Estimating σ^2

- Calculate sum of square of residuals
  - Residuals = difference between actual y_i and predicted value $\mathrm{B}x_i + A$

$$\boxed{SS_R = \sum_{i=1}^{n}(Y_i - A - Bx_i)^2}$$

Example $n = 2$

$\hat{\sigma}_{MLE} = 0$

- The MLE estimate would be:

$$\hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^{n} \left(Y_i - (\overbrace{A + Bx_i}^{\hat{\mu}_{x_i}}) \right)^2}{n}$$

$$\hat{\sigma}_{MLE} = \frac{SSR}{n}$$

- The above is biased like for normal Gaussian parameters.
- We will use a different method:

The Chi-Square distribution (Section 5.8 of textbook)

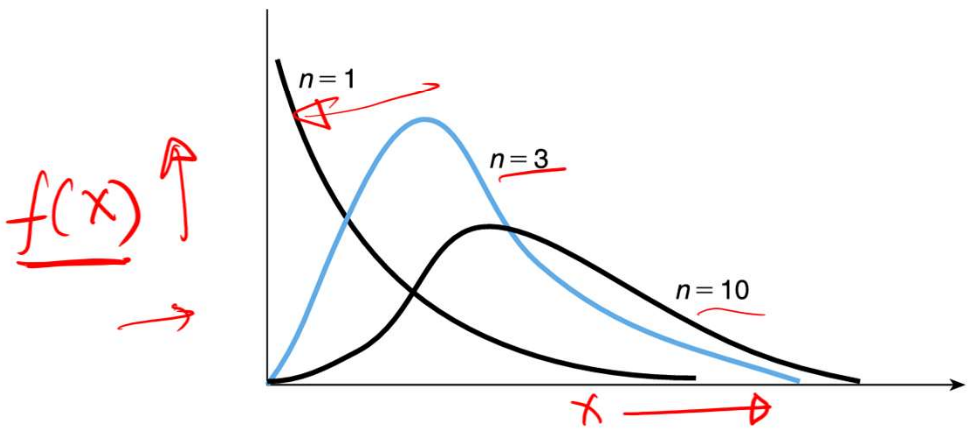
Definition. If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then X, defined by

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \tag{5.8.1}$$

is said to have a *chi-square distribution with n degrees of freedom*. We will use the notation

$$X \sim \chi_n^2$$

to signify that X has a chi-square distribution with n degrees of freedom.



Deriving the density of χ_n^2 distribution

- Use MGF.
- Consider n=1 first.

MGF of x

$$E[e^{tX}] = E[e^{tZ^2}] \text{ where } Z \sim \mathcal{N}(0,1)$$

$$= \int_{-\infty}^{\infty} e^{tx^2} f_Z(x)\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} \boxed{e^{-x^2/2}\,dx}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2(1-2t)/2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\bar{\sigma}^2}\,dx \quad \text{where } \bar{\sigma}^2 = (1-2t)^{-1}$$

$$= (1-2t)^{-1/2} \frac{1}{\sqrt{2\pi}\,\bar{\sigma}} \int_{-\infty}^{\infty} e^{-x^2/2\bar{\sigma}^2}\,dx$$

$$= (1-2t)^{-1/2}$$

General n

- $E_X[e^{\{tX\}}] = E\left[e^{t\sum_i Z_i^2}\right] = \prod_i E\left[e^{tZ_i^2}\right] = (1-2t)^{-n/2}$

$\alpha =$ $\lambda =$

- The above is MGF of gamma distribution with parameters (n/2,1/2).

A random variable is said to have a gamma distribution with parameters (α, λ), $\lambda > 0$, $\alpha > 0$, if its density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Thus, density of χ^2 distribution is

- $$\boxed{f(x) = \frac{\frac{1}{2}e^{-x/2}\left(\frac{x}{2}\right)^{(n/2)-1}}{\Gamma\left(\frac{n}{2}\right)}, \quad x > 0}$$

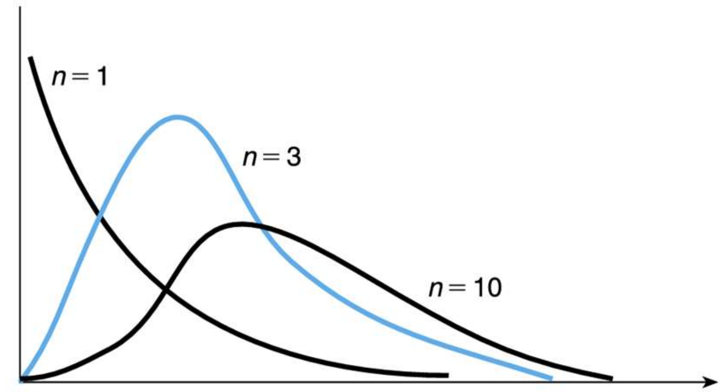
Expected value of χ_n^2 distribution [Homework]

- $E[\chi_n^2]$ = n [Can be derived from the MGF]

$\frac{\partial}{\partial t} MGF(\chi_n^2)$

- $Var[\chi_n^2]$ = 2n

- Mode = max(n-2, 0)



Estimating σ^2

- Calculate sum of square of residuals where
  - Residuals = difference between actual y_i and predicted value $\mathrm{B}x_i + A$

$$SS_R = \sum_{i=1}^{n} (Y_i - A - Bx_i)^2$$

$N(0, 1)$

$Y_i \sim N(\alpha + \beta x_i; \sigma^2)$

$$\boxed{\sum_{i=1}^{n} \frac{(Y_i - \beta x_i - \alpha)^2}{\sigma^2}} \sim \chi^2_n$$

$Z_i = \frac{Y_i - (\alpha + \beta x_i)}{\sigma}$

$Z_i \sim N(0, 1)$

- It can be show that $\boxed{\frac{SS_R}{\sigma^2}}$ follows a Chi-square distribution with n-2 degrees of freedom
  - Book has a kind of intuitive proof…

A, & B are functions of Y_i

$\therefore$ each of the n terms in SSR are not independent of each other.

Estimating σ^2

Let estimate of σ^2 be called S.

$$\hat{\sigma}^2 = S = \boxed{\frac{SS_R}{n-2}}$$

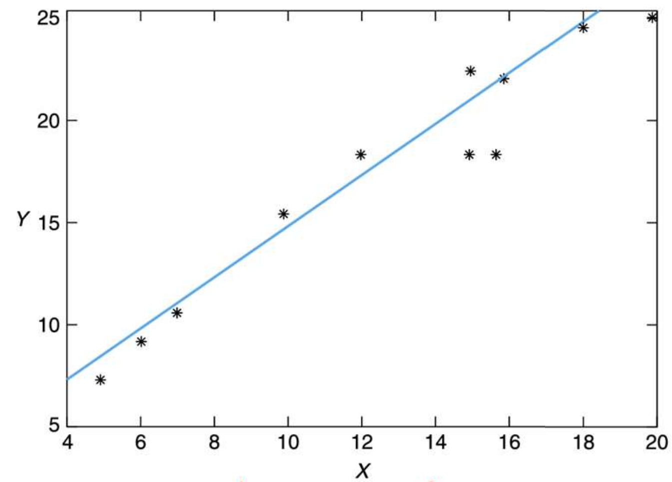
S is an unbiased estimate of σ^2. It is easy to see that $E[S] = \sigma^2$

Example 9.3.a. The following data relate x, the moisture of a wet mix of a certain product, to Y, the density of the finished product.

x_i	y_i	$x_i y_i$	x_i^2
5	7.4	37	25
6	9.3		
7	10.6		
10	15.4		
12	18.1		
15	22.2		
18	24.1		
20	24.8		

SSR

$\bar{x}$ $\frac{\sum x_i}{9}$ $\bar{Y}$ Σ Σ



% transition

Compute the least square fit. Estimate A, B, S

$$B = \frac{\sum_i x_i y_i - n\bar{x}\bar{Y}}{\sum_i x_i^2 - n\bar{x}^2}, \quad A$$

A B

$$y = 2.463 + 1.206x$$

Multi-variable linear regression

Reading material: Section 9.10 of Ross Textbook

General case: k > 1

$$f(Y \mid x_1, \ldots, x_k) \sim N(\mu_x, \sigma^2), \quad where \ \mu_x = \beta_1 x_1 + \cdots + \beta_k x_k + \beta_0$$

α

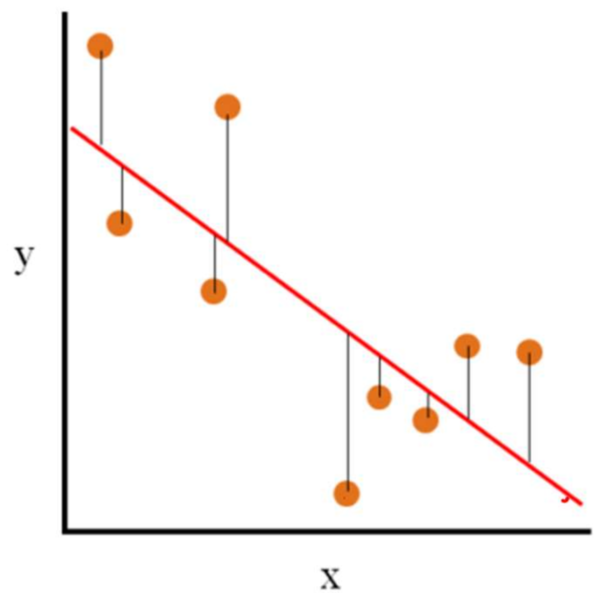
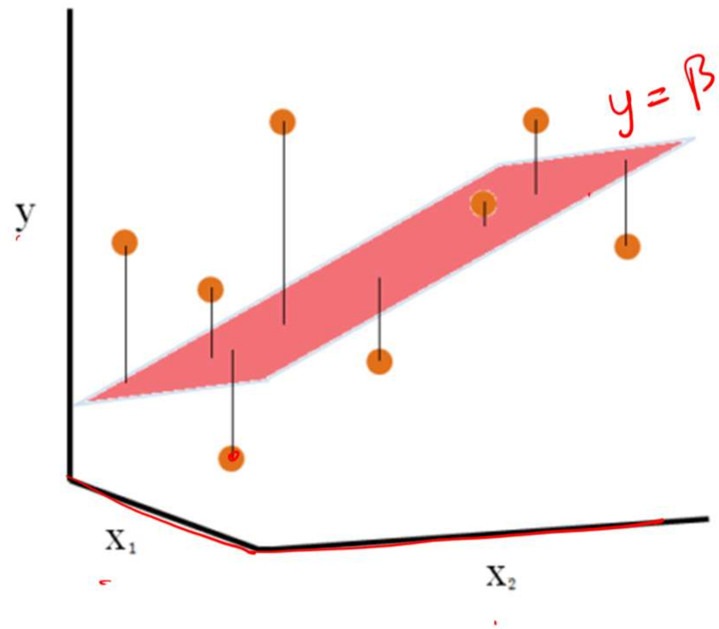
Or

$$Y = \beta_1 x_1 + \cdots + \beta_k x_k + \beta_0 + e \quad \text{where } e \sim N(0, \sigma^2)$$

Training data D will be denoted as

$$\{(x_{i1}, x_{i2}, \ldots x_{ik}, y_i) : i = 1 \ldots n\}$$

$$\{(x_i, \quad y_i) : i = 1 \ldots n\}$$

Simple Linear Regression	Multiple Linear Regression (2 Independent Variables (x_1, x_2))
	

$y = \beta_1 x_1 + \beta_2 x_2 + \beta_0$

Parameter estimation using MLE

$$LL(D) = \sum_{i=1}^{n} \log e^{-\frac{(Y_i - (B_1 x_1 + \cdots B_k x_k + B_0))^2}{\sigma^2}} - n \log [2\pi\sigma]$$

$$\frac{\partial LL}{\partial B_0} = 0 \qquad \sum_{i=1}^{n} (y_i - (B_1 x_1 + \cdots B_k x_k + B_0))(-1) = 0$$

$$\frac{\partial LL}{\partial B_1} = 0 \qquad \sum_{i=1}^{n} (y_i - (B_1 x_1 + \cdots B_k x_k + B_0)) x_1 = 0$$

$\vdots$

$$\frac{\partial LL}{\partial B_k} = 0 \qquad \sum_{i=1}^{n} (y_i - (B_1 x_1 + \cdots B_k x_k + B_0)) x_k = 0$$

Solving the MLE

$$\sum_{i=1}^{n} Y_i = nB_0 + B_1 \sum_{i=1}^{n} x_{i1} + B_2 \sum_{i=1}^{n} x_{i2} + \cdots + B_k \sum_{i=1}^{n} x_{ik} \tag{9.10.1}$$

$$\sum_{i=1}^{n} x_{i1} Y_i = B_0 \sum_{i=1}^{n} x_{i1} + B_1 \sum_{i=1}^{n} x_{i1}^2 + B_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \cdots + B_k \sum_{i=1}^{n} x_{i1} x_{ik}$$

$$\vdots$$

$$\sum_{i=1}^{k} x_{ik} Y_i = B_0 \sum_{i=1}^{n} x_{ik} + B_1 \sum_{i=1}^{n} x_{ik} x_{i1} + B_2 \sum_{i=1}^{n} x_{ik} x_{i2} + \cdots + B_k \sum_{i=1}^{n} x_{ik}^2$$

Matrix notation for k-dimensional covariates.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \rightarrow (1 \ \bar{x}_1)$$

$(n+1)\times 1$

$X \equiv n \times (k+1)$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$(k+1)\times 1$

then $\mathbf{Y}$ is an $n \times 1$, $\mathbf{X}$ an $n \times p$, $\boldsymbol{\beta}$ a $p \times 1$, and $\mathbf{e}$ an $n \times 1$ matrix where $p \equiv k + 1$.

The regression equation on this data becomes: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$X' =$ transpose of X

$$= \begin{bmatrix} n & \sum_i x_{i1} & \sum_i x_{i2} & \cdots & \sum_i x_{ik} \\ \sum_i x_{i1} & \sum_i x_{i1}^2 & \sum_i x_{i1}x_{i2} & \cdots & \sum_i x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_i x_{ik} & \sum_i x_{ik}x_{i1} & \sum_i x_{ik}x_{i2} & \cdots & \sum_i x_{ik}^2 \end{bmatrix}$$

(r, c)

$\sum_{i=1}^{n} x_{ir} x_{ic}$

Covariance matrix if $\bar{x} = 0$

and

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum_i Y_i \\ \sum_i x_{i1}Y_i \\ \vdots \\ \sum_i x_{ik}Y_i \end{bmatrix}$$

Solving the MLE

$$\sum_{i=1}^{n} Y_i = nB_0 + B_1 \sum_{i=1}^{n} x_{i1} + B_2 \sum_{i=1}^{n} x_{i2} + \cdots + B_k \sum_{i=1}^{n} x_{ik} \tag{9.10.1}$$

$$\sum_{i=1}^{n} x_{i1} Y_i = B_0 \sum_{i=1}^{n} x_{i1} + B_1 \sum_{i=1}^{n} x_{i1}^2 + B_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \cdots + B_k \sum_{i=1}^{n} x_{i1} x_{ik}$$

$$\vdots$$

$$\sum_{i=1}^{k} x_{ik} Y_i = B_0 \sum_{i=1}^{n} x_{ik} + B_1 \sum_{i=1}^{n} x_{ik} x_{i1} + B_2 \sum_{i=1}^{n} x_{ik} x_{i2} + \cdots + B_k \sum_{i=1}^{n} x_{ik}^2$$

$$\mathbf{X'XB} = \mathbf{X'Y}$$

where $\mathbf{X}'$ is the transpose of $\mathbf{X}$.

$$\mathbf{B} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

Example

Get least square estimate on this data with last column as y.

Table 9.5

	Age (years)	Elevation (1000 ft)	Rain-fall(inches)	Specific Gravity	Diameter at Breast Height (inches)
1	44	1.3	250	.63	18.1
2	33	2.2	115	.59	19.6
3	33	2.2	75	.56	16.6
4	32	2.6	85	.55	16.4
5	34	2.0	100	.54	16.9
6	31	1.8	75	.59	17.0
7	33	2.2	85	.56	20.0
8	30	3.6	75	.46	16.6
9	34	1.6	225	.63	16.2
10	34	1.5	250	.60	18.5
11	33	2.2	255	.63	18.7
12	36	1.7	175	.58	19.4
13	33	2.2	75	.55	17.6
14	34	1.3	85	.57	18.3
15	37	2.6	90	.62	18.8

$k = 4$

$$y = 11.54873 + 0.05728\,x_1 + 0.08712\,x_2 + 7.33231\,x_3$$

Evaluating goodness of a fit: The coefficient of determination

- Measure amount of variation in the data:

$$S_{YY} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

- Measure the amount of variation in the residual after a model is fit

$$SS_R = \sum_{i=1}^{n} (Y_i - A - Bx_i)^2 \quad \sum_{i=1}^{n} (Y_i - BX_i)^2 \; : \; \hat{Y}_i = BX_i$$

$$SSR = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

- The coefficient of determination

$$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} = 1 - \frac{SS_R}{S_{YY}}$$

If $\forall i \; \hat{Y}_i = \overline{Y}$ what is $R^2 = 0$

If $\forall i \; \hat{Y}_i = Y_i$ $\quad R^2 = 1$

$$0 \le R^2 \le 1$$

Demo

https://colab.research.google.com/github/rafiag/DTI2020/blob/main/0
02a_Multi_Linear_Regression_(EN).ipynb