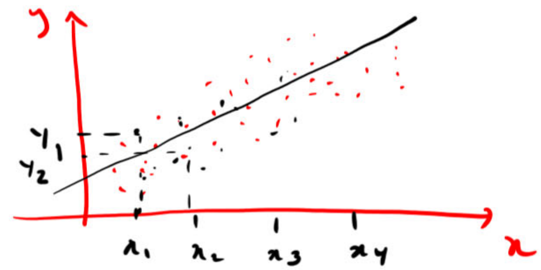
Math 4501 - Probability and Statistics II

6.5 - Regression ← we will employ MLE technique to determine the regression parameters

Simplest regression problem

$Y_i - \alpha_1 + \beta x_i = \varepsilon_i$



Given the data points

$$\rightarrow (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

estimate the parameters α and β of the linear model

$$\boxed{E[Y|x] = \underbrace{\alpha_1 + \beta x}_{\mu(x)},}$$

$$Y - \underbrace{E[Y|x]}_{\alpha_1 + \beta x} = \varepsilon \sim N(0, \sigma^2)$$

unknown

that is, fit a straight line to the given set of data.

$$E[Y|x] = \mu(x) = \alpha_1 + \beta x$$

Assumptions:

- for each particular value of x the value of Y differs from its mean by a random amount ε.
- the distribution of ε is $N(0, \sigma^2)$.

$$Y = \underbrace{\alpha_1 + \beta x + \varepsilon}$$

σ^2 is another parameter to estimate

Consequence: For the linear model described above, we have

$$\boxed{Y_i = \overbrace{\alpha_1 + \beta x_i}^{\mu(x_i)} + \varepsilon_i,}$$

$\varepsilon_i \sim N(0, \sigma^2)$

where ε_i, $i = 1, 2, \ldots, n$, are independent $N(0, \sigma^2)$ random variables.

$$\Rightarrow \boxed{Y_i \sim N(\alpha_1 + \beta x_i, \sigma^2)}$$

GOAL: Estimate

$\underline{\alpha_1}, \underline{\beta}, \underline{\sigma^2}$

$\mu(x) = \alpha_1 + \beta x$

$y = mx + b$

$y = y_0 + m(x - x_0)$

- For convenience, we set

$$\boxed{\alpha_1 = \alpha - \beta\bar{x}},$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of the observations $x_1, \ldots, x_n$.

- For each $i = 1, 2, \ldots, n$, we have that

$Y_i = \alpha_1 + \beta x + \varepsilon_i \xrightarrow{\alpha_1 = \alpha - \beta\bar{x}}$

$$\boxed{Y_i = \underbrace{\alpha + \beta(x_i - \bar{x})} + \varepsilon_i}$$

is equal to a nonrandom quantity $\alpha + \beta(x_i - \bar{x})$ plus a mean-zero normal random variable ε_i.

- The random variables $Y_1, Y_2, \ldots, Y_n$ are mutually independent normal variables with respective means

$$\alpha + \beta(x_i - \bar{x}), \quad i = 1, 2, \ldots, n$$

and unknown variance σ^2.

estimate α, β, σ^2

$$\boxed{Y_i \sim N(\alpha + \beta(x_i - \bar{x}), \sigma^2)}$$

Proposition

Under the conditions described above, the maximum likelihood estimators of α, β *and* σ^2 *are given by:*

$$\widehat{\alpha} = \bar{Y}$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\widehat{\sigma^2} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \widehat{\alpha} - \widehat{\beta}(x_i - \bar{x})\right]^2}$$

Interpretation for: $\hat{\beta} = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ = ratio of sample covariance to "sample" variance

$\overbrace{\hat{Y}_i}^{\text{estimates for } Y_i} = \hat{\alpha} - \hat{\beta}(x_i - \bar{x})$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{Y}_i\right]^2$ } average of the squares of the deviations between actual values and estimated values!

Distributions of $\widehat{\alpha}$ and $\widehat{\beta}$

- As in the preceding discussion $x_1, x_2, \ldots, x_n$ are treated as nonrandom constants.
- Since the x-values are given, when determining the distributions of $\widehat{\alpha}$ and $\widehat{\beta}$, the only random variables are $Y_1, Y_2, \ldots, Y_n$.

Proposition

Under the conditions described earlier, we have that:

$\hat{\alpha} = \bar{Y} = \frac{1}{n}\sum Y_i$

1) $\widehat{\alpha}$ is normally distributed with mean α and variance $\frac{\sigma^2}{n}$, that is, $\hat{\alpha} \sim N\left(\alpha, \frac{\sigma^2}{n}\right)$

2) $\widehat{\beta}$ is normally distributed with mean β and variance, that is, $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})}$$

$$\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\ .$$

CONSEQUENCE: $\hat{\alpha}$ and $\hat{\beta}$ are unbiased estimators for α and β, respectively, because $E[\hat{\alpha}] = \alpha$ and $E[\hat{\beta}] = \beta$

Proof : By assumption $Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2)$, that is, $Y_i \sim N(\overbrace{\alpha + \beta(x_i - \bar{x})}^{E[Y_i]}, \overbrace{\sigma^2}^{\mathrm{Var}(Y_i)})$

1) Recall that $\underbrace{\hat{\alpha} = \bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i}$

Since $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ are independent then $Y_1, Y_2, \dots, Y_m$ are independent.

Since $\hat{\alpha}$ is a linear combination of independent normally distributed random variables, then $\hat{\alpha}$ is also normally distributed, that is $\hat{\alpha} \sim N(\mu_{\hat{\alpha}}, \sigma^2_{\hat{\alpha}})$

we need to find these two values

$$\mu_{\hat{\alpha}} = E[\hat{\alpha}] = E[\bar{Y}] = E\left[\frac{1}{m}\sum_{i=1}^{m} Y_i\right] \underset{\text{linearity}}{=} \frac{1}{m}\sum_{i=1}^{m} \underbrace{E[Y_i]}_{\alpha + \beta(x_i - \bar{x})} = \frac{1}{m}\sum_{i=1}^{m} \alpha + \beta(x_i - \bar{x})$$

$$= \frac{1}{m}\underbrace{\sum_{i=1}^{m} \alpha}_{m\alpha} + \frac{\beta}{m}\underbrace{\sum_{i=1}^{m}(x_i - \bar{x})}_{m\bar{x} - m\bar{x} = 0} = \frac{1}{m} \cdot m\alpha + \frac{\beta}{m} \cdot 0 = \alpha$$

$$\sigma_{\hat\alpha}^2 = Var(\hat\alpha) = Var\left(\frac{1}{m}\sum_{i=1}^{m} Y_i\right) = \frac{1}{m^2} Var\left(\sum_{i=1}^{m} Y_i\right) = \frac{1}{m^2}\sum_{i=1}^{m} \underbrace{Var(Y_i)}_{\sigma^2}$$

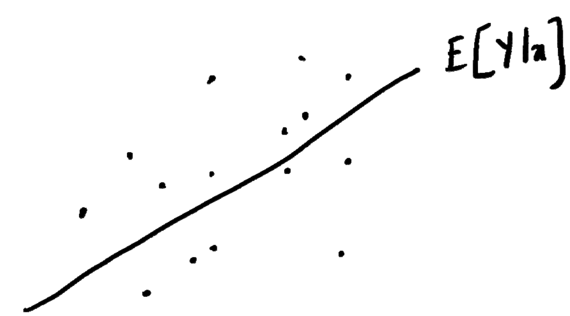
$Y_1, Y_2, \dots, Y_m$
independence

$$= \frac{1}{m^2}\underbrace{\sum_{i=1}^{m} \sigma^2}_{m\sigma^2} = \frac{1}{m^2}\cdot m\sigma^2 = \frac{\sigma^2}{m}$$

CONCLUSION: $$\boxed{\hat\alpha \sim N\left(\alpha, \frac{\sigma^2}{m}\right)}$$

DETOUR:

$\underline{\underline{x}}$ $\quad Y(x) = ???$ ← not easy



$$\boxed{E[Y|x] = \alpha + \beta(x - \bar{x})}$$

$$Y_i - E[Y|x_i] = Y_i - \alpha - \beta(x_i - \bar{x}) = \varepsilon_i$$ ← error or deviation (random)

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i$$

Proof of item 2:

Recall that $Y_i = \overbrace{\alpha + \beta(x_i - \bar{x}) + \varepsilon_i}$ $\sim N(\overbrace{\alpha + \beta(x_i-\bar{x})}^{E[Y_i]}, \sigma^2)$, $i = 1, \dots, n$, independent

σ^2: Var(Y_i)

and that

$$\hat{\beta} = \boxed{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \underset{\uparrow}{=} \frac{\sum_{i=1}^{n} Y_i (x_i - \bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

last time

because $\sum_{i=1}^{n} \bar{Y}(x_i - \bar{x}) = \bar{Y} \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{=0} = \bar{Y} \cdot 0 = 0$

Since $\hat{\beta}$ is a linear combination of $\underbrace{\text{independent normally distributed r.v.s}}_{\text{these are the } Y_i\text{'s}}$

$$\hat{\beta} = Y_1 \cdot \frac{\overbrace{(x_1 - \bar{x})}^{a_1}}{\sum (x_i - \bar{x})^2} + Y_2 \cdot \frac{\overbrace{(x_2 - \bar{x})}^{a_2}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} + \dots + Y_n \frac{\overbrace{(x_n - \bar{x})}^{a_n}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

then $\hat{\beta}$ is itself normally distributed, that is

$$\hat{\beta} \sim N\left(\mu_{\hat{\beta}}, \sigma^2_{\hat{\beta}}\right)$$

to be determined.

Then:

$$\mu_{\hat{\beta}} = E\left[\hat{\beta}\right] = E\left[\frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right] = \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2} E\left[\sum_{i=1}^{n} y_i (x_i - \bar{x})\right]$$

linearity

$$\overset{\text{linearity}}{=} \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \overbrace{E[y_i]}^{\alpha + \beta(x_i - \bar{x})}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \overbrace{\left(\alpha + \beta(x_i - \bar{x})\right)}^{E[y_i]}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} \left(\alpha (x_i - \bar{x}) + \beta (x_i - \bar{x})^2 \right)}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\alpha \overbrace{\sum_{i=1}^{n} (x_i - \bar{x})}^{=0} + \beta \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\beta \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta$$

, that is $E[\hat{\beta}] = \beta$

Finally, let us compute $\sigma^2_{\hat{\beta}}$:

$$\sigma^2_{\hat{\beta}} = \mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} Y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right) = \frac{1}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^2} \mathrm{Var}\left(\sum_{i=1}^{n} Y_i (x_i - \bar{x}) \right)$$

$Y_1, \ldots, Y_n$ independent

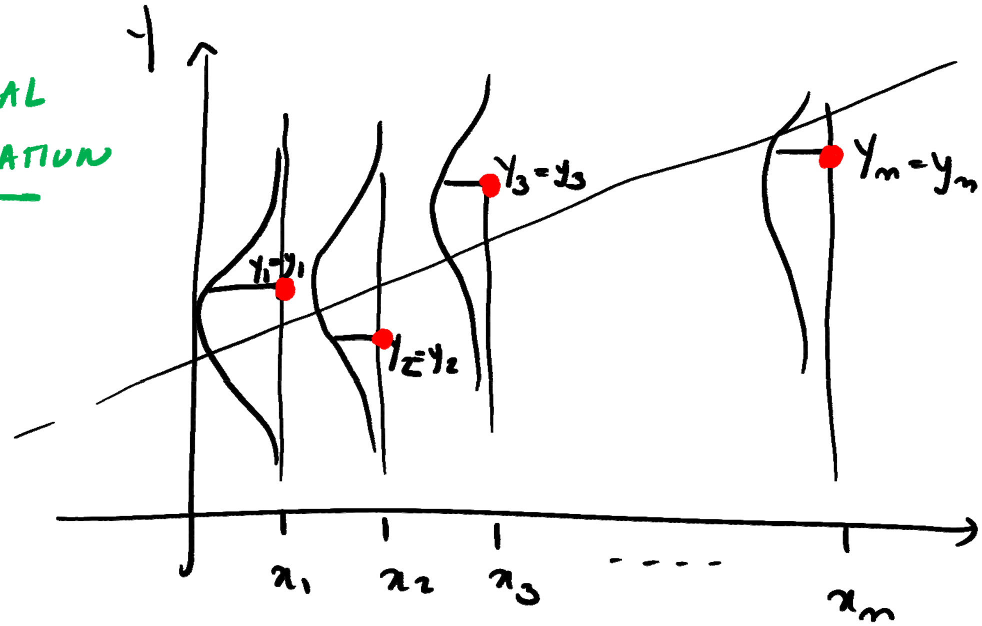
$$\overset{\downarrow}{=} \frac{1}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} \sum_{i=1}^{n} \mathrm{Var}\left(Y_i (x_i-\bar{x})\right) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 \overbrace{\mathrm{Var}(Y_i)}^{\sigma^2}}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2}$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sigma^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} = \frac{\sigma^2 \sum_{i=1}^{n}(x_i-\bar{x})^2}{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

CONCLUSION:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

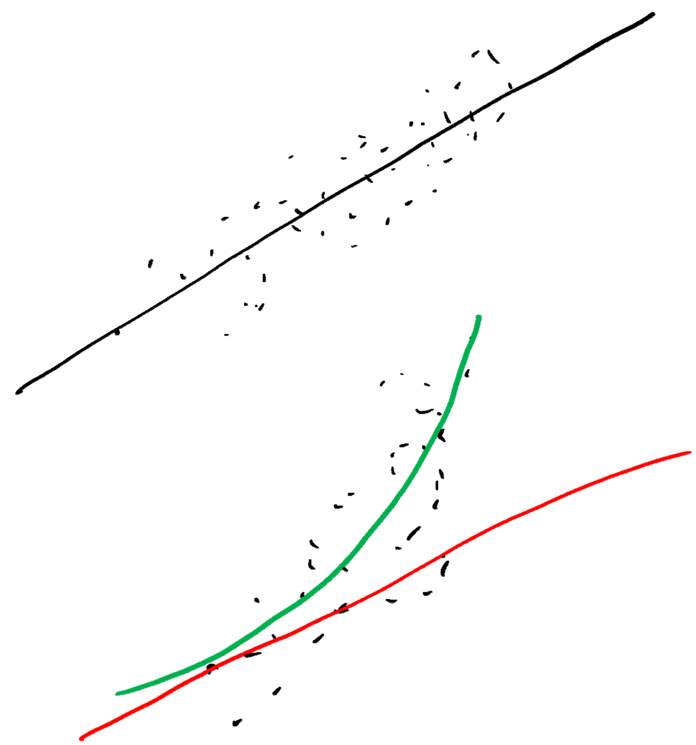
GEOMETRICAL INTERPRETATION



$\alpha + \beta x$

$\alpha e^{\beta x}$

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i$$



$$\alpha e^{\beta x}$$

$$\alpha + \beta x + \gamma x^2$$

$$y = e^{x}$$

$$y = \alpha e^{\beta x}$$

$$\to \quad z = \ln y = \ln\left(\alpha e^{\beta x}\right) = \ln \alpha + \ln e^{\beta x} = \overbrace{\ln \alpha}^{\tilde{\alpha}} + \beta x$$

$$z = \tilde{\alpha}_1 + \beta x$$

In general, when we want to estimate an unknown parameter $\theta \in \Omega$.

for a prob. distr. $f(x, \theta)$

We take a random sample $\underbrace{X_1, X_2, \ldots, X_n}_{\text{i.i.d. r.v.s}}$ and determine

an estimator $\hat{\theta} = u(X_1, X_2, \ldots, X_n)$ $\leftarrow$ can use MLE or method of moments

When we actual collect data, we get $X_1 = x_1$, $X_2 = x_2$, $\ldots$, $X_n = x_n$

$x_1, x_2, \ldots, x_n$ are real numbers!!

we can use $x_1, x_2, \ldots, x_n$ to find a point estimate $\hat{\theta} = \underbrace{u(x_1, x_2, \ldots, x_n)}_{\text{a number}}$

With regression: pick values $x_1, x_2, \dots, x_n$ non random.

For each value we propose to observe a r.v. Y_i, $i = 1, 2, \dots n$

$Y_1, Y_2, \dots, Y_n$ are mutually independent

We assume that $Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

unknown

unknown

Use MLE to find estimators

$$\hat{\alpha}(Y_1, Y_2, \dots, Y_n, x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

$$\hat{\beta}(Y_1, Y_2, \dots, Y_n, x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$\widehat{\sigma^2}(Y_1, Y_2, \dots, Y_n, x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} \left[Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x}) \right]^2$$

$\underbrace{\text{estimators for } \alpha, \beta, \text{ and } \sigma^2}_{\text{functions of the sample}}$

as r.v.s.

Finally, after observing $Y_1 = y_1$, $Y_2 = y_2$, ..., $Y_n = y_n$ we can get point estimates

ε_1 ε_2 ε_n

for $\hat{\alpha}$, $\hat{\beta}$, and $\widehat{\sigma^2}$ ← numerical values