

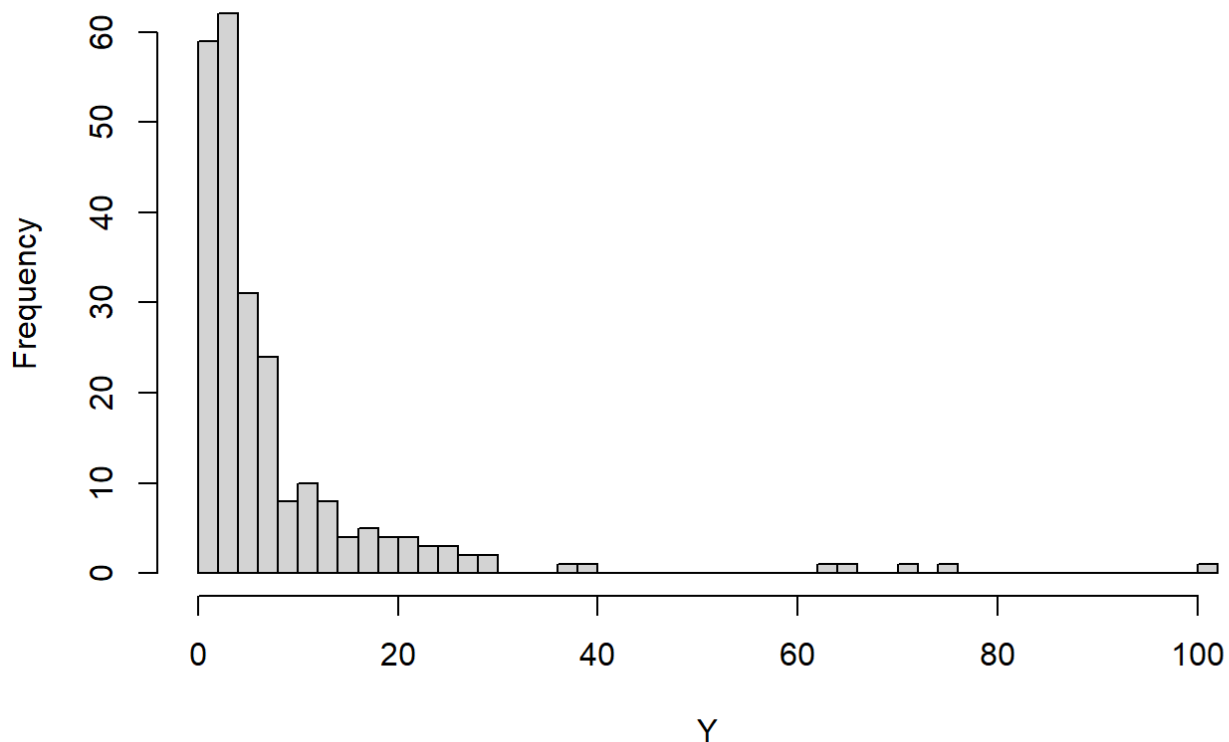
thirdlab3

2025-10-11

```
library(MASS)
data(epil)
ndata=epil
#| Variable | Description |
#| ----- | ----- |
#| `y`      | seizure count in a 2-week period |
#| `treat`  | treatment (placebo or progabide) |
#| `base`   | baseline seizure count before treatment |
#| `age`    | patient age |
#| `period` | observation period (1-4) |
#| `subject` | patient ID |
#| `trt`    | alternative treatment code |
#| 'lbase'. | Log transformed and normalize. |
#| 'lage'.  | Log transformed and normalize. |
#| 'V4'.    | Count of seizures during 4th period. |
```

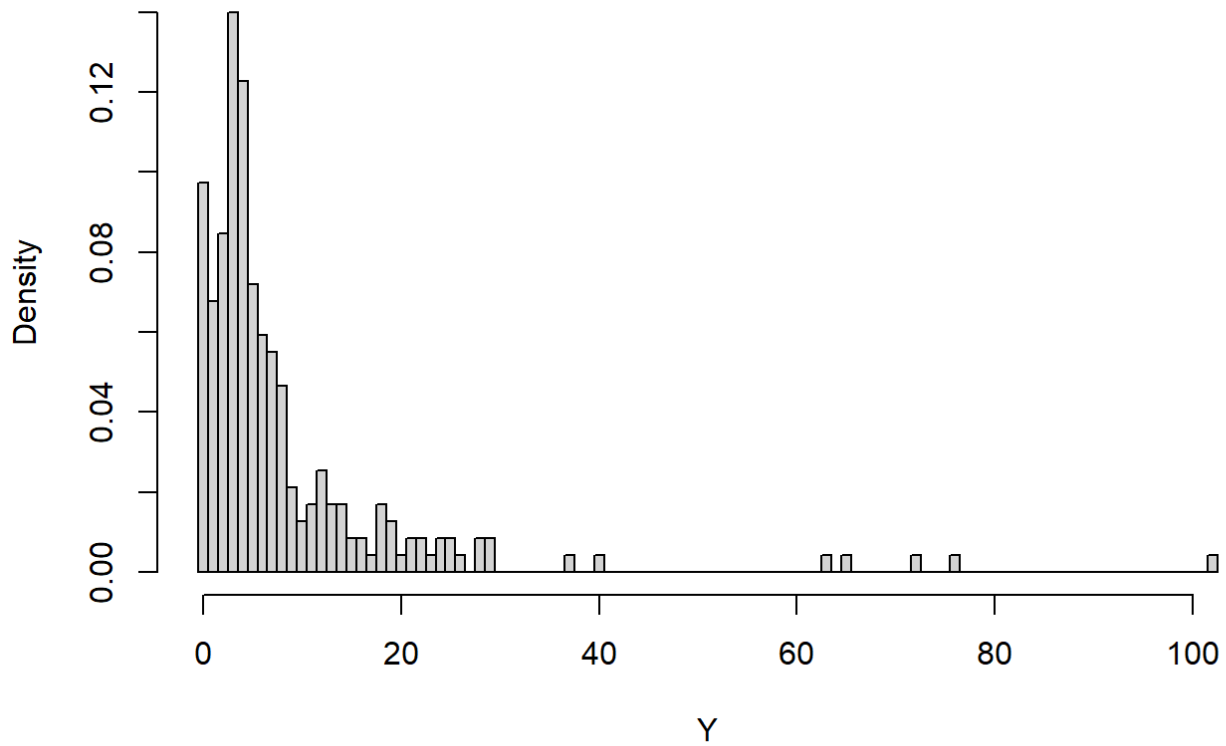
```
Y=ndata[,1]
hist(Y,breaks=40,freq=TRUE)#Default is freq=TRUE
```

Histogram of Y



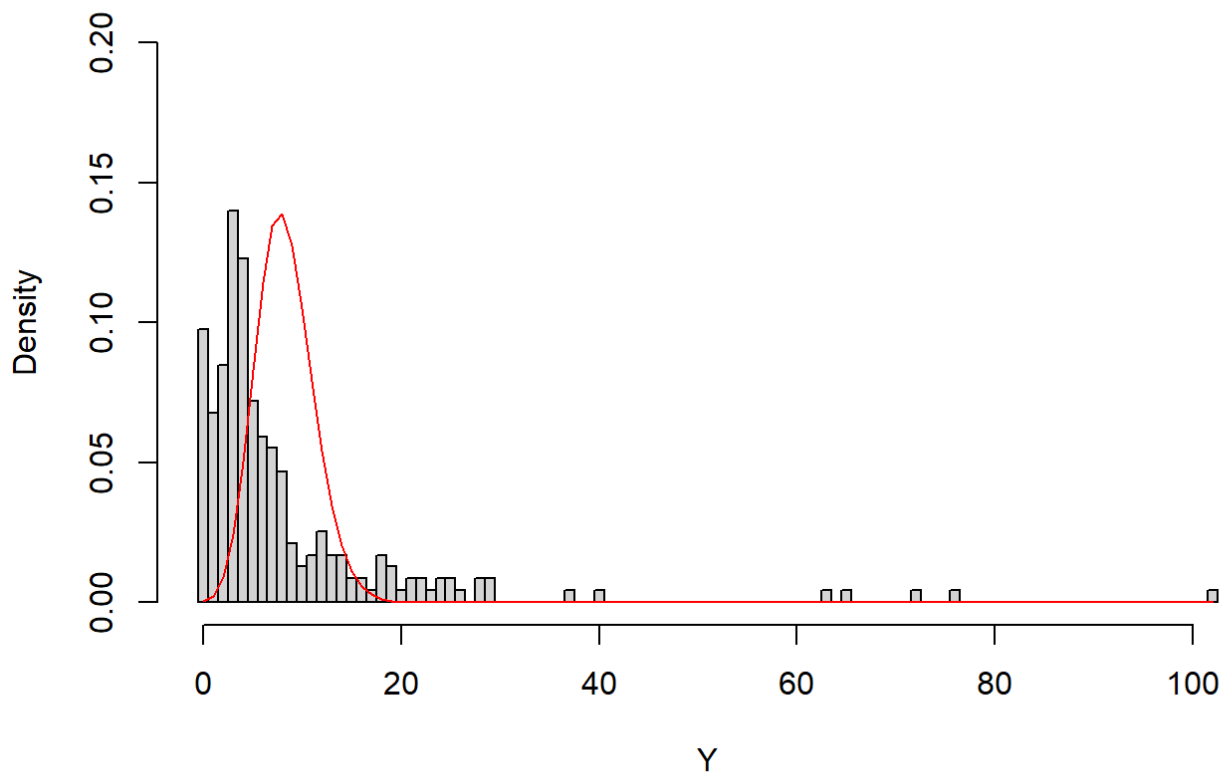
```
breaks = seq(min(Y) - 0.5, max(Y) + 0.5, by = 1)
h=hist(Y,breaks = breaks,freq=FALSE,main="PMF of Y")
```

PMF of Y



```
my=mean(Y)
mv=var(Y)
k=0:max(Y)
fitpois=fitdistr(Y,densfun = "poisson")
pois_pmf=dpois(k,lambda = fitpois$estimate)
hist(Y,breaks = breaks,freq=FALSE,ylim=c(0,0.2),main="PMF of Y")
#points(k,pois_pmf,pch=1)
lines(k,pois_pmf,lwd=1,col='red')
```

PMF of Y



```
print(fitpois$estimate)
```

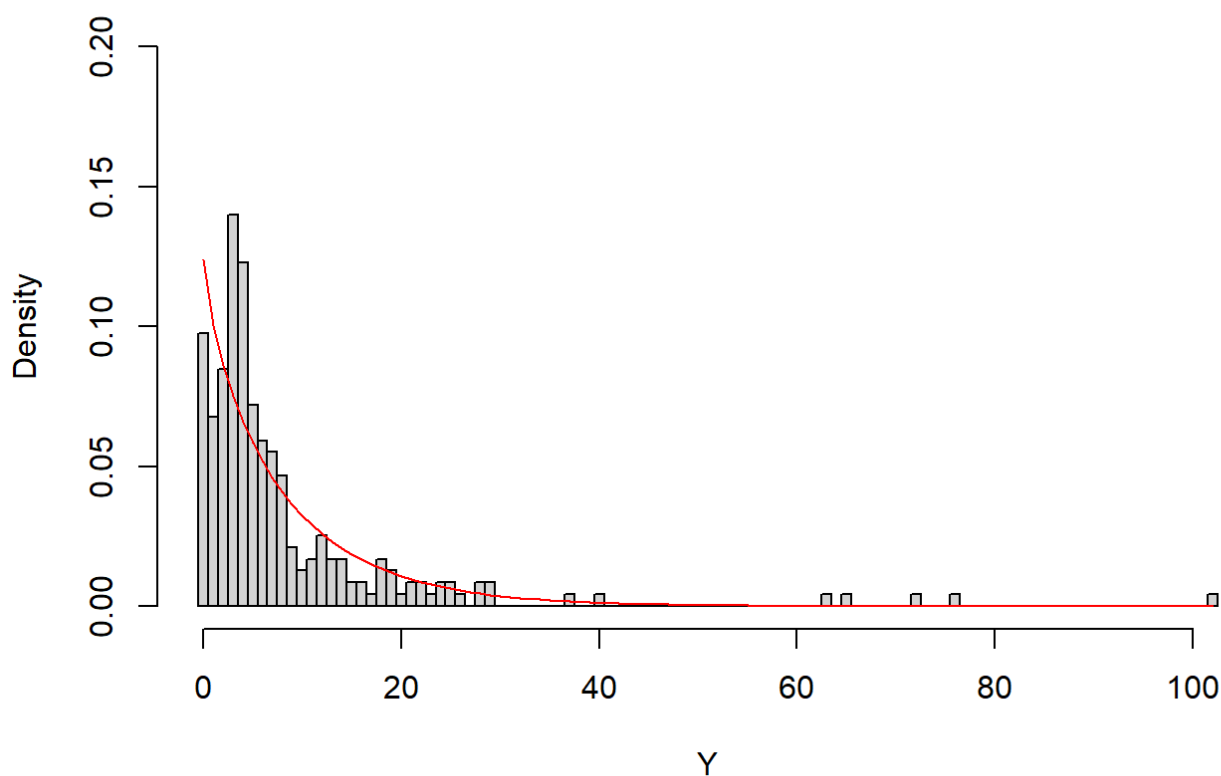
```
##    lambda  
## 8.254237
```

```
fitnbinom=fitdistr(Y,densfun = "Negative Binomial")
```

```
## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced
```

```
nbinom_pmf=dnbinom(k,size=fitnbinom$estimate[1],mu=fitnbinom$estimate[2])  
hist(Y,breaks = breaks,freq=FALSE,ylim=c(0,0.2),main="PMF of Y")  
lines(k,nbinom_pmf,lwd=1,col='red')
```

PMF of Y



```
sprintf("Poisson:Log Likelihood:%f",fitpois$loglik)
```

```
## [1] "Poisson:Log Likelihood:-1641.869823"
```

```
sprintf("Negative Binomial:Log Likelihood:%f",fitnbinom$loglik)
```

```
## [1] "Negative Binomial:Log Likelihood:-747.311184"
```

```
nr=nrow(ndata)
s=sample(x=1:nr,size=0.8*nr,replace = FALSE)
traindata=ndata[s,]
testdata=ndata[-s,]
linearmod=lm(y~lbase+lage+trt,data=traindata)
#Observe: R automatically converts categorical variables to indicator variables
ytrainpred=predict(linearmod,traindata)
ytestpred=predict(linearmod,testdata)
maetrain=mean(abs(traindata[,1]-ytrainpred))
rmsetrain=sqrt(mean((traindata[,1]-ytrainpred)^2))
maetest=mean(abs(testdata[,1]-ytestpred))
rmsetest=sqrt(mean((testdata[,1]-ytestpred)^2))
sprintf("Train MAE:%f",maetrain)
```

```
## [1] "Train MAE:5.811004"
```

```
sprintf("Train RMSE:%f",rmsetrain)
```

```
## [1] "Train RMSE:10.055714"
```

```
sprintf("Test MAE:%f",maetest)
```

```
## [1] "Test MAE:5.990024"
```

```
sprintf("Test RMSE:%f",rmsetest)
```

```
## [1] "Test RMSE:8.614641"
```

```
sprintf("Log Likelihood:%f",logLik(linearmod))
```

```
## [1] "Log Likelihood:-700.690962"
```

```
poissonlossmod=glm(y~lbase+lage+trt,family="poisson",data=traindata)
ytrainpred=predict(poissonlossmod,traindata)
ytestpred=predict(poissonlossmod,testdata)
maetrain=mean(abs(traindata[,1]-ytrainpred))
rmsetrain=sqrt(mean((traindata[,1]-ytrainpred)^2))
maetest=mean(abs(testdata[,1]-ytestpred))
rmsetest=sqrt(mean((testdata[,1]-ytestpred)^2))
sprintf("Train MAE:%f",maetrain)
```

```
## [1] "Train MAE:6.683756"
```

```
sprintf("Train RMSE:%f",rmsetrain)
```

```
## [1] "Train RMSE:13.601664"
```

```
sprintf("Test MAE:%f",maetest)
```

```
## [1] "Test MAE:7.314701"
```

```
sprintf("Test RMSE:%f",rmsetest)
```

```
## [1] "Test RMSE:13.009787"
```

```
sprintf("Log Likelihood:%f",logLik(poissonlossmod))
```

```
## [1] "Log Likelihood:-698.898003"
```

```
nblossmod=glm.nb(y~lbase+lage+trt,data=traindata)
ytrainpred=predict(nblossmod,traindata)
ytestpred=predict(nblossmod,testdata)
maetrain=mean(abs(traindata[,1]-ytrainpred))
rmsetrain=sqrt(mean((traindata[,1]-ytrainpred)^2))
maetest=mean(abs(testdata[,1]-ytestpred))
rmsetest=sqrt(mean((testdata[,1]-ytestpred)^2))
sprintf("Train MAE:%f",maetrain)
```

```
## [1] "Train MAE:6.641569"
```

```
sprintf("Train RMSE:%f",rmsetrain)
```

```
## [1] "Train RMSE:13.633040"
```

```
sprintf("Test MAE:%f",maetest)
```

```
## [1] "Test MAE:7.294945"
```

```
sprintf("Test RMSE:%f",rmsetest)
```

```
## [1] "Test RMSE:13.049174"
```

```
sprintf("Log Likelihood:%f",logLik(nblossmod))
```

```
## [1] "Log Likelihood:-519.221992"
```

```
z=log1p(Y)
ndata2=cbind(z,ndata)
```

```
nr=nrow(ndata2)
s=sample(x=1:nr,size=0.8*nr,replace = FALSE)
traindata=ndata2[s,]
testdata=ndata2[-s,]
linearmod=lm(z~lbase+lage+trt,data=traindata)
tytrainpred=predict(linearmod,traindata)
tytestpred=predict(linearmod,testdata)
ytrainpred=exp(tytrainpred)-1
ytestpred=exp(tytestpred)-1
maetrain=mean(abs(traindata[,2]-ytrainpred))
rmsetrain=sqrt(mean((traindata[,2]-ytrainpred)^2))
maetest=mean(abs(testdata[,2]-ytestpred))
rmsetest=sqrt(mean((testdata[,2]-ytestpred)^2))
sprintf("Train MAE:%f",maetrain)
```

```
## [1] "Train MAE:4.162970"
```

```
sprintf("Train RMSE:%f",rmsetrain)
```

```
## [1] "Train RMSE:9.504859"
```

```
sprintf("Test MAE:%f",maetest)
```

```
## [1] "Test MAE:5.029333"
```

```
sprintf("Test RMSE:%f",rmsetest)
```

```
## [1] "Test RMSE:10.564344"
```

```
sprintf("Log Likelihood:%f",logLik(linearmod))
```

```
## [1] "Log Likelihood:-196.984418"
```

```
poissonlossmod=glm(z~lbase+lage+trt,family="quasipoisson",data=traindata)
tytrainpred=predict(poissonlossmod,traindata)
tytestpred=predict(poissonlossmod,testdata)
ytrainpred=exp(tytrainpred)-1
ytestpred=exp(tytestpred)-1
maetrain=mean(abs(traindata[,2]-ytrainpred))
rmsetrain=sqrt(mean((traindata[,2]-ytrainpred)^2))
maetest=mean(abs(testdata[,2]-ytestpred))
rmsetest=sqrt(mean((testdata[,2]-ytestpred)^2))
sprintf("Train MAE:%f",maetrain)
```

```
## [1] "Train MAE:7.298159"
```

```
sprintf("Train RMSE:%f",rmsetrain)
```

```
## [1] "Train RMSE:13.490588"
```

```
sprintf("Test MAE:%f",maetest)
```

```
## [1] "Test MAE:8.686751"
```

```
sprintf("Test RMSE:%f",rmsetest)
```

```
## [1] "Test RMSE:16.090214"
```

```
sprintf("Log Likelihood:%f",logLik(poissonlossmod))
```

```
## [1] "Log Likelihood:NA"
```

```
nblossmod=suppressWarnings(glm.nb(z~lbase+lage+trt,data=traindata))
tytrainpred=predict(nblossmod,traindata)
tytestpred=predict(nblossmod,testdata)
ytrainpred=exp(tytrainpred)-1
ytestpred=exp(tytestpred)-1
maetrain=mean(abs(traindata[,2]-ytrainpred))
rmsetrain=sqrt(mean((traindata[,2]-ytrainpred)^2))
maetest=mean(abs(testdata[,2]-ytestpred))
rmsetest=sqrt(mean((testdata[,2]-ytestpred)^2))
sprintf("Train MAE:%f",maetrain)
```

```
## [1] "Train MAE:7.298159"
```

```
sprintf("Train RMSE:%f",rmsetrain)
```

```
## [1] "Train RMSE:13.490588"
```

```
sprintf("Test MAE:%f",maetest)
```

```
## [1] "Test MAE:8.686751"
```

```
sprintf("Test RMSE:%f",rmsetest)
```

```
## [1] "Test RMSE:16.090214"
```

```
sprintf("Log Likelihood:%f",logLik(nblossmod))
```

```
## [1] "Log Likelihood:-251.512108"
```