Lab 6 self assessment - PMD's answers

There are various reasonably appropriate analyses and some very wrong analyses. I indicate how I would analyze the data and why. I try to indicate other reasonably appropriate analyses. My answer is structured as a set of questions, with the answers I see in the data. These follow the outline in the week 6 lecture summary. Results from each part of the analysis are in blue.

Study Design:

 experimental or observational?

 observational - genotypes are characteristics of plants, not randomly assigned

(aside: Genetic knowledge suggests that expression of the target gene is the only difference the knockout genotype and the WT. This non-statistical information supports making causal claims)

 paired or not?

 not paired - no evidence of a connection between a knockout plant and a specific WT plant

 problems with independence?

 yes - cluster effects. eu = plant, ou = measurement on a plant. 12 eu's, 36 ou's.

I averaged the three measurements per plant so there are 12 rows of data.

What is the question? i.e., what is the goal of the study:

 Want to know about differences in location.

 Part 1 wants a p-value; part 2 wants a confidence interval

 Note: part 1 asks about equal means; that may not be possible.

Assumptions: all based on 12 observations, so evidence is limited

 already dealt with independence

 equal variances:



 Sure looks like unequal variances.

 numbers support this: ratio of sd's is 3.7. (sd's are 0.153 and 0.572)

 normality:

 .



 Looks pretty good, except for one unusually large value.

 Note: the QQ plot is based on the residuals, not the collection of 12 observations.

My suggestion is to carefully examine one large value - any error or relevant difference?

 You don't have any of that information, so we'll stick with analyzing all 12 observations.

Because of the unequal variances and apparent outlier, I would follow one of two paths to a test:

 1) a non-parametric test to be resistant to the outlier

 2) evaluate log transformed values

 2) would be needed if we want a confidence interval

Assumptions on log transformed alanine concentrations: Look good! Notice "outlier" gone!



Answers to the study questions:

Part 1: p-values for the comparison

using a non-parametric test:

 If an answer in terms of medians is acceptable, do a Wilcoxon rank-sum test

 p = 0.026. Evidence of a difference in medians

using a t-test on log transformed values:

 p = 0.016. Evidence of a difference in medians

 Note: If we additionally assume normal, constant variance errors, this is a difference is means.

Part2: Confidence interval for the effect of the knockout

 Need to use t-methods to get a confidence interval.

 apply to log transformed values to better satisfy assumptions. Result will be a multiplicative effect.

 on the log scale, the estimated difference is -1.04 (HM - WT), with 95% CI of (-1.84, -0.24)

 The estimated multiplicative effects are 0.35 with 95% CI of (0.16, 0.78).

 The median alanine concentration in the knockout genotype is 35% of that in the wildtype

 (95% CI: 16%, 78%).

 Note: Sometimes it is clearly (biologically) to express as a drop from reference.

 Here the knockout is 65% less than the wildtype.