EXPERIMENTAL DESIGN TERMINOLOGY [DEVORE 10.1]

DEFINITIONS & NOTATION:

The collection of I samples to determine cause & effect is an **experiment**.
A **balanced experiment** has equal-sized samples/groups.
Each data point of a sample is called an **observation** or **measurement**.
The dependent variable to be measured is called the **response**.
The manner of sample collection & grouping is called **experimental design**.
The main characteristic distinguishing all the samples is called the **factor**.
The factor's particular values or settings are called its **levels**.
Each sample corresponding to a level is called a **group**.

FACTOR A:	GROUP SIZE:	GROUPS:
Level 1	J	$x_{1\bullet}: \ x_{11},\ x_{12},\ \cdots,\ x_{1J}$
Level 2	J	$x_{2\bullet}: \ x_{21},\ x_{22},\ \cdots,\ x_{2J}$
$\vdots$	$\vdots$	$\vdots$
Level I	J	$x_{I\bullet}: \ x_{I1},\ x_{I2},\ \cdots,\ x_{IJ}$

FACTOR A:	GROUP SIZE:	GROUP MEAN:	GROUP STD DEV:
Level 1 ($x_{1\bullet}$)	J	$\overline{x}_{1\bullet}$	s_1
Level 2 ($x_{2\bullet}$)	J	$\overline{x}_{2\bullet}$	s_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Level I ($x_{I\bullet}$)	J	$\overline{x}_{I\bullet}$	s_I

This section (§10.1) & §10.2 involve only balanced experiments.

This chapter's last section (§10.3) considers unbalanced experiments.

EXAMPLES:

Suppose we wish to determine whether three light bulb brands all have similar lifetimes or not.

A sample of 5 bulbs from each brand has their lifetimes measured (in years) and recorded in the below table:

BULB BRAND:	SAMPLE SIZE:	LIFETIMES (in yrs):
Brand 1 ($x_{1\bullet}$)	5	9.22, 9.07, 8.95, 8.98, 9.54
Brand 2 ($x_{2\bullet}$)	5	8.92, 8.88, 9.10, 8.71, 8.85
Brand 3 ($x_{3\bullet}$)	5	9.08, 8.99, 9.06, 8.93, 9.02

or expressed in terms of means and standard deviations:

BULB BRAND:	SAMPLE SIZE:	MEAN LIFETIMES (in yrs):	STD DEV:
Brand 1 ($x_{1\bullet}$)	5	$\overline{x}_{1\bullet} = 9.152$	$s_1 \approx 0.2410$
Brand 2 ($x_{2\bullet}$)	5	$\overline{x}_{2\bullet} = 8.892$	$s_2 \approx 0.1406$
Brand 3 ($x_{3\bullet}$)	5	$\overline{x}_{3\bullet} = 9.016$	$s_3 \approx 0.0594$

THE PROBLEM WITH MANY-SAMPLE t-TESTS [DEVORE 10.1]

Suppose a designed experiment calls to test four independent samples:

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_A &: \quad \text{At least two of the } \mu\text{'s differ} \end{aligned}$$

One way to do this is perform $\binom{4}{2}$ independent t-tests, each at signif. level α:

$$\begin{array}{llllll} H_0^{(1)}: \mu_1 = \mu_2 & H_0^{(2)}: \mu_1 = \mu_3 & H_0^{(3)}: \mu_1 = \mu_4 & H_0^{(4)}: \mu_2 = \mu_3 & H_0^{(5)}: \mu_2 = \mu_4 & H_0^{(6)}: \mu_3 = \mu_4 \\ H_A^{(1)}: \mu_1 \neq \mu_2 & H_A^{(2)}: \mu_1 \neq \mu_3 & H_A^{(3)}: \mu_1 \neq \mu_4 & H_A^{(4)}: \mu_2 \neq \mu_3 & H_A^{(5)}: \mu_2 \neq \mu_4 & H_A^{(6)}: \mu_3 \neq \mu_4 \end{array}$$

Suppose a designed experiment calls to test four independent samples:

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_A &: \quad \text{At least two of the } \mu\text{'s differ} \end{aligned}$$

Alas, since each successive t-test is performed with the same dataset,

the **experiment-wise significance level**, α_{exp}, grows with each t-test:

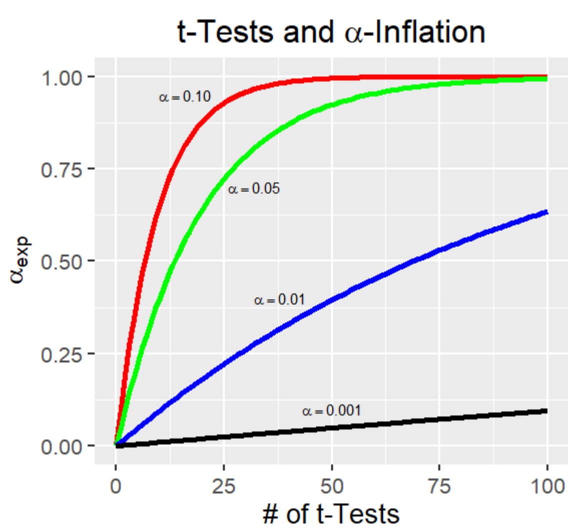
$$\begin{aligned} \alpha_{exp} &:= \mathbb{P}(\text{Committing a Type I Error in at least one } t\text{-test}) \\ &= 1 - \mathbb{P}(\text{Never Committing a Type I Error in any of the } t\text{-tests}) \\ &= 1 - \mathbb{P}\left(\bigcap_{i=1}^{6}(\text{Not Committing a Type I Error in } i^{th} \ t\text{-test})\right) \\ &\overset{IND}{=} 1 - \prod_{i=1}^{6} \mathbb{P}(\text{Not Committing a Type I Error in } i^{th} \ t\text{-test}) \\ &\overset{\alpha}{=} 1 - \prod_{i=1}^{6}(1-\alpha) \\ &= 1 - (1-\alpha)^6 \qquad \left[\alpha := \mathbb{P}\left(\text{Rejecting } H_0^{(k)} \mid H_0^{(k)} \text{ is True}\right)\right] \end{aligned}$$

Alas, with successive t-tests, α_{exp} grows (AKA **α-inflation**):

Chosen α	Resulting $\alpha_{exp} = 1-(1-\alpha)^6$
0.10	**0.4686**
0.05	**0.2649**
0.01	**0.0585**
0.001	**0.0060**

Required $\alpha = 1-(1-\alpha_{exp})^{1/6}$	Desired α_{exp}
0.0174	0.10
0.0085	0.05
0.0017	0.01
0.0002	0.001

A loose (rough) upper bound for α_{exp} is $\alpha \times$ (# t-tests): $\qquad \alpha_{exp} \leq \alpha N_{t\text{-}tests}$



To prevent α-inflation, all means should be simultaneously tested.

1-FACTOR FIXED EFFECTS LINEAR MODELS [DEVORE 10.1]

1-FACTOR FIXED EFFECTS LINEAR (STATISTICAL) MODEL (DEFINITION):

Given a 1-factor balanced experiment with $I > 2$ groups, each of size J.

Let $X_{ij} \equiv$ random variable for j^{th} measurement in the i^{th} group.

Then, the **fixed effects linear (statistical) model** for the experiment is defined as:

$$X_{ij} = \mu + \alpha_i^A + E_{ij} \qquad \text{where} \qquad E_{ij} \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$$

$$\begin{array}{rcl} \mu & \equiv & \text{population grand mean of all } I \text{ population means} \\ \alpha_i^A & \equiv & \text{deviation of } i^{th} \text{ population mean } \mu_i \text{ from } \mu \text{ due to Factor A} \\ E_{ij} & \equiv & \text{rv for error/noise applied to } j^{th} \text{ measurement in } i^{th} \text{ group} \end{array}$$

Fixed effects means all relevant levels of factor A are considered in model.

1-FACTOR LINEAR MODEL (MOTIVATING EXAMPLES):

$$X_{ij} = \mu$$

$$\mu := 3.2$$

$$\mu_1 = 3.2,\ \mu_2 = 3.2,\ \mu_3 = 3.2$$

FACTOR A:	MEASUREMENTS:			
Level 1 ($x_{1\bullet}$)	$x_{11} = 3.2$,	$x_{12} = 3.2$,	$x_{13} = 3.2$,	$x_{14} = 3.2$
Level 2 ($x_{2\bullet}$)	$x_{21} = 3.2$,	$x_{22} = 3.2$,	$x_{23} = 3.2$,	$x_{24} = 3.2$
Level 3 ($x_{3\bullet}$)	$x_{31} = 3.2$,	$x_{32} = 3.2$,	$x_{33} = 3.2$,	$x_{34} = 3.2$

$$X_{ij} = \mu + \alpha_i^A$$

$$\mu := 3.2$$

$$\alpha_1^A := -5.5,\ \alpha_2^A := -2.0,\ \alpha_3^A := 7.5$$

$$\mu_1 = -2.3,\ \mu_2 = 1.2,\ \mu_3 = 10.7$$

FACTOR A:	MEASUREMENTS:			
Level 1 ($x_{1\bullet}$)	$x_{11} = -2.3$,	$x_{12} = -2.3$,	$x_{13} = -2.3$,	$x_{14} = -2.3$
Level 2 ($x_{2\bullet}$)	$x_{21} = 1.2$,	$x_{22} = 1.2$,	$x_{23} = 1.2$,	$x_{24} = 1.2$
Level 3 ($x_{3\bullet}$)	$x_{31} = 10.7$,	$x_{32} = 10.7$,	$x_{33} = 10.7$,	$x_{34} = 10.7$

$$X_{ij} = \mu + \alpha_i^A + E_{ij}$$

$$\mu := 3.2,\ \alpha_1^A := -5.5,\ \alpha_2^A := -2.0,\ \alpha_3^A := 7.5,\ E_{ij} \overset{iid}{\sim} \text{Normal}(0, \sigma^2 := 3.24)$$

$$\mu_1 = -2.3,\ \mu_2 = 1.2,\ \mu_3 = 10.7$$

FACTOR A:	MEASUREMENTS:			
Level 1 ($x_{1\bullet}$)	$x_{11} = -1.23$,	$x_{12} = -1.17$,	$x_{13} = 0.05$,	$x_{14} = -3.08$
Level 2 ($x_{2\bullet}$)	$x_{21} = 0.54$,	$x_{22} = 1.03$,	$x_{23} = 0.62$,	$x_{24} = 1.63$
Level 3 ($x_{3\bullet}$)	$x_{31} = 13.64$,	$x_{32} = 12.30$,	$x_{33} = 11.74$,	$x_{34} = 10.60$

1-FACTOR LINEAR MODELS (POINT ESTIMATORS) [DEVORE 10.1]

1-FACTOR LINEAR MODEL (LEAST-SQUARES ESTIMATORS – LSE's):

Given a 1-factor linear model: $X_{ij} = \mu + \alpha_i^A + E_{ij}$ where $E_{ij} \overset{iid}{\sim}$ Normal$(0, \sigma^2)$ Then:

(a) The **least-squares**[♠♣] **estimators**[†‡] **(LSE's)** for the model parameters are:

$$\begin{aligned} \hat{\mu} &= \overline{x}_{\bullet\bullet} \\ \hat{\alpha}_i^A &= \overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet} \end{aligned} \quad \text{where} \quad \begin{aligned} \overline{x}_{\bullet\bullet} &\equiv \text{Grand sample mean} \\ \overline{x}_{i\bullet} &\equiv \text{Sample mean of } i^{th} \text{ group} \end{aligned}$$

(b) For these least-squares estimators, it's required that $\sum_i \hat{\alpha}_i^A = 0$.

(c) These least-squares estimators are all unbiased.

[†] A. Dean, D. Voss, D. Draguljić, *Design & Analysis of Experiments*, 2^{nd} Ed, Springer, 2017. (§3.4.3)

[‡] D.C. Montgomery, *Design & Analysis of Experiments*, 7^{th} Ed, Wiley, 2009. (§3.3.3, §3.10.1)

[♠] A.M. Legendre, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*, 1806.

[♣] Gauss, *Theoria Motus Corporum Coelestrium in Sectionibus Conicis Solem Ambientium*, 1809.

1-FACTOR LINEAR MODEL (PREDICTED RESPONSES & RESIDUALS):

Given a 1-factor linear model:

$$X_{ij} = \mu + \alpha_i^A + E_{ij} \quad \text{where} \quad E_{ij} \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$$

Then the corresponding **predicted responses**, denoted $\hat{x}_{ij}$, are:

$$\hat{x}_{ij} := \hat{\mu} + \hat{\alpha}_i^A = \overline{x}_{\bullet\bullet} + (\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet}) = \overline{x}_{i\bullet}$$

Moreover, the corresponding **residuals**, denoted x_{ij}^{res}, are:

$$x_{ij}^{res} := x_{ij} - \hat{x}_{ij} = x_{ij} - \overline{x}_{i\bullet}$$

LINEAR MODELS (BEST LINEAR UNBIASED ESTIMATORS – BLUE's):

A point estimator $\hat{\theta}$ is called a **best linear unbiased estimator (BLUE)** if:

- It estimates a parameter θ of a linear model.
- It is a linear combination of the data points: $\hat{\theta} := \sum_{k=1}^n c_k x_k$
- It is an unbiased estimator: $\mathbb{E}[\hat{\theta}] = \theta$
- It has minimum variance of all such unbiased estimators.

REMARK: BLUE's are generally easier to construct & prove than UMVUE's.

1-FACTOR LINEAR MODEL (GAUSS[1]-MARKOV[2] THEOREM):

Given a 1-factor linear model: $X_{ij} = \mu + \alpha_i^A + E_{ij}$

Moreover, suppose the following conditions are all satisfied:

$$\begin{aligned} \mathbb{E}[E_{ij}] &= 0 && \text{(errors are all centered at zero)} \\ \mathbb{V}[E_{ij}] &= \sigma^2 && \text{(errors all have the same finite variance)} \\ \mathbb{C}[E_{ij}, E_{i'j'}] &= 0 && \text{(errors are uncorrelated when } i \neq i' \text{ or } j \neq j') \end{aligned}$$

Then, the least-squares estimators $\hat{\mu}, \hat{\alpha}_i^A$ are all BLUE's.

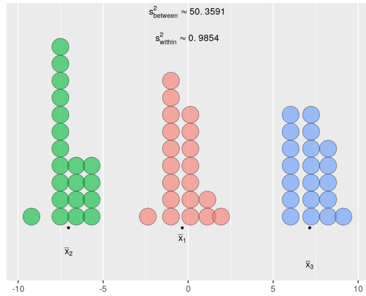
[1] C.F. Gauss, "Theoria Combinationis Observationum Erroribus Minimis Obnoxiae", (1823), 1-58.

[2] A.A. Markov, *Calculus of Probabilities*, 1^{st} Edition, 1900.

1-FACTOR ANOVA (MOTIVATION) [DEVORE 10.1]

High variance between groups

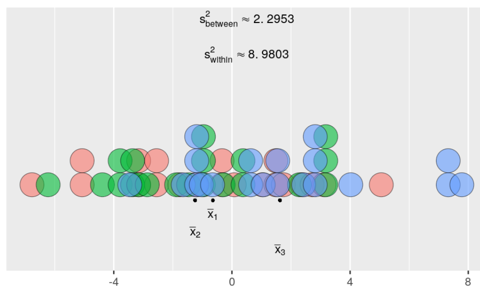
Low variance within groups



$s^2_{between}/s^2_{within} \gg 1 \iff$ Factor A clearly has a significant effect!!

Low variance between groups

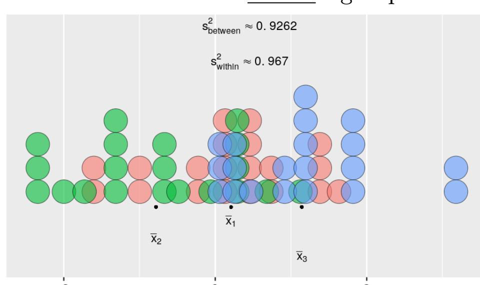
High variance within groups



$s^2_{between}/s^2_{within} \ll 1 \iff$ Factor A clearly has no significant effect!

Low variance between groups

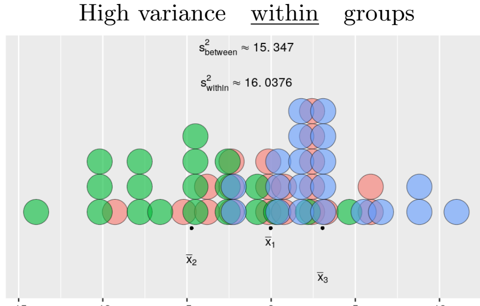
Low variance within groups



$s^2_{between}/s^2_{within} \approx 1 \implies$ Hard to tell if factor A has a significant effect...

High variance between groups

High variance within groups



$s^2_{between}/s^2_{within} \approx 1 \implies$ Hard to tell if factor A has a significant effect...

1-FACTOR ANOVA♠♣ (MOTIVATION) [DEVORE 10.1]

- **1-FACTOR ANOVA BASIC MODEL ASSUMPTIONS:**

  In order for the forthcoming ANOVA test to bear good statistical properties and to utilize the Gauss-Markov Theorem, certain assumptions regarding the samples & populations must be imposed (similarly to t-tests & F-tests):

  - All measurements on units are independent.
  - All groups are approximately normally distributed.
  - All groups have approximately same variance.

- **1-FACTOR ANOVA TEST STATISTIC:**

  Given an experiment with one factor and $I > 2$ groups.

  Moreover, suppose the 1-factor basic ANOVA assumptions are all satisfied.

  Then, the F-test using the following test statistic value: $\quad f = \dfrac{s^2_{between}}{s^2_{within}}$

  is the most-powerful test that prevents α-inflation for hypotheses: $\quad \begin{array}{ll} H_0: & \mu_1 = \mu_2 = \cdots = \mu_I \\ H_A: & \text{At least two of the } \mu\text{'s differ} \end{array}$

- **$s^2_{between}$ IN TERMS OF A MEAN SQUARE & SUM OF SQUARES:**

  The variance between groups, $s^2_{between}$, is the variance of the I group means $\overline{x}_{i\bullet}$ scaled by common group size J:

$$s^2_{between} := \frac{J \cdot \sum_i (\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet})^2}{I-1} = \frac{\sum_i \sum_j (\hat{\alpha}_i^A)^2}{I-1} := \frac{\text{SS}_A}{\nu_A} := \text{MS}_A$$

  where the **grand mean**, $\overline{x}_{\bullet\bullet}$, is the mean of the I group means, $\overline{x}_{i\bullet}$: $\quad \overline{x}_{\bullet\bullet} := \frac{1}{I}\sum_i \overline{x}_{i\bullet} = \frac{1}{IJ}\sum_i \sum_j x_{ij}$

  A large variance between groups indicates much of the observed variation is explained by the chosen Factor A.

- **s^2_{within} IN TERMS OF A MEAN SQUARE & SUM OF SQUARES:**

  The variance within the groups, s^2_{within}, is the mean of the variances of the I groups:

$$s^2_{within} := \frac{1}{I}\sum_i s_i^2 = \frac{(J-1)\cdot \sum_i s_i^2}{I(J-1)} = \frac{\sum_i \sum_j (x_{ij} - \overline{x}_{i\bullet})^2}{I(J-1)} = \frac{\sum_i \sum_j (x_{ij}^{res})^2}{I(J-1)} := \frac{\text{SS}_{res}}{\nu_{res}} := \text{MS}_{res}$$

  Effectively, a large variance within the groups indicates that much of the observed variation is not explained by the chosen Factor A. Therefore, the within variance is considered unexplained error in the experiment.

- **F-TEST STATISTIC VALUE IN TERMS OF MEAN SQUARES:**

$$f_A = \frac{s^2_{between}}{s^2_{within}} = \frac{\text{MS}_A}{\text{MS}_{res}}$$

  The test statistic value for 1-Factor ANOVA will be denoted f_A instead of f.

  In terms of the F-test notation in section 9.5, f_A is always f_+.

♠ R.A. Fisher, "The Correlation between Relatives on the Supposition of Mendelian Inheritance", *Transactions of the Royal Society of Edinburgh*, **52** (1918), 399-433.

♣ R.A. Fisher, *Statistical Methods for Research Workers*, 1925. (Ch VII)

1-FACTOR BALANCED COMPLETELY RANDOMIZED ANOVA (1F bcrANOVA) [DEVORE 10.1]

- **1F bcrANOVA (BALANCED COMPLETELY RANDOMIZED DESIGN):** As an example:
  - Collect 12 relevant experimental units (EU's): $\mathrm{EU}_1, \mathrm{EU}_2, \cdots, \mathrm{EU}_{12}$
  - Produce a random shuffle sequence using software: $(4, 12, 5, 10;\ 7, 2, 1, 11;\ 3, 6, 8, 9)$
  - Use random shuffle sequence to assign the EU's into the I levels.
  - Measure each EU appropriately (note the change in notation):

FACTOR A:	MEASUREMENTS:			
Level 1	EU_4,	EU_{12},	EU_5,	EU_{10}
Level 2	EU_7,	EU_2,	EU_1,	EU_{11}
Level 3	EU_3,	EU_6,	EU_8,	EU_9

$\overset{MEASURE}{\Longrightarrow}$

FACTOR A:	MEASUREMENTS:			
Level 1 ($x_{1\bullet}$)	x_{11},	x_{12},	x_{13},	x_{14}
Level 2 ($x_{2\bullet}$)	x_{21},	x_{22},	x_{23},	x_{24}
Level 3 ($x_{3\bullet}$)	x_{31},	x_{32},	x_{33},	x_{34}

$$\begin{aligned} \mathrm{EU}_k &\equiv \big(k^{th} \text{ experimental unit collected}\big) \\ x_{ij} &\equiv \big(\text{Measurement of } j^{th} \text{ experimental unit in } i^{th} \text{ level}\big) \\ x_{i\bullet} &\equiv \big(\text{Group of all measurements in } i^{th} \text{ level}\big) \end{aligned}$$

- **1F bcrANOVA (FIXED EFFECTS MODEL ASSUMPTIONS):**
  - ⋆ (**1 Desired Factor**) Factor A has I levels.
  - ⋆ (**All Factor Levels are Considered**) AKA Fixed Effects.
  - ⋆ (**Balanced Replication in Groups**) Each group has $J > 1$ units.
  - ⋆ (**Distinct Exp. Units**) All IJ units are distinct from each other.
  - ⋆ (**Random Assignment across Groups**)
  - ⋆ (**Independence**) All measurements on units are independent.
  - ⋆ (**Normality**) All groups are approximately normally distributed.
  - ⋆ (**Equal Variances**) All groups have approximately same variance.

  Mnemonic: **1DF AFLaC BRiG DEU | RAaG | I.N.EV**

- **1F bcrANOVA (SUMS OF SQUARES "PARTITION" VARIATION):**

$$\underbrace{\mathrm{SS}_{total}}_{Total\ Variation\ in\ Experiment} = \underbrace{\mathrm{SS}_A}_{Variation\ due\ to\ Factor\ A} + \underbrace{\mathrm{SS}_{res}}_{Unexplained\ Variation}$$

$$\textstyle\sum_{ij}(x_{ij} - \hat{\mu})^2 = \sum_{ij}(\hat{\alpha}_i^A)^2 + \sum_{ij}(x_{ij}^{res})^2$$

$$\textstyle\sum_i \sum_j (x_{ij} - \overline{x}_{\bullet\bullet})^2 = \sum_i \sum_j (\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet})^2 + \sum_i \sum_j (x_{ij} - \overline{x}_{i\bullet})^2$$

$$\underbrace{\nu}_{Total\ dof's\ in\ Experiment} = \underbrace{\nu_A}_{'Between\ Groups'\ dof's} + \underbrace{\nu_{res}}_{'Within\ Groups'\ dof's}$$

$$\nu = IJ - 1 \qquad \nu_A = I - 1 \qquad \nu_{res} = I(J-1)$$

- **1F bcrANOVA (EXPECTED MEAN SQUARES):**

$$(i)\ \mathbb{E}[\mathrm{MS}_{res}] = \sigma^2, \qquad (ii)\ \mathbb{E}[\mathrm{MS}_A] = \sigma^2 + \frac{J}{I-1}\sum_i (\alpha_i^A)^2$$

- **1F bcrANOVA (POINT ESTIMATORS OF σ^2):**
  - (i) MS_{res} is <u>always</u> an <u>unbiased</u> point estimator of the population variance: $\mathbb{E}[\mathrm{MS}_{res}] = \sigma^2$
  - (ii) If the status quo prevails, MS_A is an <u>unbiased</u> estimator of pop. variance: H_0 is indeed true $\Longrightarrow \mathbb{E}[\mathrm{MS}_A] = \sigma^2$
  - (iii) If the status quo fails, MS_A tends to <u>overestimate</u> population variance: H_0 is indeed false $\Longrightarrow \mathbb{E}[\mathrm{MS}_A] > \sigma^2$

1-FACTOR BALANCED COMPLETELY RANDOMIZED ANOVA (1F bcrANOVA) [DEVORE 10.1]

- **1F bcrANOVA (FIXED EFFECTS LINEAR MODEL):**

1F bcrANOVA Fixed Effects Linear Model		
I	$\equiv$	# groups to compare
J	$\equiv$	# measurements in each group
X_{ij}	$\equiv$	rv for j^{th} measurement taken from i^{th} group
μ_i	$\equiv$	Mean of i^{th} population or true average response from i^{th} group
μ	$\equiv$	Common population mean or true average overall response
α_i^A	$\equiv$	Deviation from μ due to i^{th} group
E_{ij}	$\equiv$	Deviation from μ due to random error
ASSUMPTIONS: $E_{ij} \overset{iid}{\sim} \text{Normal}\left(0, \sigma^2\right)$		
$X_{ij} = \mu + \alpha_i^A + E_{ij}$ where $\sum_i \alpha_i^A = 0$		
$H_0^A:$ All $\alpha_i^A = 0$ $H_A^A:$ Some $\alpha_i^A \neq 0$		

- **1F bcrANOVA (F-TEST PROCEDURE):**

1. Determine df's: $n = IJ, \quad \nu_A = I - 1, \quad \nu_{err} = I(J-1)$
2. Compute Group Means (if not provided): $\overline{x}_{i\bullet} := \underbrace{\frac{1}{J}\sum_j x_{ij}}_{\text{Given measurements}}$
3. Compute Group Variances (if not provided): $s_i^2 := \underbrace{\frac{1}{J-1}\sum_j (x_{ij} - \overline{x}_{i\bullet})^2}_{\text{Given measurements}} = \underbrace{\sqrt{J} \cdot \widehat{\sigma}_{\overline{x}_{i\bullet}}}_{\text{Given ESE's}}$
4. Compute Grand Mean: $\overline{x}_{\bullet\bullet} := \frac{1}{I}\sum_i \overline{x}_{i\bullet}$
5. Compute $\text{SS}_{res} := \sum_{ij}(x_{ij}^{res})^2 = \sum_i \sum_j (x_{ij} - \overline{x}_{i\bullet})^2 = (J-1) \cdot \sum_i s_i^2$
6. Compute $\text{SS}_A := \sum_{ij}(\hat{\alpha}_i^A)^2 = \sum_i \sum_j (\overline{x}_{i\bullet} - \overline{x}_{\bullet\bullet})^2$
7. Compute Mean Squares: $\text{MS}_{res} := \dfrac{\text{SS}_{res}}{\nu_{res}}, \qquad \text{MS}_A := \dfrac{\text{SS}_A}{\nu_A}$
8. Compute Test Statistic Value: $f_A = \dfrac{\text{MS}_A}{\text{MS}_{res}}$
9. Compute P-value: $p_A := \mathbb{P}(F > f_A) \approx 1 - \Phi_F(f_A; \nu_A, \nu_{res})$
10. Render Decision:
    (by software) If $p_A \leq \alpha$ then reject H_0^A in favor of H_A^A, else accept H_0^A.
    (by hand) If $f_A \geq f^*_{\nu_A, \nu_{res}; \alpha}$ then reject H_0^A in favor of H_A^A, else accept H_0^A.

- **1F bcrANOVA (SUMMARY TABLE):**

1F bcrANOVA Table (Significance Level α)

Variation Source	**df**	**Sum of Squares**	**Mean Square**	**F Stat Value**	**P-value**	**Decision**
Factor A	ν_A	SS_A	MS_A	f_A	p_A	Acc/Rej H_0^A
Unknown	ν_{res}	SS_{res}	MS_{res}			
Total	ν	SS_{total}				

1-FACTOR ANOVA (EFFECT SIZE MEASURES) [DEVORE 10.1]

- **EFFECT SIZE MEASURES (MOTIVATION):**

Recall when performing a hypothesis test, statistical significance does not necessarily imply practical significance. As Gravetter & Wallnau put it in §13.5 of their statistics textbook[GW]:

"the term *significant* does not necessarily mean *large*, it simply means larger than expected by chance."

Q: How to determine whether a statistically significant effect is a practical (i.e. large enough) effect??

A: Effect size measures! What follows are 3 such popular measures.

- **EFFECT SIZE MEASURES DUE TO FISHER, KELLEY & HAYS:**

YEAR	NAME	MEASURE	HOW IT COMPARES*
1925†	Fisher[GW],[H],[LH],[S]	$\hat{\eta}_A^2 := \dfrac{\text{SS}_A}{\text{SS}_{total}}$	Most biased (positively)♠ Least SD, Most RMSE♠
1935‡	Kelley	$\hat{\epsilon}_A^2 := \dfrac{\text{SS}_A - \nu_A\text{MS}_{res}}{\text{SS}_{total}}$	Least biased (negatively)♠ Most SD, Nearly Least RMSE♠
1963♣	Hays[H],[LH],[S]	$\hat{\omega}_A^2 := \dfrac{\text{SS}_A - \nu_A\text{MS}_{res}}{\text{SS}_{total} + \text{MS}_{res}}$	Moderately biased (negatively)♠ Moderate SD, Least RMSE♠

*Requires all 1-Factor ANOVA assumptions (LADR'S RAIN EV) to be satisfied.

SD ≡ **S**tandard **D**eviation, RMSE ≡ **R**oot **M**ean **S**quared **E**rror

†R.A. Fisher, *Statistical Methods for Research Workers*, 1925. (Chapter VIII, §45)

‡T.L. Kelley, "An Unbiased Correlation Ratio Measure", *Proceedings of Nat. Acad. Sciences*, **21** (1935), 554-559.

♣W.L. Hays, *Statistics for Psychologists*, 1963.

♠K. Okada, "Is Omega Squared Less Biased? A Comparison ... ", *Behaviormetrika*, **40** (2013), 129-147.

- **EFFECT SIZE MEASURES (GENERAL REMARKS):**

There are about 75 different effect size measures◇ that have been discovered!!

◇S.F. Davis (Ed.), *Handbook of Research Methods in Experimental Psychology*, 2003. (Chapter 5 by R.E. Kirk)

Moreover, realize that many of these measures are 'measures of association' and, hence, are tailored for either numerical-numerical (num-num) inference (Ch 12 & 13) or categorical-categorical (cat-cat) inference (Ch 14).

Cutoff values for "small"/"medium"/"large" effects vary by field[LH]:

– J. Cohen, *Statistical Power Analysis for Behavioral Sciences*, 1969. (§8.2)

Be very careful when interpreting values of effect size measures[S], especially for 2-Factor ANOVA or higher:

– K.E. O'Grady, "Measures of Explained Variance: Cautions and Limitations", *Psych. Bulletin*, **92** (1982), 766-777.

– C.A. Pierce, R.A. Block, H. Aguinis, "Cautionary Note on Reporting Eta-Squared Values from Multifactor ANOVA Designs", *Educational & Psychological Measurement*, **64** (2004), 916-924.

- **REFERENCES:**

$[GW]$	F.J. Gravetter L.B. Wallnau	*Statistics for the Behavioral Sciences*	7^{th} Ed	2007
$[H]$	D.C. Howell	*Statistical Methods for Psychology*	7^{th} Ed	2010
$[LH]$	R.G. Lomax D.L. Hahs-Vaughn	*Statistical Concepts : A Second Course*	4^{th} Ed	2012
$[S]$	J.P. Stevens	*Intermediate Statistics A Modern Approach*	3^{rd} Ed	2007

EX 10.1.1: Given the following 1-factor balanced experiment:

FACTOR A:	GROUP SIZE:	MEASUREMENTS:
Level 1 ($x_{1\bullet}$)	4	$x_{11},\ x_{12},\ x_{13},\ x_{14}$
Level 2 ($x_{2\bullet}$)	4	$x_{21},\ x_{22},\ x_{23},\ x_{24}$
Level 3 ($x_{3\bullet}$)	4	$x_{31},\ x_{32},\ x_{33},\ x_{34}$

(a) Formulate this experiment as a 1-Factor ANOVA fixed effects linear model.

(b) Use multivariable calculus to compute the least-squares estimators (LSE's) for this linear model.
Be sure to explain why the constraint $\sum_i \alpha_i = 0$ is necessary to impose.

(c) Show that each least-squares estimator (LSE) found in part (b) is a linear combination of the data points.

(d) Establish the sums of squares partition for the general version of this linear model with I levels each of size J.

EX 10.1.2: The lifetimes of three light bulb brands were measured and summarized into this table:

BULB BRAND	**(BULB LIFETIMES in yrs)**
Brand 1 ($x_{1\bullet}$)	9.22, 9.07, 8.95, 8.98, 9.54
Brand 2 ($x_{2\bullet}$)	8.92, 8.88, 9.10, 8.71, 8.85
Brand 3 ($x_{3\bullet}$)	9.08, 8.99, 9.06, 8.93, 9.02

A 1-Factor Balanced Completely Randomized ANOVA (1F bcrANOVA) at significance level $\alpha = 0.01$ is to be performed.

(a) Identify factor A and its levels.

(b) Determine factor A's level count, I, common group sample size, J, and degrees of freedom, ν_{res} & ν_A.

(c) State the appropriate null hypothesis H_0^A & alternative hypothesis H_A^A.

(d) Compute the cell means, $\overline{x}_{i\bullet}$.

(e) Compute the cell variances, s_i^2.

(f) Compute the grand mean, $\overline{x}_{\bullet\bullet}$.

(g) Compute the sums of squares, SS_{res} & SS_A.

(h) Compute the square means, MS_{res} & MS_A.

(i) Compute the test F-statistic, f_A, and Fisher's effect size measure, $\hat{\eta}_A^2$.

(j) By hand, lookup F cutoff value, $f^*_{\nu_A,\nu_{res};\alpha}$. By software (SW), compute resulting P-value, p_A.

(k) Render the appropriate decision. Also, interpret Fisher's effect size measure, $\hat{\eta}_A^2$.

(l) Summarize everything in an 1-Factor ANOVA table.

EX 10.1.3: Dentists use resin composites and ceramic fillings among others for cavities in teeth. The shear bond strengths of resin composite-ceramic bonds formed from three possible configurations (conventional, reversed, all-composite) were measured (in MPa) and summarized in the following table:

GROUP:	**SAMPLE SIZE:**	**MEAN:**	**STD DEV:**
Conventional	10	$\overline{x}_{1\bullet} = 10.37$	$s_1 = 1.99$
Reversed	10	$\overline{x}_{2\bullet} = 18.02$	$s_2 = 2.52$
All-Composite	10	$\overline{x}_{3\bullet} = 21.82$	$s_3 = 2.45$

This table and all the details regarding the experiment can be found in the following paper:

A. Della Bona, R. van Noort, "Shear vs. Tensile Bond Strength of Resin Composite Bonded to Ceramic", *Journal of Dental Research*, **74** (1995), 1591-1596.

A 1-Factor Balanced Completely Randomized ANOVA (1F bcrANOVA) at significance level $\alpha = 0.05$ is to be performed.

(a) Identify factor A and its levels.

(b) Determine factor A's level count, I, common group sample size, J, and degrees of freedom, ν_{res} & ν_A.

(c) State the appropriate null hypothesis H_0^A & alternative hypothesis H_A^A.

(d) Compute the grand mean, $\overline{x}_{\bullet\bullet}$.

(e) Compute the sums of squares, SS_{res} & SS_A.

(f) Compute the square means, MS_{res} & MS_A.

(g) Compute the test F-statistic, f_A, and Fisher's effect size measure, $\hat{\eta}_A^2$.

(h) By hand, lookup F cutoff value, $f^*_{\nu_A,\nu_{res};\alpha}$. By software (SW), compute resulting P-value, p_A.

(i) Render the appropriate decision. Also, interpret Fisher's effect size measure, $\hat{\eta}_A^2$.

(j) Summarize everything in an 1-Factor ANOVA table.