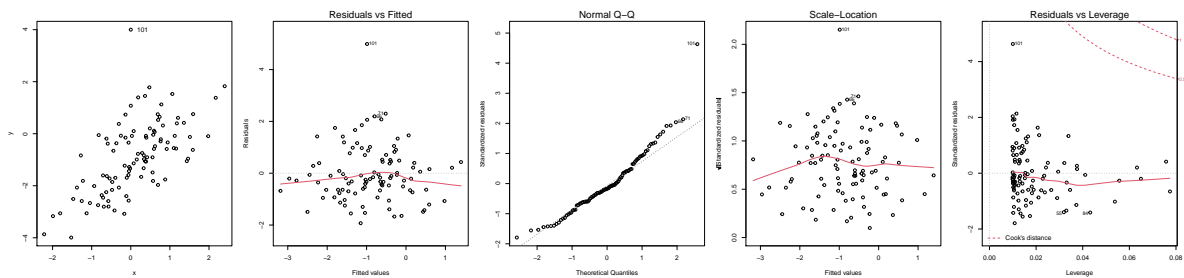# The University of British Columbia
Data Science 570 Predictive Modelling
Three types of points

1. Outlier

```
set.seed(1)
x=rnorm(100)
e=rnorm(100)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,0)
y=c(y,4)#outlier
fit2=lm(y~x)
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]+0.3,y[101], "101")
plot(fit2)
```



```
#summary(fit2)
```

2. Outlier and influential point

```
set.seed(1)
x=rnorm(100)
e=rnorm(100)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,-2)
y=c(y,50)#outlier and influential point
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]+0.3,y[101], "101")
fit2=lm(y~x)
plot(fit2)
```
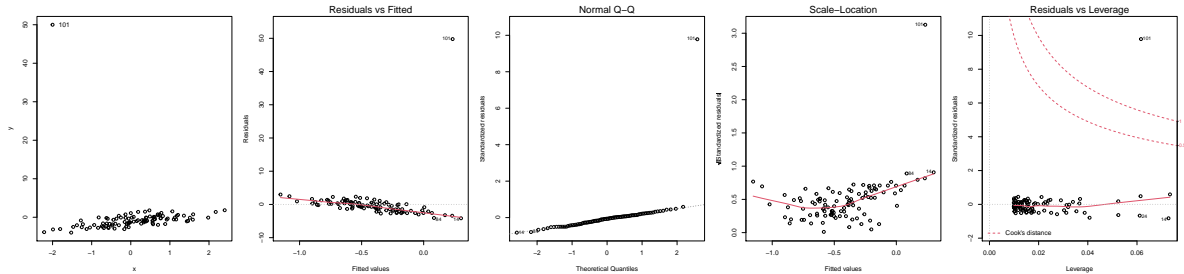


```
#summary(fit2)
cooks.distance(fit2)
```

```
##            1            2            3            4            5            6
## 1.252317e-03 1.889337e-05 2.580950e-03 1.968290e-03 1.333061e-04 2.698618e-06
##            7            8            9           10           11           12
## 1.247070e-04 4.553409e-04 6.859262e-05 1.000112e-04 3.704282e-04 6.202731e-05
##           13           14           15           16           17           18
## 4.439617e-08 2.665886e-02 1.924767e-04 2.082011e-04 1.654674e-04 4.512468e-05
##           19           20           21           22           23           24
## 2.876573e-04 1.039379e-10 3.371474e-06 9.140262e-04 9.483638e-05 1.451897e-02
##           25           26           27           28           29           30
## 3.090508e-06 2.405553e-07 1.526460e-04 5.038627e-03 9.352508e-04 2.960635e-05
##           31           32           33           34           35           36
## 8.647710e-04 3.368814e-04 3.904218e-05 9.016607e-04 3.060965e-03 1.731254e-03
##           37           38           39           40           41           42
## 4.765330e-04 2.749169e-04 1.533236e-05 3.381889e-05 1.475379e-03 1.208705e-05
##           43           44           45           46           47           48
## 4.874045e-04 2.602137e-05 2.205599e-03 1.705981e-03 7.721658e-04 5.201953e-05
##           49           50           51           52           53           54
## 7.972182e-04 3.837782e-04 2.812559e-05 5.984786e-04 4.409145e-05 4.798151e-03
##           55           56           57           58           59           60
## 2.570567e-05 9.075049e-04 1.697771e-06 3.231850e-03 3.608455e-04 2.302249e-04
##           61           62           63           64           65           66
```

```
## 1.377212e-02 1.490203e-04 4.917170e-04 1.885602e-05 1.647142e-03 6.354078e-04
##           67           68           69           70           71           72
## 1.105536e-02 6.301551e-06 5.485625e-05 7.666813e-03 1.177524e-03 6.712528e-04
##           73           74           75           76           77           78
## 1.056523e-04 1.540645e-03 4.037810e-03 1.246113e-05 3.916109e-05 4.013288e-04
##           79           80           81           82           83           84
## 4.992687e-05 8.336974e-06 1.871343e-03 8.011014e-06 6.212583e-04 1.331531e-02
##           85           86           87           88           89           90
## 1.176624e-04 2.062779e-05 2.046775e-03 6.813147e-04 6.005081e-05 2.640341e-04
##           91           92           93           94           95           96
## 6.089614e-04 9.094228e-04 1.637773e-05 3.530008e-04 5.428422e-05 1.956834e-04
##           97           98           99          100          101
## 4.376329e-04 1.587900e-03 1.899253e-03 6.562256e-04 3.143158e+00
```

3. Outlier and high leverage point

```
set.seed(1)
x=rnorm(100)
e=rnorm(100)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,3)
y=c(y,4)#outlier and high leverage point
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]-0.3,y[101], "101")
fit2=lm(y~x)
plot(fit2)
```
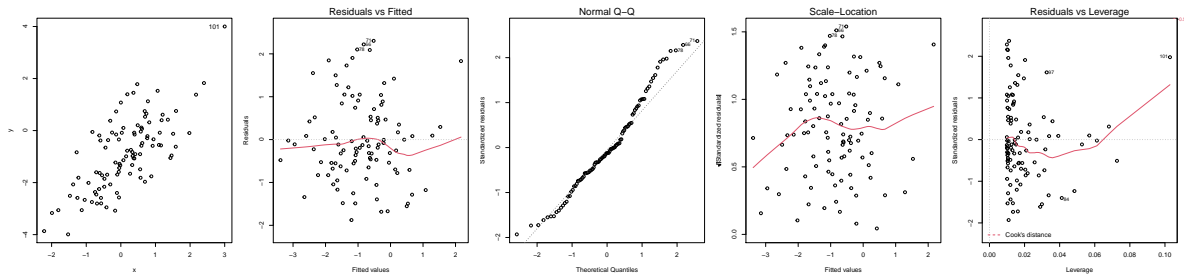


```
#summary(fit2)
```

4. Outlier, high levarage, and influential point

```
set.seed(1)
x=rnorm(100)
e=rnorm(100)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,3)#outlier, high levarage point, and influential point
y=c(y,6)
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]-0.3,y[101], "101")
fit2=lm(y~x)
plot(fit2)
```
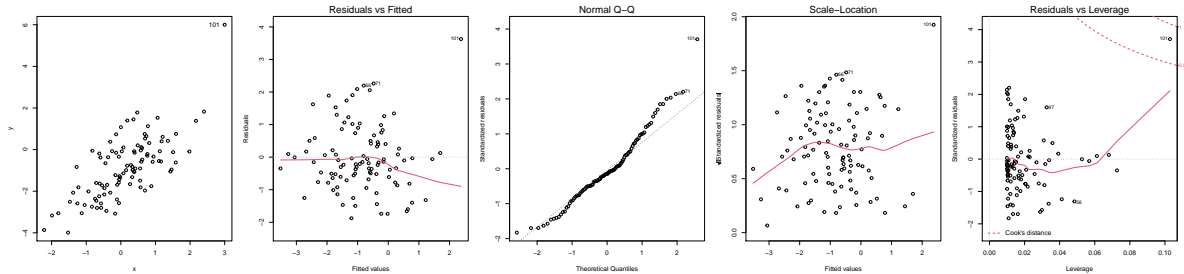


```
#summary(fit2)
cooks.distance(fit2)
```

```
##            1            2            3            4            5            6
## 2.193411e-03 5.473491e-06 6.246755e-03 1.964689e-05 2.297821e-03 3.546845e-02
##            7            8            9           10           11           12
## 2.425573e-03 4.651737e-03 6.133712e-04 1.760651e-02 1.045208e-02 1.256236e-03
##           13           14           15           16           17           18
## 1.863161e-02 4.766415e-03 1.172980e-03 6.785951e-04 4.446676e-04 1.252374e-03
##           19           20           21           22           23           24
## 1.194944e-03 3.366857e-04 3.038531e-03 1.115973e-02 2.071044e-04 2.949230e-04
##           25           26           27           28           29           30
## 1.645855e-04 2.677833e-03 7.358389e-06 5.734538e-04 2.499164e-03 6.818148e-04
##           31           32           33           34           35           36
## 1.354579e-04 1.585552e-03 1.255266e-03 1.105632e-02 4.562417e-03 1.385889e-02
##           37           38           39           40           41           42
## 3.456310e-04 1.262630e-03 6.065793e-03 1.367630e-04 1.848634e-02 8.390719e-03
##           43           44           45           46           47           48
## 1.963539e-02 1.547328e-03 8.762767e-03 3.621820e-03 2.123189e-02 3.118554e-05
##           49           50           51           52           53           54
## 8.022460e-03 2.372892e-02 8.773232e-04 4.658954e-05 5.974694e-04 8.207304e-03
##           55           56           57           58           59           60
## 3.960551e-02 4.355022e-02 6.966749e-03 2.806804e-03 1.203037e-02 1.864885e-02
##           61           62           63           64           65           66
```

```
## 6.045470e-04 2.330663e-04 6.241637e-03 3.879952e-03 2.340561e-03 2.295259e-02
##           67           68           69           70           71           72
## 5.225578e-07 3.815863e-02 1.047153e-04 1.063668e-04 2.749244e-02 4.020461e-04
##           73           74           75           76           77           78
## 9.256788e-04 4.079579e-05 3.836305e-04 1.578840e-05 4.910287e-03 2.117005e-02
##           79           80           81           82           83           84
## 5.085489e-03 1.302579e-02 9.817796e-03 5.331129e-03 7.376399e-05 3.303263e-02
##           85           86           87           88           89           90
## 1.245210e-03 1.719671e-04 1.725034e-02 2.942700e-03 1.078636e-03 4.316984e-03
##           91           92           93           94           95           96
## 6.036342e-05 7.638659e-04 8.026379e-03 3.708305e-03 3.317645e-02 6.990006e-03
##           97           98           99          100          101
## 4.304348e-02 6.484570e-03 5.331140e-03 6.295700e-04 7.881521e-01
```

5. High levarage point

```
set.seed(1)
x=rnorm(100)
e=rnorm(100)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,3)
y=c(y,0)#high levarage point
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]-0.3,y[101], "101")
fit2=lm(y~x)
plot(fit2)
```
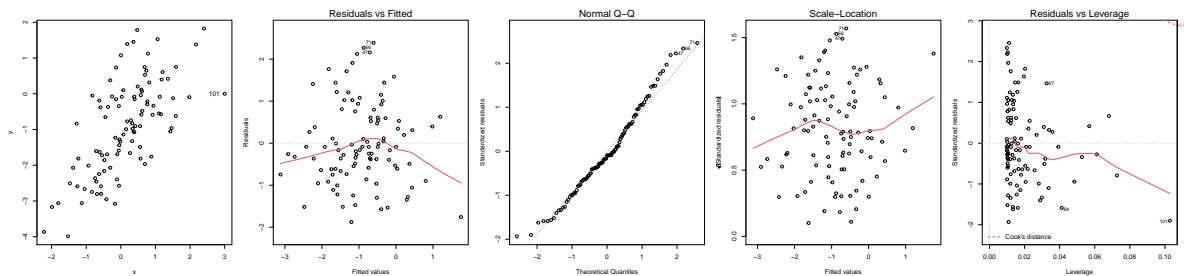


```
#summary(fit2)
```

6. High levarage and influential point

```
set.seed(1)
x=rnorm(100)
e=rnorm(100)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,4)
y=c(y,0)#high levarage point and influential point
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]-0.3,y[101], "101")
fit2=lm(y~x)
plot(fit2)
```
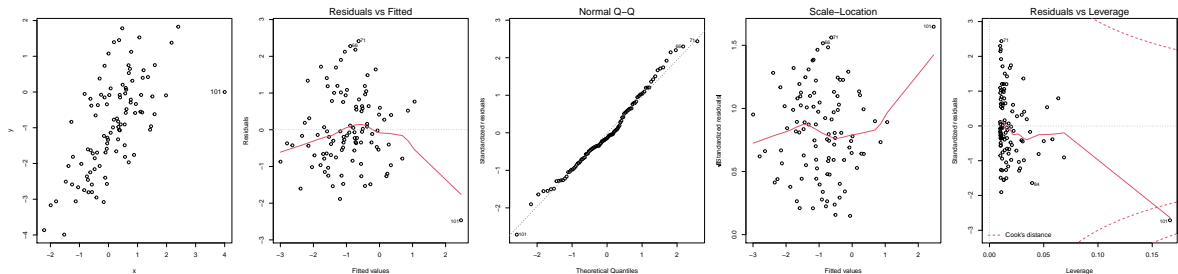


```
#summary(fit2)
cooks.distance(fit2)
```

```
##            1            2            3            4            5            6
## 3.537797e-03 6.589205e-05 9.784962e-03 2.765143e-03 1.687245e-03 3.061285e-02
##            7            8            9           10           11           12
## 3.904579e-03 7.747013e-03 1.548376e-03 1.789891e-02 2.571285e-03 7.226275e-04
##           13           14           15           16           17           18
## 1.662176e-02 3.024160e-02 4.972574e-06 6.454387e-04 3.898491e-04 1.153242e-04
##           19           20           21           22           23           24
## 3.155000e-03 1.950702e-05 1.005251e-03 1.623891e-02 1.251358e-04 4.550527e-03
##           25           26           27           28           29           30
## 3.710602e-06 3.036008e-03 1.044748e-05 5.669077e-04 3.468507e-03 2.768389e-04
##           31           32           33           34           35           36
## 1.012083e-03 1.658567e-03 2.133722e-03 1.158224e-02 6.643350e-04 1.620800e-02
##           37           38           39           40           41           42
## 6.035201e-04 1.260291e-03 2.251467e-03 5.192873e-05 2.002044e-02 8.551006e-03
##           43           44           45           46           47           48
## 1.588870e-02 7.245743e-04 1.193325e-02 5.703729e-03 2.546833e-02 1.850811e-04
##           49           50           51           52           53           54
## 8.553591e-03 1.786797e-02 1.632797e-03 1.572820e-05 2.743314e-04 1.492829e-02
##           55           56           57           58           59           60
## 2.329508e-02 1.549221e-02 6.631562e-03 6.437315e-03 9.754886e-03 2.003469e-02
##           61           62           63           64           65           66
```

```
## 2.138651e-02 1.992785e-04 9.555248e-03 4.562597e-03 4.164580e-03 2.649329e-02
##           67           68           69           70           71           72
## 5.045573e-03 2.161759e-02 2.985994e-05 8.088038e-03 3.319222e-02 4.396263e-05
##           73           74           75           76           77           78
## 2.117350e-03 2.335092e-04 3.131314e-03 1.295485e-05 4.318043e-03 2.353338e-02
##           79           80           81           82           83           84
## 6.039961e-03 1.143673e-02 1.246164e-02 5.724803e-03 1.885130e-03 5.522207e-02
##           85           86           87           88           89           90
## 2.558737e-03 2.525957e-05 2.649265e-02 3.537728e-03 6.065115e-04 3.689746e-03
##           91           92           93           94           95           96
## 2.906315e-04 4.096503e-03 3.199541e-03 6.322705e-03 1.595385e-02 5.229560e-03
##           97           98           99          100          101
## 2.999589e-02 8.587693e-03 1.559762e-03 1.097302e-03 7.343219e-01
```

7. Influential point

```
set.seed(1)
x=rnorm(100)
e=rnorm(100,0,0.1)
y=-1+x+e
fit1=lm(y~x)
#summary(fit1)
x=c(x,1.5)
y=c(y,-1)#influential point
par(mfrow=c(1,5), mai = c(1, .5, 0.2, 0.2))
plot(x,y)
text(x[101]-0.3,y[101], "101")
fit2=lm(y~x)
plot(fit2)
```
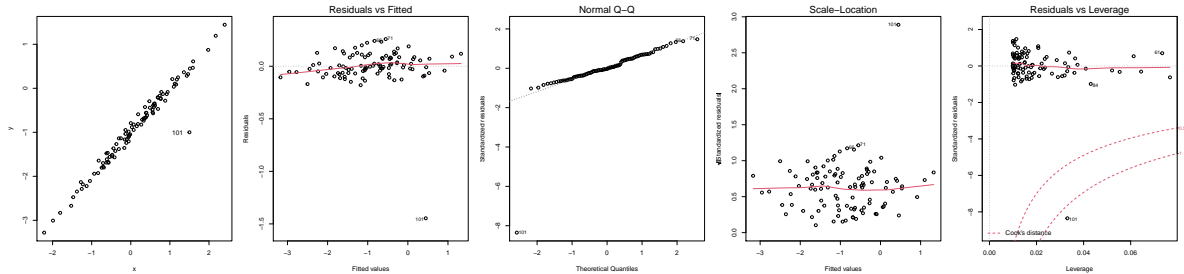


```
#summary(fit2)
cooks.distance(fit2)
```

```
##            1            2            3            4            5            6
## 1.087522e-03 9.692572e-05 3.324978e-03 3.244519e-03 2.972141e-04 1.026271e-02
##            7            8            9           10           11           12
## 1.881622e-03 3.779175e-03 9.613529e-04 6.207635e-03 5.584729e-05 7.681003e-05
##           13           14           15           16           17           18
## 5.678826e-03 1.616814e-02 2.014578e-04 1.046787e-04 4.770160e-05 3.982810e-05
##           19           20           21           22           23           24
## 1.938023e-03 3.407651e-05 4.140980e-05 7.270761e-03 2.716499e-06 3.306404e-03
##           25           26           27           28           29           30
## 9.562957e-05 1.232749e-03 3.140927e-06 4.657547e-04 9.958561e-04 7.309053e-06
##           31           32           33           34           35           36
## 1.523855e-03 3.663605e-04 1.090160e-03 3.206879e-03 8.312386e-05 4.914344e-03
##           37           38           39           40           41           42
## 1.316455e-04 2.556780e-04 1.739675e-04 2.098794e-04 5.827787e-03 3.080837e-03
##           43           44           45           46           47           48
## 4.106739e-03 5.671937e-05 3.864910e-03 1.839849e-03 9.521169e-03 3.350696e-04
##           49           50           51           52           53           54
## 2.341288e-03 4.501556e-03 8.882594e-04 9.401963e-07 1.018755e-05 5.583593e-03
##           55           56           57           58           59           60
## 4.986403e-03 1.641675e-03 2.393392e-03 2.408670e-03 2.401794e-03 6.983980e-03
##           61           62           63           64           65           66
```

```
## 1.942815e-02 1.474588e-05 4.414164e-03 1.808759e-03 1.358882e-03 9.493252e-03
##           67           68           69           70           71           72
## 3.096881e-03 4.437163e-03 3.957765e-06 9.322105e-03 1.256607e-02 2.020470e-05
##           73           74           75           76           77           78
## 1.244325e-03 1.000895e-04 1.403052e-03 6.339187e-05 1.578310e-03 8.245665e-03
##           79           80           81           82           83           84
## 2.354391e-03 3.947367e-03 3.899049e-03 2.151941e-03 1.799517e-03 2.182469e-02
##           85           86           87           88           89           90
## 1.423218e-03 1.110478e-05 1.241086e-02 9.481298e-04 5.785390e-05 8.267542e-04
##           91           92           93           94           95           96
## 6.468307e-05 3.059462e-03 3.123971e-04 3.130270e-03 2.659499e-03 1.140656e-03
##           97           98           99          100          101
## 9.465909e-03 2.663317e-03 3.646772e-04 2.826429e-04 1.193439e+00
```