1. (a)

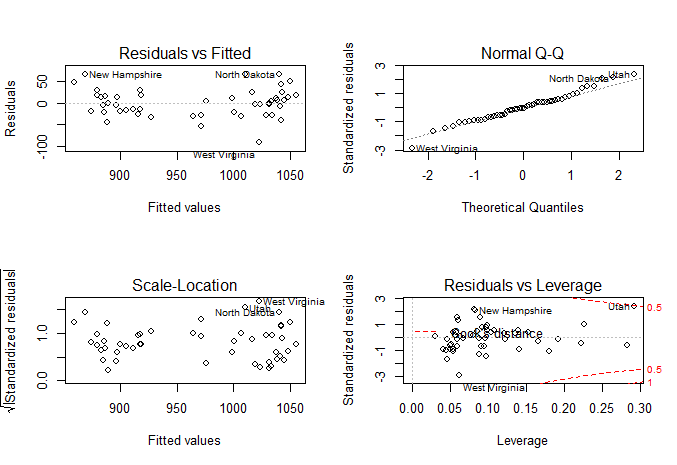
> library(faraway)

> data(sat)

> model\_1<-lm(total~expend + salary + ratio + takers, data = sat)

> par(mfrow=c(2,2))

> plot(model\_1, add.smooth=FALSE)



(b)

Since there is no curvature in the plot of residuals vs. fitted, there is no reason to doubt the linearity between the mean response and the predictors.

(c)

Since the trend is roughly flat, the assumption of constant variance still holds.

(d)

> residuals(model\_1)

Alabama Alaska Arizona Arkansas California

-2.067916 -30.212462 -27.625139 -28.860484 -15.847970

Colorado Connecticut Delaware Florida Georgia

4.029938 28.164072 13.320106 -27.227193 -20.169552

Hawaii Idaho Illinois Indiana Iowa

-16.755330 -21.751100 10.518739 -18.337458 48.732925

Kansas Kentucky Louisiana Maine Maryland

12.178680 -29.525950 -2.531446 16.494501 11.596320

Massachusetts Michigan Minnesota Mississippi Missouri

47.687788 -8.007105 42.937442 2.937404 6.237246

Montana Nebraska Nevada New Hampshire New Jersey

10.549009 5.586986 -54.149446 65.865300 -13.747579

New Mexico New York North Carolina North Dakota Ohio

-3.820086 -4.879704 -21.083117 66.571288 -31.783361

Oklahoma Oregon Pennsylvania Rhode Island South Carolina

-4.420282 29.408040 -5.311804 -1.423530 -45.075815

South Dakota Tennessee Texas Utah Vermont

24.854905 23.516451 -34.004860 65.766079 14.432938

Virginia Washington West Virginia Wisconsin Wyoming

12.293358 18.084596 -90.531418 18.043483 -40.657489

> min(residuals(model\_1))

[1] -90.53142

> which.min(residuals(model\_1))

West Virginia

48

> max(residuals(model\_1))

[1] 66.57129

> which.max(residuals(model\_1))

North Dakota

34

(e)

> hatvalues(model\_1)

Alabama Alaska Arizona Arkansas California

0.09537668 0.18030612 0.04931612 0.05382878 0.28211791

Colorado Connecticut Delaware Florida Georgia

0.03014533 0.22545191 0.05823786 0.14068586 0.09418039

Hawaii Idaho Illinois Indiana Iowa

0.05164378 0.05847998 0.13483540 0.05031937 0.05882595

Kansas Kentucky Louisiana Maine Maryland

0.10383439 0.04106460 0.08754545 0.13973600 0.05372121

Massachusetts Michigan Minnesota Mississippi Missouri

0.08871286 0.16536832 0.06069775 0.06245712 0.05589831

Montana Nebraska Nevada New Hampshire New Jersey

0.12118024 0.07953981 0.04539029 0.08283540 0.22209778

New Mexico New York North Carolina North Dakota Ohio

0.04533550 0.19157520 0.09063656 0.08104629 0.04477728

Oklahoma Oregon Pennsylvania Rhode Island South Carolina

0.05949338 0.09828873 0.12277229 0.06739042 0.09671590

South Dakota Tennessee Texas Utah Vermont

0.09658262 0.09162128 0.05405622 0.29211280 0.08566501

Virginia Washington West Virginia Wisconsin Wyoming

0.15230581 0.10816994 0.06206536 0.09807708 0.08748135

> which.max(hatvalues(model\_1))

Utah

44

(f)

Since there is no observation in either of four default diagnostic plots that is significantly apart from other observations, there is no outlier.

(g)

> cooks.distance(model\_1)

Alabama Alaska Arizona Arkansas California

9.320823e-05 4.581008e-02 7.787666e-03 9.366213e-03 2.571304e-02

Colorado Connecticut Delaware Florida Georgia

9.733862e-05 5.574806e-02 2.178814e-03 2.641399e-02 8.732755e-03

Hawaii Idaho Illinois Indiana Iowa

3.014830e-03 5.837035e-03 3.727482e-03 3.508650e-03 2.949550e-02

Kansas Kentucky Louisiana Maine Maryland

3.586302e-03 7.280787e-03 1.260173e-04 9.607378e-03 1.508793e-03

Massachusetts Michigan Minnesota Mississippi Missouri

4.543308e-02 2.846396e-03 2.372007e-02 1.146592e-04 4.562770e-04

Montana Nebraska Nevada New Hampshire New Jersey

3.265407e-03 5.480389e-04 2.731373e-02 7.989442e-02 1.297264e-02

New Mexico New York North Carolina North Dakota Ohio

1.357578e-04 1.305355e-03 9.111288e-03 7.954292e-02 9.271069e-03

Oklahoma Oregon Pennsylvania Rhode Island South Carolina

2.457687e-04 1.955161e-02 8.418630e-04 2.936388e-05 4.504204e-02

South Dakota Tennessee Texas Utah Vermont

1.367187e-02 1.148386e-02 1.306408e-02 4.715287e-01 3.991940e-03

Virginia Washington West Virginia Wisconsin Wyoming

5.990470e-03 8.318437e-03 1.081395e-01 7.340931e-03 3.247841e-02

> which(cooks.distance(model\_1) >= 1)

named integer(0)

Therefore, there is no influential point.

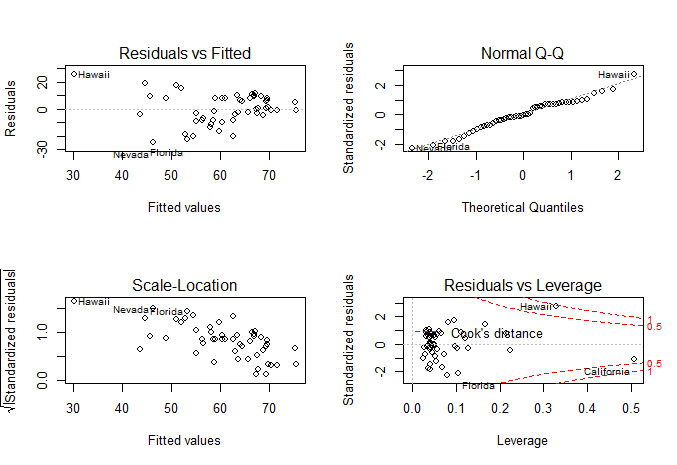
2.

(a)

> model\_2<-lm(home~usborn + income + pop, data = eco)

> par(mfrow=c(2,2))

> plot(model\_2, add.smooth=FALSE)



(b)

Since there is no curvature in the residuals vs. fitted values plot, there is no reason to doubt the linearity between

the mean response and predictors

(c)

Since the variance of residuals appears to be smaller as the fitted value gets larger, and the standardized residuals tend to decrease as the fitted value gets larger, the assumption of constant variance does not hold.

(d)

> residuals(model\_2)

Alabama Alaska Arizona Arkansas California

6.3595383 -20.4728416 -18.5626403 -0.1842986 -8.7480585

Colorado Connecticut Delaware DC Florida

-16.4932483 8.0112259 -8.2442147 -4.2766054 -22.6938775

Georgia Hawaii Idaho Illinois Indiana

-4.3006323 25.9267293 -9.8056785 8.0388930 1.2818635

Iowa Kansas Kentucky Louisiana Maine

11.1680171 -2.2591609 7.9470558 11.8838697 8.1262706

Maryland Massachusetts Michigan Minnesota Mississippi

-6.7973580 17.7351714 5.5907064 7.5394337 8.9986683

Missouri Montana Nebraska Nevada New.Hampshire

0.1793212 -4.1579433 6.2288712 -24.6034576 -13.7993498

New.Jersey New.Mexico New.York North.Carolina North.Dakota

9.0412869 -3.3997213 15.4758519 -1.0844607 9.7895416

Ohio Oklahoma Oregon Pennsylvania Rhode.Island

-1.2823468 -2.2426020 -11.4646312 4.9673351 18.6108160

South.Carolina South.Dakota Tennessee Texas Utah

0.5921193 5.6584232 -1.1673620 -2.9029093 8.1811741

Vermont Virginia Washington West.Virginia Wisconsin

-1.6289330 -8.3893877 -8.1644437 10.3142194 9.4096842

Wyoming

-19.9299245

> min(residuals(model\_2))

[1] -24.60346

> which.min(residuals(model\_2))

Nevada

29

> max(residuals(model\_2))

[1] 25.92673

> which.max(residuals(model\_2))

Hawaii

12

(e)

> hatvalues(model\_2)

Alabama Alaska Arizona Arkansas California

0.04317415 0.04057164 0.06711280 0.04891862 0.50786366

Colorado Connecticut Delaware DC Florida

0.04168408 0.21368240 0.07289825 0.22335255 0.10406213

Georgia Hawaii Idaho Illinois Indiana

0.03958591 0.32992973 0.05171509 0.05631998 0.04159111

Iowa Kansas Kentucky Louisiana Maine

0.03183286 0.02728528 0.04293110 0.03873095 0.03377196

Maryland Massachusetts Michigan Minnesota Mississippi

0.04547808 0.08247962 0.05306063 0.05033451 0.06559250

Missouri Montana Nebraska Nevada New.Hampshire

0.04074049 0.05689361 0.03198049 0.07897435 0.05541099

New.Jersey New.Mexico New.York North.Carolina North.Dakota

0.11257832 0.10202271 0.16745291 0.05056167 0.04177837

Ohio Oklahoma Oregon Pennsylvania Rhode.Island

0.09871212 0.04034840 0.02526177 0.12009472 0.09587355

South.Carolina South.Dakota Tennessee Texas Utah

0.04018629 0.03925663 0.04315216 0.12804281 0.05680171

Vermont Virginia Washington West.Virginia Wisconsin

0.03359108 0.02978440 0.02812210 0.06097812 0.03220404

Wyoming

0.03523661

> which.max(hatvalues(model\_2))

California

5

(f)

From the plot of residuals vs. fitted and the plot of scale-location, Hawaii seems like an outlier because it is apart

from other dataset

(g)

> cooks.distance(model\_2)

Alabama Alaska Arizona Arkansas California

3.595360e-03 3.482466e-02 5.009101e-02 3.462708e-06 3.025052e-01

Colorado Connecticut Delaware DC Florida

2.327546e-02 4.181229e-02 1.086661e-02 1.276658e-02 1.258600e-01

Georgia Hawaii Idaho Illinois Indiana

1.496309e-03 9.311281e-01 1.042385e-02 7.704397e-03 1.402541e-04

Iowa Kansas Kentucky Louisiana Maine

7.984732e-03 2.774502e-04 5.579963e-03 1.115879e-02 4.503112e-03

Maryland Massachusetts Michigan Minnesota Mississippi

4.347542e-03 5.809238e-02 3.486540e-03 5.980468e-03 1.146757e-02

Missouri Montana Nebraska Nevada New.Hampshire

2.683801e-06 2.084647e-03 2.496140e-03 1.062351e-01 2.229259e-02

New.Jersey New.Mexico New.York North.Carolina North.Dakota

2.202868e-02 2.756678e-03 1.090730e-01 1.243508e-04 8.220107e-03

Ohio Oklahoma Oregon Pennsylvania Rhode.Island

3.766931e-04 4.153715e-04 6.587811e-03 7.214931e-03 7.657809e-02

South.Carolina South.Dakota Tennessee Texas Utah

2.883074e-05 2.566979e-03 1.210769e-04 2.675254e-03 8.056008e-03

Vermont Virginia Washington West.Virginia Wisconsin

1.799044e-04 4.198035e-03 3.741199e-03 1.386847e-02 5.738865e-03

Wyoming

2.834634e-02

> which(cooks.distance(model\_2) >= 1)

named integer(0)

Even though when we set up the criteria to be Cook's Distance is greater than one and we did not get any influential point, but the Cook's Distance associated with Hawaii is 0.9311281, which is very close to one. Therefore, it is reasonable to conclude that Hawaii is an influential point

3.

(a)

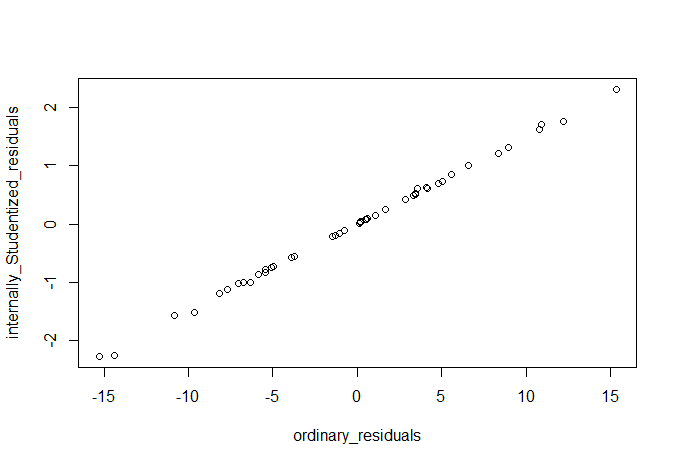
> data(swiss)

> model\_3<-lm(Fertility~Agriculture + Examination + Education + Catholic + Infant.Mortality , data = swiss)

> internally\_Studentized\_residuals <-rstandard(model\_3)

> ordinary\_residuals<-residuals(model\_3)

> plot(ordinary\_residuals,internally\_Studentized\_residuals)



(b)

Because the denominator of the internally Studentized residuals is , if we consider the formula as r = k , then each observation has slightly different value of coefficient(k). Therefore, these observations on the plot

do not exactly fall on a straight line

(c)

> externally\_Studentized\_residuals <-rstudent(model\_3)

> externally\_Studentized\_residuals

Courtelary Delemont Franches-Mnt Moutier Neuveville Porrentruy

0.84584030 0.08684212 1.01074888 1.31796269 1.81614211 -2.36721827

Broye Glane Gruyere Sarine Veveyse Aigle

0.61438409 1.66113713 0.15298493 0.50049516 0.51646133 0.73437530

Aubonne Avenches Cossonay Echallens Grandson Lausanne

0.07248412 0.42104970 -0.54614038 -0.74687721 0.01511285 0.02925763

La Vallee Lavaux Morges Moudon Nyone Orbe

0.61245225 0.24191247 0.47865752 -1.60418331 -0.71558271 -0.99565881

Oron Payerne Paysd'enhaut Rolle Vevey Yverdon

-0.15431971 0.24875019 0.07942498 -0.21167094 -0.77601876 -1.02128559

Conthey Entremont Herens Martigwy Monthey St Maurice

-0.11516331 -1.12779737 -0.18491467 -0.85417952 -0.56297196 -1.20389559

Sierre Sion Boudry La Chauxdfnd Le Locle Neuchatel

2.44522679 1.22241299 0.69339739 -1.00648626 0.60322648 1.74541155

Val de Ruz ValdeTravers V. De Geneve Rive Droite Rive Gauche

0.73018972 -0.83329565 0.03781782 -1.54311369 -2.39447054

(d)

> critical\_value\_0 <- qt(0.05/2, df=df.residual(model\_3)-1, lower=FALSE)

> critical\_value\_0

[1] 2.021075

> which(abs(externally\_Studentized\_residuals) > critical\_value\_0)

Porrentruy Sierre Rive Gauche

6 37 47

(e)

> critical\_value\_1 <- qt(0.05/(2\*nobs(model\_3)), df=df.residual(model\_3)-1, lower=FALSE)

> critical\_value\_1

[1] 3.529468

> which(abs(externally\_Studentized\_residuals) > critical\_value\_1)

named integer(0)

4.

(a)

> par(pty="s")

> par(mfrow=c(3,3))

> for(i in 1:9){

+ qqnorm(rgeom(50,0.4))

+ }

Plot see next page

(b)

(1) The overall trend is not a linear line, but a concaved up curvature.

(2) The spread of points is discrete rather than continuous

(3) The data is not evenly spread across the plot