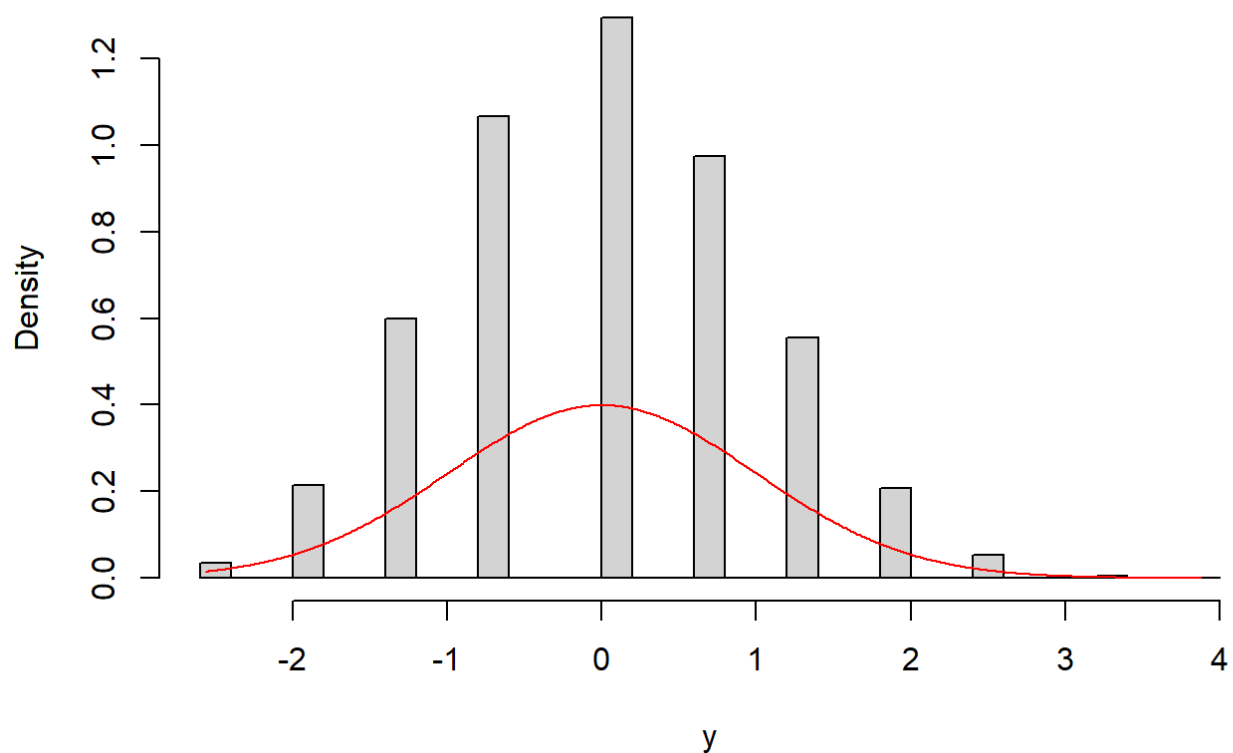# fourthlab

2025-10-25

```r
p=0.4
n=10
nsamples=10000
samples=numeric(nsamples)
for(i in 1:nsamples){
  samples[i]=rbinom(1,size=n,p=p)
}
x=samples
y=scale(x)
k=seq(min(y),max(y),by=0.01)
gpmf=dnorm(k,mean = mean(y),sd=sd(y))
hist(y,breaks=40,freq=FALSE)
lines(k,gpmf,lwd=1,col="red")
```
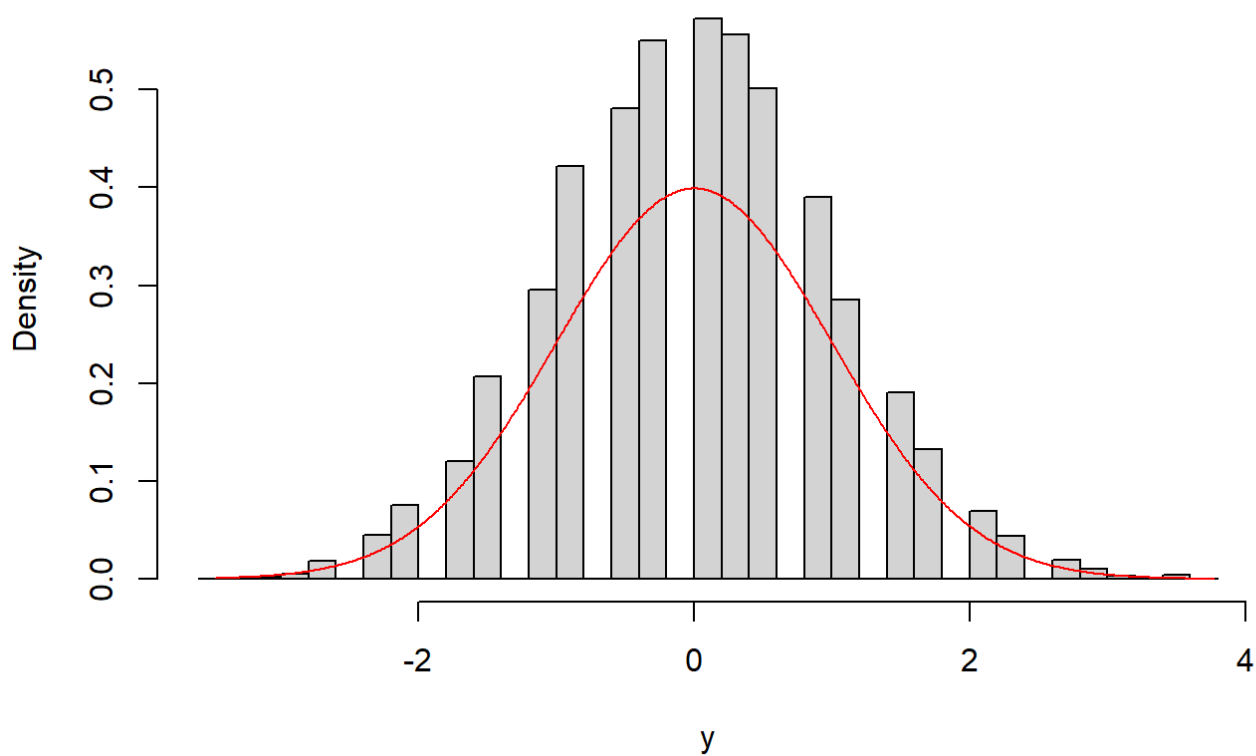
Histogram of y

```
p=0.4
n=50
nsamples=10000
samples=numeric(nsamples)
for(i in 1:nsamples){
    samples[i]=rbinom(1,size=n,p=p)
}
x=samples
y=scale(x)
k=seq(min(y),max(y),by=0.01)
gpmf=dnorm(k,mean = mean(y),sd=sd(y))
hist(y,breaks=40,freq=FALSE)
lines(k,gpmf,lwd=1,col="red")
```

Histogram of y


```
p=0.4
n=100
nsamples=10000
samples=numeric(nsamples)
for(i in 1:nsamples){
    samples[i]=rbinom(1,size=n,p=p)
}
x=samples
y=scale(x)
k=seq(min(y),max(y),by=0.01)
gpmf=dnorm(k,mean = mean(y),sd=sd(y))
hist(y,breaks=40,freq=FALSE)
lines(k,gpmf,lwd=1,col="red")
```

Histogram of y


```
p=0.4
n=500
nsamples=10000
samples=numeric(nsamples)
for(i in 1:nsamples){
  samples[i]=rbinom(1,size=n,p=p)
}
x=samples
y=scale(x)
k=seq(min(y),max(y),by=0.01)
gpmf=dnorm(k,mean = mean(y),sd=sd(y))
hist(y,breaks=40,freq=FALSE)
lines(k,gpmf,lwd=1,col="red")
```
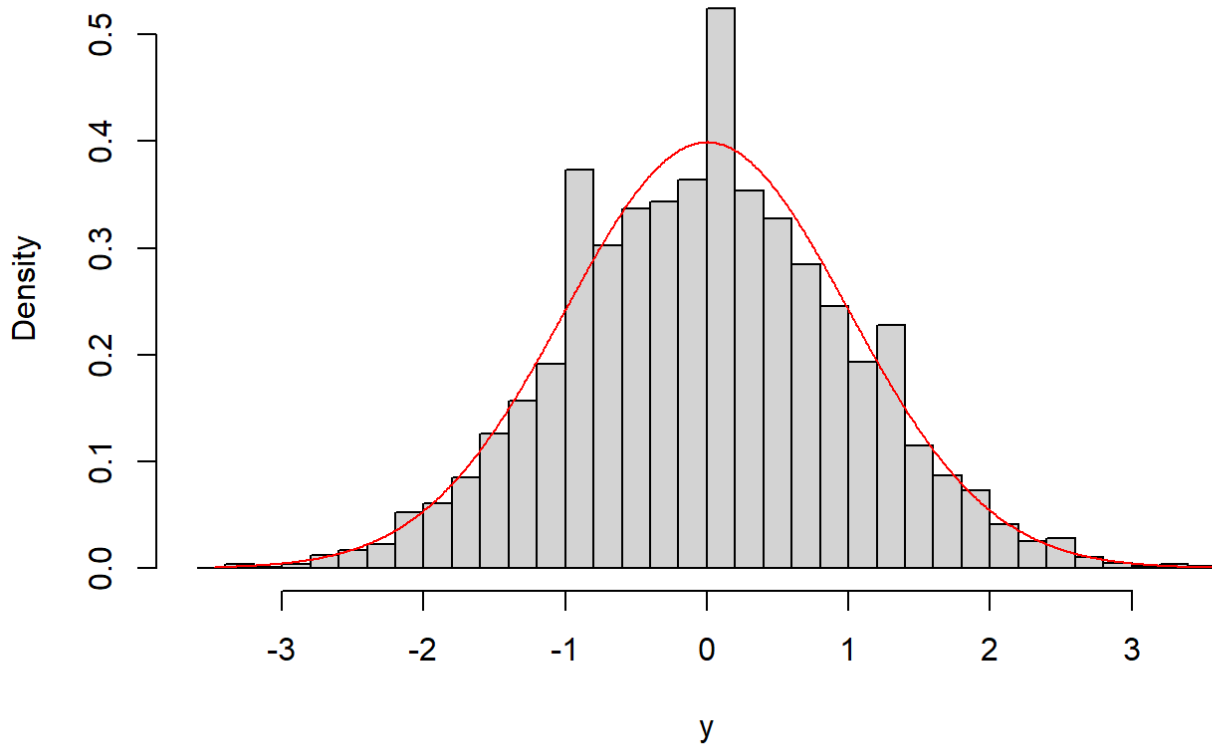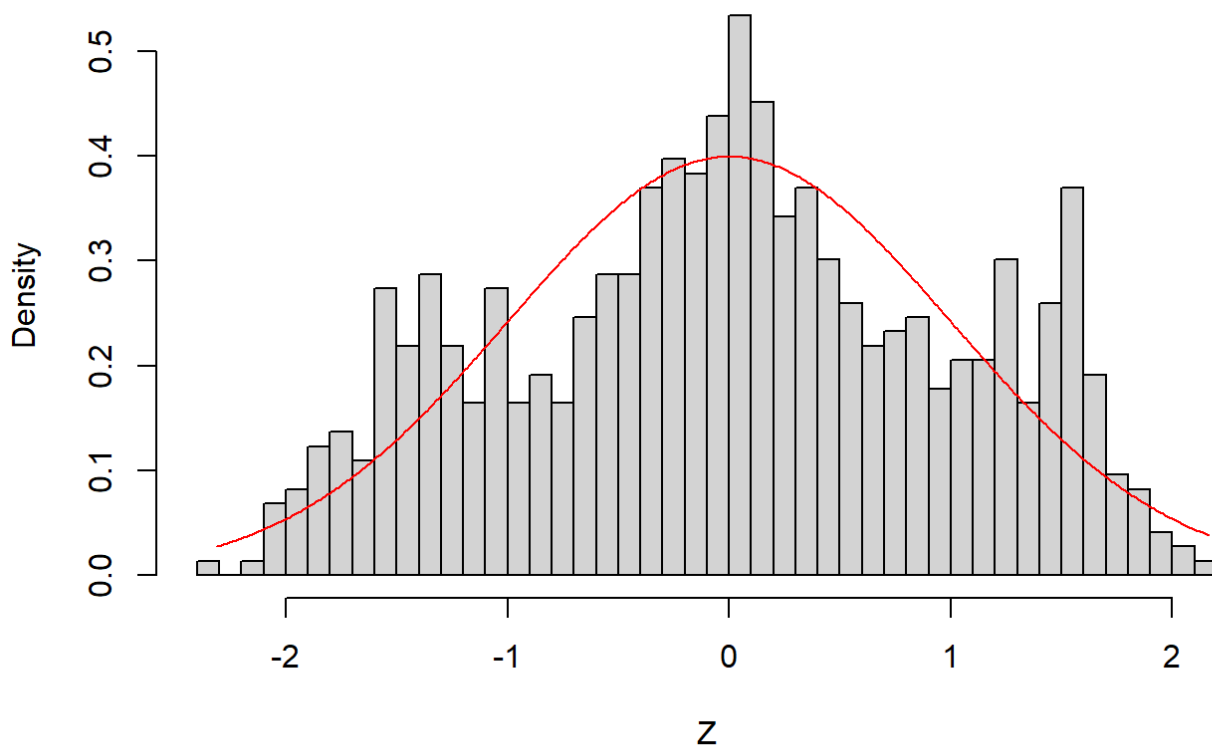
Histogram of y


```r
library(MASS)
df=read.csv("bikerental.csv")[,2:13]
Y=df[,12]
Z=scale(Y)
k=seq(min(Z),max(Z),by=0.01)
fit_norm=fitdistr(Z,densfun = "normal")
gpdf=dnorm(k,mean=fit_norm$estimate[1],sd=fit_norm$estimate[2])
hist(Z,breaks = 40,freq=FALSE,main="Density Histogram")
lines(k,gpdf,col="red")
```

Density Histogram

```
X=df[,1:11]
Xs=as.data.frame(scale(X))
mcor=numeric(11)
for (i in 1:11){
  mcor[i]=abs(cor(Xs[,i],Z))
}
si=order(mcor,decreasing=TRUE)
st=si[1:2]
```
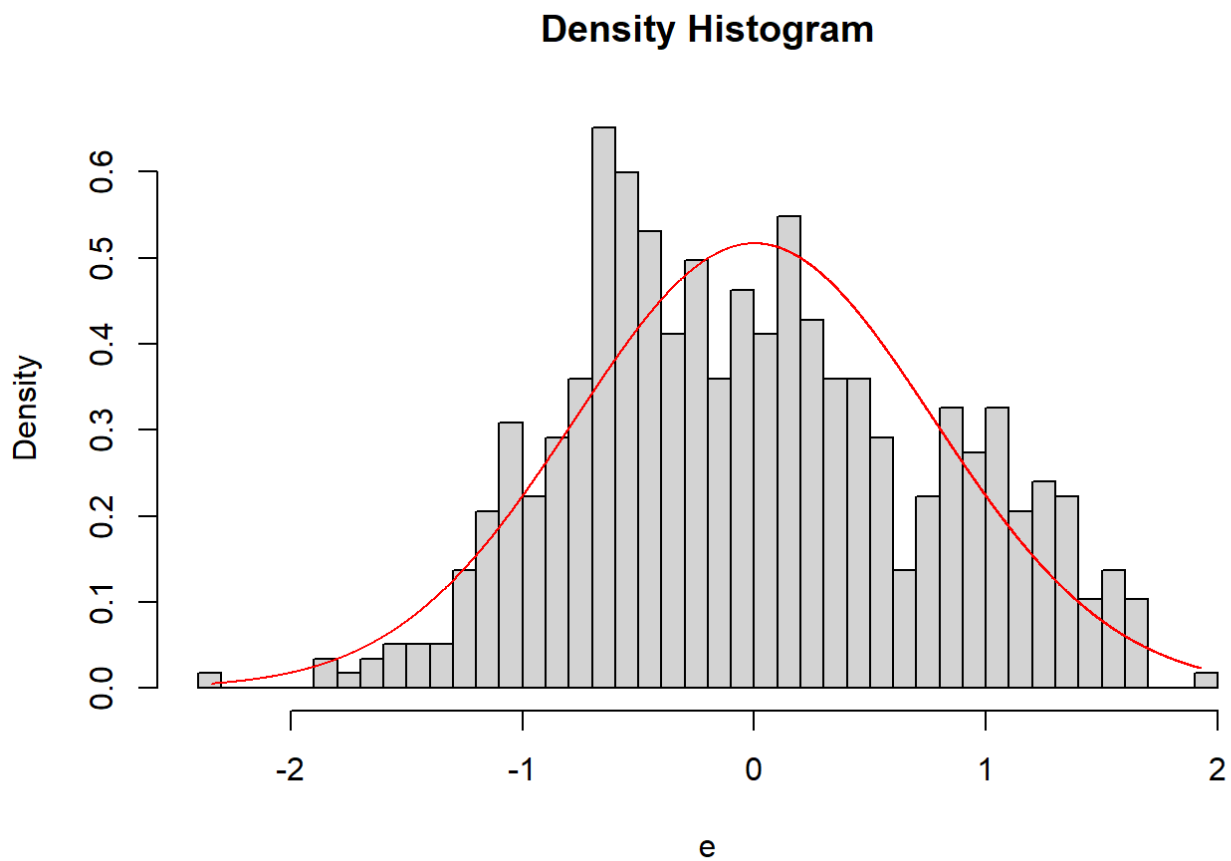
```
Xn=Xs[,st]
ndata=cbind(Xn,Z)
nr=nrow(ndata)
s=sample(x=1:nr,size=0.8*nr,replace = FALSE)
traindata=ndata[s,]
testdata=ndata[-s,]
mymod=lm(Z~.,data=traindata)
ytrainpred=predict(mymod,traindata)
ytestpred=predict(mymod,testdata)
trainmae=mean(abs(traindata[,3]-ytrainpred))
testmae=mean(abs(testdata[,3]-ytestpred))
sprintf("Train MAE:%f",trainmae)
```

```
## [1] "Train MAE:0.637937"
```

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

```
## [1] "Test MAE:0.645620"
```

```
e=traindata[,3]-ytrainpred
emy=mean(e)
emsd=sd(e)
ek=seq(min(e), max(e),by = 0.001)
emd=dnorm(ek,mean=emy,sd=emsd)
hist(e,breaks = 40,freq=FALSE,main="Density Histogram")
lines(ek,emd,lwd=1,col='red')
```


Density Histogram

```
k=0
trainmae=numeric(55)
testmae=numeric(55)
for(i in 1:10){
  t=i+1
  for(j in t:11){
st=c(i,j)
Xn=Xs[,st]
ndata=cbind(Xn,Z)
nr=nrow(ndata)
s=sample(x=1:nr,size=0.8*nr,replace = FALSE)
traindata=ndata[s,]
testdata=ndata[-s,]
mymod=lm(Z~.,data=traindata)
ytrainpred=predict(mymod,traindata)
ytestpred=predict(mymod,testdata)
k=k+1
trainmae[k]=mean(abs(traindata[,3]-ytrainpred))
testmae[k]=mean(abs(testdata[,3]-ytestpred))
  }
}
print(trainmae)
```

```
##  [1] 0.5960723 0.7374361 0.7508974 0.7518365 0.7510825 0.7034259 0.6109153
##  [8] 0.6319980 0.7334710 0.7322528 0.6245360 0.6902630 0.6827894 0.6793881
## [15] 0.6284046 0.4378438 0.4368827 0.6989564 0.6472395 0.7809827 0.7713535
## [22] 0.7758707 0.7553478 0.6239511 0.6286171 0.7720358 0.7625447 0.8115451
## [29] 0.8148305 0.7768751 0.6455372 0.6386611 0.8068146 0.7917083 0.8134088
## [36] 0.7710704 0.6452240 0.6371343 0.8054245 0.8083729 0.7869699 0.6424942
## [43] 0.6332564 0.8072370 0.7856938 0.6184959 0.6220174 0.7689280 0.7545960
## [50] 0.6134461 0.6314793 0.6379572 0.6160859 0.6309554 0.7850938
```

```
print(testmae)
```

```
##  [1] 0.5477490 0.7648647 0.7527893 0.7458047 0.7583896 0.7573349 0.6662903
##  [8] 0.5871631 0.7576348 0.7867936 0.6804744 0.6703936 0.7033692 0.7025708
## [15] 0.6889175 0.4542717 0.4390733 0.6476542 0.6781323 0.7746016 0.8107794
## [22] 0.8143830 0.6945383 0.6705071 0.6305827 0.7652136 0.8062430 0.8169890
## [29] 0.8053639 0.7596590 0.6382159 0.6305419 0.8035856 0.7966896 0.8206631
## [36] 0.7828600 0.6273114 0.6191319 0.8342747 0.7375318 0.7373656 0.6407747
## [43] 0.6489612 0.8250276 0.8264801 0.6174664 0.5753845 0.8094079 0.7858629
## [50] 0.7073101 0.5961818 0.6204979 0.6189797 0.6431669 0.7699637
```