```
---
title: "1P22CS020_ML_ASSIGNMENT_2"
author: "Nisha Mohanraj"
date: "2023-09-17"
output: html_document
---
```

1. In this exercise you will create some simulated data and will fit simple linear regression models to it. Make sure to use set.seed(1) prior to starting part (a) to ensure consistent results.rnorm

a. Using the rnorm() function, create a vector, “x”, containing 100 observations drawn from a N(0,1) distribution. This represents a feature, X.
b. Using the rnorm() function, create a vector, “eps”, containing 100 observations drawn from a N(0,0.25) distribution.
c. Using “x” and “eps”, generate a vector “y” according to the model
$Y=-1+0.5X+\varepsilon$.
What is the length of the vector “y” ? What are the values of β0 and β1 in this linear model ?

d. Create a scatterplot displaying the relationship between “x” and “y”. Comment on what you observe.

e. Fit a least squares linear model to predict “y” using “x”. Comment on the model obtained. How do β^0 and β^1 compare to β0 and β1 ?

f. Display the least squares line on the scatterplot obtained in (d). Draw the population regression line on the plot, in a different color. Use the legend() function to create an appropriate legend.

Answers

````
```{r}
set.seed(1)

#a
x=rnorm(100,0,1)

#b
eps=rnorm(100,0,sqrt(0.25))

#c
y=-1+0.5*x+eps
length(y)
#β0=-1 and β1 =0.5


#d
plot(x,y)
# The model shows the linear relation ship


#e
fit=lm(y~x)
summary(fit)
#β^o and β^1 are practically equal to βo and β1. As expected, β^o and β^1 are
statistically significant


#f
plot(x,y)
abline(fit,col="red",lty=1,lwd=2)

abline(-1,0.5,col="green",lty=2,lwd=3)
````
````

```
legend("bottomright", c(" Least square", "pop-Regression"), col = c("red", "green"), lty = c(1, 2))

```

2. This problem involves the “Boston” data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

a. For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response ? Create some plots to back up your assertions.

```{r}
library(MASS)
data("Boston")
attach(Boston)
View(Boston)


```
Question-A

```{r}
fit_zn=lm(crim~zn)
summary(fit_zn)
```

```{r}
fit_indus=lm(crim~indus)
summary(fit_indus)
```

```{r}
fit_chas=lm(crim~chas)
summary(fit_chas)
```
```{r}
fit_nox=lm(crim~nox)
summary(fit_nox)
```
```{r}
fit_rm=lm(crim~rm)
summary(fit_rm)
```

```{r}
fit_age=lm(crim~age)
summary(fit_age)
```
```{r}
fit_dis=lm(crim~dis)
summary(dis)
```
```{r}
fit_rad=lm(crim~rad)
summary(fit_rad)
```
```{r}
fit_tax=lm(crim~tax)
summary(fit_tax)
```

````
```{r}
fit_pratio=lm(crim~ptratio)
summary(fit_pratio)
```
```{r}
fit_black=lm(crim~black)
summary(fit_black)
```
```{r}
fit_lstat=lm(crim~lstat)
summary(fit_lstat)
```
```{r}
fit_medv=lm(crim~medv)
summary(fit_medv)

plot(crim,zn)

plot(crim,dis)

plot(crim,medv)
```
````

b. Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis $H0:\beta j=0$?

````
```{r}
fit_all=lm(crim~.,data = Boston)

summary(fit_all)

##Based on the Multiple linear Regression , only zn, dis, rad, black, and medv have a
significant association with crim (p-value is below 0.05) which means we can reject the
null hypothesis
```
````