To compare the least squares fitted equation with that of the Principal components, in this example we use Monte Carlo simulation to generate a set of observations from a bivariate normal distribution.

a) Use function `mvrnorm` from library `MASS` to generate 250 observations from a bivariate normal with means 70 and 162, standard deviations 3 and 14, and correlation -0.80.

b) Fit a least squares line to the data set.

c) Make a scatterplot.

d) Find principal components of variables x and y

e) Add PC1 axis to the scatterplot. Which fitted line is closest to the data set?

f) Add PC2 axis to the scatterplot.

g) Add an ellipse that encloses the 95% of the data set.

```
# parameters
mx = 70
sdx = 3
my = 162
sdy = 14
rho = -0.80
mu = c(mx,my)

# create covariance matrix
cova = rho*sdx*sdy
#  -33.6
aux = c(sdx^2,cova,cova,sdy^2)
sigma = matrix(aux,nrow=2)
sigma
#       [,1]  [,2]
#[1,]   9.0 -33.6
#[2,] -33.6 196.0

# correlation matrix
cov2cor(sigma)
#      [,1] [,2]
#[1,]  1.0 -0.8
#[2,] -0.8  1.0

# generate bivariate normal observations
library(MASS)
set.seed(5)
n = 250
d0 = mvrnorm(n,mu,sigma)
d0 = data.frame(d0)
x = d0[,1]
y = d0[,2]

# least squares line
plot(y~x,pch=19,cex=0.6)

m1 = lm(y~x)
abline(m1)
abline(h=my,lty=2)
abline(v=mx,lty=2)
grid()

# same scaling
plot(y~x,pch=19,cex=0.60,xlim=c(20,120),ylim=c(110,210))
abline(m1)
grid()
```

```
# principal components
pc1 = prcomp(d0)
names(pc1)         # "sdev"     "rotation" "center"   "scale"    "x"
d1 = pc1$x

# eigenvals
pc1$sdev^2
# [1] 199.50901   3.26403

# eigenvectors
rot = pc1$rotation
#          PC1       PC2
#X1  0.1684984 0.9857019
#X2 -0.9857019 0.1684984

# 1st PC axis, largest variance
slope1 = rot[2,1]/rot[1,1]
int1 = my - mx*slope1

# 2nd PC axis, smallest variance
slope2 = rot[2,2]/rot[1,2]
int2 = my - mx*slope2

# PC1 axis
abline(int1,slope1, col="red", lty=2, lwd=2)

# ls lines for orthogonal and vertical distances

# PC2 axis added
abline(int2,slope2, col="blue", lty=2)
legend("topright",c("PC1","PC2","LSq"),lty=c(1,1),col=c("red","blue",1))

# which line fits best?

# enclose 95% of obs in an ellipse
library(mixtools)
alpha = 0.05
ellipse(mu,sigma,alpha,2000,col="green")

# ellipse with 2000 points resolution
```

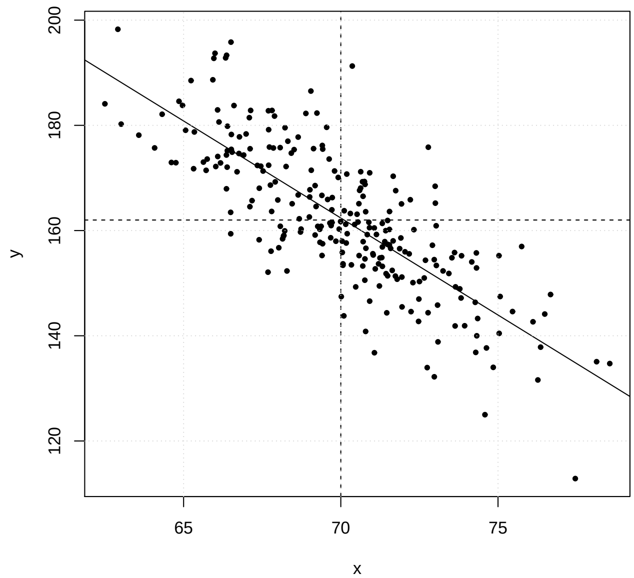

Figure 1: Fitted OLS line

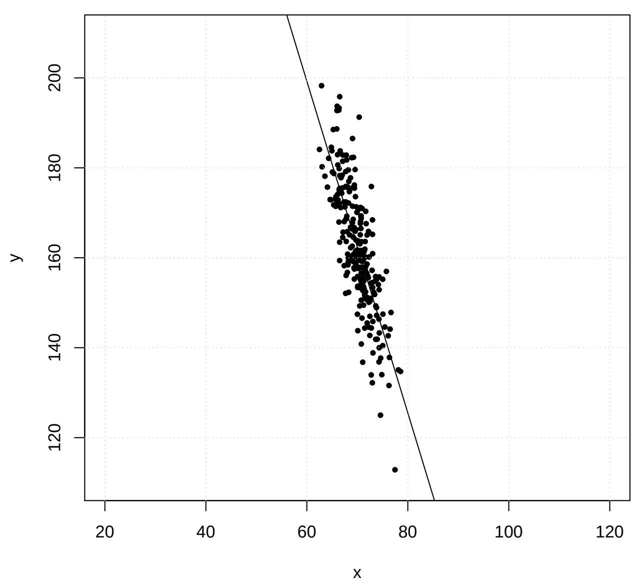

Figure 2: Fitted OLS line in a box (no stretching)

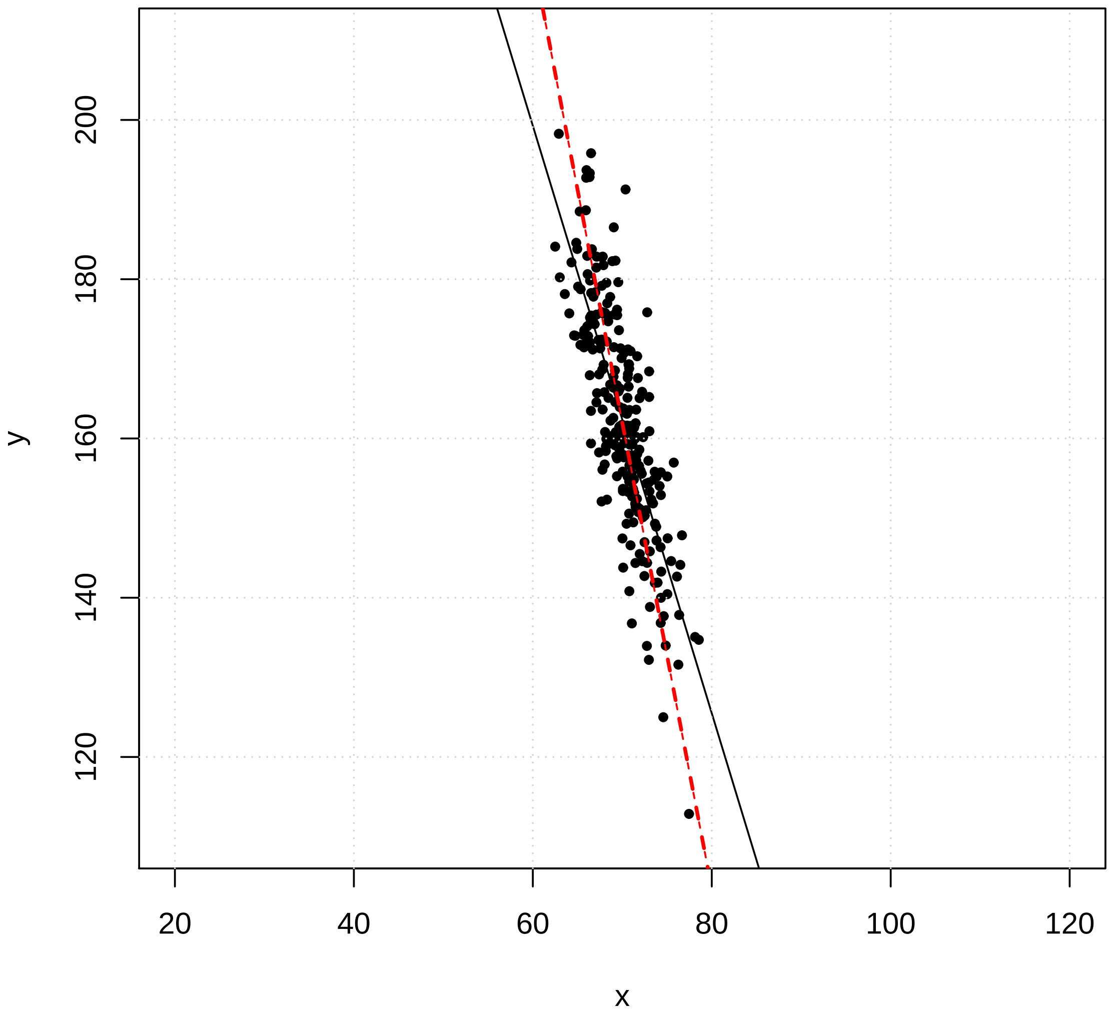

Figure 3: Fitted lines for vertical and orthogonal distances

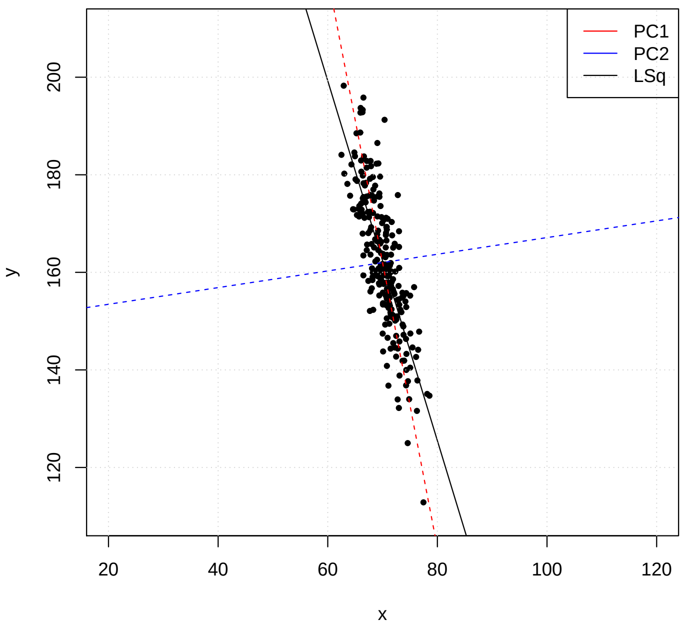

Figure 4: Scatterplot on principal components axes

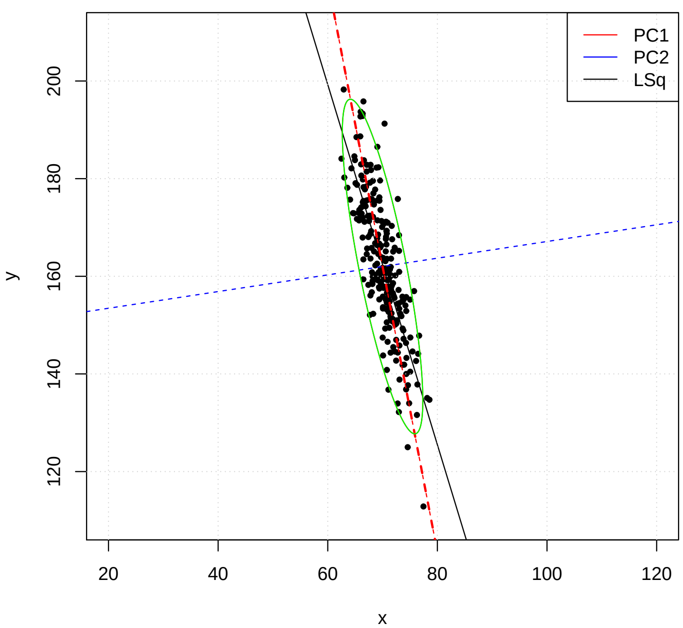

Figure 5: Scatterplot with 95% data enclosed in an ellipse