STAT504 Assignment3

Anuradha Ramachandran

2023-02-13

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(plot3D)
```

Problem 1

1.1 Best Linear Predictor

$L = argmin_{\beta} E[(Y - X_i^T \beta)^2]$
Differentiating L with respect to β and equating to 0.
$-2E[X_i(Y - X_i^T \beta)] = 0$
$E[X_i Y_i] = E[X_i X_i^T]\beta$
$\beta = E[X_i X_i^T]^{-1} E[X_i Y_i]$

1.2 If CEF is linear then it equals the linear regression

$Y_i = X_i^T \beta^* + \epsilon_i$
$X_i Y_i = X_i X_i^T \beta^* + X_i \epsilon_i$
$E[X_i Y_i] = E[X_i X_i^T]\beta^* + E[X_i \epsilon_i]$ Since ϵ_i is uncorrelated with any function of $X_i, E[X_i \epsilon_i] = 0$
$\beta^* = E[X_i X_i^T]^{-1} E[X_i Y_i] = \beta_{OLS}$

1.3 Linear regression is the best linear approximation to the CEF

$L = argmin_{\beta} E[(E[X_i Y_i] - X_i^T \beta)^2]$
From part(1.1), $\beta = E[X_i X_i^T]^{-1} E[X_i Y_i]$
Substitute $Y_i = E[Y_i|X_i]$ in β formula.
$\beta = E[X_i X_i^T]^{-1} E[X_i E[Y_i|X_i]] = E[X_i X_i^T]^{-1} E[E[X_i Y_i|X_i]] = E[X_i X_i^T]^{-1} E[X_i Y_i]$

1.4 Saturated Regression

CEF can be written in terms of Linear combination of covariates such as: $X = [1, X_{1i}, X_{2i}, X_{1i}X_{2i}]^T$
$\beta = [\beta_0, \beta_1, \beta_2, \beta_{12}]^T$

CEF = $X^T\beta$
Since X_{1i}, X_{2i} are binary random variables, CEF can be written as:
$E[Y_i|X_{1i}=0, X_{2i}=0] = \beta_0$
$E[Y_i|X_{1i}=1, X_{2i}=0] = \beta_0 + \beta_1$
$E[Y_i|X_{1i}=0, X_{2i}=1] = \beta_0 + \beta_2$
$E[Y_i|X_{1i}=1, X_{2i}=1] = \beta_0 + \beta_1 + \beta_2 + \beta_{12}$
Meaning of parameters,
β_0: the CEF value when both covariates are zero
$\beta_1 : E[Y_i|X_{1i}=1, X_{2i}=0] - E[Y_i|X_{1i}=0, X_{2i}=0]$, change in CEF when X_{1i} is changed by one unit.
$\beta_2 : E[Y_i|X_{1i}=0, X_{2i}=1] - E[Y_i|X_{1i}=0, X_{2i}=0]$, change in CEF when X_{2i} is changed by one unit.
$\beta_{12} : E[Y_i|X_{1i}=1, X_{2i}=1] + E[Y_i|X_{1i}=0, X_{2i}=0] - E[Y_i|X_{1i}=0, X_{2i}=1] - E[Y_i|X_{1i}=1, X_{2i}=0]$, is the interaction coefficient and measures how each slope changes when the value of X_{1i}, X_{2i} is changed by one unit.

1.5 Quadratic regression

If CEF = $E[Y_i|X_{1i}] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2$
CEF can be estimated using linear regression as $E[Y_i|X_{1i}] = X_i^T\beta$, where
$X = [1, X_{1i}, X_{1i}^2]^T$ and
$\beta = [\beta_0, \beta_1, \beta_2]^T$
Now β can be estimated using OLS formula.

1.6 FWL theorem

Consider $cov(Y_i, D_i^{\perp X_i}) = cov(\tau_r D_i + X_i^T\beta_r + e_{r,i}, D_i^{\perp X_i})$
$cov(Y_i, D_i^{\perp X_i}) = cov(\tau_r D_i + X_i^T\beta_r, D_i^{\perp X_i}) + cov(e_{r,i}, D_i^{\perp X_i})$
$cov(e_{r,i}, D_i^{\perp X_i}) = 0$
$cov(Y_i, D_i^{\perp X_i}) = \tau_r cov(D_i, D_i^{\perp X_i}) + \beta_r cov(X_i^T, D_i^{\perp X_i})$
$cov(X_i^T, D_i^{\perp X_i}) = 0$
$cov(Y_i, D_i^{\perp X_i}) = \tau_r var(D_i^{\perp X_i})$
$\tau_r = \frac{cov(Y_i, D_i^{\perp X_i})}{var(D_i^{\perp X_i})} = \frac{cov(Y_i^{\perp X_i} + X_i^T E[X_i X_i^T]^{-1} E[X_i Y_i], D_i^{\perp X_i})}{var(D_i^{\perp X_i})}$
$\tau_r = \frac{cov(Y_i^{\perp X_i}, D_i^{\perp X_i})}{var(D_i^{\perp X_i})}$

1.7 OVB theorem

From FWL theorem, $\tau_r = \frac{cov(Y_i^{\perp X_i}, D_i^{\perp X_i})}{var(D_i^{\perp X_i})}$
$\tau_r = \frac{cov((\tau D_i + X_i^T\beta + \gamma Z_i + e_i)^{\perp X_i}, D_i^{\perp X_i})}{var(D_i^{\perp X_i})}$
By FWL property,$(X_i^T\beta)^{\perp X_i} = 0$
$\tau_r = \frac{cov(\tau D_i^{\perp X_i} + \gamma Z_i^{\perp X_i}, D_i^{\perp X_i})}{var(D_i^{\perp X_i})}$
$\tau_r = \frac{\tau cov(D_i^{\perp X_i}, D_i^{\perp X_i})}{var(D_i^{\perp X_i})} + \gamma\frac{cov(Z_i^{\perp X_i}, D_i^{\perp X_i})}{var(D_i^{\perp X_i})}$
$\tau_r = \tau + \gamma\delta$
γ is the predictive impact of Z on Y and δ is the imbalance of Z among levels of D. (Source: Scribe notes Lecture 6)
OVB theorem, explains the relationship between coefficient estimates when certain variables are omitted or controlled. (Source: Angirst and Pischlke book section 3.2.2).

1.8 Partialling out property

(a)
$V = X_1 + X_2$
$V_i^{\perp Z} = V_i - Z_i^T E[Z_i Z_i^T]^{-1} E[Z_i V_i]$
$V_i^{\perp Z} = X_{1i} + X_{2i} - Z_i^T E[Z_i Z_i^T]^{-1} E[Z_i (X_{1i} + X_{2i})]$
$V_i^{\perp Z} = X_{1i} - Z_i^T E[Z_i Z_i^T]^{-1} E[Z_i X_{1i}] + X_{2i} - Z_i^T E[Z_i Z_i^T]^{-1} E[Z_i X_{2i}] = X_{1i}^{\perp Z} + X_{2i}^{\perp Z}$

(b) $e^{\perp Z} = e_i - Z_i^T E[Z_i Z_i^T]^{-1} E[Z_i e_i]$
e_i is uncorrelated with Z, $E[Z_i e_i] = 0$
$e^{\perp Z} = e$

(c) OVB with partial R^2 Let $Z = [Z_1, \ldots, Z_p]^T$
Consider $Z_1^{\perp Z}$ is the residual of the best linear predictor of Z_1 on Z.
The BLP of of Z_1 on Z is given by
$Z_1 = Z^T \beta + e$ where e $= Z_1^{\perp Z}$
$Z_1^{\perp Z} = Z_1 - Z^T \beta$ where
$Z^T = [Z_1, \ldots, Z_p]$
$\beta = \frac{cov(Z, Z_1)}{var(Z_1)} = [\frac{cov(Z_1, Z_1)}{var(Z_1)}, \frac{cov(Z_2, Z_1)}{var(Z_1)}, \ldots, \frac{cov(Z_p, Z_1)}{var(Z_1)}]^T$
$\beta = [1, 0, \ldots, 0]^T$
Hence $Z_1^{\perp Z} = Z_1 - Z_1 = 0$
Similarly $Z_2^{\perp Z}, \ldots, Z_p^{\perp Z} = 0$
$Z^{\perp Z} = [Z_1^{\perp Z}, \ldots, Z_p^{\perp Z}]^T = 0$

1.9 Partial R^2

$R^2_{y \sim Z+D+X} = 1 - \frac{var(Y^{\perp Z,D,X})}{var(Y)}$
$R^2_{y \sim D+X} = 1 - \frac{var(Y^{\perp D,X})}{var(Y)}$
Substituting $R^2_{y \sim Z+D+X}$ and $R^2_{y \sim D+X}$ in the definition of $R^2_{y \sim Z|D,X}$, we get
$R^2_{y \sim Z|D,X} = \frac{var(Y^{\perp D,X}) - var(Y^{\perp Z,D,X})}{var(Y^{\perp D,X})} = 1 - \frac{var(Y^{\perp Z,D,X})}{var(Y^{\perp D,X})}$
$Z^{\perp D,X} = Y^{\perp D,X} - Y^{\perp Z,D,X}$
$R^2_{y \sim Z|D,X} = \frac{var(Z^{\perp D,X})}{var(Y^{\perp D,X})}$
By definition of R^2,
$R^2_{y \sim Z|D,X} = cor(Y^{\perp D,X}, Z^{\perp D,X})^2 = cor(Z^{\perp D,X}, Y^{\perp D,X})^2 = R^2_{Z \sim Y|D,X}$

1.10

$(Bias)^2 = (\gamma\delta)^2$
$\gamma = \frac{cov(Y^{\perp D,X}, Z^{\perp D,X})}{var(Z^{\perp D,X})}$
$\delta = \frac{cov(D^{\perp X}, Z^{\perp X})}{var(D^{\perp X})}$
$(Bias)^2 = (\gamma\delta)^2 = \frac{cov(Y^{\perp D,X}, Z^{\perp D,X})^2}{var(Z^{\perp D,X})^2} \frac{cov(D^{\perp X}, Z^{\perp X})^2}{var(D^{\perp X})^2}$
$(\gamma\delta)^2 = \frac{cov(Y^{\perp D,X}, Z^{\perp D,X})^2 var(Y^{\perp D,X})}{var(Z^{\perp D,X})^2 var(Y^{\perp D,X})} \frac{cov(D^{\perp X}, Z^{\perp X})^2 var(Z^{\perp X})}{var(D^{\perp X})^2 var(Z^{\perp X})}$
$(\gamma\delta)^2 = cor^2(Y^{\perp D,X}, Z^{\perp D,X}) cor^2(D^{\perp X}, Z^{\perp X}) \frac{var(Z^{\perp X})}{var(Z^{\perp D,X})} \frac{var(Y^{\perp D,X})}{var(D^{\perp X})}$
$(\gamma\delta)^2 = R^2_{y \sim Z|D,X} R^2_{D \sim Z|X} \frac{1}{1 - R^2_{D \sim Z|X}} \frac{var(Y^{\perp D,X})}{var(D^{\perp X})}$

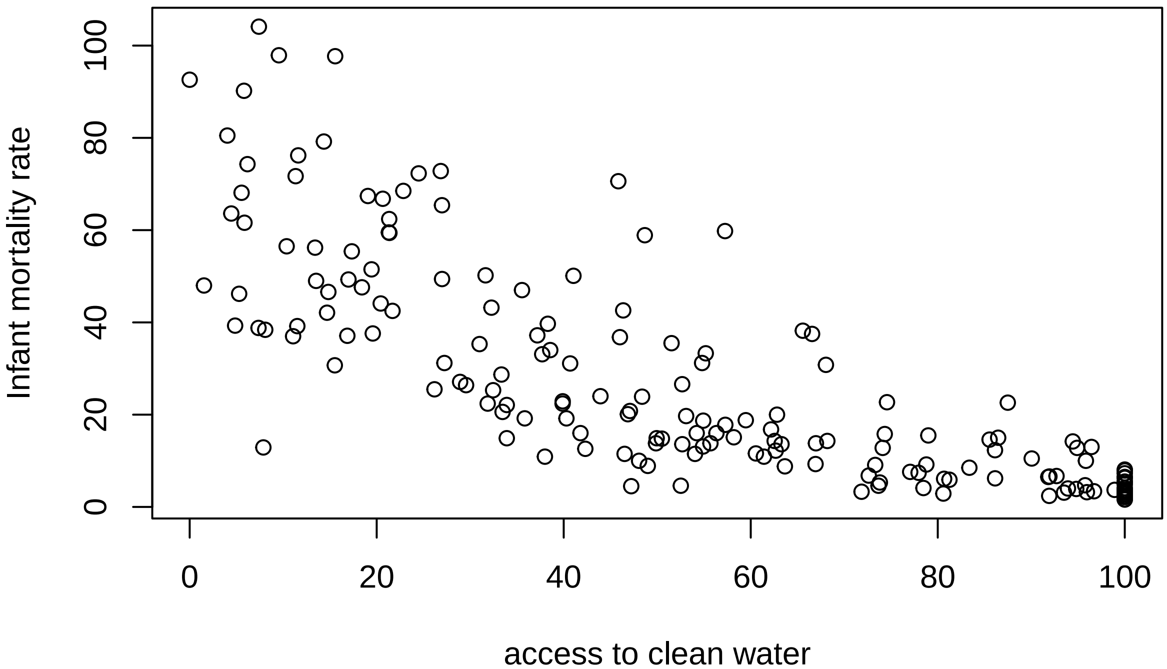
Problem 2

```
# loading data
data = read.csv("/Users/anuram/Library/Mobile Documents/com~apple~CloudDocs/MS Stats/Winter 2023/STAT50
```

2.1 (a)

Yes, there seems to be a association between infant mortality rate and access to clean water. There seems to be negative association between them, ie., as access to clean water increases, infant mortality rate decreases.

```
plot(data$epi_watsup, data$wdi_mortinftot, xlab = "access to clean water",
    ylab = "Infant mortality rate")
```



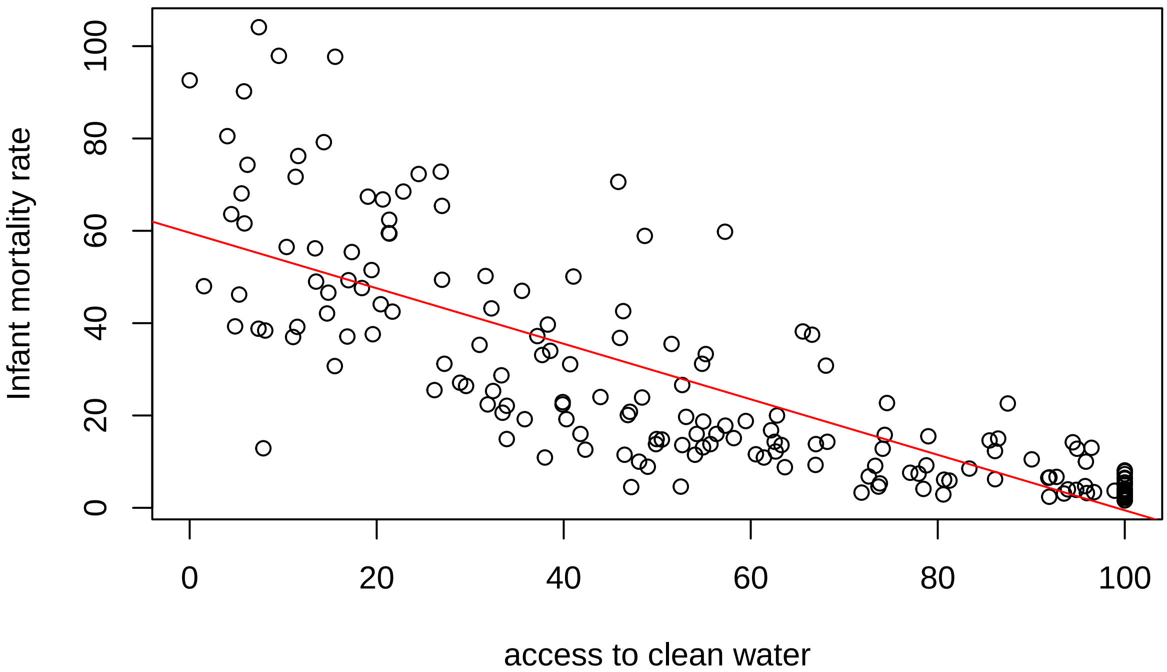
2.1 (b) Simple Linear Regression

Estimates of the simple linear regression model: Intercept: 59.5832: On average 60% infant mortality rate is observed when there is no access to clean water. Slope: -0.6013: When the access to clean water increases by 1 unit, on average a 6% decrease in infant mortality rate is observed.

```
ols = lm(data$wdi_mortinftot ~ data$epi_watsup)
theta.hat = coef(ols)["data$epi_watsup"]
theta.hat
```

```
## data$epi_watsup
##       -0.601256
```

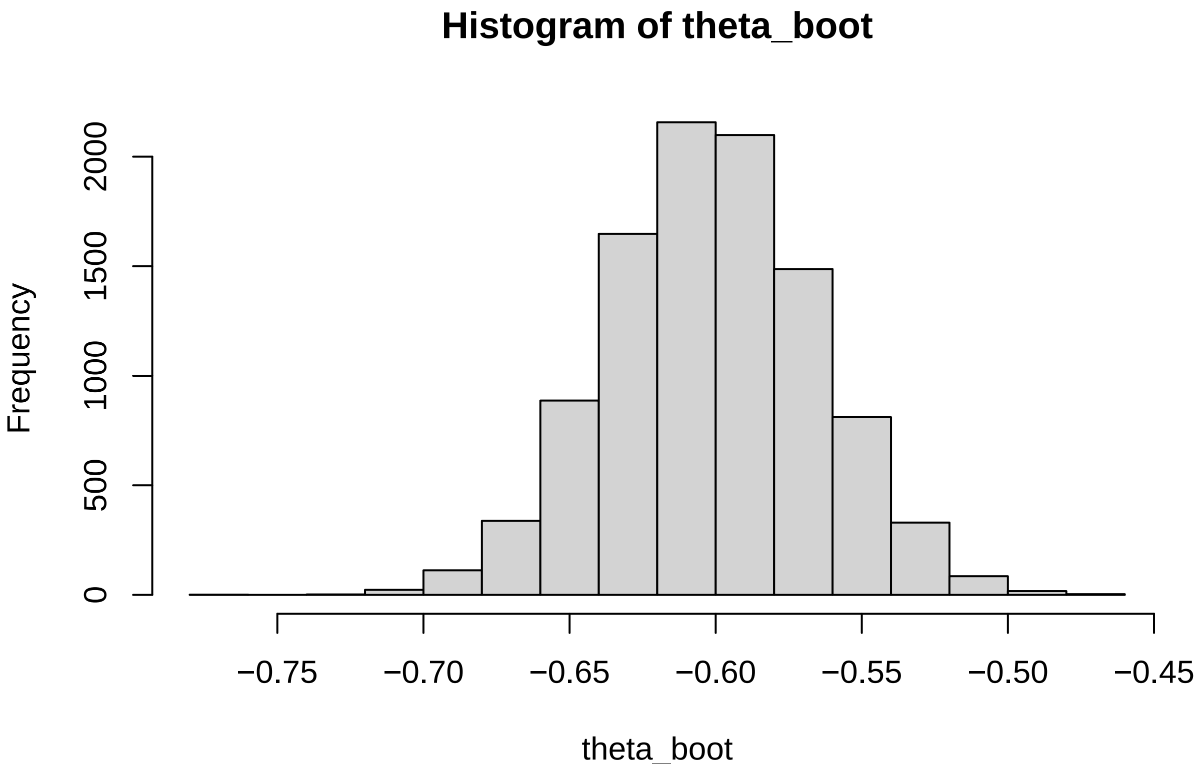
```
plot(data$epi_watsup, data$wdi_mortinftot, xlab = "access to clean water",
    ylab = "Infant mortality rate")
abline(ols, col = "red")
```



2.1 (c) 95% CI using non-parametric bootstrap

Regression coefficients fall within (-0.66 ,-0.54) range, 95% of the time.

```
set.seed(123)
# bootstrap empirical distribution
B = 10^4
n = 184
theta_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = data[idx_boot, ]
    ols_boot = lm(wdi_mortinftot ~ epi_watsup, data = data_boot)
    theta_boot[i] = coef(ols_boot)["epi_watsup"]
}
hist(theta_boot)
```




```r
# 95% confidence interval
se_boot <- sd(theta_boot)
z_crit <- qnorm(0.95)
low <- theta.hat - z_crit * se_boot
up <- theta.hat + z_crit * se_boot
c(low, up)
```

```
## data$epi_watsup data$epi_watsup
##      -0.6597247      -0.5427872
```

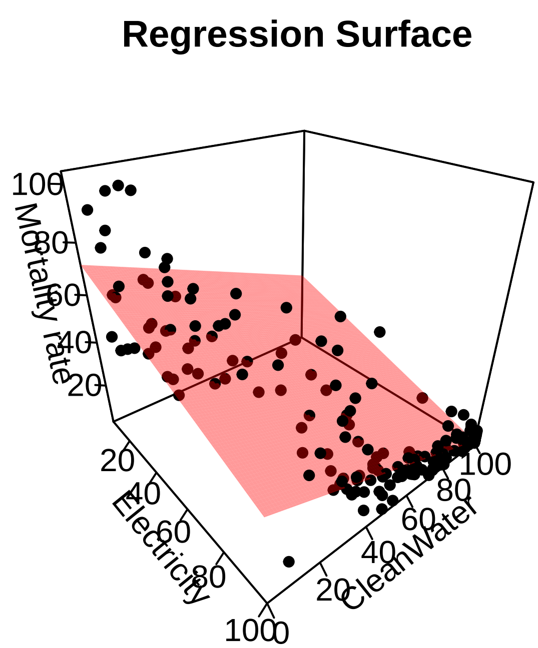
2.1 (d) Two covariate model

Linear model The linear model of infant mortality rate as a factor of both access to water and electricity is: $Infant\_mortality\_rate = 74.1860 - 0.3496(Access\_to\_cleanwater) - 0.3673(Access\_to\_electricity)$ The magnitude of coefficient of access to clean water using the two covariate model(0.36) is lower than the univariate model(0.60). $\beta_{yx.1z}$ is the coefficient in the linear model that explains only the effect of access to water on infant mortality rate without including the effects of electricity.

```
##
## Call:
## lm(formula = wdi_mortinftot ~ epi_watsup + wdi_accelectr, data = data)
##
## Coefficients:
##   (Intercept)     epi_watsup  wdi_accelectr
##       74.1860        -0.3496        -0.3673
```

```r
grid.size = 100
# x grid
x.grid    = seq(min(data$epi_watsup), max(data$epi_watsup), length.out = grid.size)
```

```r
# z grid
z.grid    = seq(min(data$wdi_accelectr), max(data$wdi_accelectr), length.out = grid.size)

# expand all combinations
grid.data = expand.grid(wdi_accelectr = z.grid, epi_watsup = x.grid)

# create predictions
y.hat.grid    <- matrix(predict(ols_2, newdata = grid.data),
                        nrow = grid.size,
                        ncol = grid.size)

# scatter plot with regression surface
scatter3D(x = data$wdi_accelectr, y = data$epi_watsup, z = data$wdi_mortinftot,  pch = 20, col="black",
          xlab = "Electricity", ylab = "CleanWater", zlab = "Mortality rate",  ticktype = "detailed",
          surf = list(x = z.grid, y = x.grid, z = y.hat.grid, col = "red", alpha = 0.4),
          phi = 30, theta = 50, # control the angle of the plot
          main = "Regression Surface")
```



Scatter plot

2.1 (e) FWL theorem

(i) linear model of X__i = epi__watsup on Zi = wdi accelectr

$X_i = -2.929 + 0.750 * Z_i$

```r
# linear model of X_i = epi watsup on Zi = wdi accelectr
mxz = lm(epi_watsup ~ wdi_accelectr, data = data)
coef(mxz)
```

```
##   (Intercept) wdi_accelectr
##    -2.9288002     0.7500156
```

```r
# residuals of this regression
x.z = resid(mxz)
```

(ii) linear model of Y_i = wdi_mortinftot on Zi = wdi accelectr
$Y_i = 75.209 - 0.629 * Z_i$

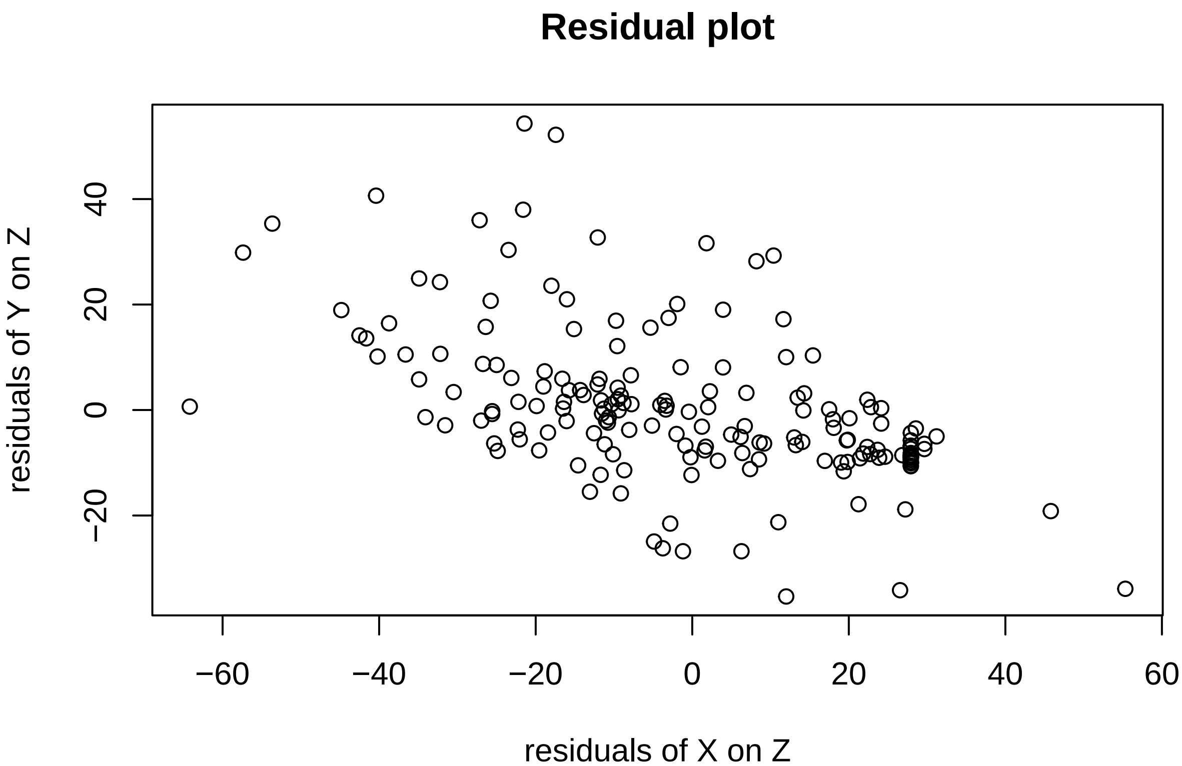
```
# linear model of Y_i = wdi_mortinftot on Zi = wdi
# accelectr
myz = lm(wdi_mortinftot ~ wdi_accelectr, data = data)
coef(myz)
```

```
##   (Intercept) wdi_accelectr
##    75.2099778    -0.6295686
```

```
# residuals of this regression
y.z = resid(myz)
```

(iii) Scatter plot of residuals
$y.z = 1.391127e^{-15} - 0.3496 * x.z$
The coefficient of $x.z$ is same as $\hat{\beta_{yx.1z}} = -0.349$

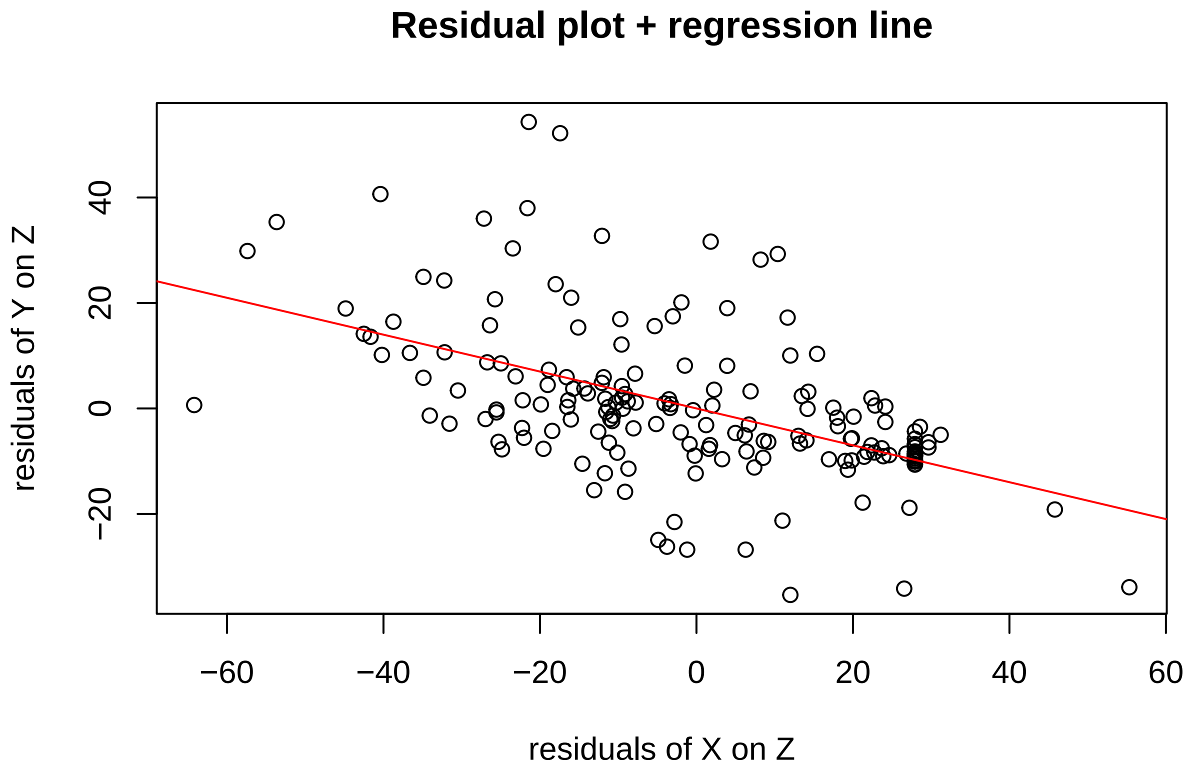
```
# scatter plot of residuals
plot(y.z ~ x.z, xlab = "residuals of X on Z", ylab = "residuals of Y on Z",
    main = "Residual plot")
```




```
# Regress y.z on x.z
residreg = lm(y.z ~ x.z)
coef(residreg)
```

```
##   (Intercept)           x.z
##  1.391127e-15 -3.496213e-01
```

```r
# scatter plot of residuals + regression line
plot(y.z ~ x.z, xlab = "residuals of X on Z", ylab = "residuals of Y on Z",
    main = "Residual plot + regression line")
abline(residreg, col = "red")
```



2.1 (f) OVB theorem

(i) Regress Zi = wdi accelectr on X_i = epi_watsup

$Z = 39.7519829 + 0.6850051 * X$

$\hat{\beta}_{zx.1} = 0.685$ is the slope of the linear model of access to electricity (Z) on access to clean water(X). This means that for every unit increase in access to clean water, the access to electricity increases by 0.685 units.

```r
# linear model of Zi = wdi accelectr on X_i = epi_watsup
mzx = lm(wdi_accelectr ~ epi_watsup, data = data)
coef(mzx)
```

```
## (Intercept)  epi_watsup
##  39.7519829   0.6850051
```

```r
# beta_yx1z is the coefficient of X from regression Y on
# X+Z
beta_yx1z = coef(ols_2)["epi_watsup"]
beta_yx1z
```

(ii)

```
## epi_watsup
## -0.3496213
```

```r
# beta_yx1 is the coefficient of X from regression Y on X
beta_yx1 = coef(ols)["data$epi_watsup"]

# beta_yz1x is the coefficient of Z from regression Y on
# X+Z
beta_yz1x = coef(ols_2)["wdi_accelectr"]

# beta_zx1 is the coefficient of X from regression Z on X
beta_zx1 = coef(mzx)["epi_watsup"]

# calculated beta_yx1z from OVB theorem
beta_yx1z_Calc = beta_yx1 - beta_yz1x * beta_zx1
beta_yx1z_Calc
```

```
## data$epi_watsup
##      -0.3496213
```

```r
# residual from Z on X
res_zx = resid(mzx)
# residual from Y on X
res_yx = resid(ols)
# correlation between residual of Y on X and Z on X
cor2yx.zx = cor(res_yx, res_zx)^2

# correlation between z on x
cor2_zx = cor(data$wdi_accelectr, data$epi_watsup)^2

# standard deviation of residuals from Y on X
sd_yx = sd(res_yx)

# standard deviation of X
sd_x = sd(data$epi_watsup)

# calculated beta_yx1z from partial R^2 formulation
beta.yx1z_Calc2 = beta_yx1 + sqrt((cor2yx.zx * cor2_zx)/(1 -
    cor2_zx)) * sd_yx/sd_x
beta.yx1z_Calc2
```

(iii)

```
## data$epi_watsup
##      -0.3496213
```

2.2 Quadractic model

2.2(a) OLS with by regressing Y on X and X^2

$Infant\_mortalityrate = 71.7508 - 1.2555 * Access\_to\_cleanwater + 0.0059 * (Access\_to\_cleanwater)^2$

Estimated coefficients of this model are:
Intercept = 71.7508
Access to $cleanwater$ = -1.2555
Access to $cleanwater^2$ = 0.0059

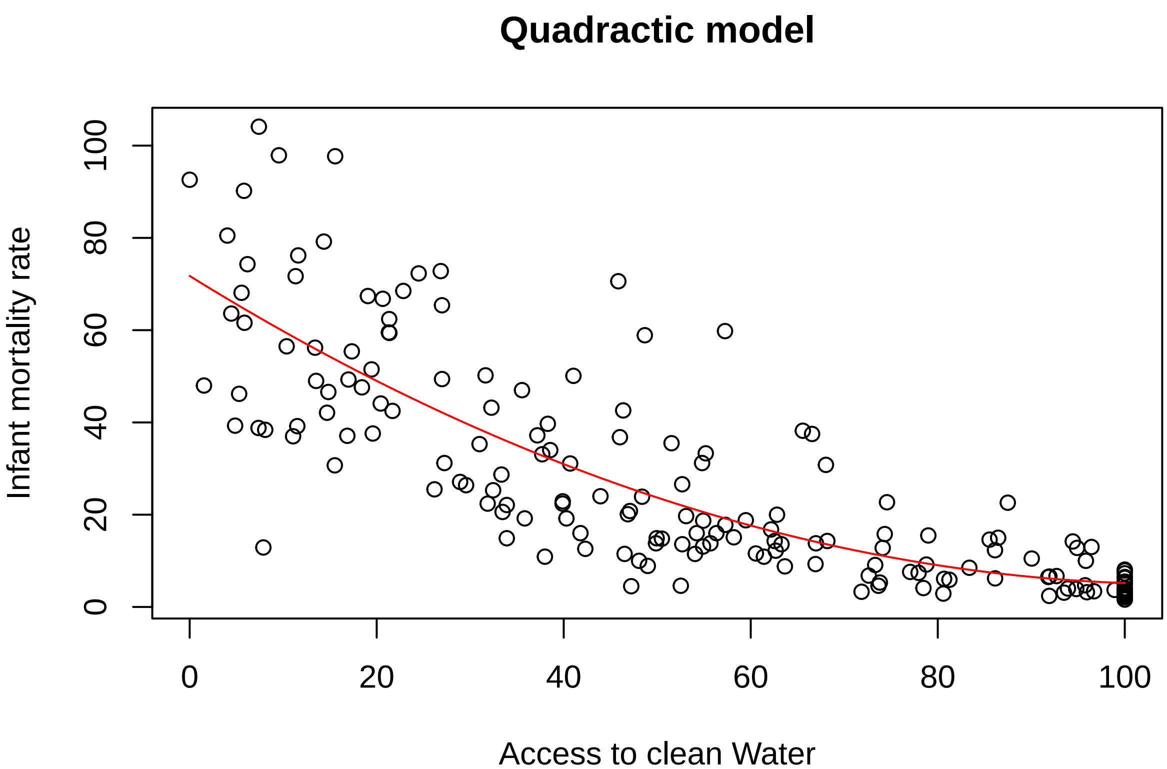
The coefficient of X(Access to clean water) $\hat{\beta}_{yx.1.x^2}$ cannot be directly interpreted as impact of change in X_i on Y_i

The change of Y impacted by change in X $= -1.255 + 2 * 0.0059(Access\_to\_cleanwater)$.

```r
# Quadractic plot with X and X^2
data$epi_watsup_2 = data$epi_watsup^2
quad.ols = lm(wdi_mortinftot ~ epi_watsup + epi_watsup_2, data = data)
quad.ols
```

```
##
## Call:
## lm(formula = wdi_mortinftot ~ epi_watsup + epi_watsup_2, data = data)
##
## Coefficients:
##  (Intercept)    epi_watsup  epi_watsup_2
##    71.750805     -1.255507      0.005898
```

```r
# sequence of access to clean water values
cleanWater.seq = seq(0, 100, 0.5)
predictedModel = predict(quad.ols, list(epi_watsup = cleanWater.seq,
    epi_watsup_2 = cleanWater.seq^2))
plot(data$epi_watsup, data$wdi_mortinftot, xlab = "Access to clean Water",
    ylab = "Infant mortality rate", main = "Quadractic model")
lines(cleanWater.seq, predictedModel, col = "red")
```



2.2(b) Average partial derivative of X on Y $E[Y_i|X_i] = \beta_{y1.xx^2} + \beta_{yx.1x^2}X_i + \beta_{yx^2.1x}X_i^2$

$\frac{\partial E[Y_i|X_i]}{\partial X_i} = \beta_{yx.1x^2} + 2\beta_{yx^2.1x}X_i$

$APD_{yx} = E[\frac{\partial E[Y_i|X_i]}{\partial X_i}] = \beta_{yx.1x^2} + 2\beta_{yx^2.1x}E[X_i]$

2.2(c) Estimated Average partial derivative of X on Y $A\hat{P}D_{yx} = -1.2555 + 2 * 0.0059 * E_n[X_i] = -0.6033$

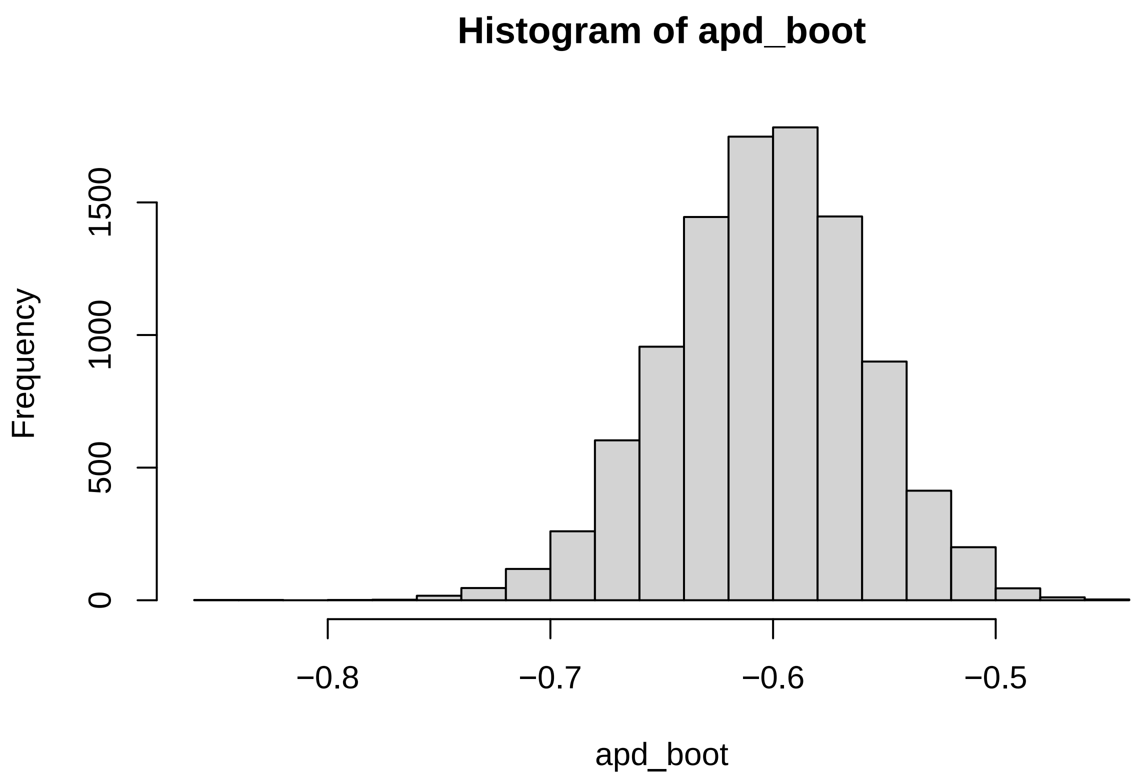
The 95% confidence interval for estimated average partial derivation of X on Y is (-0.68, -0.53)

APD for simple linear regression is -0.601. APD for quadratic model is -0.603. Hence APD of both models are similar.

```r
# empirical mean = En[Xi]
emp.mean = mean(data$epi_watsup)
# Estimated average partial derivative of X on Y
est.apd = coef(quad.ols)["epi_watsup"] + 2 * coef(quad.ols)["epi_watsup_2"] *
    emp.mean
est.apd
```

```
## epi_watsup
## -0.6032968
```

```r
set.seed(123)
# bootstrap empirical distribution
B = 10000
n = 184
apd_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = data[idx_boot, ]
    ols_boot = lm(wdi_mortinftot ~ epi_watsup + epi_watsup_2,
        data = data_boot)
    betayx.1x_2 = coef(ols_boot)["epi_watsup"]
    betayx_2.1x = coef(ols_boot)["epi_watsup_2"]
    empr.mean = mean(data_boot$epi_watsup)
    apd_boot[i] = betayx.1x_2 + 2 * betayx_2.1x * empr.mean
}
hist(apd_boot)
```




```
# 95% confidence interval
se_boot <- sd(apd_boot)
z_crit <- qnorm(0.95)
low <- est.apd - z_crit * se_boot
up <- est.apd + z_crit * se_boot
c(low, up)
```

```
## epi_watsup epi_watsup
## -0.6758397 -0.5307539
```

2.2(d) Estimated Average partial derivative of X on Y using numerical approximation The estimated APD of X on Y using numerical approximation = -0.6033 = estimated APD from plug in estimator.

```
# plug-in estimate of APD using numerical derivative
h = 1e-04
# E[Y|X+h]
data$epi_watsup_p = data$epi_watsup + h
yp = predict(quad.ols, list(epi_watsup = data$epi_watsup_p, epi_watsup_2 = data$epi_watsup_p^2))

# E[Y|X-h]
data$epi_watsup_n = data$epi_watsup - h
yn = predict(quad.ols, list(epi_watsup = data$epi_watsup_n, epi_watsup_2 = data$epi_watsup_n^2))

# dx = E[Y|X+h] - E[Y|X-h]/2h
dx = (yp - yn)/(2 * h)

# est.APD_2 = mean(dx)
est.apd2 = mean(dx)
est.apd2
```

```
## [1] -0.6032968
```

2.2(e) Quadractic ols with both access to clean water and electricity

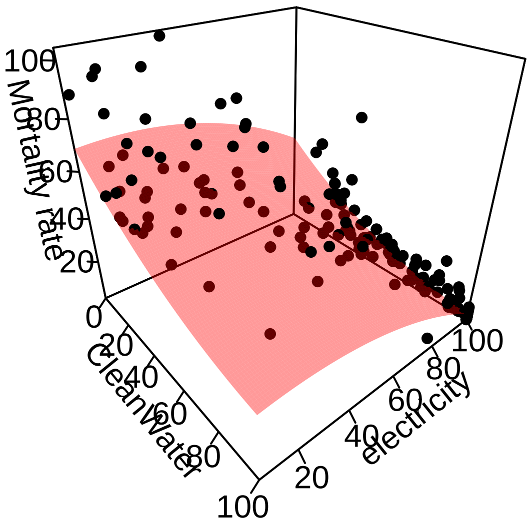
$Y_i = 68.1133 - 0.7064 * X_i + 0.0031 * X_i^2 + 0.1570 * Z_i - 0.0039 * Z_i^2$
The coefficients are: Intercept: 68.113336
$X_i = epi\_watsup : -0.706417$
$X_i^2 = epi\_watsup^2 : 0.003103$
$Z_i = wdi\_accelectr : 0.157026$
$Z_i^2 = wdi\_accelectr^2 : -0.003961$

```
# Quadractic plot with X, X^2, Z and Z^2
data$wdi_accelectr_2 = data$wdi_accelectr^2
quad.ols2 = lm(wdi_mortinftot ~ epi_watsup + I(epi_watsup^2) +
    wdi_accelectr + I(wdi_accelectr^2), data = data)
quad.ols2
```

```
##
## Call:
## lm(formula = wdi_mortinftot ~ epi_watsup + I(epi_watsup^2) +
##     wdi_accelectr + I(wdi_accelectr^2), data = data)
##
## Coefficients:
##        (Intercept)          epi_watsup     I(epi_watsup^2)       wdi_accelectr
##          68.113336           -0.706417            0.003103            0.157026
## I(wdi_accelectr^2)
##          -0.003961
```

```
#Scatterplot
grid.size = 100
# x grid
x.grid    = seq(min(data$epi_watsup), max(data$epi_watsup), length.out = grid.size)

# z grid
z.grid    = seq(min(data$wdi_accelectr), max(data$wdi_accelectr), length.out = grid.size)

# expand all combinations
grid.data = expand.grid( epi_watsup = x.grid, wdi_accelectr = z.grid)

y.hat.grid   = matrix(predict(quad.ols2, newdata = grid.data),
                       nrow = grid.size,
                      ncol = grid.size)

# scatter plot with regression surface
scatter3D(x = data$epi_watsup, y = data$wdi_accelectr,z = data$wdi_mortinftot,  pch = 20, col="black",
          ylab = "electricity", xlab = "CleanWater", zlab = "Mortality rate",  ticktype = "detailed",
          surf = list(x = x.grid, y = z.grid, z = y.hat.grid, col = "red", alpha = 0.4),
          phi = 30, theta = 50, # control the angle of the plot
          main = "Regression Surface")
```

Regression Surface



2.2(f) Average partial derivative of X on Y adjusting for Z

The estimated APD using numerical approximation is -0.3632. The regession coefficient $\hat{\beta}_{yx.1z} = -0.3496$. Hence estimated APD and $\hat{\beta}_{yx.1z}$ are similar. The estimated APD using plug in principle is -0.3632.

```
# plug-in estimate of APD using numerical approximation
h = 1e-04
# E[Y|X+h]
data$epi_watsup_p = data$epi_watsup + h
yp = predict(quad.ols2, list(epi_watsup = data$epi_watsup_p,
    wdi_accelectr = data$wdi_accelectr))

# E[Y|X-h]
data$epi_watsup_n = data$epi_watsup - h
yn = predict(quad.ols2, list(epi_watsup = data$epi_watsup_n,
    wdi_accelectr = data$wdi_accelectr))

# dx = E[Y|X+h] - E[Y|X-h]/2h
dx = (yp - yn)/(2 * h)

# est.APD_3 = mean(dx)
est.apd3 = mean(dx)
est.apd3
```

```
## [1] -0.3632335
```

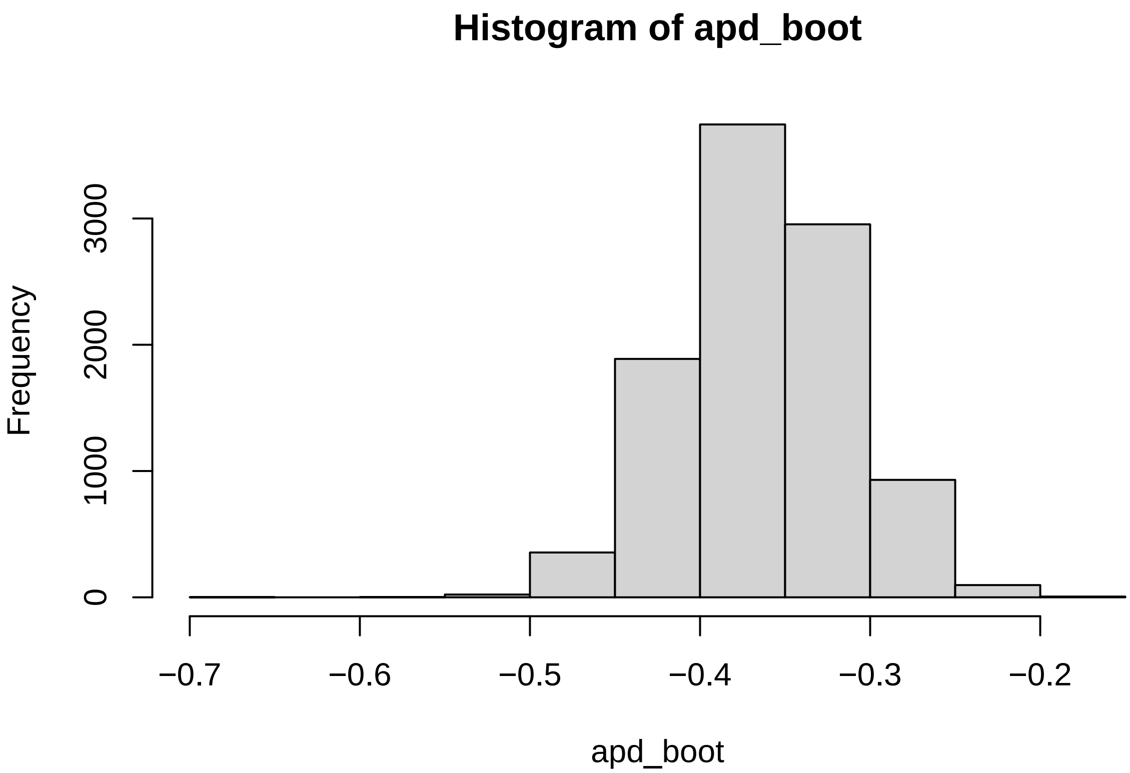
```
# APD estimate using plug in estimator
est.apd4 = coef(quad.ols2)["epi_watsup"] + 2 * coef(quad.ols2)["I(epi_watsup^2)"] *
    mean(data$epi_watsup)
est.apd4
```

```
## epi_watsup
## -0.3632335
```

2.2(g) 95% confidence interval for APD__yx.Z

The 95% confidence interval for estimated APD =(-0.445, -0.282).

```
# bootstrap empirical distribution
set.seed(123)
B = 10000
n = 184
apd_boot = rep(NA, B)
for (i in 1:B) {
    idx_boot = sample(1:n, size = n, replace = T)
    data_boot = data[idx_boot, ]
    ols_boot = lm(wdi_mortinftot ~ epi_watsup + I(epi_watsup^2) +
        wdi_accelectr + I(wdi_accelectr^2), data = data_boot)
    apd_boot[i] = coef(ols_boot)["epi_watsup"] + 2 * coef(ols_boot)["I(epi_watsup^2)"] *
        mean(data_boot$epi_watsup)
}
hist(apd_boot)
```




```
# 95% confidence interval
se_boot <- sd(apd_boot)
z_crit <- qnorm(0.95)
low <- est.apd4 - z_crit * se_boot
up <- est.apd4 + z_crit * se_boot
c(low, up)
```

```
## epi_watsup epi_watsup
## -0.4445949 -0.2818720
```

2.3 Causal interpretation Yes, we can intrepret the estimates from the regression as causal effect of access to clean water in infant mortality, only when conditioned on access to electricity.