

Exercise 26.37 of PSL3e

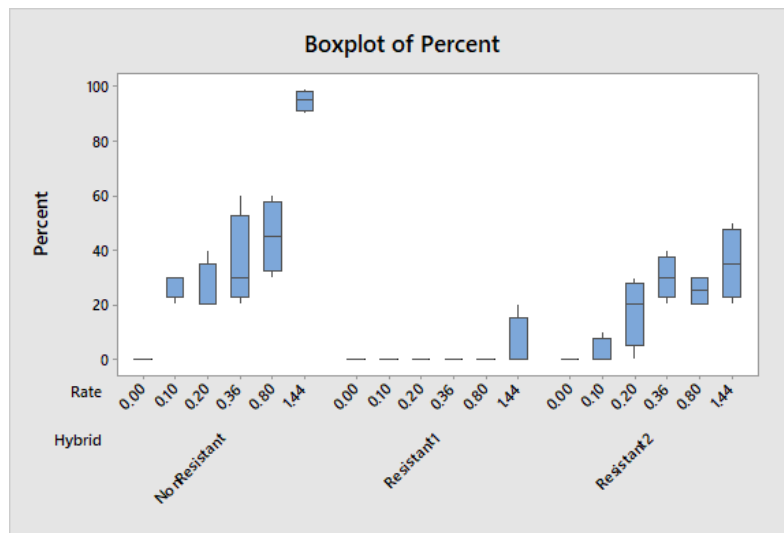
Data: Resistance of corn hybrids (produced by genetic engineering) to a certain herbicide, glusofinate. In this study, resistance is indicated by low degrees (percents) of leaf burn (necrosis) after application of the herbicide. Three corn hybrids (two resistant and one non-resistant) were used, and these should presumably be considered as blocks (because variety cannot be assigned). Six different application rates (doses) of the herbicide were used, ranging from zero (control) to 1.44 kg/ha. The dataset has four replications per combination of application rate and hybrid; it is not described what these are, or what the observational unit is (e.g. plant or plot).

Model: The two-way ANOVA model for these data can be written:

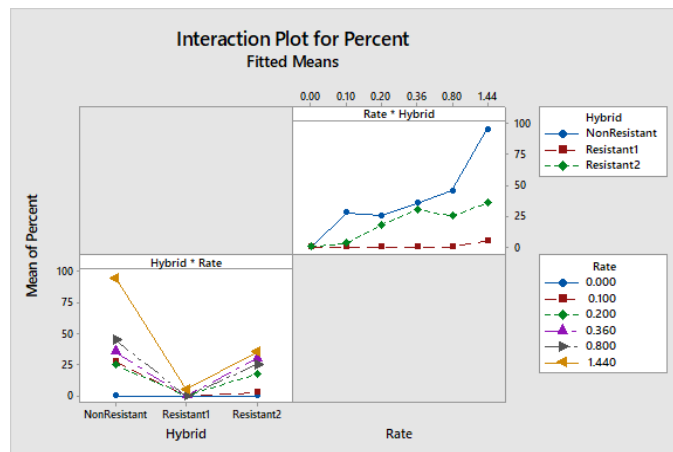
$$\text{percent}_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where the errors (ε_{ijk}) are i.i.d. from $N(0, \sigma)$, and the indices i and j correspond to the row (application rate) and column (corn hybrid) factors, with $I = 6$, $J = 3$, $n_{ij} = 4$ and $N = 72$.

The requested means and graph for (a) can be obtained as part of the two-way ANOVA analysis, but we start with a descriptive plot of the raw data. Even with only 4 observations per group, a summarized graphical display (per group) such as the box plots can be useful.



Many of the boxes are rather strangely shown as a horizontal line, but this reflects that the observations are all equal (and in this case zero). We note immediately that the hybrid **Resistant1** has hardly any non-zero observations. We carry out the analysis in the ANOVA-General Linear Model menu (with some parts of the output omitted) to get better access to means, factorial plots and residuals.



General Linear Model: Percent versus Hybrid, Rate

Factor Information

Factor	Type	Levels	Values
Hybrid	Fixed	3	NonResistant, Resistant1, Resistant2
Rate	Fixed	6	0.00, 0.10, 0.20, 0.36, 0.80, 1.44

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Hybrid	2	16482	8240.85	128.40	0.000
Rate	5	13955	2791.01	43.49	0.000
Hybrid*Rate	10	10318	1031.85	16.08	0.000
Error	54	3466	64.18		
Total	71	44221			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8.01128	92.16%	89.70%	86.07%

Fits and Diagnostics for Unusual Observations

Obs	Percent	Fit	Resid	Std Resid	
9	40.00	25.00	15.00	2.16	R
13	60.00	35.00	25.00	3.60	R
15	20.00	35.00	-15.00	-2.16	R
17	60.00	45.00	15.00	2.16	R
20	30.00	45.00	-15.00	-2.16	R
48	20.00	5.00	15.00	2.16	R
57	0.00	17.50	-17.50	-2.52	R
69	50.00	35.00	15.00	2.16	R
71	20.00	35.00	-15.00	-2.16	R

R Large residual

Means

Term	Fitted Mean	SE Mean
Hybrid*Rate		
NonResistant 0.00	0.00	4.01
NonResistant 0.10	27.50	4.01
NonResistant 0.20	25.00	4.01
NonResistant 0.36	35.00	4.01
NonResistant 0.80	45.00	4.01
NonResistant 1.44	94.75	4.01
Resistant1 0.00	0.00	4.01
Resistant1 0.10	-0.00	4.01
Resistant1 0.20	-0.00	4.01
Resistant1 0.36	-0.00	4.01
Resistant1 0.80	-0.00	4.01
Resistant1 1.44	5.00	4.01
Resistant2 0.00	0.00	4.01
Resistant2 0.10	2.50	4.01
Resistant2 0.20	17.50	4.01
Resistant2 0.36	30.00	4.01
Resistant2 0.80	25.00	4.01
Resistant2 1.44	35.00	4.01

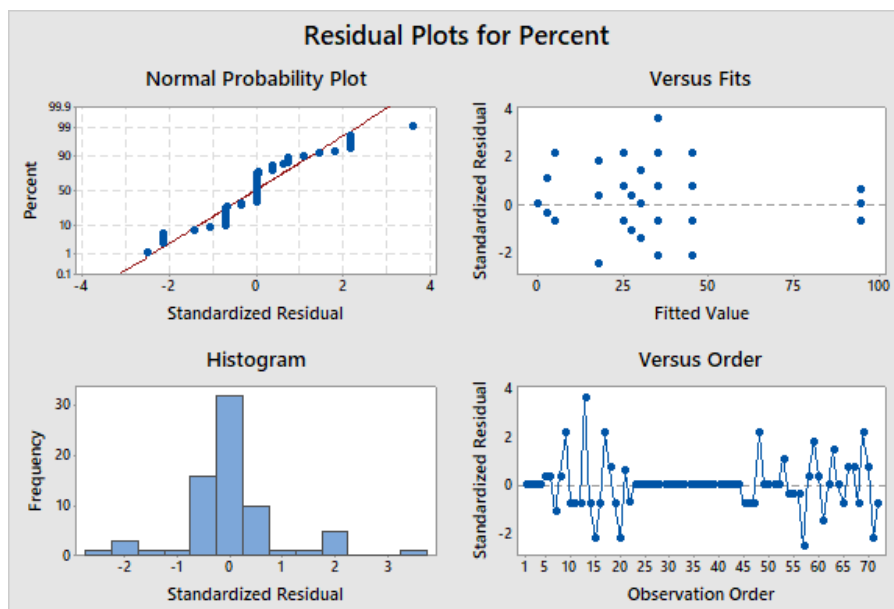
(a) The interaction plot with rate on the x -axis and curves for each hybrid shows three clearly non-parallel curves. One of these (for the hybrid **Resistant1**) is essentially flat at zero, and we already discovered the reason for this from the boxplot. The two other curves show increasing values with rate, both of them starting at zero for rate zero (which is logical: without any herbicide application there should not be any leaf burn). The curve for the non-resistant hybrid has larger means across all non-zero rates, with the largest differences at rates 0.10 and 1.44.

In this case, there does not seem any value in further describing the main effects because all the relevant information can be read off the interaction plot.

(b) The two-way ANOVA model has the assumptions:

- normal distribution of errors,
- same standard deviations (or variance) of errors across the data,
- independence of the errors (or observations),
- the same mean for the 4 observations within each combination of hybrid and rate.

Our primary tools for model checking are different residuals plots, shown below in Minitab's default display.



The four-in-one residual panel shows several unusual features. First, the normal probability plot has a quite discrete character, with many replicates of residuals with the same value. In particular, there seems to be many residuals equal to zero. The histogram has a very strong peak at zero and the plot against order shows a block of zeros between 20 and 40. It is clear that a normal distribution does not describe the residuals well. Also the plots of standardized residuals against fitted values shows discrete patterns. The observation order here is determined by the factor and does therefore not have any additional meaning, but it is clear that very many residuals are equal to zero (in fact, 30 out of 72 values). The reason for this is that 7 out of the 18 groups have only zero percents. Among the residuals, there is also one quite large value (3.60 for observation 13).

This leads to the main issues with the model assumptions: the data are too discrete, and the variance is not at all constant across the data because large parts of the data show no variation (in particular, the corn hybrid **Resistant1**). Therefore the two-way ANOVA model does not meet its assumptions, and the validity of the analysis is uncertain (to say the least).

As an addendum, we briefly discuss some possible approaches for analysis of these data. One option is to average the replicates (whose meaning is not clear from the description provided anyway) to a single value per cell, and then use Friedman's test. That could give an overall comparison of the six application rates within the three corn hybrids, or conversely a comparison of the three corn hybrids within the six application rates. However, the interaction cannot be explored after averaging, so this does not seem very attractive.

Another option could be to restrict the analysis to the two other hybrids than **Resistant1** (which anyway shows clearly different results, for which no statistical assessment seems needed) and to non-zero rates. This improves the residual plots considerably, and the model assumptions may be met to an acceptable degree (not shown).