COMPLETELY RANDOM DESIGN (CRD)

Description of the Design

-Simplest design to use.

-Design can be used when experimental units are essentially homogeneous.

-Because of the homogeneity requirement, it may be difficult to use this design for field experiments.

-The CRD is best suited for experiments with a small number of treatments.

Randomization Procedure

-Treatments are assigned to experimental units completely at random.

-Every experimental unit has the same probability of receiving any treatment.

-Randomization is performed using a random number table, computer, program, etc.

Example of Randomization

-Given you have 4 treatments (A, B, C, and D) and 5 replicates, how many experimental units would you have?

1 D	2 D	3 B	4 C	5 D	6 C	7 A	8 A	9 B	10 D
11 C	12 B	13 A	14 B	15 C	16 B	17 C	18 D	19 A	20 A

-Note that there is no “blocking” of experimental units into replicates.

-Every experimental unit has the same probability of receiving any treatment.

Advantages of a CRD

1. Very flexible design (i.e. number of treatments and replicates is only limited by the available number of experimental units).

2. Statistical analysis is simple compared to other designs.

3. Loss of information due to missing data is small compared to other designs due to the larger number of degrees of freedom for the error source of variation.

Disadvantages

1. If experimental units are not homogeneous and you fail to minimize this variation using blocking, there may be a loss of precision.

2. Usually the least efficient design unless experimental units are homogeneous.

3. Not suited for a large number of treatments.

Fixed vs. Random Effects

-The choice of labeling a factor as a fixed or random effect will affect how you will make the F-test.

-This will become more important later in the course when we discuss interactions.

Fixed Effect

-All treatments of interest are included in your experiment.

-You cannot make inferences to a larger experiment.

Example 1: An experiment is conducted at Fargo and Grand Forks, ND. If location is considered a fixed effect, you cannot make inferences toward a larger area (e.g. the central Red River Valley).

Example 2: An experiment is conducted using four rates (e.g. ½ X, X, 1.5 X, 2 X) of a herbicide to determine its efficacy to control weeds. If rate is considered a fixed effect, you cannot make inferences about what may have occurred at any rates not used in the experiment (e.g. ¼ x, 1.25 X, etc.).

Random Effect

-Treatments are a sample of the population to which you can make inferences.

-You can make inferences toward a larger population using the information from the analyses.

Example 1: An experiment is conducted at Fargo and Grand Forks, ND. If location is considered a random effect, you can make inferences toward a larger area (e.g. you could use the results to state what might be expected to occur in the central Red River Valley).

Example 2: An experiment is conducted using four rates (e.g. ½ X, X, 1.5 X, 2 X) of an herbicide to determine its efficacy to control weeds. If rate is considered a random effect, you can make inferences about what may have occurred at rates not used in the experiment (e.g. ¼ x, 1.25 X, etc.).

Analysis of the Fixed Effects Model

Notation

- Statistical notation can be confusing, but use of the Y-dot notation can help simplify things.

- The dot in the Y-dot notation implies summation across over the subscript it replaces.

- For example,

$$y_{i.} = \sum_{j=1}^{n} y_{ij} = \text{Treatment total, where n = number of observations in a treatment}$$

$$\bar{y}_{i.} = y_{i.}/n = \text{Treatment mean}$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} = \text{Experiment total, where a = number of treatments}$$

$$\bar{y}_{..} = y_{..}/N = \text{Experiment mean, where N = total number of observations in the experiment.}$$

Linear Additive Model for the CRD

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where: Y_{ij} is the j^{th} observation of the i^{th} treatment,
μ is the population mean,
τ_i is the treatment effect of the i^{th} treatment, and
ε_{ij} is the random error.

-Using this model we can estimate τ_i or ε_{ij} for any observation if we are given Y_{ij} and μ.

<u>Example</u>

	Treatment 1	Treatment 2	Treatment 3	
	4	9	8	
	5	10	11	
	6	11	8	
$Y_{i.}$	15	30	27	$Y_{..} = 72$
$\overline{Y}_{i.}$	5	10	9	$\overline{Y}_{..} = 8$
$\overline{Y}_{i.} - \overline{Y}_{..}$	-3	2	1	

-We can now write the linear model for each observation (Y_{ij}).

-Write in μ for each observation.

	Treatment 1	Treatment 2	Treatment 3	
	4 = 8	9 = 8	8 = 8	
	5 = 8	10 = 8	11 = 8	
	6 = 8	11 = 8	8 = 8	
$Y_{i.}$	15	30	27	$Y_{..} = 72$
$\overline{Y}_{i.}$	5	10	9	$\overline{Y}_{..} = 8$
$\overline{Y}_{i.} - \overline{Y}_{..}$	-3	2	1	

-Write in the respective τ_i for each observation where $\tau_i = \overline{Y}_{i.} - \overline{Y}_{..}$

	Treatment 1	Treatment 2	Treatment 3	
	4 = 8 – 3	9 = 8 + 2	8 = 8 + 1	
	5 = 8 – 3	10 = 8 + 2	11 = 8 + 1	
	6 = 8 – 3	11 = 8 + 2	8 = 8 + 1	
$Y_{i.}$	15	30	27	$Y_{..} = 72$
$\overline{Y}_{i.}$	5	10	9	$\overline{Y}_{..} = 8$
$\overline{Y}_{i.} - \overline{Y}_{..}$	-3	2	1	

-Write in the ε_{ij} for each observation.

	Treatment 1	Treatment 2	Treatment 3	
	4 = 8 – 3 - 1	9 = 8 + 2 - 1	8 = 8 + 1 - 1	
	5 = 8 – 3 + 0	10 = 8 + 2 + 0	11 = 8 + 1 + 2	
	6 = 8 – 3 + 1	11 = 8 + 2 + 1	8 = 8 + 1 - 1	
$Y_{i.}$	15	30	27	$Y_{..}$ = 72
$\bar{Y}_{i.}$	5	10	9	$\bar{Y}_{..}$ = 8
$\bar{Y}_{i.} - \bar{Y}_{..}$	-3	2	1	

-Note for each treatment $\sum \varepsilon_{ij} = 0$.

-If you are asked to solve for τ_3, what is the answer?

-If you are asked to solve for ε_{23}, what is the answer?

-Question: If you are given just the treatment totals ($Y_{i.}$'s), how would you fill in the values for each of the observations such that the Error SS = 0.

Answer: Remember that the Experimental Error is the failure of observations treated alike to be the same. Therefore, if all treatments have the same value in each replicate, the Experimental Error SS =0.

Example

Given the following information, fill in the values for all Y_{ij}'s such that the Experimental Error SS = 0.

	Treatment 1	Treatment 2	Treatment 3	
$Y_{i.}$	15	30	27	$Y_{..}$ = 72
$\bar{Y}_{i.}$	5	10	9	$\bar{Y}_{..}$ = 8

Answer

	Treatment 1	Treatment 2	Treatment 3	
	5	10	9	
	5	10	9	
	5	10	9	
$Y_{i.}$	15	30	27	$Y_{..} = 72$
$\overline{Y}_{i.}$	5	10	9	$\overline{Y}_{..} = 8$

- Note in the previous two examples that $\sum \tau_i = 0$. This is true for all situations.
- Given
  $H_0 : \mu_1 = \mu_2 = ... = \mu_t$
  $H_A : \mu_i \neq \mu_{i'}$ for at least one pair of treatments (i,i')
  $$\frac{\sum_{i=1}^{t} \mu_i}{t} = \mu$$ (i.e., the sum of the treatment means divided by the number of treatments equals the experiment mean).

- This definition implies that $\sum_{i=1}^{t} \tau_i = 0 = (\overline{Y}_{i.} - \overline{Y}_{..})$.

- The hypothesis written above can be rewritten in terms of the treatment effects τ_i as:

  $H_0 : \tau_1 = \tau_2 = ... = \tau_a = 0$
  $H_A : \tau_i \neq 0$ for at least i.
- Thus, when we are testing the null hypothesis that all treatments means are the same, we are testing at the same time the null hypothesis that all treatment effects, τ_i, are zero.

Partitioning the Total Sum of Squares

- Remember that:
  $\mu = \overline{Y}_{..}$
  $\tau_i = (\overline{Y}_{i.} - \overline{Y}_{..})$
  $\varepsilon_{ij} = (Y_{ij} - \overline{Y}_{i.})$

- Thus, $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$ can be rewritten as: $Y_{ij} = \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.})$.

- The Analysis of Variance is derived from the partitioning of the corrected Total Sum of Squares.

$$\text{Total Sum of Squares} = \sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \overline{Y}_{..})^2$$

and

$$\sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{a} \sum_{j=1}^{n} [(\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.})]^2$$

$$= r\sum_{i=1}^{t} (\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \overline{Y}_{i..})^2 + 2\sum_{i=1}^{t} \sum_{j=1}^{r} (\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.})$$

The last term of the equation equals zero because $\sum_{j=1}^{r} (Y_{ij} - \overline{Y}_{i.}) = \sum \varepsilon_{ij} = 0$.

Thus, $\sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \overline{Y}_{..})^2 = r\sum_{i=1}^{t} (\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i=1}^{t} \sum_{j=1}^{r} (Y_{ij} - \overline{Y}_{i..})^2$, which is

Total Sum of Squares = Treatment Sum of Squares + Error Sum of Squares

<u>ANOVA for Any Number of Treatments with Equal Replication</u>

Given the following data:

	Treatment			
Replicate	A	B	C	
1	23	42	47	
2	36	26	43	
3	31	47	43	
4	33	34	39	
$Y_{i.}$	123	149	172	$Y_{...}$=444
$\sum Y_{ij}^2$	3,875	5,805	7,428	

Step 1. Write the hypotheses to be tested.

$H_o : \mu_1 = \mu_2 = \mu_3$

$H_A : \mu_1 = \mu_2 \neq \mu_3$

or

$\mu_1 \neq \mu_2 = \mu_3$

or

$\mu_1 \neq \mu_2 \neq \mu_3$

H_o: All three means are equal.

H_A: At least one of the means is different from the other means.

Step 2. Calculate the Correction Factor.

$$CF = \frac{Y_{..}^2}{rt} = \frac{444^2}{4*3} = 16{,}428.0$$

Step 3. Calculate the Total SS

$$TotalSS = \sum Y_{ij}^2 - CF$$

$$= (23^2 + 36^2 + 31^2 + ... + 39^2) - CF$$

$$= 17{,}108 - 16{,}428$$

$$= 680.0$$

Step 4. Calculate the Treatment SS (TRT SS)

$$TRTSS = \sum \frac{Y_{i.}^2}{r} - CF$$

$$= \left(\frac{123^2}{4} + \frac{149^2}{4} + \frac{172^2}{4} \right) - 16{,}428$$

$$= 16728.5 - 16428.0$$

$$= 300.5$$

Step 5. Calculate the Error SS

Error SS = Total SS – Treatment SS

= 680 – 300.5

= 379.5

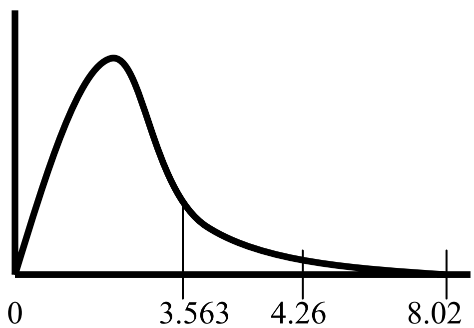
Step 6. Complete the ANOVA table

Sources of variation	Df	SS	MS	F
Treatment	t-1 = 2	300.5	150.25	3.563^{NS}
Error	t(r-1) = 9	379.5	42.167	
Total	rt-1 = 11	680.0		

Step 7. Look up Table F-values.

$F_{0.05;2,9} = 4.26$
$F_{0.01;2,9} = 8.02$

Step 8. Make conclusions.



-Since F-calc (3.563) < 4.26 we fail to reject Ho: $\mu_1 = \mu_2 = \mu_3$ at the 95% level of confidence.
-Since F-calc (3.563) < 8.02 we fail to reject Ho: $\mu_1 = \mu_2 = \mu_3$ at the 99% level of confidence.

Step 9. Calculate Coefficient of Variation (CV).

$$\%CV = \frac{s}{\overline{Y}} * 100$$

Remember that the Error MS = s².

$$\%CV = \frac{\sqrt{42.167}}{\left(\frac{444}{4*3}\right)} * 100$$

$$= (6.494/37) * 100$$

$$= 17.6\%$$

ANOVA for Any Number of Treatments with Unequal Replication

Given the following data:

	Treatment				
Replicate	A	B	C	D	
1	2.0	1.7	2.0	2.1	
2	2.2	1.9	2.4	2.2	
3	1.8	1.5	2.7	2.2	
4	2.3		2.5	1.9	
5	1.7		2.4		
$Y_{i.}$	10	5.1	12	8.4	$Y_{..}$=35.5
$\sum Y_{ij}^2$	20.26	8.75	29.06	17.7	

Step 1. Write the hypotheses to be tested.

$H_o : \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_A: At least one of the means is different from one of the other means.

Step 2. Calculate the Correction Factor.

$$CF = \frac{Y_{..}^2}{\sum r_i} = \frac{35.5^2}{17} = 74.132$$

Step 3. Calculate the Total SS

$$TotalSS = \sum Y_{ij}^2 - CF$$

$$= (2.0^2 + 2.2^2 + 1.8^2 + ... + 1.9^2) - CF$$

$$= 75.77 - 74.132$$

$$= 1.638$$

Step 4. Calculate the Treatment SS (TRT SS)

$$TRTSS = \sum \frac{Y_{i.}^2}{r_i} - CF$$

$$= \left(\frac{10^2}{5} + \frac{5.1^2}{3} + \frac{12^2}{5} + \frac{8.4^2}{4} \right) - 74.132$$

$$= 75.110 - 74.132$$

$$= 0.978$$

Step 5. Calculate the Error SS

Error SS = Total SS – Treatment SS

= 1.638 – 0.978

= 0.660

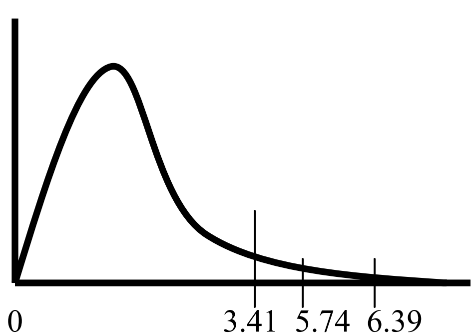
Step 6. Complete the ANOVA table

Sources of variation	Df	SS	MS	F
Treatment	t-1 = 3	0.978	0.326	6.392**
Error	By subtraction = 13	0.660	0.051	
Total	Total number of observations -1 = 16	1.638		

Step 7. Look up Table F-values.

$F_{0.05;3,13} = 3.41$
$F_{0.01;3,13} = 5.74$

Step 8. Make conclusions.



- Since F-calc (6.392) > 3.41 we reject Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ at the 95% level of confidence.
- Since F-calc (6.392) > 5.74 we reject Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ at the 99% level of confidence.

Step 9. Calculate Coefficient of Variation (CV).

$$\%CV = \frac{s}{\overline{\overline{Y}}} * 100$$

<u>Remember that the Error MS = s^2.</u>

$$\%CV = \frac{\sqrt{0.051}}{\left(\frac{35.5}{17}\right)} * 100$$

$$= (0.2259 / 2.088) * 100$$

$$= 10.82\%$$

<u>ANOVA with Sampling (Equal Number of Samples Per Experimental Unit)</u>

Linear Model $Y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \delta_{ijk}$

Where: Y_{ijk} is the k^{th} sample of the j^{th} observation of the i^{th} treatment,
μ is the population mean,
τ_i is the treatment effect of the i^{th} treatment,
ε_{ij} is the random error, and
δ_{ijk} is the sampling error.

<u>ANOVA table</u>

SOV	Df	F
Treatment	t-1	Treatment MS/Experimental Error MS
Experimental error	(tr-1) - (t-1)	
Sampling Error	(trs-1) - (tr-1)	
Total	trs-1	

Facts about ANOVA with Sampling

-There are two sources of variation that contribute to the variance appropriate to comparisons among treatment means.

1. Sampling Error = variation among sampling units treated alike (σ_s^2).
2. Experimental Error = variation among experimental units treated alike ($\sigma_s^2 + r\sigma_E^2$).

-The Experimental Error MS is expected to be larger than the Sampling Error MS.

-If the Experimental Error variance component is not important, the Sampling Error MS and the Experimental Error MS will be of the same order of magnitude.

-If the Experimental Error variance component is important, the Experimental Error MS will be much larger than the Sampling Error MS.

Example

	Temperature								
	8°			12°			16°		
	Pot number			Pot number			Pot number		
Plant	1	2	3	1	2	3	1	2	3
1	3.5	2.5	3.0	5.0	3.5	4.5	5.0	5.5	5.5
2	4.0	4.5	3.0	5.5	3.5	4.0	4.5	6.0	4.5
3	3.0	5.5	2.5	4.0	3.0	4.0	5.0	5.0	6.5
4	4.5	5.0	3.0	3.5	4.0	5.0	4.5	5.0	5.5
$Y_{ij.}$	15.0	17.5	11.5	18.0	14.0	17.5	19.0	21.5	22.0
$Y_{i..}$		44.0			49.5			62.5	

$Y_{...}=156.0$

Note i = treatment, j = replicate, and k = sample.

Step 1. Calculate correction factor:

$$\frac{Y_{...}^2}{rts} = \frac{156^2}{3(3)(4)} = 676$$

Step 2. Calculate the Total SS:

$$TotalSS = \sum Y_{ijk}^2 - CF$$

$$= \left(3.5^2 + 4.0^2 + 3.0^2 + ... + 5.5^2\right) - CF$$

$$= 712.5 - 676.0$$

$$= 36.5$$

Step 3. Calculate the Treatment SS:

$$TreatmentSS = \sum \frac{Y_{i..}^2}{rs} - CF$$

$$= \left(\frac{44^2}{3(4)} + \frac{49.5^2}{(3(4)} + \frac{62.5^2}{3(4)}\right) - 676.0$$

$$= 691.04 - 676.0$$

$$= 15.042$$

Step 4. Calculate the <u>SS</u> <u>A</u>mong <u>E</u>xperimental <u>U</u>nits <u>T</u>otal (SSAEUT)

$$SSAEUT = \sum \frac{Y_{ij.}^2}{s} - CF$$

$$= \left(\frac{15^2}{4} + \frac{17.5^2}{4} + \frac{11.5^2}{4} + ... + \frac{22.0^2}{4}\right) - 676.0$$

$$= 699.25 - 676.0$$

$$= 23.25$$

Step 5. Calculate the Experimental Error SS:

Experimental Error SS = SSAEUT – SS TRT

= 23.25 – 15.042

= 8.208

Step 6. Calculate the Sampling Error SS:

Sampling Error SS = Total SS – SSAEUT

= 36.5 – 23.25

= 13.25

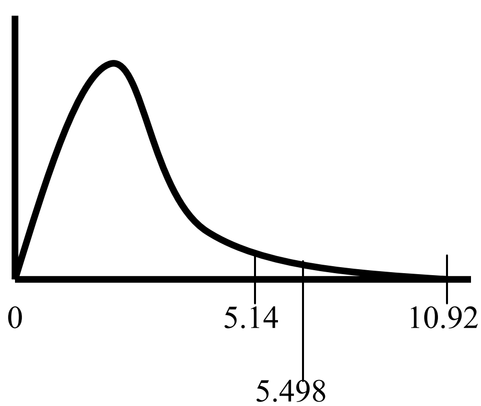
Step 7. Complete the ANOVA Table:

SOV	Df	SS	MS	F
Treatment	t-1 = 2	15.042	7.521	5.498*
Experimental Error	(tr-1) - (t-1) = 6	8.208	1.368	
Sampling Error	(trs-1) - (tr-1) = 27	13.25		
Total	trs-1 = 35	36.5		

Step 8 Look up Table F-values.

$F_{0.05;2,6} = 5.14$
$F_{0.01;2,6} = 10.92$

Step 8. Make conclusions.



- Since F-calc (5.498) > 5.14 we reject Ho: at the 95% level of confidence.
- Since F-calc (5.498) < 10.92 we fail to reject Ho: at the 99% level of confidence.

ANOVA When the Number of Subsamples are Not Equal

$$TotalSS = \sum Y_{ijk}^2 - \frac{Y_{...}^2}{total\#ofobservations}$$ df = #observations – 1

$$TreatmentSS = \sum \frac{Y_{i..}^2}{r_j s_k} - \frac{Y_{...}^2}{total\#ofobs.}$$ df = # treatments –1

$$SSAEUT = \sum \frac{Y_{ij.}^2}{s_k} - \frac{Y_{...}^2}{total\#ofobs.}$$ df = # Experimental units – 1

SS Experimental Error = SSAEUT – SS TRT   df = SSAEUT df – TRT df

SS Sampling Error = Total SS – SSAEUT   df = Total df – SSAEUT df

Assumptions Underlying ANOVA

- Experimental errors are random, independently, and normally distributed about a mean of zero and with a common variance (i.e. treatment variances are homogenous).
- The above assumption can be express as NID(0, σ^2).
- Departure from this assumption can affect both the level of significance and the sensitivity of F- or t-tests to real departures from H_o:
- This results in the rejection of Ho when it is true (i.e. a Type I Error) more often than α calls for.
- The power of the test also is reduced if the assumption of NID(0, σ^2) is violated.
- Violation of the assumption NID(0, σ^2) with the fixed model is usually of little consequence because ANOVA is a very robust technique.
- Violation of the basic assumptions of ANOVA can be investigated by observing plots of the residuals.
- Residuals will be discussed in more detail when Transformations are discussed later in the semester.