Lecture 9

Completely Randomized Designs

Reading: Oehlert Chapter 3; Dean-Voss-Draguljić Chapter 3

DSA 8020 Statistical Methods II

March 6-10, 2023

Whitney Huang

Clemson University

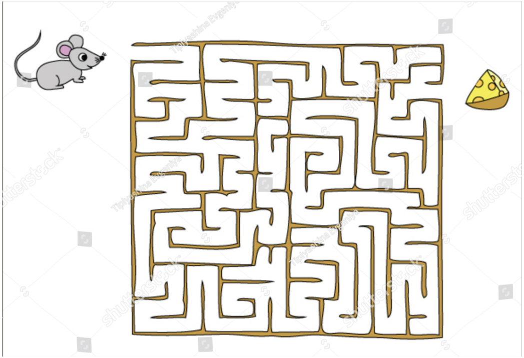
Agenda

1. Completely Randomized Designs

2. Checking Model Assumptions

Navigational Learning and Memory in Mice

An experiment was conducted to determine if experience has an effect on the time it takes for mice to run a maze. Four treatment groups, consisting of mice having been trained on the maze one, two, three and four times were run through the maze and their times recorded.



Source: https://www.shutterstock.com/image-vector/find-your-way-cheese-mouse-maze-232569073

Completely Randomized Designs

A completely randomized design (CRD) has

- g different treatment groups
- g known treatment group sizes $n_1, n_2, \cdots, n_g$ with $\sum_{i=1}^{g} n_i = N$
- Completely random assignment of treatments to the experimental units

This is the basic experimental design; everything else is a modification

- Easiest to analyze
- Most resilient when things go wrong
- Often sufficient

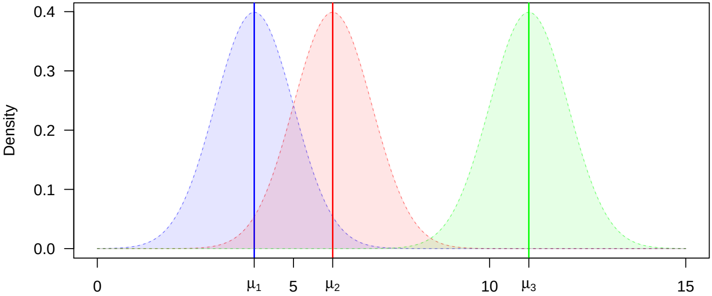
Inference

- Any evidence means (i.e., $\{\mu_1, \mu_2, \cdots, \mu_g\}$) are not all the same? ⇒ ANOVA
- Which ones differ? ⇒ Multiple comparisons
- Estimates/confidence intervals of means and differences

Statistical Model: Means Model

Let Y_{ij} be the random variable that represents the response for the j^{th} experimental unit to treatment i. Let $\mu_i = \mathrm{E}(Y_{ij})$ be the mean response for the i^{th} treatment. We have

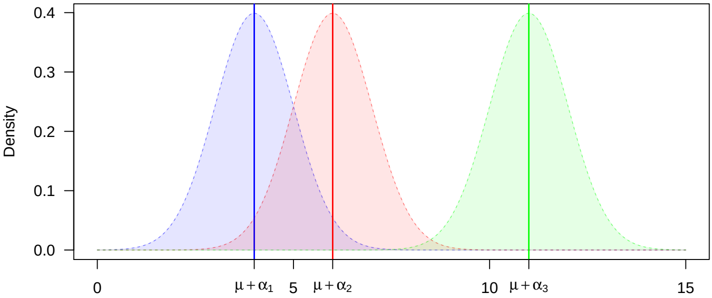
$$\boxed{Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i, \quad \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)}$$



Effects Model

Alternatively, we could let $\mu_i = \mu + \alpha_i$, which leads to

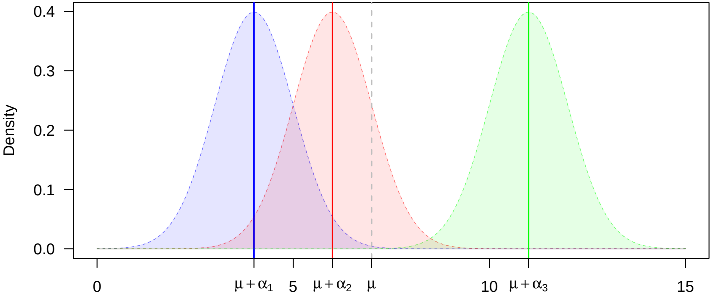
$$\boxed{Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i, \quad \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)}$$



Overparameterized. Need to add a constraint so that the parameters are estimable.

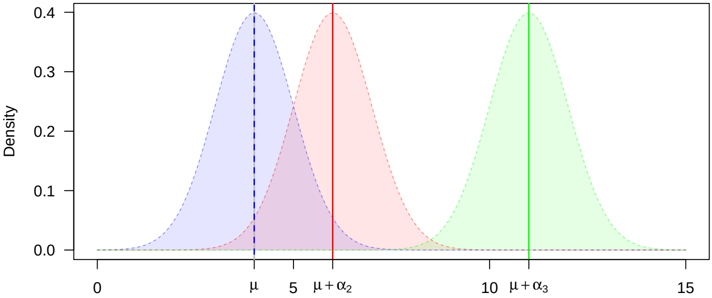
Effects Model Cont'd

Suppose we let $\sum_{i=1}^{g} n_i\alpha_i = 0$



Effects Model Cont'd

Suppose we let $\alpha_1 = 0$



Data Layout & the Dot Notation

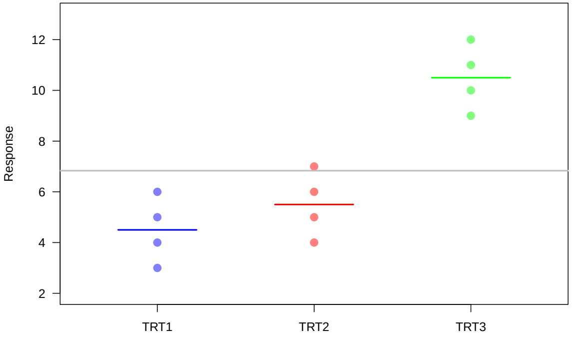
y_{ij} is the observed response for the j^{th} experimental unit to treatment i.

Treatment	Observations				Totals	Averages
1	y_{11}	y_{12}	$\cdots$	y_{1n_1}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	$\cdots$	y_{2n_2}	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
g	y_{g1}	y_{g2}	$\cdots$	y_{gn_g}	$y_{g\cdot}$	$\bar{y}_{g\cdot}$
					$y_{\cdot\cdot}$	$\bar{y}_{\cdot\cdot}$

ANOVA

Decomposition of y_{ij}: $y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$

$$\Rightarrow \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{..})^2}_{\text{SS}_T} = \underbrace{\sum_{i=1}^{g} n_i(\bar{y}_{i.} - \bar{y}_{..})^2}_{\text{SS}_{TRT}} + \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2}_{\text{SS}_E}$$



ANOVA Table

Source	df	SS	MS	EMS
Treatment	$g-1$	SS_{TRT}	$\text{MS}_{TRT} = \frac{\text{SS}_{TRT}}{g-1}$	$\sigma^2 + \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{g-1}$
Error	$N-g$	SS_E	$\text{MS}_E = \frac{\text{SS}_E}{N-g}$	σ^2
Total	$N-1$	SS_T		

$$\text{SS}_T = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{g} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$\text{SS}_{TRT} = \sum_{i=1}^{g} n_i (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{g} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

$$\text{SS}_E = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{g} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{g} \frac{y_{i.}^2}{n_i} = \text{SS}_T - \text{SS}_{TRT}$$

F-Test

Testing for treatment effects

$$H_0 : \alpha_i = 0 \quad \text{for all } i$$
$$H_a : \alpha_i \neq 0 \quad \text{for some } i$$

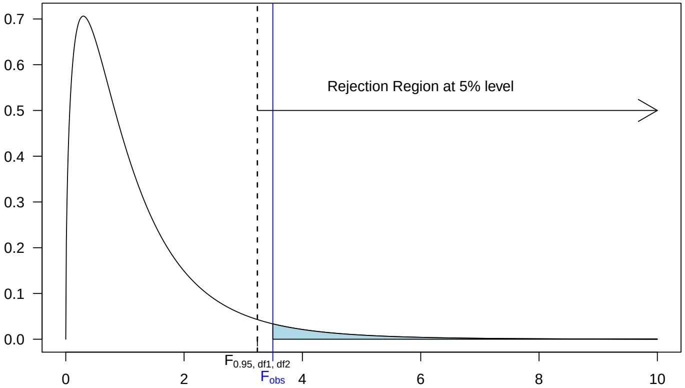
Test statistics: $F = \frac{\text{MS}_{TRT}}{\text{MS}_E}$. Under H_0, the test statistic follows an F-distribution with $g-1$ and $N-g$ degrees of freedom
Reject H_0 if

$$F_{obs} > F_{g-1,N-g;\alpha}$$

for an α-level test, $F_{g-1,N-g;\alpha}$ is the $100 \times (1-\alpha)\%$ percentile of a central F-distribution with $g-1$ and $N-g$ degrees of freedom.

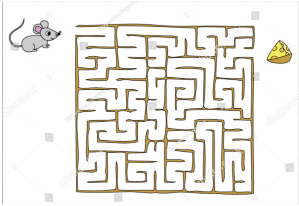
The P-value of the F-test is the probability of obtaining F at least as extreme as F_{obs}, that is, $\text{P}(F > F_{obs}) \Rightarrow$ reject H_0 if P-value $< \alpha$.

F Distribution and the F-Test



Mice Example Revisited

An experiment was conducted to determine if experience has an effect on the time it takes for mice to run a maze. Four treatment groups, consisting of mice having been trained on the maze one, two, three and four times were run through the maze and their times recorded.



Source: https://www.shutterstock.com/image-vector/find-your-way-cheese-mouse-maze-232569073

Training runs	1	2	3	4
n_i	5	5	5	5
$\bar{y}_{i\cdot}$	9.14	7.24	6.76	5.18
s_i^2	0.308	0.418	0.313	0.262

Example Cont'd

Training runs	1	2	3	4
n_i	5	5	5	5
$\bar{y}_{i\cdot}$	9.14	7.24	6.76	5.18
s_i^2	0.308	0.418	0.313	0.262

- Write down the model.
- Fill out the ANOVA table and test whether the time to run the maze is affected by training. Use a significant level of .05.

Model Assumptions

All models are wrong, but some are useful–G.E.P Box

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

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

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

Checking Model Assumptions

We need to check if these assumptions reasonably met

Model:

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

Data:

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

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:

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

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

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

- Studentized residual (externally Studentized residual)

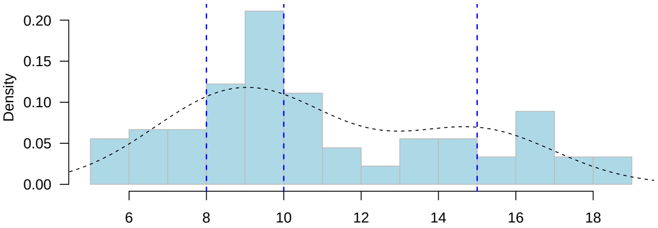
$$t_{ij} = s_{ij}\sqrt{\frac{N-g-1}{N-g-s_{ij}^2}}$$

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

Assessing Normality

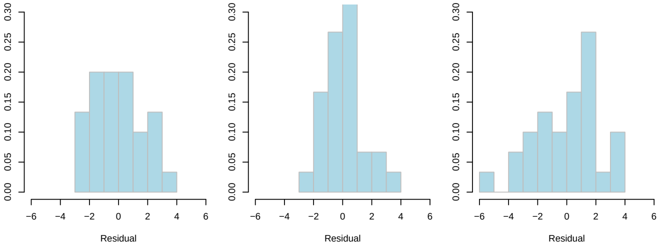
We DO NOT assume all $y'_{ij}s$ come from the same normal distribution, instead we assume $\epsilon'_{ij}s$ come from the same normal distribution ⇒ Not informative to plot a histogram for all the data—-treatment effects lead to non-normality

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

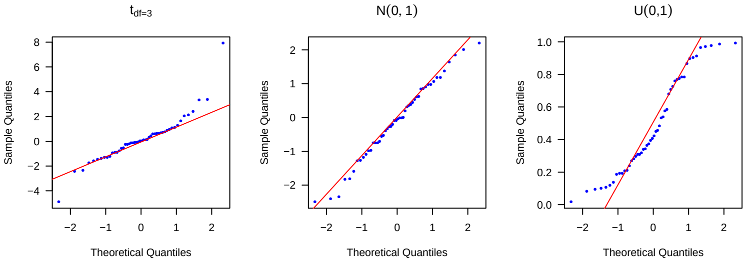
- If sample sizes are large, histograms of residuals can be constructed from each treatment separately



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

Normal Quantile Plots

Plots $r_{(k)}$ versus $\Phi^{-1}(\frac{k}{n+1})$, $k = 1, \cdots, n$, where $r_{(k)}$ is the k^{th} ordered residual and $\Phi^{-1}(\frac{k}{n+1})$ 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 QQ-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

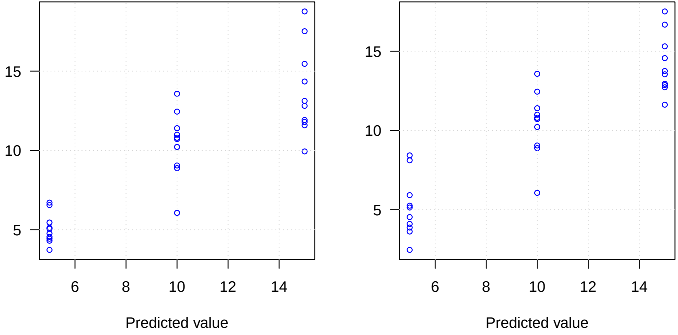
- We can test for equal variance, but some tests rely heavily on normality assumption:
  - Hartley's test
  - Bartlett's test
  - Cochran's C test
- F-test is reasonably robust to unequal variance if $n_i's$ are equal balanced design, or nearly so
- "*If you have to test for equality of variances, your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

Levene's Test

1. Compute $r_{ij} = y_{ij} - \bar{y}_{i.}$
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



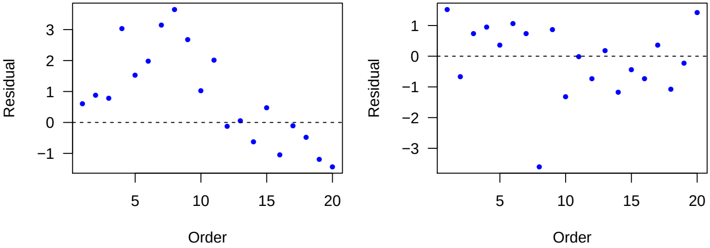
Use this residual versus predicted value (treatment) plot to assess equal variance assumption and search for possible outliers

Remarks on Assessing Constant Variance Assumption

- Checking constant variance assumption: Assess the assumption qualitatively, don't just rely on 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