Topic 2 Simple Linear Regression

- SLR relating two variables, X & Y
  - linearity:

$$\mathrm{E}(Y|X=x)=\beta_0+\beta_1 x$$

  - Homoscedasticity:

$$\mathrm{Y}ar(Y|X=x)=\sigma^2$$

=> An equivalent way:

$$Y=\beta_0+\beta_1 X+e$$

Where:

  - $$E(e|X=x)=0$$
  - $$\mathrm{V}ar(e|X=x)=\sigma^2$$

=> a *working model* <= *modeling assumption*, rather than a statement of reality

  - two general models: the real model and the model you used to fit the data
    - in reality, not know the general relationship all the time (maybe sometimes very good understanding but mostly not)
  - A true model: **Non**-existence
- Some refs:
  - ***Parameters 参数***: unobserved <u>quantities</u> that characterize the model, here in the linear model as β_0, β_1, and σ^2
    - in frequentist statistics, these are considered <u>fixed constants</u>
      <=> because they are the representatives of the whole population
  - ***Estimators 估计值***: denoted with a "hat" ($\hat{}$) => $\hat{\beta}_0$, $\hat{\beta}_1$, & $\hat{\sigma}^2$
    => Estimators are functions of the data => therefore <u>random variables</u> => statistics

- ***Fitted values 拟合值***: predictions of the outcome $\mu(x) = \mathrm{E}(Y|X=x) = \beta_0 + \beta_1 x$, then:

$$\hat{y}_1 = \hat{\mu}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

here, the errors e_1, ..., e_n are **random variables** but not ~~parameters~~, because they are unobserved

PS:

  - Since: $y_i = \mathrm{E}(Y|X=x_i) + e_i$
    => e_i: standard error 标准误差, AKA the *vertical distance between the point* y_i *and the mean function* $\mathrm{E}(Y|X=x_i)$
    - Why e_i exists: $\sigma^2 > 0$, the observed value of the ith response (y_i) will not typically equal its $\mathrm{E}(Y|X=x_i)$
    - e_i => dependent on unknown parameters and not observable quantities

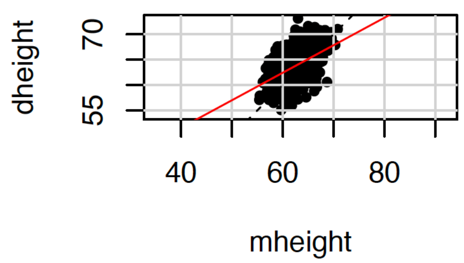
=> (2+1) assumptions concerning the errors

  - $\mathrm{E}(Y|X=x_i) = 0$
  - errors are all *independent*
  - errors are often assumed to be normally distributed
    但是：normality is much stronger than needed
    => The normality assumption is used *primarily* to obtain tests and confidence statements with **small samples**

- ***Residuals 残差***: estimates of the errors

$$\hat{e}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

- e.g. Heights data




```
## beta0.(Intercept) beta1.mheight
##          29.917437      0.541747
##    sigma
## 2.266311
```

- Coefficients: interpretation
  - ***Slope 斜率***: rate of change of the mean of Y as a function of X (Y的平均的X的方程的变化率)

$$\hat{\beta}_1 = \frac{\mathrm{d}}{\mathrm{d}x}\hat{\mu}(x) = \hat{\mu}(x+1) - \hat{\mu}(x)$$

  - ***Intercept 截距***: estimated mean of Y when $X = 0$

$$\hat{\beta}_0 = \hat{\mu}(0)$$

检查数据range (极差/全距)：如果截距meaningless，就有助于center the predictor

$$X_c = X - \bar{x}$$

- Ordinary least squares (OLS) estimation 普通最小二乘法估计
Def: 回归分析当中最常用估计β（回归系数）的方法是OLS，它基于误差值之上
估计β_0 & β_1有许多方法，最普遍的是OLS，其**最小化残差平方和 residual sum of squares (RSS)**：

$$\mathrm{RSS}(\beta_0, \beta_1) = \sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2$$

  - 滥用符号 abuse of notation：将符号用于固定但未知的量

最小化器 minimizers (Criterion for least squares) 可以表示为：

$$\hat{\beta}_1 = \frac{\mathrm{SXY}}{\mathrm{SXX}} = r_{xy}\frac{\mathrm{SD}_y}{\mathrm{SD}_x} = r_{xy}(\frac{\mathrm{SXY}}{\mathrm{SXX}})^{1/2}, \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

Table 2.1 Definitions of Symbols[a]

Quantity	Definition	Description
$\bar{x}$	$\sum x_i/n$	Sample average of x
$\bar{y}$	$\sum y_i/n$	Sample average of y
SXX	$\sum(x_i-\bar{x})^2 = \sum(x_i-\bar{x})x_i$	Sum of squares for the xs
SD_x^2	$\mathrm{SXX}/(n-1)$	Sample variance of the xs
SD_x	$\sqrt{\mathrm{SXX}/(n-1)}$	Sample standard deviation of the xs
SYY	$\sum(y_i-\bar{y})^2 = \sum(y_i-\bar{y})y_i$	Sum of squares for the ys
SD_y^2	$\mathrm{SYY}/(n-1)$	Sample variance of the ys
SD_y	$\sqrt{\mathrm{SYY}/(n-1)}$	Sample standard deviation of the ys
SXY	$\sum(x_i-\bar{x})(y_i-\bar{y}) = \sum(x_i-\bar{x})y_i$	Sum of cross-products
s_{xy}	$\mathrm{SXY}/(n-1)$	Sample covariance
r_{xy}	$s_{xy}/(\mathrm{SD}_x\mathrm{SD}_y)$	Sample correlation

[a]In each equation, the symbol Σ means to add over all n values or pairs of values in the data.

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})x_i}$$

在这种情况下，OLS产生的是参数的*estimates*，而不是参数的实际值

- To Derive OLS:
  - RSS: $\mathrm{RSS}(\beta_0,\beta_1)=\sum_{i=1}^n(y_i-\beta_0-\beta_1x_i)^2$
  - 找到最小值一种方法是对β_0和β_1进行微分，将导数设置为0，然后求解
    $\frac{\partial \mathrm{RSS}(\beta_0,\beta_1)}{\beta_0}=-2\sum_{i=1}^n(y_i-\beta_0-\beta_1x_i)=0$
    $\frac{\partial \mathrm{RSS}(\beta_0,\beta_1)}{\beta_1}=-2\sum_{i=1}^n x_i(y_i-\beta_0-\beta_1x_i)=0$
  - 整理，得：SLR中的*normal equations*
    $\beta_0 n+\beta_1\sum x_i=\sum y_i$
    $\beta_0\sum x_i+\beta_1\sum x_i^2=\sum x_iy_i$
  - 所以：
    $\hat\beta_0=\bar y-\hat\beta_1\bar x$
    $\beta_1=\frac{\sum x_iy_i-\bar y\sum x_i}{\sum x_i^2-\bar x\sum x_i}=\frac{\sum x_iy_i-n\bar x\bar y}{\sum x_i^2-n\bar x^2}=\frac{SXY}{SXX}$
    - Since: $SXX=\sum(x_i-\bar x)^2=\sum x_i^2-n\bar x^2$
      $SXY=\sum(x_i-\bar x)(y_i-\bar y).=\sum x_iy_i-n\overline{(xy)}$

    此外，令$c_i=\frac{x_i-\bar x}{\mathrm{SXX}}$，$d_i=\frac{1}{n}-c_ix_i$，则
    $\hat\beta_1=\frac{\sum(x_i-\bar x)(y_i-\bar y)}{\mathrm{SXX}}=\sum(\frac{x_i-\bar x}{\mathrm{SXX}})=\sum c_iy_i$
    $\hat\beta_0=\bar y-\hat\beta_1\bar x=\sum(\frac{i}{n}-c_i\bar x)=\sum d_iy_i$
    因此：$\hat y_i=\hat\beta_0+\hat\beta_1x_i=\sum(d_i+c_ix_i)y_i$，也为$y_i$的线性组合

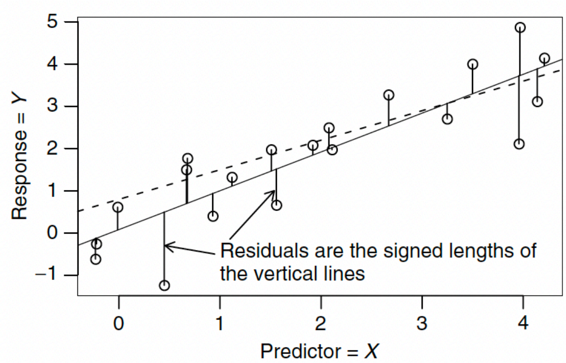


Figure 2.2 A schematic plot for OLS fitting. Each data point is indicated by a small circle. The solid line is the OLS line. The vertical lines between the points and the solid line are the residuals. Points below the line have negative residuals, while points above the line have positive residuals. The true mean function shown as a dashed line for these simulated data is $E(Y|X=x)=0.7+.8x$.

=> Estimating the variance σ^2

共同方差common variance/残差均方residual mean square σ^2通常估计为：

$$\sigma^2=\frac{\mathrm{RSS}}{n-2}$$

其中，

$$\mathrm{R}SS = \mathrm{R}SS(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \hat{e}_i^2$$

为最小化残差平方和 minimized residual sum of squares，或 simply RSS

- 为什么自由度为*n-2*？因为：
  如果误差是高斯分布（正态分布）(小样本量)，则$\hat{\sigma}^2$是卡方分布的倍数

$$\hat{\sigma}^2 = \frac{\hat{\sigma}^2}{n-2}\chi^2(n-2)$$

  在这种情况下$\mathrm{E}(\hat{\sigma}^2) = \sigma^2$, or $n-2$使得方差估计值无偏差 unbiased

  - 自由度（*df*）为m的χ^2随机变量的均值为m

$$\mathrm{E}\left(\hat{\sigma}^2 \mid X\right) = \frac{\sigma^2}{n-2}\mathrm{E}\left[\chi^2(n-2)\right] = \frac{\sigma^2}{n-2}(n-2) = \sigma^2$$

  PS: normality is NOT required for this result to hold

- Properties of estimators
  Review: **estimators** - <u>Functions</u> of the data (formulas, algorithms). They can always be applied to data regardless of the generating model.
  => Properties
  - *Numerical properties 数值性质*
    1. 回归线始终穿过数据的中心：

$$\hat{\mu}(\bar{x}) = \hat{\beta}_0 + \hat{\beta}_1\bar{x} = \bar{y}$$

       - 证明：$\mathrm{E}(Y|X=\bar{x}) = \bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1\bar{x} = \bar{y}$
    2. 残差始终平均为零 average out to zero：

$$\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i = 0$$

       前提：平均方程<u>有截距</u> —— 没有的话可能$\neq 0$
  - *Statistical properties 统计性质*
    1. 如果线性假设 linearity assumption 成立，则 OLS 估计量是无偏的

$$\mathrm{E}(\hat{\beta}_0|X) = \beta_0, \quad \mathrm{E}(\hat{\beta}_1|X) = \hat{\beta}_1$$

- 这个性质不需要样本为i.i.d
- 证明：
  $$\begin{aligned}\mathrm{E}(\hat{\beta}_1|X) = \mathrm{E}(\sum c_i y_i|X=x_i) &= \sum c_i \mathrm{E}(y_i|X=x_i) \\ &= \sum c_i(\beta_0+\beta_1 x_i) \\ &= \beta_0 \sum c_i + \beta_1 \sum c_i x_i\end{aligned}$$
  因为$\sum c_i = 0$ & $\sum c_i x_i = 1$，因此$\mathrm{E}(\hat{\beta}_1|X) = \hat{\beta}_1$
  而对β_0与之类似

2. $\hat{\beta}_0$ & $\hat{\beta}_1$ -> negatively correlated
   Proof:

   - $\hat{\beta}$的方差：
     因为y_i被假定为独立的，因此
     $$\begin{aligned}\mathrm{Var}(\hat{\beta}_1|X) = \mathrm{Var}(\sum c_i y_i|X=x_i) &= \sum c_i^2 \mathrm{Var}(y_i|X=x_i) \\ &= \sigma^2 \sum c_i^2 = \sigma^2/\mathrm{SXX}\end{aligned}$$
     在如上计算中，采用了$\sum c_i^2 = \sum (x_i-\bar{x})^2/\mathrm{SXX}^2 = 1/\mathrm{SXX}$
     对$\hat{\beta}_0$来说：
     $$\mathrm{Var}(\hat{\beta}_0|X) = \mathrm{Var}(\bar{y} - \hat{\beta}_1 \bar{x}|X)$$
     $$= \mathrm{Var}(\bar{y}|X) + \bar{x}^2 \mathrm{Var}(\hat{\beta}_1|X) - 2\bar{x}\mathrm{Cov}(\bar{y}, \hat{\beta}_1|X)$$
     因此计算协方差：（因为y_i各自独立，且$\sum c_i = 0$）
     $$\mathrm{Cov}(\bar{y}, \hat{\beta}_1|X) = \mathrm{Cov}(\frac{1}{n}\sum y_i, \sum c_i y_i|X)$$
     $$= \frac{1}{n}\sum c_i \mathrm{Cov}(y_i, y_j|X) = \frac{\sigma^2}{n}\sum c_i = 0$$
     因此：
     $$\mathrm{Var}\left(\hat{\beta}_0 \mid X\right) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\mathrm{SXX}}\right)$$
     最终，
     $$\begin{aligned}\mathrm{Cov}\left(\hat{\beta}_0, \hat{\beta}_1 \mid X\right) &= \mathrm{Cov}\left(\bar{y} - \hat{\beta}_1\bar{x}, \hat{\beta}_1 \mid X\right) \\ &= \mathrm{Cov}\left(\bar{y}, \hat{\beta}_1 \mid X\right) - \bar{x}\,\mathrm{Cov}\left(\hat{\beta}_1, \hat{\beta}_1 \mid X\right) \\ &= 0 - \sigma^2 \frac{\bar{x}}{\mathrm{SXX}} \\ &= -\sigma^2 \frac{\bar{x}}{\mathrm{SXX}}\end{aligned}$$
     - 进一步应用这些结果，得到拟合值的方差

$$
\begin{aligned}
\operatorname{Var}(\hat{y} \mid X=x) &= \operatorname{Var}\left(\hat{\beta}_0+\hat{\beta}_1 x \mid X=x\right) \\
&= \operatorname{Var}\left(\hat{\beta}_0 \mid X=x\right) \\
&+ x^2 \operatorname{Var}\left(\hat{\beta}_1 \mid X=x\right) \\
&+ 2x \operatorname{Cov}\left(\hat{\beta}_0, \hat{\beta}_1 \mid X=x\right) \\
&= \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\mathrm{SXX}}\right) \\
&+ \sigma^2 x^2 \frac{1}{\mathrm{SXX}} - 2\sigma^2 x \frac{\bar{x}}{\mathrm{SXX}} \\
&= \sigma^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\mathrm{SXX}}\right)
\end{aligned}
$$

- 所以：对于未来值$\tilde{y}_*$有
  $\operatorname{Var}\left(\tilde{y}_* \mid X=x_*\right)=\sigma^2\left(\frac{1}{n}+\frac{\left(x_*-\bar{x}\right)^2}{\mathrm{SXX}}\right)+\sigma^2$

进一步计算相关系数 correlated coefficient:
$\rho(\hat{\beta}_0, \hat{\beta}_1|X) = \frac{\operatorname{Cov}(\hat{\beta}_0|X, \hat{\beta}_1|X)}{\sqrt{\operatorname{Var}(\hat{\beta}_0|X)\operatorname{Var}(\hat{\beta}_1|X)}} = \frac{-\bar{x}}{\sqrt{\mathrm{SXX}/n+\bar{x}^2}}$
如果SXX中反映的预测变量的变化相对于$\bar{x}$较小，则相关性将接近±1

3. *Gauss-Markov theorem 高斯-马尔可夫定理* (optimality 最优性)：OLS估计量在所有无偏线性估计量 all unbiased linear estimators 中具有最小方差
   => BLUE: Best Linear unbiased estimator 最佳线性无偏估计
4. 如果线性成立且误差呈高斯分布：
   $e_i \mid X \sim \operatorname{NID}\left(0, \sigma^2\right) \quad i=1, \ldots, n$
   则OLS估计量$\hat{\beta}_0$ & $\hat{\beta}_1$也是高斯分布
5. 如果样本独立同分布 iid 且很大，则OLS估计量是渐近联合高斯分布的 asymptotically jointly Gaussian

- Central limit theorem (CLT) 中心极限定理
  根据CLT，如果
  1. 线性模型成立 holds
  2. 样本是iid
  3. 样本量大

  因此，asymptotically 渐近地 (approximately 近似地):

$$
\frac{\hat{\beta}_0-\beta_0}{\operatorname{se}\left(\hat{\beta}_0 \mid X\right)} \sim N(0,1), \quad \frac{\hat{\beta}_1-\beta_1}{\operatorname{se}\left(\hat{\beta}_1 \mid X\right)} \sim N(0,1)
$$

  因为SE需要被估计，有限n的更好近似是具有$n-2$自由度的t-分布

$$\frac{\hat{\beta}_0-\beta_0}{\widehat{\mathrm{se}}\left(\hat{\beta}_0 \mid X\right)} \sim t(n-2), \quad \frac{\hat{\beta}_1-\beta_1}{\widehat{\mathrm{se}}\left(\hat{\beta}_1 \mid X\right)} \sim t(n-2)$$

如果误差本身是高斯分布的，那么这些近似值是精确的

因此，基于如上估计，CI可以被描述为

$$\begin{aligned} \mathrm{CI}_0 &= \left\{\hat{\beta}_0 \pm t(\alpha/2, n-2)\widehat{\mathrm{se}}\left(\hat{\beta}_0 \mid X\right)\right\} \\ \mathrm{CI}_1 &= \left\{\hat{\beta}_1 \pm t(\alpha/2, n-2)\widehat{\mathrm{se}}\left(\hat{\beta}_1 \mid X\right)\right\} \end{aligned}$$

这些区间具有以下性质：

$$\mathrm{P}\left(\mathrm{CI}_0 \ni \beta_0\right) \approx 1-\alpha, \quad \mathrm{P}\left(\mathrm{CI}_1 \ni \beta_1\right) \approx 1-\alpha$$

解释：

- 覆盖概率 coverage probability 是在相同样本量下多次独立重复同一实验的结果
- 这些间隔是边际的 marginal（它们不保证同时覆盖 simultaneous coverage），并且它们的有效性取决于许多假设，如上所述。

- t-检验：推断两个变量是否彼此相关
  $H_0: \beta_1 = 0$ (Y **un**correlated with X)
  $H_1: \beta_1 \neq 0$ (Y correlated with X)
  - null hypothesis:

$$T=\frac{\hat{\beta}_0-0}{\hat{se}(\hat{\beta}_1|X)} \sim t(n-2)$$

    检验统计量T以SE为单位测量$\hat{\beta}_1$
    t-test拒绝：$\mid T \mid > t(1-\alpha/2, n-2)$

  解释：

  - 一小部分α会给出false positive
  - 拒绝原假设提供了X和Y之间关系的证据，但并不能保证这一点
  - 不拒绝原假设并不能证明变量X和Y不相关。 这只是意味着没有足够的证据表明它们是这样的
- 拟合值 fitted value: interpolate/extrapolate the fitted model for a new covariate value $x*$ 协变量值
  => $\hat{y}* = \hat{\mu}(x*) = \hat{\beta}_0 + \hat{\beta}_1 x*$
  拟合的均方误差 mean square error (MSE) 为：$\mathrm{MSE}[\hat{y}*] = \mathrm{E}[\hat{\mu}(x* - \mu(x*)|x*]^2$

(在多元回归的情况下，具体的表达式更容易写出来)

所以：$y*$的真实值为$y* = \beta_0 + \beta_1 x * + e*$
<= 其中，$e*$为附着于未来值的随机误差，大概方差为σ^2
=> 在估计系数的更常见情况下，预测误差变异性 prediction error variability 将具有由系数估计的不确定性引起的<u>第二个分量</u>
=> $\operatorname{Var}\left(\tilde{y}_* \mid x_*\right)=\sigma^2+\sigma^2\left(\frac{1}{n}+\frac{\left(x_*-\bar{x}\right)^2}{\mathrm{SXX}}\right)$
其中：

- 第一个σ^2：由于$e*$造成的变异性
- 剩下的项：估计线性系数造成的误差

=> 因此，如果$x*$与x_i十分类似，则第二个项将逐渐比第一个项变得更小；
类似地，如果$x*$与x_i变得很不一样，则第二个项将dominate
即：OLS预测为$\tilde{y}_*=\tilde{\mu}\left(x_*\right)+\tilde{e}_*=\hat{\mu}\left(x_*\right)+0=\hat{y}_*$ （OLS预测值与拟合值一致）
但是：$\tilde{y}_*=\tilde{\mu}\left(x_*\right)+\tilde{e}_*=\hat{\mu}\left(x_*\right)+0=\hat{y}_*$
这里的σ^2被称作***irreducible error 不可避免误差***

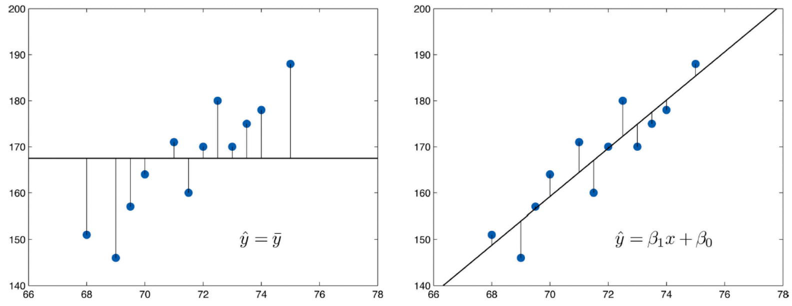
- 值得注意的是：在$x*$处的预测值的标准误差 SE of prediction (sepred)
  $\operatorname{sepred}\left(\tilde{y}_* \mid x_*\right)=\sigma\left(1+\frac{1}{n}+\frac{\left(x_*-\bar{x}\right)^2}{\mathrm{SXX}}\right)^{1/2}$
  其遵循*t*- distribution
  但是：当估计E($Y|X = x*$)时（具有特定身高的母亲的所有女儿的平均身高），由拟合值$\hat{y} = \beta_0 + \beta_1 x*$估计，标准差为：
  $\operatorname{sefit}\left(\hat{y} \mid x_*\right)=\hat{\sigma}\left(\frac{1}{n}+\frac{\left(x_*-\bar{x}\right)^2}{\mathrm{SXX}}\right)^{1/2}$
  其使用*F*-distribution且*df*也为*n-2* —— *2F(α; 2, n − 2)*
  <= 纠正<u>关于两个估计</u>而不是一个估计的同时推断
  - *F* table in `R`: `pf`
  - MLR: *p′F(α; p′, n − p′)*

- Coefficient of determination 决定系数: *R²*
  OLS 拟合导致方差分解 variance decomposition (ANOVA, Analysis of variance)
  - SYY: 平方综合 total sum of squares，观察到的响应总变化、忽略任何和所有预测变量
  - RSS: *unexplained* variation

  => SSreg: *sum of squares due to regression 回归平方和* —— `SSreg = SYY - RSS`
  =>$SSreg = SYY - \left(SYY - \frac{(SYY)^2}{SXX}\right) = \frac{(SXY)^2}{SXX}$
  => $\frac{\text{SSreg}}{\text{SYY}} = 1 - \frac{\text{RSS}}{\text{SYY}}$
  AKA: $\frac{1}{n}\mathrm{SYY} = \frac{1}{n}\mathrm{RSS} + \frac{1}{n}\mathrm{SSreg}$
  => $\frac{1}{n}\sum_{i=1}^n\left(y_i-\bar{y}\right)^2=\frac{1}{n}\sum_{i=1}^n\left[y_i-\hat{\mu}\left(x_i\right)\right]^2+\frac{1}{n}\sum_{i=1}^n\left[\hat{\mu}\left(x_i\right)-\bar{y}\right]^2$
  R^2: fraction of variance (FVE) of Y explained by X 由X解释的Y的方差分数

$R^2 = \frac{\text{SSreg}/n}{\text{SYY}/n} = 1 - \frac{\text{RSS}/n}{\text{SYY}/n}$
$=> 0 \leq R^2 \leq 1$

=> R^2: scale-free one-number summary of the strength of the relationship between the x_i and the y_i in the data

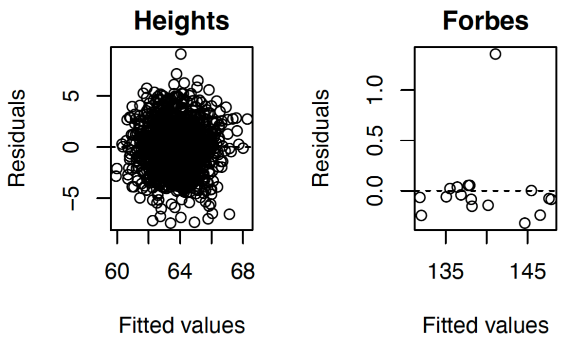


(a): Set I (b): Set II

Other properties:

- In SLR: $R^2 = [\text{Cov}(X, Y)]^2 = r_{xy}^2$
- R^2 only measure *model fit*
  R^2 => a perfect fit of the model to the data (<u>overfitting</u>)
- for large n, R^2 approximates FVE
- Adjusted R^2:
  $R^2 = 1 - \frac{\text{RSS}/df}{\text{SYY}/(n-1)}$
  => adjusted $R^2 < R^2$ (always)
  & $\approx R^2$ for large n
  & adjusted R^2 can be negative

- Residuals:
  - typically plotted against the fitted values (especially in multiple regression)



  - plot of residuals: find failures of assumptions
  - Curvature in residual plot -> the fitted mean function is <u>inappropriate</u>

=> Log transformation model

$\log Y = \beta_0 + \beta_1 + e$
$Y = \exp(\beta_0 + \beta_1 x + e) = \exp(\beta_0) \cdot \exp(\beta_1 x) \cdot \exp e$

    - in the original scale:
      the model is exponential 模型在原始尺度上呈指数增长
      the error is multiplicative 误差与原始比例相乘
    - 解释：$\exp(\beta_1)$表示X每增加一个单位，Y的变化百分比
      $\exp(\beta_1) = \frac{\exp[\beta_1(x+1)]}{\exp(\beta_1 x)}$

  => Log-log transformation model
  taking the log of both the response & the predictor:
  $\log Y = \beta_0 + \beta_1 \log x + e$
  $\quad Y = \exp(\beta_0) \cdot x^{\beta_1} \cdot \exp(e)$

    - 该模型是原始尺度的幂律 power law of the original scale
    - 如果β_1接近整数，则最容易解释 (quad regression)

- SUMMARY:
    - OLS是一种将线性模型拟合到数据的方法。它具有特定的数值属性。 其统计特性取决于数据生成机制 data generating mechanism
    - OLS的典型推论（例如t检验、CI）在很大程度上取决于建模假设，特别是简单随机样本（独立同分布 iid 观察值）
    - 模型拟合可能非常重要，但不太适合预测
    - 注意模型解读