Homework2-PracticeAbstract

2023-01-23

An experiment was conducted to (a) investigate the effects of a drug, human granulocyte colony stimulating factor (gcsf), on the growth of certain cells in human blood and (b) to determine if the effect of the drug is influenced by age. Thirty-eight subjects were randomized to one of three dose levels of gcsf: 0, 30, or 300 micrograms. Blood samples were collected at three times: prior to drug administration, one day after the drug treatment, and five days after treatment. The number of cell colonies per ml of blood (colony forming units) was recorded.

The data are provided in an Excel file gcsf.xlsx on the class website. The study variables are: id: subject id number gcsf: drug dose (in 0, 30, or 300 micrograms) age: age of subject in years cfu0: day 0 colony forming units (baseline) cfu1: day 1 colony forming units cfu5: day 5 colony forming units

The above data format is “wide” (e.g., one record per subject). The data are saved in the worksheet, gcsfWide. The dataset is also formatted in “long” format in the worksheet, gcsfLong, with variables: id, days (1, 2, 5), gcsf (0, 30, 300), age and cfu. Like Assignment 01, the objective is to investigate the scientific question or questions. Please analyze these data and report your results in an abstract of at most 500 words. Please also submit one table or figure to support your analysis.

#read in the data  
library(ggplot2)  
library(readxl)  
library(knitr)  
library(car)

## Loading required package: carData

library(rstatix)

##   
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':  
##   
## filter

library(ggpubr)  
library(lme4)

## Loading required package: Matrix

library(lmerTest)

## Warning: package 'lmerTest' was built under R version 4.2.2

##   
## Attaching package: 'lmerTest'

## The following object is masked from 'package:lme4':  
##   
## lmer

## The following object is masked from 'package:stats':  
##   
## step

library(effects)

## Warning: package 'effects' was built under R version 4.2.2

## lattice theme set by effectsTheme()  
## See ?effectsTheme for details.

data\_wide <- read\_excel('gcsf.xlsx', sheet = 'gcsfWide')  
data\_long <- read\_excel('gcsf.xlsx', sheet = 'gcsfLong')

#look at summary statistics  
summary(data\_long)

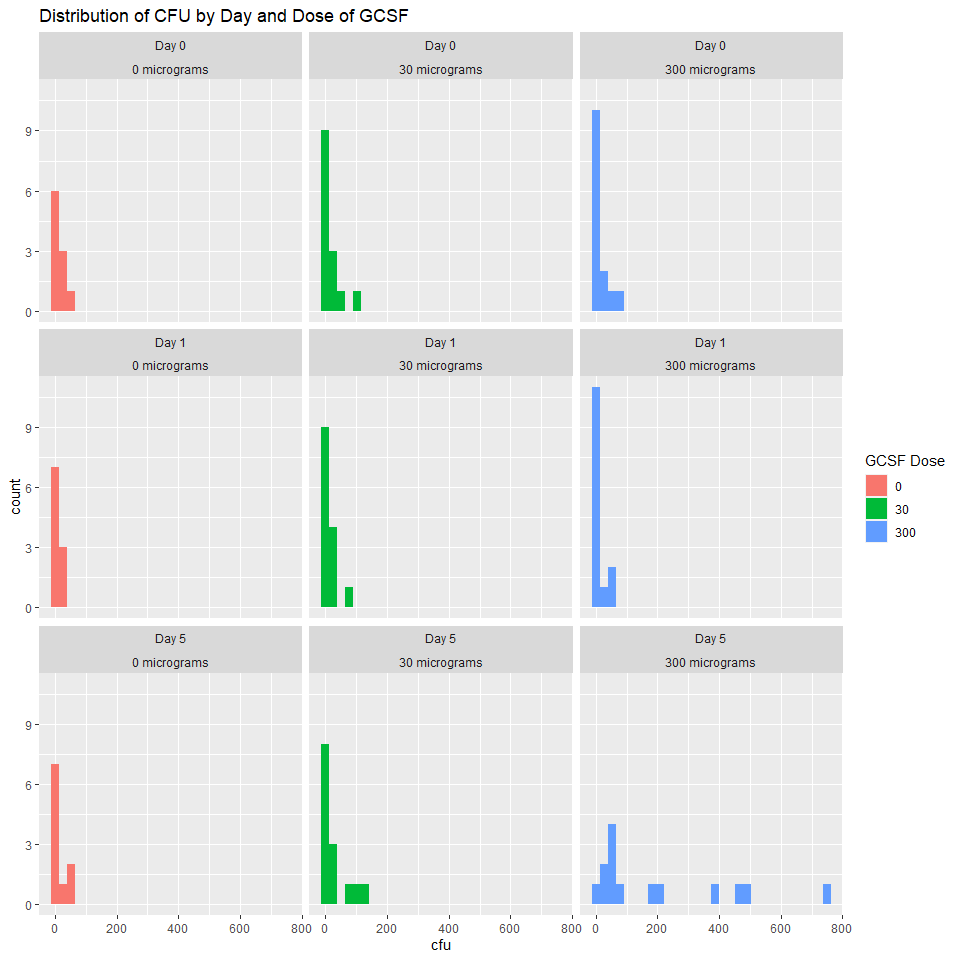
## id days gcsf age cfu   
## Min. : 1.0 Min. :0 Min. : 0.0 Min. :20.00 Min. : -1.00   
## 1st Qu.:10.0 1st Qu.:0 1st Qu.: 0.0 1st Qu.:23.00 1st Qu.: 2.00   
## Median :19.5 Median :1 Median : 30.0 Median :50.00 Median : 8.50   
## Mean :19.5 Mean :2 Mean :121.6 Mean :48.58 Mean : 38.72   
## 3rd Qu.:29.0 3rd Qu.:5 3rd Qu.:300.0 3rd Qu.:74.00 3rd Qu.: 29.75   
## Max. :38.0 Max. :5 Max. :300.0 Max. :79.00 Max. :749.00

#change id, gcsf, and days to factor variables  
data\_long$gcsf\_c <- as.factor(data\_long$gcsf)  
data\_long$days\_c <- as.factor(data\_long$days)  
data\_long$id\_c <- as.factor(data\_long$id)  
  
#create label variable for plot  
data\_long$day\_label <- NA  
data\_long$day\_label[data\_long$days == 0] <- 'Day 0'  
data\_long$day\_label[data\_long$days == 1] <- 'Day 1'  
data\_long$day\_label[data\_long$days == 5] <- 'Day 5'  
  
data\_long$dose\_label <- NA  
data\_long$dose\_label[data\_long$gcsf == 0] <- '0 micrograms'  
data\_long$dose\_label[data\_long$gcsf == 30] <- '30 micrograms'  
data\_long$dose\_label[data\_long$gcsf == 300] <- '300 micrograms'

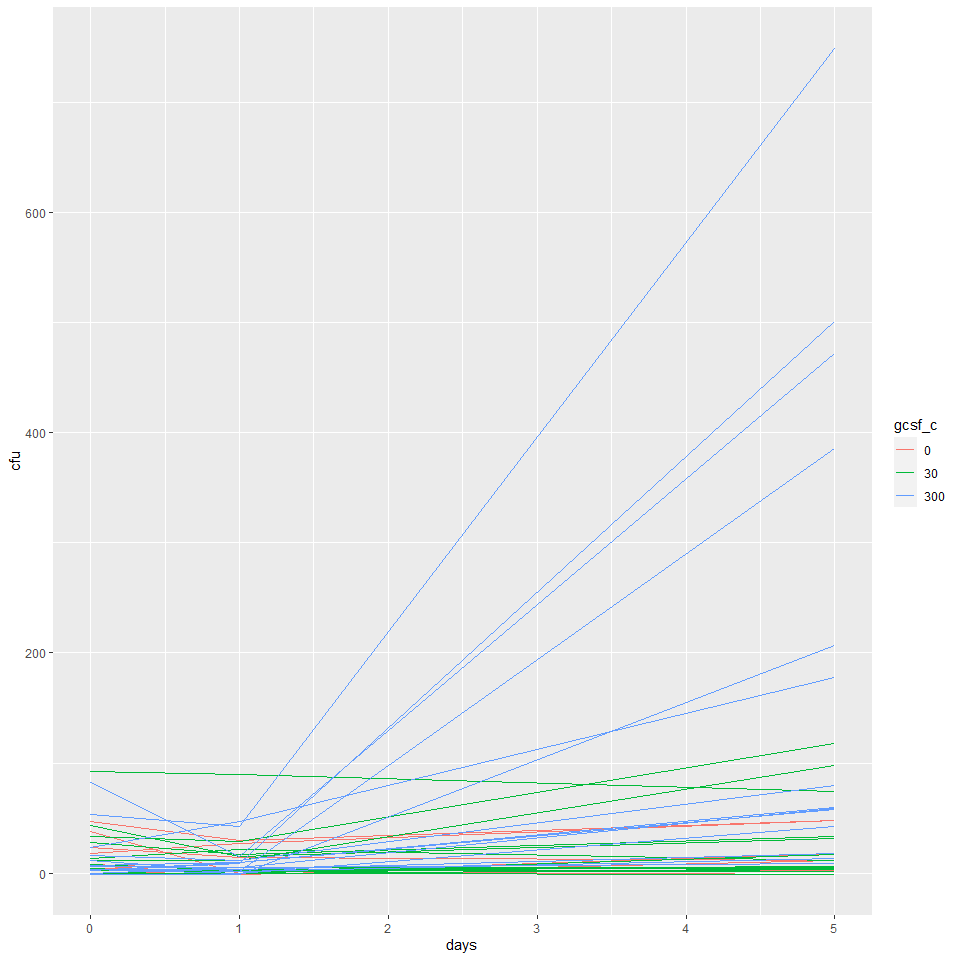
With this data, a repeated measure ANCOVA could be fit to test for difference in means with the cfu as the output variable and gcsf, age, time, and id as input variables. The assumptions of repeated measures ANCOVA include normality of the groups.

#look at distribution of cfu by treatment group  
ggplot(aes(x = cfu, fill = gcsf\_c), data = as.data.frame(data\_long)) +  
 geom\_histogram() +   
 facet\_wrap(c('day\_label','dose\_label')) +  
 ggtitle("Distribution of CFU by Day and Dose of GCSF") +  
 labs(fill = "GCSF Dose")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#spaghetti plot  
ggplot(data = data\_long, aes(x = days, y = cfu, group = id, colour = gcsf\_c)) + geom\_line()



Based on the above plot, the normality assumption of repeated measures ANCOVA is not satisfied. As an alternate approach, two separate non-parametric tests will be used to determine if the change in

#get summary stats  
summary\_stats <- data\_long %>%  
 group\_by(days, gcsf) %>%  
 get\_summary\_stats(cfu, type = "median\_iqr")  
  
overall <- data\_long %>% group\_by(gcsf) %>% get\_summary\_stats(cfu, type = "median\_iqr")  
  
summary\_stats[10:12,] <- data.frame(days = rep(NA,3), overall)  
  
  
#get summary of age by treatment group  
summary\_age<- data\_long %>%  
 group\_by(gcsf) %>%  
 get\_summary\_stats(age, type = "median\_iqr")

#create summary table  
table1 <- data.frame(summary\_stats) %>%   
 select(-'variable') %>%   
 reshape(idvar = "gcsf",  
 timevar = "days",  
 direction = "wide")

## Warning in reshapeWide(data, idvar = idvar, timevar = timevar, varying =  
## varying, : there are records with missing times, which will be dropped.

#table1$median.Overall <- overall$median  
#table1$iqr.Overall <- overall$iqr  
  
table1\_1 <- table1 %>% select(-c('n.1','n.5','n.NA'))  
  
table1\_1$Age\_Median <- summary\_age$median  
table1\_1$Age\_IQR <- summary\_age$iqr  
  
#reorder columns  
table1\_1 <- table1\_1 %>% select(gcsf, n.0, Age\_Median, Age\_IQR,  
 median.0, iqr.0,  
 median.1, iqr.1,  
 median.5, iqr.5,  
 median.NA, iqr.NA)  
  
#rename columns for table  
colnames(table1\_1) <- c("GCSF Dose (micrograms)","N",  
 'Age Median', 'Age IQR',  
 "Day 0 CFU Median", "Day 0 CFU IQR",  
 "Day 1 CFU Median", "Day 1 CFU IQR",  
 "Day 5 CFU Median", "Day 5 CFU IQR",  
 "Overall CFU Median", "Overall CFU IQR"  
 )  
  
kable(table1\_1, caption = 'Table 1: Summary of CFU by Day and Treatment Group')

Table 1: Summary of CFU by Day and Treatment Group

| GCSF Dose (micrograms) | N | Age Median | Age IQR | Day 0 CFU Median | Day 0 CFU IQR | Day 1 CFU Median | Day 1 CFU IQR | Day 5 CFU Median | Day 5 CFU IQR | Overall CFU Median | Overall CFU IQR |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 10 | 49 | 51 | 7.5 | 19.25 | 2.5 | 10.00 | 6.5 | 14.25 | 3.5 | 16.00 |
| 30 | 14 | 50 | 52 | 7.0 | 22.00 | 4.0 | 15.25 | 8.5 | 29.25 | 6.0 | 24.25 |
| 300 | 14 | 50 | 49 | 4.5 | 12.50 | 7.0 | 9.25 | 70.0 | 295.25 | 11.0 | 53.75 |

#perform analysis - mixed effects  
#unique intercept by pt id  
#unique treatment effect per measurement time  
#age and treatment interaction  
model1 <- lmerTest::lmer(cfu ~ age\*gcsf\_c + (1+gcsf\_c|days\_c) + (1|id\_c), data = data\_long)

## boundary (singular) fit: see help('isSingular')

summary(model1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: cfu ~ age \* gcsf\_c + (1 + gcsf\_c | days\_c) + (1 | id\_c)  
## Data: data\_long  
##   
## REML criterion at convergence: 1308  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.1892 -0.1642 -0.0322 0.1306 6.0528   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## id\_c (Intercept) 1206.788 34.739   
## days\_c (Intercept) 2.805 1.675   
## gcsf\_c30 33.676 5.803 1.00   
## gcsf\_c300 11147.196 105.580 1.00 1.00  
## Residual 5940.207 77.073   
## Number of obs: 114, groups: id\_c, 38; days\_c, 3  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)  
## (Intercept) 11.47490 38.49812 31.95222 0.298 0.768  
## age 0.01907 0.70299 31.99838 0.027 0.979  
## gcsf\_c30 33.56644 50.18004 31.65988 0.669 0.508  
## gcsf\_c300 43.94247 79.42549 4.74677 0.553 0.605  
## age:gcsf\_c30 -0.53365 0.91239 31.99838 -0.585 0.563  
## age:gcsf\_c300 0.40798 0.93153 31.99838 0.438 0.664  
##   
## Correlation of Fixed Effects:  
## (Intr) age gcs\_30 gc\_300 ag:\_30  
## age -0.886   
## gcsf\_c30 -0.765 0.679   
## gcsf\_c300 -0.465 0.429 0.423   
## ag:gcsf\_c30 0.682 -0.770 -0.882 -0.331   
## ag:gcsf\_300 0.668 -0.755 -0.513 -0.570 0.581  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see help('isSingular')

#model does not converge  
#try no age/treatment interaction  
model2 <- lmerTest::lmer(cfu ~ age + gcsf\_c + (1+gcsf\_c|days\_c) + (1|id\_c), data = data\_long)

## boundary (singular) fit: see help('isSingular')

## Warning: Model failed to converge with 1 negative eigenvalue: -4.6e-03

summary(model2)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [  
## lmerModLmerTest]  
## Formula: cfu ~ age + gcsf\_c + (1 + gcsf\_c | days\_c) + (1 | id\_c)  
## Data: data\_long  
##   
## REML criterion at convergence: 1312.2  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.1017 -0.1578 -0.0726 0.1226 6.1900   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## id\_c (Intercept) 1136.6 33.713   
## days\_c (Intercept) 0.0 0.000   
## gcsf\_c30 55.6 7.457 NaN   
## gcsf\_c300 11506.5 107.268 NaN 1.00  
## Residual 5941.2 77.079   
## Number of obs: 114, groups: id\_c, 38; days\_c, 3  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)  
## (Intercept) 14.55012 24.74197 34.00014 0.588 0.560  
## age -0.04433 0.35740 34.00014 -0.124 0.902  
## gcsf\_c30 7.65079 23.51316 23.05811 0.325 0.748  
## gcsf\_c300 63.79681 66.10480 2.25429 0.965 0.426  
##   
## Correlation of Fixed Effects:  
## (Intr) age gcs\_30  
## age -0.701   
## gcsf\_c30 -0.535 -0.001   
## gcsf\_c300 -0.190 -0.001 0.372  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## boundary (singular) fit: see help('isSingular')

#model does not converge either

#plots would be useful if model converged properly   
  
# plot(Effect("gcsf\_c", model2))  
# plot(Effect("age", model2))  
#   
# lattice::dotplot(ranef(model2))$days\_c

#because the mixed effects model does not converge,   
#consider standard ANCOVA without time = 1  
  
#test assumptions  
#verify covariates are independent of treatment  
anova\_model\_indep <- aov(age ~ gcsf, data = data\_wide)  
anova\_model\_indep2 <- aov(cfu0 ~ gcsf, data = data\_wide)  
#view summary of anova model  
summary(anova\_model\_indep)

## Df Sum Sq Mean Sq F value Pr(>F)  
## gcsf 1 0 0.1 0 0.99  
## Residuals 36 24401 677.8

summary(anova\_model\_indep2)

## Df Sum Sq Mean Sq F value Pr(>F)  
## gcsf 1 11 10.9 0.021 0.885  
## Residuals 36 18293 508.1

#verify homogeneity of variance  
leveneTest(cfu5~as.factor(gcsf), data = data\_wide)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 2 5.574 0.007916 \*\*  
## 35   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#note that this assumption is violated  
 #ANCOVA is fairly robust to assumption violations, so we will proceed.  
  
#conduct ANCOVA  
#interaction model  
ancova\_model <- aov(cfu5 ~ cfu0 + gcsf\*age, data = data\_wide)  
Anova(ancova\_model, type="III")

## Anova Table (Type III tests)  
##   
## Response: cfu5  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 6974 1 0.4255 0.518707   
## cfu0 175314 1 10.6966 0.002516 \*\*  
## gcsf 33037 1 2.0157 0.165055   
## age 37 1 0.0023 0.962359   
## gcsf:age 5681 1 0.3466 0.560028   
## Residuals 540858 33   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#no interaction  
ancova\_model2 <- aov(cfu5 ~ cfu0 + gcsf + age, data = data\_wide)  
Anova(ancova\_model2, type="III")

## Anova Table (Type III tests)  
##   
## Response: cfu5  
## Sum Sq Df F value Pr(>F)   
## (Intercept) 23062 1 1.4347 0.2392888   
## cfu0 181301 1 11.2787 0.0019443 \*\*   
## gcsf 297133 1 18.4845 0.0001362 \*\*\*  
## age 5374 1 0.3343 0.5669309   
## Residuals 546539 34   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Effect of Different Doses of Human Granulocyte Colony Stimulating Factor (GCSF) on Blood Cell Colony Forming Unit (CFU) Count

### Background

Human granulocyte colony stimulating factor (GCSF) is produced by various tissues in the human body and plays a role in the production of blood cells. There are pharmaceutical versions of GCSF that may aid in the production of colony forming units in human blood. This could be useful in helping patients recover from chemotherapy by promoting production of blood cells after chemotherapy halted the production. It is hypothesized that the effect of GCSF may vary by age.

### Approach

Thirty-eight subjects were randomized to one of three dose levels of GCSF (0, 30, or 300 micrograms). Number of cell colonies per ml of blood (CFU) was recorded three times for each patient, before randomization (day 0), one day after treatment (day 1), and five days after treatment (day 5). An ANCOVA model was used to determine if the dose of GCSF effected the CFU from day 0 to day 5 and if there was an age and dose interaction. Summary statistics (median and IQR) were reported for each day and dose level.

### Results

The median CFU on day 5 for 300 microgram dose was much larger than the other days and treatment groups (see Table 1). The difference in median CFU for day 1 vs day 0 for 0, 30, and 300 micrograms was -5.0, -3.0, and -2.5, respectively. The difference in median CFU for day 5 vs day 0 for 0, 30, and 300 micrograms was -1.0, 1.5, and 65.5, respectively. Based on the ANCOVA, there is no evidence of a change in treatment effect by age (p-value = 0.56) after adjusting for dose of GCSF and day 0 CFU. The ANCOVA model was refit without the interaction to more precisely estimate the treatment effect. There is very strong evidence of a difference in CFU by dose of GCSF (p-value = 0.0001) after adjusting for age and day 0 CFU.

### Conclusions

There is no evidence of a difference in treatment effect by age, and there is very strong evidence of a difference in treatment effect by dose. It is clear that at day 5, the 300 microgram dose of GCSF had a much larger effect on CFU count than the 30 and 0 microgram doses.

### References

1. Granulocyte colony stimulating factor (G-CSF). Granulocyte colony stimulating factor (G-CSF) | Cancer information | Cancer Research UK. (2020, September 28). Retrieved January 25, 2023, from <https://www.cancerresearchuk.org/about-cancer/treatment/drugs/g-csf>