Topic 10: *Fixed*, *Random*, and *Mixed Models*:

The purpose of this chapter is to introduce ANOVA models appropriate to different **experimental objectives**

10.2.1: ***Model I ANOVA*** *or* ***fixed model***

1. Treatment effects are additive and **fixed by the researcher**
2. The researcher is only interested in these **specific treatments** and will limit his conclusions to them.
3. Model $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
   where α_i will be **the same** if the experiment is repeated

$$\boxed{\sum \tau_i = 0 \quad \text{and} \quad H_0\text{: is } \tau_1 = \dots = \tau_i = 0 \quad \text{vs.} \quad H_A\text{: some } \tau_i \neq 0}$$

4. When H_o is false three will be an additional component in the variance between treatments = $r\sum\tau_i^2/(t-1)$

10.2.2: ***Model II ANOVA*** *or* ***random model*** or **components of variance model**

1. The treatments are a ***random sample*** from a larger population of treatments for which the mean is zero and the variance is σ^2_t
2. The **objective** of the researcher ***is to extend the conclusions*** based on the sample of treatments ***to ALL treatments*** in the population
3. Here the **treatment effects** are random variables (s_i) and ***knowledge about the particular* ones** investigated is relatively ***useless***
4. Model $Y_{ij} = \mu + s_i + \varepsilon_{ij}$
   where s_i will be **different** if the experiment is repeated

$$\boxed{\sum s_i \neq 0 \quad s_i \text{ are } N(0, \sigma^2_s) \quad H_0\text{: is } s^2_s = 0 \quad \text{vs. } H_A\text{: } s^2_s \neq 0}$$

5. When the null hypothesis is false there will be an additional component of variance equal to $r\sigma^2_s$.
6. The researcher wants to test the presence and estimate the magnitude of the added variance component among groups: σ^2_s.
7. For One Way ANOVA, the computation is the same for the **fixed** and **random** Models. However, the **objectives** and the **conclusions** are different. The computations following the initial significance test are also different.

10. 2. 3. Differences between fixed and random-effects model

1) **The objectives are different**.

For the Fixed model: In the fertilizer experiment each fertilizer is of specific importance. They are not a random sample. The purpose of the study is to compare these treatments. The purpose of a fixed model is to test the hypothesis that the treatment effects are the same.

For the Random model: In the breeding study the objective is to estimate the variance of a breed. The sires are merely **a sample** from which inferences are to be made concerning the **population**.

The purpose of a random model is to estimate σ^2_s and σ^2_ε, the variance components (**H_o $\sigma^2_s = 0$ vs. H_1 $\sigma^2_s > 0$**)

2) **The sampling procedures are different**.

For the Fixed model: treatments are not randomized but are selected purposefully by the experiment. If the experiment is repeated, the 4 fertilizers must be used again and only the random errors are changed, i.e., α_i's are assumed to be constants and do not change, only ε_{ij}'s change.

For the Random model: treatments are randomly selected and the variance in the population of treatments contributes to the total sum of squares. In the 2nd experiment, the 4 sires most likely will be changed from experiment to experiment. Thus not only the errors are changeable but the sire effect, s_i, are changeable.

3) **The expected sums of the effects are different**.

For the Fixed model: $\Sigma\alpha_i = \Sigma(\mu-\mu_i) = 0$.

For the Random model: $\Sigma s_i = \Sigma(\mu-\mu_i) \neq 0$

4) **The expected variances are different**

For the Fixed model:

Var. of Y_{ij} = variance of $\varepsilon_{ij} = \sigma^2_\varepsilon$

For the Random model:

Var. of Y_{ij} = variance of s_i + variance of $\varepsilon_{ij} = \sigma^2_s + \sigma^2_e$

(σ^2_s and σ^2_ε are called **variance components)**

Table 1. Expected mean of **one-way** classification experiments.

			EMS= Expected MS	
Source	df	MS	Fixed Model	Random Model
Treatment	a-1	MST	$\sigma^2_\varepsilon + r\,\Sigma\alpha^2_i/(a-1)$	$\sigma^2_\varepsilon + r\,\sigma^2_s$
Exp. error	a (r-1)	MSE	σ^2_ε	σ^2_ε

To calculate $\sigma^2_s = (MST - \sigma^2_\varepsilon)/r$

10. 3. Two-way Classification Experiments: fixed, random or mixed

$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$,

Suppose three different scenarios

Fix. Breeder interested in a particular set of varieties in a particular set of locations

Random: interested in a random samples of varieties released during different decades in a random set of locations representing a particular region

Mixed: interested in a fix varieties released in a random set of locations representing a particular region

	Fix	**Random**	**Mixed**
μ	μ of specific varieties	μ of all varieties in that period	μ of specific varieties
α_i	$\mu_{i.}-\mu$, effect Var ith $\Sigma\alpha_i = 0$	random ~N(0, σ^2_α) $\Sigma\alpha_i \neq 0$	$\mu_{i.}-\mu$, effect Var ith $\Sigma\alpha_i = 0$
β_j	$\mu_{.j} - \mu$, effect Loc jth $\Sigma\beta_j = 0$	random ~N(0, σ^2_β) $\Sigma\beta_j \neq 0$	~N(0, σ^2_β) random $\Sigma\beta_j \neq 0$
$(\alpha\beta)_{ij}$	specific inter. effect ith var. & jth loc.	random ~N(0, $\sigma^2_{\alpha\beta}$) from all possible int.	random ~N(0, $\sigma^2_{\alpha\beta}$) from all possible int.
Variance	σ^2_ε	$\sigma^2_\alpha+\sigma^2_\beta+\sigma^2_{\alpha\beta}+\sigma^2_\varepsilon$	$\sigma^2_\beta+\sigma^2_{\alpha\beta}+\sigma^2_\varepsilon$
Objective	Test H about α_i, β_j, & $(\alpha\beta)_{ij}$	Estimate σ^2_α, σ^2_β, & $\sigma^2_{\alpha\beta}$	Test H about α_i. Estimate σ^2_β, & $\sigma^2_{\alpha\beta}$

10. 3. 4. Expected mean squares and F tests.

EMS: algebraic expressions specifying the model parameters estimated by the calculated MS. They are used to determine the appropriate error variances for an F test.

EMS include:

- The error variances
- Functions of variances of random effects
- Functions of sums of squares and products of fixed effects

Expected mean squares of two-way classification experiments with a varieties, b locations, and r replications.

		Fixed	Random	Mixed Model
Source	df	Model	Model	Loc.=Rand. Var.=Fixed
Loc.	b-1	$\sigma^2_\varepsilon + \frac{ar\Sigma\beta^2}{(b-1)}$	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta} + ar\sigma^2_\beta$	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta} + ar\sigma^2_\beta$
Variety	a-1	$\sigma^2_\varepsilon + \frac{br\Sigma\alpha^2}{(a-1)}$	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta} + br\sigma^2_\alpha$	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta} + br\Sigma\alpha^2/(a-1)$
L * V	(b-1)(a-1)	$\sigma^2_\varepsilon + \frac{r\Sigma\Sigma(\alpha\beta)^2}{(b-1)(a-1)}$	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta}$	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta}$
Error	ba(r-1)	σ^2_ε	σ^2_ε	σ^2_ε

Based on the expected mean squares an appropriate denominator mean square can be found to form an F test concerning location, variety and interaction.

Appropriate F tests for various models.

		Fixed	Random	Mixed (V=fix)
Location	b-1	MSL/**MSE**	MSL/MSLV	MSL/ MSLV
Variety	a-1	MSV/**MSE**	MSV/MSLV	MSV/MSLV
Interaction	(b-1)(a-1)	MSLV/**MSE**	MSLV/**MSE**	MSLV/**MSE**

Note that the appropriate F tests change, depending on the model.

If one factor in a factorial is RANDOM, then even if we have a significant interaction we might be interested in testing the main fixed effects!

ST&D is outdated in EMS formulas: do not use Tables from p260 or p380.

10. 3. 4. 1. Rules for Determining Expected Mean Squares (EMS)

EMS for balanced factorial, nested, or nested factorial, experiment. Incomplete designs such as **Latin squares** are **excluded**

RULE 1. The error term $\varepsilon_{ij...m}$, is written as $\varepsilon_{(ij...)\,m}$, where the subscript m denotes the replication subscript. For the two-factor model= $\varepsilon_{(ij)\,k}$.

RULE 2. In addition to the error term the model contains all the main effects and any interactions that the **experimenter assumes exist**.

RULE 3. For each term in the model, divide the subscripts into three classes:

a) **Live**: subscripts not in parentheses.
b) **Dead**: subscripts in parentheses (usually nested factors).
c) **Absent**: subscripts not present in a particular term.

RULE 4. **Degrees of freedom**. The Number of d.f. for any term is the product of: N of levels of dead subscript **x** N of levels **minus 1** for each live subscript.

RULE 5. Each effect has either a variance component (random effect) or a fixed factor (fixed effect) associated with it.
If an interaction contains at least one random effect, the entire interaction is considered as random.

RULE 6. **EMS**. prepare a Table with a **row** for each **model component** and a **column** for each **subscript**. Over each subscript, write the number of levels of the factor and whether the factor is fixed (F) or random (R). Replicates are random.

a) Write **1** if a **dead subscript** in the row matches the subscript in the column

b) If any of the subscripts on the row = subscript in the column, write **0** if the column is headed by a **fixed factor** and **1** if the column is headed by a **random factor**. For **interactions** with **at least 1 random factor** write **1** for columns with those subscripts.

c) In the remaining positions, write the N of levels shown above the column heading:

Fixed or Random	F	R	R
Number of levels	a	b	r
Factor	i	j	k
α_i	**0**	**b**	**r**
β_j	**a**	**1**	**r**
$(\alpha\beta)_{ij}$	**1**	**1**	**r**
$\varepsilon_{(ij)\,k}$	**1**	**1**	**r**

d) To obtain the EMS: 1st cover all columns headed by live subscripts on that component. Then, in each row that contains at least the same subscripts as those on the component being considered, take the product of the visible numbers and multiply by the appropriate fixed or random factor from Rule 5. The sum of these quantities is the expected mean of the model component being considered.

Example for a mixed model two-way ANOVA

RULE 1. error term = $\varepsilon_{(ij)\,k}$.

RULE 2. here the **experimenter assumes** α_i, ß$_j$, and $(\alpha$ß$)_{ij}$.

RULE 3. in $(\alpha\beta)_{ij}$ i and j are live and k is absent

in $\varepsilon_{(ij)\,k}$ k is live and i and j are dead.

RULE 4. df $(\alpha\beta)_{ij}$= (a-1)(b-1);

df $\varepsilon_{(ij)\,k}$=ab(r-1)

RULE 5. A variance component has Greek letters as subscripts: $\sigma^2_{ß}$, $\sigma^2_{\alpha ß}$.
A fixed effect is always represented by:

$$\frac{\sum_{i=1}^{a}\alpha_i^2}{a-1}$$

RULE 6.

Fix or Rand.	F	R	R	F	R	R	F	R	R	Expected Mean Squares
N of levels	a	b	r	a	b	r	a	b	r	
Factor	i	j	k	i	j	k	i	j	k	
α_i				0			0	b	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta} + br\Sigma\alpha^2/(a-1)$
β_j					1		a	1	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta} + ar\sigma^2_\beta$
$(\alpha\beta)_{ij}$				1	1		1	1	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta}$
$\varepsilon_{(ij)\,k}$	1	1		1	1	1	1	1	1	σ^2_ε

- To find E(MS$_A$):
  - cover column i.
  - The product of the visible numbers in the rows that contain at least subscript i are **br** (row 1), r (row 3), and 1 (row 4).
  - Note that i is missing in row 2.
- To find E(MS$_{A*B}$)
  - cover columns i and j.
  - The product of the visible numbers in the rows that contain at least subscripts i and j are r (row 3), and 1 (row 4). Note that i and j are not present simultaneously in rows 1 and 2.

Use the above rules to complete the following EMS

Exercise 1 three-way ANOVA

a levels of factor A (fixed),
b levels of factor B (random)
c levels of factor **C**(random)
r replicates.

The researcher wants to include all interactions

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

Fixed or Random Number of levels Factor					**Expected Mean Squares**	**F**
α_i						
β_j						
γ_k						
$(\alpha\beta)_{ij}$						
$(\alpha\gamma)_{ik}$						
$(\beta\gamma)_{jk}$						
$(\alpha\beta\gamma)_{ijk}$						
$\varepsilon_{(ijk)l}$						

Exercise 2 with nested effects (Answer page 10)

2–way factorial A (fixed) x B (random) with C (random) nested in AxB with d replications. Subscripts in parentheses indicate the nested components

Fix or Rand. N of levels Factor					**Expected Mean Squares**	**F**
α_i						
β_j						
$(\alpha\beta)_{ij}$						
$\gamma(\alpha\beta)_{k(ij)}$						
$\varepsilon_{(ijk)l}$						

10. 4. Expected Mean squares for three-way ANOVA

a levels of factor A (fixed), b levels of factor B (random), **c** levels of factor **C**(random), and **r** replicates.

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

Fixed or Random Number of levels Factor	F a i	R b j	R c k	R r l	**Expected Mean Squares**	**F**
α_i	0	b	c	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta\gamma} + br\sigma^2_{\alpha\gamma} + cr\sigma^2_{\alpha\beta} + bcr\Sigma\alpha^2/(a-1)$	**?**
β_j	a	1	c	r	$\sigma^2_\varepsilon +$ **$r\sigma^2_{\alpha\beta\gamma}$** $+ ar\sigma^2_{\beta\gamma} +$ **$cr\sigma^2_{\alpha\beta}$** $+ acr\sigma^2_\beta$	**?**
γ_k	a	b	1	r	$\sigma^2_\varepsilon +$ **$r\sigma^2_{\alpha\beta\gamma}$** $+ ar\sigma^2_{\beta\gamma} +$ **$br\sigma^2_{\alpha\gamma}$** $+ abr\sigma^2_\gamma$	**?**
$(\alpha\beta)_{ij}$	**1**	1	c	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta\gamma} + cr\sigma^2_{\alpha\beta}$	$MS_{\alpha\beta}/$**$MS_{\alpha\beta\gamma}$**
$(\alpha\gamma)_{ik}$	**1**	b	1	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta\gamma} + br\sigma^2_{\alpha\gamma}$	$MS_{\alpha\gamma}/$**$MS_{\alpha\beta\gamma}$**
$(\beta\gamma)_{jk}$	a	1	1	r	$\sigma^2_\varepsilon +$ **$r\sigma^2_{\alpha\beta\gamma}$** $+ ar\sigma^2_{\beta\gamma}$	$MS_{\beta\gamma}/$**$MS_{\alpha\beta\gamma}$**
$(\alpha\beta\gamma)_{ijk}$	**1**	1	1	r	$\sigma^2_\varepsilon + r\sigma^2_{\alpha\beta\gamma}$	$MS_{\alpha\beta\gamma}/$**MS_ε**
$\varepsilon_{(ijk)l}$	1	1	1	1	σ^2_ε	

The 1's in bold would be zero in ST&D. The variances in bold would be absent in ST&D.

No exact test for the main effect A, B, or C!

Option 1: use SAS **RANDOM** and **TEST option**

Identical results to those presented in the previous table can be obtained in SAS using the **RANDOM** statement in PROC GLM.

TEST option in the RANDOM statement: determines what tests are appropriate, and provides F ratios and probabilities for these tests.

Contrasts are not corrected by the RANDOM statement, and there is no way for the user to specify a synthetic denominator for a contrast (need to learn **PROC MIXED**).

All random effects and all interactions including at least one random effect must be explicitly designated as random in the SAS RANDOM statement

Example: **A** (fix) 3 levels, **B** (random) 5 levels, **C** (random) 2 levels, 1 rep.
model Y= A B C A*B A*C B*C A*B*C;
random B C A*B A*C B*C A*B*C;.

```
Source    Type III Expected Mean Square
A         Var(Error) + Var(A*B*C) + 5 Var(A*C) + 2 Var(A*B) + Q(A)
B         Var(Error) + Var(A*B*C) + 3 Var(B*C) + 2 Var(A*B) + 6 Var(B)
C         Var(Error) + Var(A*B*C) + 3 Var(B*C) + 5 Var(A*C) + 15 Var(C)
A*B       Var(Error) + Var(A*B*C) + 2 Var(A*B)
A*C       Var(Error) + Var(A*B*C) + 5 Var(A*C)
B*C       Var(Error) + Var(A*B*C) + 3 Var(B*C)
A*B*C     Var(Error) + Var(A*B*C)
```

10. 5. Approximate F tests

Satterthwaite (1946) method of **linear combinations of mean squares:**

$MS' = MS_r + \cdots + MS_S$ and $MS'' = MS_U + \cdots + MS_V$

Mean squares are chosen so that no MS appear simultaneously in MS' and MS'', and E(MS') - E(MS'') is equal to the effect (the model parameter or variance component) considered in the null hypothesis. The test statistic is:

$$F = MS'/MS''$$

which is distributed approximately as $F_{p,\,q}$,where ***p*** and ***q*** are the ***effective degrees of freedom*.**

$$p = \frac{(MS_r + \cdots + MS_S)^2}{MS_r^2/df_r + \cdots + MS_S^2/df_S} \qquad q = \frac{(MS_u + \cdots + MS_V)^2}{MS_U^2/df_U + \cdots + MS_V^2/df_V}$$

p and *q* may not be integers so interpolate in the F tables may be required.

In the 3-factor mixed effects model discussed above, an appropriate test statistic for H_0: $\alpha_1 = \ldots = \alpha_t = 0$ would be

$$F = \frac{MS_A + MS_{ABC}}{MS_{AB} + MS_{AC}}$$

$$F = \frac{\sigma_\varepsilon^2 + r\sigma_{\alpha\beta\gamma}^2 + br\sigma_{\alpha\gamma}^2 + cr\sigma_{\alpha\beta}^2 + bcr\sum\frac{\alpha^2}{a-1} + \sigma_\varepsilon^2 + r\sigma_{\alpha\beta\gamma}^2}{\sigma_\varepsilon^2 + r\sigma_{\alpha\beta\gamma}^2 + br\sigma_{\alpha\gamma}^2 + cr\sigma_{\alpha\beta}^2 \quad + \quad \sigma_\varepsilon^2 + r\sigma_{\alpha\beta\gamma}^2}$$

This is an F test for Factor A, represented by the term

$bcr\sum\frac{\alpha^2}{a-1}$ that is the only one term that is not repeated in the numerator and the denominator

d.f. for F: from equations for ***p*** and ***q***. Since no MS appears in both the numerator and denominator, they are independent.

It is better to define additive linear combinations in the numerator and denominator than to subtract (e.g $MS_A / (MS_{AB} + MS_{AC} - MS_{ABC})$) (ST&D pg. 380) [but SAS subtracts]

Additional examples with nested effects

The rules for EMS are also applicable when a factorial experiment includes subsampling. Subscripts in parentheses indicate the nested components

Answer Exercise 2**:** 2–way factorial A (fixed) x B (random) with C replications (random) nested in AxB with d subsamples per replication.

Fix or Rand.	F	R	R	R	**Expected Mean Squares**	**F**
N of levels	a	b	c	d		
Factor	i	j	k	l		
α_i	0	B	c	d	$\sigma^2_\varepsilon + d\sigma^2_{\gamma(\alpha\beta)} + cd\sigma^2_{\alpha\beta} + bcd\Sigma\alpha^2/(a-1)$	$MS_\alpha/MS_{\alpha\beta}$
β_j	a	1	c	d	$\sigma^2_\varepsilon + d\sigma^2_{\gamma(\alpha\beta)} + \boldsymbol{cd\sigma^2_{\alpha\beta}} + acd\sigma^2_\beta$	$MS_\beta/MS_{\alpha\beta}$
$(\alpha\beta)_{ij}$	**1**	1	c	d	$\sigma^2_\varepsilon + d\sigma^2_{\gamma(\alpha\beta)} + cd\sigma^2_{\alpha\beta}$	$MS_{\alpha\beta}/MS_{\gamma(\alpha\beta)}$
$\gamma(\alpha\beta)_{k(ij)}$	1	1	1	d	$\sigma^2_\varepsilon + d\sigma^2_{\gamma(\alpha\beta)}$	$MS_{\gamma(\alpha\beta)}/MS_\varepsilon$
$\varepsilon_{(ijk)l}$	1	1	1	1	σ^2_ε	

Blocks nested in locations and locations RANDOM.

Eleven varieties of barley tested in 2 locations selected at random from the Sacramento Valley. In each location the experiment is organized as a RCBD with 13 blocks.

Fixed or Random	F	R	R	**Expected Mean Squares**	**F**
Number of levels	11	2	13		
Factor	i	j	k		
α_i (variety)	0	2	13	$\sigma^2_\varepsilon+13\sigma^2_{\alpha\beta}+26\Sigma\alpha^2/4$	$MS_\alpha/$ $MS_{\alpha\beta}$
β_j (loct.)	11	1	13	$\sigma^2_\varepsilon+13\sigma^2_{\alpha\beta}+11\sigma^2_{\gamma(j)}+143\sigma^2_\beta$	$(MS_\beta+MS_\varepsilon)/(MS_{\alpha\beta}+MS_{\gamma(j)})$
$\gamma_{k(j)}$ (Block within location)	11	1	1	$\sigma^2_\varepsilon+11\sigma^2_{\gamma(j)}$	$MS_{\gamma(j)}/MS_\varepsilon$
$(\alpha\beta)_{ij}$	**1**	1	13	$\sigma^2_\varepsilon+13\sigma^2_{\alpha\beta}$	$MS_{\alpha\beta}/MS_\varepsilon$
$\varepsilon_{k(ij)}$	1	1	1	σ^2_ε	

```
Source      Expected Mean Square
Loc           Var(Error)+ 11
Var(block(loc))+13Var(loc*var)+143Var(loc)
Var           Var(Error)+ 13 Var(loc*var)+26Var(var)
loc*var       Var(Error)+ 13 Var(loc*var)
block(loc)    Var(Error)+ 11 Var(block(loc))
Error         Var(Error)
```

The genotypes need to be tested using the loc* var interaction variance that makes a lot of sense if we are trying to make a generalization valid to all locations!