Some properties about SS_{Reg}:

(1)$SS_{Reg} = \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$. When $\hat{\beta}_1 = 0$, i.e., the regression line is horizontal, $SS_{Reg} = 0$.

(2)SS_{Reg} may be considered a measure of that part of variability of the y_i which is associated with the regression line. The larger SS_{Reg} in relation to $SSTO$, the greater is the effect of the regression relation in accounting for the total variation in the observations. A measure is defined as the ratio of SS_{Reg} to $SSTO$. We will discuss more about this in next section.

3.4.1 Mean Squares and Expected Mean Squares

$$\begin{aligned} MS_{Reg} &= \frac{SS_{Reg}}{1} = SS_{Reg} \\ MS_{Error} &= \frac{SS_{Error}}{n-2} = \frac{RSS}{n-2} = \hat{\sigma}^2 = s^2 \\ E(MS_{Error}) &= \sigma^2 \\ E(MS_{Reg}) &= \sigma^2 + \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

We can show the last equation based on what have learned: **Proof:**

$$\begin{aligned} E(MS_{reg}) &= E(SS_{reg}) = E(\hat{\beta}_1^2 \sum (x_i - \bar{x})^2) \\ &= E(\hat{\beta}_1^2) \sum (x_i - \bar{x})^2 \\ &= [Var(\hat{\beta}_1) + E(\hat{\beta}_1)^2] \sum (x_i - \bar{x})^2 \\ &= \sigma^2 + \beta_1^2 \sum (x_i - \bar{x})^2 \end{aligned}$$

3.4.2 Distributions of Sum of Squares

Under the assumptions of simple linear regression and normality,

(1)

$$SS_{Error}/\sigma^2 \sim \chi^2_{n-2}$$

(2)Under the null $\beta_1 = 0$,

$$SS_{Reg}/\sigma^2 \sim \chi^2_1$$

(3)SS_{Error} and SS_{Reg} are independent.

Therefore, under the null

$$F = \frac{MS_{Reg}}{MS_{Error}} \sim F_{1,n-2}$$

It can be shown the this F statistic is the same as the t^2 for $\beta_1 = 0$ v.s. $\beta_1 \neq 0$. The “anova” function in R can be used to obtain the ANOVA table.

```
> anova(lm(rating~complaints, data=attitude))
Analysis of Variance Table

Response: rating
           Df Sum Sq Mean Sq F value    Pr(>F)
complaints  1 2927.6 2927.58  59.861 1.988e-08 ***
Residuals  28 1369.4   48.91
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

3.5 Correlation and Regression

A natural measure of the effect of x in reducing the variation in y, i.e., in reducing the uncertainty in predicting y, is to express the reduction in variation ($SSTO - SS_{Error} = SS_{Reg}$). as a proportion of the total variation:

$$r^2 = \frac{SS_{Reg}}{SSTO} = 1 - \frac{SS_{Error}}{SSTO}$$

The measure r^2 is called the **coefficient of determination**. Since $0 \leq SS_{Error} \leq SSTO$, it follows that:

$$0 \leq r^2 \leq 1$$

The square-root with sign is called correlation coefficient. Let

$$\begin{aligned} s_{xx} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \\ s_{yy} &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \\ s_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \end{aligned}$$

Then the correlation coefficient between the $x's$ and $y's$ is

$$r = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$$

The slope of the least squares line is

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}}$$

and therefore

$$r = \hat{\beta}_1 \sqrt{\frac{s_{xx}}{s_{yy}}}$$

It can also be proved that

$$\frac{\hat{y} - \bar{y}}{\sqrt{s_{yy}}} = r \frac{\hat{x} - \bar{x}}{\sqrt{s_{xx}}}$$

The interpretation is as follows:

Suppose that $r > 0$ and that x is one standard deviation greater than its average; then the predicted value of y is r standard deviations bigger than its average.

3.6 Regression Approach to ANOVA with Categorical Predictors

Is this possible? Yes. With the help of dummy variables. Beyond the scope of 120C.

3.7 Diagnostics for the simple linear regression model

The assumptions for simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where ϵ_i is a random error term such that

(1) $E(\epsilon_i) = 0$

(2) $Var(\epsilon_i) = \sigma^2$

(3) ϵ_i and ϵ_j are independent for all i and j.

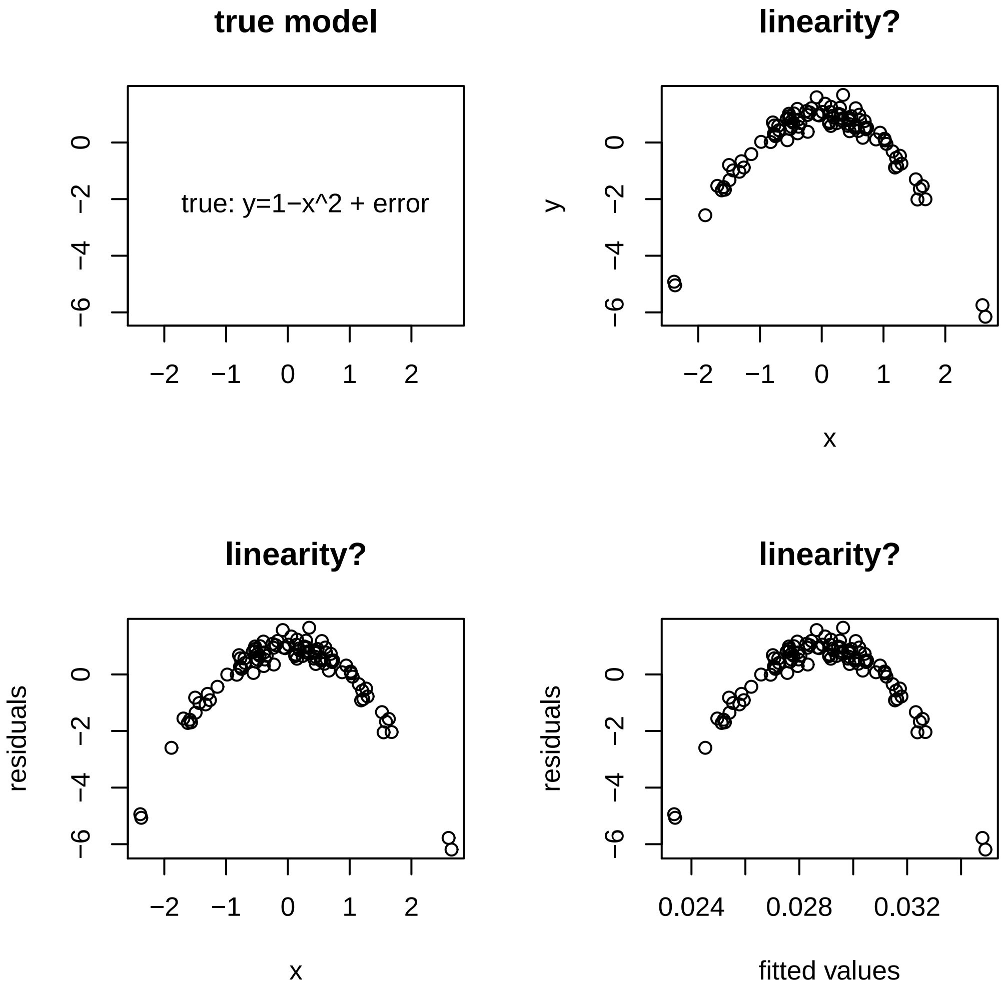
The simple linear regression model consists of two components: the **systematic** component and the **random component**. The diagnostics of the simple linear regression model also focus on these two components:

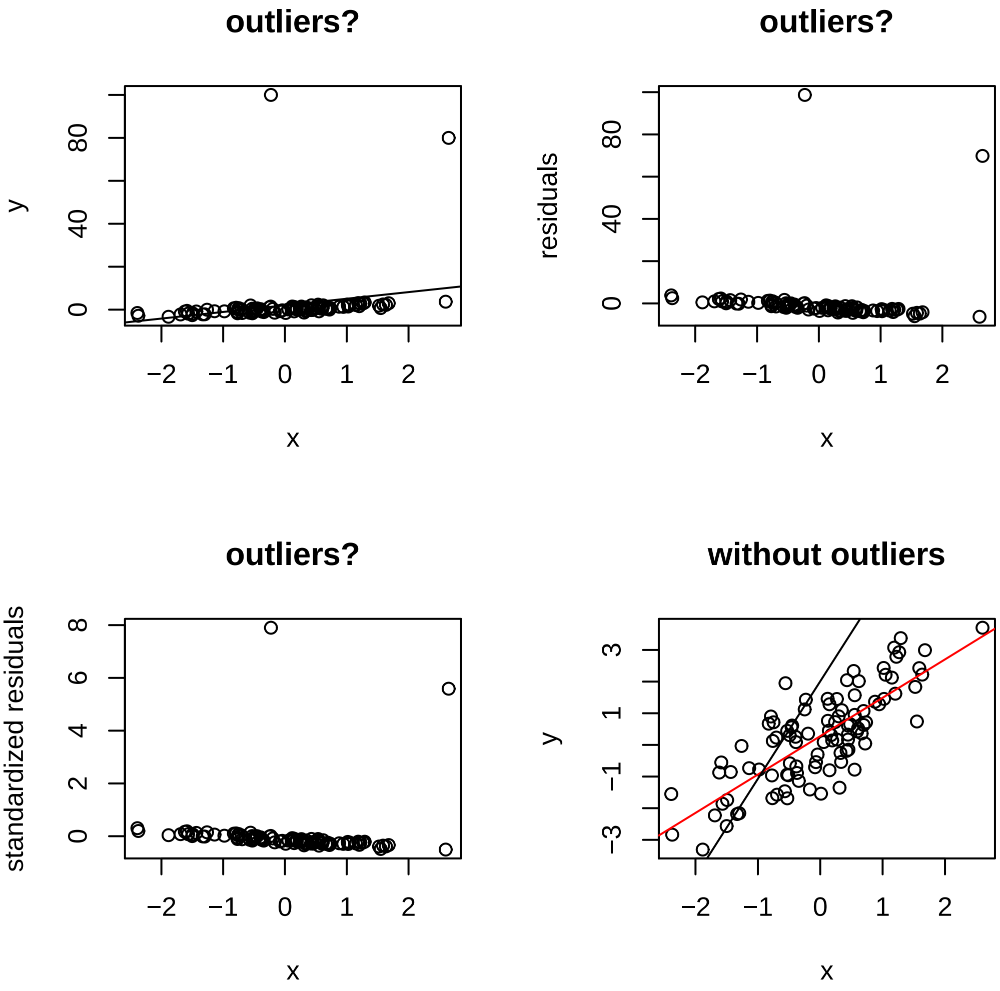
- Diagnostics for the systematic component (or the "mean" model)
- Diagnostics for the random component

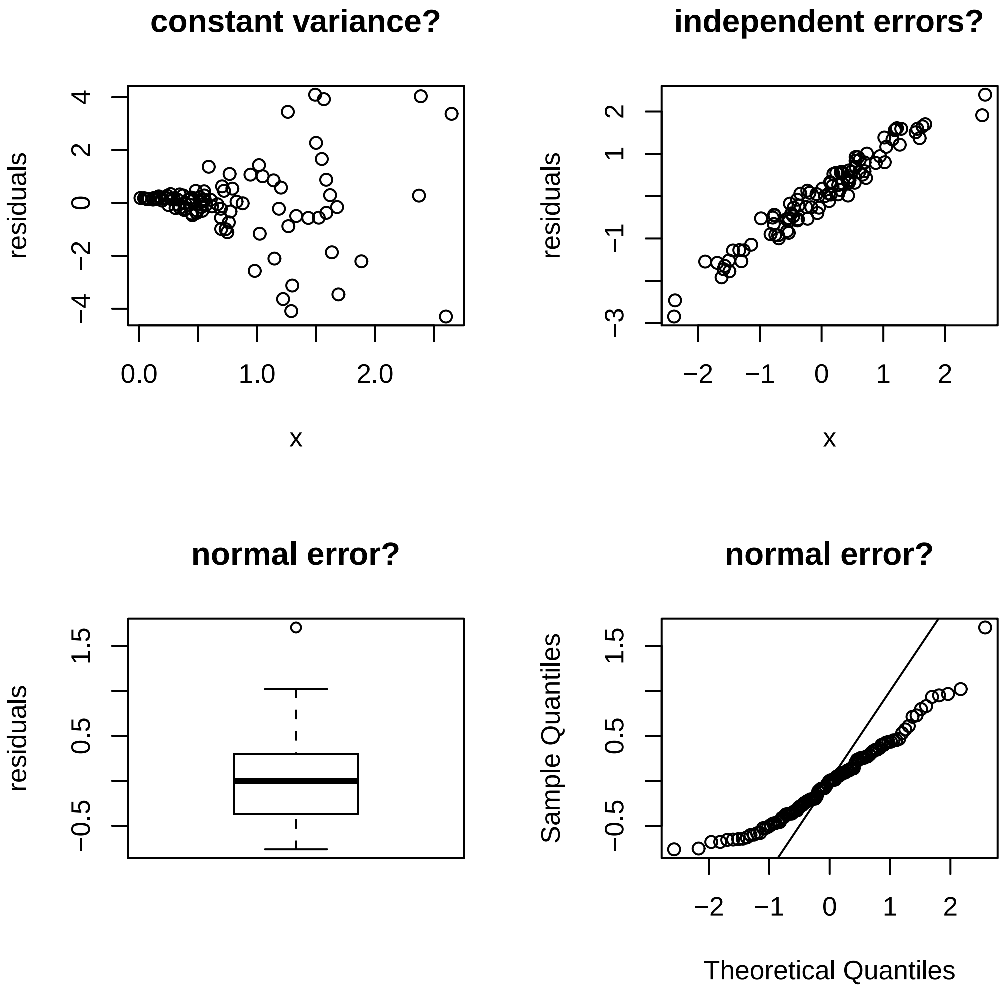
Most dialogistic procedures are **visual** and **subjective**, and focus on the **residuals** from the fitted model $e_i = y_i - \hat{y}_i$

3.7.1 Departures from Model to Be Studied by Residuals

- The regression function is not linear (systematic)
- The model fits all but one or a few outliers (systematic)
- The error terms do no have constant variance (random)
- The error terms are not independent (random)
- The error terms are not normally distributed (random)


true model
true: y=1−x^2 + error
0
−2
−4
−6
−2
−1
0
1
2
linearity?
y
x
linearity?
residuals
x
linearity?
residuals
fitted values
0.024
0.028
0.032


outliers?
y
0
40
80
x
−2
−1
0
1
2
outliers?
residuals
0
40
80
x
−2
−1
0
1
2
outliers?
standardized residuals
0
2
4
6
8
x
−2
−1
0
1
2
without outliers
y
−3
−1
1
3
x
−2
−1
0
1
2


constant variance?
residuals
4
2
0
−2
−4
0.0
1.0
2.0
x
independent errors?
residuals
2
1
−1
−3
−2
−1
0
1
2
x
normal error?
residuals
1.5
0.5
−0.5
normal error?
Sample Quantiles
1.5
0.5
−0.5
−2
−1
0
1
2
Theoretical Quantiles

3.7.2 Diagnostics for the systematic component:

(1)Linearity: Is x approximately related to y in a linear fashion?

Methods to check for linearity:

- In simple linear regression, we can simply plot y vs. x and look for patterns
- An alternative way is to look at residuals vs. x and look for pattern. Under linearity, the residuals will be scattered around the horizontal line at 0. This strategy can also be used for multiple linear models.

(2)Outliers: Does one of the observations standout in the y space? Methods to look for possible outliers:

- Plot residuals v.s. x or $\hat{y}$
- Plot the standardized residuals ($e_i/\hat{\sigma}$) vs. the fitted values $\hat{y}$

  Consider the variance of e_i, which is defined as

  $$\frac{\sum(e_i - \bar{e})^2}{n-2} = \frac{\sum e_i^2}{n-2} = \frac{RSS}{n-2} = \hat{\sigma}^2$$

  $e_i/\hat{\sigma}^2$ is called the standardized residuals.

  Plotting standardized residuals is particularly helpful for distinguishing outlying observations, since it is easy to identify residuals that lie many standard deviations from zero. Most observations (roughly 95%) should be within 2 standard deviations from 0.
- Check the influence of individual observations. Does a single (few) observation(s) greatly effect the estimate of parameters (especially β's)?
- Check the predicted C.I. for suspected outliers. If one observation is suspected to be outlier, refit the model without it. Then treat the observation as new and look at the the 95% C.I. based on the fitted model.

(3)Leverage: Does any observation standout in the x space?

-Histogram or boxplot of the covariate x

-More sophisticated measure later

3.7.3 Diagnostics for the random component

(1) Constant variance?

- Plot the residual vs. the fitted value $\hat{y}$ or x. If constant variance holds, the vertical spread of the plot should roughly equal at each $\hat{y}$ value

  If the residuals have constant variance, then they are said to be homoscedastic. If the residuals have nonconstant variance, then they are said to be heteroscedastic.

(2) Independent errors? Plot residuals v.s. x or fitted values $\hat{y}$ to examine whether there is trend in the residuals.

(3) Normality? Small departures from normality is fine. Should check whether there is major departures.

- A boxplot of residuals. Other graphical tools, such as histogram, can also be used.
- Quantile-quantile plot of residuals (normal probability plot).

3.7.4 Fixes for non-constant variance

The distribution theory is fairly robust to violation for the normality assumption; however, non-constant variance can lead to incorrect inference and result in less efficient estimate.

Here we introduce two ways to fix this problem.

(a) Weighted least square (WLS)

Suppose $var(\epsilon_i) = \rho_i^2\sigma^2$. We fit the following model

$$\rho_i^{-1}y_i = \rho_i^{-1}\beta_0 + \rho_i^{-1}\beta_1 x_i + \rho_i^{-1}\epsilon_i$$

This model will result in constant variance in the residuals.

But in most situations, ρ_i is usually not available.

(b) Transformation

Suppose the variance is a function of the mean of y, i.e., $var[y] = \sigma^2(\mu)$, where $\mu = E[y]$.

Example: $y \sim Poisson(\mu)$. Then $E[y] = \mu$, $var[y] = \mu$.

Consider the sample means of a random sample $y_1, \cdots, y_n$, where the variance depends on the mean of y. By the central limit theorem,

$$\sqrt{n}(\bar{y} - \mu) \rightarrow_d N(0, \sigma^2(\mu))$$

In the Possion example, $\sqrt{n}(\bar{y} - \mu) \rightarrow_d N(0, \mu)$.

Consider a transformation function g. By the delta method,

$$\sqrt{n}[g(\bar{y}) - g(\mu)] \rightarrow_d N(0, [g'(\mu)]^2\sigma^2(\mu))$$

To obtain constant variance, $[g'(\mu)]^2\sigma^2(\mu))$ must be free from μ, i.e., it must be a constant.

$$\begin{aligned}
[g'(\mu)]^2\sigma^2(\mu)) &= c \\
&\Rightarrow g'(\mu) = \frac{\sqrt{c}}{\sigma(\mu)} \\
&\Rightarrow g(\mu) = \int \frac{d\mu}{\sigma(\mu)}
\end{aligned}$$

Back to the Poisson example,

$$g(\mu) = \int \mu^{-1/2} d\mu = 2\mu^{1/2}$$

If we let $g(y) = \sqrt{y}$, then

$$\sqrt{n}[\sqrt{\bar{y}} - \sqrt{\mu}] \rightarrow_d N(0, 1/4)$$

Some practical transformations:

log, power, box-cox, exponential

4 Multiple Regression

We will extend the simple linear regression to the situation when several (more than one) predictors should be considered.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

.

If p is large, it is convenient to consider vector-valued random variables.

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \epsilon_2 \\ \cdots & \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \epsilon_n \end{aligned}$$

4.1 Vector-Valued Random Variables (Random Vectors)

4.1.1 Definition

Back to the simple linear regression:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \ i = 1, \cdots, n$$

It can be rewritten in the matrass form:

$$\begin{aligned} Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} &= \begin{pmatrix} \beta_0 + \beta_1 x_{11} + \epsilon_1 \\ \beta_0 + \beta_1 x_{21} + \epsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \epsilon_n \end{pmatrix} \\ &= \begin{pmatrix} 1 \ x_{11} \\ 2 \ x_{21} \\ \vdots \\ n \ x_{n1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \\ &= X\beta + \epsilon \end{aligned}$$