

## Statistical modeling for phenotypic traits

1. First linear models
2. Linear models with Plant\_type
3. Linear models with asreml library
4. Linear models with Soil variable
5. Linear models with Soil variable with Plant\_type

# UCPH\_StatisticalAnalysis

Elise

2024-06-09

## Statistical modeling for phenotypic traits

The explanatory variable (X) refers to the genotype, which is a categorical variable. The response variables (Y) are the phenotypic data of the variables dataframe.

In this case, the variables are:

And the genotypes are:

```
print(variables)
```

```
## [1] "FW_shoot_g"      "Plant_height_cm"
```

```
unique(endpoint$Genotype)
```

```
## [1] EPPN2_H EPPN_T  EPPN1_H EPPN1_L EPPN3_L EPPN4_H EPPN2_L EPPN3_H EPPN4_L  
## 9 Levels: EPPN_T EPPN1_H EPPN1_L EPPN2_H EPPN2_L EPPN3_H EPPN3_L ... EPPN4_L
```

### 1. First linear models

Firstly, we model the  $Y = X + r + c + e$  Where - Y is the phenotypic trait; - X the genotype (fixed effect); - r the row effect (fixed or random); - c the column effect (fixed or random);

```

check_factor_levels <- function(data, factors) {
  for (factor in factors) {
    num_levels <- nlevels(as.factor(data[[factor]]))
    if (num_levels < 2) {
      stop(paste("Factor", factor, "must have at least two levels but has", num_levels))
    } else {
      print(paste("Factor", factor, "has", num_levels, "levels"))
    }
  }
}

# Example of factors to check in your dataset
factors_to_check <- c("Genotype", "Row", "Column", "Soil")
check_factor_levels(endpoint_clean, factors_to_check)

```

```

## [1] "Factor Genotype has 9 levels"
## [1] "Factor Row has 12 levels"
## [1] "Factor Column has 9 levels"
## [1] "Factor Soil has 2 levels"

```

```

print("Erreur dans `contrasts<-`(`*tmp*`, value = contr.funs[1 + isOF[nn]]) :
      les contrastes ne peuvent être appliqués qu'aux facteurs ayant au moins deux niveaux")

```

```

## [1] "Erreur dans `contrasts<-`(`*tmp*`, value = contr.funs[1 + isOF[nn]]) : \n les
      contrastes ne peuvent être appliqués qu'aux facteurs ayant au moins deux niveaux"

```

## 2. Linear models with Plant\_type

Model with X as Plant\_type instead of Genotype, and row and column effects as random effects.  
Plant\_type is defined as H for Hybrid, L for pure Line and T for Tester.

## 3. Linear models with asreml library

```

modasreml <- asreml(fixed = DW_shoot_g ~ Genotype,
  random = ~ Row + Column,
  residual = ~ NULL,
  data = endpoint_clean)

```

```

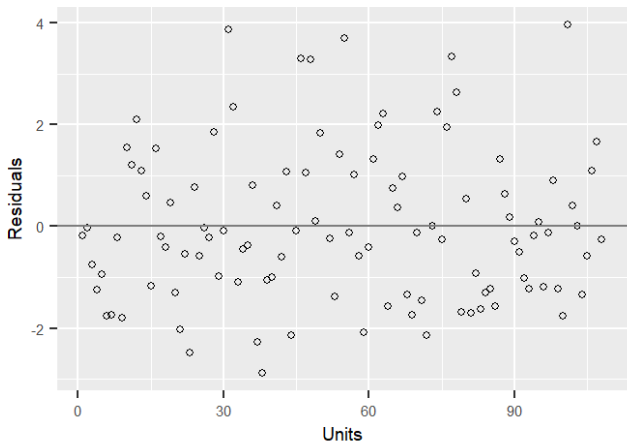
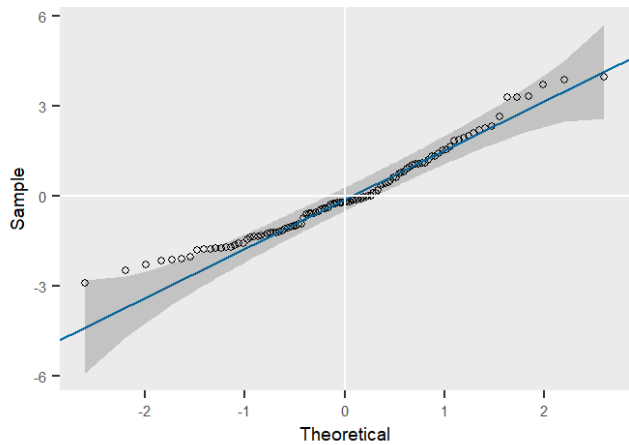
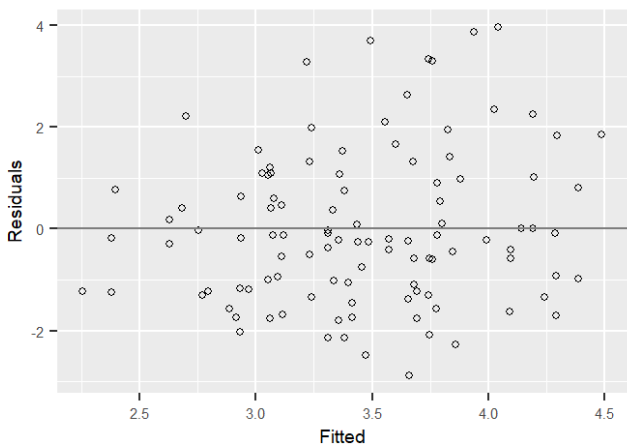
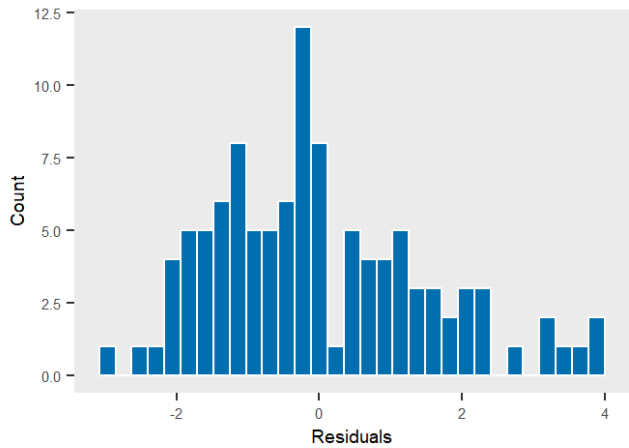
## ASReml Version 4.2 09/06/2024 17:04:10
##           LogLik      Sigma2    DF    wall
##  1      -110.8407      2.459644   98  17:04:10
##  2      -109.7897      2.537136   98  17:04:10 ( 1 restrained)
##  3      -109.6061      2.567283   98  17:04:10 ( 1 restrained)
##  4      -109.5973      2.570370   98  17:04:10 ( 1 restrained)
##  5      -109.5968      2.570553   98  17:04:10 ( 1 restrained)

```

```

plot(modasreml)

```



```
summary(modasrem1)$varcomp
```

##	component	std.error	z.ratio	bound	%ch
## Column	2.687275e-01	0.2464866	1.090232	P	0
## Row	8.472656e-07	NA	NA	B	NA
## units!R	2.570553e+00	0.4046865	6.351961	P	0

4. Linear models with Soil variable

Model with Soil as explicative variable.

5. Linear models with Soil variable with Plant\_type