ANOVA

APPLIED STATISTICS (STAT200)

TWO-FACTOR ANOVA WITH $K_{ij} = 1$

SECTION 11.4 (DEVORE & BERK 2012)

Example: Suppose we wish to study permeability of woven material used to construct automobile airbags, an experiment related to the ability to absorb energy. We perform an experiment using $I = 4$ temperature levels and $J = 3$ levels of fabric.

When factor A consists of I levels and factor B consists of J levels, there are IJ different combinations of levels of he two factors, each called a treatment.
We denote with K_{ij} the number of observations of the treatment consisting of factor A at level i and factor B at level j.

In this section we focus on the case $K_{ij} = 1$, so that the data consists of IJ observations. This setting is also termed two-factor ANOVA without replication.

NOTATION AND DEFINITIONS

We use double subscripts to identify random variables and observed values. Let:

- X_{ij}= the random variable denoting the measurement when factor A is held at level i and factor B is held at level j.
  x_{ij}= the observed value of X_{ij}. They are usually presented in a two-way table
- $\bar{X}_{i.} = \frac{\sum_{j=1}^{J} X_{ij}}{J}$ is the average of data obtained when factor A is held at level i, with observed values $\bar{x}_{i.}$
- $\bar{X}_{.j} = \frac{\sum_{i=1}^{I} X_{ij}}{I}$ is the average of data obtained when factor B is held at level j, with observed values $\bar{x}_{.j}$
- $\bar{X}_{..} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}}{IJ}$ is the grand mean, with observed value $\bar{x}_{..}$

Totals rather than averages are denoted without the horizontal bar. For example $x_{.j} = \sum_{i=1}^{I} x_{ij}$.

Example: two-way table

		Washing treatment				
		1	2	3	4	Total
Brand of Pen	1	.97	.48	.48	.46	2.39
	2	.77	.14	.22	.25	1.38
	3	.67	.39	.57	.19	1.82
	Total	2.41	1.01	1.27	.90	5.59

THE MODEL – FIRST TRY

We define μ_{ij} as the true average response when factor A is at level i and factor B at level j, resulting in IJ mean parameters. Then the model equation is

$$X_{ij} = \mu_{ij} + \varepsilon_{ij}$$

where ε_{ij} is the random amount by which the observed value differs from its expectation. All ε_{ij}'s are assumed normal and independent with common variance σ^2.

Note: there is no valid test procedure for this choice of parameters! Under the alternative hypothesis, the μ_{ij}'s can assume any value and σ^2 can be any positive value ⇒ There are $IJ + 1$ freely varying parameters, but only IJ observations. Thus: if we use x_{ij} to compute μ_{ij}, we have no way to estimate σ^2.

⇒ We must specify a model that is realistic but involving few(er) parameters.

THE MODEL – AN ADDITIVE MODEL

Assume the existence of I parameters $\alpha_1, \dots, \alpha_I$ and J parameters $\beta_1, \dots, \beta_J$ such that

$$X_{ij} = \alpha_i + \beta_j + \varepsilon_{ij}$$

for $i = 1, \dots, I; j = 1, \dots, J$. Hence $\mu_{ij} = \alpha_i + \beta_j$. Including σ^2, there are now $I + J + 1$ model parameters, so if $I \geq 3$, $J \geq 3$, there will be fewer parameters than observations.

This new model is called an additive model, because each mean response μ_{ij} is the sum of an effect due to factor *A* at level i (α_i) and an effect due to factor *B* at level j (β_j).

For an additive model holds: If we consider two levels i and i' of factor *A*, then

$$\mu_{ij} - \mu_{i'j} = (\alpha_i + \beta_j) - (\alpha_{i'} + \beta_j) = \alpha_i - \alpha_{i'}$$

which is independent of the level j for factor *B*. Analogously, $\mu_{ij} - \mu_{ij'} = \beta_j - \beta_{j'}$ for two levels j and j' of factor *B*.

THE MODEL – ADDITIVE VS NON-ADDITIVE MODELS

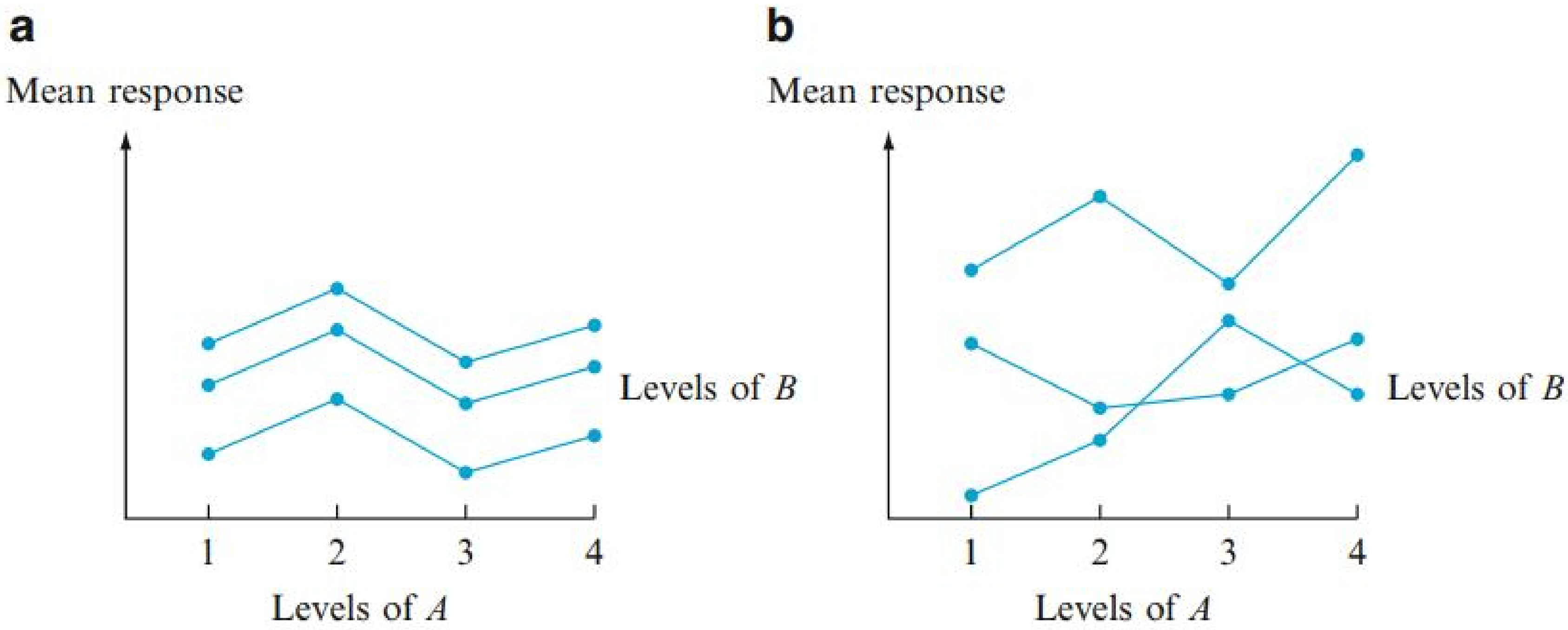


Figure 11.7 Mean responses for two types of model: (a) additive; (b) nonadditive

Observe that α_i and β_j are not uniquely determined!

	$\boldsymbol{\beta_1 = 1}$	$\boldsymbol{\beta_2 = 4}$
$\boldsymbol{\alpha_1 = 1}$	$\boldsymbol{\mu_{11} = 2}$	$\boldsymbol{\mu_{12} = 5}$
$\boldsymbol{\alpha_1 = 2}$	$\boldsymbol{\mu_{21} = 3}$	$\boldsymbol{\mu_{22} = 6}$

	$\boldsymbol{\beta_1 = 2}$	$\boldsymbol{\beta_2 = 5}$
$\boldsymbol{\alpha_1 = 0}$	$\boldsymbol{\mu_{11} = 2}$	$\boldsymbol{\mu_{12} = 5}$
$\boldsymbol{\alpha_1 = 1}$	$\boldsymbol{\mu_{21} = 3}$	$\boldsymbol{\mu_{22} = 6}$

Subtracting a value *c* from all α_i's and adding *c* to all β_j's results in the same μ_{ij}. We can fix this problem defining the final model

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

where $\sum_{i=1}^{I} \alpha_i = 0, \sum_{j=1}^{J} \beta_j = 0$, and the ε_{ij}'s are assumed independent, normally distributed, with mean 0 and common variance σ^2.

Interpretation: μ is the true grand mean, α_i is the effect of factor *A* at level i, and β_j is the effect of factor *B* at level j. Unbiased estimators for these parameters are $\hat{\mu} = \bar{x}_{..}$, $\hat{\alpha}_i = \bar{x}_{i.} - \bar{x}_{..}$, and $\hat{\beta}_j = \bar{x}_{.j} - \bar{x}_{..}$.

TWO HYPOTHESES

There are two different hypotheses of interest in a two-factor experiment with $K_{ij} = 1$. These are:

1. The first null hypothesis H_{0A} states that the different levels of factor A have no effect on true average response, i.e.
$$H_{0A}: \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0 \text{ versus } H_{aA}: \text{at least one } \alpha_i \neq 0$$
2. The second null hypothesis H_{0B} asserts that the different levels of factor B have no effect on true average response, i.e.
$$H_{0B}: \beta_1 = \beta_2 = \cdots = \beta_J = 0 \text{ versus } H_{aB}: \text{at least one } \beta_j \neq 0$$

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}^2 - \frac{1}{IJ}X_{..}^2, \text{df} = \text{IJ} - 1$$

$$SSA = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_{i.} - \bar{X}_{..})^2 = \frac{1}{J}\sum_{j=1}^{J} X_{i.}^2 - \frac{1}{IJ}X_{..}^2, df = I - 1$$

$$SSB = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{X}_{.j} - \bar{X}_{..})^2 = \frac{1}{J}\sum_{j=1}^{J} X_{.j}^2 - \frac{1}{IJ}X_{..}^2, df = J - 1$$

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J}(X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2, df = (I-1)(J-1)$$

And the fundamental identity becomes $SST = SSA + SSB + SSE$

TESTING PROCEDURES – F RATIOS AND HYPOTHESES

Forming *F* ratios as in single-factor ANOVA, we can show that:

1. if H_{0A} is true, the corresponding *F* ratio has an *F* distribution with numerator $df = (I - 1)$ and denominator $df = (I - 1)(J - 1)$;
2. an analogous result applies when testing H_{0B}.

Hypotheses	Test statistic Value	Rejection Region
H_{0A} versus H_{aA}	$f_A = \frac{MSA}{MSE}$	$f_A \geq F_{\alpha,I-1,(I-1)(J-1)}$
H_{0B} versus H_{aB}	$f_B = \frac{MSB}{MSE}$	$f_B \geq F_{\alpha,J-1,(I-1)(J-1)}$

The plausibility of using the *F* test can be demonstrated by determining the expected mean squares. After some manipulation, one obtains:

$$E(MSE) = \sigma^2$$

$$E(MSA) = \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I}\alpha_i^2$$

$$E(MSB) = \sigma^2 + \frac{I}{J-1}\sum_{j=1}^{J}\beta_j^2$$

Then follows:

- MSE always is an unbiased estimator of σ^2

TESTING PROCEDURES – EXPECTED MEAN SQUARES

- When H_{0A} is true, MSA is an unbiased estimator of σ^2.
  $\Rightarrow f_A$ is a ratio of two unbiased estimators of σ^2
- When H_{0A} is false, MSA tends to overestimate σ^2
  $\Rightarrow H_{0A}$ should be rejected when the observed ratio f_A is too large
- Similar comments apply to MSB and H_{0B}.
  When H_{0B} is true, MSB is an unbiased estimator of $\sigma^2 \Rightarrow f_B$ is a ratio of two unbiased estimators of σ^2.
  When H_{0B} is false, MSB tends to overestimate $\sigma^2 \Rightarrow H_{0B}$ should be rejected when the observed ratio f_B is too large

How can plausibility of the normality and constant variance assumptions be investigated?

1. Define the predicted values (also called fitted values)
$$\hat{x}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{x}_{..} + (\bar{x}_{i.} - \bar{x}_{..}) + \left(\bar{x}_{.j} - \bar{x}_{..}\right) = \bar{x}_{i.} + \bar{x}_{.j} + \bar{x}_{..}$$
and the residuals
$$\hat{\varepsilon}_{ij} = x_{ij} - \hat{x}_{ij} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$$
2. We can check the normality assumption by testing the residuals for normality. This applies to other ANOVA settings as well.
3. Moreover, a normal plot of residuals allows to visually assess normality.

4. We can easiest check the constant variance assumption visually by plotting the residuals against the fitted values.
We will see more sophisticated procedures in the linear regression setting.

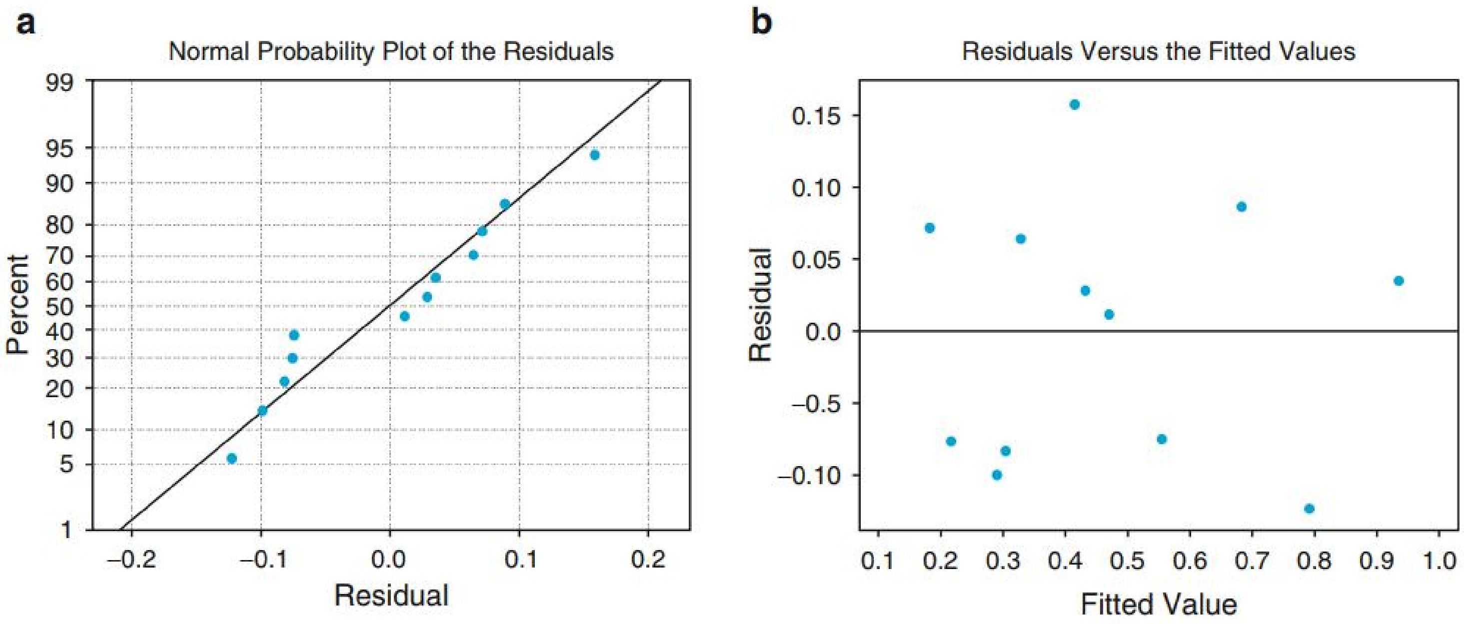


Figure 11.9 Plots from MINITAB for Example 11.13

When either H_{0A} and H_{0B} has been rejected, Tukey's procedure can be used to identify significant differences between the levels of the factor under investigation. Similar to the one-way setting, the principle idea is:

1. For comparing levels of factor *A*, obtain $Q_{\alpha,I,(I-1)(J-1)}$
   For comparing levels of factor *B*, obtain $Q_{\alpha,J,(I-1)(J-1)}$
2. Compute $w = Q$ (estimated sd of the sample means being compared)
   $$= \begin{cases} Q_{\alpha,I,(I-1)(J-1)} \cdot \sqrt{MSE/J} & \text{for factor A comparisons} \\ Q_{\alpha,J,(I-1)(J-1)} \cdot \sqrt{MSE/I} & \text{for factor B comparisons} \end{cases}$$
3. Arrange the sample means in increasing order, underscore those pairs differing by less than w. Pairs of sample means not underscored by the same line as corresponds to a pairs of population judged as significantly different

TWO-FACTOR ANOVA WITHOUT REPLIC. – IMPLEMENTATION IN R

As before, we focus on the simplest function for carrying out an ANOVA:

- Use the function `aov`
- One argument necessary: the model specification, which has a form involving two predictors in this setting

  `response ~ predictor1 + predictor2`
- `aov` creates an object of class "`aov`" for summarizing the results. Use the `summary` function to show the results

As before: if you are not sure how a test works, try a simple example.

`- script on screen -`

A particular setup of a two-way ANOVA without replication is the so-called randomized block experiment.

For such an experiment, assume we interested in carrying out a single-factor ANOVA. That is we want to test for the presence of effects due to the I different treatments under study. Then we proceed as follows:

1. Choose the IJ subjects or experimental units
2. Allocate the treatments in a random fashion:
   - Choose J subjects at random for the first treatment
   - Choose J subjects at random for the second treatment
   - And so on until we reach the last treatment

If one proceeds like this, it may happen that subjects exhibit differences with respect to other characteristics that may affect the observed response.

Example: some patients may be healthier than others. If all healthy patients are coincidentally assigned to first treatment, and all sick patients to another treatment, this could easily bias the results severely.

In general, the presence or absence of a significant F-value may be due to differences related to other characteristics rather than to the presence or absence of factor effects!
That is also the reason why we introduced the paired t-test previously.

One generalization of the paired experiment to $I > 2$ is called a randomized block experiment. This experiment is designed as follows:

- We construct an extraneous factor, blocks, by dividing the IJ units into J groups with I units in each group
- This "blocking" is done in such a way that within each block, the I units are homogeneous with respect to other factors thought to affect the response
- Within each homogeneous block, the I treatments are randomly assigned to the I subjects in the block

RANDOMIZED BLOCK EXPERIMENTS - EXAMPLE

Example: A consumer product-testing organization wishes to compare the annual power consumption for five different brands of dehumidifier. Thus, "brand" is the factor of interest.

Power consumption depends on the humidity level. Therefore, one monitors products from each brand at four different humidity levels (moderate to heavy). Thus, we are blocking the observations on humidity level, because we are in principle not interested in showing a "humidity effect".

In total, 20 dehumidifiers are distributed. Within each humidity level, brands are randomly assigned to five selected locations (with similar humidity).

RANDOMIZED BLOCK EXPERIMENTS - EXAMPLE

Table 11.6 Power consumption data for Example 11.15

	Blocks (humidity level)					
Treatments (brands)	**1**	**2**	**3**	**4**	$x_{i\cdot}$	$\bar{x}_{i\cdot}$
1	685	792	838	875	3190	797.50
2	722	806	893	953	3374	843.50
3	733	802	880	941	3356	839.00
4	811	888	952	1005	3656	914.00
5	828	920	978	1023	3749	937.25
$x_{\cdot j}$	3779	4208	4541	4797	17,325	

Table 11.7 ANOVA table for Example 11.15

Source of Variation	df	Sum of Squares	Mean Square	f
Treatments (brands)	4	53,231.00	13,307.75	$f_A = 95.57$
Blocks	3	116,217.75	38,739.25	$f_B = 278.20$
Error	12	1671.00	139.25	
Total	19	171,119.75		

RANDOMIZED BLOCK EXPERIMENT – IMPLEMENTATION IN R

Nothing changes here: ANOVA with two factors, now named blocks and treatments – but R does not care how you name factors.

Hence, simply use the same command which you saw just before for the two-factor ANOVA without replication.

TWO-FACTOR ANOVA WITH $K_{ij} > 1$

SECTION 11.5 (DEVORE & BERK 2012)

INTERACTION

In previous section, we assumed to have an additive structure with

$$\mu_{ij} = \mu + \alpha_i + \beta_j, \ \ \sum_i \alpha_i = \sum_j \beta_j = 0.$$

This grants, for example, that $\mu_{ij} - \mu_{ij'}$ is independent of the level i of the first factor.

When additivity does not hold, we say that there is interaction between the different levels of the factors.

Example: the effect of changing factor B from level 2 to level 4 varies when the level of factor A is fixed at 1 or 2, respectively. More precisely,

$$\mu_{12} - \mu_{14} = 3 \text{ and } \ \mu_{22} - \mu_{24} = -5$$

- Recall: in the setting with $K_{ij} = 1$ treated previously, the assumption of additivity allowed us to obtain an estimator of σ^2 that was unbiased whether or not H_0 is true
- When $K_{ij} > 1$ for at least one (i, j) pair, a valid estimator of σ^2 can be obtained without assuming additivity
- In specifying the appropriate model and deriving test procedures, we will focus on the case $K_{ij} = K > 1$, so the number of observations per "cell" (for each combination of levels) is constant
- Extending the setting to varying numbers of observations per "cell" is straightforward, but slightly complicates the notation

INTERACTION – PARAMETERS

We introduce a new set of parameters, revealing the role of interaction more clearly:

$$\mu = \frac{1}{IJ}\sum_i \sum_j \mu_{ij}, \qquad \bar{\mu}_{i.} = \frac{1}{J}\sum_j \mu_{ij}, \qquad \bar{\mu}_{.j} = \frac{1}{I}\sum_i \mu_{ij}$$

We also define

1. As before, the effect of factor A at level i: $\alpha_i = \bar{\mu}_{i.} - \mu$
2. As before, the effect of factor B at level j: $\beta_j = \bar{\mu}_{.j} - \mu$
3. New effect:

$$\gamma_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$$

from which follows

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

Note: The model is additive if and only if all γ_{ij}'s $= 0$.

- The γ_{ij}'s are referred to as the interaction parameters
- The α_i's (β_j's) are called the main effect for factor *A* (main effects for factor *B*)

Although there are I α_i's, J β_j's, and IJ γ_{ij}'s in addition to μ, the conditions

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_j \gamma_{ij} = \sum_i \gamma_{ij} = 0$$

imply that only IJ of these new parameters are independently determined:

1. μ,
2. $I - 1$ of the α_i's,
3. $J - 1$ of the β_j's, and
4. $(I - 1)(J - 1)$ of the γ_{ij}'s.

In this new setting with interaction, there are three sets of hypotheses that will be considered:

$$H_{0A}: \alpha_1 = \cdots = \alpha_I = 0 \text{ versus } H_{aA}: \text{at least one } \alpha_i \neq 0$$

$$H_{0B}: \beta_1 = \cdots = \beta_J = 0 \text{ versus } H_{aB}: \text{at least one } \beta_j \neq 0$$

$$H_{0AB}: \gamma_{ij} = 0 \text{ for all } i, j \text{ versus } H_{aAB}: \text{at least one } \gamma_{ij} \neq 0$$

- The no-interaction hypothesis H_{0AB} is usually tested first. If H_{0AB} is not rejected, then the other two hypotheses can be tested to see whether the main effects are significant

INTERACTION – HYPOTHESES

- However: once H_{0AB} is rejected, we believe that the effect of factor *A* at any particular level depends on the level of *B* (and vice versa). It then does not make sense to test H_{0A} and H_{0B} (in general)
- In case of interaction, it may be appropriate to carry out one-way ANOVAs to compare levels of *A* separately for each level of *B* as part of the "post-hoc testing"

Example: Factor *A* involves four kinds of glue, Factor *B* involves three types of material. The response is strength of the glue joint. Assume that the strength of the glues depend on which material is being glued, i.e. significant interaction is present.
Then it makes sense to carry out three separate one-way ANOVA analyses, one for each material.

We have to use triple subscripts for both random variables and observed values, with X_{ijk} and x_{ijk} referring to the k^{th} observation when factor A is at level i and factor B is at level j.

The model equation is then given by

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

with $i = 1, \dots, I;\ j = 1, \dots, J;\ k = 1, \dots, K$. As usually:

- the ε_{ijk}'s are independent and normally distributed, each with mean 0 and constant variance σ^2
- a dot in place of a subscript means that we have summed over all values of that subscript. A horizontal bar denotes averaging

TESTING PROCEDURES – RELEVANT FORMULAS

$$SST = \sum_i \sum_j \sum_k \left(X_{ijk} - \bar{X}_{...}\right)^2, \qquad df = IJK - 1$$

$$SSA = \sum_i \sum_j \sum_k (\bar{X}_{i..} - \bar{X}_{...})^2, \qquad df = I - 1$$

$$SSB = \sum_i \sum_j \sum_k \left(\bar{X}_{.j.} - \bar{X}_{...}\right)^2, \qquad df = J - 1$$

$$SSE = \sum_i \sum_j \sum_k \left(X_{ijk} - \bar{X}_{ij.}\right)^2, \qquad df = IJ(K - 1)$$

$$SSAB = \sum_i \sum_j \sum_k \left(\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}\right)^2, \qquad df = (I - 1)(J - 1)$$

The fundamental identity becomes $SST = SSA + SSB + SSAB + SSE$, implying that the interaction sum of squares *SSAB* can be obtained by subtraction.

TESTING PROCEDURES – F RATIOS AND HYPOTHESES

Each of the three mean square ratios can be shown to have an *F* distribution when the associated H_0 is true. This yields the following level α test procedures.

Hypotheses	Test Statistic Value	Rejection Region
H_{0A} versus H_{aA}	$f_A = \dfrac{MSA}{MSE}$	$f_A \geq F_{\alpha,I-1,IJ(K-1)}$
H_{0B} versus H_{aB}	$f_B = \dfrac{MSB}{MSE}$	$f_B \geq F_{\alpha,J-1,IJ(K-1)}$
H_{0AB} versus H_{aAB}	$f_{AB} = \dfrac{MSAB}{MSE}$	$f_{AB} \geq F_{\alpha,(I-1)(J-1),IJ(K-1)}$

The plausibility of using the *F* test can again be demonstrated by determining the expected mean squares. After some manipulation, one obtains:

$$E(MSE) = \sigma^2$$

$$E(MSA) = \sigma^2 + \frac{JK}{I-1}\sum_{i=1}^{I}\alpha_i^2$$

$$E(MSB) = \sigma^2 + \frac{IK}{J-1}\sum_{j=1}^{J}\beta_j^2$$

$$E(MSAB) = \sigma^2 + \frac{K}{(I-1)(J-1)}\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_{ij}^2$$

The expected mean squares suggest that each set of hypotheses can be tested using the appropriate ratio of mean squares with MSE denominator.

TESTING PROCEDURES – EXPECTED MEAN SQUARES

- MSE always is an unbiased estimator of σ^2
- When H_{0A} (H_{0B} / H_{0AB}) is true, MSA (MSB / MSAB) is an unbiased estimator of σ^2.
  $\Rightarrow$ f_A (f_B / f_{AB}) is a ratio of two unbiased estimators of σ^2
- When H_{0A} (H_{0B} / H_{0AB}) is false, MSA (MSB / MSAB) tends to overestimate σ^2.
  $\Rightarrow$ H_{0A} (H_{0B} / H_{0AB}) should be rejected when the observed ratiof_A (f_B / f_{AB}) is too large

TESTING PROCEDURES – ASSUMPTIONS, POST-HOC

- For investigating normality, check the residuals
  1. graphically (QQ-plot)
  2. formally with a test (e.g. Shapiro-Wilk)
- For investigating homoscedasticity, check the residuals
  1. graphically (plot against predicted values)
  2. formally with a test (Levene's test)
- Post-hoc testing: Tukey's HSD

EXAMPLE – YIELD DATA

Example 11.16 from Devore & Berk: the setting considered is characterized by

- three different varieties of tomato (Harvester, Ife No. 1, and Pusa Early Dwarf) and
- four different plant densities (10, 20, 30, and 40 thousand plants per hectare)

are considered for planting in a particular region. Questions to be investigated are

1. if either variety or plant density affects yield and
2. if specific combinations of variety and plant density affects yield

Each combination of variety and plant density is used in three different plots.

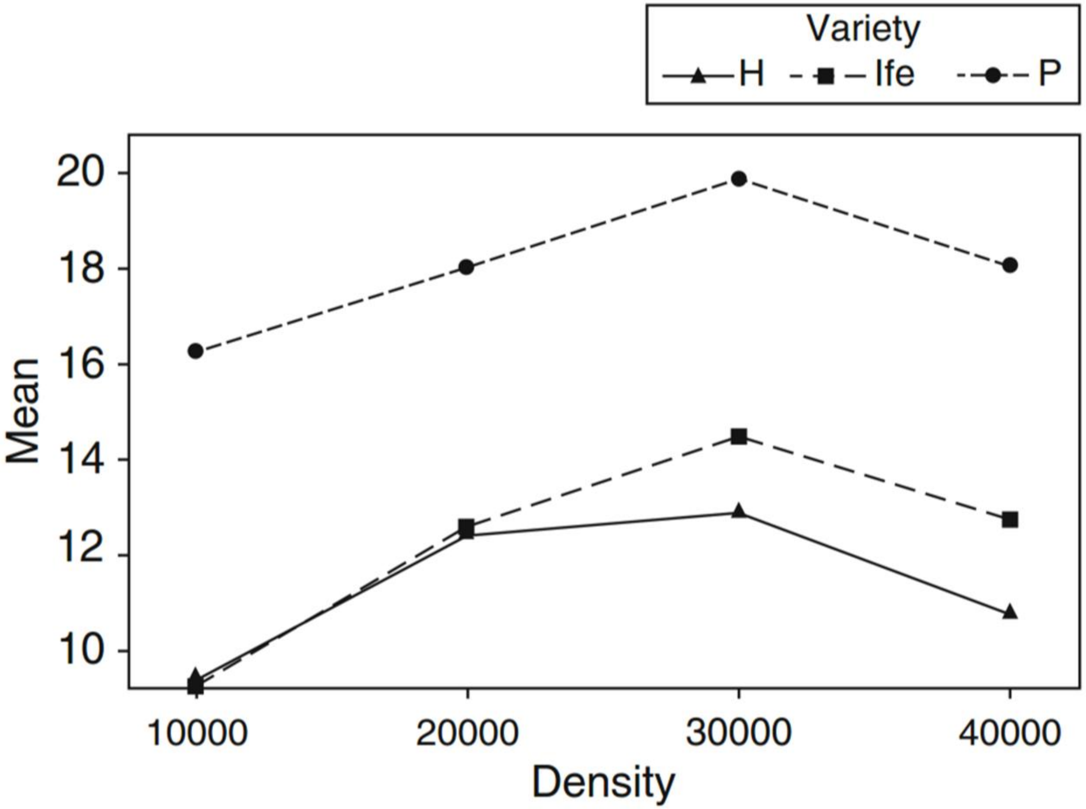
EXAMPLE – YIELD DATA

Data in tabular form

Table 11.8 Yield data for Example 11.16

	Planting Density													
Variety	**10,000**			**20,000**			**30,000**			**40,000**			$x_{i\cdot\cdot}$	$\bar{x}_{i\cdot\cdot}$
H	10.5	9.2	7.9	12.8	11.2	13.3	12.1	12.6	14.0	10.8	9.1	12.5	136.0	11.33
Ife	8.1	8.6	10.1	12.7	13.7	11.5	14.4	15.4	13.7	11.3	12.5	14.5	146.5	12.21
P	16.1	15.3	17.5	16.6	19.2	18.5	20.8	18.0	21.0	18.4	18.9	17.2	217.5	18.13
$x_{\cdot j \cdot}$	103.3			129.5			142.0			125.2			500.00	
$\bar{x}_{\cdot j \cdot}$	11.48			14.39			15.78			13.91				13.89

Interaction plot



Results

Table 11.9 ANOVA table for Example 11.17

Source of Variation	df	Sum of Squares	Mean Square	f
Varieties	2	327.60	163.8	$f_A = 103.02$
Density	3	86.69	28.9	$f_B = 18.18$
Interaction	6	8.03	1.34	$f_{AB} = .84$
Error	24	38.04	1.59	
Total	35	460.36		

TWO-FACTOR ANOVA WIT REPLICATIONS – IMPLEMENTATION IN R

- Use the function `aov` (simplest approach)
- One argument necessary: the model specification, which has a form involving two predictors and an interaction term in this setting

```
response ~ pred1 + pred2 + pred1 : pred2
```

or simpler

```
response ~ pred1 * pred2
```

- `aov` creates an object of class "`aov`", use `summary` for showing the results

Example: `- script on screen -`