**Residual Analysis: Theory**

- Theory: define the residual for the \( i^{th} \) observation \((x_i, y_i)\) as
  \[
  \hat{\varepsilon}_i = r_i = y_i - \hat{y}_i, \quad \hat{y}_i = x_i^T \hat{\beta},
  \]
  the \( i^{th} \) row of model matrix \( X \).
  \( \hat{\varepsilon}_i \) contains information given by the model; \( r_i \) is the "difference" between \( y_i \) (observed) and \( \hat{y}_i \) (fitted) and contains information on possible model inadequacy.

- Vector of residuals \( r = (r_1, \ldots, r_N)^T = y - X \hat{\beta} = (I - H)Y \)

- Under the model assumption \( E(y) = X \beta \), it can be shown that
  
  (a) \( E(r) = 0 \), \quad \therefore \text{\( X \beta \) is correct model}
  
  (b) \( r \) and \( \hat{y} \) are independent.
  
  (c) variances of \( r_i \) are nearly constant for "nearly balanced" designs.

Note variance of \( r_i \):

- \( h_{ii} \) (leverage) = \( 1 - h_{ii} \)
- Mahalanobis dist. btwn design pts & \( \hat{\varepsilon} \)

\[
Y = X\beta + \varepsilon = \hat{Y} + \hat{\varepsilon}
\]

- Under fitted model: \( Y = X_i \beta_i + \varepsilon_i \) (If \( h_{ii} = X_i(x_i x_i') \))

**Residual Plots**

- Plot \( r_i \) vs. \( \hat{y}_i \) (see Figure 1): It should appear as a parallel band around 0. Otherwise, it would suggest model violation. If spread of \( r_i \) increases as \( \hat{y}_i \) increases, error variance of \( y \) increases with mean of \( y \). May need a transformation of \( y \). (Will be explained in future lecture.)

- Plot \( r_i \) from replicates per treatment (see Figure 2): to see if error variance depends on treatment.

- Plot \( r_i \) vs. \( x_i \): If not a parallel band around 0, relationship between \( y_i \) and \( x_i \) not fully captured, revise the \( X\beta \) part of the model.

- Plot \( r_i \) vs. time sequence: to see if there is a time trend or autocorrelation over time.
Plot of $r_i$ vs. $\hat{y}_i$

Figure 1: $r_i$ vs. $\hat{y}_i$, Pulp Experiment

Plot of $r_i$ vs. treatment

Figure 2: $r_i$ vs. treatment, Pulp Experiment
Y_1, …, Y_n iid \sim cdfs F \leftrightsquigarrow Box-(Whisker) Plot

- A powerful graphical display (due to Tukey) to capture the location, dispersion, skewness and extremity of a distribution. See Figure 3.

- Q_1 = \text{lower quartile (25\% quantile)}, Q_3 = \text{upper quartile (75\% quantile)}, Q_2 = \text{median (50\% quantile, estimate of location parameter)} is the white line in the box. Q_1 and Q_3 are boundaries of the black box.

- IQR = \text{interquartile range (length of box)} = Q_3 - Q_1: \text{measure of dispersion}.

- Minimum and maximum of observed values within

$$[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$$

are denoted by two whiskers. Any values outside the whiskers are regarded as extreme values and displayed (possible outliers).

- If Q_1 and Q_3 are not symmetric around the median, it indicates skewness.

- Side-by-side box plots (LNp. 3-2~3) are useful to compare the difference between the distributions of several groups of data.

---

**Box-(Whisker) Plot**

![Box-Whisker Plot](image)

**Figure 3: Box-Whisker Plot**
Normal Probability Plot $\rightarrow$ Q-Q plot ($LM, LN_p 7.15 \sim 16$)

Original purpose: To test if a distribution is normal, e.g., if the residuals follow a normal distribution (see Figure 5).

- Can be used to identify outliers.
- Used to identify significant effects $\beta_i$'s $\rightarrow$ replace $t$-tests

More powerful use in factorial experiments (discussed in Units 5 and 6).

- Let $r_{(1)} \leq \ldots \leq r_{(N)}$ be the ordered residuals. The cumulative probability for $r_{(i)}$ is $p_i = (i - 0.5)/N$. Thus the plot of $p_i$ vs. $r_{(i)}$ should be S-shaped as in Figure 4(a) if the errors are normal. By transforming the scale of the horizontal axis, the S-shaped curve is straightened to be a line (see Figure 4(b)).

- Normal probability plot of residuals:

$$
\Phi^{-1}\left(\frac{i - 0.5}{N}\right), \quad r_{(i)}, \quad i = 1, \ldots, N, \quad \Phi = \text{normal cdf.}
$$

If the errors are normal, it should plot roughly as a straight line. See Figure 5.

Regular and Normal Probability Plots of Normal

Figure 4: Normal Plot of $r_i$, Pulp Experiment

$Z_i, \ldots, Z_N \text{ iid } \sim N(\mu, \sigma^2)$

$Z_i = \varepsilon_i W_i + \mu, W_i = (Z_i - \mu)/\sigma$ iid $\sim N(0, 1)$

Normal probability plot of $W_i$'s $\Rightarrow W_{(i)}$ vs. $\Phi^{-1}: y = \chi$

Normal probability plot of $Z_i$'s $\Rightarrow Z_{(i)}$ vs. $\Phi^{-1}: y = \sigma x + \mu$
Normal Probability Plot: Pulp Experiment

![Graph](image)

**Figure 5:** Normal Plot of $r_i$, Pulp Experiment

- **Reading:** textbook, 2.6

---

**Pulp Experiment Revisited**

- Compare the 2 scenarios:
  - (S1) plant has only 4 operators (or only interested in these 4 operators)
  - (S2) 4 operators randomly sampled from a large population of operators

- $\tau_i$'s: parameters (unknown fixed values)
  - interest: difference btwn the 4 specific $\tau_i$'s
- $\bar{\tau}_i$'s: random variables
  - interest: difference btwn all operators in this population

- In the pulp experiment the effects $\tau_i$ are called *fixed effects* because the interest was in comparing the four *specific* operators in the study. If these four operators were chosen randomly from the population of operators in the plant, the interest would usually be in the variation among all operators in the population. Because the observed data are from operators randomly selected from the population, the variation among operators in the population is referred to as *random effects*. 

---

jointly made by Jeff Wu (GT, USA) and S.-W. Cheng (NTHU, Taiwan)
One-way random effects model (REM): \( \text{FEM in } L\bar{N}_p = 3-4 \& 3-8 \)

\[ y_{ij} = \mu + \tau_i + \varepsilon_{ij} \]

where \( \varepsilon_{ij} \): independent error terms with \( N(0, \sigma^2) \),

\( \tau_i \): independent \( N(0, \sigma^2_\tau) \),

and \( \tau_i \) and \( \varepsilon_{ij} \) are independent (Why? Give an example.).

\( \sigma^2 \) and \( \sigma^2_\tau \) are the two variance components of the model.

The variance among operators in the population is measured by \( \sigma^2_\tau \).

\[ \text{REM: } E(Y) = \Omega \omega \]

\( \mu \) is independent \( \text{REM: } \mathcal{N}(\Omega \omega, \sigma^2\Omega) \)

\( \text{FEM: } y_{ij} \sim \mathcal{N}(\tau_i + \varepsilon_{ij}, \sigma^2) \)

\[ \text{FEM, } E(Y) = \chi^2 \]

\[ \text{FEM, } \text{indep. } \mathcal{N}(\tau_i + \varepsilon_{ij}, \sigma^2) \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]

\[ \text{REM: } \sigma^2_\tau \]

\[ \text{REM: } \sigma^2 \]
ANOVA Tables ($n_i = n$)

- We can apply the same ANOVA and $F$-test in the fixed effects case for analyzing data.

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>$k - 1$</td>
<td>$SSTr$</td>
<td>$MSTr = \frac{SSTr}{k - 1}$</td>
</tr>
<tr>
<td>residual</td>
<td>$N - k$</td>
<td>$SSE$</td>
<td>$MSE = \frac{SSE}{N - k}$</td>
</tr>
<tr>
<td>total</td>
<td>$N - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pulp Experiment

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>E(EMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>3</td>
<td>1.34</td>
<td>0.447</td>
<td>$\sigma^2 + \frac{\sigma^2}{k}$</td>
</tr>
<tr>
<td>residual</td>
<td>16</td>
<td>1.70</td>
<td>0.106</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>total</td>
<td>19</td>
<td>3.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- However, we need to compute the expected mean squares under the alternative of $\sigma^2 > 0$,
- (i) for sample size determination, and
- (ii) to estimate the variance components. ($\sigma^2 \& \sigma^2$)

### Expected Mean Squares for Treatments

- Equation (1) holds independent of $\sigma^2$,
  - $E(\text{MSE}) = E\left(\frac{SSE}{N - k}\right) = \sigma^2$.  

- Under the alternative: $\sigma^2 > 0$, and for $n_i = n$,
  - $E(\text{MSTR}) = E\left(\frac{SSTr}{k - 1}\right) = \sigma^2 + \frac{n\sigma^2}{k}$

- For unequal $n_i$’s, $n$ in (2) is replaced by
  - $n' = \frac{1}{k - 1} \left[ \sum_{i=1}^{k} n_i - \frac{\sum_{i=1}^{k} n_i^2}{k} \right]$.
  
- $\text{SSE}/(N - k)$ : an unbiased estimator of $\sigma^2$
  - $\text{SSE}$ only contains information of error var. component $\sigma^2$
  - $E\left(\frac{SSTr}{k - 1}\right) = \sigma^2$

- $\text{SSTr}$ contains information about factor var. component $\sigma^2$
  - $\text{SSE}$ only contains error var. component $\sigma^2$
  - $E(\text{SSTR})$ of FEM
  - (cf. $E(\text{SSTR})$ of FEM in LNp.3-6)
**Variance components: estimation of $\sigma^2$ and $\sigma^2_i$**

- From equations (1) and (2) in LNP. 3-35, we obtain the following unbiased estimates of the variance components:

  $$\hat{\sigma}^2 = \frac{MSE}{n} \quad \text{and} \quad \hat{\sigma}^2_i = \frac{MSTr - MSE}{n}$$

  Note that $\hat{\sigma}^2_i \geq 0$ if and only if $MSTr \geq MSE$, which is equivalent to $F \geq 1$. Therefore a negative variance estimate $\hat{\sigma}^2_i$ occurs only if the value of the $F$-statistic is less than 1. Obviously the null hypothesis $H_0$ is not rejected when $F \leq 1$. Since variance cannot be negative, a negative variance estimate is replaced by 0. This does not mean that $\sigma^2_i$ is zero. It simply means that there is not enough information in the data to get a good estimate of $\sigma^2_i$.

- For the pulp experiment, $n = 5$, $\hat{\sigma}^2 = 0.106$, $\hat{\sigma}^2_i = (0.447 - 0.106)/5 = 0.068$, i.e., sheet-to-sheet variance (within same operator) is 0.106, which is about 50% higher than operator-to-operator variance 0.068.

  Implications on process improvement: try to reduce the two sources of variation, also considering costs.
Estimation of Overall Mean $\eta$  

- In REM, $\eta$, the population mean, is often of interest.  
  From $\bar{E}(y_{ij}) = \eta$, we use the estimate
  
  $\hat{\eta} = \bar{y}$.

- For balanced data, $\hat{\eta} = \bar{y}$.

- $\text{Var}(\hat{\eta}) = \text{Var}(\bar{y} + \bar{\varepsilon}) = \frac{\sigma^2}{k} + \frac{\sigma^2}{N}$, where $N = \sum_{i=1}^{k} n_i$.

- $\hat{\eta} = \bar{y}_c = \bar{y} + \bar{\varepsilon}_c \sim N(\eta, \sigma^2/k + \sigma^2/N)$.

- In FEM, $E(y_{ij}) = \mu_i = \eta_i$.

- Same as the $\hat{\eta}$ in FEM under sum coding, but in the case of FEM
  $\hat{\eta} = (\mu_1 + \cdots + \mu_k)/k$.

- $s.e.(\hat{\eta}) = \frac{\text{MST}_r}{nk}$.

- For $n_i = n$, $\text{Var}(\hat{\eta}) = \frac{\sigma^2}{k} + \frac{n\sigma^2}{nk} = \frac{1}{nk} (\sigma^2 + n\sigma^2_{\varepsilon})$.

- Using (2) in LNp.3-35, $\frac{\text{MST}_r}{nk}$ is an unbiased estimate of $\text{Var}(\hat{\eta})$.

- Confidence interval for $\eta$:
  
  $\bar{y}_c$ and $\text{MST}_r$ are independent.

- In the pulp experiment, $\hat{\eta} = 60.40$, $\text{MST}_r = 0.447$, and the 95% confidence interval for $\eta$ is
  
  $60.40 \pm 3.182 \sqrt{\frac{0.447}{5 \times 4}} = [59.92, 60.88]$.

- Compare REM and split-plot designs.

Reading: textbook, 2.5