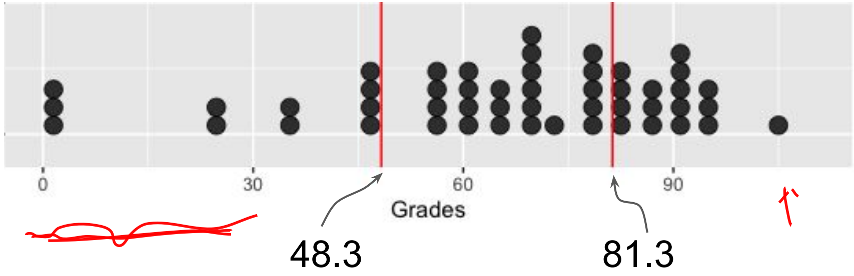
Simple linear regression

Chapter 14 of Rice

08/03/2021

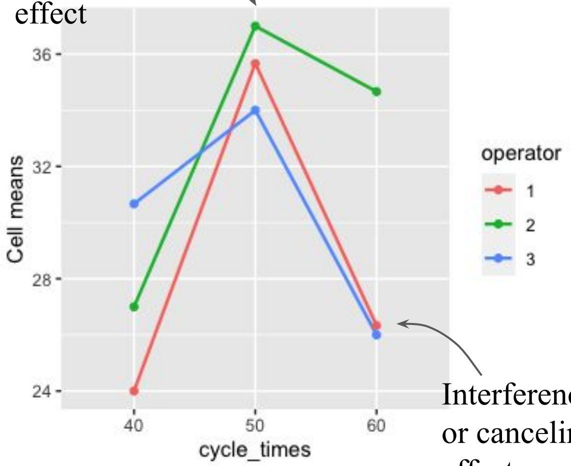
Midterm grades

Homework	25%	(7 assignments in total; lowest score dropped)
Labs	15%	(3 lab quizzes)
Midterm Exam	20%	
Final Exam	40%	
Participation	2%	(extra)



In the previous lecture,

Synergism, or compounding effect



Interference, or canceling effect

- Two-way ANOVA:

$$SS_{\text{Tot}} = SS_A + SS_B + SS_{AB} + SS_E.$$

  - Under $H_0: \alpha_1 = \cdots = \alpha_I = 0$, $SS_A/\sigma^2 \sim \chi^2_{I-1}$.
  - Under $H_0: \beta_1 = \cdots = \beta_J = 0$, $SS_B/\sigma^2 \sim \chi^2_{J-1}$.
  - Under $H_0:$ all δ_{ij}'s are zero, $SS_{AB}/\sigma^2 \sim \chi^2_{(I-1)(J-1)}$.
  - $SS_E/\sigma^2 \sim \chi^2_{n-IJ}$.

- One-way MANOVA:
  - Model assumption $(j = 1, \cdots, n_i,\ i = 1, \cdots, k)$:

$$\boldsymbol{Y}_{ij} = \underbrace{\boldsymbol{\mu}}_{\text{common mean level}} + \underbrace{\boldsymbol{\alpha}_i}_{\text{unique effect due to treatment } i} + \underbrace{\boldsymbol{\epsilon}_{ij}}_{\overset{\text{iid}}{\sim} N(\mathbf{0},\, \Sigma)}.$$

  - Wilk's lambda: $\boldsymbol{\Lambda}^* = |\mathbf{E}|/|\mathbf{B} + \mathbf{E}|$.
  -

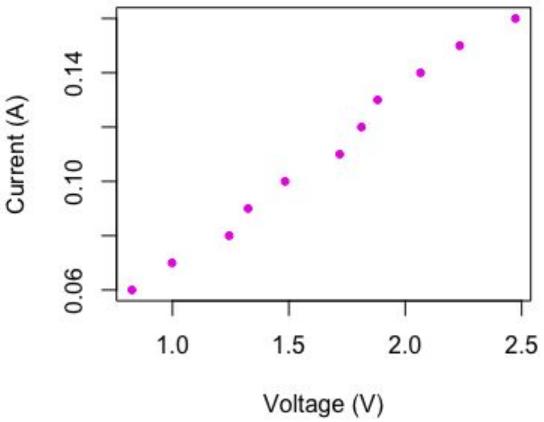
```
res.man <- manova(cbind(v1, v2, v3) ~ factor(Group),
                  data = dataset)
summary(res.man, 'Pillai')
```

Ohm’s law

Example 1. In 1825, Georg Ohm conducted experiments on resistance. He found that his data could be modeled through the equation:

$$I = \frac{V}{R},$$

where I is the current through the conductor in units of amperes, V is the voltage measured *across* the conductor in units of volts.



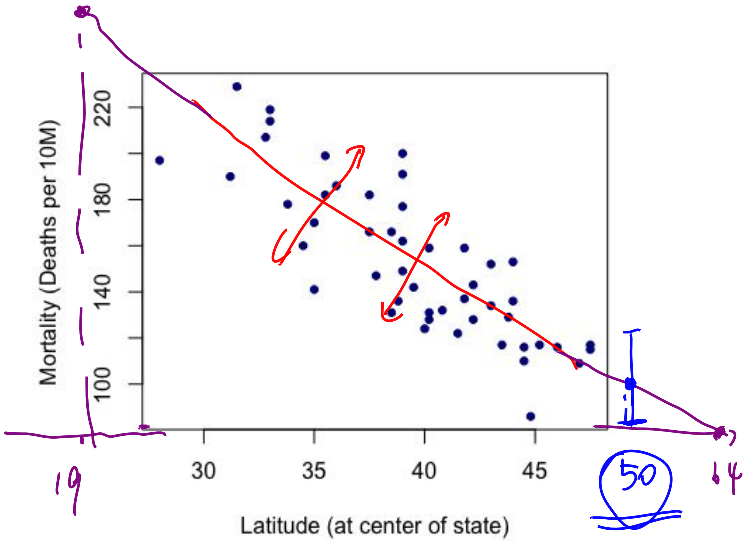
- Can you generalize this physics law if you were Ohm?
- How to find a constant R such that the resulting line fits the data best?
- How to account for measurement errors?

The most important of the early quantitative descriptions of the physics of electricity.

Skin cancer mortality vs. Latitude

Example 2. During the 50s, data were collected to examine the relationship between the mortality rate due to skin cancer (number of deaths per 10 million people) and the latitude at the center of each of 48 states in the United States (Alaska and Hawaii were not yet states. And, Washington, D.C. was included in the data set even though it is not technically a state.)

Alaska: 64.2008°N
Hawaii: 19.8968°N



- How to find a line that fits the data best?
- How to account for the variations?
- How to predict for the two un-observed states?

Simple linear regression (SLR)

Dependent	Independent
Response variable	Predictor variable
Y_1	X_1
Y_2	X_2
$\vdots$	$\vdots$
Y_n	X_n

Model assumption $(i = 1, \cdots, n)$:

$$Y_i = \underbrace{\beta_0}_{\text{common mean level}} + \underbrace{\beta_1 X_i}_{\text{slope times predictor variable}} + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0, \sigma^2)} .$$

intercept

$X_i = 0$

1. Linearity - Plotting Y_i vs X_i
2. Normality - QQ plot
3. Zero mean in error terms
4. Homoscedasticity
5. Independence

3–5: Residual plot

Heteroscedastic

Find estimators for β_0 and β_1

14.1 of Rice

08/03/2021

$\to \sum_{i=1}^{n}(x_i-\bar{x}_n)^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2$

$MSE = E(\hat{\theta}_n - \theta)^2$

Method of least squares

$\to \sum_{i=1}^{n}(x_i-\bar{x}_n)(Y_i-\bar{Y}_n) = \sum_{i=1}^{n} x_iY_i - \bar{Y}_n\sum_{i=1}^{n} x_i - \bar{x}_n\sum_{i=1}^{n} Y_i + n\bar{x}_n\bar{Y}_n$

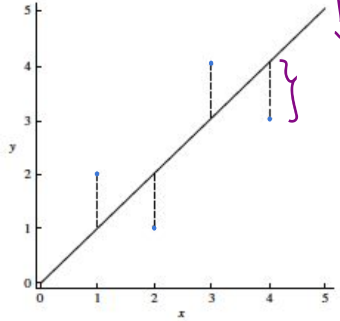
$n\bar{Y}_n\bar{x}_n \quad n\bar{x}_n\bar{Y}_n$

$= \sum_{i=1}^{n} x_iY_i - n\bar{x}_n\bar{Y}_n.$

Response variable	Predictor variable
Y_1	X_1
Y_2	X_2
$\vdots$	$\vdots$
Y_n	X_n

Proposition A. Under the SLR assumptions, find estimators for β_0 and β_1 such that they minimize the sum of squared vertical deviations:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2.$$



$\beta_0 + \beta_1 x$

$Y_i - \beta_0 - \beta_1 x_i$

Solution:

$$\begin{cases} \frac{\partial S}{\beta_0} = \sum_{i=1}^{n} -2(Y_i - \beta_0 - \beta_1 x_i) = 2\left[\sum_{i=1}^{n} Y_i - n\beta_0 - \beta_1\sum_i x_i\right] = 0 \\ \frac{\partial S}{\partial\beta_1} = \sum_{i=1}^{n} -2x_i(Y_i - \beta_0 - \beta_1 x_i) = 2\left[\sum_{i=1}^{n} x_iY_i - \beta_0\sum_{i=1}^{n} x_i - \beta_1\sum_{i=1}^{n} x_i^2\right] = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum_i x_i \sum_i Y_i - n\beta_0\sum_{i=1}^{n} x_i - \beta_1\left(\sum_{i=1}^{n} x_i\right)^2 = 0 \\ n\sum_{i=1}^{n} x_iY_i - n\beta_0\sum_{i=1}^{n} x_i - n\beta_1\sum_{i=1}^{n} x_i^2 = 0 \end{cases}$$

Subtract

$$\Rightarrow \hat{\beta}_1 = \frac{n\sum_i x_iY_i - \sum_i x_i \sum_i Y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \boxed{\frac{\sum_i (x_i-\bar{x}_n)(Y_i-\bar{Y}_n)}{\sum_i (x_i-\bar{x}_n)^2}}$$

$$\hat{\beta}_0 = \boxed{\overline{Y}_n - \hat{\beta}_1 \overline{X}_n}$$

$$= \overline{Y}_n - \overline{X}_n \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)}{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}$$

$$= \boxed{\frac{\sum_i X_i^2 \sum_i Y_i - \sum_i X_i \sum_i X_i Y_i}{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}}$$

Maximum likelihood estimation

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

Proposition B. Under the SLR assumptions, calculate $\sup_{\Theta} L(\beta_0,\ \beta_1,\ \sigma^2)$ and find MLEs of β_0, β_1 and σ^2.

Solution: $Y_i \mid \beta_0, \beta_1, \sigma^2 \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$

$\Theta = \{\beta_0 \in \mathbb{R}, \beta_1 \in \mathbb{R}, \sigma^2 > 0\}$, $\quad L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}}$,

$\dim \Theta = 3$.

Response variable	Predictor variable
Y_1	X_1
Y_2	X_2
$\vdots$	$\vdots$
Y_n	X_n

$$\ell(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2.$$

minimize $S(\beta_0, \beta_1)$

$$\begin{cases} \hat{\beta}_0 = \overline{Y}_n - \hat{\beta}_1 \overline{X}_n \\ \hat{\beta}_1 = \dfrac{\sum_{i=1}^n (Y_i - \overline{Y}_n)(X_i - \overline{X}_n)}{\sum_{i=1}^n (X_i - \overline{X}_n)^2} \end{cases}$$

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{n}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$$

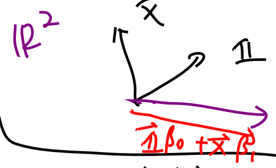
$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.$$

Geometric approach

$$Y_1 = x_1\beta_1 + \varepsilon_1$$
$$Y_2 = x_2\beta_1 + \varepsilon_2$$

$\vec{a} \perp \vec{b} \rightarrow \vec{a}^T\vec{b} = 0$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\beta_0 + \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\beta_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$



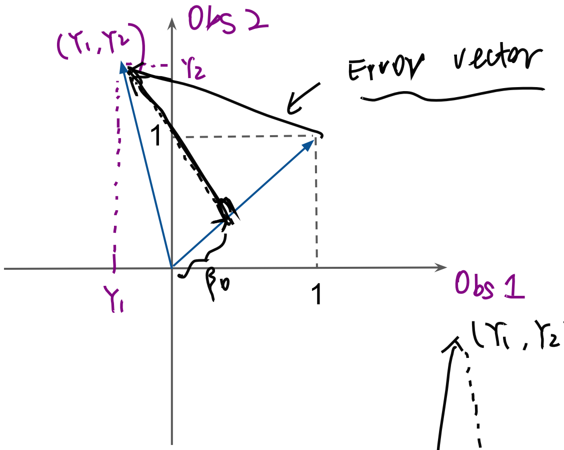
$$(1,1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2$$

We start with $n=2$:

$$Y_1 = \beta_0 + \epsilon_1$$
$$Y_2 = \beta_0 + \epsilon_2$$

$$\rightarrow \underbrace{\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}}_{\vec{Y}} = \underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\beta_0}_{\mathbb{1}_2 \beta_0} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

Find the best value of β_0.

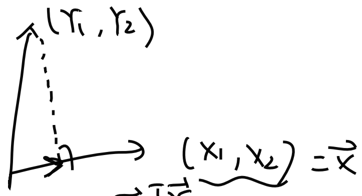


To make perpendicular error vector:

$$(\vec{Y} - \mathbb{1}\beta_0)^T \mathbb{1} = 0$$

$$\vec{Y}^T\mathbb{1} - \beta_0 \underbrace{\mathbb{1}^T\mathbb{1}}_{2} = 0$$

$$\Rightarrow \beta_0 = \frac{\vec{Y}^T\mathbb{1}}{2} = \frac{\mathbb{1}^T\vec{Y}}{2} = \frac{\mathbb{1}^T\vec{Y}}{\mathbb{1}^T\mathbb{1}}$$



$$\Rightarrow \hat{\beta}_1 = \frac{\vec{X}^T\vec{Y}}{\vec{X}^T\vec{X}}$$

$X = (\vec{v}_1, \dots \vec{v}_k)$

Geometric approach

$\vec{Y} \perp \mathcal{L} \iff \vec{Y} \perp \vec{v}$, for all $\vec{v} \perp \mathcal{L}$

linear subspace

$\iff \vec{Y} \perp (\vec{v}_1, \vec{v}_2, \dots \vec{v}_m)$ basis for $\mathcal{L}$

We now look at $n=3$:

X^TX invertible $\iff$ rank$(X) = k$

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\
Y_3 &= \beta_0 + \beta_1 X_3 + \epsilon_3
\end{aligned}
$$

Find the best values of β_0 and β_1.

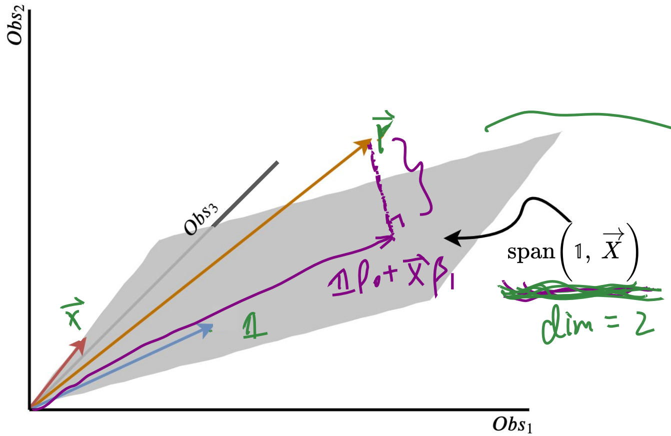
$$
\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}}_{\vec{Y}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \beta_1 + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}
$$

$$\mathbb{1}\beta_0 + \vec{X}\beta_1$$

$$= (\mathbb{1}, \vec{X}) \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$= X \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

where $X = (\mathbb{1}, \vec{X}) = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{pmatrix}$



$$\boxed{X^T \left[\vec{Y} - X \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \right] = 0}$$

$$X^T\vec{Y} - X^TX \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = 0 \Rightarrow \boxed{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = (\underbrace{X^TX}_{2\times 2})^{-1} X^T \vec{Y}}$$

Geometric approach

Proposition C. We generalize to any $n \geq 2$:

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n
\end{aligned}
$$

Find the best values of β_0 and β_1.

$\mathbb{R}^n$

$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$ $\quad \mathrm{rank}(X) = 2$

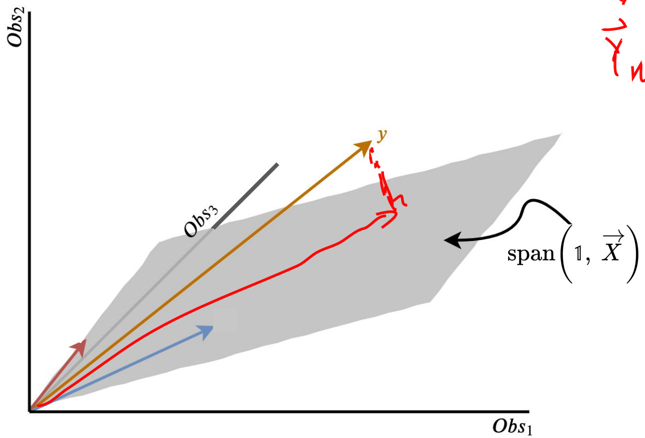
$$\vec{Y}_n = \vec{\mathbb{1}}_n \beta_0 + \vec{X} \beta_1 + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Find ~~vector~~ $\mathbb{1}_n\beta_0 + \vec{X}\beta_1$ in $\mathrm{span}(\mathbb{1}, \vec{X})$ such that

$$X^T(\vec{Y} - \mathbb{1}\beta_0 - \vec{X}\beta_1)$$

$$= X^T(\vec{Y} - X\vec{\beta}) = 0$$

$$\Rightarrow \vec{\beta} = (X^TX)^{-1} X^T \vec{Y}$$



Geometric approach

$\hat\beta_0 = \bar Y_n - \bar X_n \hat\beta_1$

$\hat\beta_1 = \dfrac{\sum_i (X_i - \bar X_n)(Y_i - \bar Y_n)}{\sum_i (X_i - \bar X_n)^2}$

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \dfrac{1}{ad-bc}$

Corollary C. Method of least squares and the geometric approach give the same estimators:

$$\begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

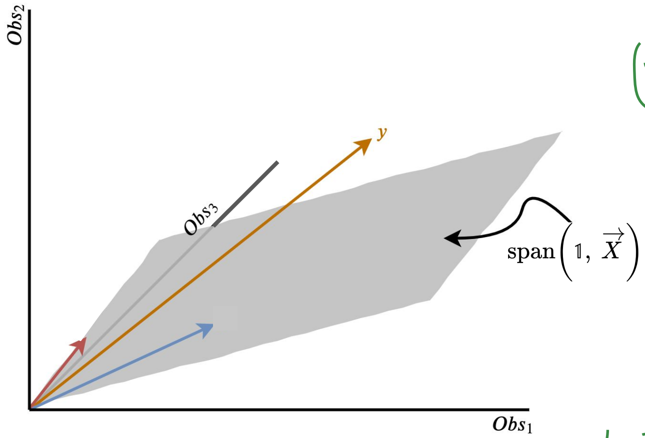
Proof.

$X^TX = \begin{pmatrix} 1, & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}$

$(X^TX)^{-1} = \begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix} \dfrac{1}{n\sum_i x_i^2 - (\sum_i x_i)^2}.$

$X^TY = \begin{pmatrix} 1 & \cdots & 1 \\ x_1, & \cdots & x_n \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum_i Y_i \\ \sum_i x_i Y_i \end{pmatrix}.$

$(X^TX)^{-1}X^TY = \dfrac{1}{n\sum_i x_i^2 - (\sum_i x_i)^2} \begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix} \begin{pmatrix} \sum_i Y_i \\ \sum_i x_i Y_i \end{pmatrix}$



$$= \begin{pmatrix} \dfrac{\sum_i x_i^2 \sum y_i - \sum_i x_i \sum_i x_i y_i}{n \sum_i x_i^2 - \left(\sum_i x_i\right)^2} \\ \\ \dfrac{-\sum_i x_i \sum_i y_i + n \sum x_i y_i}{n \sum x_i^2 - \left(\sum_i x_i\right)^2} \end{pmatrix}$$

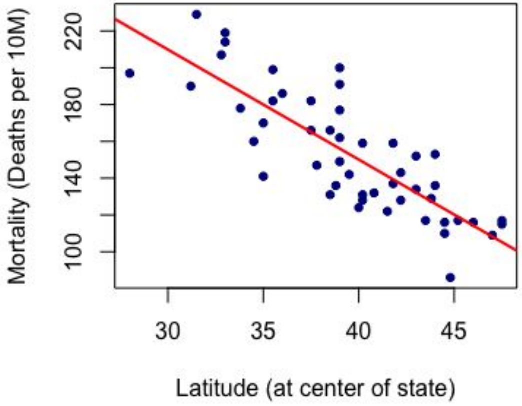
$n \sum_i (x_i - \bar{x}_n)(y_i - \bar{y}_n)$

$n \sum_i (x_i - \bar{x}_n)^2$

$$= \begin{pmatrix} \hat{\beta}_0 \\ \\ \hat{\beta}_1 \end{pmatrix}.$$

Calculate in R

Example 2. During the 50s, data were collected to examine the relationship between the mortality rate due to skin cancer (number of deaths per 10 million people) and the latitude at the center of each of 48 states in the United States (Alaska and Hawaii were not yet states. And, Washington, D.C. was included in the data set even though it is not technically a state.)



$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_n)(X_i - \bar{X}_n)}{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$ ← cov$(\vec{Y}, \vec{X})$ ← var$(\vec{X})$

$$\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{X}_n = \frac{\sum_i Y_i \sum_i X_i^2 - \sum_i X_i \sum_i X_i Y_i}{n\sum_i X_i^2 - (\sum_i X_i)^2}$$

Mort Lat

```
> beta1 <- cov(Mort, Lat)/var(Lat)
> beta0 <- mean(Mort)-beta1*mean(Lat)
> c('beta0' = beta0, 'beta1' = beta1)
     beta0      beta1
389.189351  -5.977636
```

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$X = (1, \vec{X})$

```
> X <- cbind(1,Lat)
> y <- Mort
> solve(t(X)%*%X)%*%t(X)%*%y
          [,1]
     389.189351
Lat   -5.977636
```

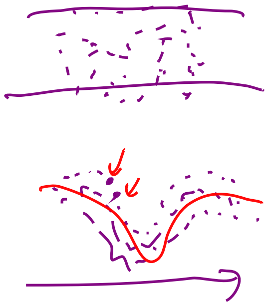
$(X^TX)^{-1}X^TY$

solve$(X) = X^{-1}$

Diagnostics

Model assumption ($i = 1, \cdots, n$):

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0, \sigma^2)} . \quad \Rightarrow \quad \text{Residuals } \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$



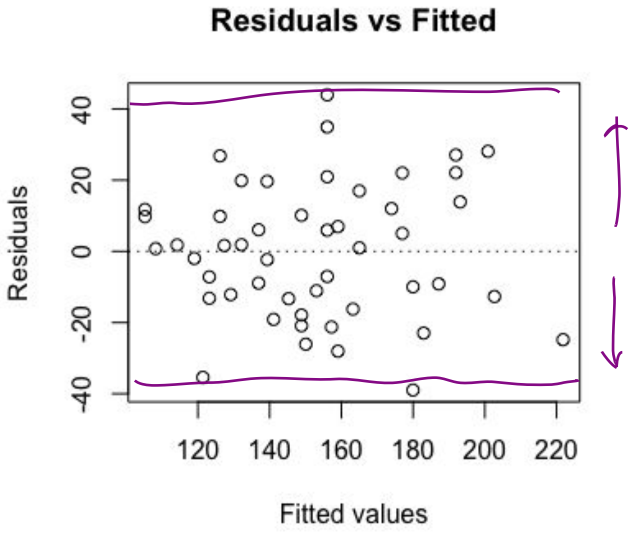
1. Linearity - Plotting Y_i vs X_i
2. Normality - QQ plot
3. Zero mean in error terms
4. Homoscedasticity
5. Independence

$\left.\begin{matrix} 3. \\ 4. \\ 5. \end{matrix}\right\}$ Residual plot

```
Fitted <- beta0 + beta1*Lat
Res <- y - beta0 - beta1*Lat

plot(Fitted, Res, xlab='Fitted values', ylab='Residuals',
     main='Residuals vs Fitted')
abline(h=0, lty=3)
```

Lat

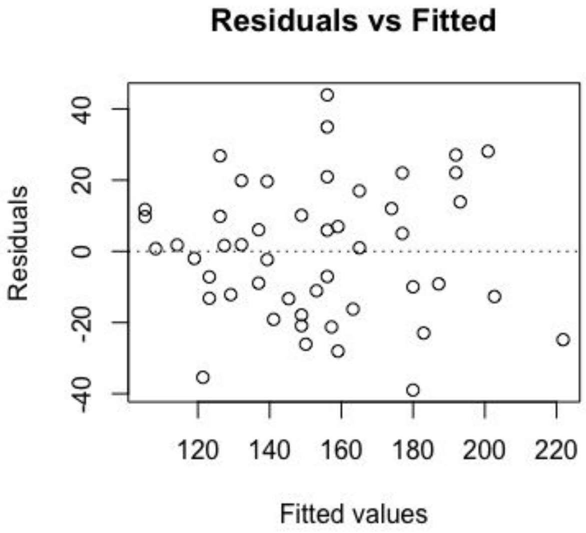


Diagnostics

Model assumption $(i = 1, \cdots, n)$:

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0,\sigma^2)} . \quad \Rightarrow \quad \text{Residuals } \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

1. Residuals are pretty **symmetrically** distributed, tending to cluster towards the middle of the plot.
2. There are no clear **patterns**.

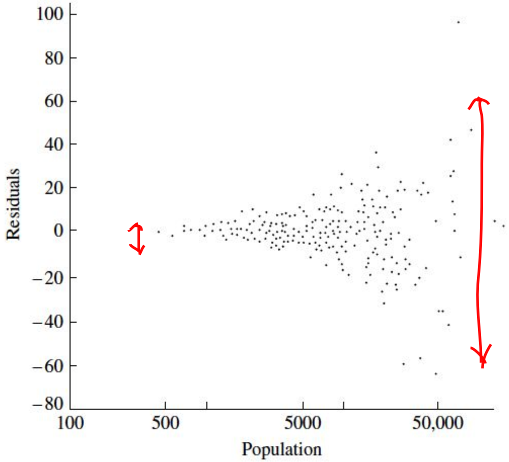


Diagnostics

$\sqrt{Y_i} = \beta_0 + \beta_1 x_i + \varepsilon_i$

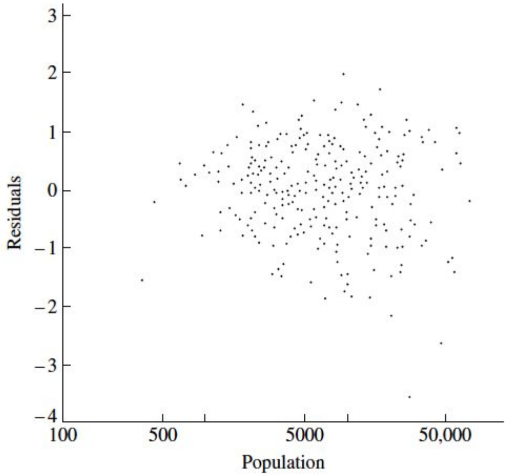
Model assumption $(i = 1, \cdots, n)$:

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0,\sigma^2)} . \quad \Rightarrow \quad \text{Residuals } \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$



Heteroscedasticity:
Non-constant error variance

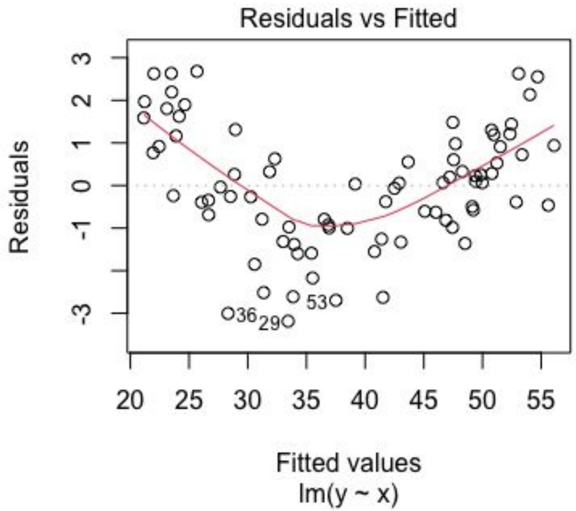
Remedial actions:
- ➔ $\sqrt{Y}$
- ➔ $\log(Y)$



Diagnostics

Model assumption ($i = 1, \cdots, n$):

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0,\sigma^2)} . \quad \Rightarrow \quad \text{Residuals } \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

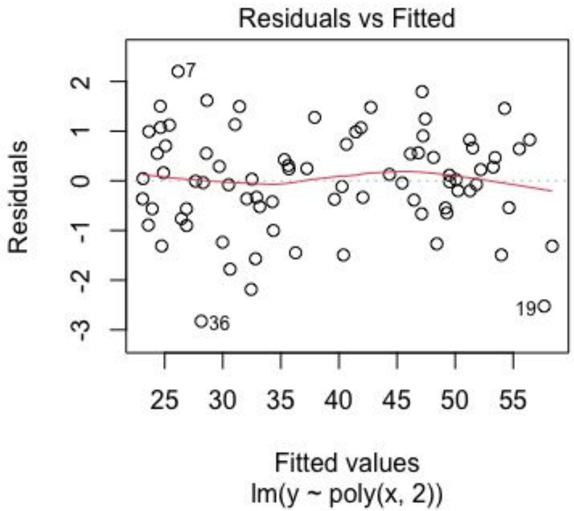


Non-constant mean +
Non-independence

Remedial actions:

➔ Include polynomial powers of X_i

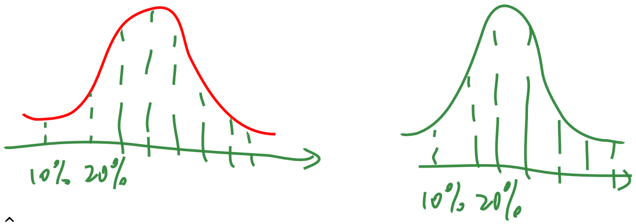
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$



Diagnostics

Model assumption ($i = 1, \cdots, n$):

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0,\sigma^2)} . \quad \Rightarrow \quad \text{Residuals } \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

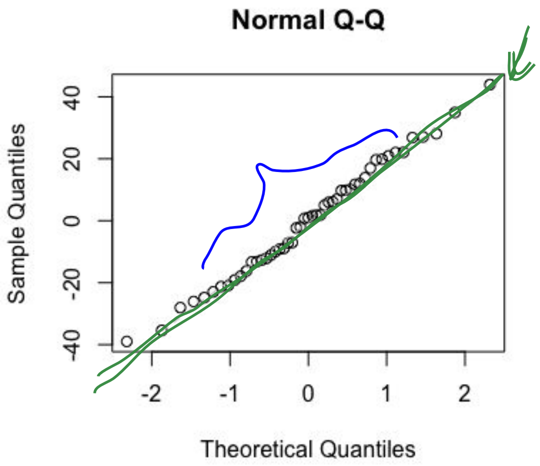


Comparing the sample quantiles of the residuals with theoretical quantiles

1. Linearity - Plotting Y_i vs X_i
2. Normality - QQ plot
3. Zero mean in error terms
4. Homoscedasticity
5. Independence

(items 3–5) } Residual plot

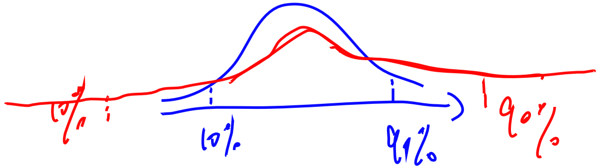
```
qqnorm(Res, main = 'Normal Q-Q')
qqline(Res, lty=2, col='gray')
```

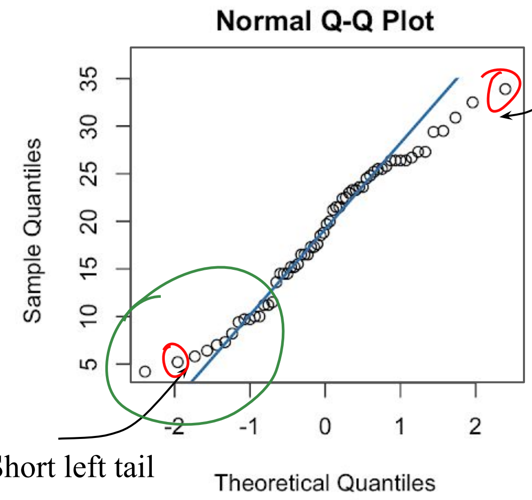


Diagnostics

Model assumption $(i = 1, \cdots, n)$:

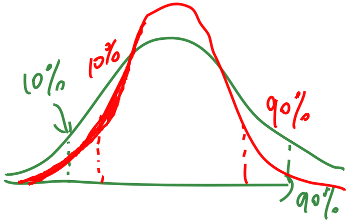
$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0,\sigma^2)} . \quad \Rightarrow \quad \text{Residuals } \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

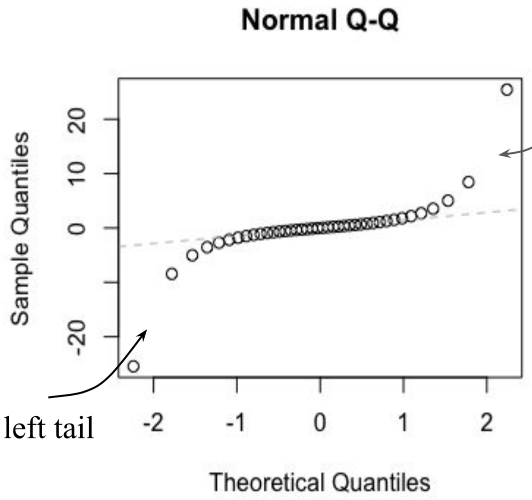




Short right tail

Short left tail





Fat right tail

Fat left tail

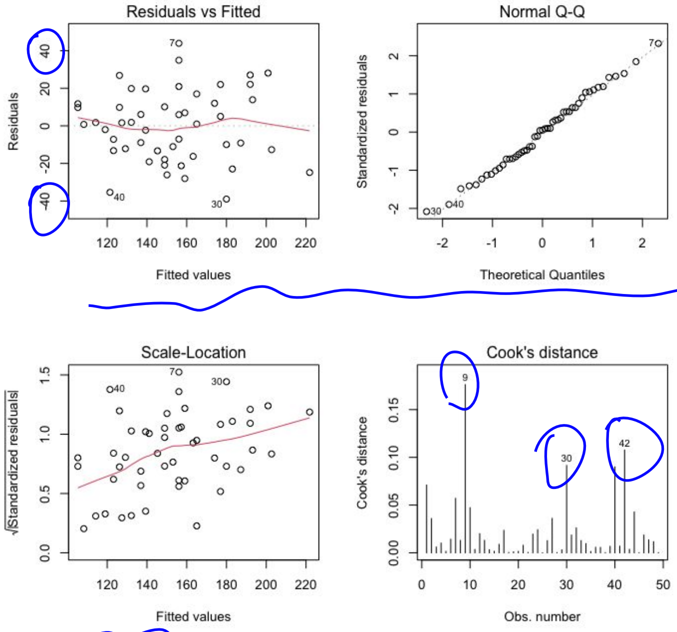
Fit & Diagnostics in R

Model assumption $(i = 1, \cdots, n)$:

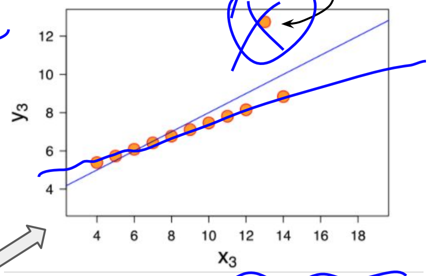
$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\epsilon_i}_{\overset{\text{iid}}{\sim} N(0,\sigma^2)} .$$

Residuals $\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$

```
> fit <- lm(Mort ~ Lat)
> fit$coefficients
(Intercept)         Lat
 389.189351   -5.977636
>
> par(mfrow=c(2,2))
> plot(fit, which = 1)
> plot(fit, which = 2)
> plot(fit, which = 3)
> plot(fit, which = 4)
> par(mfrow=c(1,1))
```



Influential point



Sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$

14.2 of Rice

08/03/2021

Mean and variance

Proposition D. Under the SLR assumptions,

$$E\left(\hat{\beta}_0\right) = \beta_0, \text{ var } \left(\hat{\beta}_0\right) = \frac{n^{-1}\sum_i X_i^2}{\sum_i \left(X_i - \bar{X}_n\right)^2}\sigma^2,$$

$$E\left(\hat{\beta}_1\right) = \beta_1, \text{ var } \left(\hat{\beta}_1\right) = \frac{1}{\sum_i \left(X_i - \bar{X}_n\right)^2}\sigma^2,$$

$$\text{cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = \frac{-\bar{X}_n}{\sum_i \left(X_i - \bar{X}_n\right)^2}\sigma^2.$$

Sampling distribution

$\varepsilon_i \sim N(\ )$

Theorem D. Under the SLR assumptions,

$$\frac{\hat{\beta}_k - \beta_k}{\text{se}\left(\hat{\beta}_k\right)} \sim t_{n-2},\ k = 0, 1.$$

$\hat{\beta}_0 \quad \hat{\beta}_1 \sim N(\ )$

$$95\% \text{ exact confidence interval for } \beta_k:$$

$$\hat{\beta}_k \pm t_{n-2}(\alpha/2) \cdot \text{se}\left(\hat{\beta}_k\right)$$

$$\text{RSS} = \sum_{i=1}^{n} \hat{\epsilon}_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 \leftarrow \chi^2_{n-2}$$

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

$$\text{se}\left(\hat{\beta}_0\right) = \frac{n^{-1}\sum_i X_i^2}{\sum_i \left(X_i - \bar{X}_n\right)^2}\hat{\sigma}^2$$

$$\text{se}\left(\hat{\beta}_1\right) = \frac{1}{\sum_i \left(X_i - \bar{X}_n\right)^2}\hat{\sigma}^2$$

Tomorrow ...

Multiple linear regression

- Generalized the SLR results
- Implement in R