Diagnostics

- *regression diagnostics*: check model assumptions to suggest further improvement after fitting. *The building of an empirical model is an iterative process*. During the process, it is required to check whether *the current fitted model is consistent with data*.
- **Q**: what assumptions needed to be checked?

  model: $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$

  - error structure: errors independent, equal variance, normally distributed
  - mean structure: whether $E(\boldsymbol{Y})=\boldsymbol{X\beta}$ is a correct structure
  - unusual observations: whether *some* observations do not fit the model
- two types of diagnostic techniques: numerical and graphical

Residual

- recall (residuals)
  - prediction: $\hat{\boldsymbol{Y}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{HY}$, $\quad \boldsymbol{H}$: hat matrix
  - residuals: $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X\beta} + (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\varepsilon} = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\varepsilon}$

    (Note: errors and residuals are different. **Q**: what difference?)
  - $\text{var}(\hat{\boldsymbol{\varepsilon}}) = \text{var}((\boldsymbol{I}-\boldsymbol{H})\boldsymbol{\varepsilon}) = (\boldsymbol{I}-\boldsymbol{H})^2\sigma^2 = (\boldsymbol{I}-\boldsymbol{H})\sigma^2$

    $\Rightarrow$ even though $\boldsymbol{\varepsilon}$ is uncorrelated and equal variance, $\hat{\boldsymbol{\varepsilon}}$ may be not

    (Note: $\boldsymbol{H}$ depends on $\boldsymbol{X}$ only)

- residuals are useful for detecting lack of fit and checking model assumptions

  (**Q**: Why residuals can do the works?

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

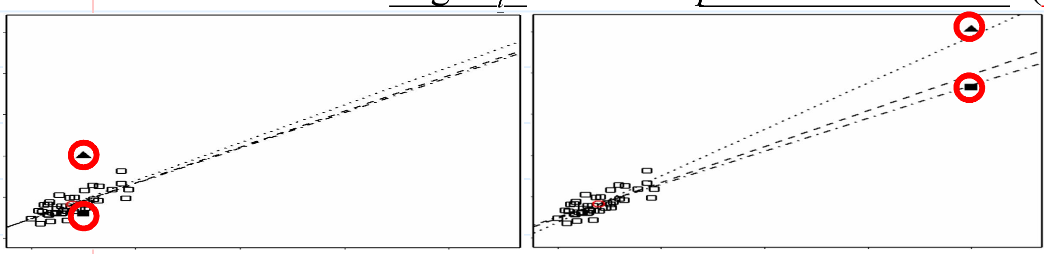
  $$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = (\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{H}_1\boldsymbol{X}_2\boldsymbol{\beta}_2) + ((\boldsymbol{I}-\boldsymbol{H}_1)\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}) = \hat{\boldsymbol{Y}}_{\boldsymbol{X}_1} + \hat{\boldsymbol{\varepsilon}}_{\boldsymbol{X}_1}$$)

Leverage

- leverage: $h_i \equiv \boldsymbol{H}_{ii}$ (Note 1. $\text{var}(\hat{\varepsilon}_i)=(1-h_i)\sigma^2$. Note 2. h_i is known before observing $\boldsymbol{Y}$)

  $\boldsymbol{x}_i$ whose h_i is large $\Rightarrow$ $\text{var}(\hat{\varepsilon}_i)$ small $\Rightarrow$ fitted model has to force to fit close to y_i

  $\boldsymbol{x}_i$ whose h_i is small $\Rightarrow$ $\text{var}(\hat{\varepsilon}_i)$ large $\Rightarrow$ in this $\boldsymbol{x}_i$, model cannot fit so well
  - h_i roughly determines how close $(\boldsymbol{x}_i, y_i)$ to the regression surface (i.e., $(\boldsymbol{x}_i, \hat{y}_i)$)
  - observations with large h_i's should be *paid more attention*. (**Q**: why?)



Q: why $\boldsymbol{x}_i$ with large leverage has stronger influence on fit?

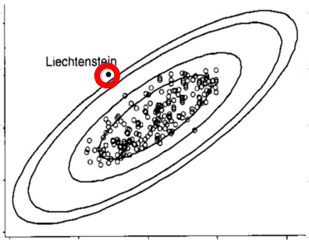
  - for linear model with an intercept, its fitted model must pass the point $(\bar{\boldsymbol{x}}, \bar{y})$

$$E(y) = \beta_0 + \sum_{i=1}^{p-1}\beta_i x_i = \left(\beta_0 + \sum_{i=1}^{p-1}\beta_i\bar{x}_i\right) + \sum_{i=1}^{p-1}\beta_i(x_i - \bar{x}_i) = \beta_0' + \sum_{i=1}^{p-1}\beta_i(x_i - \bar{x}_i)$$

$$\Rightarrow \quad \hat{\beta}_0' = \bar{y} \quad \Rightarrow \quad E(y) = \bar{y} + \sum_{i=1}^{p-1}\hat{\beta}_i(x_i - \bar{x}_i)$$

Q: why is it called leverage?

- some facts about leverage:
  - h_i corresponds to a Mahalanobis distance defined by $\mathcal{X}$ ($\boldsymbol{X}$ without intercept), actually, $$h_i = 1/n + [1/(n-1)]\,(\mathcal{X}_i - \bar{\mathcal{X}})^T\, \hat{\Sigma}_{\mathcal{X}}^{-1}\, (\mathcal{X}_i - \bar{\mathcal{X}}),$$
    where $\hat{\Sigma}_{\mathcal{X}}$ is the estimated covariance of $\mathcal{X}$
    - $\Rightarrow$ "extreme" $\mathcal{X}_i$ has large leverage
    - $\Rightarrow$ a little change of y_i value on point with large leverage will change the fit a lot



  - $\sum_{i=1}^{n} h_i = p$, and $1 \geq h_i \geq 1/n \;\; \forall\, i$

    (**Q**: what h_i's are too large? Note: an average leverage is p/n
    $\Rightarrow$ large leverage $>> p/n$ $\Rightarrow$ "rule of thumb": if $h_i > 2p/n$, look closer)
  - $\text{var}(\hat{\boldsymbol{Y}}) = \text{var}(\boldsymbol{HY}) = \boldsymbol{H}\sigma^2 \;\Rightarrow\; \text{var}(\hat{y}_i) = h_i\sigma^2$
- (*internally*) *studentized residuals* r_i's:

  because $\text{var}(\hat{\varepsilon}_i) = (1-h_i)\sigma^2$, let $r_i = \hat{\varepsilon}_i \,/\, [(1-h_i)^{1/2}\hat{\sigma}]$,

  then $\text{var}(r_i) \approx 1$ (if model assumptions are correct)
  - non-constant variance removed
  - dependence is very small in practice
  - sum of r_i's is not zero
  - r_i is slightly correlated with $\hat{y}_i$.
  - studentized residuals are preferred in residual plots
  - if there is some underlying heteroscedasticity (i.e., violation of $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$) in the errors, studentization cannot correct it

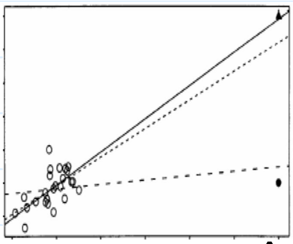
❖ **Reading**: Faraway (2005, 1st ed.), 4.2.1 ❖ **Further reading**: D&S, 8.1

Outlier

- an outlier is a point that does not fit the current model (**Q**: possible cause?)

  $\Rightarrow$ usually, large residual (**Q**: why?)
- **Q**: is there a problem if (raw studentized) residuals are used to detect outliers?

  $\Rightarrow$ outliers may affect the fit (see plot)



- **idea**: exclude i^{th} observation and re-compute the estimates to get $\hat{\beta}_{(i)}$ and $\hat{\sigma}^2_{(i)}$, where (i) denotes that the i^{th} case has been excluded. Then, consider $y_i - \hat{y}_{(i)}$, where $\hat{y}_{(i)} = x_i^T \hat{\beta}_{(i)}$. (**Q**: why is it better in detecting outliers?)

  $$\text{var}(y_i - \hat{y}_{(i)}) = \sigma^2(1 + \boldsymbol{x}_i^T(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\boldsymbol{x}_i)$$ (Hint. prediction of future observation)
- *jackknife* (or *externally studentized*, or *crossvalidated*) *residuals*

  $$t_i = (y_i - \hat{y}_{(i)}) / [(1 + \boldsymbol{x}_i^T(\boldsymbol{X}_{(i)}^T\boldsymbol{X}_{(i)})^{-1}\boldsymbol{x}_i)^{1/2}\,\hat{\sigma}_{(i)}]$$

  which are distributed as $t_{(n-1)-p}$ under null, if model is correct and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$
- a simpler way to calculate t_i (avoid doing n regression)

  $$t_i = \hat{\varepsilon}_i / [(1-h_i)^{1/2}\,\hat{\sigma}_{(i)}] = r_i\,((n-p-1)/(n-p-r_i^2))^{1/2}$$
- test for outliers
  - given a specific case i, conclude an outlier if $|t_i| > t_{n-1-p}^{(\alpha/2)}$

- in practice, a few (or all) t_i's will be tested ⇒ problem of *multiple testing*: need to adjust the significance level of the test accordingly

$$H_0\text{: no outlier in the } n \text{ observations} \quad \text{against} \quad H_1\text{: at least one outlier}$$

$$1-\alpha^* = 1-\text{Prob(Type I error}|H_0) = \text{Prob(all tests accept}|H_0)$$

$$= 1-\text{Prob(at least one rejected}|H_0) \geq 1-\Sigma_i \text{Prob(test } i \text{ rejects}|H_0) = 1-n\alpha$$

⇒ conclude an outlier if $|t_i| > t_{n-p-1}^{(\alpha/2n)}$

⇒ it's conservative, tends *not* to label points as outliers (especially when n large)

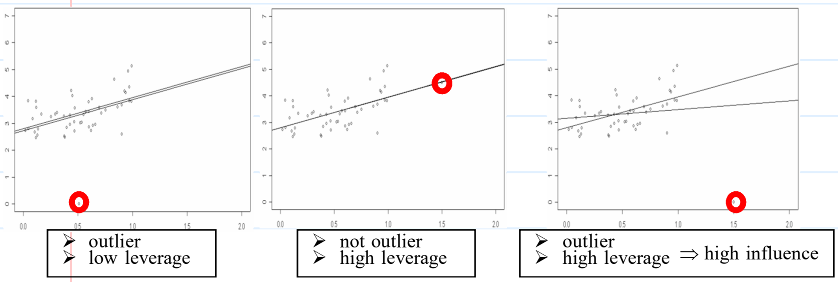
- some problems about outliers:
  - two or more outliers next to each other can hide each other
  - an outlier in one model may not be an outlier in another when the variables have been changed or transformed ⇒ reinvestigate outliers when model changed
  - the error distribution may not be Normal ⇒ larger residuals may be expected.
  - for large datasets, individual outliers are usually much less of a problem from the viewpoint of fit. In this case, it's still worth identifying outliers if these types of points are worth knowing about in the particular application. For large datasets, we need only worry about clusters of outliers.

- **Q**: What should be done if some observations are identified as outliers?
  - check for a data entry error first
  - Examine the physical context (sometimes, outliers may have physical significance)
  - exclude the point from the analysis
    - try to re-include later if model changed
    - if exclude permanently, report
  - use robust regression when outliers cannot be identified as mistakes or aberrations
  - dangerous to exclude them in an automatic manner

❖ **Reading**: Faraway (2005, 1st ed.), 4.2.2 ❖ **Further reading**: D&S, 8.1

Influential observation

- Each observations have different influence/contribution to the fitted model. Our fitted model should not change too much (i.e., robust) just because of adding/dropping a specific observation.
- *influential point*: one whose removal from data would *cause large change* in the *fit*.
- an influential point *may or may not* be an outlier and *may or may not* have large leverage but it will tend to have at least one of those two properties.



Q$_1$: what points are more influential?

Q$_2$: what information should be included in the detection of influential points?

- measures of influence (**Q**: how to numerically characterize "large change in fit"?)
  - change in coefficients: $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ (**Q**: how large is large?)
  - change in fit: $\hat{Y} - \hat{Y}_{(i)}$ (**Q**: how large is large?)
  - Cook's statistics/distances (scale and unit free):

$$D_i = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^T(\boldsymbol{X}^T\boldsymbol{X})(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}) / (p\hat{\sigma}^2)$$

$$= (\hat{Y} - \hat{Y}_{(i)})^T(\hat{Y} - \hat{Y}_{(i)}) / (p\hat{\sigma}^2)$$

$$= (1/p)\, r_i^2\, (h_i/(1-h_i))$$

$\Rightarrow$ it's a combination of residual and leverage. (**Q**: what are the effects of residual and leverage on Cook's statistic?)

$\Rightarrow$ **Q**: how large is large? If assume $\boldsymbol{X}$ is multivariate Normal, can do a test on D_i. However, normality may not be a reasonable assumption in practice.

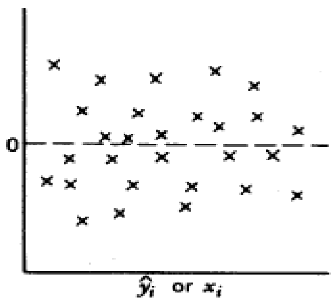
  - Others: DFFITS, Atkinson's modified Cook's statistics
- suggestion for identifying influential points: use relatively large D_i as an indication of a possible problem, then examine their $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ and/or $\hat{Y} - \hat{Y}_{(i)}$.

❖ **Reading**: Faraway (2005, 1st ed.), 4.2.3 ❖ **Further reading**: D&S, 8.3, 8.4

Residual plots

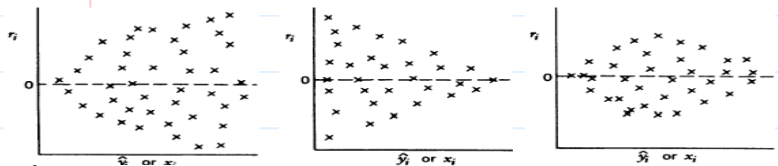
- *residual plots*: plot residuals (or absolute values of residuals) against (i) $\hat{y}$, (ii) x_k (for predictors in model and not in model), (iii) combination (or transformation) of x_k's, (iv) time order (if available), (v) any other quantities relevant to residuals
(**Q**: why draw residual plots?)

- in residual plots,
  - find *overall* patterns from the shape of *all* points (cf., residuals used in checking outliers or influential points $\Rightarrow$ identifying *individually unusual* point)
  - check assumptions: (i) non-constant variance; (ii) incorrectly specified mean structure (i.e., $E(\boldsymbol{Y}) = \boldsymbol{X\beta}$)
  - rather subjective
- a satisfactory residual plot (null plot)
  - constant variance
  - no curvature in the mean of residuals

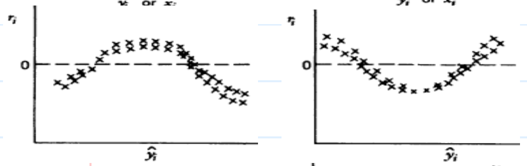


Note: one satisfactory residual plot cannot guarantee the residual plots for other variables will be satisfactory

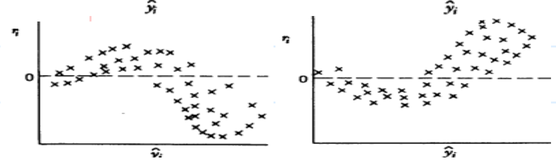
- some possible patterns in unsatisfactory residual plots:



evidence of non-constant variance

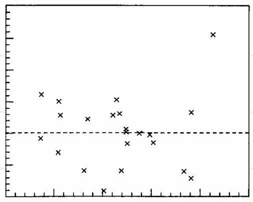


curvature in the mean of residuals
$\Rightarrow$ evidence of incorrectly specified mean structure

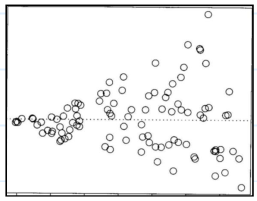


evidence of non-constant variance and incorrectly specified mean structure

- unfortunately, in real data set, it's rare the pattern is so clear (**Q**: what will you conclude from the residual plot on the right?)



- in models with many terms or models with complex non-linear mean structure, cannot necessarily associate shapes in a residual plot with a particular problem with the assumptions, e.g.,



  true model: $E(Y)=|x_1|/[2+(1.5+x_2)^2]$ with constant variance

  fitted model: $E(Y)=\beta_0+\beta_1 x_1+\beta_2 x_2$

- possible remedies for unsatisfactory residual plots

unsatisfactory residual plot	plot residuals against ...		
	$\hat{y}$	x_k	time order
non-constant variance	1. weighted least square 2. transform y	1. weighted least square 2. transform y	weighted least square
curvature in mean structure	1. add extra term 2. transform y	1. add extra term of x_k 2. transform y	add term of time in model

❖ **Reading**: F, 4.1.1 ❖ **Further reading**: D&S, 2.5

Non-constant variance

- if not sure, plot absolute values of residuals against $\hat{y}$, x_k's, time order
- when non-constant variance exists, $\hat{\beta}_{OLS}$ will be more variable than the best estimates ($\hat{\beta}_{OLS}$ unbiased but not BLUE) and $\hat{\sigma}$ wrong ($\Rightarrow$ test and C.I. inaccurate)

- It's better try to understand the cause of non-constant variance before taking any remedies, e.g., (1) larger response have more "room" to vary, (2) response constrained to lie between a maximum and a minimum, (3) response from Poisson distribution or binomial distribution, …

  $\Rightarrow$ discovering reasons to support the remedies you are going to take

- remedies for non-constant variance
  - weighted least squares
    - need weighting information (may from plotting residual vs. x_k) or
    - model the form of Σ and using IRWLS
  - transform Y (may use information from plotting residual vs. $\hat{y}$)
    - idea: find a transformation h such that $\text{var}(h(y_x))$ is a constant, (**Q**: how? Hint: δ-method)

$$h(y_x)=h(E(y))+(y-E(y))h'(E(y))+\ldots \Rightarrow \text{var}(h(y)) \approx \text{var}(y)[h'(E(y))]^2=c$$

$$\text{hope var}(h(y)) \text{ to be a constant } c \Rightarrow h'(E(y_x)) \propto 1/(\text{var}(y_x))^{1/2}$$

$$h(E(y)) \equiv \int 1/(\text{var}(y))^{1/2}\, d(E(y))$$

      - Example 1: $\text{var}(y_x) \propto [E(y_x)]^2 \Rightarrow$ suggest $h(y)=\log(y)$
      - Example 2: $\text{var}(y_x) \propto E(y_x) \Rightarrow$ suggest $h(y)=y^{1/2}$
    - Note: in residual plot, tend to see $[\text{var}(y_x)]^{1/2}$ rather than $\text{var}(y_x)$ (example?)

- practical problems:
  - if $y_i \leq 0$, for some i, square root or log transformations fail $\Rightarrow$ can do transformation on y_i+d, where d is some small amount s.t. $y_i+d>0$ for all i
  - transformation may make interpretation difficult
- example of transformations

sqrt(y_i)	var(y_i) $\propto$ E(y_i)	useful for count data from Poisson distribution
log(y_i)	var(y_i) $\propto$ [E(y_i)]^2	very common, good candidate if the range of $\boldsymbol{Y}$ is very broad
$1/y_i$	var(y_i) $\propto$ [E(y_i)]^4	appropriate when responses are "bunched" near zero, but, in markedly decreasing numbers, large responses do occur
$\sin^{-1}$(sqrt(y_i))	var(y_i) $\propto$ E(y_i)(1 - E(y_i))	for binomial proportions

- do nothing $\Rightarrow$ because (i) $\hat{\boldsymbol{\beta}}_{OLS}$ is still unbiased, although not BLUE; (ii) tests and C.I. inaccurate, but bootstrap may be used to get more accurate results
- use generalized linear model (e.g., Poisson/binomial $y \Rightarrow$ var($y_{\boldsymbol{x}}$): function of E($y_{\boldsymbol{x}}$))

• formal test for non-constant variance
- regressing absolute residuals on $\hat{y}$ or x_k's

- data with replication $\Rightarrow$ can estimate variances of distinct $\boldsymbol{x}_i$'s and test their homogeneity (see D&S, 2.2)

- data without replication $\Rightarrow$ assign variance a model, test whether parameters in the model equal zero (see Weinberg (2005), 8.3.2)

- formal test may be good at detecting a particular kind of non-constant variance (depending on the alternative hypothesis), but always do the residual plots

❖ **Reading**: Faraway (2005, 1st ed.), 4.1.1 ❖ **Further reading**: D&S, 2.2, 13.6

Curvature in the mean of residuals

• related to the concept of lack-of-fit (tests for lack-of-fit can be used if possible), i.e., the current model, E($\boldsymbol{Y}$)=$\boldsymbol{X\beta}$, may need to be modified for *achieving better fitting*

• A simple test for curvature: test whether a plot of residuals versus a quantity $\boldsymbol{U}$ (e.g., $\hat{y}$ or x_k's) is a null plot or has curvature

$\Rightarrow$ refit the original mean structure with an additional term $\boldsymbol{U}^2$ added

$\Rightarrow$ significant t-test for $\boldsymbol{U}^2$ suggests curvature (be aware of collinearity between $\boldsymbol{U}^2$ and other terms in original mean structure)

• **Q**: how to identify *why* the non-linearity happened?
- plot residuals against $\hat{y} \Rightarrow$ can tell you whether some problems exist, but cannot tell you why

- plot residuals against x_k's or y against x_k's $\Rightarrow$ may tell you why this problem happened, but in multivariate regression there may exist correlation between predictors, then it's difficult to find why
- *added variable* (*partial regression*) plots
  - recall:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon = (X_1\beta_1 + H_1X_2\beta_2) + ((I-H_1)X_2\beta_2 + \varepsilon)$$

  1. regress Y on all predictors except x_k $\Rightarrow$ get residuals $\hat{\varepsilon}_Y(x_k)$
  2. regress x_k on all predictors except x_k $\Rightarrow$ get residuals $\hat{\varepsilon}_{x_k}$
     - $\hat{\varepsilon}_Y(x_k)$: part of Y not explained by all predictors except x_k
     - $\hat{\varepsilon}_{x_k}$: part of x_k not explained by all predictors except x_k
  3. plot $\hat{\varepsilon}_Y(x_k)$ versus $\hat{\varepsilon}_{x_k}$
  - the slope of a fitted line to the added variable plot is $\hat{\beta}_k$ and intercept=0 (the line passes (0, 0))
  - a strong relationship between the plotted quantities corresponds to a strong adjusted relationship between y and x_k
  - can be used to check if new predictors should be included
- partial residual plots
  - plot $\hat{\varepsilon} + \hat{\beta}_k x_k$ versus x_k $\Rightarrow$ same interpretation as added variable plots
  - $y - \Sigma_{j\neq k}\hat{\beta}_j x_j = \hat{y} + \hat{\varepsilon} - \Sigma_{j\neq k}\hat{\beta}_j x_j = \hat{\varepsilon} + \hat{\beta}_k x_k$

- remedies for curvature $\Rightarrow$ adjust the mean structure, E(Y)=$X\beta$, for better fitting
  - many many modeling techniques in addition to linear regression can be adopted (GLM, additive model, nonparametric regression, ACE, AVAS, regression trees, regression spline, MARS)
  - add more (polynomial or cross product) terms
    - may identify required terms from residual plot, added variable plot, or partial residual plot (polynomial model will be introduced in further lecture)
  - transformation of response or predictors. idea behind the approach:

    (i) a statistical model is a local approximation of the underlying system

    (ii) when the mean structure of the underlying system is non-linear and complex, a linear approximation over a relatively wide range of X may be inadequate (e.g.,

$$E(Y) = \beta_0 x_1^{\beta_1} x_2^{\beta_2})$$

    (iii)we sometimes can find suitable transformations of data that will permit a non-linear model to be better approximated (after transformation) by a linear one (e.g.,

$$E(\log(Y)) \approx \log(\beta_0) + \beta_1 \log(x_1) + \beta_2 \log(x_2))$$

- some examples:

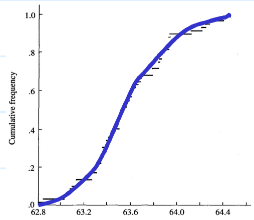
transformation		non-linear model
log(y)	log(x)	$E(y) = \alpha \prod x_j^{\beta_j}$
log(y)	x	$E(y) = \alpha \exp(\Sigma \beta_j x_j)$
y	log(x)	$E(y) = \alpha + \Sigma \beta_j \log(x_j)$
1/y	1/x	$E(y) = 1/[\alpha + \Sigma (\beta_j/x_j)]$
1/y	x	$E(y) = 1/(\alpha + \Sigma \beta_j x_j)$
y	1/x	$E(y) = \alpha + \Sigma \beta_j (1/x_j)$

- There exists numerical method for finding a suitable transformation to improve the fit and/or remedy non-constant variance (e.g., Box-Cox transformation, future lectures)

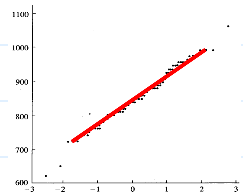
❖ **Reading**: F, 4.3, 7.2.4 ❖ **Further reading**: D&S, 8.2

Various plots and tests for diagnostics

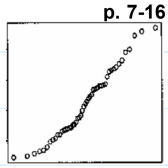
- Q-Q plot
  - **Q**: we often see the statement "$z_1, z_2, \ldots, z_m$ are i.i.d. from a cdf F", how to examine if F is an appropriate distribution assumption for z_i's? (Hint: examine the similarity btw cdf and empirical cdf)



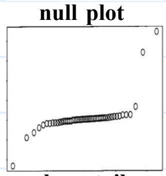
  - normal (probability) plot: assessing normality assumption of $\boldsymbol{\varepsilon}$ (Note: tests and C.I. depend on normality assumption)
    1. sort the data $\hat{\varepsilon}_{(1)} \le \hat{\varepsilon}_{(2)} \le \cdots \le \hat{\varepsilon}_{(n)}$
    2. plot $\hat{\varepsilon}_{(i)}$ against $\Phi^{-1}(i/(n+1))$, where Φ is the cdf of N(0, 1)
    - If the residuals are normally distributed, an approximately straight-line relationship will be observed (null plot)



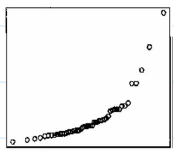
- non-normality: long-tail, short-tail, asymmetric
  - worst case is long-tail; mild non-normality can safely be ignored; the larger the sample size, the less troublesome the non-normality
  - for long-tail, (i) use test based on other distributions, or bootstrap, or permutation tests (ii) for estimation, use robust methods (e.g., least absolute deviation instead of least square)
  - asymmetric, transform $\boldsymbol{Y}$ (e.g., Box-Cox method)
  - short-tail can be reasonably ignored
- formal tests exists (such as Kolmogorov-Smirnov test), but not as flexible as the Q-Q plot



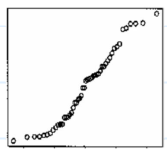
null plot



long tail



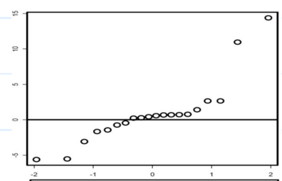
asymmetric

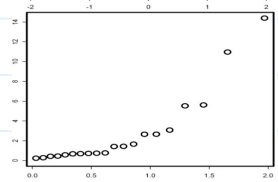


short tail

- normal plot can be applied to identify extreme values (e.g., in residuals, leverages, Cook's statistics, …): in the case, not interested in a straight line relationship, but rather looking for points *that depart from the straight line*

- half-normal plot
  1. sort the absolute data $|\hat{\varepsilon}|_{(1)} \le |\hat{\varepsilon}|_{(2)} \le \cdots \le |\hat{\varepsilon}|_{(n)}$
  2. plot $|\hat{\varepsilon}|_{(i)}$ against $\Phi^{-1}((n+i)/(2n+1))$
  - usually used to identify "extreme" values
  - can be used to examine residuals, leverages, Cook's statistics, treatment effects (especially for experimental data without replicates)





- diagnostic of correlated errors when a time order is available
  - plot $\hat{\varepsilon}$ against time
  - plot $\hat{\varepsilon}_{i+1}$ against $\hat{\varepsilon}_i$, when i related to time
  - use formal tests like the Durbin-Watson or runs test

$$DW = \frac{\sum_{i=2}^{n} \left(\hat{\epsilon}_i - \hat{\epsilon}_{i-1}\right)^2}{\sum_{i=1}^{n} \hat{\epsilon}_i^2}$$

    - $0 \le DW \le 4$
    - positively correlated $\Rightarrow DW \to 0$
    - negatively correlated $\Rightarrow DW \to 4$
    - under null (i.e., correlation=0) $\Rightarrow DW \approx 2$
    - null distribution depends on $\boldsymbol{X}$
  - use GLS when you have correlated errors

❖ **Reading**: F, 4.1.2, 4.1.3 ❖ **Further reading**: D&S, 2.4, 2.7, chapter 7