BIEN175C : Stat-3 Assignment

Viraj Khatri

2023-06-12

Problem Statement BIEN 175C Stats Assignment 2

Researchers are investigating the effect of Lecanemab, a humanized igG1 monoclonal antibody, to remove amyloid-beta aggregates in individuals diagnosed with Alzheimer's disease. 90 patients were randomly assigned to one of three groups, placebo, treatment with 10mg per kilogram of body weight, or treatment with 10mg per kilogram of body weight every week. Patients were followed up with PET three months from onset of treatment and Centiloid units derived for each. Note: a Centiloid unit (CL) of 0 represents a young healthy individual, a CL of 100 represents a degree of amyloid-beta aggregates similar to those present in mild to moderate dementia due to Alzheimer's disease. Perform the correct statistical analysis on the data below showing your work, generate a question statement, hypotheses statements, and give your interpretation of the data analysis. Show your work. (11 pts)

My answers starts from here

I answer the questions asked in the problem statement here onwards. Wherever required, I write R-codes to do statistical analysis.

For the simplicity of the rest of the analysis, I call placebo group shown in the first column to be Group-0, and the group with 10mg Lecanemab to be Group-1 and the 10 mg Lecanemab per week to be Group-2. This naming make sense because placebo doesn't receive any treatment so 0 also denotes the treatment statues i.e. no treatment. While Group-1 is given treatment once and Group-2 is given treatment every week.

Generating a question statement

The question is basically to ask whether Lecanemab an effect on individuals diagnosed with Alzheimer disease?

Generating a hypothesis statement

The null hypothesis is that Lecanemab does not affect the individuals with Alzheimer meaning that thre is no difference between the treatment groups and the placebo group.

The alternate hypothesis is that it does affect individuals with Alzheimers meaning that there is a significant difference between the treatment groups and the placebo group.

Formally,

Let the average Centiloid unit(CL) of Placebo or Group-0 is μ_0 and for those who got 10 mg per 100kg bodyweigt i.e. Group-1 is μ_1 and for the third group which got 10mg per 100 kg bodyweight per week i.e. Group-2 is μ_2

$$H_0 : \quad \mu_0 = \mu_1 = \mu_2$$

$$H_1 : \quad \mu_0 \neq \mu_1 \quad \text{or} \quad \mu_1 \neq \mu_2 \quad \text{or} \quad \mu_0 \neq \mu_2 \quad \text{or} \quad \mu_0 \neq \mu_1 \neq \mu_2$$

Data Analysis

We can test this hypothesis using ANOVA test which tests the group average of a given variable in multiple groups.

Table 1: CL Readings

Placebo	Lecanemab/2weeks	Lecanemab/week
96	34	11
83	20	11
97	64	12
74	21	14
90	41	16
72	53	13
85	60	11
72	30	12
91	33	11
100	37	16
98	63	15
87	28	10
99	65	15
76	21	16
79	68	12
79	29	19
89	50	14
97	22	14
85	44	17
81	22	12
97	60	20
93	27	12
93	31	19
87	64	17
96	32	10
100	59	19
92	61	11
90	28	15
99	67	16
98	64	20

Since our groups are randomly generated, out data generating process is random, therefore the assumption required for ANOVA test is satisfied. We can just go ahead and run the test.

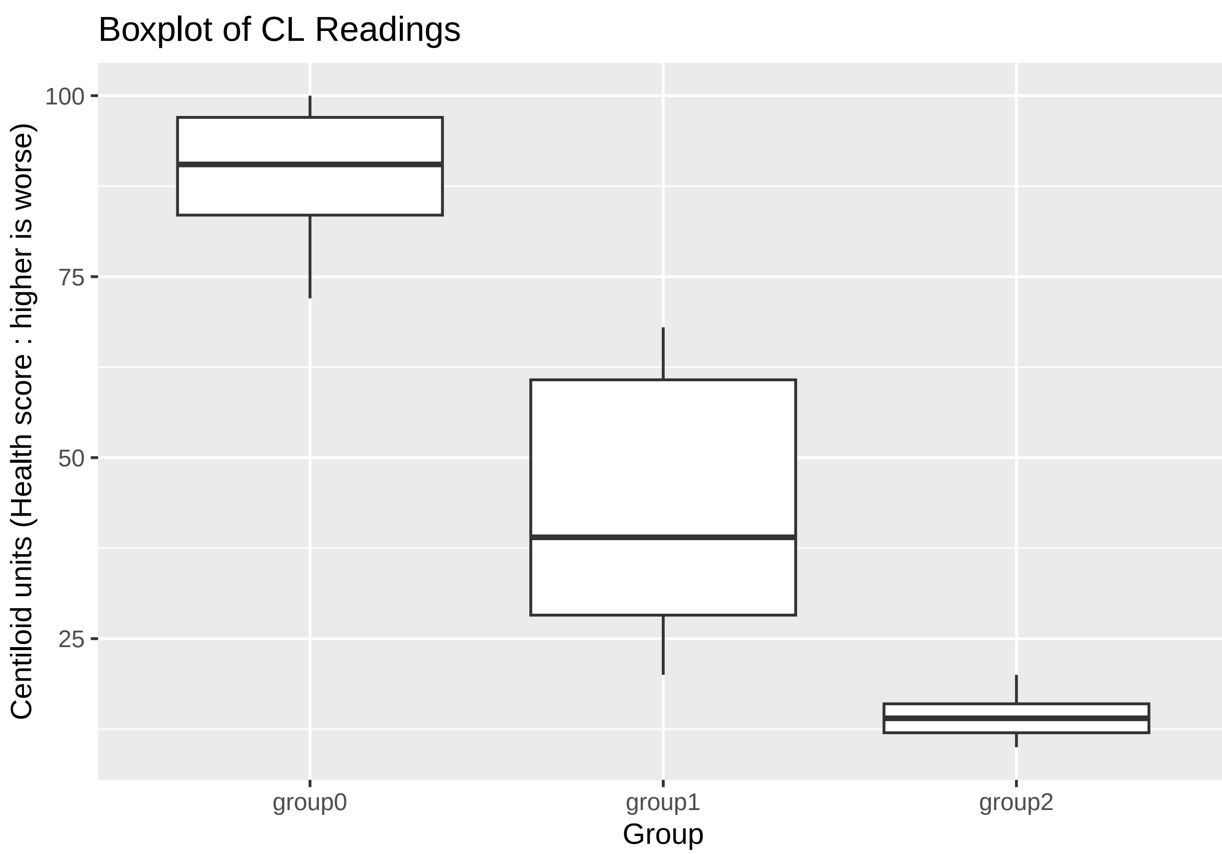
Storaing data in R-vectors

```
group0=c(96, 83, 97, 74, 90, 72, 85, 72, 91, 100, 98, 87, 99, 76, 79, 79, 89, 97
                  , 85, 81, 97, 93, 93, 87, 96, 100, 92, 90, 99, 98)
group1=c(34, 20, 64, 21, 41, 53, 60, 30, 33, 37, 63, 28, 65, 21, 68, 29, 50, 22,
                                 44, 22, 60, 27, 31, 64, 32, 59, 61, 28, 67, 64)
group2=c(11, 11, 12, 14, 16, 13, 11, 12, 11, 16, 15, 10, 15, 16, 12, 19, 14, 14,
                                 17, 12, 20, 12, 19, 17, 10, 19, 11, 15, 16, 20)

#generate a dataframe to combine all groups together
df=data.frame(group0, group1,group2)
```

Descriptive analysis

```
#install.packages("tidyverse")
library(ggplot2)
```

```r
# Convert the data frame to long format for plotting
df_long = tidyr::gather(df, group, value)

# Plot the boxplot
ggplot(df_long, aes(x = group, y = value)) +
  geom_boxplot() +
  xlab("Group") +
  ylab("Centiloid units (Health score : higher is worse)") +
  ggtitle("Boxplot of CL Readings")
```



Interpretation

The box plot shows five point summary : minimum, 1st quartile or 25th percentile, median, 3rd quartile or 75th percentile and maximum values. We see that the minimum of Placebo (group0) is higher than the maximum of group1. This clearly says that the people with treatment (group1) have better health (lower CL score) than without treatment. Same story goes on when we give higher doeses of treatment i.e. group2 people are healthier than group1 people. However, all these are descriptive evidences, they are not statistically sound proofs to say that the drug works. Therefore, we need to conduct a hypothesis test.

Hypothesis

Rewriting, the hypothesis:

$$H_0 : \quad \mu_0 = \mu_1 = \mu_2$$

$$H_1 : \quad \mu_0 \neq \mu_1 \quad \text{or} \quad \mu_1 \neq \mu_2 \quad \text{or} \quad \mu_0 \neq \mu_2 \quad \text{or} \quad \mu_0 \neq \mu_1 \neq \mu_2$$

ANOVA test

Our strategy is to show that at least one of the group means are not equal. If we can show that $\mu_0 \neq \mu_1$ then we can reject the null hypothesis and hence we are done.

Conducting Anova test

```
result1 = aov(group0~group1, data=df)
summary(result1) # Display the ANOVA table
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## group1       1  302.5  302.53   4.345 0.0464 *
## Residuals   28 1949.6   69.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation of the results

We clearly see that the *p-value* of group1 is less than 0.05, therefore, with 95% confidence we can say that μ_0 is not equal to μ_1. Hence, we can reject the null hypothesis. Therefore, we conclude that Lecanemab drug has impact on individuals diagnosed with Alzheimer's.

Simple Linear Regression

To further confirm our conclusion from ANOVA test, we use regressions. Let's define a variable call it `treat` which takes value 0 for group0 or Placebo group, takes value 1 for group1 and takes value 2 for group2. We represent it by X or independent or explanatory variable in the following regression for an individual i:

$$y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

The variable y_i represents value of CL score for the individual i and X_i is the `treat` variable for example if individual i belongs to Placebo group then $X_i = 0$. The X variable takes values in set $\{0,1,2\}$. The coefficients β_0 and $beta_1$ are real numbers and ϵ is the error term.

Hypothesis

Our equivalent hypothesis can be written as :

$$H_0 : \beta_1 = 0 \quad \text{and} \quad H_1 : \beta_1 \neq 0$$

```
# Create a new variable "treat" with numerical values
df_long$treat <- ifelse(df_long$group == "group0", 0,
                        ifelse(df_long$group == "group1", 1,
                               ifelse(df_long$group == "group2", 2, NA)))


reg=lm(value ~ treat, data=df_long)
summary(reg)
```

```
##
## Call:
## lm(formula = value ~ treat, data = df_long)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.922  -5.235   2.494   8.494  19.078
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   86.339      1.996   43.26   <2e-16 ***
## treat        -37.417      1.546  -24.20   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.98 on 88 degrees of freedom
## Multiple R-squared:  0.8694, Adjusted R-squared:  0.8679
## F-statistic: 585.7 on 1 and 88 DF,  p-value: < 2.2e-16
```

Inrerpretation of the results

We find that the *p-value* of the coefficient of `treat` is less than 0.0001, therefore, with (1-0.0001)x100=99.99 percent confidence, we can say that, treatment has decreased the CL score, therefore improved health of the individuals.

Multiple Linear Regression

Now let's define dummy variables or binary variables X_1 and X_2. The variable X_1 takes value 1 if the individual belongs to group1, 0 otherwise. The variable X_2 takes value 1 if the individual belongs to group2, 0 otherwise. Note that for the individual in group0, both $X_1 = X_2 = 0$.

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

Hypothesis

Our equivalent hypothesis can be written as :

$$H_0 : \beta_1 = \beta_2 = 0 \quad \text{and} \quad H_1 : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0$$

```
summary(lm(df_long$value ~ df_long$group))
```

```
##
## Call:
## lm(formula = df_long$value ~ df_long$group)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.267  -5.708   0.250   6.808  24.733
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           89.167      2.070   43.08   <2e-16 ***
## df_long$groupgroup1  -45.900      2.927  -15.68   <2e-16 ***
## df_long$groupgroup2  -74.833      2.927  -25.57   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.34 on 87 degrees of freedom
## Multiple R-squared:  0.8843, Adjusted R-squared:  0.8816
## F-statistic: 332.4 on 2 and 87 DF,  p-value: < 2.2e-16
```

Interpretation

We note that the *p-value* of both the coefficient of group1 and group2 are less than 0.0001 that means that with more than 99.99 confidence, we can say that both type of treatment (given to group1 and group2) decreases CL score and hence improves the health.

Numerically, keeping all other things constant, with more than 99.99 percent confidence, when we give treatment to group1 i.e. we give a treatment of 10mg Lecanemab per 100kg body weight then the CL score decreases by 45.9 units compared to Placebo. This is the interpretation of $\beta_1 = -45.9$.

Further, keeping all other things constant, with more than 99.99 percent confidence, when we give treatment to group2 i.e. we give a treatment of 10mg Lecanemab per 100kg body weight per week then the CL score decreases by 74.83 units compared to Placebo. This is the interpretation of $\beta_2 = -74.83$.

Final Conclusion

My visualization analysis and formal test shows that the drug has positive impact of health (measured by Centiloid unit score) on individuals diagnosed with Alzheimer.