

Linear and Generalized Linear Models (4433LGLM6Y)

ANOVA

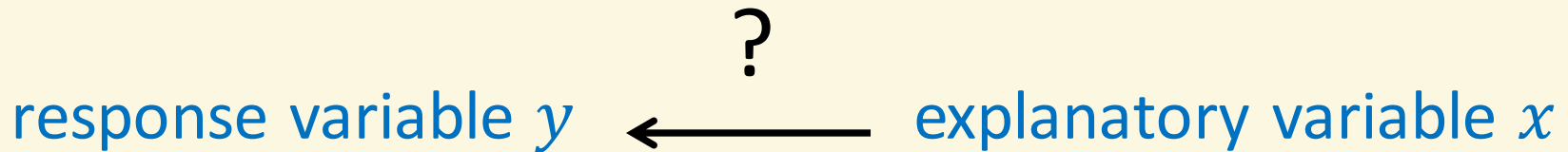
Meeting 2

Dr. Jos Hageman



WAGENINGEN UNIVERSITY
WAGENINGEN UR

A qualitative explanatory variable

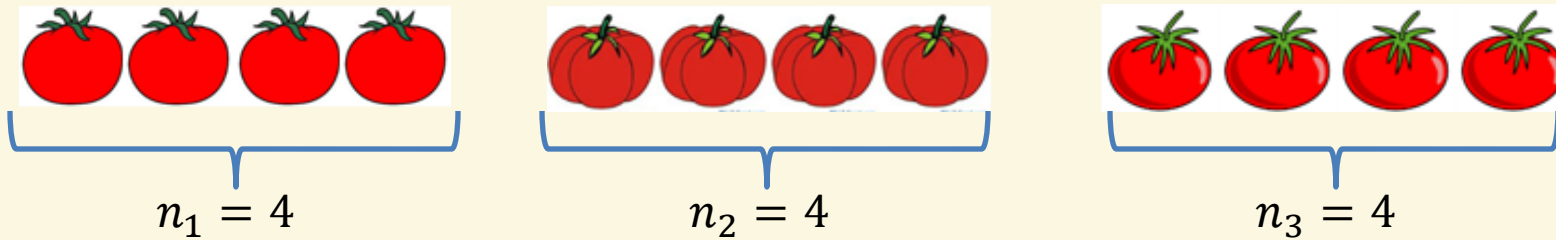


In regression y and x are both quantitative variables.

Now, suppose that y is quantitative, but x is qualitative.

Example: sweet taste of tomatoes

Does sweet taste differ between three types of tomato?



3 random samples from 3 populations: round, beef, or cherry tomatoes.

Response y = sweet taste, as scored by judges (scale 0 to 100).

Explanatory variable x = type of tomato (round, beef, or cherry).

Sweet taste of tomatoes, aim of experiment



Response y = taste

Explanatory variable x = type of tomato

Aim:

Inference about systematic differences in taste between the three types of tomato.

Provide estimates, standard errors, tests, and confidence intervals.

| | taste | type |
|----|-------|------|
| 1 | 25.44 | r |
| 2 | 28.10 | r |
| 3 | 46.46 | r |
| 4 | 36.96 | r |
| 5 | 24.83 | b |
| 6 | 28.47 | b |
| 7 | 48.15 | b |
| 8 | 31.78 | b |
| 9 | 53.42 | c |
| 10 | 70.87 | c |
| 11 | 57.07 | c |
| 12 | 38.08 | c |

In more general terms

t populations of units, with t random samples of sizes $n_1 \dots n_t$.

For the tomatoes: $t = 3, n_1 = 4, n_2 = 4$ and $n_3 = 4$.

For each sampled unit response y is measured.

The (unknown) population means for y are $\mu_1 \dots \mu_t$.

Are there differences between $\mu_1 \dots \mu_t$?

If so, how large are these differences?

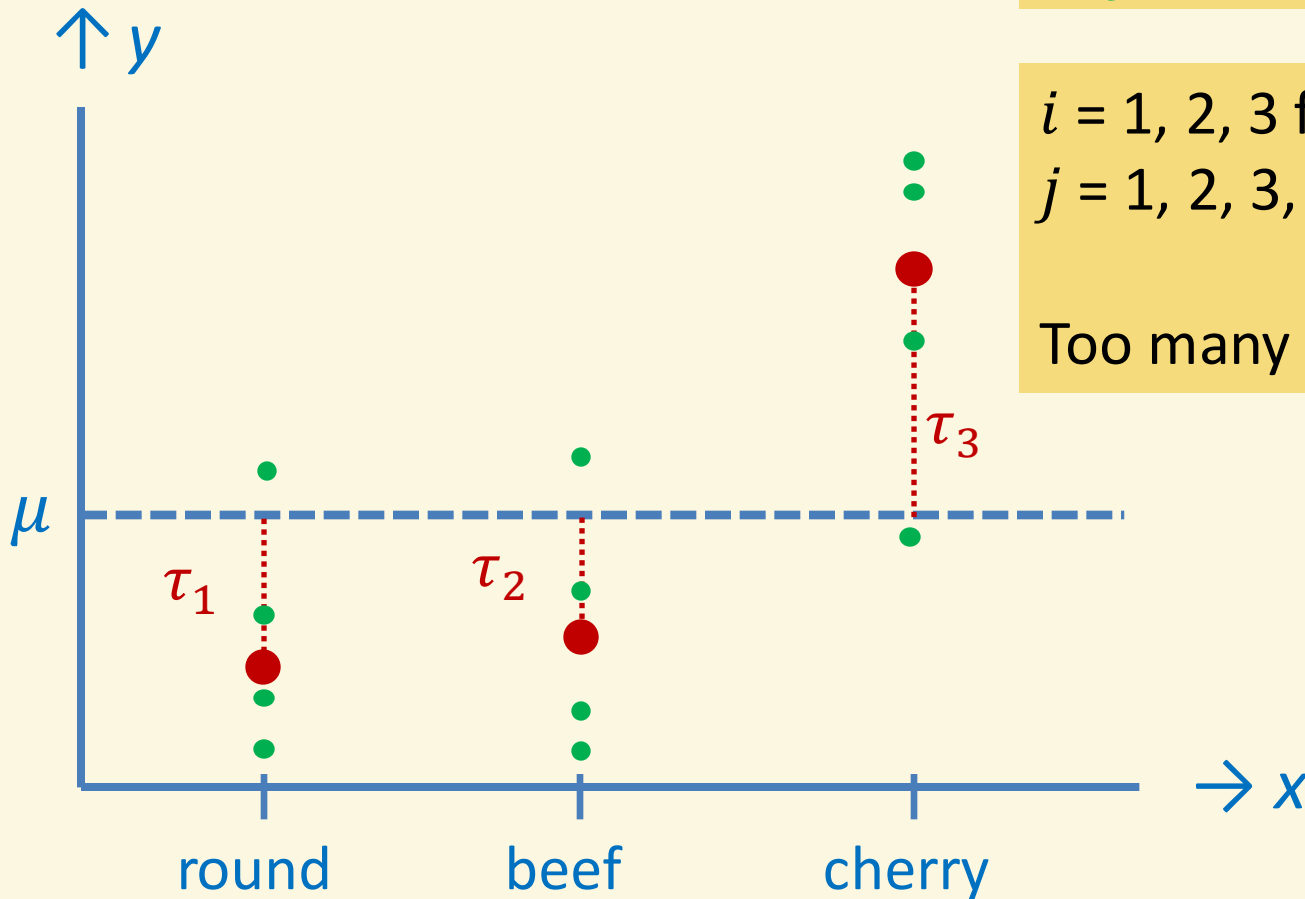
Provide tests, estimates, standard errors, confidence intervals.

Building the model

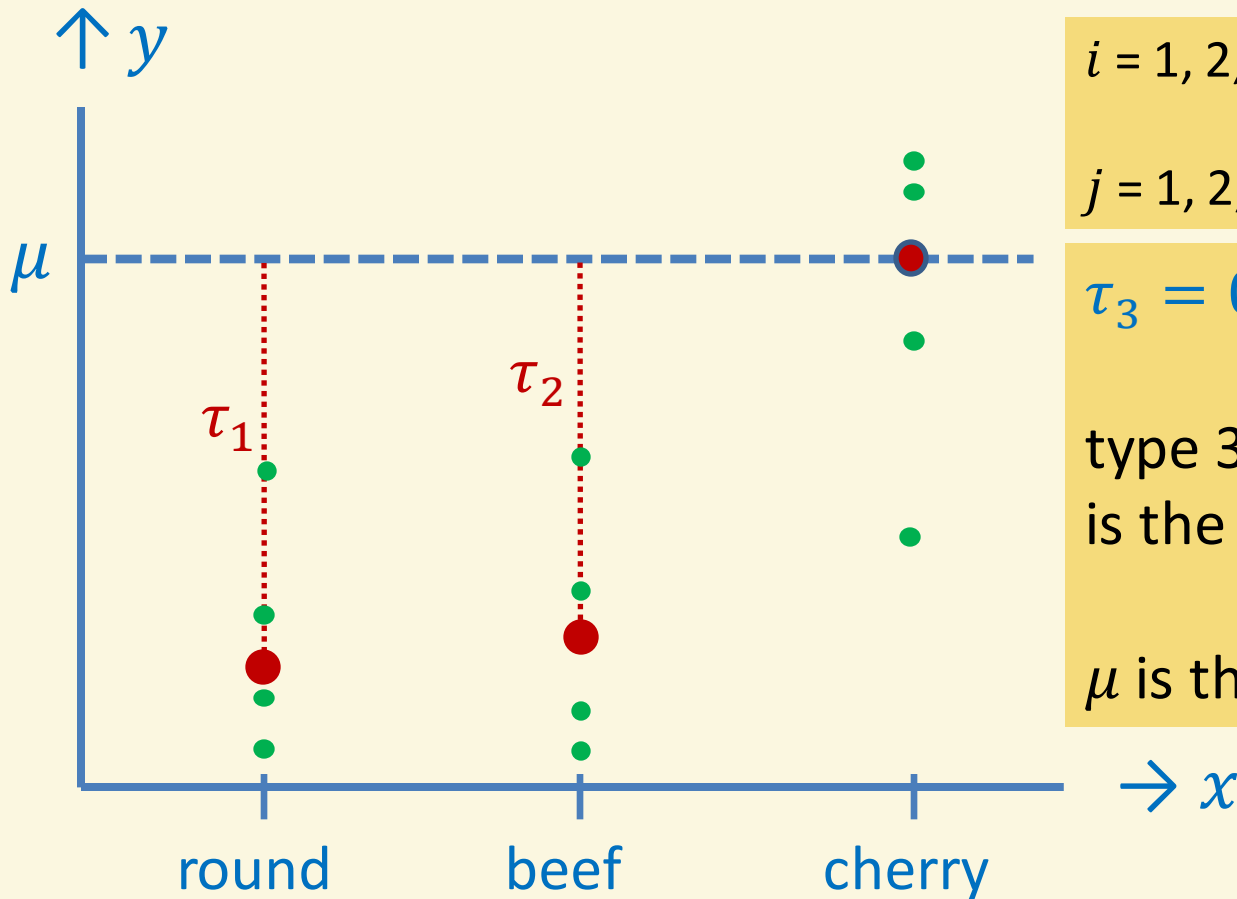
$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

$i = 1, 2, 3$ for type
 $j = 1, 2, 3, 4$ for tomato

Too many parameters!



Cornerstone representation



$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

$i = 1, 2, 3$ for type

$j = 1, 2, 3, 4$ for tomato

$$\tau_3 = 0$$

type 3 (cherry tomato)
is the **reference**

μ is the mean of type 3

Alternative model representations

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

$$\sum_{i=1}^t \tau_i = 0$$

alternative side condition:

sum-to-zero condition

μ is the overall mean of the three types

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

means model

no side condition, just three means

Cornerstone representation: interpretation of model parameters

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

$i = 1, 2, 3$ for type
 $j = 1, 2, 3, 4$ for tomato

$$\tau_3 = 0$$

type 3 (cherry tomato)
is the **reference**

type 1 and 2 relative to 3

$$\mu_1 = \mu + \tau_1 \quad \text{mean type 1}$$

$$\mu_2 = \mu + \tau_2 \quad \text{mean type 2}$$

$$\mu_3 = \mu \quad \text{mean type 3}$$

$$\mu = \mu_3 \quad \text{reference level}$$

$$\tau_1 = \mu_1 - \mu_3 \quad \text{type 1 versus 3}$$

$$\tau_2 = \mu_2 - \mu_3 \quad \text{type 2 versus 3}$$

Model assumptions about error terms ϵ

Same assumptions as in regression:

ϵ 's are mutually independent,
i.e. we need random samples from the populations

ϵ 's normally distributed around 0 with constant variance σ_ϵ^2

For the tomato example there are four parameters in the model:

μ, τ_1, τ_2 and σ_ϵ^2 (assuming that cherry is the reference, i.e. $\tau_3 = 0$).

Least squares estimation of μ and τ 's

Formally, again we can minimize the sum of squares

$$SS = \sum_{i=1}^t \sum_{j=1}^{n_i} \left(y_{ij} - (\mu + \tau_i) \right)^2$$

We simply get a sample mean as estimators for μ ,
and differences between sample means as estimators for τ 's:

$$\hat{\mu} = \bar{y}_{3.}, \quad \hat{\tau}_1 = \hat{\mu}_1 - \hat{\mu}_3 = \bar{y}_{1.} - \bar{y}_{3.}, \dots \quad \text{for the tomatoes.}$$

Residuals and fitted values

Again, we have fitted values and residuals:

$$\hat{y}_{ij} = \hat{\mu}_i = \hat{\mu} + \hat{\tau}_i = \bar{y}_{i.} ,$$

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.} .$$

As before, minimum SS is the sum of squares for error (or residual)

$$SSE = \sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^t \sum_{j=1}^{n_i} e_{ij}^2 .$$

Estimation of σ_{ϵ}^2

We can estimate the variance σ_{ϵ}^2 from each sample:

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2, i = 1 \dots t.$$

These estimates are pooled:

$$\begin{aligned} \hat{\sigma}_{\epsilon}^2 &= \frac{(n_1 - 1)s_1^2 + \dots + (n_t - 1)s_t^2}{(n_1 - 1) + \dots + (n_t - 1)} = \\ &= \frac{1}{n_1 + \dots + n_t - t} \sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = \frac{1}{N - t} \sum_{i=1}^t \sum_{j=1}^{n_i} e_{ij}^2 = MSE \end{aligned}$$

where $N = n_1 + \dots + n_t$ is the total sample size.

For the tomatoes $N = n_1 + n_2 + n_3 = 4 + 4 + 4 = 12$ and $t = 3$.

One-way ANOVA = regression with dummies

| | taste | type | x_1 | x_2 |
|----|-------|------|-------|-------|
| 1 | 25.44 | r | 1 | 0 |
| 2 | 28.10 | r | 1 | 0 |
| 3 | 46.46 | r | 1 | 0 |
| 4 | 36.96 | r | 1 | 0 |
| 5 | 24.83 | b | 0 | 1 |
| 6 | 28.47 | b | 0 | 1 |
| 7 | 48.15 | b | 0 | 1 |
| 8 | 31.78 | b | 0 | 1 |
| 9 | 53.42 | c | 0 | 0 |
| 10 | 70.87 | c | 0 | 0 |
| 11 | 57.07 | c | 0 | 0 |
| 12 | 38.08 | c | 0 | 0 |

x_1 and x_2 are dummy variables for round and beef tomatoes.

No dummy for cherry tomatoes is needed (as you will see).

Multiple regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

is the same as one-way ANOVA.

Compare parameters in ANOVA and regression

| | | one-way ANOVA: mean taste | Regression: mean taste | | |
|--------|-----|---------------------------------|---------------------------|-------|---|
| model | i | $\mu_i = \mu + \tau_i$ | x_1 | x_2 | $\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ |
| round | 1 | $\mu_1 = \mu + \tau_1$ | 1 | 0 | $\mu_1 = \beta_0 + \beta_1 * 1 + \beta_2 * 0 = \beta_0 + \beta_1$ |
| beef | 2 | $\mu_2 = \mu + \tau_2$ | 0 | 1 | $\mu_2 = \beta_0 + \beta_1 * 0 + \beta_2 * 1 = \beta_0 + \beta_2$ |
| cherry | 3 | $\mu_3 = \mu + 0 = \mu$ | 0 | 0 | $\mu_3 = \beta_0 + \beta_1 * 0 + \beta_2 * 0 = \beta_0$ |

- β_0 in regression same as μ in ANOVA = μ_3 = mean taste for cherry
- β_1 in regression same as τ_1 in ANOVA = $\mu_1 - \mu_3$
- β_2 in regression same as τ_2 in ANOVA = $\mu_2 - \mu_3$

Sums of squares in ANOVA table

One-way ANOVA can be fitted by regression.
Again we have sums of squares SST , SSR , SSE .
But SSR will now be called SS_{Treat} .

Again, SS relate to response y , **systematic** and **random** parts:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

total sum of squares SST

treatment sum of squares SS_{Treat}

error (or residual) sum of squares SSE

ANOVA table, tomato example (in R)



Analysis of Variance Table

Response: taste

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|---------|
| TomType | 2 | 1187.5 | 593.75 | 4.7114 | 0.03981 |
| Residuals | 9 | 1134.2 | 126.02 | | |
| Total | 11 | 2322.0 | | | |

SSTreat

SSE

SST

‘Tomtype’ is the name given in the R program to the factor with three levels for the types of tomato

‘Residuals’ is how R indicates the error part

F-test for tomato example

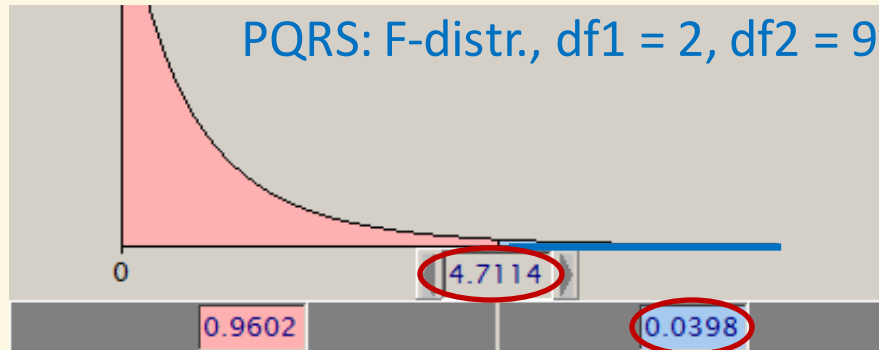


$F = MSR/MSE$ becomes $F = MSTreat/MSE$.

- F-test was test for predictive value, $H_0: \beta_1 = \beta_2 = 0$.
- β_1, β_2 in regression are the same as τ_1, τ_2 in one-way ANOVA
- So, $H_0: \beta_1 = \beta_2 = 0$ is the same as $H_0: \tau_1 = \tau_2 (= \tau_3) = 0$.
- $H_0: \tau_1 = \tau_2 (= \tau_3) = 0$ is the same as $H_0: \mu_1 = \mu_2 = \mu_3$.

This is a test whether the three types of tomato have the same population mean for sweet taste (same expected taste) or not.

F-test for tomato example, P-value



Under H_0 F follows an F-distribution, $df_1 = 2$, $df_2 = 11 - 2 = 9$.

Large values of F lead to rejection.

P-value is 0.0398 (area to the right of outcome 4.7114) and below 0.05.

So, $F = 4.7114$ is an unlikely outcome when H_0 is true.

H_0 is rejected: it is shown that at least two of the types of tomato have different population means for taste (different expected taste).

What next?



F-test shows differences among the three population means for taste of the types of tomato.

Are all means different, or are two means the same and is the third different from the other two?

To find out, additionally pairwise comparisons between the three sample means are performed.

Pairwise comparison by t-test, tomato example

For example, round vs. cherry :

$$H_0: \mu_1 - \mu_3 = 0 \text{ vs. } H_a: \mu_1 - \mu_3 \neq 0$$

$$t = (\hat{\mu}_1 - \hat{\mu}_3) / se(\hat{\mu}_1 - \hat{\mu}_3) = (\bar{y}_{1.} - \bar{y}_{3.}) / sd(\bar{y}_{1.} - \bar{y}_{3.}) =$$

$$(\bar{y}_{1.} - \bar{y}_{3.}) / \sqrt{\frac{2\hat{\sigma}_\epsilon^2}{4}} = (\bar{y}_{1.} - \bar{y}_{3.}) / \sqrt{\frac{2 * MSE}{4}}$$

Refer to t-distribution, $11 - 2 = 9$ degrees of freedom (from SSE).

Details of $sd(\bar{y}_{1.} - \bar{y}_{3.})$

$$sd(\bar{y}_{1.} - \bar{y}_{3.}) = \sqrt{Var(\bar{y}_{1.} - \bar{y}_{3.})} =$$

$$\sqrt{Var(\bar{y}_{1.}) + Var(\bar{y}_{3.})} = \sqrt{\frac{\sigma_{\epsilon}^2}{n_1} + \frac{\sigma_{\epsilon}^2}{n_3}}$$

For the tomatoes: $n_1 = 4, n_3 = 4$ and $sd(\bar{y}_{1.} - \bar{y}_{3.}) = \sqrt{2\sigma_{\epsilon}^2/4}$.

t-test, tomato example, R



Round versus cherry ($\mu_1 - \mu_3$) is in the output from R below (where cherry tomatoes are the reference).

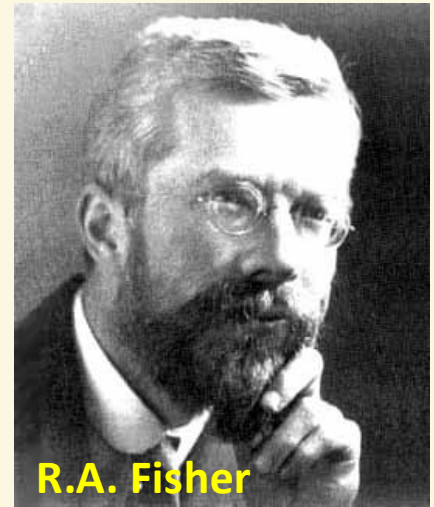
Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 54.861 | 5.613 | 9.774 | 4.33e-06 |
| tomtype2r | -20.619 | 7.938 | -2.598 | 0.0289 |
| tomtype2b | -21.555 | 7.938 | -2.715 | 0.0238 |

The P-value is 0.029 and below 0.05.

We have shown that expected sweet taste is higher for cherry tomatoes than for round tomatoes.

Fisher's LSD method



A difference, e.g. round vs. cherry, is significant when

$$\text{abs}((\bar{y}_1 - \bar{y}_3) / \text{sd}(\bar{y}_1 - \bar{y}_3)) > t_0$$

t_0 from t-distribution with $\alpha/2 = 0.025$ to the right, df from SSE

Same as:

$$\text{abs}(\bar{y}_1 - \bar{y}_3) > t_0 \text{ se}(\hat{\tau}_1)$$

$t_0 \text{ se}(\hat{\tau}_1)$ is called the least significant difference (LSD).

The LSD for the tomato example

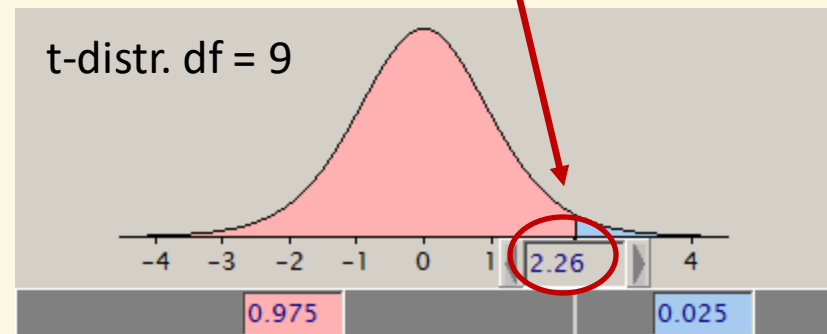


For equal sample sizes, say n , LSD is the same for all pairs:

$$LSD = t_0 \sqrt{2MSE/n}.$$

For the tomato example: $n = 4$, $MSE = 126.02$ and $t_0 = 2.26$

$$LSD = 2.26 \sqrt{2 * 126.02/4} = 17.94$$



The LSD as a yardstick



When the LSD, e.g. 17.94 for the tomatoes, fits in between two means, e.g. $\bar{y}_1.$ and $\bar{y}_3.$, the corresponding treatments are significantly different.

Fisher's F-protected LSD



When pairwise comparisons by the LSD method are only performed after a significant result with the F-test, the LSD method is called **F-protected**.

All pairwise comparisons, tomato example, R

```
$`emmeans`
```

| TomType | emmean | SE | df | lower.CL | upper.CL |
|---------|--------|------|----|----------|----------|
| 1 | 34.2 | 5.61 | 9 | 21.5 | 46.9 |
| 2 | 33.3 | 5.61 | 9 | 20.6 | 46.0 |
| 3 | 54.9 | 5.61 | 9 | 42.2 | 67.6 |

Confidence level used: 0.95

```
$contrasts
```

| contrast | estimate | SE | df | t.ratio | p.value |
|----------|----------|------|----|---------|---------|
| 1 - 2 | 0.935 | 7.94 | 9 | 0.118 | 0.9088 |
| 1 - 3 | -20.619 | 7.94 | 9 | -2.598 | 0.0289 |
| 2 - 3 | -21.555 | 7.94 | 9 | -2.715 | 0.0238 |



Results with Fisher's LSD method

with routine 'emmeans' from
library 'emmeans'

Details confidence intervals for differences

For each pairwise difference a 0.95 confidence interval may be constructed, e.g. for $(\mu_1 - \mu_3)$ for round versus cherry :

$$((\hat{\mu}_1 - \hat{\mu}_3) \pm t_0 \text{ se}(\hat{\mu}_1 - \hat{\mu}_3)).$$

This is the same as:

$$((\bar{y}_1 - \bar{y}_3) \pm LSD)$$

For round versus cherry: $(-20.62 \pm 17.94) = (-38.6, -2.7)$.

Notation of significant differences, tomato example

| type | means | |
|------|-------|---|
| ro | 34.24 | a |
| be | 33.31 | a |
| ch | 54.86 | b |

Common letter implies means are not significantly different.

| type | | | |
|------|--------------------|--------------------|--------------------|
| | ro | be | ch |
| | 34.24 ^a | 33.31 ^a | 54.86 ^b |

Common superscript implies means are not significantly different.

| type | | | |
|------|-------|-------|-------|
| | be | ro | ch |
| | 33.31 | 34.24 | 54.86 |
| | <hr/> | | |

Common underline implies means are not significantly different (means need to be ordered from low to high).

Type I error

Suppose that on average all three types of tomato taste the same: $\mu_1 = \mu_2 = \mu_3$.

For each t-test that we perform, we have a small probability (the type I error α) to wrongly decide that two treatments are different, usually $\alpha = 0.05$

With several tests, the probability that at least once we wrongly decide that two treatments are different will be larger than 0.05.

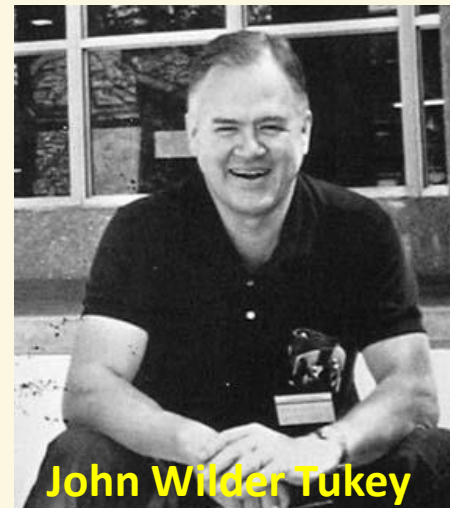
This has worried some statisticians in the past.

Experimentwise error

Assuming that all treatments are the same, the probability to decide at least once that two treatments are different is called the **experiment wise error**.

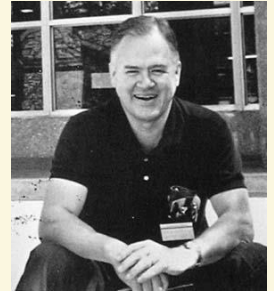
Methods have been proposed to control the experimentwise error, i.e. make sure that experimentwise error is below e.g. 0.05.

One popular method to do so is **Tukey's** range test.



Tukey's range test

Two treatments differ significantly, when the difference between their sample means exceeds yardstick W .



$$|\bar{y}_i - \bar{y}_j| > W,$$

$$\text{where } W = q \sqrt{\frac{s_{\varepsilon}^2}{n}}.$$

q depends on number of treatments (t), df from SSE , and desired α_{exp} .

q is from the so-called studentized range distribution (table 10 in O&L).

For unequal sample sizes: replace $1/n$ by harmonic mean $(\frac{1}{n_i} + \frac{1}{n_j})/2$ (approximate Tukey-Kramer method).

Two way ANOVA



Taste of tomatoes revisited



| taste type | | | ripe |
|------------|-------|---|------|
| 1 | 46.46 | r | lo |
| 2 | 36.96 | r | lo |
| 3 | 25.44 | r | sh |
| 4 | 28.10 | r | sh |
| 5 | 48.15 | b | lo |
| 6 | 31.78 | b | lo |
| 7 | 24.83 | b | sh |
| 8 | 28.47 | b | sh |
| 9 | 57.07 | c | lo |
| 10 | 70.87 | c | lo |
| 11 | 53.42 | c | sh |
| 12 | 38.08 | c | sh |

Response y = sweet taste

Two factors now:

Type of tomato:

beef, cherry or round tomatoes.

Ripening duration: long or short.

How to handle two factors together?

Exploring structure in treatments

We want to explore the structure in the 6 treatments.

For instance, is the effect of ripening the same for the three types of tomato?

If so, we can say something about ripening without having to bother about the type of tomato.

If not, we have to describe the effect of ripening for each type of tomato separately.

Interaction

When the effect of ripening depends upon the type of tomato, we say that there is **interaction** between the experimental factors for tomato type and ripening.

We want to know more about interaction.

The one-way ANOVA is of little use in this respect, because it does not explore the structure of the treatments.

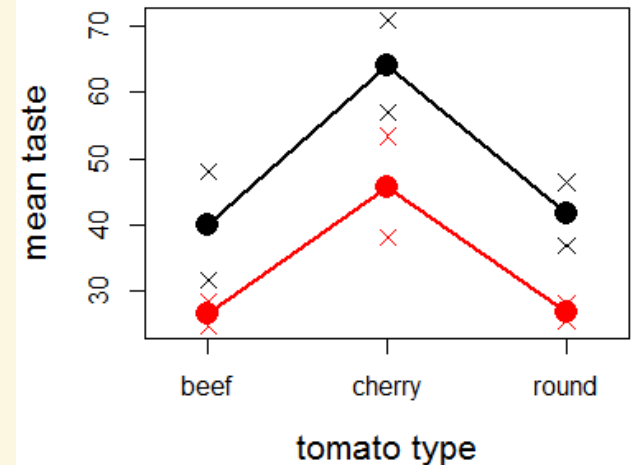
A profile plot for the tomato example

- calculate the 6 sample means
- plot means vs. type, showing levels of ripe
- connect points for same level of ripe



profile plot suggests no interaction

| type | ripe | |
|------|-------|-------|
| | lo | sh |
| b | 39.97 | 26.65 |
| c | 63.97 | 45.75 |
| r | 41.71 | 26.77 |



Two-way ANOVA model without interaction

Model without interaction is the ‘additive model’:

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

Easy and **concise conclusions**, e.g. for tomatoes same effect of duration can be reported for all three types of tomato.

More accurate inference, e.g. for tomatoes long vs. short averaged over types, and not per type.

But the model has to be appropriate, otherwise conclusions can be **misleading**, e.g. for tomatoes when effect of duration would be very different for (some) types of tomato.

Two-way ANOVA model with interaction

Two ways of writing model with interaction:

Means Model: $y_{ijk} = \mu_{ij} + \epsilon_{ijk}$ (no factorial structure, one-way ANOVA)

Effects Model: $y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \epsilon_{ijk}$ (factorial structure)

$\tau\beta_{ij}$ is an **interaction term** (read $\tau\beta$ as a single symbol)

correction on top of “additive” part $\mu + \tau_i + \beta_j$ with **main effects**

effect factor A may differ at different levels of factor B and vice versa

interaction is explicit in the effects model

Effects model with interaction, tomato example

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$



| | <i>ripe</i> | |
|-------------|-------------|-------|
| <i>type</i> | lo | sh |
| b | 39.97 | 26.65 |
| c | 63.97 | 45.75 |
| r | 41.71 | 26.77 |

| | <i>ripe</i> | |
|-------------|---|---|
| <i>type</i> | lo | sh |
| b | $\mu + \tau_1 + \beta_1 + \tau\beta_{11}$ | $\mu + \tau_1 + \beta_2 + \tau\beta_{12}$ |
| c | $\mu + \tau_2 + \beta_1 + \tau\beta_{21}$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ |
| r | $\mu + \tau_3 + \beta_1 + \tau\beta_{31}$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ |

sample means \bar{y}_{ij} . estimates for population means $\mu_{ij} = \mu + \tau_i + \beta_j + \tau\beta_{ij}$

Effects model with interaction, tomato example

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$



| <i>type</i> | <i>ripe</i> | |
|-------------|---|---|
| | lo | sh |
| b | $\mu + \tau_1 + \beta_1 + \tau\beta_{11}$ | $\mu + \tau_1 + \beta_2 + \tau\beta_{12}$ |
| c | $\mu + \tau_2 + \beta_1 + \tau\beta_{21}$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ |
| r | $\mu + \tau_3 + \beta_1 + \tau\beta_{31}$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ |

we have $1 + 3 + 2 + 6 = 12$ parameters.

but we need only six parameters to represent six population means

Cornerstone representation – step 1



| type | ripe | | type | ripe | |
|------|---|---|------|-------|-------|
| | lo | sh | | lo | sh |
| b | $\mu + \tau_1 + \beta_1 + \tau\beta_{11}$ | $\mu + \tau_1 + \beta_2 + \tau\beta_{12}$ | b | μ | μ |
| c | $\mu + \tau_2 + \beta_1 + \tau\beta_{21}$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ | c | μ | μ |
| r | $\mu + \tau_3 + \beta_1 + \tau\beta_{31}$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ | r | μ | μ |

R would take beef & long as a reference (ordering levels alphabetically)

μ is the population mean for type beef and duration long

Cornerstone representation – step 2



| type | ripe | | type | ripe | |
|------|---|---|------|----------------|--------------------------|
| | lo | sh | | lo | sh |
| b | $\mu + \tau_1 + \beta_1 + \tau\beta_{11}$ | $\mu + \tau_1 + \beta_2 + \tau\beta_{12}$ | b | μ | $\mu + \beta_2$ |
| c | $\mu + \tau_2 + \beta_1 + \tau\beta_{21}$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ | c | $\mu + \tau_2$ | $\mu + \tau_2 + \beta_2$ |
| r | $\mu + \tau_3 + \beta_1 + \tau\beta_{31}$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ | r | $\mu + \tau_3$ | $\mu + \tau_3 + \beta_2$ |



use main effects τ_2, τ_3, β_2 for population means in 1st row & 1st column

no need for τ_1, β_1 , i.e. $\tau_1 = 0, \beta_1 = 0$

Cornerstone representation – step 3



| type | ripe | | | type | ripe | |
|------|---|---|---|------|----------------|---|
| | lo | sh | | | lo | sh |
| b | $\mu + \tau_1 + \beta_1 + \tau\beta_{11}$ | $\mu + \tau_1 + \beta_2 + \tau\beta_{12}$ | → | b | μ | $\mu + \beta_2$ |
| c | $\mu + \tau_2 + \beta_1 + \tau\beta_{21}$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ | | c | $\mu + \tau_2$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ |
| r | $\mu + \tau_3 + \beta_1 + \tau\beta_{31}$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ | | r | $\mu + \tau_3$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ |

interactions $\tau\beta_{22}, \tau\beta_{32}$, on top of main effects, for remaining two means

no need for interactions in 1st row & 1st column,

i.e. $\tau\beta_{11} = \tau\beta_{12} = \tau\beta_{21} = \tau\beta_{31} = 0$

Cornerstone representation



| type | ripe | | | type | ripe | |
|------|---|---|---|------|----------------|---|
| | lo | sh | | | lo | sh |
| b | $\mu + \tau_1 + \beta_1 + \tau\beta_{11}$ | $\mu + \tau_1 + \beta_2 + \tau\beta_{12}$ | → | b | μ | $\mu + \beta_2$ |
| c | $\mu + \tau_2 + \beta_1 + \tau\beta_{21}$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ | | c | $\mu + \tau_2$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ |
| r | $\mu + \tau_3 + \beta_1 + \tau\beta_{31}$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ | | r | $\mu + \tau_3$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ |

$\mu, \tau_2, \tau_3, \beta_2, \tau\beta_{22}, \tau\beta_{32}$ are left to reproduce six means exactly

Estimates with cornerstone representation -1

| type | ripe | |
|------|-------|-------|
| | lo | sh |
| b | 39.97 | 26.65 |
| c | 63.97 | 45.75 |
| r | 41.71 | 26.77 |

| type | ripe | |
|------|----------------|---|
| | lo | sh |
| b | μ | $\mu + \beta_2$ |
| c | $\mu + \tau_2$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ |
| r | $\mu + \tau_3$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ |



$\hat{\mu}$ is sample mean taste beef / long: 39.97

$\hat{\tau}_2$ is taste difference of cherry vs beef
after long ripening: $63.97 - 39.97 = 24.00$

$\hat{\tau}_3$ is taste difference of round vs beef
after long ripening: $41.71 - 39.97 = 1.74$

$\hat{\beta}_2$ is taste diff of short vs long ripening for
beef tomatoes: $26.65 - 39.97 = -13.32$

| type | ripe | |
|------|-------------|-------------------|
| | lo | sh |
| b | 39.97 | 39.97 -13.32 |
| c | 39.97+24.00 | 39.97+24.00-13.32 |
| r | 39.97+ 1.74 | 39.97+ 1.74-13.32 |

We have reproduced the sample means in the 1st row and 1st column.
But the two means in the lower right corner are not OK yet.

Estimates with cornerstone representation - 2



| type | ripe | |
|------|-------|-------|
| | lo | sh |
| b | 39.97 | 26.65 |
| c | 63.97 | 45.75 |
| r | 41.71 | 26.77 |

| type | ripe | |
|------|----------------|---|
| | lo | sh |
| b | μ | $\mu + \beta_2$ |
| c | $\mu + \tau_2$ | $\mu + \tau_2 + \beta_2 + \tau\beta_{22}$ |
| r | $\mu + \tau_3$ | $\mu + \tau_3 + \beta_2 + \tau\beta_{32}$ |

We use the interaction terms to reproduce the last two means

| type | ripe | |
|------|-------------|------------------------|
| | lo | sh |
| b | 39.97 | 39.97 -13.32 |
| c | 39.97+24.00 | 39.97+24.00-13.32-4.90 |
| r | 39.97+ 1.74 | 39.97+ 1.74-13.32-1.62 |

$\widehat{\tau\beta_{22}}$ added to $\hat{\mu} + \hat{\tau}_2 + \hat{\beta}_2$ for mean of cherry-short: $45.75 - (39.97 + 24.00 - 13.32) = -4.90$

$\widehat{\tau\beta_{32}}$ added to $\hat{\mu} + \hat{\tau}_3 + \hat{\beta}_2$ for mean of round-short: $26.77 - (39.87 + 1.74 - 13.32) = -1.62$

More about the cornerstone representation

Different statistical software may choose a different level as a reference, usually by default the first or the last level of a factor.

By default R takes the first level (after alphabetical ordering)

Other side conditions than corner-stone are possible, such as the sum-to-zero condition, but have become more rare.

In all cases, interaction parameters are of first interest.

F-test in means model

Means model: $y_{ijk} = \mu_{ij} + \epsilon_{ijk}$ essentially a one-way ANOVA

Analysis of Variance Table

Response: taste

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|--------|
| type:ripe | 5 | 1919.88 | 383.98 | 5.7288 | 0.028 |
| Residuals | 6 | 402.15 | 67.03 | | |



R output

6 - 1 = 5 df,
because
6 (= 3 * 2)
treatments

$H_0: \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22} = \mu_{31} = \mu_{32}$ vs.
 H_a : at least two means different

P-value = 0.028 < 0.05, reject H_0 , some means differ.

Due to type, ripe, and / or their interaction?

Splitting up $SSTreat$

$$\underline{y}_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \underline{\epsilon}_{ijk}, \quad \underline{\epsilon}_{ijk} \sim N(0, \sigma_\epsilon^2), \text{ ind.}$$

We want an F-test for interaction.

Therefore, we split $SSTreat$ into three separate SS:

SSA SS for main effects A

SSB SS for main effects B

$SSAB$ SS for interaction between A and B

$$SSTreat = SSA + SSB + SSAB$$

Splitting *SSTreat*, tomato example



We start with the SS for interaction.

Compare SSEs from models with and without interaction:

CM : $y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \epsilon_{ijk}$ (interaction model)

RM: $y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$ (additive model)

```
> CM.ia <- lm(taste ~ type + ripe + type:ripe)
> deviance(CM.ia)
[1] 402.1538
> RM.ia <- lm(taste ~ type + ripe)
> deviance(RM.ia)
[1] 414.6399
> SS.ia <- deviance(RM.ia) - deviance(CM.ia)
[1] 12.48602
```

So: $SS_{type \times ripe} = 414.6399 - 402.1538 = 12.49$

Splitting *SSTreat*, tomato example

SS for main effect of *type*

Restrict attention to additive model

$$\text{CM: } y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

$$\text{RM: } y_{ijk} = \mu + \beta_j + \epsilon_{ijk}$$

```
> CM.m1 <- lm(taste ~ type + ripe)
> deviance(CM.m1)
[1] 414.6399
> RM.m1 <- lm(taste ~ ripe)
> deviance(RM.m1)
[1] 1602.059
> SS.type <- deviance(RM.m1) - deviance(CM.m1)
[1] 1187.419
```



SS for main effect of *ripe*

Restrict attention to additive model

$$\text{CM: } y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

$$\text{RM: } y_{ijk} = \mu + \tau_i + \epsilon_{ijk}$$

```
> CM.m2 <- lm(taste ~ type + ripe)
> deviance(CM.m2)
[1] 414.6399
> RM.m2 <- lm(taste ~ type)
> deviance(RM.m2)
[1] 1134.615
> SS.ripe <- deviance(RM.m2) - deviance(CM.m2)
[1] 719.975
```

So: $SS_{\text{type}} = 1187.42$ and $SS_{\text{ripe}} = 719.98$

$SS_{\text{type}} + SS_{\text{ripe}} + SS_{\text{type} \times \text{ripe}} = 1187.42 + 719.98 + 12.49 = 1919.88 = SS_{\text{Treat}}$

Default steps to make SS in R



```
> anova(lm(taste ~ type + ripe + type:ripe))
```

Analysis of Variance Table

Response: taste

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|--------|
| type | 2 | 1187.42 | 593.71 | 8.8579 | 0.0162 |
| ripe | 1 | 719.98 | 719.98 | 10.7418 | 0.0169 |
| type:ripe | 2 | 12.49 | 6.24 | 0.0931 | 0.9124 |
| Residuals | 6 | 402.15 | 67.03 | | |

SS are: type I SS =
accumulated SS =
sequential SS

In the R table SS are constructed like this:

$$y_{ijk} = \mu + \epsilon_{ijk}$$

+ type $\rightarrow 1187.42$

$$y_{ijk} = \mu + \tau_i + \epsilon_{ijk}$$

+ ripe $\rightarrow 719.98$

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}$$

+ type x ripe $\rightarrow 12.49$

$$y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \epsilon_{ijk}$$

F-test for interaction

```
> anova(lm(taste ~ type + ripe + type:ripe))
```

Analysis of Variance Table

Response: taste

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|--------|
| type | 2 | 1187.42 | 593.71 | 8.8579 | 0.0162 |
| ripe | 1 | 719.98 | 719.98 | 10.7418 | 0.0169 |
| type:ripe | 2 | 12.49 | 6.24 | 0.0931 | 0.9124 |
| Residuals | 6 | 402.15 | 67.03 | | |



$H_0: \tau\beta_{11} = \tau\beta_{12} = \dots = \tau\beta_{32} = 0$ versus H_a : not all $\tau\beta_{ij} = 0$

test statistic: $F = \frac{MStype \times ripe}{MSE} = \frac{6.24}{67.030} = 0.093$

P-value = 0.91 > 0.05, from F-distr., df1 = 2, df2 = 6, do not reject H_0

effect of ripening upon expected sweetness cannot be shown to be different across tomato types and vice versa

additive model can be assumed (Occam's razor)

F-tests for main effects, type

```
> anova(lm(taste ~ type + ripe + type:ripe))
```

Analysis of Variance Table

Response: taste

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|--------|
| type | 2 | 1187.42 | 593.71 | 8.8579 | 0.0162 |
| ripe | 1 | 719.98 | 719.98 | 10.7418 | 0.0169 |
| type:ripe | 2 | 12.49 | 6.24 | 0.0931 | 0.9124 |
| Residuals | 6 | 402.15 | 67.03 | | |



$H_0: \tau_1 = \tau_2 = \tau_3 = 0$ vs. H_a : not all $\tau_i = 0$

Test statistic: $F = \frac{MStype}{MSE} = 8.86$

P-value = 0.016 < 0.05, from F-distr. df1 = 2, df2 = 6, reject H_0 ,

expected sweetness differs among types (irrespective of ripe)

F-tests for main effects, ripe

```
> anova(lm(taste ~ type + ripe + type:ripe))
```

Analysis of Variance Table

Response: taste

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|--------|
| type | 2 | 1187.42 | 593.71 | 8.8579 | 0.0162 |
| ripe | 1 | 719.98 | 719.98 | 10.7418 | 0.0169 |
| type:ripe | 2 | 12.49 | 6.24 | 0.0931 | 0.9124 |
| Residuals | 6 | 402.15 | 67.03 | | |



$$H_0: \beta_1 = \beta_2 = 0 \quad \text{vs.} \quad H_a: \text{not all } \beta_j \neq 0$$

$$\text{Test statistic: } F = \frac{MS_{\text{ripe}}}{MSE} = 10.74$$

P-value = 0.017 < 0.05, from F-distr. df1 = 1, df2 = 6, reject H_0 ,
expected sweetness differs between long and short ripening
(irrespective of type)

What comes after the F-tests in two-way ANOVA?

For example, in an experiment about weight loss with two factors for three diets and two sexes, how do we proceed after the F-test for interaction?

We perform pairwise comparisons of means.

For instance with Fisher's LSD method or Tukey's range test.

Which table(s) of means we inspect depends upon the outcome of the test for interaction.