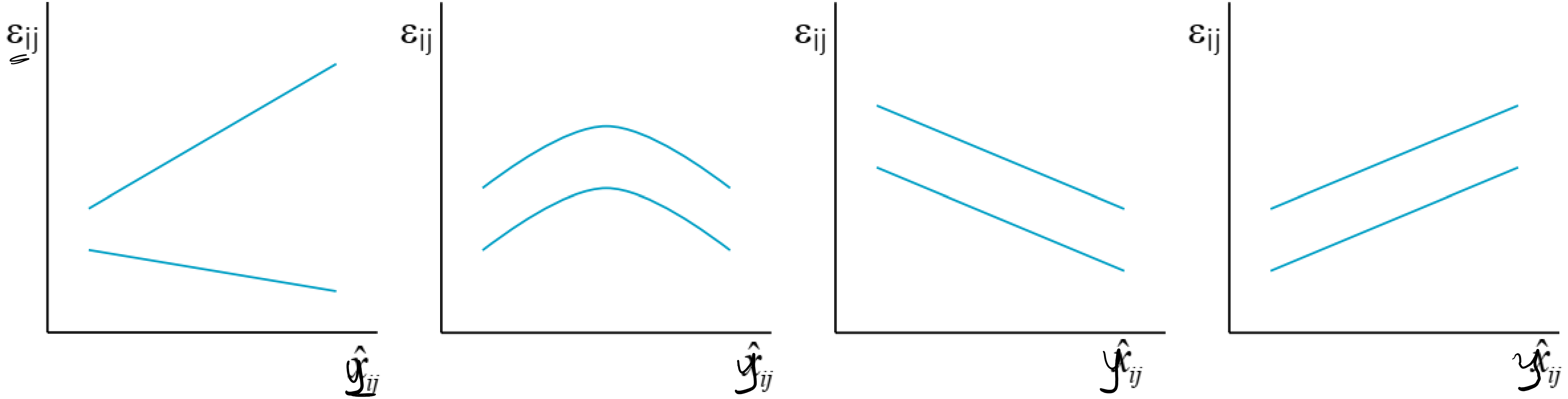


Model is not adequate, a quadratic term (may be interaction term) is needed in the model.

If the model is adequate and the assumptions are satisfied, the errors or residuals, ϵ_{ij} , should have NO structure.

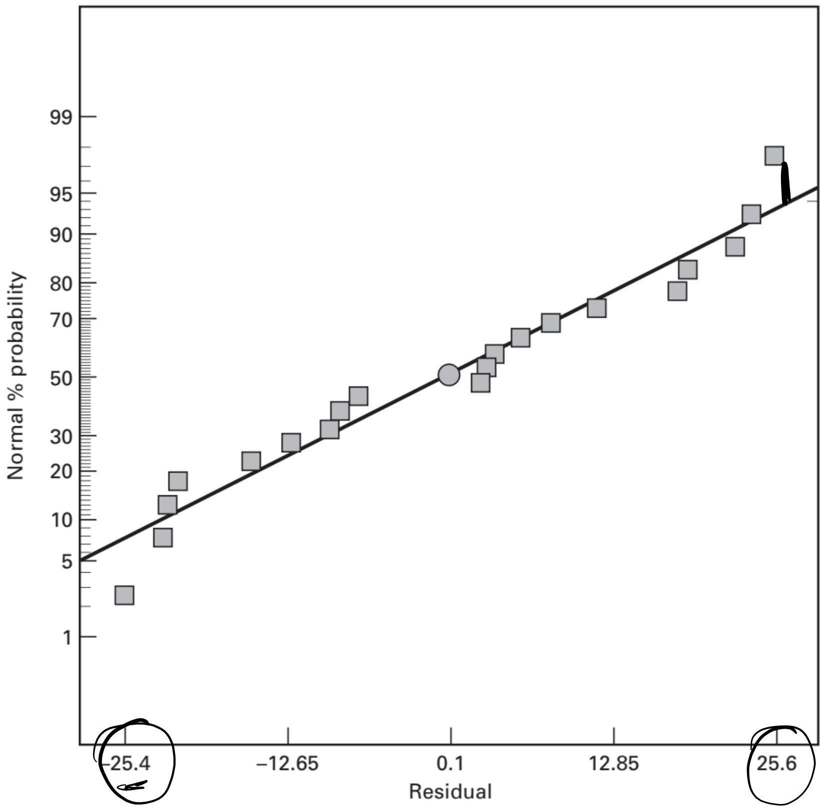
Model Adequacy Checking

Dot Diagram of Residuals (Errors) vs Model Predictions



If the model is adequate and the assumptions are satisfied, the errors or residuals, ϵ_{ij}, should be INDEPENDENT of observations

ANOVA: Model Adequacy Checking



Etch Rate Data and Residuals from Example 3.1[a]

Power (w)	1	2	3	4	5	$\hat{y}_{ij} = \bar{y}_{i.}$
	Observations (j)					
	23.8	−9.2	−21.2	−12.2	18.8	
160	575 (13)	542 (14)	530 (8)	539 (5)	570 (4)	551.2
	−22.4	5.6	2.6	−8.4	22.6	
180	565 (18)	593 (9)	590 (6)	579 (16)	610 (17)	587.4
	−25.4	25.6	−15.4	11.6	3.6	
200	600 (7)	651 (19)	610 (10)	637 (20)	629 (1)	625.4
	18.0	−7.0	8.0	−22.0	3.0	
220	725 (2)	700 (3)	715 (15)	685 (11)	710 (12)	707.0

[a]The residuals are shown in the box in each cell. The numbers in parentheses indicate the order in which each experimental run was made.

A rough check for outliers may be made by examining the **standardized residuals**

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}} \qquad \textbf{(3.18)}$$

If the errors ϵ_{ij} are $N(0, \sigma^2)$, the standardized residuals should be approximately normal with mean zero and unit variance. Thus, about 68 percent of the standardized residuals should fall within the limits ± 1, about 95 percent of them should fall within ± 2, and virtually all of them should fall within ± 3. A residual bigger than 3 or 4 standard deviations from zero is a potential outlier.

For the tensile strength data of Example 3.1, the normal probability plot gives no indication of outliers. Furthermore, the largest standardized residual is

$$d_1 = \frac{e_1}{\sqrt{MS_E}} = \frac{25.6}{\sqrt{333.70}} = \frac{25.6}{18.27} = 1.40$$

which should cause no concern.