ANOVA

1. **ANOVA:**

To do ANOVA in R, use:

* *Name of your model = aov (data~group, data=Name of your data)*
* *summary(Name of your model)*

If you reject the null hypothesis, you need to do Tukey test as below:

*TukeyHSD(Name of your model, conf.level= …..)*

*plot(TukeyHSD(Name of your model, conf.level= …..), las=1 or 2)*

To do residual plots, you first need to compute the residual as:

*Name of your residual = resid(lm(data~group, data=Name of your data))*

You can plot the residual against the treatment as:

*stripchart(Residuals~group)*

You can also compute the fitted values as:

*Name of your fitted values = fitted(lm(data~group, data=Name of your data))*

Another way to get the all the residual diagnostic plots is to use:

*plot(name of your model)*

Example 1: The aim of a randomized experiment was to estimate the effect of the fatty acid CPFA on the level of a certain protein in rat livers. The CPFA levels were set at at 50, 150, 300, 450, 600 (ignore the units). Only one level of the CPFA could be investigated in a day’s work, so a control group (no CPFA) was investigated each day as well. The data are presented in the table below.

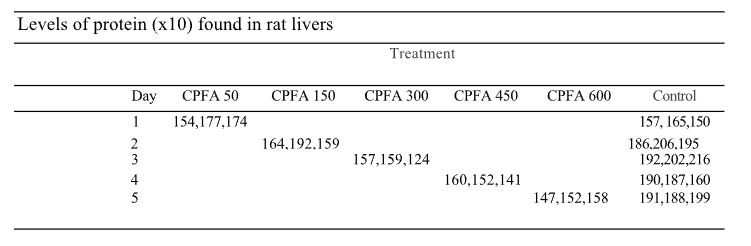


Table 1: CPFA level

Test the null hypothesis that the mean levels of proteins are the same for all ten treatment-day combinations. Then, do the Tukey analysis if the null hypothesis is rejected.

Check the assumptions of the model plot plotting the residuals against the treatment, the residuals against the fitted values, and the QQ plot of the residuals.

Example 1: (cont’d) A posteriori, is there evidence that the use of control groups on different days was necessary? You are being asked to compare the model with ten different means (the complete model) with a reduced model consisting of six different means.

First, you will need to recode the Treatment column so that all 5 control groups are lumped into one group (simply just rename them in Excel and save the file as a different name). Then, perform ANOVA analysis on this new data.

Compare this new reduced model with the original model. To find the extra sum of square, compute:

Extra SS = SSE (Reduced model)-SSE (Full model)

DF (extra SS)= DF (SSE of Reduced model)- DF (SSE of Full model)

To ask R to do it, use:

*anova(Name of your reduced mode, Name of your full model)*

If the test is significant, we reject the reduced model and accept the full model. In plain terms, keeping

lumping the five controls into a single treatment increases significantly the random error.

1. **Kruskal-Wallis test:**

To do KW test, use:

*kruskal.test(data~group,data=Name of your data)*

Example 2: Redo example 1 using KW test.