# Linear Models: The less than full rank model — inference

# Recap

By now you should be familiar with the less than full rank model. To recap, it is:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where the errors $\boldsymbol{\varepsilon}$ have mean $\mathbf{0}$, variance $\sigma^2 I$, and (sometimes) are assumed to be jointly normally distributed.

In a less than full rank model, $r(X) < p$, and this means that $X^T X$ is singular. In turn, this means that not all quantities associated with the model can be estimated.

# Testability

The quantities that can be estimated are called estimable. In this section, we wish to develop procedures for testing hypotheses for the less than full rank model.

As you might expect, not all hypotheses can be tested. In general, if you cannot estimate the value of something, it is difficult to test any hypotheses about it!

Hypotheses which can be tested are called *testable*. This is defined as follows.

### Definition

A hypothesis $H_0$ is testable if there exists a set of estimable functions $\mathbf{c}_1^T\boldsymbol{\beta}, \mathbf{c}_2^T\boldsymbol{\beta}, \ldots, \mathbf{c}_m^T\boldsymbol{\beta}$ such that $H_0$ is true if and only if

$$\mathbf{c}_1^T\boldsymbol{\beta} = \mathbf{c}_2^T\boldsymbol{\beta} = \ldots = \mathbf{c}_m^T\boldsymbol{\beta} = \mathbf{0}$$

and $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m$ are linearly independent.

Note that the hypothesis itself does not have to explicitly be of this form, it merely needs to be equivalent to it.

How do we determine if a hypothesis is testable?

We are only looking at linear hypothesis (nonlinear hypotheses are outside the scope of this course), so in general we can write any hypothesis as

$$H_0 : C\boldsymbol{\beta} = \mathbf{0}$$

where $C$ is of dimension $n \times p$.

Suppose that $C$ is of rank $m$. We take $m$ linearly independent rows of $C$ as $\mathbf{c}_1^T, \mathbf{c}_2^T, \ldots, \mathbf{c}_m^T$.

Because the rest of the rows of $C$ are dependent on these rows, $H_0$ is equivalent to

$$\mathbf{c}_1^T \boldsymbol{\beta} = \mathbf{c}_2^T \boldsymbol{\beta} = \ldots = \mathbf{c}_m^T \boldsymbol{\beta} = \mathbf{0}.$$

It can be shown that the maximum number of linearly independent estimable functions in a less than full rank model is $r(X) = r$. So we need $m \leq r$.

Now recall that a linear function $\mathbf{c}^T\beta$ is estimable if and only if

$$\mathbf{c}^T(X^TX)^cX^TX = \mathbf{c}^T.$$

This leads us to conclude that $H_0$ is testable if and only if

$$C(X^TX)^cX^TX = C.$$

**Example.** Consider the one-way classification model with fixed effects and $k = 3$. The linear model that we use is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}.$$

Suppose we want to test the hypothesis that the means of all three populations are equal. This is equivalent to $H_0 : \tau_1 = \tau_2 = \tau_3$. However, this hypothesis is true if and only if

$$\tau_1 - \tau_2 = 0$$

and

$$\tau_2 - \tau_3 = 0.$$

So we can express this hypothesis as $H_0 : C\boldsymbol{\beta} = \mathbf{0}$ where

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}.$$

Now $\tau_1 - \tau_2$ is a contrast, so it is known to be estimable. Similarly, $\tau_2 - \tau_3$ is estimable. The rows of $C$ are obviously linearly independent, so $H_0$ is testable.

Once we have determined that a hypothesis is testable, how can we test it?

We look back at the full rank case. In this case, the hypothesis $H_0 : C\boldsymbol{\beta} = \mathbf{0}$ is a special case of the general linear hypothesis $C\boldsymbol{\beta} = \boldsymbol{\delta}^*$, with $\boldsymbol{\delta}^* = 0$. The $F$ statistic that we would use to test this is

$$\frac{(C\mathbf{b})^T[C(X^TX)^{-1}C^T]^{-1}C\mathbf{b}/r}{SS_{Res}/(n-p)},$$

which under the null hypothesis has an $F$ distribution with $r$ and $n-p$ degrees of freedom, where $r = r(C)$.

Obviously, we cannot do this in the less than full rank case, because $X^T X$ does not have an inverse. However, we can use a conditional inverse.

Because we now use $r = r(X)$, we replace $r$ by $r(C) = m$.

The other change is that $\frac{SS_{Res}}{n-p}$ is no longer the estimator of the variance, $s^2$. We change this to the estimator, $s^2 = \frac{SS_{Res}}{n-r}$.

Therefore our proposed statistic for testing this hypothesis in the less than full rank model is

$$\frac{(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b}/m}{s^2},$$

which under the null hypothesis should follow an $F$ distribution with $m$ and $n-r$ degrees of freedom.

The following theorems show that this does indeed work.

## Theorem

*Let $\mathbf{y} = X\boldsymbol{\beta} + \varepsilon$ be a linear model $X$ is $n \times p$, $r(X) < p$, and $\varepsilon$ is a normal random vector with mean $\mathbf{0}$ and variance $\sigma^2 I$. Suppose that $C\boldsymbol{\beta} = \mathbf{0}$ is testable, where $C$ is $m \times p$ with rank $m \leq r$. Then*

$$\frac{(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b}}{\sigma^2}$$

*has a noncentral $\chi^2$ distribution with $m$ degrees of freedom and noncentrality parameter*

$$\lambda = \frac{(C\boldsymbol{\beta})^T[C(X^TX)^cC^T]^{-1}C\boldsymbol{\beta}}{2\sigma^2}.$$

**Proof.** Since the hypothesis is testable, $C\mathbf{b}$ is not dependent on the conditional inverse that we use to calculate $\mathbf{b}$. Since

$$C\mathbf{b} = C(X^TX)^cX^T\mathbf{y},$$

we can apply expectation and variance rules to determine that $C\mathbf{b}$ is a normal random vector with mean

$$C(X^TX)^cX^TX\boldsymbol{\beta} = C\boldsymbol{\beta}$$

and variance

$$C(X^TX)^cX^T\sigma^2IX(X^TX)^cC^T = C(X^TX)^cC^T\sigma^2.$$

We can then apply a well-known corollary for the distribution of a quadratic form to derive the theorem.

This quadratic form follows a similar pattern to the full rank model: if the null hypothesis is true and $C\boldsymbol{\beta} = \mathbf{0}$, then it has an ordinary $\chi^2$ distribution. However, if the null hypothesis is false, we would expect it to be larger than such a distribution, so we would use a right-tailed test.

What about the denominator? We have already shown that $\frac{s^2}{\sigma^2}$ has a $\chi^2$ distribution with $n - r$ degrees of freedom.

The only thing we need to complete our theory is that the numerator and denominator correspond to independent $\chi^2$ distributions.

## Theorem

*Let $\mathbf{y} = X\beta + \varepsilon$ be a less than full rank linear model, where $\varepsilon$ is a normal random vector with mean $\mathbf{0}$ and variance $\sigma^2 I$. Suppose that $C\beta = \mathbf{0}$ is testable. Then $C\mathbf{b}$ is independent of $s^2$.*

**Proof.** We use the result for independence between a random vector and a quadratic form. We know that

$$C\mathbf{b} = C(X^T X)^c X^T \mathbf{y}.$$

We can also write

$$SS_{Res} = \mathbf{y}^T[I - X(X^TX)^cX^T]\mathbf{y}.$$

Since

$$\begin{aligned}
BVA &= C(X^TX)^cX^T\sigma^2 I[I - X(X^TX)^cX^T] \\
&= [C(X^TX)^cX^T - C(X^TX)^cX^TX(X^TX)^cX^T]\sigma^2 \\
&= [C(X^TX)^cX^T - C(X^TX)^cX^T]\sigma^2 \\
&= 0,
\end{aligned}$$

$C\mathbf{b}$ is independent of $SS_{Res}$, and hence of $s^2$.

**Example.** Let us look at the carbon removal example from the previous section. To recap, there are three methods of removing carbon from wastewater which we wish to compare. The data is:

| AF | FS | FCC |
|---|---|---|
| 34.6 | 38.8 | 26.7 |
| 35.1 | 39.0 | 26.7 |
| 35.3 | 40.1 | 27.0 |

We wish to test whether the populations have the same mean. In other words, we want to test $H_0 : \tau_1 = \tau_2 = \tau_3$. This can be written in matrix form as $H_0 : C\boldsymbol{\beta} = \mathbf{0}$, where

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}.$$

We have previously calculated

$$(X^T X)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 0 \\ 35 \\ 39.3 \\ 26.8 \end{bmatrix}.$$

Therefore

$$
\begin{aligned}
C(X^TX)^cC^T &= \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} \end{bmatrix} \\
&= \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}
\end{aligned}
$$

and

$$
\begin{aligned}
[C(X^TX)^cC^T]^{-1} &= \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \\
C\mathbf{b} &= \begin{bmatrix} -4.3 \\ 8.2 \end{bmatrix} \\
(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b} &= \begin{bmatrix} -4.3 & 8.2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -4.3 \\ 8.2 \end{bmatrix} \\
&= 241.98.
\end{aligned}
$$

We have previously calculated $s^2 = 0.217$. Therefore our $F$ statistic, with 2 and $9 - 3 = 7$ degrees of freedom, is

$$\frac{241.98/2}{0.217} = 558.41.$$

As the critical value for this distribution at the 0.01 level is 9.55, we can reject $H_0$ quite firmly! Thus the populations are not all the same. It still is possible that *some* of the populations are the same, but not all of them.

# Treatment contrasts

In a one-way classification model, the hypothesis that we will most often want to test is a treatment contrast; in other words, we want to test

$$H_0 : \sum_{i=1}^{k} a_i \tau_i = 0$$

where $\sum_i a_i = 0$. We have seen that the left-hand side of this hypothesis is estimable, and so this hypothesis is testable.

The most common example of a treatment contrast test is

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k,$$

which tests whether all populations have equal means. If they do, then the distributional assumptions on the errors imply that the response variable is identically distributed amongst all populations.

We can write this test as a multiple hypothesis on treatment contrasts:

$$H_0 : \tau_1 - \tau_2 = 0, \tau_1 - \tau_3 = 0, \ldots, \tau_1 - \tau_k = 0.$$

A treatment contrast hypothesis is easily tested using machinery developed in the previous section. However, the formula can be simplified because of the special form of a treatment contrast.

In particular, we can write $X^T X$ as

$$X^T X = \begin{bmatrix} n & n_1 & n_2 & \dots & n_k \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ n_k & 0 & 0 & \dots & n_k \end{bmatrix}$$

It is easy to check that $r(X) = k$, and in particular

$$(X^TX)^c = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_1} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_k} \end{bmatrix}.$$

Also we have

$$X^T\mathbf{y} = \begin{bmatrix} \sum_{ij} y_{ij} \\ \sum_j y_{1j} \\ \sum_j y_{2j} \\ \vdots \\ \sum_j y_{kj} \end{bmatrix}.$$

Multiplying out gives us

$$\mathbf{b} = (X^T X)^c X^T \mathbf{y} = \begin{bmatrix} 0 \\ \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix}.$$

This means that if we know or are given $s^2$, the only other information that we need to test treatment contrast hypotheses is the means, and number, of the samples from the various populations. We do not need the samples themselves!

**Example.** A tennis ball manufacturer is studying the life span of a newly developed tennis ball on five different surface. The response is the number of hours that the ball is used before it is judged to be dead. A study is conducted and the following data obtained:

| Surface | Clay | Grass | Composition | Wood | Asphalt |
|---|---|---|---|---|---|
| Mean | 6.2 | 6.8 | 6.4 | 5 | 4.4 |
| Number | 20 | 22 | 24 | 21 | 25 |

We are also given $s^2 = 8.87$.

Suppose we want to test if the lifespan of the ball is different on hard surfaces (wood and asphalt) vs. soft surfaces. This contrast gives the hypothesis

$$H_0 : \frac{1}{3}(\tau_1 + \tau_2 + \tau_3) - \frac{1}{2}(\tau_4 + \tau_5) = 0$$

with corresponding matrix

$$C = \left[ \begin{array}{cccccc} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{2} & -\frac{1}{2} \end{array} \right].$$

Now

$$
\begin{aligned}
C\mathbf{b} &= \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ 6.2 \\ 6.8 \\ 6.4 \\ 5 \\ 4.4 \end{bmatrix} \\
&= 1.77 \\
(X^T X)^c &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{20} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{22} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{21} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{25} \end{bmatrix}
\end{aligned}
$$

Putting it together gives

$$[C(X^TX)^cC^T]^{-1} = 26.92$$

and the $F$ statistic

$$\frac{(1.77 \times 26.92 \times 1.77)/1}{8.87} = 9.47.$$

Since the critical value for an $F$ distribution with 1 and $112 - 5 = 107$ degrees of freedom is 6.88 at the 0.01 level, we can reject the null hypothesis that the ball lasts as long on hard and soft surfaces.

# Two-factor models

Up to now we have given examples which are one-way classification models, i.e. we only worry about the levels of one factor.

However, the theory we have developed is not restricted to such models. As long as we can express our model in a linear model form, we can apply the theory.

In this section, we look at two-factor models, but the ideas can be easily implemented for any number of factors.

In a basic two-factor model, we assume that each level of each factor affects the overall mean $\mu$ by a fixed amount. If we name these effects $\tau_i$ for the $i$th level of factor 1 and $\beta_j$ for the $j$th level of factor 2, then our model is

$$y_{ijk} = \mu+\tau_i+\beta_j+\varepsilon_{ijk}, \qquad i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, n_{ij}.$$

In matrix form (with 1 sample from each combination of factor levels):

$$
X = \left[\begin{array}{c|cccc|cccc}
1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\
1 & 1 & 0 & \dots & 0 & 0 & 1 & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots \\
1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\
\hline
1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\
1 & 0 & 1 & \dots & 0 & 0 & 1 & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots \\
1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\
\hline
\vdots & & & & & & & & \\
\hline
1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\
1 & 0 & 0 & \dots & 1 & 0 & 1 & & 0 \\
\vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots \\
1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1
\end{array}\right], \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_a \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_b \end{bmatrix}
$$

This model is known as the *additive* model. This is because it assumes that the effects from each factor level can be added up to produce the overall effect.

Many of the hypotheses that we can test in a one-factor model can be tested for each factor in an additive two-factor model.

**Theorem**

*In an additive two-factor model, every contrast in the $\tau$'s is estimable.*

Similarly, every contrast in the $\beta$'s is also estimable/testable.

The most common hypotheses that we will want to test is

$$\tau_1 = \tau_2 = \ldots = \tau_a$$

and

$$\beta_1 = \beta_2 = \ldots = \beta_b.$$

Because they are all composed of treatment contrasts for one factor, they are testable. We can use the theory already developed to test them.

**Example.** We want to determine the time taken to dissolve a capsule in a biological fluid. A study is conducted with 1 sample from each combination of factor levels and the following data found:

| | | Fluid type | |
|---|---|---|---|
| | | Gastric | Duodenal |
| Capsule | A | 39.5 | 31.2 |
| type | B | 47.4 | 44 |

The linear model is

$$\mathbf{y} = \begin{bmatrix} 39.5 \\ 47.4 \\ 31.2 \\ 44 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

We want to test the hypotheses that there is no difference in the response for the levels of each of the two factors.

The first factor (fluid type) gives the hypothesis

$$H_0 : \tau_1 = \tau_2 \text{ or } \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix} \boldsymbol{\beta} = \mathbf{0}.$$

This gives the $C$ matrix, and calculations give

$$\frac{(C\mathbf{b})^T [C(X^TX)^c C^T]^{-1} C\mathbf{b}/1}{s^2} = \frac{34.22}{6} = 5.7.$$

Since the critical value at 95% is 161.45 (we only have 1 and 1 degrees of freedom!), we cannot reject the null hypothesis. The capsule type hypothesis is tested in a similar way.

# Interaction

In some cases, it is possible that *interaction* between factors may occur.

Interaction happens when the level of one factor affects the effect of the levels of another factor.

So, for example, if the effect of factor 1 when factor 2 is at level 1 is different from the effect of factor 1 when factor 2 is at level 2, then there is interaction.

**Example.** Suppose that we are studying the effect of pressure and temperature on viscosity, and the *actual* means of the response variable for each of the combinations are given by

| | | Pressure | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Temperature | 1 | 4 | 6 | 4 | 3 |
| | 2 | 8 | 2 | 7 | 5 |

When the pressure is at level 1, changing the temperature from level 1 to level 2 results in an increase of viscocity of 4.

However, when the pressure is at level 2, changing the temperature from level 1 to level 2 results in a *decrease* of viscosity of 4!

In this case, the factors interact.

If, on the other hand, the actual means were

| | | Pressure | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Temperature | 1 | 4 | 6 | 4 | 3 |
| | 2 | 8 | 10 | 8 | 7 |

then there would be no interaction between the factors. Even though the factors themselves are significant, the *combination* of factor levels has no effect apart from the individual factor effects.

An additive model assumes that there is no interaction between the factors, so the effects of the factor levels can be measured in isolation from the other factor(s).

If we have interaction, or want to test whether there is interaction, we must use a different model.

We add an interaction term for each combination of factor levels.

Our model then becomes

$$y_{ijk} = \mu + \tau_i + \beta_j + \xi_{ij} + \varepsilon_{ijk},$$

where $\xi_{ij}$ is an interaction term which quantifies the effect of factor 1 being at level $i$ at the same time that factor 2 is at level $j$.

**Example.** Consider the previous example. If we allow an interaction term, $\mathbf{y}$ stays the same, but the linear model becomes

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \xi_{11} \\ \xi_{12} \\ \xi_{21} \\ \xi_{22} \end{bmatrix}.$$

In a two-factor model with interaction, we are often interested in testing whether there is interaction or not.

However, testing the presence of interaction is not quite as straightforward as it may seem. It seems like we would want to test the hypothesis

$$H_0 : \xi_{11} = \xi_{12} = \ldots = \xi_{1b} = \xi_{21} = \ldots = \xi_{ab},$$

but it turns out that this is not correct. We illustrate with an example.

**Example.** Suppose we have a two-factor model with the following actual means:

| | | Factor I | |
|---|---|---|---|
| | | 1 | 2 |
| Factor II | 1 | 6 | 5 |
| | 2 | 6 | 5 |

There is clearly no interaction between the factors.

One possible parameter set is

$$\mu = 0, \tau_1 = 5, \tau_2 = 4, \beta_1 = \beta_2 = 1, \xi_{ij} = 0 \qquad \forall i, j.$$

However, an equally valid parameter set is

$$\mu = 0, \tau_1 = 2, \tau_2 = 1, \beta_1 = 3, \beta_2 = 2,$$
$$\xi_{11} = 1, \xi_{12} = 2, \xi_{21} = 1, \xi_{22} = 2.$$

Therefore, testing that the interaction terms are all equal is not sufficient.

The following theorem tells us how to test for no interaction.

### Theorem

*For the linear model*

$$y_{ijk} = \mu + \tau_i + \beta_j + \xi_{ij} + \varepsilon_{ijk},$$

*there is no interaction if and only if*

$$(\xi_{ij} - \xi_{ij'}) - (\xi_{i'j} - \xi_{i'j'}) = 0$$

*for all* $i \neq i', j \neq j'$.

This theorem generates $ab(a-1)(b-1)$ equations. However, it can be shown that all but $(a-1)(b-1)$ of them are redundant.

**Example.** In a two-factor design with two levels in each factor, the theorem shows that there is no interaction if and only if

$$\begin{aligned}
(\xi_{11}-\xi_{12})-(\xi_{21}-\xi_{22}) &= 0\\
(\xi_{21}-\xi_{22})-(\xi_{11}-\xi_{12}) &= 0\\
(\xi_{12}-\xi_{11})-(\xi_{22}-\xi_{21}) &= 0\\
(\xi_{12}-\xi_{21})-(\xi_{12}-\xi_{11}) &= 0
\end{aligned}$$

It is easy to see that all of these equations are equivalent, so we need only test one (the first one, say). This gives the hypothesis $H_0 : C\boldsymbol{\beta} = \mathbf{0}$, where

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \xi_{11} \\ \xi_{12} \\ \xi_{21} \\ \xi_{22} \end{bmatrix}.$$

# Interaction considera(c)tion

Some things to consider when testing for interaction:

If we have one sample per combination of factors, it is impossible to account for or test for interaction.

This can be seen by noting that $r(X) = n$ and therefore $n - r$, the second number of degrees of freedom of our $F$ statistic, is 0.

This is because we are essentially treating each combination of factors as a separate population. If we have one sample from each population, then we have no way to estimate the variance!

Even if we test for interaction and find that there is none, we theoretically should still use the residual sum of squares from the full model with interaction, unless there is a convincing data-related reason to think that there is no interaction.

This follows from the same reasoning as using $SS_{Res}$ for the full model in sequential tests: we cannot be sure that there is no interaction, we just haven't found any!

However, for practical purposes, this may take away too many degrees of freedom from $SS_{Res}$. So if you find no interaction, it's OK to use an additive model.

Interaction can exist in models with three or more factors. Certainly we would be interested in interaction between two factors.

Technically, it is also possible to have interaction between three or more factors!

However, this is hard to test for. R does do it, but you can probably get away with looking at only two-factor interaction for now.

It is possible to to do analysis of covariance (ANCOVA) using the framework that we have built up. We simply have one (or more) categorical predictors and one (or more) continuous predictors. For example:

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}.$$

Then we apply the less than full rank model results.

This model is an additive model, which means that it assumes no interaction. Interaction in this case means that the slopes of the regression lines are different for each population, whereas this model assumes that the slopes are the same (but the intercepts may be different).