Friday R Study

KERAMA EZEKIEL JAMES

2023-05-25

#This is another workout of the R analysis  
library(epicalc)

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

## Loading required package: foreign

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

## Loading required package: survival

## Loading required package: MASS

## Warning: package 'MASS' was built under R version 4.2.3

## Loading required package: nnet

#The zap command is used to clear all the contents from the environment  
zap()  
#Loading of the dataset to be used   
data("Decay")  
#Data to be used we use the use command  
use(Decay)  
#To describe the data used we use the des() command  
des()

##   
## No. of observations = 436   
## Variable Class Description   
## 1 decay numeric Any decayed tooth   
## 2 strep numeric CFU of mutan strep.

#To get the summary of the entire dataset we use the summ() command  
summ()

##   
## No. of observations = 436  
##   
##   
## Var. name obs. mean median s.d. min. max.   
## 1 decay 436 0.63 1 0.48 0 1   
## 2 strep 436 95.25 105 53.5 0.5 152.5

data("ANCtable")  
use(ANCtable)  
death <- factor(death)   
levels(death) <- c("no","yes")  
anc <-factor(anc)  
levels(anc) <- c("old","new")  
clinic <- factor(clinic)  
levels(clinic) <- c("A","B")  
pack()  
.data

## death anc clinic Freq  
## 1 no old A 176  
## 2 yes old A 12  
## 3 no new A 293  
## 4 yes new A 16  
## 5 no old B 197  
## 6 yes old B 34  
## 7 no new B 23  
## 8 yes new B 4

glm(death ~ anc + clinic, binomial,weights = Freq,data = .data)

##   
## Call: glm(formula = death ~ anc + clinic, family = binomial, data = .data,   
## weights = Freq)  
##   
## Coefficients:  
## (Intercept) ancnew clinicB   
## -2.7273 -0.1503 0.9863   
##   
## Degrees of Freedom: 7 Total (i.e. Null); 5 Residual  
## Null Deviance: 447.7   
## Residual Deviance: 430.9 AIC: 436.9

str(death)

## Factor w/ 2 levels "no","yes": 1 2 1 2 1 2 1 2

str(clinic)

## Factor w/ 2 levels "A","B": 1 1 1 1 2 2 2 2

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

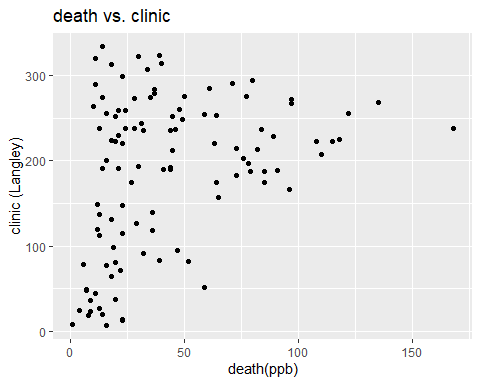
##   
## Attaching package: 'ggplot2'

## The following object is masked \_by\_ '.GlobalEnv':  
##   
## .data

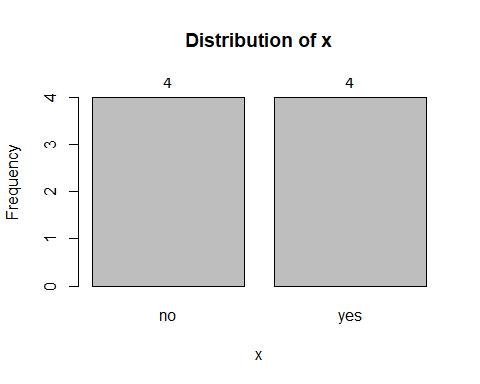
## The following object is masked from 'package:epicalc':  
##   
## alpha

ggplot(airquality, aes(x = Ozone, y = Solar.R)) +   
 geom\_point() +   
 labs(title = "death vs. clinic", x = "death(ppb)", y = "clinic (Langley)")

## Warning: Removed 42 rows containing missing values (`geom\_point()`).

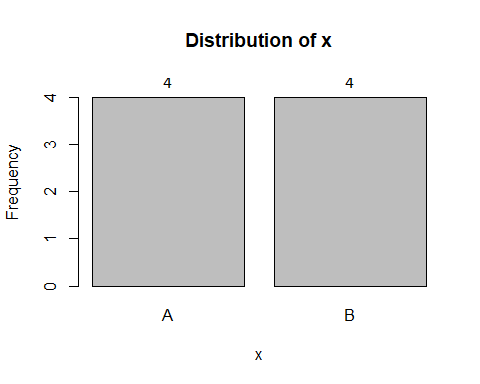


summ(death)



## Frequency Percent Cum. percent  
## no 4 50 50  
## yes 4 50 100  
## Total 8 100 100

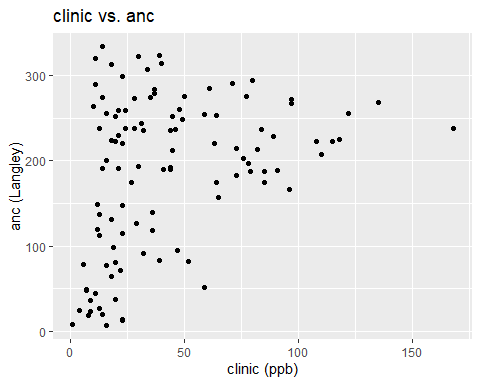
summ(clinic)



## Frequency Percent Cum. percent  
## A 4 50 50  
## B 4 50 100  
## Total 8 100 100

ggplot(airquality, aes(x = Ozone, y = Solar.R)) +   
 geom\_point() +   
 labs(title = "clinic vs. anc", x = "clinic (ppb)", y = "anc (Langley)")

## Warning: Removed 42 rows containing missing values (`geom\_point()`).



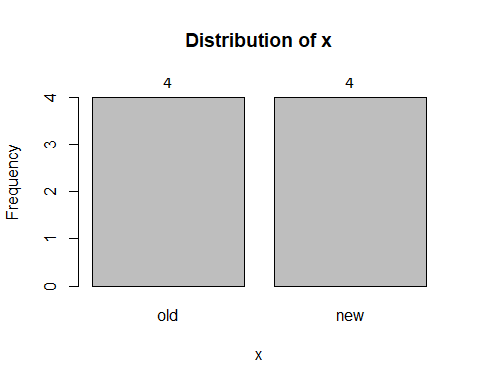
names(death)

## NULL

str(anc)

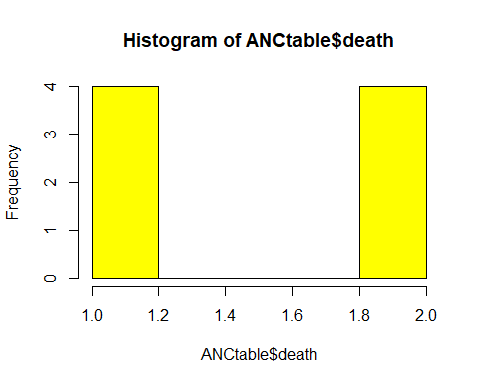
## Factor w/ 2 levels "old","new": 1 1 2 2 1 1 2 2

summ(anc)

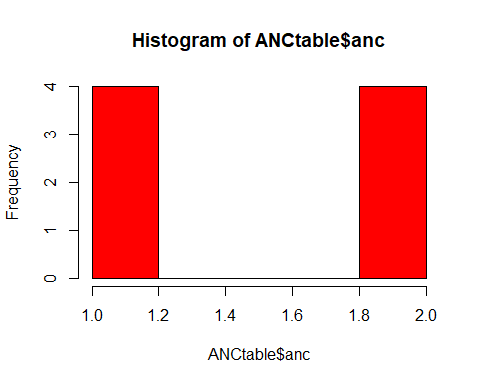


## Frequency Percent Cum. percent  
## old 4 50 50  
## new 4 50 100  
## Total 8 100 100

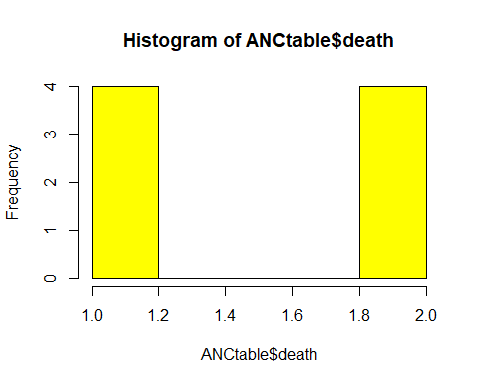
hist(ANCtable$death,col = "yellow")



hist(ANCtable$anc,col = "red")



str(hist(ANCtable$death,col = "yellow"))

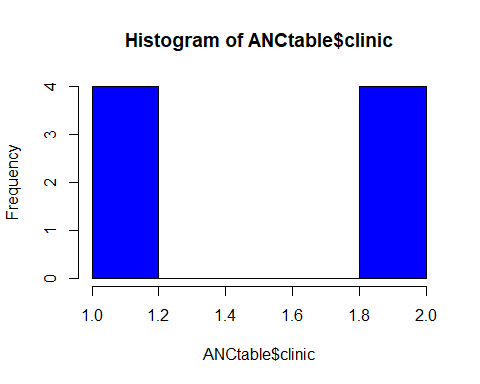


## List of 6  
## $ breaks : num [1:6] 1 1.2 1.4 1.6 1.8 2  
## $ counts : int [1:5] 4 0 0 0 4  
## $ density : num [1:5] 2.5 0 0 0 2.5  
## $ mids : num [1:5] 1.1 1.3 1.5 1.7 1.9  
## $ xname : chr "ANCtable$death"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

# Model Ozone and Solar.R  
model1 <- lm(death ~ clinic, data = ANCtable)  
summary(model1)

##   
## Call:  
## lm(formula = death ~ clinic, data = ANCtable)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.5 -0.5 0.0 0.5 0.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.500e+00 6.455e-01 2.324 0.0591 .  
## clinic 2.355e-16 4.082e-01 0.000 1.0000   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5774 on 6 degrees of freedom  
## Multiple R-squared: 1.183e-30, Adjusted R-squared: -0.1667   
## F-statistic: 7.1e-30 on 1 and 6 DF, p-value: 1

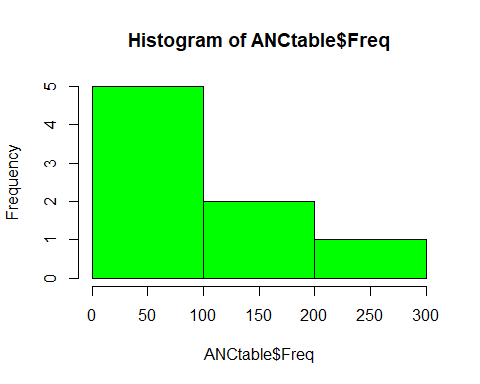
hist(ANCtable$clinic,col = "blue")



str(hist(ANCtable$clinic,col = "blue"))

## List of 6  
## $ breaks : num [1:6] 1 1.2 1.4 1.6 1.8 2  
## $ counts : int [1:5] 4 0 0 0 4  
## $ density : num [1:5] 2.5 0 0 0 2.5  
## $ mids : num [1:5] 1.1 1.3 1.5 1.7 1.9  
## $ xname : chr "ANCtable$clinic"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

hist(ANCtable$Freq,col = "green")  
str(hist(ANCtable$Freq,col = "green"))



## List of 6  
## $ breaks : num [1:4] 0 100 200 300  
## $ counts : int [1:3] 5 2 1  
## $ density : num [1:3] 0.00625 0.0025 0.00125  
## $ mids : num [1:3] 50 150 250  
## $ xname : chr "ANCtable$Freq"  
## $ equidist: logi TRUE  
## - attr(\*, "class")= chr "histogram"

#How to come up with the names of all the variables in the entire dataset  
names(ANCtable)

## [1] "death" "anc" "clinic" "Freq"

#How to come up with the structral information of the entire dataset  
str(ANCtable)

## 'data.frame': 8 obs. of 4 variables:  
## $ death : int 1 2 1 2 1 2 1 2  
## $ anc : int 1 1 2 2 1 1 2 2  
## $ clinic: int 1 1 1 1 2 2 2 2  
## $ Freq : int 176 12 293 16 197 34 23 4

#How to come up with the summary of the entire dataset  
summ(ANCtable)

##   
## No. of observations = 8  
##   
##   
## Var. name obs. mean median s.d. min. max.   
## 1 death 8 1.5 1.5 0.53 1 2   
## 2 anc 8 1.5 1.5 0.53 1 2   
## 3 clinic 8 1.5 1.5 0.53 1 2   
## 4 Freq 8 94.38 28.5 111.15 4 293

#How to come up wih the sum of the number of columns of the entire dataset  
ncol(ANCtable)

## [1] 4

#How to come up with the summation of the number of the rows of the entire vdataset in the R software  
nrow(ANCtable)

## [1] 8

#How to view the entire dataset in an R software  
View(ANCtable)  
#How to to cme up with the summation of the death variable in the ANCtable dataset in the R software  
sum(ANCtable$death)

## [1] 12

#Getting the summation of the clinic variable in the ANCtable dataset  
sum(ANCtable$clinic)

## [1] 12

.data$condition <- c(1,1,2,2,3,3,4,4)  
data2 <- reshape(.data,timevar = "death", v.names = "Freq",idvar = "condition",direction = "wide")