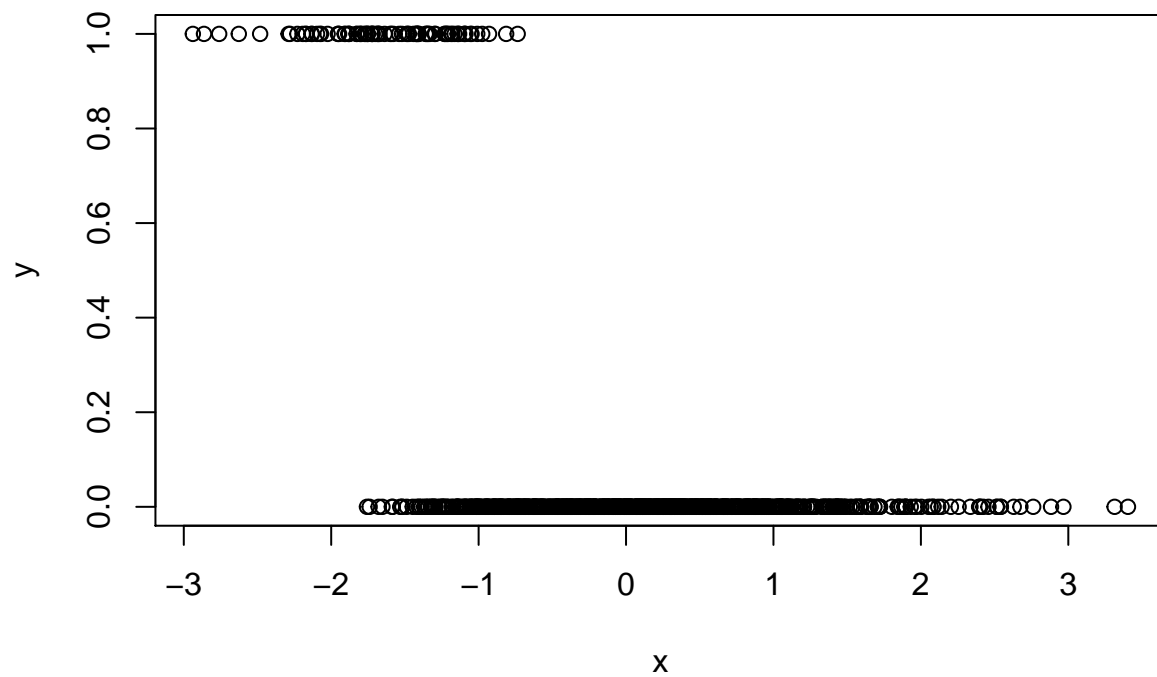


R Notebook

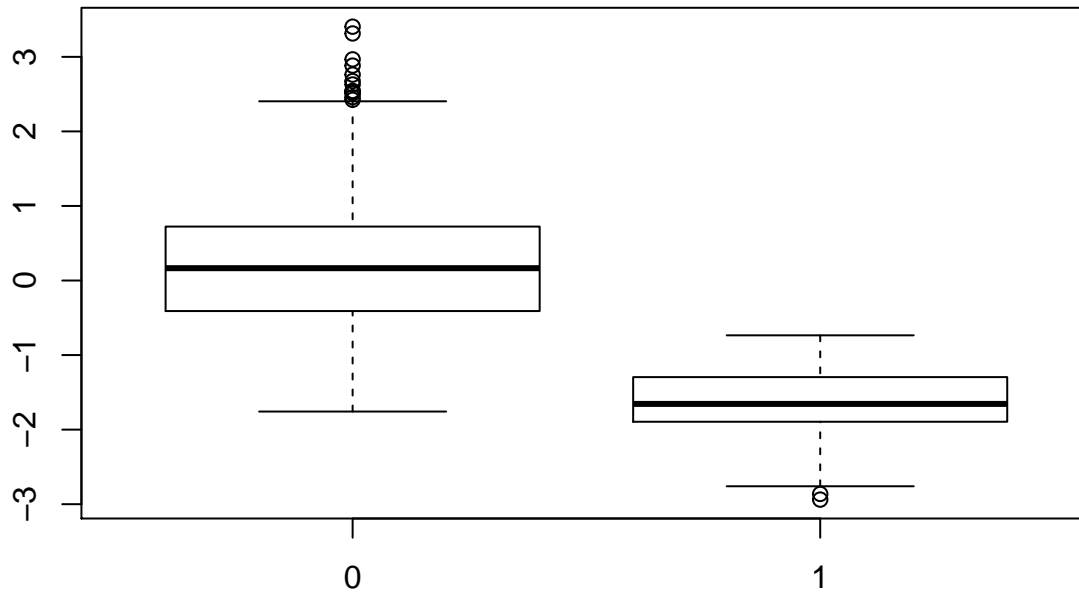
```
options(scipen=999)
set.seed(1234)
LR2 <- read.table(file="./LR2.csv", header = TRUE, sep = ",")
names(LR2)
```

```
## [1] "y" "x"
```

```
attach(LR2)
plot(x,y)
```



```
boxplot(x~y)
```



Assignment 2

Exercise 1

$$\Pr(Y = 1|X = x) = \Phi(\beta_0 + \beta_1 x)$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}t^2} dt$$

$$\Phi(z) = P(Z \leq z), Z \sim \mathcal{N}(0, 1)$$

Thus $\Phi(\beta_0 + \beta_1 x) = P(Z \leq z)$

Write an R function that computes the maximum likelihood estimate, $\text{\texttt{\textbackslash mathscr{L}}} \text{\texttt{\textbackslash left(\texttt{\textbackslash beta_0, \texttt{\textbackslash beta_1} \texttt{\textbackslash right)}}$, along with bootstrapped errors.

```
# objective function
probit_mle_b <- function(x,y) {

  # probit link
  #
  probit <- function(b,x,y) {
    n <- length(y)
    ll <- 0
    for(i in 1:n) {
      z <- b[1]+b[2]*x[i]
      z <- pnorm(z, mean=0, sd=1, log.p = FALSE)
      ll <- ll + z*(y[i]==1) + z*(y[i]==0)
      # print(ll)
    }
    # print(ll)
    return(-ll)
  }
}
```

```

# mle
#
obj = optim(c(0,0), probit, x=x, y=y)

coef1 <- obj$par[1]
coef2 <- obj$par[2]

## Bootstrap
##

B <- 100

b_boot = matrix(rep(0,2*B),B,2)
n <- length(y)
for (i in 1:B){
  # indices for the i-th bootstrap subsample
  ind_ = sample(n,n,replace=TRUE)
  # input vector in the subsample
  xb = x[ind_]
  # output vector in the subsample
  yb = y[ind_]

  # compute the maximum likelihood estimates
  obj = optim(c(0,0), probit, x=xb, y=yb)

  b_boot[i,1] = obj$par[1]
  b_boot[i,2] = obj$par[2]
}

return(
  list(
    coefficients = c(coef1,coef2),
    standard_errors = c(sd(b_boot[,1]),sd(b_boot[,2])),
    predict = ifelse(pnorm(coef1+coef2*x,0,1)>1/2,1,0)
  )
)
}

# Apply the probit estimator to LR2
est <- probit_mle_b(LR2$x,LR2$y)
est$coefficients

## [1] 15.582031 2.621094

glm.est <- glm(y~x,family=binomial(link = "probit"))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

# summary(glm.est)
plot(x, y)
abline(h = 1, lty=2)
abline(h = 0, lty=2)
x0 <- seq(min(x),max(x),length.out = length(x))
y0 <- sort(predict(glm.est,list(x),type="response"))

```

```
lines(x0,y0,lwd=1.5,col="dodgerblue")
```

