Simple linear regression.

Model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, 2, \cdots n.$

Data : $\{(y_i, x_i),\ i = 1, 2, \cdots n\}$.

Assumptions:
1. x_i's are non-stochastic.
2. y_i's are stochastic.
3. $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.
4. $\beta_0, \beta_1, \sigma^2$ are unknown.

Problem:
1. Estimation of $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$
2. Prediction of y values for some new x values
3. Doing testing for $\beta_0, \beta_1, \sigma^2$

Matrix representation.

$$\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \underset{\sim}{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad \underset{\sim}{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Model: $\underset{\sim}{y} = [\underset{\sim}{1}\ \underset{\sim}{x}]\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$.

$\underset{\sim}{\epsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$.

Let $\underset{n \times 2}{[\underset{\sim}{1}\ \underset{\sim}{x}]} = X \qquad \underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n) \Rightarrow$ Y_i's are independent but not identically distributed.

$E(y_i) = \beta_0 + \beta_1 x_i$

Least square estimation of $(\beta_0, \beta_1) = \underset{\sim}{\beta}$

Least square condition: $$S(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \quad \text{--- (1)}$$

$$= (\underset{\sim}{Y} - X\underset{\sim}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim}{\beta}) \quad \text{--- (2)}$$

$S(\underset{\sim}{\beta})$ has to be minimize w.r.t. $(\beta_0, \beta_1) = \underset{\sim}{\beta}$.

$$\left.\frac{\partial S(\beta)}{\partial \beta_0}\right|_{\hat\beta_0} = 0, \quad \left.\frac{\partial S(\beta)}{\partial \beta_1}\right|_{\hat\beta_1} = 0 \quad \text{solve}$$

$$S_{xy} = \sum_{i=1}^{n}(y_i - \bar y)(x_i - \bar x)$$

HW
$$\hat\beta_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat\beta_0 = \bar y - \hat\beta_1 \bar x$$

$$S_{xy} = \sum_{i=1}^{n}(y_i - \bar y)x_i = \sum_{i=1}^{n}(x_i - \bar x)y_i = \sum_{i=1}^{n}(x_i - \bar x)(y_i - \bar y)$$

$$\sum (x_i - \bar x)(y_i - \bar y) = \sum_{i=1}^{n}(x_i - \bar x)y_i - \bar y \underbrace{\sum_{i=1}^{n}(x_i - \bar x)}_{0}$$

$$\left.\frac{\partial S(\underset{\sim}{\beta})}{\partial \underset{\sim}{\beta}}\right|_{\hat\beta} = \underset{\sim}{0} \implies \text{Normal equations.}$$

$$X^T y = X^T X \hat\beta$$

$$\boxed{\hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{y}}$$

Estimator of (β_0, β_1).

$$\frac{\partial(\underset{\sim}{b}^T A)}{\partial \underset{\sim}{b}} = A^T$$

$$\frac{\partial\, \underset{\sim}{b}^T A \underset{\sim}{b}}{\partial \underset{\sim}{b}} = b^T(A + A^T)$$

$$X = [\underset{\sim}{1}\ \underset{\sim}{x}]$$

we need to know the
→ distribution of the estimator
→ properties of the estimator.

$$\hat\beta_1 = \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^{n}\frac{(x_i - \bar x)y_i}{S_{xx}} = \sum_{i=1}^{n}\left[\frac{(x_i - \bar x)}{S_{xx}}\right] y_i$$

let $c_i = \dfrac{(x_i - \bar x)}{S_{xx}}$ and $\underset{\sim}{c} = (c_1\, c_2 \cdots c_n)^T$

$\Rightarrow \hat\beta_1 = \underset{\sim}{c}^T \underset{\sim}{y}$. also we know $\underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$.

$$\hat\beta_1 \sim N\left(\underset{\sim}{c}^T X \underset{\sim}{\beta},\ \underset{\sim}{c}^T(\sigma^2 I_n)\underset{\sim}{c}\right) = N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right).$$

$$\underset{\sim}{c}^T X \underset{\sim}{\beta} = [\underset{\sim}{c}^T \underset{\sim}{1},\ \underset{\sim}{c}^T \underset{\sim}{x}]\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} = (0\ \ 1)\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} = \beta_1$$

$$\underset{\sim}{c}^T \sigma^2 I_n \underset{\sim}{c} = \sigma^2 \underset{\sim}{c}^T \underset{\sim}{c} = \sigma^2 \frac{\sum_{i=1}^{n}(x_i - \bar x)^2}{S_{xx}^2} = \frac{\sigma^2}{S_{xx}}.$$

Linear estimator. If an estimator $T(\underset{\sim}{Y})$ can be expressed as a linear combination of $\underset{\sim}{Y}$ then it is called a linear estimator.

i.e. $T(\underset{\sim}{Y}) = \sum_{i=1}^{n} \alpha_i y_i = \underset{\sim}{\alpha}^T \underset{\sim}{Y}$

1. $\hat{\beta}_1$ is a linear estimator of β_1
2. $\hat{\beta}_1$ is an unbiased estimator of β_1
3. $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$.

HW.

1. $\hat{\beta}_0$ is a linear estimator.
2. $\hat{\beta}_0$ is an unbiased estimator of β_0.
3. $\hat{\beta}_0 \sim N\left(\hat{\beta}_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$= \bar{y} - \bar{x}\frac{S_{xy}}{S_{xx}} = \bar{y} - \bar{x}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{S_{xx}} y_i = \cancel{\sum_{i=1}^{n}\frac{1}{n}}$$

$$= \sum_{i=1}^{n}\left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{S_{xx}}\right) y_i = \sum_{i=1}^{n} b_i y_i = \underset{\sim}{b}^T \underset{\sim}{Y}.$$

$$b_i = \left(\frac{1}{n} - \frac{\bar{x}(x_i-\bar{x})}{S_{xx}}\right)$$

$\Rightarrow \hat{\beta}_0$ is a linear estimator of $\hat{\beta}$.

HW Show that the regression line passes through $(\bar{x}, \bar{y})$.

Prediction ~~lie~~ line.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \longrightarrow \text{old / new}$$

$$= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x$$

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

$\Rightarrow$ Regression line passes through $(\bar{x}, \bar{y})$.

Estimated error.

$$e_i = y_i - \hat{y}_i$$

$$= (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$i = 1, 2 \cdots n$.

$\underset{\sim}{e} = (e_1\, e_2 \cdots e_n)^T$ can be used for the estimation of σ^2.

Consider.

$$\underset{\sim}{e}^T \underset{\sim}{e} = (\underset{\sim}{y} - \underset{\sim}{\hat{y}})^T (\underset{\sim}{y} - \underset{\sim}{\hat{y}}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

$$= (\underset{\sim}{y} - X\hat{\beta})^T (\underset{\sim}{y} - X\hat{\beta}) \quad \text{as} \quad \underset{\sim}{\hat{y}} = X \underset{\sim}{\hat{\beta}} \quad \text{as } \hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y}$$

$$= X (X^T X)^{-1} X^T y$$

$$= P_X y$$

$$= (\underset{\sim}{y} - P_X \underset{\sim}{y})^T (y - P_X \underset{\sim}{y})$$

$$= \underset{\sim}{y}^T (I - P_X)^T (I - P_X) \underset{\sim}{y}$$

$$= \underset{\sim}{y}^T (I - P_X) \underset{\sim}{y}$$

square, symmetric, idempotent matrix.

Note $P_X = P_X^T$. ??

$P_X = P_X^2$.

P_X is an othogonal projection matrix $\mathcal{C}(X)$

Matrix representation.

$$\underset{\sim}{\hat{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{Y}.$$

$$\underset{\sim}{\hat{Y}} = X\hat{\beta} = X(X^T X)^{-1} X^T \underset{\sim}{Y}. = P_X \underset{\sim}{Y}$$

Let $P_x = X(X^T X)^{-1} X^T$.

hence $P_x = P_X^T$ symmetric
$P_x^2 = P_X$. idempotent.

$$\underset{\sim}{e} = Y - \hat{Y} = (I_n - P_X)\underset{\sim}{Y}. \quad \text{estimated error.}$$

$$Y = (P_x + I_n - P_x)\underset{\sim}{Y}.$$

$$\Rightarrow Y = P_X \underset{\sim}{Y} + (I_n - P_x)\underset{\sim}{Y}. = \underbrace{\underset{\sim}{\hat{Y}}}_{\text{Prediction}} + \underbrace{\underset{\sim}{e}}_{\text{estimated error.}}$$

$$Y \sim N(X\beta, I_n\sigma^2)$$

$$E(\underset{\sim}{\hat{\beta}}) = E[(X^T X)^{-1} X^T Y]$$
$$= (X^T X)^{-1} X^T E(Y)$$
$$= (X^T X)^{-1}(X^T X)\underset{\sim}{\beta}$$
$$= \underset{\sim}{\beta}. \quad \text{(unbiased estimator.)}$$

$P_X = X(X^T X)^{-1} X^T$ is the orthogonal projection matrix of $\mathcal{C}(X)$. $X = [\underset{\sim}{1}, \underset{\sim}{x}]$

$$\begin{cases} \underset{\sim}{\hat{Y}}^T \underset{\sim}{e} \\ = (P_X \underset{\sim}{Y})^T (I_n - P_x)\underset{\sim}{Y} \\ = \underset{\sim}{Y}^T P_x^T (I_n - P_x)\underset{\sim}{Y} \\ = \underset{\sim}{Y}^T \, 0 \, \underset{\sim}{Y} = 0. \end{cases}$$

$$\begin{cases} \underset{\sim}{\hat{Y}} \perp \underset{\sim}{e} \\ \underset{\sim}{\hat{Y}} \in \mathcal{C}(X) \\ \underset{\sim}{e} \in (\mathcal{C}(X))^\perp \end{cases}$$

$\hat{Y}, \underset{\sim}{e}$ are uncorrelated

$$\begin{cases} \operatorname{cov}(\underset{\sim}{\hat{Y}}, \underset{\sim}{e}) \\ = \operatorname{cov}(P_x \underset{\sim}{Y}, (I_n - P_x)\underset{\sim}{Y}). \\ = P_X \sigma^2 I_n (I_n - P_x)^T. \\ = \sigma^2 P_X (I_n - P_x)^T. \\ = \sigma^2 \, 0_{\text{matrix}} = 0 \text{ matrix.} \end{cases}$$

$$\hat{\underset{\sim}{y}} = X\hat{\beta} = P_X \underset{\sim}{y} \quad \Big| \quad P_X X = X.$$

$$\underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$$

$$\hat{\underset{\sim}{y}} \sim N\left(P_X X \underset{\sim}{\beta},\ P_X(\sigma^2 I_n) P_X^T\right)$$

$$= N\left(X\underset{\sim}{\beta}, \sigma^2 P_X\right)$$

$$\Rightarrow E(\underset{\sim}{y}) = E(\hat{\underset{\sim}{y}})$$

but $\underline{\hat{y}}$ are no more independat.

$$\underset{\sim}{e} = (I_n - P_X)\underset{\sim}{y}.$$ <u>estimated error.</u>

$$\underset{\sim}{e} \sim N\left((I_n - P_X)X\beta,\ (I_n - P_X)(\sigma^2 I_n)(I_n - P_X)^T\right)$$

$$= N\left(\underset{\sim}{0}, \sigma^2(I_n - P_X)\right)$$

$$E(\underset{\sim}{e}) = 0 = E(\underset{\sim}{\epsilon})$$

but $\underset{\sim}{e}$ is no more ~~ind~~ <u>iid.</u>

$$\underset{\sim}{e} = (I_n - P_x)\underset{\sim}{Y} \sim N\left(\underset{\sim}{0}, \sigma^2(I_n - P_x)\right).$$

$$\sum_{i=1}^{n} e_i^2 = \min_{\underset{\sim}{\beta}} (\underset{\sim}{Y} - X\beta)^T(\underset{\sim}{Y} - X\underset{\sim}{\beta})$$

$$= (Y - X\hat{\underset{\sim}{\beta}})^T(Y - X\hat{\underset{\sim}{\beta}}).$$

$$= \underset{\sim}{e}^T \underset{\sim}{e}.$$

$$= \left[(I_n - P_x)\underset{\sim}{Y}\right]^T\left[(I_n - P_x)Y\right]$$

$$= \underset{\sim}{Y}^T (I_n - P_x)\underset{\sim}{Y}.$$

$$= \sigma^2 \left(\frac{\underset{\sim}{Y}}{\sigma}\right)^T (I_n - P_x)\left(\frac{\underset{\sim}{Y}}{\sigma}\right).$$

$$= \sigma^2(\underset{\sim}{Z}^T (I_n - P_x)\underset{\sim}{Z}.) \sim \sigma^2 \chi^2_{df = (n-2)}$$

$$ncp\ \left(\frac{\underset{\sim}{X\beta}}{\sigma}\right)^T (I_n - P_x)\left(\frac{X\beta}{\sigma}\right).$$

We know $\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n).$

$$\underset{\sim}{Z} = \frac{\underset{\sim}{Y}}{\sigma} \sim N\left(\left(\frac{X\underset{\sim}{\beta}}{\sigma}\right), I_n\right).$$

$$(I_n - P_x)^T = (I_n - P_x) \qquad X = [\underset{\sim}{1}\ \underset{\sim}{x}]$$

$$(I - P_x)^2 = (I_n - P_x).$$

$$\text{Rank}(I_n - P_x) = n - \text{rank}(P_x)$$

$$= n - 2$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, 2 \ldots n.$$

if all x_i's are same. $x_i = c$.

$$\left.\begin{aligned} E(Y_i) &= \beta_0 + \beta_1 c. \\ V(Y_i) &= \sigma^2 \end{aligned}\right\} \text{Normal}$$

$$ncp\ \left(\frac{X\underset{\sim}{\beta}}{\sigma}\right)^T \underbrace{(I_n - P_x)\left(\frac{X\underset{\sim}{\beta}}{\sigma}\right)}$$

$$= \frac{1}{\sigma^2}\underset{\sim}{\beta}^T X^T (I_n - P_x) X \underset{\sim}{\beta} = \frac{1}{\sigma^2}\underset{\sim}{\beta}^T\left[X^T X - X^T P_x X\right]\underset{\sim}{\beta}. = \frac{1}{\sigma^2}\underset{\sim}{\beta}^T O \underset{\sim}{\beta}.$$

matrix

$$= 0_{\text{Sum}}.$$

$$\sum_{i=1}^{n} e_i^2 = \underset{\sim}{e}^T \underset{\sim}{e} \sim \sigma^2 \chi^2_{n-2,\ ncp=0}.$$

$$E(\underset{\sim}{e}^T \underset{\sim}{e}) = \sigma^2 (n-2)$$

$$\Rightarrow \quad E\left(\frac{e^T e}{n-2}\right) = \sigma^2$$

$$\Rightarrow \quad E\left(\frac{\underset{\sim}{Y}^T (I_n - P_x) \underset{\sim}{Y}}{n-2}\right) = \sigma^2.$$

$$\underset{\sim}{Y}^T (I_n - P_x) \underset{\sim}{Y}.$$

$$= \sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

$$= S_{yy} - \frac{S_{xy}^2}{S_{xx}}.$$

$\rightarrow$ $\hat{\sigma}^2_{LS} = \dfrac{Y^T (I_n - P_x) Y}{n-2}$ is an unbiased estimator of σ^2.

$\rightarrow$ we can obtain the estimated error variance without fitting the model.

$\rightarrow$ $\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$ Model.

HW. ① obtain ~~[illegible]~~ Maximum likelihood estimation of $\beta_0, \beta_1, \sigma^2$.

② show that ① $E(\hat{\beta}_{0\,mle}) = E(\hat{\beta}_{0\,LS})$

② $E(\hat{\beta}_{1\,mle}) = E(\hat{\beta}_{1\,LS})$.

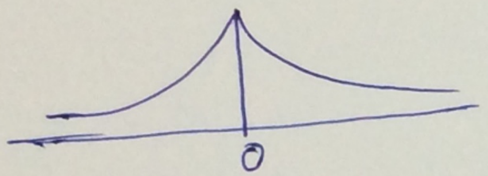
③ $E(\hat{\sigma}^2_{mle}) \neq E(\hat{\sigma}^2_{LS})$.

$$\hat{\sigma}^2_{mle} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

HW.

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

$\epsilon_i \sim$ double exponential $(0, \lambda)$



$$f(\epsilon) = \frac{1}{2\lambda} e^{-\frac{|\epsilon|}{\lambda}} \quad \lambda > 0, \ \epsilon \in \mathbb{R}.$$

Find the LS as well as MLE of $\beta_0, \beta_1, \lambda$, (computer).

$$\prod_{i=1}^{n} f(y_i | x_i) = \prod_{i=1}^{n} \frac{1}{2\lambda} e^{-\frac{|y_i - \beta_0 - \beta_1 x_i|}{\lambda}}$$

$$= \left(\frac{1}{2\lambda}\right)^n e^{-\frac{1}{\lambda} \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|}$$

absolute distance

Bivariate normal, testing, prediction.

Data = $\{(x_i, y_i).\ i = 1, 2, \dots n\}$.

Assumption: (i) x_i, y_i both are stochastic.

(ii) (x_i, y_i) are independently distributed.

(iii) $(x_i, y_i) \overset{iid}{\sim}$ Bivariate Normal $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$.

The regression on $Y|X=x$ and $X|Y=y$ both are applicable.

The conditional distribution of $Y|X=x$ can be obtained as

$$Y|x \sim N\big(\underbrace{E(Y|X)}_{\text{Regression line.}},\ V(Y|X)\big).$$

By definition $Y|X=x$ has p.d.f. $f(y|x) = \dfrac{f(x,y)}{f_X(x)}$.

$f_X(x) \sim N(\mu_x, \sigma_x^2)$ and $f_Y(y) \sim N(\mu_y, \sigma_y^2)$.

$$f_{XY}(x,y) = \frac{e^{-\frac{1}{2}\left\{\frac{1}{(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\}}}{2\pi\,\sigma_x\,\sigma_y\,\sqrt{1-\rho^2}}.$$

$$= \frac{e^{-\frac{1}{2}\left(\frac{[Q]}{1-\rho^2}\right)}}{2\pi\,\sigma_x\sigma_y\sqrt{1-\rho^2}}.$$

$$Q = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right).$$

$$= \left\{\left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{y-\mu_y}{\sigma_y}\right)\left(\frac{x-\mu_x}{\sigma_x}\right) + \rho^2\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\} + (1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2.$$

$$f(x,y) = \frac{e^{-\frac{1}{2}\frac{1}{(1-\rho^2)}\left[\left\{\left(\frac{y-\mu_y}{\sigma_y}\right) - \rho\left(\frac{x-\mu_x}{\sigma_x}\right)\right\}^2 + (1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]}}{2\pi\sqrt{1-\rho^2}\,\sigma_x\sigma_y}$$

$$= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}{\sqrt{2\pi}\,\sigma_x}. \quad \frac{e^{-\frac{1}{2}\frac{1}{(1-\rho^2)\sigma_y^2}\left[\left\{y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right\}^2\right]}}{\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_y}.$$

$$\frac{f(x,y)}{f_X(x)} = \frac{e^{-\frac{1}{2}\frac{\left[y-\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right)\right]^2}{(1-\rho^2)\sigma_y^2}}}{\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_y} \Rightarrow$$

$$Y|X \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x),\ (1-\rho^2)\sigma_y^2\right)$$

$$E(Y|X) = \underbrace{\mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x}_{\beta_0} + \underbrace{\rho\frac{\sigma_y}{\sigma_x}}_{\beta_1}x.$$

$$\underline{V(Y|x) \le V(Y)}$$

More general cases can be addressed.

$$(y, \underset{\sim}{x}) \sim N\left[\begin{pmatrix}\mu_y \\ \underset{\sim}{\mu_x}\end{pmatrix}, \left(\begin{array}{c|c}\sigma_y^2 & \underset{\sim}{\sigma_{yx}}^T \\ \hline \underset{\sim}{\sigma_{yx}} & \Sigma_x\end{array}\right)\right]$$

$$y \mid \underset{\sim}{X} = \underset{\sim}{x} \sim N\left(\mu_y + \underset{\sim}{\sigma_{yx}}^T \Sigma_x^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu_x}),\ \sigma_y^2 - \underset{\sim}{\sigma_{yx}}^T \Sigma_x^{-1} \underset{\sim}{\sigma_{yx}}\right).$$

$$(\underset{\sim}{y}, \underset{\sim}{x}) \sim N\left[\begin{pmatrix}\underset{\sim}{\mu_y} \\ \underset{\sim}{\mu_x}\end{pmatrix}\begin{pmatrix}\Sigma_{yy} & \Sigma_{yx}^T \\ \Sigma_{yx} & \Sigma_{xx}\end{pmatrix}\right]$$

$$\underset{\sim}{Y} \mid \underset{\sim}{X} = \underset{\sim}{x} \sim N\left(\underset{\sim}{\mu_y} + \Sigma_{yx}^T \Sigma_{xx}^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu_x}),\ \Sigma_{yy} - \Sigma_{yx}^T \Sigma_{xx}^{-1} \Sigma_{yx}\right).$$

→ Feature prediction.

→ Missing data analysis. $\begin{pmatrix} y_3 \\ y_5 \end{pmatrix} \Big| (x_1\, x_2\, x_3\, x_4\, x_5,\ y_1\, y_2\, y_4)$.

x_1, x_2, $\cancel{x_3}$, $\cancel{x_4}$, $\cancel{x_5}$, x_6, y_1, y_2, $\cancel{y_3}$, y_4, $\cancel{y_5}$

~~H~~ F P L S

1, 2, - 6 - 7 n.

Hx Px | Hx Px | Hx Px

2 6 7

H P

F L S

$(X_i\, Y_i) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

To test the independence between Y and X we need to test. (13)

$H_0\ \beta_1 = 0$ vs $H_1: \beta_1 \neq 0.$

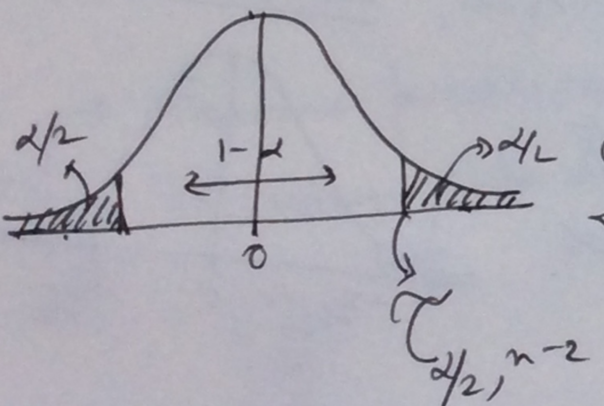
$\Leftrightarrow H_0\ \rho = 0$ vs $H_1: \rho \neq 0.$ ρ is population correlation coefficient

Consider the sample correlation coefficient.

$$\hat{\rho} = r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}.$$

Under H_0

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2},\ \text{rep} = 0.$$



We reject H_0 in favour of H_1 at level α
if $|T(\text{observed})| > \tau_{\alpha/2, n-2}$.

uncorrelated $\not\Rightarrow$ independence. → Exception (BN, B.Bernoulli)

independence $\Rightarrow$ uncorrelated.

(x_1, y_1)
(x_2, y_2)
(x_3, y_3)
$\vdots$
(x_n, y_n)

$$Y_i \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$$

if $(x_i, y_i) \overset{iid}{\sim} N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

$$\begin{cases} H_0: \rho = \rho_0 \neq 0 \\ H_1: \rho \neq \rho_0. \end{cases}$$

define.

$$Z = \frac{1}{2} \log \frac{1+r}{1-r} = \tanh^{-1}(r).$$

$$\mu_Z = \frac{1}{2} \log \frac{1+\rho_0}{1-\rho_0} = \tanh^{-1}(\rho_0)$$

$$\sigma_Z^2 = (n-3)^{-1}$$

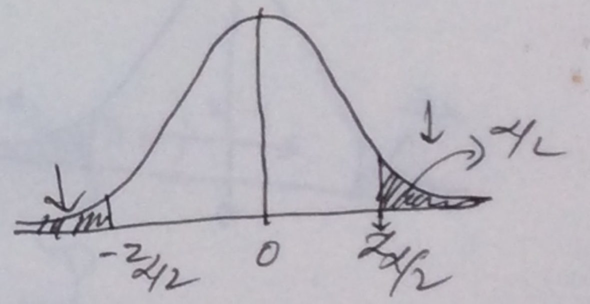
Define the test statistic $W = \dfrac{Z - \mu_Z}{\sigma_Z}$.

$W \sim N(0,1)$ under H_0 when $n \uparrow \infty$. Large sample.

We reject H_0 in favour of H_1 if

$$|W(\text{observed})| > Z_{\alpha/2}$$

at level α.



(x_1, y_1)
(x_2, y_2)
(x_3, y_3)
$\vdots$
(x_n, y_n)

$y_i \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$

if $(x_i, y_i) \overset{iid}{\sim} N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

$$\begin{cases} H_0: \rho = \rho_0 \neq 0 \\ H_1: \rho \neq \rho_0. \end{cases}$$

define.

$$Z = \frac{1}{2} \log \frac{1+r}{1-r} = \tanh^{-1}(r).$$

$$\mu_Z = \frac{1}{2} \log \frac{1+\rho_0}{1-\rho_0} = \tanh^{-1}(\rho_0).$$

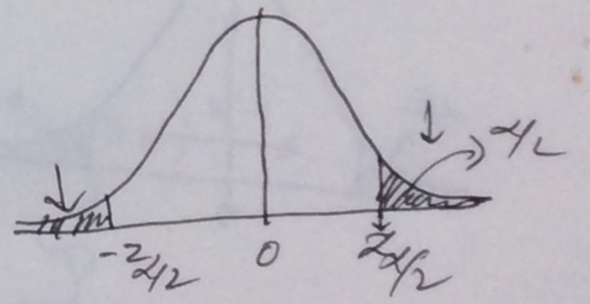
$$\sigma_Z^2 = (n-3)^{-1}$$

Define the ~~st~~ test statistic $W = \dfrac{Z - \mu_Z}{\sigma_Z}$.

$W \sim N(0,1)$ under H_0 when $n \uparrow \infty$. Large sample.

We reject H_0 in favour of H_1 if

$|W(\text{observed})| > Z_{\alpha/2}$ at level α.

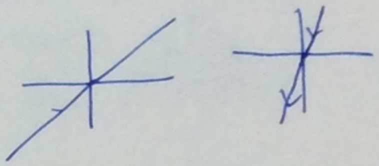


Testing for simple linear regression.

$\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n) \iff y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ independent.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$H_0: \beta_0 = b_0$ vs $H_1: \underline{b_0 \neq b_0}$. if $b_0 = 0$.



<u>Step 1</u>. β_0 is estimated by $\hat{\beta}_0 = \bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}$

<u>Step 2</u>. $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$

under Ho.

$$\hat{\beta}_0 \sim N\left(b_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)\right)$$

<u>Step 3</u>.

$$\frac{\hat{\beta}_0 - b_0}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)}} \sim N(0,1) \text{ under Ho.}$$

<u>Step 4</u>: As σ^2 is unknown σ^2 is estimated by $\hat{\sigma}^2$

where $\hat{\sigma}^2 = \frac{e^T e}{n-2} = \left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) / (n-2)$.

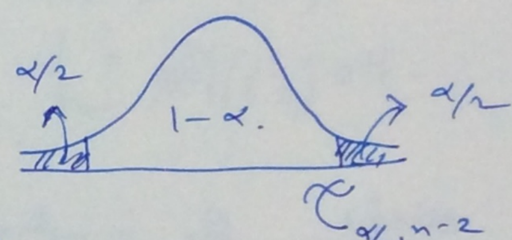
<u>Step 5.</u> Get the distribution of test statistic under H_0.

$$T = \frac{\hat{\beta}_0 - b_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t_{df = n-2,\ ncp = 0} \quad \text{under } H_0.$$

$$\underset{\sim}{e}^T \underset{\sim}{e} \sim \sigma^2 \chi^2_{n-2,\ ncp=0}$$

$$E(\underset{\sim}{e}^T \underset{\sim}{e}) = (n-2)\sigma^2$$

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



<u>Step 6:</u> We reject H_0 in favor of H_1 if

$$|T(\text{observed})| > \tau_{\alpha/2,\, n-2} \quad \text{at level-}\alpha$$

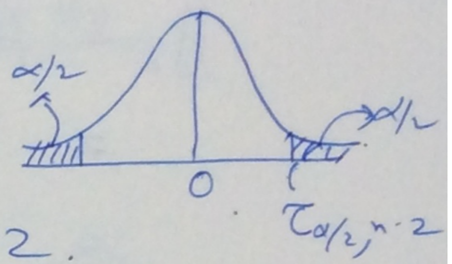
$\alpha = 0.05$ or 0.01.

HW (1) Test $H_0: \beta_1 = b_1$ vs $H_1: \beta_1 \neq b_1$ / 95% CI of β_1

HW (2) Test $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 > \sigma_0^2$ / 95% CI of σ^2

95% CI of β_0 theory

we know $$\hat{\beta}_0 \sim N\left(\beta_0,\ \sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)\right).$$



as σ^2 unknown.

$$T = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t_{n-2}.$$

$\alpha = 0.05$

$$P\left(-\tau_{\alpha/2,n-2} < T < \tau_{\alpha/2,n-2}\right) = 1-\alpha.$$

$$\Rightarrow P\left(-\tau_{\alpha/2,n-2} < \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} < \tau_{\alpha/2,n-2}\right) = 1-\alpha.$$

$$\Rightarrow P\left(\underbrace{\hat{\beta}_0 - \tau_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}_{L(X,Y)} < \beta_0 < \underbrace{\hat{\beta}_0 + \tau_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}_{U(X,Y)}\right) = 1-\alpha.$$

Prediction through regression.

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$ for some x_0, which is nonstochastic.

$$= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0$$

$$= \bar{y} + \hat{\beta}_1 (x_0 - \bar{x}) \quad \Rightarrow \quad \hat{y}_0 \sim N(\beta_0 + \beta_1 x_0, \quad)$$

$$E(\hat{y}_0) = E\left(\bar{y} + \hat{\beta}_1 (x_0 - \bar{x})\right) = E(\bar{y}) + E(\hat{\beta}_1)(x_0 - \bar{x})$$

$$= E(\hat{\beta}_0) + E(\hat{\beta}_1) x_0 = \beta_0 + \beta_1 x_0$$

$$V(\hat{y}_0) = V(\bar{y}) + (x_0 - \bar{x})^2 \operatorname{Var}(\hat{\beta}_1) + 2(x_0 - \bar{x}) \operatorname{cov}(\bar{y}, \hat{\beta}_1).$$

$$= \underbrace{\frac{\sigma^2}{n} + (x_0 - \bar{x})^2 \frac{\sigma^2}{S_{xx}}} + 2(x_0 - \bar{x}) \cdot 0.$$

$\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I_n)$

$\bar{Y} = \frac{1}{n} \underset{\sim}{1}^T \underset{\sim}{Y}.$

$\hat{\beta}_1 = \underset{\sim}{c}^T \underset{\sim}{Y}$

$$\operatorname{cov}(\bar{Y}, \hat{\beta}_1) = \frac{1}{n} \underset{\sim}{1}^T (\sigma^2 I_n) \underset{\sim}{c}$$

$$= \frac{\sigma^2}{n} \underset{\sim}{1}^T \underset{\sim}{c} \qquad c_i = \frac{(x_i - \bar{x})}{S_{xx}}$$

$$= \frac{\sigma^2}{n} \cdot 0.$$

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ ~~where~~ $x_0 \notin$ ~~...~~

$$\hat{Y}_0 \sim N\left(\beta_0 + \beta_1 x_0,\ \sigma^2\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)\right)$$

Case I. $95\% = 100(1-\alpha)\%$ CI of $E(\hat{Y}_0)$.

$$\hat{y}_0 - t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)} < E(\hat{Y}_0) < \hat{y}_0 + t_{\alpha/2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)}$$

Case. Lets us assume $Y_0 \sim N(\beta_0 + \beta_1 x_0, \sigma^2)$ and (y_0, x_0) is independent from $\{(y_1, x_1)\ (y_2, x_2) \cdots (y_n, x_n)\}$

$100(1-\alpha)\% = 95\%$ Prediction interval of ~~$\hat{y}_0$~~ Y_0

$$Y_0 - \hat{Y}_0 \sim N\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)\right)$$

$$T_0 = \frac{Y_0 - \hat{Y}_0}{\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)}} \sim t_{n-2}$$

Prediction interval of Y_0 is $\hat{y}_0 \mp t_{\alpha/2, n-2}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)}$