October 27, 2020

1. Please discuss what a p-value is, how it should be used, and the parameters that influence it. You answer should rely heavily on the information in the ASA statement on p-values.

In general p-value stands for probability of achieving outcomes at least as extreme as the observed results of a statistical hypothesis test, assuming null hypothesis is true. P-value gives level of significance of the result. According to ASA its scientist’s dirty secret that statistical analysis of scientific hypothesis testing is based on weak foundation. Statistical techniques have many flaws which is not due to the wrongly used or misused p-value but mistake done in data analysis process by poorly trained analysts. According to ASA “Valid scientific conclusions based on p-values and related statistics cannot be drawn without at least knowing how many and which analyses were conducted.”. It means p-value can be useful only if not misinterpreted and hence should considered under principles by ASA which as below: 1. “P-values can indicate how incompatible the data are with a specified statistical model”: Since models are built under a set of assumptions, as long as these assumptions hold, a small p-value suggests a model that is incompatible with the null hypothesis. 2. “P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone”: p-value is not a self-explanation but is statement in relations to specific hypothesis. 3. “Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold”: study design and quality measurement should be considered in model design as blind consideration of p-value is the worst idea. 4. “Proper inference requires full reporting and transparency”: Multiple analysis of data is required to identify proper result by p-value. I.e. multiple tests should be done for to check validity of p-value obtained. 5. “A p-value, or statistical significance, does not measure the size of an effect or the importance of a result”: Sample size must be observed well before taking p-value (no matter tiny or very high) to the significant consideration i.e. effect of size plays a great role in p-value. 6. “By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis”: It means an approximate 0.05 p-value is not a strong evidence against the null hypothesis.

Overall, for effective use of p-value proper qualitative as well as quantitative assessments should be done by carefully looking into the data and applying a variety of tests to the data including being careful to the outlier and its removal as well as plotting data and checking cautiously each value obtained by different tests. After considering all possible factors p-value should be used to consider the result obtained. Arbitrary use of p-value is the worst way to conclude results statistically.

1. Please discuss the assumptions of parametric tests and nonparametric tests: When should each be used? Please list each of the nonparametric tests we’ve covered to date and its parametric analogue.

* Parametric tests assumption:

1. Data are normally distributed
2. Sample is randomly selected from a defined population.
3. Population variances are same (homogeneity of variances)
4. There is linear relationship in data. Nonparametric tests assumption:
5. Data are not normally distributed.
6. The level of data may be nominal or ordinal.

Parametric tests should be used if there is specific assumption are made regarding population and data are found normally distributed and nonparametric tests should be used if there is no specific assumption can be made regarding population parameters or distribution i.e. if there is definite outlier. Simply when assumptions of parametric tests are not meet nonparametric tests are used. If the mean represents the center of distribution correctly and the sample size is sufficiently large then parametric tests should be preferred because they are more efficient. If the median represents the middle of the distribution better, even with a large sample nonparametric test should be preferred.

nonparametric tests and its parametric analogue covered to date

|  |  |
| --- | --- |
| Nonparametric tests | Parametric tests |
| One-sample Wilcoxon | One sample z test and one-sample t test |
| Chi-square test | No equivalent test |
| Shapiro–Wilk test | Shapiro–Wilk test(also semi parametric) |
| Kolmogorov–Smirnov test |  |
| Fligner test | leveneTest |
| Mann-Whitney test (Wilcoxon rank-sum test) |  |
| Spearman correlation | Pearson correlation test |
| Independent means T-Test Wilcoxon’s test (Wilcoxon Signed-Rank Test) | Dependent means T-Test (repeated measures t-test) |
| Kruskal-Wallis test | One-Way independent measures ANOVA |
| Friedman’s test | One-way repeated measures ANOVA |

1. You are teaching a graduate course in statistics and interested in determining if there is a change in dopamine levels throughout your course. You are the first to do this type of study. 18 students agreed to submit urinalysis that will allow you to infer dopamine levels before the class, 4 weeks into the class, and 8 weeks into the class. Please test this hypothesis with the most appropriate statistical test for the data. Be sure to test all assumptions to ensure that the data are appropriate for the test

Variable Description ID : Subject ID TimePoint: Time the urinalysis was assessed dopamine: Dopamine levels for that Subject ID at that TimePoint

#install.packages("ez")  
#install.packages("kableExtra")  
  
library(ez)

## Warning: package 'ez' was built under R version 4.0.3

## Registered S3 methods overwritten by 'lme4':  
## method from  
## cooks.distance.influence.merMod car   
## influence.merMod car   
## dfbeta.influence.merMod car   
## dfbetas.influence.merMod car

library(car)

## Loading required package: carData

library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':  
##   
## recode

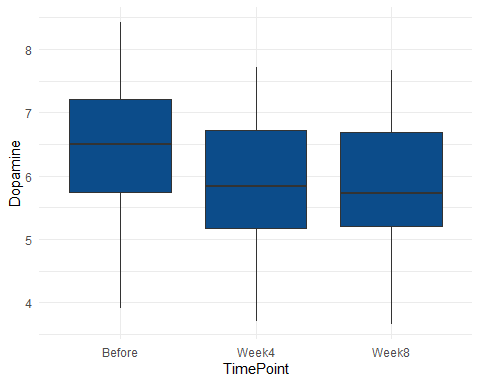
## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

# Data sets  
StudentData <- read.csv("dopamineTime.csv")  
summary(StudentData)

## ID TimePoint Dopamine   
## Min. : 1.0 Length:54 Min. :3.660   
## 1st Qu.: 5.0 Class :character 1st Qu.:5.205   
## Median : 9.5 Mode :character Median :6.045   
## Mean : 9.5 Mean :6.009   
## 3rd Qu.:14.0 3rd Qu.:6.848   
## Max. :18.0 Max. :8.430

# Prametric tests: Assuming data is normally distributed  
  
# Hypothesis:   
#H0: All means are equal (there is no effect of classes time on the level of dopamine in students)  
#H1: All means are not equal (there is effect of classes time on the level of dopamine in students)  
  
  
## Checking assumptions qualitatively  
ggplot(StudentData) +  
 aes(x = TimePoint, y = Dopamine) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 theme\_minimal()



# Checking assumptions quantitatively   
by(StudentData$Dopamine,StudentData$TimePoint,shapiro.test)

## StudentData$TimePoint: Before  
##   
## Shapiro-Wilk normality test  
##   
## data: dd[x, ]  
## W = 0.9819, p-value = 0.9675  
##   
## ------------------------------------------------------------   
## StudentData$TimePoint: Week4  
##   
## Shapiro-Wilk normality test  
##   
## data: dd[x, ]  
## W = 0.97687, p-value = 0.9121  
##   
## ------------------------------------------------------------   
## StudentData$TimePoint: Week8  
##   
## Shapiro-Wilk normality test  
##   
## data: dd[x, ]  
## W = 0.97733, p-value = 0.9183

### Checking homogeneity of variance  
leveneTest(StudentData$Dopamine~StudentData$TimePoint,StudentData, center=mean)

## Warning in leveneTest.default(y = y, group = group, ...): group coerced to  
## factor.

## Levene's Test for Homogeneity of Variance (center = mean)  
## Df F value Pr(>F)  
## group 2 0.0365 0.9642  
## 51

#The test reveals a p-value greater than 0.05, indicating that there is no significant difference between the group variance  
  
  
# Running ANOVA using aov  
x=aov(StudentData$Dopamine~StudentData$TimePoint,StudentData)  
  
# summary of ANOVA results  
summary(x)

## Df Sum Sq Mean Sq F value Pr(>F)  
## StudentData$TimePoint 2 4.32 2.160 1.664 0.2  
## Residuals 51 66.21 1.298

library(effectsize)  
## Effect Size  
eta\_squared(x, partial = T, ci = 0.95)

## Parameter | Eta2 (partial) | 1e+02% CI  
## -----------------------------------------------------  
## StudentData$TimePoint | 0.06 | [0.00, 0.20]

omega\_squared(x, partial = T, ci = 0.95)

## Parameter | Omega2 (partial) | 1e+02% CI  
## --------------------------------------------------------  
## StudentData$TimePoint | 0.02 | [-0.04, 0.17]

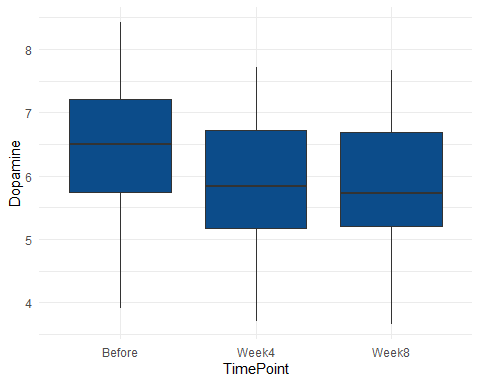
## Repeated measures ANOVA  
ezANOVA(data=StudentData,dv=.(Dopamine), wid=.(ID), within=.(TimePoint), detailed=T)

## Warning: Converting "ID" to factor for ANOVA.

## Warning: Converting "TimePoint" to factor for ANOVA.

## $ANOVA  
## Effect DFn DFd SSn SSd F p p<.05  
## 1 (Intercept) 1 17 1950.124817 65.8660833 503.3261 4.542024e-14 \*  
## 2 TimePoint 2 34 4.319544 0.3458556 212.3206 6.167367e-20 \*  
## ges  
## 1 0.96716226  
## 2 0.06124278  
##   
## $`Mauchly's Test for Sphericity`  
## Effect W p p<.05  
## 2 TimePoint 0.3809941 0.0004439621 \*  
##   
## $`Sphericity Corrections`  
## Effect GGe p[GG] p[GG]<.05 HFe p[HF] p[HF]<.05  
## 2 TimePoint 0.617663 3.887911e-13 \* 0.6418105 1.442841e-13 \*

## Equivalent Nonparametric Tests : Assuming the data is not normally distributed  
  
# Hypothesis:   
#H0: All means are equal (there is no effect of classes time on the level of dopamine in students)  
#H1: All means are not equal (there is effect of classes time on the level of dopamine in students)  
  
## Checking assumptions qualitatively  
ggplot(StudentData) +  
 aes(x = TimePoint, y = Dopamine) +  
 geom\_boxplot(fill = "#0c4c8a") +  
 theme\_minimal()



# Checking assumptions quantitatively using fligner test Checking homogeneity of variance  
  
fligner.test(StudentData$Dopamine~StudentData$TimePoint,StudentData)

##   
## Fligner-Killeen test of homogeneity of variances  
##   
## data: StudentData$Dopamine by StudentData$TimePoint  
## Fligner-Killeen:med chi-squared = 0.092594, df = 2, p-value = 0.9548

fligner.test(StudentData$Dopamine~StudentData$TimePoint)

##   
## Fligner-Killeen test of homogeneity of variances  
##   
## data: StudentData$Dopamine by StudentData$TimePoint  
## Fligner-Killeen:med chi-squared = 0.092594, df = 2, p-value = 0.9548

# Kruskal-Wallis test   
kruskal.test(StudentData$Dopamine~StudentData$TimePoint,StudentData)

##   
## Kruskal-Wallis rank sum test  
##   
## data: StudentData$Dopamine by StudentData$TimePoint  
## Kruskal-Wallis chi-squared = 2.9771, df = 2, p-value = 0.2257

# Frideman test  
  
friedman.test(as.matrix(StudentData))

##   
## Friedman rank sum test  
##   
## data: as.matrix(StudentData)  
## Friedman chi-squared = 108, df = 2, p-value < 2.2e-16

Conclusion: By doing qualitative and quantitative assessment there is no normality violation found and hence the data is normally distributed for which parametric test is preferred. Parametric tests: Shapiro-Wilk normality test: **p(0.9183)>0.05**

Levene’s Test: **p(0.9642)>F(0.0365)** and **Df(2)** Both of which are expected and parametric tests are good to be considered. ANOVA: **p(0.2)>0.05** but **Df(2)** and hence sample size is small so does not indicate too much significance ezANOVA: **p(4.542024e-14)<.05 \*.** Indicates statistical significance and hence Null Hypothesis is rejected means students with class and class with long duration do have dropped dopamine and hence stressed much.

Nonparametric tests: assuming data is not normally distributed Fligner-Killeen test**: df(2),**

**p(0.9548)>0.05** which is expected Kruskal-Wallis: **df(2),** **p(0.2257)>0.05** which is expected

friedman.test: **p(2.2e-16)<0.05** and **Df(2)** andalso gives similar results indicates again statistical significance and hence **Null Hypothesis is rejected** means students with class and class with long duration do have dropped dopamine and hence stressed much.