STP 530
Lecture 10: Advanced Diagnostics

Yi Zheng, Ph.D.
yi.isabel.zheng@asu.edu

ARIZONA STATE UNIVERSITY
SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES

November 2, 2023

Regression diagnostics are used to determining whether a fitted regression model adequately represents the data.

Previous diagnostic items introduced:

1. The regression function is linear.
2. No excessive outliers.
3. The error terms have constant variance (homoskedasticity).
4. The error terms are normally distributed.
5. The error terms are independent among each other.
6. Multicollinearity.

Remedial measures

1. **Nonlinear relationship:** transform X, or add polynomial terms.
2. **Outliers:** remove outliers and discuss separately.
3. **Non-constant variance of the error term:** transform Y; Weighted least square estimation; Robust estimators; Bootstrap.
4. **Non-normal distribution of the error term:** transform Y; Bootstrap.
5. **Dependent cases:** use advanced models (e.g., time series modeling, hierarchical linear models for nested data); Bootstrap.
6. **Multicollinearity:** caused by interaction/polynomial terms — center predictors; otherwise — remove some predictor(s).

Outline

Recall: Multiple regression in matrix notation

The hypothesized regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X}_{n\times p} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

$$\boldsymbol{\beta}_{p\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \boldsymbol{\varepsilon}_{n\times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

The fitted regression model:

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$$

where **b** is an estimator of the $\boldsymbol{\beta}$ vector.

Recall: Matrix approach to least square estimation of multiple regression

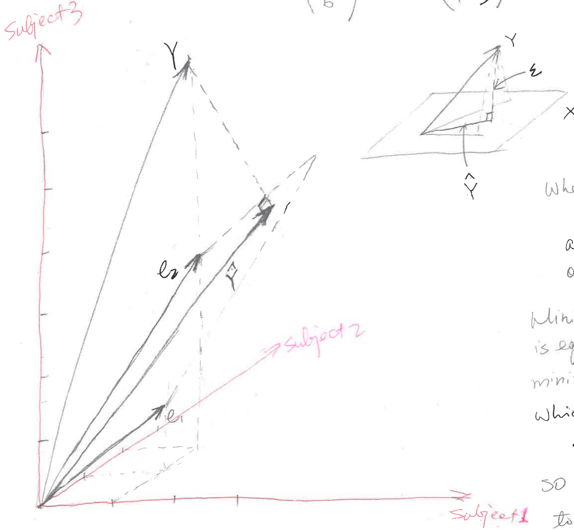
Least square estimation:

Solve for model parameters $\beta_0, \beta_1, \cdots, \beta_{p-1}$ that minimize SSE $= \sum_{i=1}^{n} \varepsilon_i^2$, which is equivalent with **minimizing** $\|\boldsymbol{\varepsilon}\| := \sqrt{\sum_{i=1}^{n} \varepsilon_i^2}$, **the length of the vector** $\boldsymbol{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$.

Also note that the fitted regression function $\hat{\mathbf{Y}} = \mathbf{Xb}$ implies that $\hat{\mathbf{Y}}$ lies in the space spanned by the columns of $\mathbf{X}$. Thus the length of $\boldsymbol{\varepsilon} = \mathbf{Y} - \hat{\mathbf{Y}}$ is minimized when $\hat{\mathbf{Y}}$ is the **projection** of $\mathbf{Y}$ onto the space spanned by the columns of $\mathbf{X}$, or equivalently, $\boldsymbol{\varepsilon}$ is orthogonal to the space spanned by the columns of $\mathbf{X}$:

$$\mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{0}$$

$$Y = \begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \quad \hat{Y} = X\beta = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$



$$= \begin{pmatrix} \beta_0 + \beta_1 \\ \beta_0 + 2\beta_1 \\ \beta_0 + 3\beta_1 \end{pmatrix}$$

$$= \beta_0 e_1 + \beta_1 e_2$$

Where $e_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $e_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

are the two basis vectors of the X plane.

Minimizing $SSE = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2$ is equivalent to minimizing $||\varepsilon|| = \sqrt{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2}$ which is length of

$$\varepsilon = Y - \hat{Y}$$

so ε must be orthogonal to the X plane

Recall: From the orthogonal relationship we have

$$\mathbf{X}'\varepsilon = \mathbf{X}'(\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{0}$$

Thus,

$$\mathbf{X}'\mathbf{Y} - \mathbf{X}'\hat{\mathbf{Y}} = \mathbf{0}$$

$$\mathbf{X}'\hat{\mathbf{Y}} = \mathbf{X}'\mathbf{Y}$$

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

Solve for **b**:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

where

$$\mathbf{b} := \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}$$

is the least square estimator of β.

Recall: The Hat-Matrix H

We define the Hat-Matrix $\mathbf{H}$ by

$$\hat{\mathbf{Y}} = \mathbf{HY}$$

Because $\hat{\mathbf{Y}} = \mathbf{Xb}$ and $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, we have

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

So that Hat-Matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Note the Hat-Matrix is symmetric and idempotent:

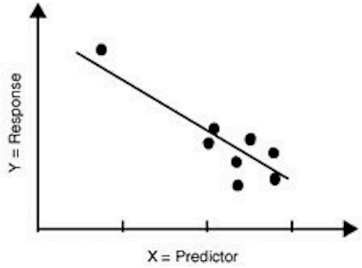
$$\mathbf{HH} = \mathbf{H}$$

The leverage h_{ii}

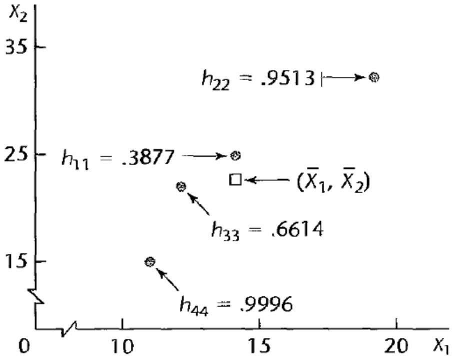
The main diagonal elements of the Hat-matrix **H** are called the **leverage** of each case, respectively, denoted by h_{ii}.

Simple linear regression:

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SS_{xx}}$$



Multiple regression:



Example data

Duncan {carData} R Documentation

Duncan's Occupational Prestige Data

Description

The `Duncan` data frame has 45 rows and 4 columns. Data on the prestige and other characteristics of 45 U. S. occupations in 1950.

Usage

```
Duncan
```

Format

This data frame contains the following columns:

type

Type of occupation. A factor with the following levels: `prof`, professional and managerial; `wc`, white-collar; `bc`, blue-collar.

income

Percentage of occupational incumbents in the 1950 US Census who earned $3,500 or more per year (about $36,000 in 2017 US dollars).

education

Percentage of occupational incumbents in 1950 who were high school graduates (which, were we cynical, we would say is roughly equivalent to a PhD in 2017)

prestige

Percentage of respondents in a social survey who rated the occupation as “good” or better in prestige

Source

Duncan, O. D. (1961) A socioeconomic index for all occupations. In Reiss, A. J., Jr. (Ed.) *Occupations and Social Status*. Free Press [Table VI-1].

```
> library(car)
> data(Duncan)
> head(Duncan)
           type income education prestige
accountant prof     62        86       82
pilot      prof     72        76       83
architect  prof     75        92       90
author     prof     55        90       76
chemist    prof     64        86       90
minister   prof     21        84       87
> m <- lm(prestige ~ education + income + type, data=Duncan)
> # Leverage
> hatvalues(m)
        accountant              pilot          architect
        0.05827491         0.07534370         0.07998300
            author            chemist           minister
        0.07061418         0.05869037         0.19120532
         professor            dentist           reporter
        0.07253152         0.11343770         0.25183588
          engineer         undertaker             lawyer
        0.06707952         0.08212343         0.09832374
         physician     welfare.worker            teacher
        0.09491067         0.09018720         0.08765839
         conductor         contractor      factory.owner
        0.36635189         0.22340093         0.14270993
     store.manager             banker         bookkeeper
        0.23091051         0.08195668         0.23514410
       mail.carrier   insurance.agent        store.clerk
        0.17194020         0.17797803         0.20877139
         carpenter        electrician        RR.engineer
        0.04860362         0.09885514         0.31468290
         machinist     auto.repairman            plumber
        0.06122920         0.04902871         0.08249185
```

```
> # Manually calculate the hat-values (leverage)
> X <- model.matrix(m)
> head(X)
           (Intercept) education income typeprof typewc
accountant           1        86     62        1      0
pilot                1        76     72        1      0
architect            1        92     75        1      0
author               1        90     55        1      0
chemist              1        86     64        1      0
minister             1        84     21        1      0
> H <- X %*% solve(t(X) %*% X) %*% t(X)
> data.frame(hatvalues = hatvalues(m), H.diag = diag(H))
                hatvalues     H.diag
accountant     0.05827491 0.05827491
pilot          0.07534370 0.07534370
architect      0.07998300 0.07998300
author         0.07061418 0.07061418
chemist        0.05869037 0.05869037
minister       0.19120532 0.19120532
professor      0.07253152 0.07253152
dentist        0.11343770 0.11343770
reporter       0.25183588 0.25183588
engineer       0.06707952 0.06707952
undertaker     0.08212343 0.08212343
lawyer         0.09832374 0.09832374
physician      0.09491067 0.09491067
welfare.worker 0.09018720 0.09018720
teacher        0.08765839 0.08765839
conductor      0.36635189 0.36635189
contractor     0.22340093 0.22340093
factory.owner  0.14270993 0.14270993
store.manager  0.23091051 0.23091051
banker         0.08195668 0.08195668
```

Leverage h_{ii}

Properties of h_{ii}:

$$0 \le h_{ii} \le 1 \qquad\qquad \sum_{i=1}^{n} h_{ii} = p$$

A leverage value is usually considered to be large if it is more than two times the mean leverage value $\bar{h}$ where

$$\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n} = \frac{p}{n}$$

So the rule-of-thumb cutoff value for leverage is $2p/n$.

Using h_{ii} to identify extrapolation

When there are 3 or more predictors in the model, it's hard to directly observe which value combination on the predictors may be extrapolating from the scope of the given dataset.

We can calculate the leverage value for the new data using

$$h_{\text{new}} = \mathbf{X}_{\text{new}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{\text{new}}$$

where $\mathbf{X}_{\text{new}} = [1, X_{\text{new},1}, X_{\text{new},2} \cdots X_{\text{new},p-1}]$ is the row vector of the new data values on the predictors. $\mathbf{X}$ is the given data matrix.

If h_{new} is well within the range of the leverage values for the cases in the given data set, no extrapolation is involved. If h_{new} is much larger than the leverage values for the cases in the given dataset, extrapolation is indicated.

Using h_{ii} to identify extrapolation

```
> # Identify extrapolation with hat values
> X.new <- matrix(c(1, 10, 80, 0, 0), nrow = 1)
> X.new
     [,1] [,2] [,3] [,4] [,5]
[1,]    1   10   80    0    0
> head(X)
           (Intercept) education income typeprof typewc
accountant           1        86     62        1      0
pilot                1        76     72        1      0
architect            1        92     75        1      0
author               1        90     55        1      0
chemist              1        86     64        1      0
minister             1        84     21        1      0
> X.new %*% solve(t(X) %*% X) %*% t(X.new)
          [,1]
[1,] 0.3986564
> max(hatvalues(m))
[1] 0.3663519
```

Outline

Various types of residuals

Name in Textbook	Formula	Name in R packages	Function in R packages
Raw residual	$e_i = Y_i - \hat{Y}_i$		
Semi-studentized residual	$e_i^* = e_i / s$		
Studentized residual	$r_i = \dfrac{e_i}{s\{e_i\}} = \dfrac{e_i}{s\sqrt{1-h_{ii}}}$	Standardized residual	`rstandard()`
Deleted residual	$d_i = Y_i - \hat{Y}_{i(i)} = \dfrac{e_i}{1-h_{ii}}$		
Studentized deleted residual	$t_i = \dfrac{d_i}{s\{d_i\}} = \dfrac{e_i}{s_{(i)}\sqrt{1-h_{ii}}} = e_i\sqrt{\dfrac{n-p-1}{\mathrm{SSE}(1-h_{ii})-e_i^2}}$	Studentized residual	`rstudent()`

Recall: The formula of the standard error of a residual

The residual vector: $\underset{n\times 1}{\mathbf{e}} = [e_1, e_2, \cdots, e_n]' = \mathbf{Y} - \hat{\mathbf{Y}}$

The variance-covariance matrix of the residual vector:

$$
\begin{aligned}
\boldsymbol{\sigma}^2\{\mathbf{e}\} =& \boldsymbol{\sigma}^2\{\mathbf{Y} - \hat{\mathbf{Y}}\} \\
=& \boldsymbol{\sigma}^2\{(\mathbf{I} - \mathbf{H})\mathbf{Y}\} \\
=& (\mathbf{I} - \mathbf{H})\boldsymbol{\sigma}^2\{\mathbf{Y}\}(\mathbf{I} - \mathbf{H})' \qquad \text{(Because } \mathbf{I} - \mathbf{H} \text{ is a constant matrix)} \\
=& (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{H})' \\
=& \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})' \\
=& \sigma^2(\mathbf{I} - \mathbf{H}) \qquad \text{(Because } \mathbf{I} - \mathbf{H} \text{ is symmetric and idempotent)}
\end{aligned}
$$

The deleted residuals

Obtain the residual for the ith case using the $\hat{Y}$ predicted by the model **fitted with all cases except the ith one**:

$$d_i = Y_i - \hat{Y}_{i(i)}$$

An algebraically equivalent expression for d_i that does not require repeatedly fitting the regression model omitting the ith case is:

$$d_i = \frac{e_i}{1 - h_{ii}}$$

Estimated variance of the deleted residuals

Note that a deleted residual is the prediction error for a new case. So the variance of d_i is the variance of $\hat{Y}$ for the prediction interval:

$$s^2\{d_i\} = s^2_{(i)}(1 + \mathbf{X}_i'(\mathbf{X}_{(i)}'\mathbf{X}_{(i)})^{-1}\mathbf{X}_i)$$

where $s^2_{(i)}$ is the estimated error variance of the model fitted when the ith case is omitted. $\mathbf{X}_{(i)}$ is the predictor matrix omitting the ith case.

An algebraically equivalent expression for $s^2\{d_i\}$ is:

$$s^2\{d_i\} = \frac{s^2_{(i)}}{1 - h_{ii}}$$

The studentized deleted residuals

$$t_i = \frac{d_i}{s\{d_i\}} = \frac{\dfrac{e_i}{1-h_{ii}}}{\dfrac{s_{(i)}}{\sqrt{1-h_{ii}}}} = \frac{e_i}{s_{(i)}\sqrt{1-h_{ii}}}$$

An algebraically equivalent expression for t_i without requiring repeatedly fitting the regression model omitting the ith case is:

$$t_i = e_i\sqrt{\frac{n-p-1}{\mathrm{SSE}(1-h_{ii})-e_i^2}}$$

And $t_i \sim t(n-p-1)$, because the deleted residual was obtained by a model fitted on a sample size of $n-1$.

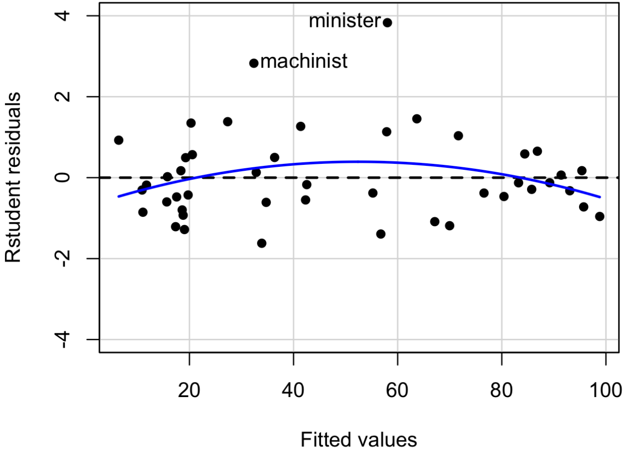
Various types of residuals

Name in Textbook	Formula	Name in R packages	Function in R packages
Raw residual	$e_i = Y_i - \hat{Y}_i$		
Semi-studentized residual	$e_i^* = e_i / s$		
Studentized residual	$r_i = \dfrac{e_i}{s\{e_i\}} = \dfrac{e_i}{s\sqrt{1-h_{ii}}}$	Standardized residual	`rstandard()`
Deleted residual	$d_i = Y_i - \hat{Y}_{i(i)} = \dfrac{e_i}{1-h_{ii}}$		
Studentized deleted residual	$t_i = \dfrac{d_i}{s\{d_i\}} = \dfrac{e_i}{s_{(i)}\sqrt{1-h_{ii}}} = e_i\sqrt{\dfrac{n-p-1}{\mathrm{SSE}(1-h_{ii})-e_i^2}}$	Studentized residual	`rstudent()`

$\widehat{\text{Prestige}} = b_0 + b_1(\text{Education}) + b_2(\text{Income}) + b_3(\text{TypeProf}) + b_4(\text{TypeWC})$

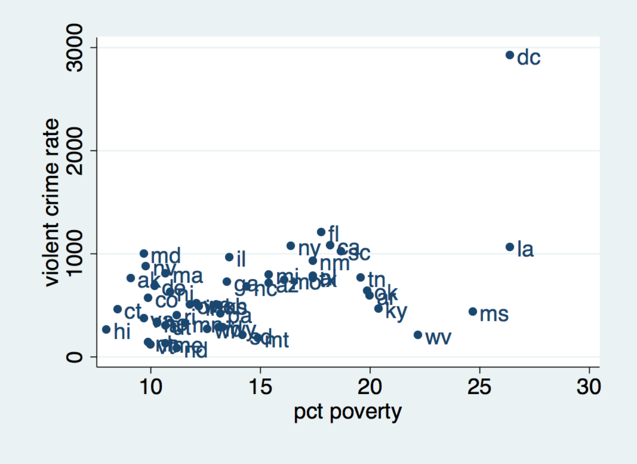
Note: The base level of “Type” is blue collar.

```
residualPlots(m, ~1, type="rstudent",
                id=list(labels=row.names(Duncan)), pch=16)
```



Outline

- What is the influence **DC** has on the fitted regression line?
- What is the influence **LA** has on the fitted regression line?



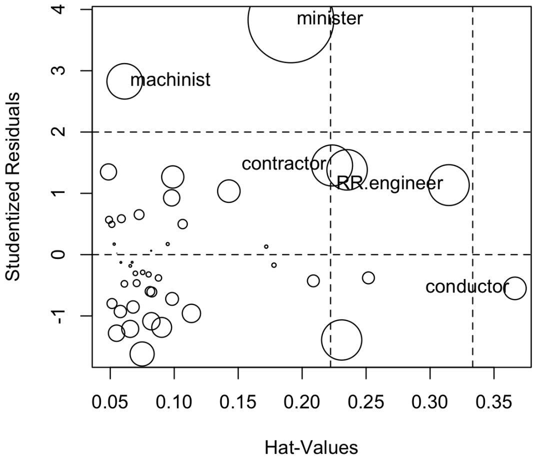
Outlier, leverage point, and influential point

- **Outlier**: an observation with a large standardized/studentized **residual**.
- **Leverage Point**: an observation with an extreme value on the **predictor** variable(s). (Leverage points have the **potential** to significantly influence the estimated slopes.)
- **Influential Point**: an influential point significantly influences the estimate of regression coefficients. (Influence can be thought of as the product of leverage and outlierness.)

One plot for all three measures:

e.g., R code:

```
car::influencePlot(m,id=list(labels=row.names(Duncan)))
```



The areas of the circiles are proportional to the **Cook's Distance**, a measure of the influence of each data point on the fitted model (see the following slide).

DFFITS

DFFITS_i measures the difference between the fitted value $\hat{Y}_i$ for the ith case when all n cases are used in fitting the regression model and the predicted value $\hat{Y}_{i(i)}$ for the ith case when the model is fitted omitting the ith case.

$$\text{DFFITS}_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{s_{(i)}\sqrt{h_{ii}}}$$

Note that $s_i\sqrt{h_{ii}}$ is the estimated standard deviation of $\hat{Y}_i$ (See proof in textbook comments as well as Lecture 5). Here $s_{(i)}$ is used to replace s_i to adjust for the "delete-case" scenario. So DFFITS_i is considered a standardized measure.

A rule-of-thumb: Consider a case influential if the absolute value of DFFITS exceeds 1 for small to medium data sets and $2\sqrt{p/n}$ for large data sets.

DFFITS (Cont'd)

An algebraically equivalent expression for DFFITS_i that does not require repeatedly fitting the regression model omitting the ith case is:

$$\text{DFFITS}_i = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}}$$

Note t_i is the studentized deleted residual. This shows that influence is a "product" of leverage and outlierness.

```
> dffits.duncan <- dffits(m)
> data.frame(Duncan, DFFITS=round(dffits.duncan, 2))
           type income education prestige DFFITS
accountant prof     62        86       82  -0.03
pilot      prof     72        76       83  -0.08
architect  prof     75        92       90  -0.10
author     prof     55        90       76  -0.13
chemist    prof     64        86       90   0.15
minister   prof     21        84       87   1.86
professor  prof     64        93       93   0.18
dentist    prof     80       100       90  -0.34
reporter     wc     67        87       52  -0.22
engineer   prof     72        86       88  -0.03
undertaker prof     42        74       57  -0.33
lawyer     prof     76        98       89  -0.24
physician  prof     76        97       97   0.06
```

Cook's Distance (D_i)

Cook's distance measure is an aggregate measure of the influence of the ith case on all n fitted values:

$$D_i = \frac{\sum_{j=1}^{n}(\hat{Y}_j - \hat{Y}_{j(i)})^2}{p * s^2}$$

An algebraically equivalent expression for Cook's distance D_i that does not require repeatedly fitting the regression model omitting the ith case is:

$$D_i = r_i^2 * \frac{h_{ii}}{1 - h_{ii}} * \frac{1}{p}$$

Note r_i is the studentized residual. This also shows that influence is a "product" of leverage and outlierness.

For interpreting Cook's distance, reference D_i to the $F(p, n-p)$ distribution and identify the corresponding percentile. If the percentile value of D_i is over 50 percent, the ith case has a major influence on the fit of the regression model.

```
> Cooks.d <- cooks.distance(m)
>
> p <- 5
> n <- nrow(Duncan)
> percentile <- 100 * pf(q=Cooks.d, df1=p, df2=n-p)
>
> data.frame(Duncan, Cooks.d=round(Cooks.d, 3), percentile=round(percentile, 1))
               type income education prestige Cooks.d percentile
accountant     prof     62        86       82   0.000        0.0
pilot          prof     72        76       83   0.001        0.0
architect      prof     75        92       90   0.002        0.0
author         prof     55        90       76   0.003        0.0
chemist        prof     64        86       90   0.004        0.0
minister       prof     21        84       87   0.517       23.8
professor      prof     64        93       93   0.007        0.0
dentist        prof     80       100       90   0.024        0.0
reporter         wc     67        87       52   0.010        0.0
engineer       prof     72        86       88   0.000        0.0
undertaker     prof     42        74       57   0.021        0.0
lawyer         prof     76        98       89   0.012        0.0
physician      prof     76        97       97   0.001        0.0
welfare.worker prof     41        84       59   0.028        0.0
teacher        prof     48        91       73   0.003        0.0
conductor        wc     76        34       38   0.036        0.1
contractor     prof     53        45       76   0.118        1.2
```

DFBETAS

DFBETAS is a measure of the influence of the ith case on each regression coefficient b_k $(k = 0, 1, \cdots, p-1)$.

$$\text{DFBETAS}_{k(i)} = \frac{b_k - b_{k(i)}}{s_{(i)}\sqrt{c_{kk}}} \qquad k = 0, 1, \cdots, p-1$$

where c_{kk} is the kth diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$.

Note the estimated variance-covariance matrix of $\mathbf{b}$ is $s^2(\mathbf{X}'\mathbf{X})^{-1}$. So $s_i\sqrt{c_{kk}}$ is the standard deviation of b_k. Here $s_{(i)}$ is used to replace s_i to adjust for the "delete-case" scenario. So $\text{DFBETAS}_{k(i)}$ is considered a standardized measure.

A rule-of-thumb: Consider a case influential if the absolute value of DFBETAS exceeds 1 for small to medium data sets and $2/\sqrt{n}$ for large data sets.

```
> dfbeta.duncan <- dfbetas(m)
>
> cutoff <- 2 * sqrt(1/nrow(Duncan))
> cutoff
[1] 0.2981424
>
> # Original data and dfbeta values for the cases with large influence on the coefficient of "education"
>
> data.frame(Duncan[which(abs(dfbeta.duncan[,"education"]) > cutoff), ], dfbeta.duncan[which(abs(dfbeta.dun
can[,"education"]) > cutoff), ])
              type income education prestige X.Intercept. education.1    income.1  typeprof     typewc
minister      prof     21        84       87    0.4506174   0.5717133 -1.56313689 0.5344010  0.2306229
contractor    prof     53        45       76    0.4638609  -0.6683542  0.09462482 0.6919298  0.4123061
factory.owner prof     60        56       81    0.2018532  -0.3302683  0.09462197 0.3431930  0.1849055
store.manager prof     42        44       45   -0.5134601   0.6154798  0.06380083 -0.7124710 -0.4390898
>
> # Original data and dfbeta values for the cases with large influence on the coefficient of "income"
>
> data.frame(Duncan[which(abs(dfbeta.duncan[,"income"]) > cutoff), ], dfbeta.duncan[which(abs(dfbeta.duncan
[,"income"]) > cutoff), ])
            type income education prestige X.Intercept. education.1   income.1    typeprof     typewc
minister    prof     21        84       87  0.450617411   0.5717133 -1.5631369  0.53440098  0.2306229
bookkeeper    wc     29        72       39 -0.008658934   0.2841271 -0.3699922 -0.08690809  0.3749392
RR.engineer   bc     81        28       67 -0.105385329  -0.1645380  0.7067142 -0.26893843 -0.2715007
```

Outline

```
> m <- lm(prestige ~ education + income + type, data=Duncan)
> summary(m)

Call:
lm(formula = prestige ~ education + income + type, data = Duncan)

Residuals:
    Min      1Q  Median      3Q     Max 
-14.890  -5.740  -1.754   5.442  28.972 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -0.18503    3.71377  -0.050  0.96051    
education     0.34532    0.11361   3.040  0.00416 ** 
income        0.59755    0.08936   6.687 5.12e-08 ***
typeprof     16.65751    6.99301   2.382  0.02206 *  
typewc      -14.66113    6.10877  -2.400  0.02114 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.744 on 40 degrees of freedom
Multiple R-squared:  0.9131, Adjusted R-squared:  0.9044 
F-statistic:   105 on 4 and 40 DF,  p-value: < 2.2e-16
```

```
> # Remove "minister"
>
> Duncan.r <- Duncan[-(which(row.names(Duncan)=="minister")), ]
>
> # Fit the regression model with the reduced data
>
> m.r <- lm(prestige ~ education + income + type, data=Duncan.r)
> summary(m.r)

Call:
lm(formula = prestige ~ education + income + type, data = Duncan.r)

Residuals:
     Min       1Q   Median       3Q      Max 
-17.0521  -6.4105  -0.7819   4.6552  23.5212 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.62984    3.22841  -0.505  0.61651    
education     0.28924    0.09917   2.917  0.00584 ** 
income        0.71813    0.08332   8.619 1.44e-10 ***
typeprof     13.43111    6.09592   2.203  0.03355 *  
typewc      -15.87744    5.28357  -3.005  0.00462 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.413 on 39 degrees of freedom
Multiple R-squared:  0.9344, Adjusted R-squared:  0.9277 
F-statistic:   139 on 4 and 39 DF,  p-value: < 2.2e-16
```

Remedial measures

1. **Nonlinear relationship:** transform X, or add polynomial terms.
2. **Influential points:** remove influential points and discuss separately.
3. **Non-constant variance of the error term:** transform Y; Weighted least square estimation; Robust estimators; Bootstrap.
4. **Non-normal distribution of the error term:** transform Y; Bootstrap.
5. **Dependent cases:** use advanced models (e.g., time series modeling, hierarchical linear models for nested data); Bootstrap.
6. **Multicollinearity:** caused by interaction/polynomial terms — center predictors; otherwise — remove some predictor(s).