Lecture

Checking Model Assumptions

Textbook: Chapter 6

STAT 8050 Design and Analysis of Experiments
February 20, 2020

Whitney Huang
Clemson University

Model Assumptions

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

Model Assumptions

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed

Model Assumptions

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed

Model Assumptions

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed
- Errors have constant variance

Model Assumptions

CLEMSON UNIVERSITY

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed
- Errors have constant variance

Model Assumptions

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed
- Errors have constant variance
- Errors are independent

Model Assumptions

Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed
- Errors have constant variance
- Errors are independent

Model Assumptions

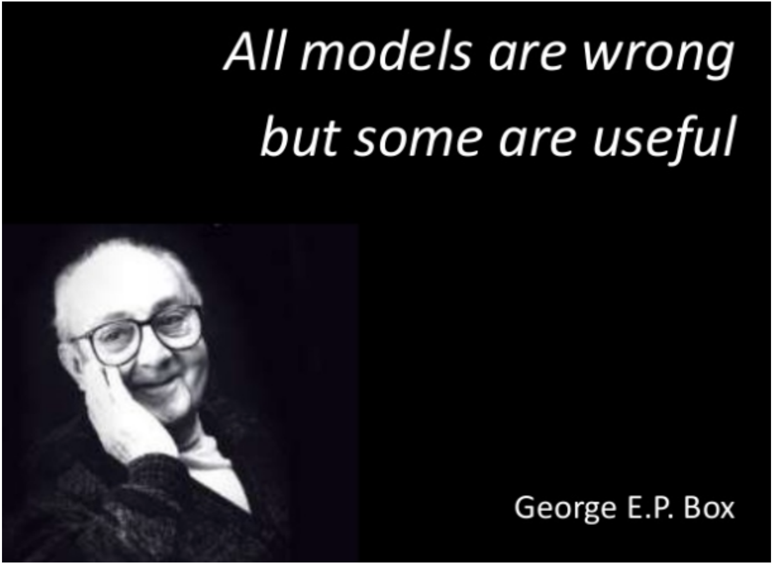
Model:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

We make the following assumptions:

- Errors normally distributed
- Errors have constant variance
- Errors are independent

$$\Rightarrow \epsilon_{ij} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2)$$


All models are wrong
but some are useful
George E.P. Box

What If Assumptions are Violated?

If the assumptions are not true, our statistical inferences might not be valid, for example,

- A confidence interval might not cover with the stated coverage rate
- A test with nominal type I error could actually have a larger or smaller type I error rate

What If Assumptions are Violated?

If the assumptions are not true, our statistical inferences might not be valid, for example,

- A confidence interval might not cover with the stated coverage rate
- A test with nominal type I error could actually have a larger or smaller type I error rate

We need good strategy for checking model assumptions, i.e., $\epsilon_{ij} \stackrel{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2)$.

Checking Model Assumptions

We need to check if these assumptions reasonably met

Model:

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

Checking Model Assumptions

We need to check if these assumptions reasonably met

Model:

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

Data:

$$\begin{array}{ccccc} y_{ij} & = & (\bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..})) & + & (y_{ij} - \bar{y}_{i.}) \\ y_{ij} & = & \hat{y}_{ij} & + & \hat{\epsilon}_{ij}\,(r_{ij}) \\ \text{observed} & = & \text{predicted} & + & \text{residual} \end{array}$$

Residuals are our "estimates" of unobservable errors $\epsilon'_{ij}s$

We will conduct model diagnostics using **residual** and **predicted** values.

Residuals

We will use residuals to assess the model assumptions.

- Raw residual:

$$r_{ij} = y_{ij} - \hat{y}_{ij}, \text{ where } \hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i = \bar{y}_{i\cdot}$$

Residuals

We will use residuals to assess the model assumptions.

- Raw residual:

$$r_{ij} = y_{ij} - \hat{y}_{ij}, \text{ where } \hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i = \bar{y}_{i\cdot}$$

- Standardized residual (internally Studentized residual) adjusts r_{ij} for its estimated standard deviation

$$s_{ij} = \frac{r_{ij}}{\sqrt{\mathrm{MS}_E(1 - \frac{1}{n_i})}}$$

Residuals

We will use residuals to assess the model assumptions.

- Raw residual:

$$r_{ij} = y_{ij} - \hat{y}_{ij}, \text{ where } \hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_i = \bar{y}_{i.}$$

- Standardized residual (internally Studentized residual) adjusts r_{ij} for its estimated standard deviation

$$s_{ij} = \frac{r_{ij}}{\sqrt{\text{MS}_E(1 - \frac{1}{n_i})}}$$

- Studentized residual (externally Studentized residual)

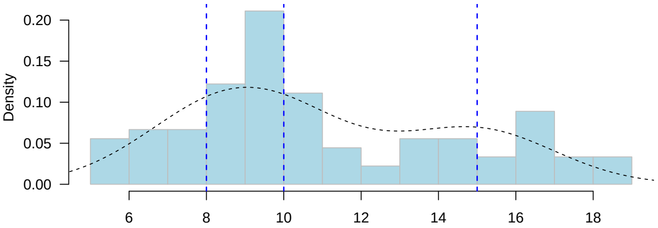
$$t_{ij} = s_{ij}\sqrt{\frac{N-g-1}{N-g-s_{ij}^2}}$$

$t_{ij} \sim t_{df=N-g-1}$ if the model is correct $\Rightarrow$ can be used to identify outliers

Assessing Normality

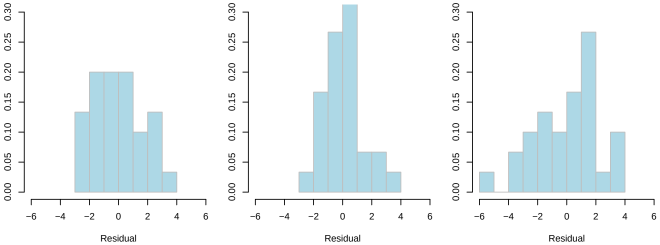
We DO NOT assume all $y'_{ij}s$ come from the same normal distribution, instead we assume $\epsilon'_{ij}s$ come from the same normal distribution $\Rightarrow$ Not informative to plot a histogram for all the data—-treatment effects lead to non-normality

Example: Suppose $g = 3$, $(\mu_1, \mu_2, \mu_3) = (8, 10, 15)$ and $\epsilon'_{ij}s \sim \mathrm{N}(0, 2^2)$



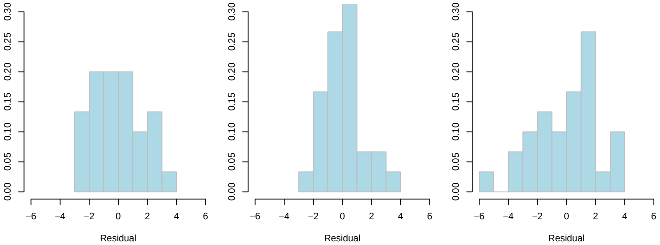
Assessing Normality Cont'd

- If sample sizes are large, histograms of residuals can be constructed from each treatment separately



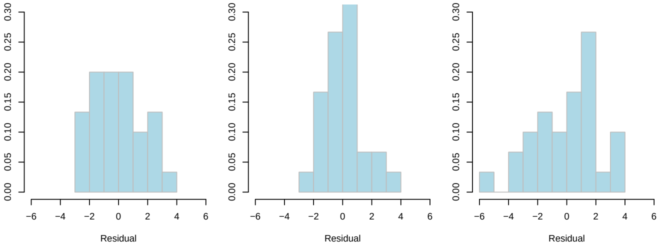
Assessing Normality Cont'd

- If sample sizes are large, histograms of residuals can be constructed from each treatment separately



Assessing Normality Cont'd

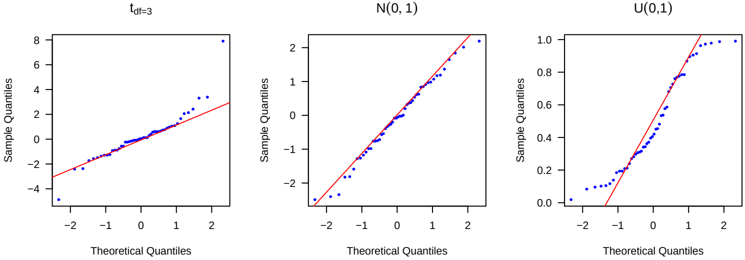
- If sample sizes are large, histograms of residuals can be constructed from each treatment separately



- Also, if sample sizes are large, normal probability plots or normal quantile plots can be generated for each treatment

Normal Probability Plot (NPP)

The NPP plots $r_{(k)}$ versus $\Phi^{-1}(\frac{k-3/8}{n+1/4})$, $k=1,\cdots,n$, where $r_{(k)}$ is the k^{th} ordered residual and $\Phi^{-1}(\frac{k-3/8}{n+1/4})$ is its corresponding (standard) normal score.



Remarks on Assessing Normality

- Assessing normality
  - Formal tests (e.g., Shapiro–Wilk test, Anderson–Darling test) are usually not useful:

    With small sample sizes, one will never be able to reject H_0, with large sample sizes, one will constantly detect little deviations that have no practical effect

  - Assess normal assumption graphically using normal probability plots or histograms
- Dealing with Non-normality
  - Use non-parametric procedure such as Kruskal–Wallis test (1952)
  - Transformation such as Box-Cox (1964)
- F-test is robust to non-normality

Assessing Equal Variance

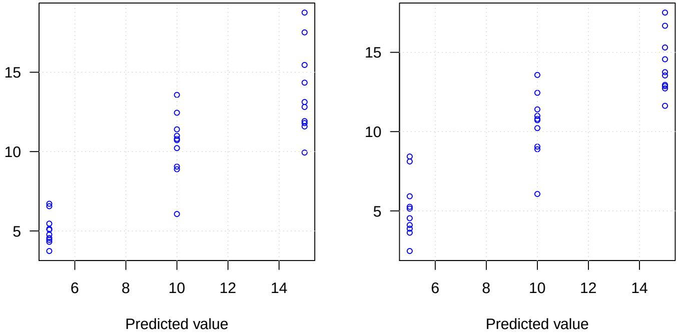
- We can test for equal variance, but some tests rely heavily on normality assumption:
  - Hartley's test
  - Bartlett's test
  - Cochran's C test
- F-test is reasonably robust to unequal variance if $n_i's$ are equal, or nearly so
- "*If you have to to test for equality of variances, your best bet is Levene's test.*" – Gary Oehlert

Levene's Test

1. Compute $r_{ij} = y_{ij} - \bar{y}_{i\cdot}$
2. Treat the $|r_{ij}|$ as data and use the ANOVA F-test to test H_0 that the groups have the same average value of $|r_{ij}|$
3. If $\frac{\text{MS}_{TRT}}{\text{MS}_E} > F_{g-1,N-g-1;\alpha} \Rightarrow$ reject H_0
4. Modified Levene's (Brown-Forsythe) test: use $d_{ij} = |y_{ij} - \tilde{y}_i|$, the absolute deviations from the group medians instead of $|r_{ij}|$

Fairly robust to non-normality and unequal sample size

Diagnostic Plot for Non-Constant Variance



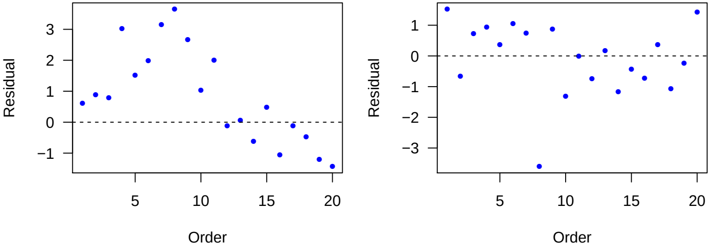
Use this residual versus predicted value (treatment) plot to assess equal variance assumption and search for possible outliers

Remarks on Assessing Constant Variance Assumption

- Checking constant variance assumption: Assess the assumption qualitatively, don't just rely no tests
- Dealing with unequal variance
  - Variance-stabilizing transformations
  - Account unequal variance in the model
- F-test is reasonably robust to unequal variance if we have (nearly) balanced designs

Assessing Dependence

Independence is often argued via randomization. However, plotting residuals versus run order or spatial location can give information on lack of independence.



Durbin–Watson statistic is a simple numerical method for checking serial dependence:

$$\text{DW} = \frac{\sum_{k=1}^{n-1}(r_k - r_{k+1})^2}{\sum_{k=1}^{n} r_k^2}$$

Example: Balloon Experiment (taken from Dean and Voss Exercise 3.12)

The experimenter (Meily Lin) had observed that some colors of birthday balloons seem to be harder to inflate than others. She ran this experiment to determine whether balloons of different colors are similar in terms of the time taken for inflation to a diameter of 7 inches. Four colors were selected from a single manufacturer. An assistant blew up the balloons and the experimenter recorded the times with a stop watch. The data, in the order collected, are given in Table 3.13, where the codes 1, 2, 3, 4 denote the colors pink, yellow, orange, blue, respectively.

Table 3.13 Times (in seconds) for the balloon experiment

Time order	1	2	3	4	5	6	7	8
Coded color	1	3	1	4	3	2	2	2
Inflation time	22.0	24.6	20.3	19.8	24.3	22.2	28.5	25.7
Time order	9	10	11	12	13	14	15	16
Coded color	3	1	2	4	4	4	3	1
Inflation time	20.2	19.6	28.8	24.0	17.1	19.3	24.2	15.8
Time order	17	18	19	20	21	22	23	24
Coded color	2	1	4	3	1	4	4	2
Inflation time	18.3	17.5	18.7	22.9	16.3	14.0	16.6	18.1
Time order	25	26	27	28	29	30	31	32
Coded color	2	4	2	3	3	1	1	3
Inflation time	18.9	16.0	20.1	22.5	16.0	19.3	15.9	20.3