Problem Set 1

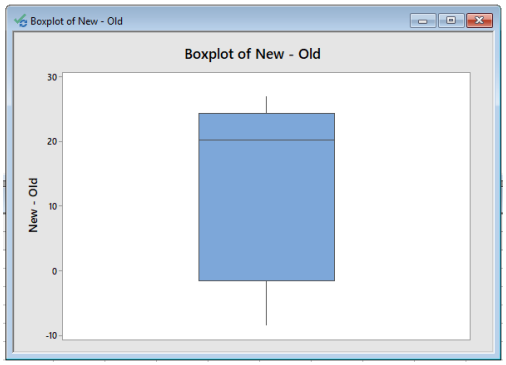
Nick Handelman
2/19/2018

1 Problem 1

1.1 Hypotheses

Old: weight (in pounds) of grass clippings from grass grown with the old fertilizer
New: weight (in pounds) of grass clippings from grass grown with the new fertilizer
$H_0 : \mu_{new} = \mu_{old}$
$H_a : \mu_{new} > \mu_{old}$

1.2 Boxplot and Summary Statistics - Minitab Output



Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
New	20	205.36	4.77	1.07
Old	20	192.03	10.54	2.36

Estimation for Paired Difference

Mean	StDev	SE Mean	95% Lower Bound for μ_difference
13.32	12.87	2.88	8.35

μ_difference: mean of (New - Old)

Test

Null hypothesis H_0: μ_difference = 0
Alternative hypothesis H_1: μ_difference > 0

T-Value	P-Value
4.63	0.000

1.3 Choice of Statistical Technique

A one-sided paired t-test is most appropriate for the problem. The goal of the analysis is to "evaluate whether the new fertilizer tends to produce *more* grass than the old fertilizer." So a one sided test is appropriate. While each fertilizer is tested on different plots of land, the plots of land are adjacent and are subject to the same environmental conditions. Thus, I believe this is analogous to testing the fertilizers on the same plots of land. So, a paired t-test is appropriate.
The boxplot shows that the differences are skewed. However, with a sample size of 20, the paired t-test is robust to the skewness. The paired t-test is not resistant to outliers. The boxplot shows there are no outliers, so this is not a concern. The fertilizers are used on plots of land adjacent to each other, so I'm slightly concerned that each plot could be contaminated by fertilizer from the adjacent plot. However, if this were a real concern, I would expect the boxplot to be much more tightly clustered around 0.

1.4 Conclusions

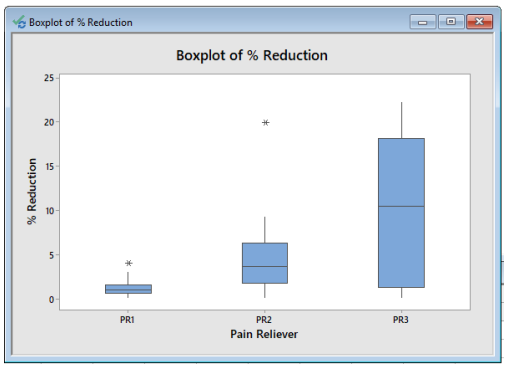
With a 95% confidence level and a p-value of less than 0.001, I reject H_0 and accept that the mean difference is greater than 0. Randomization is not used in the study design, so it is not appropriate to say that the new fertilizer caused the increase in pounds of grass clippings. It is appropriate to say that the new fertilizer strongly correlates to increased pounds of grass clippings.

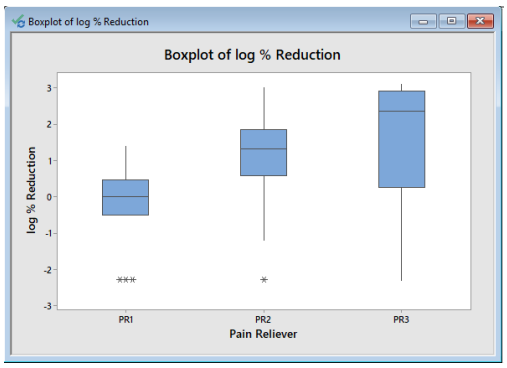
2 Problem 2

2.1 Hypotheses

$H_0 : \mu_{log(PR1)} = \mu_{log(PR2)} = \mu_{log(PR3)}$
H_a : at least one of the means differs from the others

2.2 Boxplots and Summary Statistics - Minitab Output





Descriptive Statistics: log % Reduction

Statistics

Variable	Pain Reliever	N	Mean	StDev	Minimum	Maximum
log % Reduction	PR1	15	-0.306	1.169	-2.303	1.386
	PR2	15	0.989	1.460	-2.303	2.991
	PR3	15	1.622	1.621	-2.303	3.105

One-way ANOVA: log % Reduction versus Pain Reliever

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Pain Reliever	2	28.98	14.489	7.10	0.002
Error	42	85.75	2.042		
Total	44	114.73			

Means

Pain Reliever	N	Mean	StDev	95% CI
PR1	15	-0.306	1.169	(-1.050, 0.439)
PR2	15	0.989	1.460	(0.245, 1.734)
PR3	15	1.622	1.621	(0.878, 2.367)

Pooled StDev = 1.42890

2.3 Choice of Statistical Technique

A One-Way ANOVA on the given data is my first thought. From the % reduction box plots, all of the groups are approximately normal, but they violate the ANOVA assumption of equal variance. The % reduction boxplots show that the variance increases as the mean increases across groups, so I performed a log transformation on the data. The log % reduction boxplots show that the transformed data have closer variances. While the transformed data for PR3 is quite skewed, this is acceptable since the group sample sizes are moderate and equal. The transformed PR1 and PR2 have outliers, but they are all within 2 standard deviations of their means so I don't consider them severe. From the problem description, I see no reason to think that independence within or between groups has been violated. Thus, a One-Way ANOVA on the log transformed data is appropriate.
Minitab's One-Way ANOVA has an option to "assume equal variances". I ran a test for equal variances and the p-values of the multiple comparisons test and Levene's test were 0.652 and 0.820, respectively. At a 95% confidence level, I did not reject the equal variances test null hypothesis that all variances are equal. So, I ran the One-Way ANOVA assuming equal variances.

2.4 Conclusions

With a 95% confidence level and a p-value of less than 0.003, I reject H_0 and accept that, for the transformed data, at least one of the group means differs from the others. If the transformed group distributions were

all symmetric, the claim could be made that, for the untransformed data, at least one of the group medians differs from the others. However, the boxplots indicate that the transformed PR1 and PR3 distributions are non-normal. So, that claim is questionable. The patients are assigned randomly to each of the three groups, so it is acceptable to make the claim that the different pain relievers are the cause of rejecting H_0.