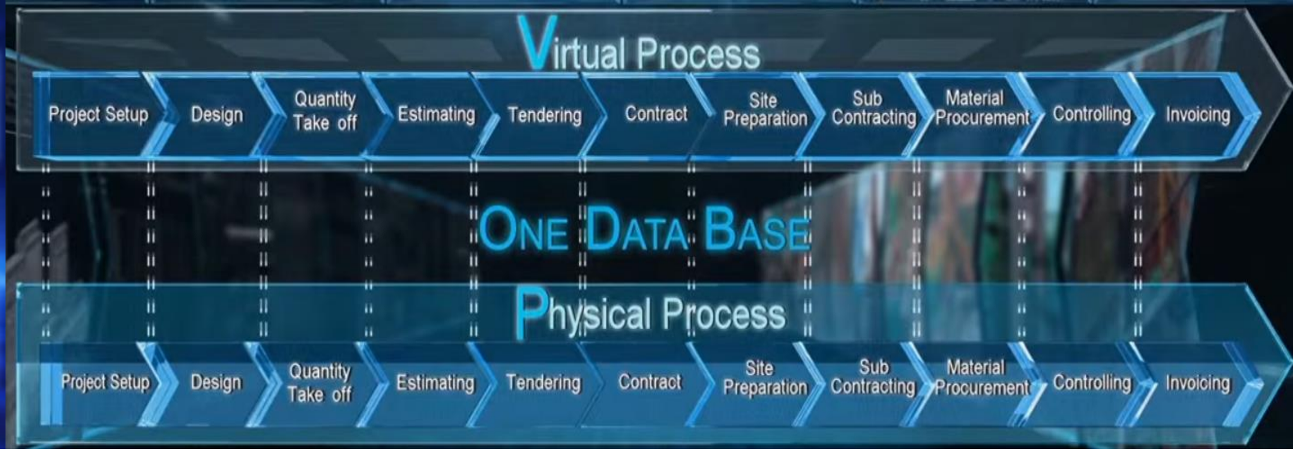


(420) 통계분석에 R/ Shiny 활용

- 통계분석에 R/Shiny 활용 사례 및 Source code
- Shiny web 시스템 활용
- 확대 적용

Shiny DEMO

통계분석 학습 및 data 실습	**대부분 통계분석 기법**	**shiny_ttest_2.r**
통계분석 기법을 업무에 활용할 수 기본 웹 시스템	**t-test, one-way ANOVA 시작하여 계속 추가**	**shiny_useCase_01.r**
Open solution 이해	**계속 발굴**	https://pecostats.shinyapps.io/DIY_ANOVA/ …

Source : shiny_ttest.r

free service level target

1 업무 담당자가 실제 사용 가능한 Process level Use case

2 솔루션 (w/ Open source) 활용 web

통계 분석 기법 활용

- 공분산
- 독립 t-Test (일표본, 대응표본, 독립표본)
- ANOVA (one-way, two-way, MANOVA)
- 요인분석 (PCA/FA)
- 상관분석
- 신뢰도 분석
- 회귀분석 / 다중 회귀분석
- 로지스틱
- 판별분석
- 군집분석 / * 경로분석 / 구조분석

Machine Learning 기본활용

- Linear Regression
- Multi variables Regression
- Logistic Claasification
- Softmax
- CNN / RNN

3 Digital Factory Knowledge

- Digital Transformation strategy
- ERP, SCM : w/ Open source
- PLM
- MOM
- IoT – BDA Pilot : w/ Raspberry Pi

Source :

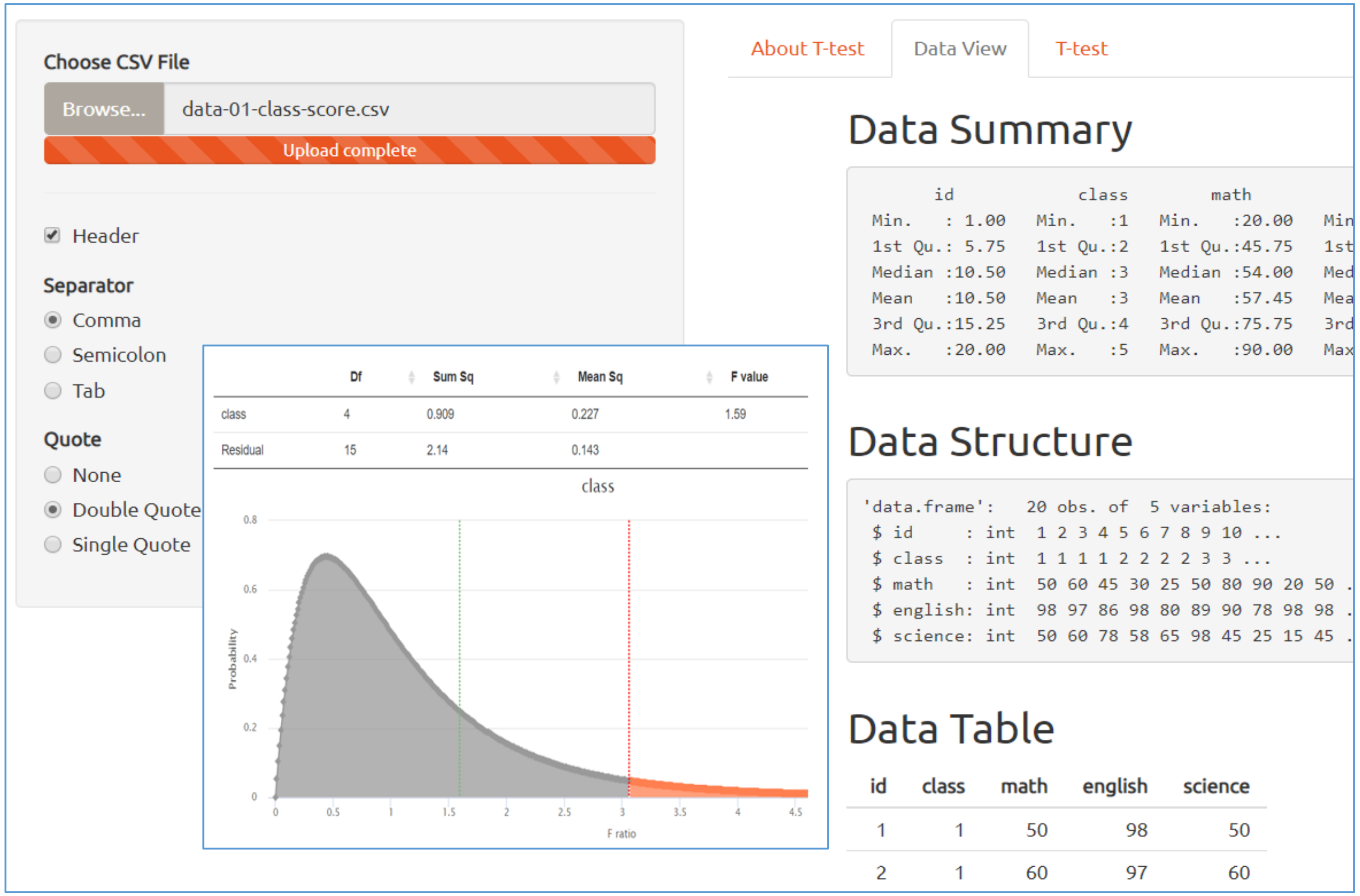
2-1 통계 분석 활용

분석 기법

- 공분산
- 독립 t-Test (일표본, 대응표본, 독립표본)
- ANOVA (one-way, two-way, MANOVA)
- 요인분석 (PCA/FA)
- 상관분석
- 신뢰도 분석
- 회귀분석 / 다중 회귀분석
- 로지스틱
- 판별분석
- 군집분석

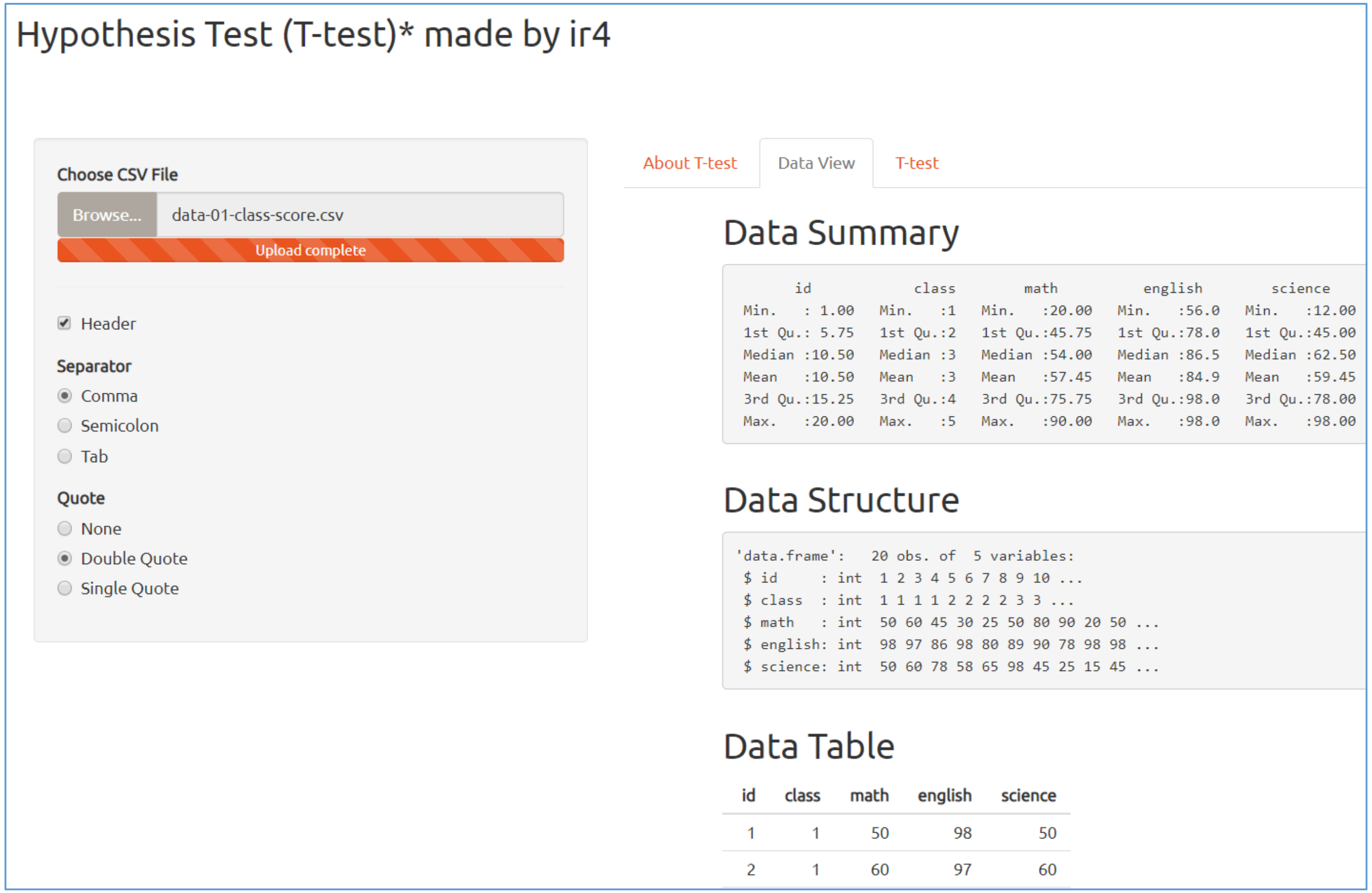
* 경로분석 / 구조분석

분석 화면



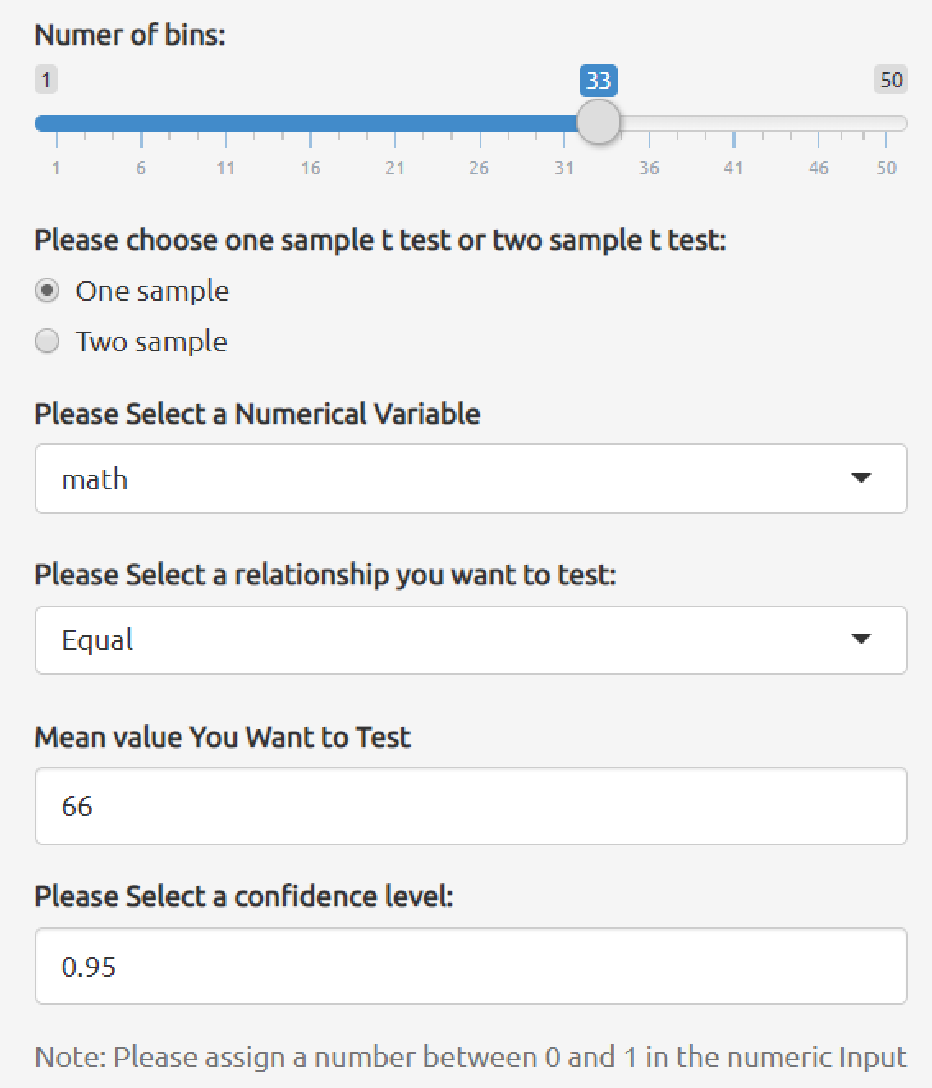
Source :

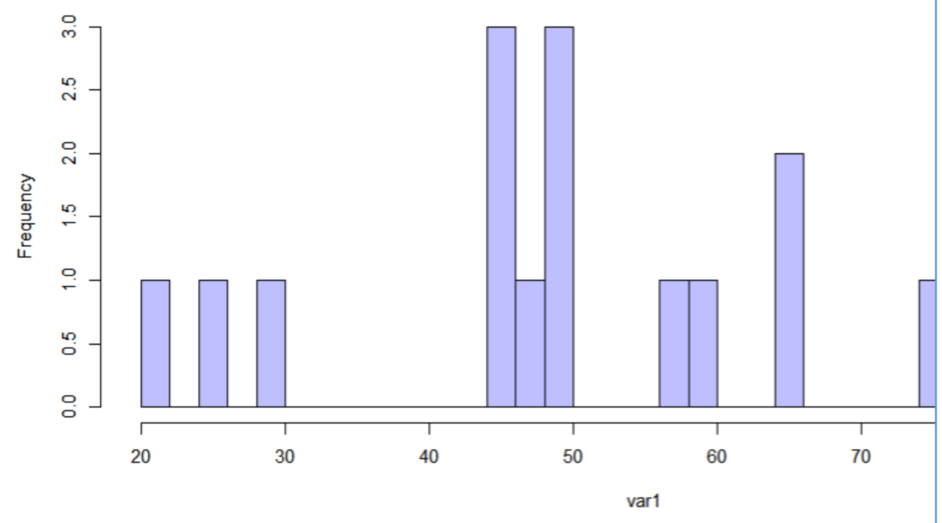
t-test



Source : shiny_ttest.r

t-test – one sample





Key summary statistics

The observed sample statistics were:

mean	standard_deviation	standard_error
57.45	20.30	4.54

Hypothesis of the t-test

We are testing the null hypothesis that the mean of population equals to the value you set

The observed t test statistic :

t -1.883676

A low P value suggests that your sample provides enough evidence that you can reject the null hypothesis for the entire population.

[1] 0.0750069

Source : shiny_ttest.r

t-test – two sample

Numer of bins:

1 — 33 — 50

Please choose one sample t test or two sample t test:

- ○ One sample
- ◉ Two sample

Please Select a Numerical Variable

math

Please Select a Numerical Variable

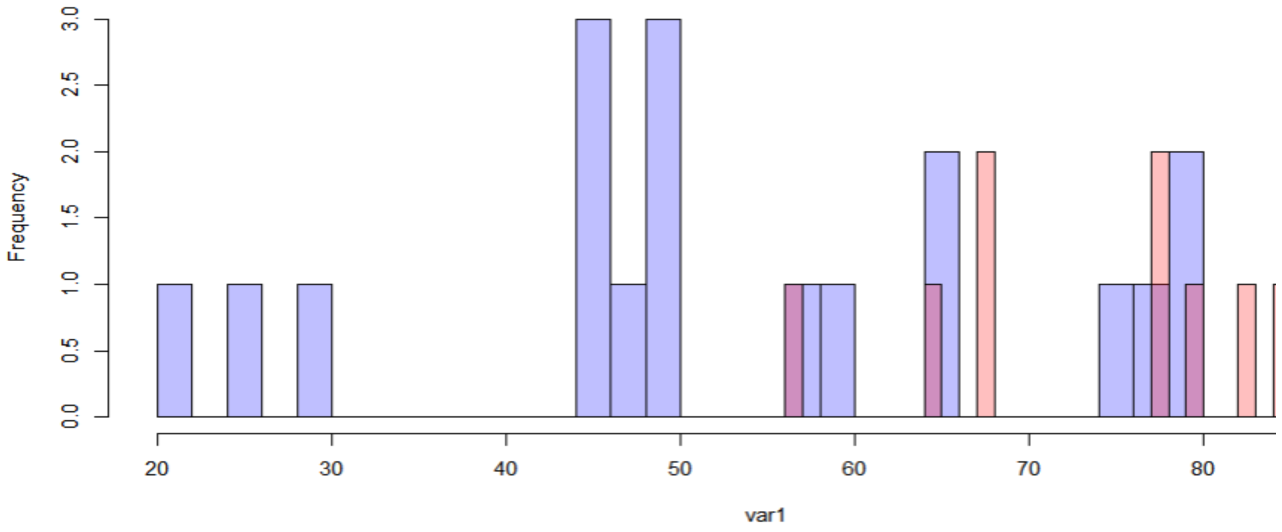
english

Are the two samples have equal variance:

- ○ Yes
- ◉ No

Please Select a relationship you want to test:

Equal



Key summary statistics

The observed sample statistics were:

mean	standard_deviation	standard_error
57.45	20.30	4.54
84.90	12.88	2.88

Hypothesis of the t-test

We are testing the null hypothesis that the mean of population equals to the value you set

The observed t test statistic :

t -5.106904

A low P value suggests that your sample provides enough evidence that you can reject the null hypothesis for the entire population.

[1] 1.437841e-05

Source : shiny_ttest.r

t-test, one-way ANOVA

upload files & summary data

upload the file

Browse... data-01-test-score.csv

Upload complete

default max. file size is 5 MB

☐ Header

Select analytics

◉ One t-test

○ oneWay ANOVA

About file | Data | Summary | Analytics

V1
x1,x2,x3,y
73,80,75,152
93,88,93,185
89,91,90,180
96,98,100,196
73,66,70,142
53,46,55,101
69,74,77,149
47,56,60,115
87,79,90,175

Source : shiny_useCase_01.r

one-way ANOVA

분석	Plot

종속 변수

요인

사후 분석

Source :

요인분석

분석

변수

분석 방법

요인 회전 방법 선택 : 베리맥스, 오블리민, …

Plot

- KMO와 Bartlett의 검정
- 설명된 총분산
- 스크리 도표
- 회전된 성분행렬

Source :

ui Output

```
### use case upload file, select analytics & report plot
library(shiny)
ui <- fluidPage(
    titlePanel(h4("upload files & summary data")),
    sidebarLayout(

        sidebarPanel(

            fileInput("file","upload the file"),
            helpText("default max. file size is 5 MB"),
            tags$hr(),

            checkboxInput(inputId = "header",label = "Header", value = FALSE),
            br(),
            radioButtons(inputId = "analytics",label="Select analytics",choices =
c("One t-test"=',', "oneWay ANOVA" = 'aov'))),

        mainPanel(
            uiOutput("tb")
    )  ))
##

server <- function(input, output, session){

    data <- reactive({
        file1 <- input$file
        if(is.null(file1))  { return() }
```

```
# read.table(file=file1$datapath,sep=input$sep,header =
input$header,stringsAsFactors = input$stringsAsFactors)
        read.table(file=file1$datapath)              })

    output$filedf <- renderTable({
        if(is.null(data())){return()}
        input$file     })
    output$sum <- renderTable({
        if(is.null(data())){return()}
        summary(data())     })
    output$table <- renderTable({
        if(is.null(data())) {return()}
        data()     })

    output$analytics <- renderTable({
        if(is.null(data())) {return()}
        data()     })

    output$tb <- renderUI ({
        if(is.null(data())) {return()}
        # h5("Powered by", tags$img(src='RStudio-Ball.png', height=200,
width=200))
        else
            x <- data()
            tabsetPanel(
                tabPanel("About file", tableOutput("filedf")),
                tabPanel("Data", tableOutput("table")),
                tabPanel("Summary", tableOutput("sum")),
                tabPanel("Analytics", tableOutput("analytics"))
            )        })   }
## shinyApp
shinyApp(ui, server)
```

Source : shiny_useCase_01.r

1. Process level Use case

생산 최적화, 데이터 분석 등 이미 확정된 목표를 가지고,
직원들의 창의성으로 지속적으로 무한 경쟁에서 승자가 되도록

통계 분석 솔루션

AI, ML 분석 솔루션

ERP – PLM - SCM

MOM – IoT/BDA

솔루션, system Infra 가 갖추어져 있다고 가치 도출이 되는가 ?

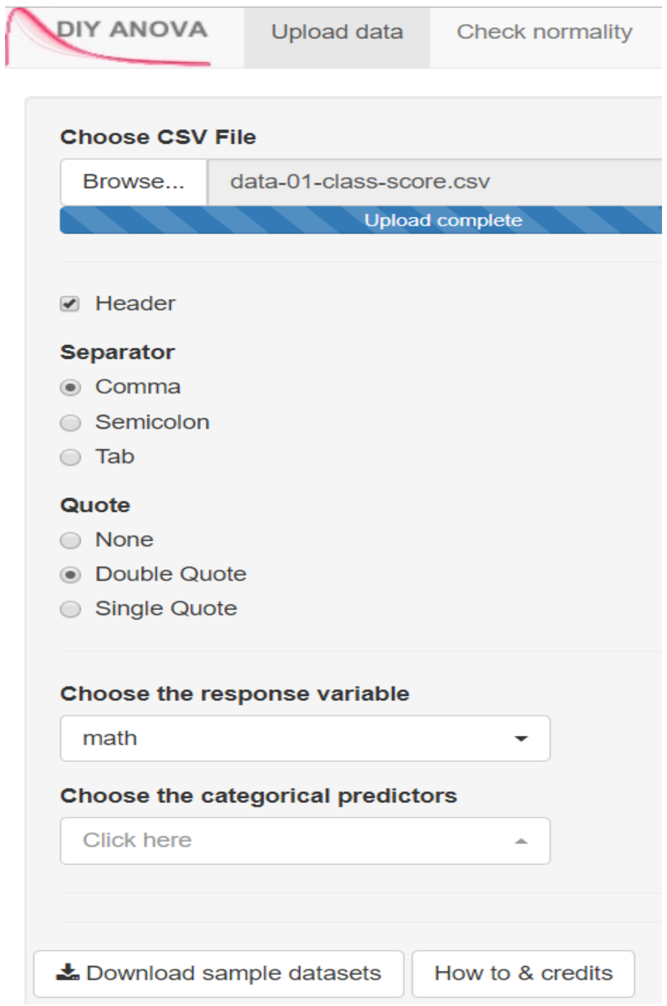
Source :

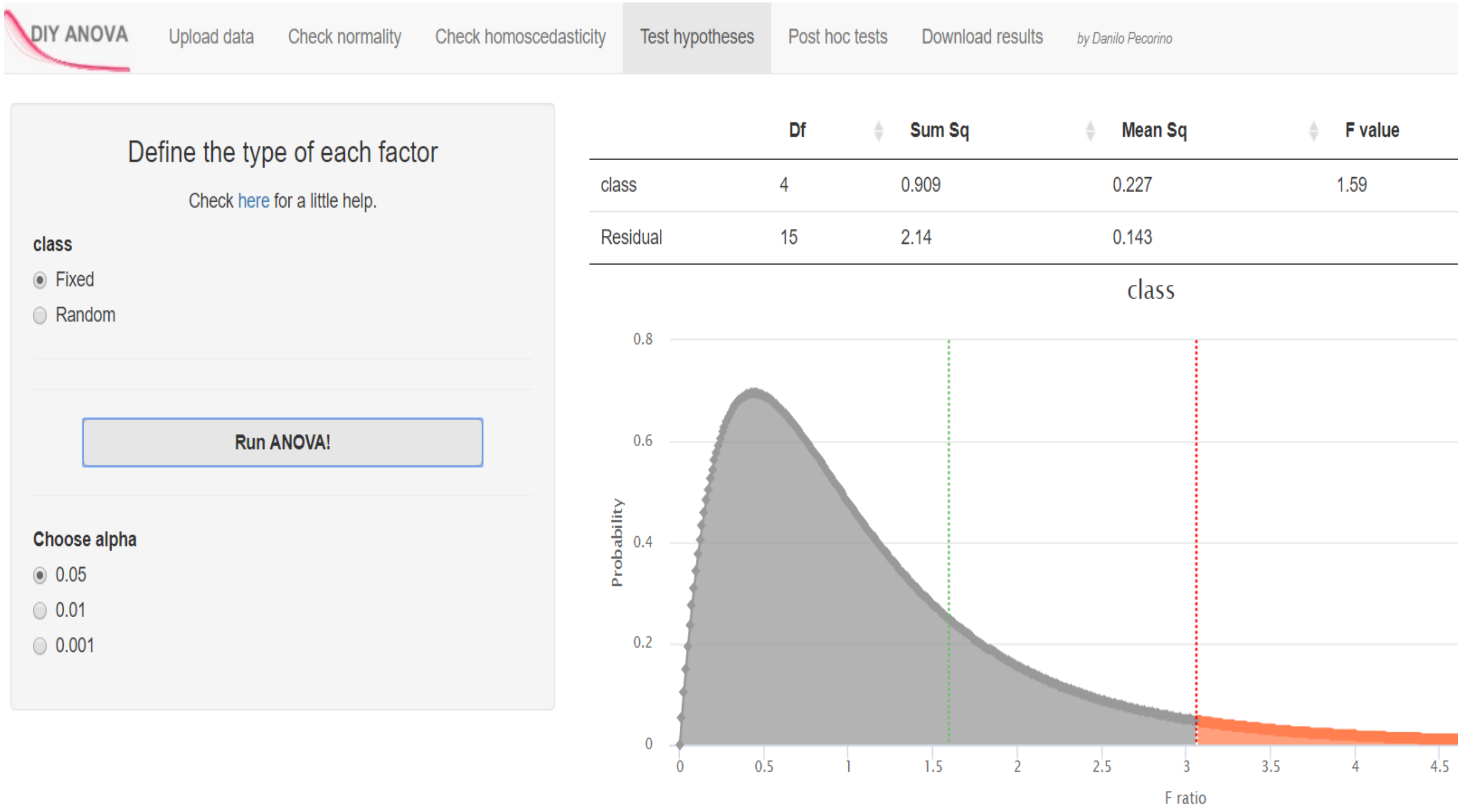
2-2 ML 분석 활용

분석 기법	분석 화면
• Linear Regression • Multi variables Regression • Logistic Claasification • Softmax • CNN / RNN	

Source :

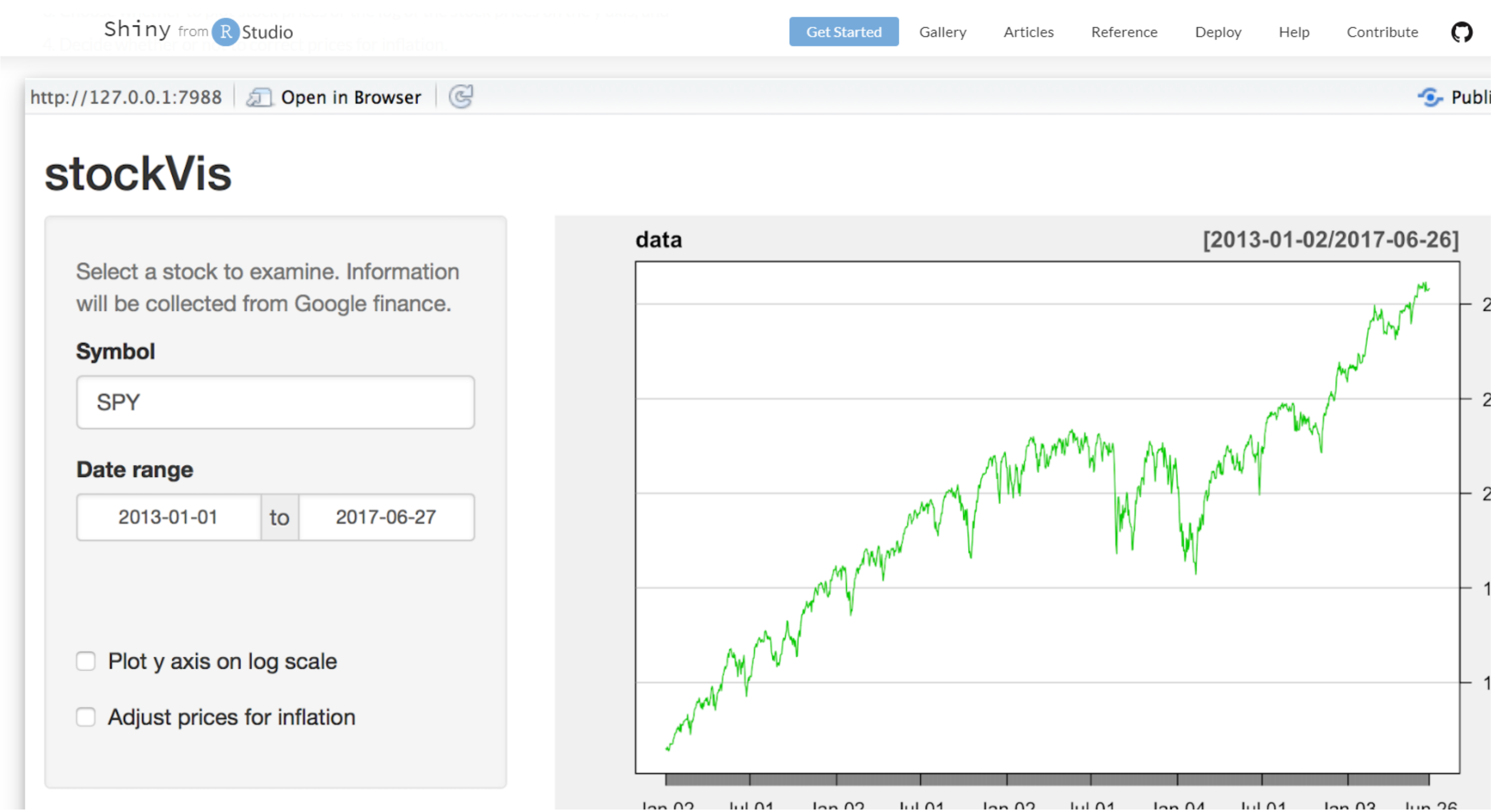
3-1. Open solution : Shiny - ANOVA





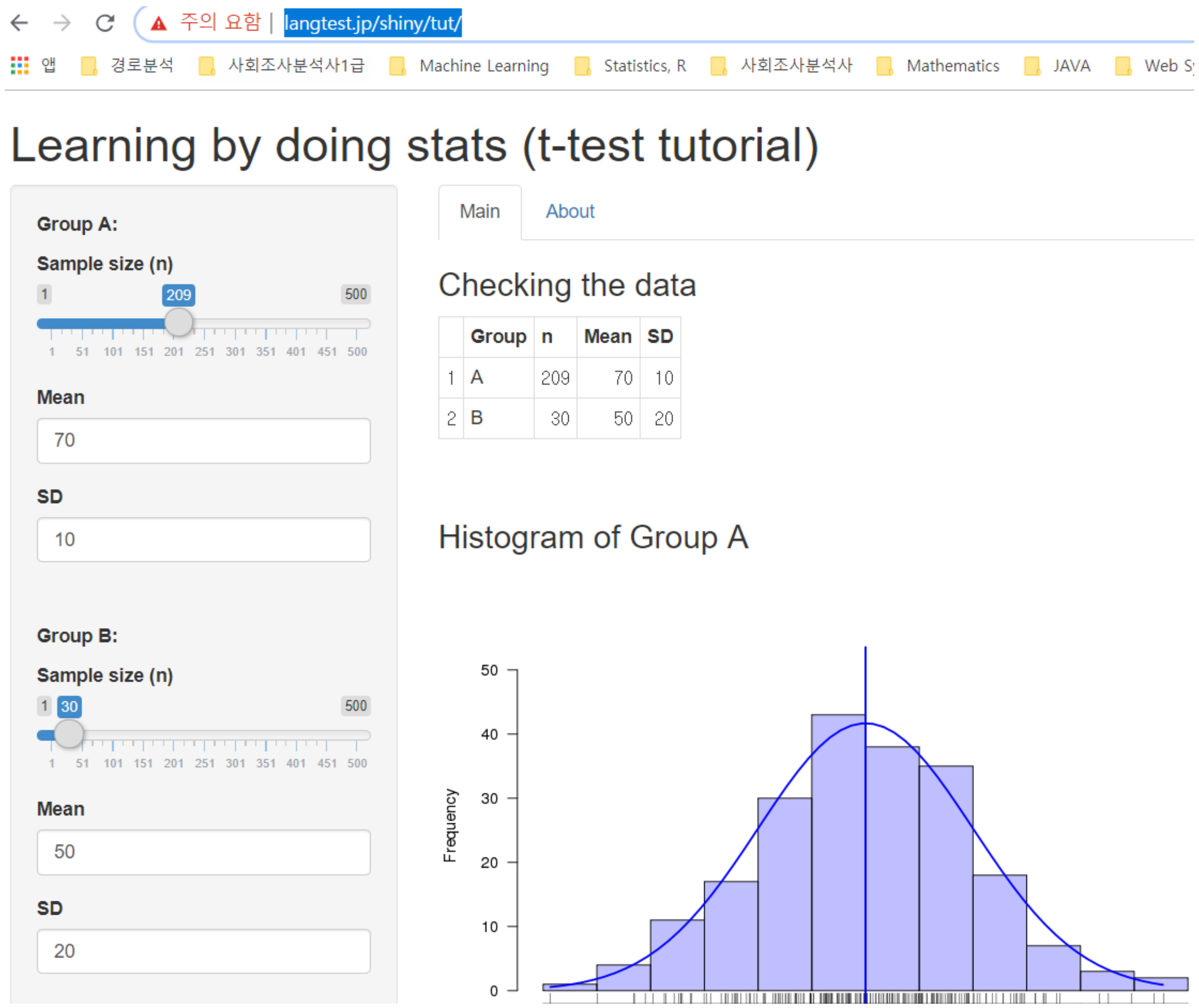
Source : https://pecostats.shinyapps.io/DIY_ANOVA/

reactive data



Source : https://shiny.rstudio.com/tutorial/written-tutorial/lesson6/

turorial



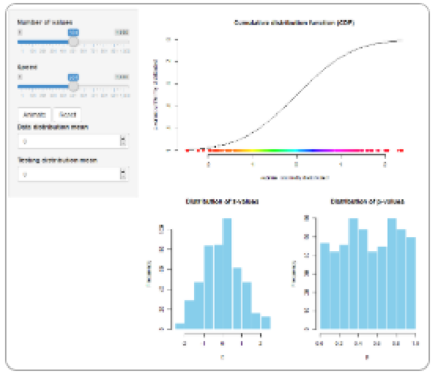
Source : http://langtest.jp/shiny/tut/

shiny widgets

ShinyApps: Experience Statistics | p-checker | p-hacker | About | Links | nicebread.de | Contact

Shiny widgets Experience statistics

Show all | p-values | teaching | p-value | plot | diagnosticity | power | clustering | Bayes | paper | outlier | plot | regression | game | quiz | distribution | histogram

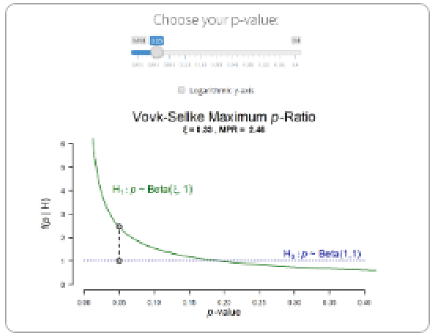


From z to p: A visualization of the link between the two distributions

by Sven Hilbert

p-value plot

Ever wondered why p-values are uniformly distributed if the H0 is accurate, even though the data

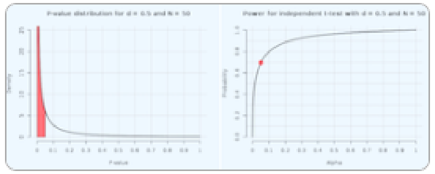


Exploring the diagnosticity of the p-value

by Erik-Jan van Kesteren and Eric-Jan Wagenmakers

p-value plot diagnosticity

Explore the Vovk-Sellke Maximum p-Ratio, a measure that indicates the maximum diagnosticity of a given p-value.

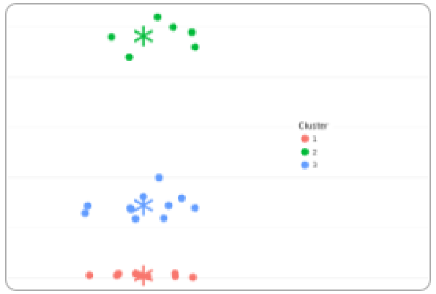


P-value distribution and power curves for an independent two-tailed t-test

by Daniel Lakens

p-value power

Plot the theoretical p-value distribution and power curve for an independent t-test based on the effect size, sample size, and alpha.



Univariate k-Means Clustering with elbow method

by Jan Freyberg

clustering

Identify how many clusters your one-dimensional data can be grouped in and how much variance you can explain with these clusters by using the "elbow method"

Source : http://shinyapps.org/