14 Nov - Lec 9

TWO-WAY ANOVA

In the one-way ANOVA we wanted to evaluate the effect of a categorical covariate (factor) on a continuous response.
This framework can be extended to the case of two or more factors

Example: we want to study the survival time of N mice subject to one of $K=3$ types of poison and one of $J=4$ types of treatment.
Hence, for each mouse, we have a combination of poison-treatment
DATA: $(y_i\,;\ \text{poison}_i\,;\ \text{treatment}_i)$ $\quad i=1,\dots,n=48$

Goal of the study is to understand the effect of the two factors on the response variable: understand if the distribution of the survival time varies depending on the level of the covariates.
In the example, it could be interesting to evaluate:

1. the MARGINAL EFFECT of the first factor (poison)
   i.e.: do all poisons have the same efficacy?
2. the MARGINAL EFFECT of the second factor (treatment)
   i.e.: do all treatments have the same efficacy?

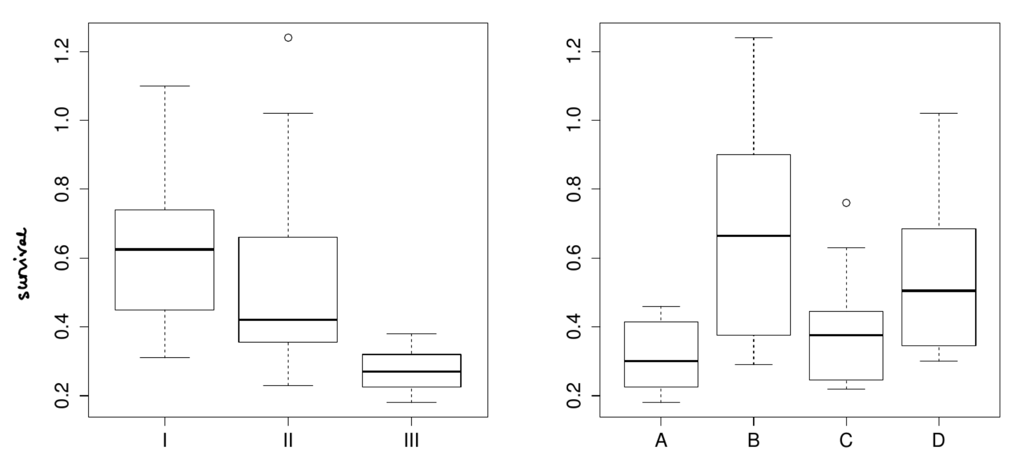
questions of the one-way ANOVA

3. the effect of poisons CONDITIONALLY on the treatment
   i.e.: if we fix the type of treatment, do different poisons have an effect on the survival time?
3. the effect of different treatments CONDITIONALLY on the poison.
   i.e.: if we fix the type of poison, do different treatments have different effect on the survival time?
4. the INTERACTION between the two factors
   i.e.: do different treatments have a different effect on the survival time depending on the type of poison?
   In the absence of interaction, one would simply choose the treatment with the largest effect, regardless of the type of poison.
   In the presence of interaction, a particular treatment could be preferable in combination with a particular poison.

If we denote with (I, II, III) the levels (types) of the poison factor and with (A, B, C, D) the levels of treatment

$\text{poison}_i \in \{I, II, III\}$ $\quad K=3$
$\text{treatment}_i \in \{A, B, C, D\}$ $\quad J=4$

MARGINAL EFFECT:
as in the one-way ANOVA, we study the group-specific means



in this case, we can do two separate analyses.

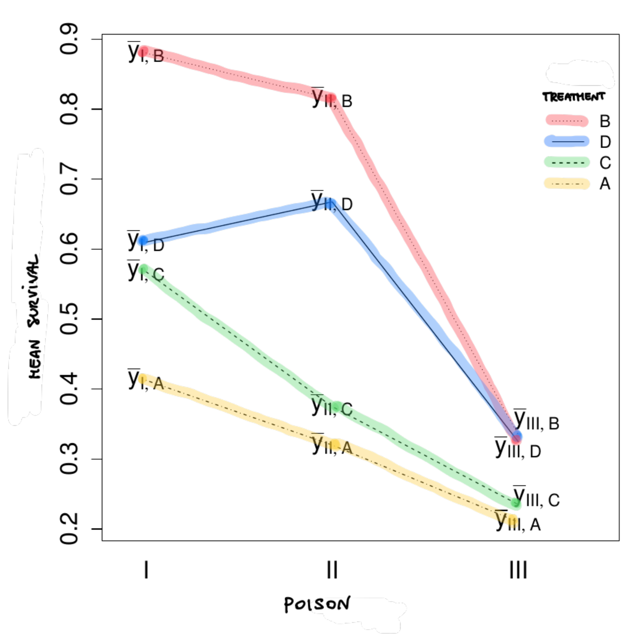
JOINT EFFECT:
we study the mean for each combination of poison/treatment

	A	B	C	D
I	0.41	0.88	0.57	0.61
II	0.32	0.81	0.38	0.67
III	0.21	0.33	0.23	0.33

$\bar{y}_{I,A}$ $\quad \bar{y}_{III,D}$

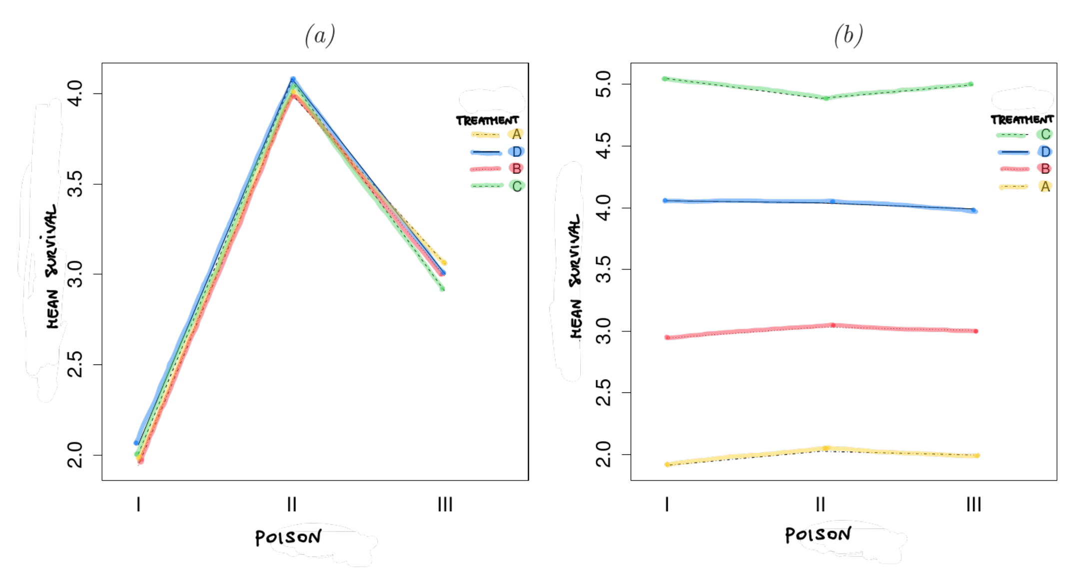
each entry in the table is the mean survival of the group poison + treatment

We can plot these means:



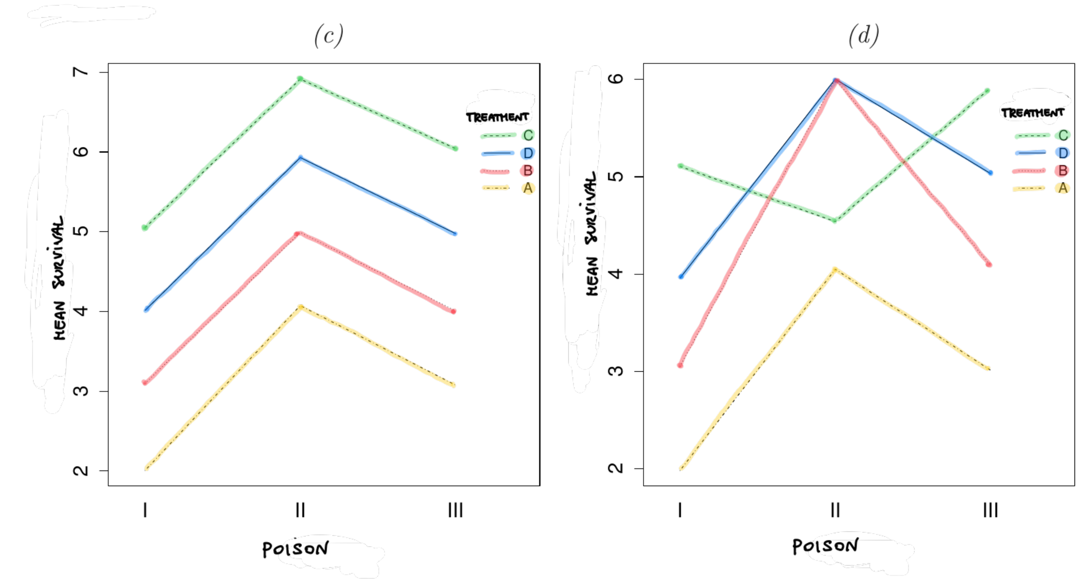
- each point is the mean of the group poison + treatment
- we connect points of the same treatment.

This type of plot shows different patterns depending on the effect of each factor



- lines corresponding to different treatments are equal:
  NO DIFFERENCE IN THE EFFECT OF TREATMENTS
- different poisons have different survival time:
  DIFFERENT EFFECT OF POISONS

- separate lines for different treatments
  DIFFERENCE IN THE EFFECT OF TREATMENTS
- for each treatment, the means corresponding to different poisons are equal
  NO DIFFERENCE IN THE EFFECT OF DIFFERENT POISONS



- lines corresponding to different treatments are different
  DIFFERENCE IN THE EFFECT OF DIFFERENT TREATMENTS
- different poisons have different survival time:
  DIFFERENCE IN THE EFFECT OF DIFFERENT POISONS
- NO INTERACTION: lines are parallel.
  The effect of the treatment is constant across poisons.

- lines corresponding to different treatments are different
  DIFFERENT EFFECT OF TREATMENTS
- different poisons have different survival time:
  DIFFERENT EFFECT OF POISONS
- INTERACTION: the type of poison affects the efficacy of the treatment
  e.g.: treatments A, B, D are more effective with poison II. Moreover, the efficacy of B and D is better than A, for all poisons.
  Treatment C is better with poisons I and III.

We can express these scenarios with a linear model:

- plot (a) corresponds to a model
  $$Y_i = f(\text{poison}_i) + \varepsilon_i \qquad \text{ONE-WAY ANOVA}$$
- plot (b) corresponds to a model
  $$Y_i = g(\text{treatment}_i) + \varepsilon_i \qquad \text{ONE-WAY ANOVA}$$
- plot (c) corresponds to a model
  $$Y_i = f(\text{poison}_i) + g(\text{treatment}_i) + \varepsilon_i \qquad \text{TWO-WAY ANOVA WITHOUT INTERACTION}$$
- plot (d) corresponds to a model
  $$Y_i = f(\text{poison}_i) + g(\text{treatment}_i) + h(\text{poison}_i \cdot \text{treatment}_i) + \varepsilon \qquad \text{TWO-WAY ANOVA WITH INTERACTION}$$

To formalise the model we need to encode each factor using DUMMY VARIABLES
Define the following variables, for $i=1,\dots,n$ $\quad$ (n = sample size)

$$P_{i,I} = \begin{cases} 1 & \text{if poison}_i = I \\ 0 & \text{otherwise} \end{cases} \qquad t_{i,A} = \begin{cases} 1 & \text{if treatment}_i = A \\ 0 & \text{otherwise} \end{cases}$$

$$P_{i,II} = \begin{cases} 1 & \text{if poison}_i = II \\ 0 & \text{otherwise} \end{cases} \qquad t_{i,B} = \begin{cases} 1 & \text{if treatment}_i = B \\ 0 & \text{otherwise} \end{cases}$$

$$P_{i,III} = \begin{cases} 1 & \text{if poison}_i = III \\ 0 & \text{otherwise} \end{cases} \qquad t_{i,C} = \begin{cases} 1 & \text{if treatment}_i = C \\ 0 & \text{otherwise} \end{cases}$$

$$t_{i,D} = \begin{cases} 1 & \text{if treatment}_i = D \\ 0 & \text{otherwise} \end{cases}$$

TWO-WAY ANOVA WITHOUT INTERACTION

The total number of dummy var. is $J+K = 4+3 = 7$.
However, similarly to the one-way ANOVA, if we include the intercept we need to define X so to avoid multicollinearity.
We have to remove one dummy for each factor: the removed level will be the reference group.
Hence the number of parameters is $\quad 1 + (J-1) + (K-1) = 1+3+2 = 6$
The linear model then is

$$Y_i = \mu + \underbrace{\alpha_{II} P_{i,II} + \alpha_{III} P_{i,III}}_{\text{effect of the first factor}} + \underbrace{\gamma_B t_{i,B} + \gamma_C t_{i,C} + \gamma_D t_{i,D}}_{\text{effect of the second factor}} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

Let us compute the expectation for units in a group treatment/poison:

- if poison = I and treatment = A
  these are the reference groups for which we removed the dummy
  here $\mathbb{E}[Y_i] = \mu$ $\quad$ intercept
- if poison = I and treatment = B
  $\mathbb{E}[Y_i] = \mu + \gamma_B$
- if poison = II and treatment = A
  $\mathbb{E}[Y_i] = \mu + \alpha_{II}$
- if poison = II and treatment = B
  $\mathbb{E}[Y_i] = \mu + \alpha_{II} + \gamma_B$

Hence, in general:

- μ: mean of the reference group
- α_{II} (and α_{III}): difference in the expected survival between poison II and poison I
  (between poison III and I)
- γ_B (and γ_C and γ_D): difference in the expected survival between treatment B and treatment A
  (between treatment C and A, and between treatment D and A)

Notice that with this formulation, the effect of each poison and of each treatment is fixed

With this model, both factors have an individual additive effect.

- Suppose we want to test whether different types of poison do not have different effects:
  test on a subset of coefficients
  $$\begin{cases} H_0: \alpha_{II} = \alpha_{III} = 0 \\ H_1: \overline{H_0} \end{cases}$$
  The reduced model in this case assumes that
  $$Y_i = \mu + \gamma_B t_{i,B} + \gamma_C t_{i,C} + \gamma_D t_{i,D} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$
  If we do not reject H_0, all poisons have the same effect.

- Similarly, if we want to test whether all treatments are equal:
  $$\begin{cases} H_0: \gamma_B = \gamma_C = \gamma_D = 0 \\ H_1: \overline{H_0} \end{cases}$$
  The reduced model in this case assumes that
  $$Y_i = \mu + \alpha_{II} P_{i,II} + \alpha_{III} P_{i,III} + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$
  If we do not reject H_0, all treatments have the same effect.

- Finally, if we want to test whether nor the poison nor the treatment type have different effects:
  $$\begin{cases} H_0: \alpha_{II} = \alpha_{III} = \gamma_B = \gamma_C = \gamma_D = 0 \\ H_1: \overline{H_0} \end{cases}$$
  And the reduced model is
  $$Y_i = \mu + \varepsilon_i \qquad \varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$$

TWO-WAY ANOVA WITH INTERACTION

Consider the same dummy variables defined before $(P_{i,II}\,;\ P_{i,III}\,;\ t_{i,B}\,;\ t_{i,C}\,;\ t_{i,D})$
Now we need to take into account every possible combination of poison/treatment.
Interaction is modeled by products of the dummy variables:

$$\begin{aligned} Y_i = \mu &+ \alpha_{II} P_{i,II} + \alpha_{III} P_{i,III} + \\ &+ \gamma_B t_{i,B} + \gamma_C t_{i,C} + \gamma_D t_{i,D} + \\ &+ \delta_1 \cdot P_{i,II} \cdot t_{i,B} + \delta_2 \cdot P_{i,III} \cdot t_{i,B} + \delta_3 \cdot P_{i,II} \cdot t_{i,C} + \delta_4 \cdot P_{i,III} \cdot t_{i,C} + \delta_5 \cdot P_{i,II} \cdot t_{i,D} + \delta_6 \cdot P_{i,III} \cdot t_{i,D} + \varepsilon_i \end{aligned}$$

The total number of parameters here is $1 + (K-1) + (J-1) + (K-1)(J-1) = 1 + 2 + 3 + 2\cdot 3 = 12 = J \cdot K$.
Hence now we have one parameter for each group (combination poison/treatment).

Notice that the two-way ANOVA model without interaction is nested.
Hence we can test the absence of interaction as

$$\begin{cases} H_0: \delta_1 = \delta_2 = \delta_3 = \delta_4 = \delta_5 = \delta_6 = 0 \\ H_1: \overline{H_0} \end{cases}$$