Lecture 13

§1 Regression towards the mean

1. Regression towards the mean (向均数回归 / 均值回归)

可以用以下方式对 $\hat{y}$ 进行转换:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\Rightarrow \hat{y} = (\bar{y} - \hat{\beta}_1 \bar{x}) + \left(r \frac{S_y}{S_x}\right) x$$

(r 为 correlation coefficient, $r\frac{S_y}{S_x} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \cdot \frac{\sqrt{\frac{1}{n-2} S_{yy}}}{\sqrt{\frac{1}{n-2} S_{xx}}} = \frac{S_{xy}}{S_{xx}} = \hat{\beta}_1$)

$$\Rightarrow \frac{\hat{y} - \bar{y}}{S_y} = r \frac{x - \bar{x}}{S_x}$$

由于通常 $|r| < 1$, 上式表示了 $\hat{y}$ 的 standardized value < x 的 standardized value.
这被称为 Regression towards the mean

2. 一个例子

Generally, regression refers to going back to a previous state.
In the 1800s, Francis Galton's data analysis described how, among other things:

- Children of tall parents have a disproportionate tendency to be shorter than their parents
- Children of short parents have a disproportionate tendency to be taller than their parents

He labelled this "regression" because from generation to generation we appeared to be returning to a kind of previous state (the average height). This conclusion turned out to be wrong.
However, elsewhere in his career Francis was instrumental in bringing statistics to science, business, and politics. In 1859, his half-cousin Charles Darwin wrote of one of Francis's publications that "I do not think I ever in all my life read anything more interesting and original."

An intuitive explanation of regression towards the mean

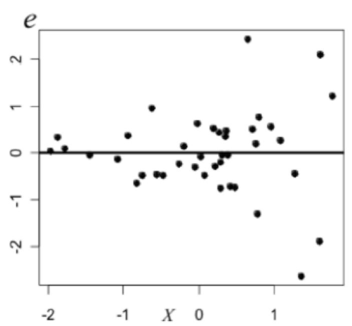
Imagine modelling height as a random variable with:

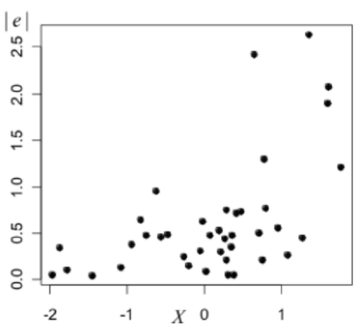
- A systemic part to take into account genetics, and
- A random part (environment etc)

The shortest individuals in a sample are likely to be the shortest because both the above parts are low. However, their parents or children can't be expected to have a low random part: it's random. Hence an apparent movement towards the mean.

§2 Transformations

若 Check 5: constant variance 没有通过, 则可以通过 transformation 来解决

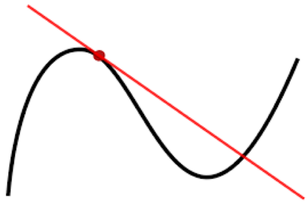




1. The Delta method

若 Y 有一分布 with mean μ & variance σ_Y^2. 令 $Z=f(Y)$（对 Y 进行一个 transformation）.
则 Z 可以近似表示为：

$$Z=f(Y)\approx f(\mu)+(Y-\mu)f'(\mu)$$



则可以求出

$$E(Z)\approx f(\mu)+E[Y-\mu]f'(\mu)=f(\mu)$$

$$var(Z)\approx var((Y-\mu)f'(\mu))=\sigma_Y^2[f'(\mu)]^2$$ （方差变为 $f(\mu)^2$ 倍）

以此近似 function of r.v. 的期望与方差

2、Linear regression 中的 Delta method

在 SLR 中，假设：

① $E(Y_i)=\mu_i$

② $var(Y_i)\propto V(\mu_i)$（Y_i 的方差正比于 μ_i [x_i 处 y_i 的均值] 的 function）

我们希望找到 transformation $Z=f(Y)$ 使得 $var(Z)\approx const$.

根据 Delta method，有

$$var(Z)\approx[f'(\mu)]^2 var(Y)$$
$$\propto[f'(\mu)]^2 V(\mu)\approx const\ c$$

$$\Rightarrow [f'(\mu)]^2\approx\frac{c}{V(\mu)}$$

$$\Rightarrow f'(\mu)\propto\frac{1}{\sqrt{V(\mu)}}$$

$$\Rightarrow f(\mu)\propto\int\frac{1}{\sqrt{V(\mu)}}d\mu$$

注：① 对 x 进行 transformation 可用于改进 nonlinearity

② 将 y 化为 $f(y)$ 后，为了保证 linearity，通常考虑对 x 进行相同的 transformation

2. Square root transformation

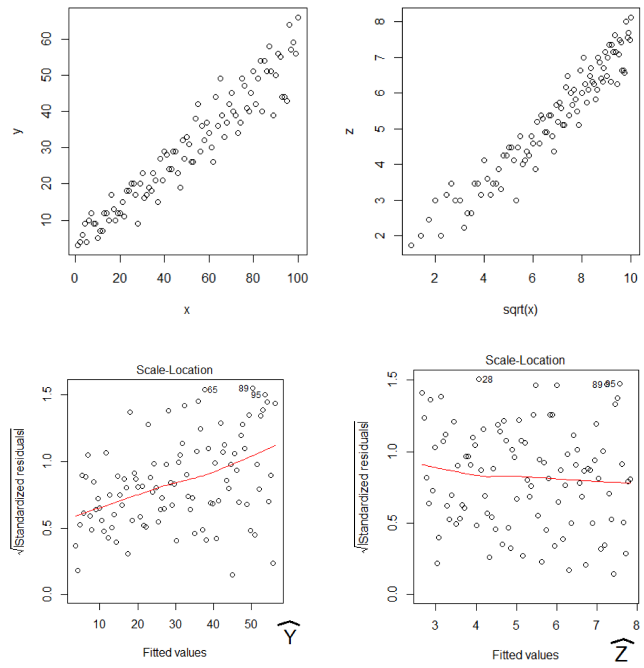
（不妨考虑 $Y_i\sim Poi(\mu_i)$ 的情况，此时有 $var(Y_i)=E(Y_i)=\mu_i$）

若 $var(Y)\propto V(\mu)=\mu$（Y_i 的方差正比于 μ_i），则

$$f'(\mu)\propto\frac{1}{\sqrt{\mu}}$$

$$\Rightarrow f(\mu)\propto\sqrt{\mu}$$

即令 $Z=\sqrt{Y}$ 可以将 $var(Y_i)$ 化为 constant



3、Logarithmic transformation

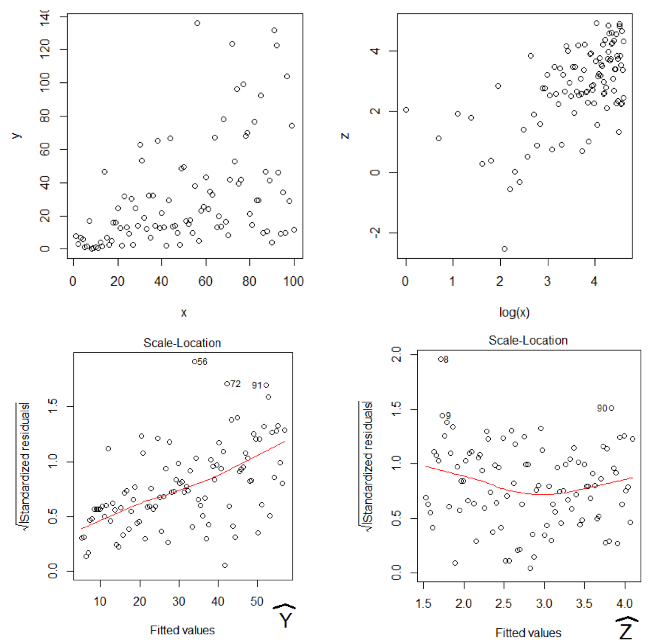
（不妨考虑 $Y_i \sim Exp(\lambda)$ 的情况，此时有 $E(Y)=\lambda^{-1}$，$var(Y)=\lambda^{-2}=(E(Y))^2$ ）

若 $var(Y) \propto V(\mu)=\mu^2$ （Y_i 的方差正比于 μ_i^2 ），则

$$f'(\mu) \propto \frac{1}{\mu}$$

$$\Rightarrow f(\mu) \propto \ln \mu$$

即令 $Z=\ln Y$ 可以将 $var(Y_i)$ 化为 constant



注：① 若 data 包含了 negative number，则可以使用 $\log(Y+k)$ 进行 transformation

② 对 log-transformed Y 的解释：

若仅 transform Y，则 model 变为：

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

$$Y = e^{\beta_0} e^{\beta_1 X} e^{e}$$

即 X 增加 1 个单位，Y 增加 e^{β_1} 个单位

Suppose we were plotting time-to-breakdown (Y) versus voltage (x, in kiloVolts), for some equipment. We fit

$$\widehat{\log Y} = 19 - 0.51X$$

So a 1-kV increase in voltage changes the estimated mean of Y by $e^{-0.51} = 0.6$. So if the voltage increases from 27kV to 28kV, the time to breakdown estimate is 60% of what it was. Ensure that the transformation leads to reasonable interpretations for your problem under study.

③ 对 log-transformed X 的解释：

若仅 transform X，则 model 变为：

$$Y = \beta_0 + \beta_1 \log(X) + e$$

若 X 增加为 k 倍，则 Y 增加 $\beta_1 \log k$

$$E(Y_{original}) = \beta_0 + \beta_1 \log(X)$$

$$E(Y_{new}) = \beta_0 + \beta_1 \log(kX)$$

$$E(Y_{new}) - E(Y_{original}) = \beta_1 \log k$$

4. Reciprocal transformation

若 $var(Y) \propto V(\mu) = \mu^4$ （Y_i 的方差正比于 μ_i^4），则

$$f'(\mu) \propto \frac{1}{\mu^2}$$

$$\Rightarrow f(\mu) \propto \frac{1}{\mu}$$

即令 $Z = 1/Y$ 可以将 $var(Y_i)$ 化为 constant

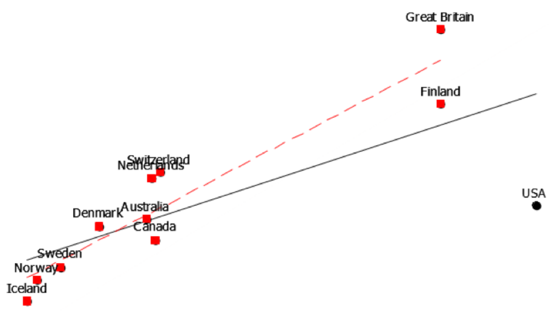
§3 处理 violated assumptions

1. 改变 underlying model

Last week, considering §3.2, we discussed briefly how we could change our underlying model, possibly swapping SLR for something more complicated, e.g.:

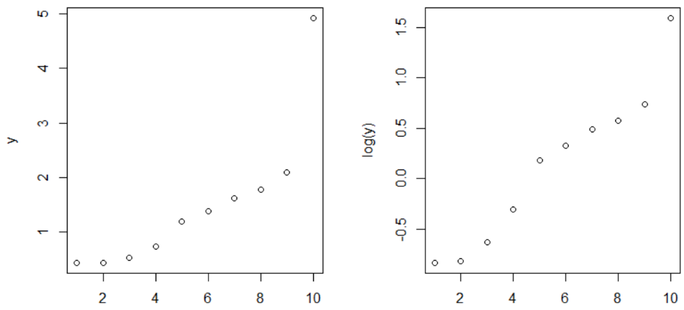
- Models that allow non-normal errors
- Nonlinear models to capture trends or unusual points
- Robust methods — reduce effects of outliers (median / quantile regression)

There are usually several options; remember you can report results with- and without outliers as well. For example in the cigarette dataset of Lecture Notes 5:

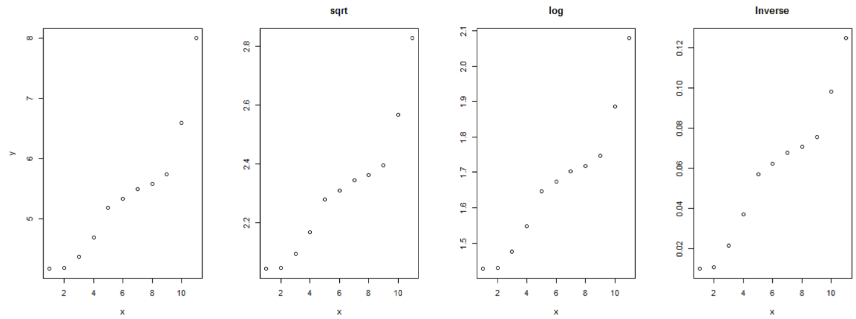


2. 利用 transformation 处理 outliers

Whereas, our current investigations in §3.3 suggest that if outliers are occurring at the tails of a skewed distribution, we might mitigate their effects through a transformation. For example, notice the effect of a logarithmic transformation:



Here we see how the inverse transform can powerfully bring in outliers.



3. 利用 transformation 处理 non-normality

Besides addressing outliers, we've seen how transforming Y can help* with nonconstant variance and nonlinearity. It can also help with error non-normality: recall slide 54 from Lecture Notes 5. In case the errors are not normal:

- CLT says that linear combinations of r.v.s are normally distributed, even if original r.v.s aren't
- Our estimators of β_0 and β_1 are linear combinations of r.v.s, so tests and CIs for them are robust against non-normality, as long as they are not too skewed and there aren't extreme outliers
- Prediction intervals aren't robust against non-normality

* Transforming X can be useful as well. For example, if X is very right-skewed, perform a log, $\sqrt{}$ or $1/x$ transformation.

4. Assumption for inference 的优先级

① $E(e) = 0$

② errors 间的 independency

③ constant variance （位次较后，因为当每个 x 有 similar number of observation 时，regression 对 noncostant variance robust）

④ Normality （尽管对于 PIs 很重要）