LSD for crop yield

Pranav Gopalkrishna, 1940223

2021-09-22

# LSD model definition

Latin square design or LSD is an experimental design. Hence, it is applied for the data collection process, not the analysis. To do this, we must have at least three factors and a response. For best results, all factors must have equal number of levels. Two of these factors may be usable in classify the experimental units. If yes, then these are the blocking factors. Together, they form a grid for experimental units to sit in. The third factor is considered as the treatment, and each level is applied randomly to the experimental units so that every cell of th grid i.e. blocking factor combo is subjected to every treatment at only once. It is assumed that there is no interaction between any of the factors.

# Data

This data set contains crop yield data, wherein crops are grown in 3 different plots, A, B and C, and there are three levels of soil density (low, medium, high). These are the blocking factors.

On top of this, we also have fertilizer types being applied to each plot of crops. Every plot and soil density category have been treated by every fertilizer type once.  
setwd("~/Documents/Study/computerScience/programming/r/data/")  
myData = read.csv("cropYieldTruncated.csv")[-3]  
head(myData)

## yield block density fertilizer  
## 1 90 A low N  
## 2 95 A low P  
## 3 107 A low NP  
## 4 92 A medium N  
## 5 89 A medium P  
## 6 92 A medium NP

# Goal

Find out the significance of fertilizer type’s effect on yield, with plot and soil density as the blocking factors. This is to account for the effect soil density may have on the crops, and the potential unaccounted effects on the crops in different plots.

b1 = myData$block  
b2 = myData$density  
t = myData$fertilizer  
y = myData$yield  
myData = data.frame(y, t, b1, b2)

# ANOVA test

model = lm(y~., myData)  
aovModel = aov(model)  
aovModel

## Call:  
## aov(formula = model)  
##   
## Terms:  
## t b1 b2 Residuals  
## Sum of Squares 1977.444 1031.781 1689.439 19057.779  
## Deg. of Freedom 2 2 2 65  
##   
## Residual standard error: 17.12298  
## Estimated effects may be unbalanced

# Conclusions

Since the p-value for the treatment (fertilizer type) is below the significance level 0.05, we may reject the null hypothesis for fertilizer type.

# Post hoc analysis

library(lsmeans)

## Loading required package: emmeans

## Warning: package 'emmeans' was built under R version 3.6.2

## The 'lsmeans' package is now basically a front end for 'emmeans'.  
## Users are encouraged to switch the rest of the way.  
## See help('transition') for more information, including how to  
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.

lsmeans(model, "t")

## t lsmean SE df lower.CL upper.CL  
## N 93.3 3.5 65 86.3 100  
## NP 105.3 3.5 65 98.3 112  
## P 100.6 3.5 65 93.6 108  
##   
## Results are averaged over the levels of: b1, b2   
## Confidence level used: 0.95

pairs(lsmeans(aovModel, "t"))

## contrast estimate SE df t.ratio p.value  
## N - NP -12.02 4.96 65 -2.423 0.0472  
## N - P -7.30 4.95 65 -1.476 0.3091  
## NP - P 4.72 4.95 65 0.954 0.6086  
##   
## Results are averaged over the levels of: b1, b2   
## P value adjustment: tukey method for comparing a family of 3 estimates

Given a 0.05 significance level, only N and NP have significantly different means.