Lecture 3 - 24 Oct 2023

- EXACT DISTRIBUTION of $\hat{\beta}_1(Y)$ and $\hat{\beta}_2(Y)$

> **Preliminary result 1**
> Given $Y_1, \dots, Y_n$ independent with distribution $Y_i \sim N(\mu_i, \sigma^2)$ $i=1,\dots,n$ and a sequence of known constants a_i, $i=1,\dots,n$,
> $\sum_{i=1}^n a_i Y_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \ \sigma^2 \sum_{i=1}^n a_i^2\right)$

We have seen that $\hat{\beta}_1$ and $\hat{\beta}_2$ are linear combinations of $Y_1, \dots, Y_n$ of the form

$$\hat{\beta}_1 = \sum_{i=1}^n v_i Y_i \qquad \hat{\beta}_2 = \sum_{i=1}^n w_i Y_i$$

hence $\hat{\beta}_1(Y)$ and $\hat{\beta}_2(Y)$ are exactly Gaussian-distributed r.v. (see res. 1)
Moreover, the expression of the two estimators are the same we obtained with OLS.
In fact, the Gaussian linear model is a special case. Hence the properties we computed still hold.
In particular, we computed

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 \qquad var(\hat{\beta}_1) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$\mathbb{E}[\hat{\beta}_2] = \beta_2 \qquad var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

The exact distributions are then easily obtained as

$$\hat{\beta}_1(Y) \sim N\left(\beta_1; \ \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)\right)$$

$$\hat{\beta}_2(Y) \sim N\left(\beta_2; \ \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

- EXACT DISTRIBUTION of $\hat{\sigma}^2(Y)$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

it is possible to show that $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$ Chi-squared with n-2 degrees of freedom

In general, for a χ^2_v r.v., the expected value is v

$$\mathbb{E}\left[\frac{n\hat{\sigma}^2}{\sigma^2}\right] = (n-2) \Rightarrow \mathbb{E}[\hat{\sigma}^2] = \frac{(n-2)}{n}\sigma^2$$

hence again we obtain an unbiased estimator as

$$S^2 = \frac{n}{n-2}\hat{\sigma}^2 \qquad \mathbb{E}[S^2] = \frac{n}{n-2}\mathbb{E}[\hat{\sigma}^2] = \frac{n}{n-2}\cdot\frac{n-2}{n}\sigma^2 = \sigma^2$$

and $\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$.

Moreover, it is possible to show that $\hat{\sigma}^2 \perp\!\!\!\perp (\hat{\beta}_1, \hat{\beta}_2)$
(hence, also $S^2 \perp\!\!\!\perp (\hat{\beta}_1, \hat{\beta}_2)$)

INFERENCE ABOUT β

We have derived the exact distributions of the estimators.
With these distributions we can test statistical hypotheses, compute confidence intervals.

Examples

Test: $\begin{cases} H_0: \beta_j = b \\ H_1: \beta_j \neq b \end{cases}$ $\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j > 0 \end{cases}$ $j = 1,2$

Confidence interval, of level $1-\alpha$: $\hat{B}_j(Y)$ such that $\mathbb{P}\left(\hat{B}_j(Y) \ni \beta_j\right) = 1-\alpha \quad \forall \beta_j \in \mathbb{R}$

Recall that: $\hat{\beta}_1 \sim N(\beta_1, V(\hat{\beta}_1))$, $\hat{\beta}_2 \sim N(\beta_2, V(\hat{\beta}_2))$, $\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}$

where $V(\hat{\beta}_1) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$, $V(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

We need to find a pivotal quantity.
PIVOTAL QUANTITY : a transformation of the data (and of the parameter) whose distribution does not depend on the parameter (hence is completely known).

> **Preliminary result 2**
> If $Z \sim N(0,1)$ and $W \sim \chi^2_v$ independent, then $\frac{Z}{\sqrt{W/v}} \sim t_v$.
> (Student's t with v degrees of freedom)
> ↳ · symmetric distrib.
> · heavier tails than a normal
> · for large v it is very close to a normal

Since $\hat{\beta}_j \sim N(\beta_j, V(\hat{\beta}_j))$, the simplest (and most intuitive) transformation is

$$\Rightarrow \frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}} \sim N(0,1)$$

however, $V(\hat{\beta}_j)$ includes σ^2 which is unknown

In place of $V(\hat{\beta}_j)$ we use an estimate* $\hat{V}(\hat{\beta}_j) = \frac{S^2}{\sigma^2}V(\hat{\beta}_j)$ (e.g. $\hat{V}(\hat{\beta}_2) = \frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$)

* estimator if I we look at it as a function of Y_i to study the distributive properties

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}}$$ what is its distribution?

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} = \frac{\overbrace{\frac{\hat{\beta}_j - \beta_j}{\sqrt{V(\hat{\beta}_j)}}}^{\sim N(0,1)}}{\sqrt{\frac{S^2}{\sigma^2}} \sim \sqrt{\frac{\chi^2_{n-2}}{n-2}}}$$

moreover, $\hat{\beta}_j \perp\!\!\!\perp S^2$

$$\Rightarrow T_j \sim t_{n-2}$$

- CONFIDENCE INTERVAL for β_j

· t is symmetric $t_{\cdot, \frac{\alpha}{2}} = -t_{\cdot, 1-\frac{\alpha}{2}}$
· if n is large t is similar to a Gaussian

$$\mathbb{P}\left(-t_{n-2;1-\frac{\alpha}{2}} < T_j < \underbrace{t_{n-2;1-\frac{\alpha}{2}}}_{\text{quantile } 1-\frac{\alpha}{2} \text{ of a } t_{n-2} \text{ distrib.}}\right) = 1-\alpha$$

$$\mathbb{P}\left(-t_{n-2;1-\frac{\alpha}{2}} < \frac{\hat{\beta}_j(Y) - \beta_j}{\sqrt{\hat{V}(\hat{\beta}_j)}} < t_{n-2;1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\mathbb{P}\left(\hat{\beta}_j(Y) - \sqrt{\hat{V}(\hat{\beta}_j)}\cdot t_{n-2;1-\frac{\alpha}{2}} < \beta_j < \hat{\beta}_j(Y) + \sqrt{\hat{V}(\hat{\beta}_j)}\, t_{n-2;1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\mathbb{P}(\beta_j \in \hat{B}(\underline{Y})) = 1-\alpha$$

$\hat{B}(\underline{Y})$ is a random interval. After observing the data we can compute its realization by substituting the estimators (function of $(Y_1,\dots,Y_n)$ r.v.) with their estimates (function of $(y_1,\dots,y_n)$ realization).

We obtain $\beta_j \in \hat{\beta}_j \pm t_{n-2;1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{\beta}_j)}$.

That is

$$\beta_1 \in \hat{\beta}_1 \pm t_{n-2;1-\frac{\alpha}{2}}\sqrt{S^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}$$

$$\beta_2 \in \hat{\beta}_2 \pm t_{n-2;1-\frac{\alpha}{2}}\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

- HYPOTHESIS TEST on β_j

$\begin{cases} H_0: \beta_j = b \\ H_1: \beta_j \neq b \end{cases}$

$$T_j = \frac{\hat{\beta}_j(Y) - b}{\sqrt{\hat{V}(\hat{\beta}_j)}} \overset{H_0}{\sim} t_{n-2} \text{ (under } H_0)$$

T_j is a random variable. After observing $y_1, \dots, y_n$ we can compute its realization, t_j^{obs}.

- fixed significance level α: $\alpha = \mathbb{P}(\text{reject } H_0 \mid H_0 \text{ true})$

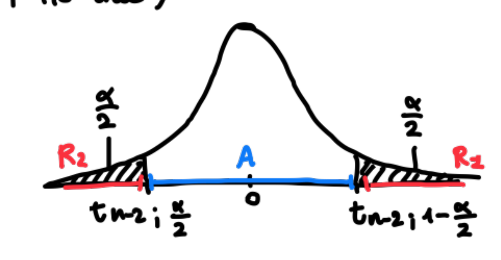
$\mathbb{P}_{H_0}(|T_j| > t_{n-2;1-\frac{\alpha}{2}}) = \alpha$

the acceptance region is

$A = (t_{n-2;\frac{\alpha}{2}}, \ t_{n-2,1-\frac{\alpha}{2}})$

if $t_j^{obs} \in A \Rightarrow$ we do not reject H_0

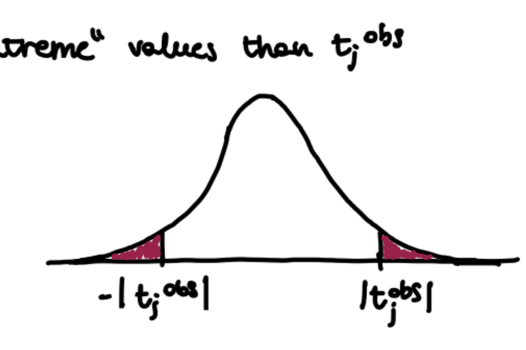
if $t_j^{obs} \notin A \Rightarrow$ we reject H_0



- p-value

it is the probability of observing "more extreme" values than t_j^{obs}

$$\alpha^{obs} = \mathbb{P}_{H_0}(|T_j| > |t_j^{obs}|) = 2\cdot\mathbb{P}_{H_0}(T_j > |t_j^{obs}|)$$



connection between the two types of test

- if $\alpha^{obs} < \alpha \Rightarrow$ reject H_0 at level α
- if $\alpha^{obs} > \alpha \Rightarrow$ do not reject H_0 at a level α

In practical applications, these methods are useful tools to investigate relevant applicative questions. For example:

· does the covariate x have a significative effect on Y?

The effect of x on Y is summarised by the coefficient β_2.
Hence this question can be formalized by the statistical test

$\begin{cases} H_0: \beta_2 = 0 \rightarrow \text{no effect} \\ H_1: \beta_2 \neq 0 \end{cases}$

Indeed the model $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$
under H_0 becomes $Y_i = \beta_1 + \varepsilon_i$ (x has no impact on Y).

INFERENCE ABOUT THE MEAN OF Y : "PREDICTION"

We observe (x_i, y_i) for $i=1,\dots,n$.
Consider an additional unit observed at a value x_*. We want to make a prediction about the value Y_* of the response variable corresponding to x_*.
The model is $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, i.e. $\mathbb{E}[Y_i] = \mu_i = \beta_1 + \beta_2 x_i$
hence $Y_* = \beta_1 + \beta_2 x_* + \varepsilon_*$, with $\mu_* = \beta_1 + \beta_2 x_*$
The predicted value is $\hat{y}_* = \hat{\beta}_1 + \hat{\beta}_2 x_*$.
$\Rightarrow$ the prediction $\hat{y}_*$ is an estimate of the parameter μ_*.
If we consider the estimators $\hat{\beta}_1(Y)$ and $\hat{\beta}_2(Y)$, we obtain the corresponding estimator $\hat{\mu}_* = \hat{\mu}_*(Y)$ of the mean of Y_* (it is a r.v.).
We can study the distribution of $\hat{\mu}_*$.

$$\hat{\mu}_* = \hat{\beta}_1 + \hat{\beta}_2 x_* = \bar{Y} - \hat{\beta}_2\bar{x} + \hat{\beta}_2 x_* = \bar{Y} + \hat{\beta}_2(x_* - \bar{x})$$

$$= \frac{1}{n}\sum_{i=1}^n Y_i + (x_* - \bar{x})\sum_{i=1}^n w_i Y_i \qquad \text{since } \hat{\beta}_2 = \sum_{i=1}^n w_i Y_i \text{ with } w_i = \frac{(x_i - \bar{x})}{\sum_{k=1}^n (x_k - \bar{x})^2}$$

$$= \sum_{i=1}^n \left(\frac{1}{n} + (x_* - \bar{x})w_i\right)Y_i$$

$\Rightarrow \hat{\mu}_*$ is a linear combination of $Y_1, \dots, Y_n$
$\Rightarrow \hat{\mu}_*$ has normal distribution $\quad \hat{\mu}_* \sim N(\dots, \dots)$

$$\mathbb{E}[\hat{\mu}_*] = \mathbb{E}[\hat{\beta}_1 + \hat{\beta}_2 x_*] \overset{\text{linearity}}{=} \beta_1 + \beta_2 x_* = \mu_* \qquad \text{unbiased}$$

homoscedastic $var(Y_i) = \sigma^2 \ \forall i$

$$var(\hat{\mu}_*) = var\left(\sum_{i=1}^n \left(\frac{1}{n} + (x_* - \bar{x})w_i\right)Y_i\right) \overset{\text{ind.}}{=} \sum_{i=1}^n \left(\frac{1}{n} + (x_* - \bar{x})w_i\right)^2\sigma^2 =$$

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

$$= \frac{1}{n}\sigma^2 + \sigma^2(x_* - \bar{x})^2\sum_{i=1}^n w_i^2 + \cancel{2\sigma^2(x_* - \bar{x})\sum_{i=1}^n w_i} =$$

where $\sum_{i=1}^n w_i = 0$

$$= \sigma^2\left(\frac{1}{n} + (x_* - \bar{x})^2 \underbrace{\frac{1}{\sum_{k=1}^n (x_k - \bar{x})^2}}_{\sum_{i=1}^n w_i^2}\right) =$$

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

$$\Rightarrow \hat{\mu}_* \sim N\Bigg(\mu_*, \underbrace{\sigma^2\left(\frac{1}{n} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}_{V(\hat{\mu}_*)}\Bigg) = N(\mu_*, V(\hat{\mu}_*))$$

$$\Rightarrow \frac{\hat{\mu}_* - \mu_*}{\sqrt{V(\hat{\mu}_*)}} \sim N(0,1)$$

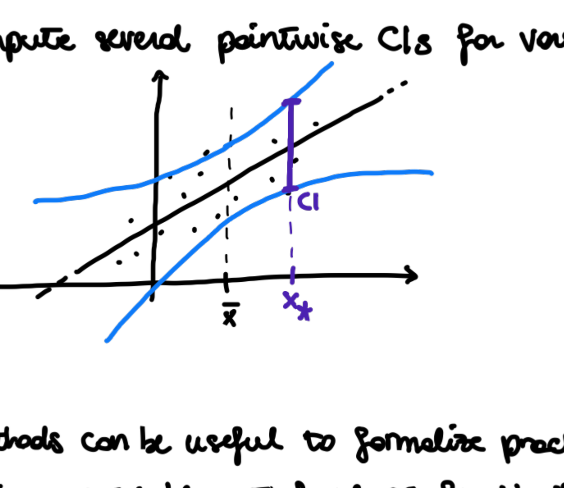
since $V(\hat{\mu}_*)$ involves the unknown σ^2, similarly to what we have done for $\hat{\beta}_j$, we substitute $V(\hat{\mu}_*)$ with $\hat{V}(\hat{\mu}_*)$, obtaining

$$\frac{\hat{\mu}_* - \mu_*}{\sqrt{\hat{V}(\hat{\mu}_*)}} \sim t_{n-2} \qquad \text{where } \hat{V}(\hat{\mu}_*) = S^2\left(\frac{1}{n} + \frac{(x_* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

Hence to perform inference we follow the same reasoning made for $\hat{\beta}_j$.
For example, a CI for μ_* is

$$\hat{y}_* \pm t_{n-2;1-\frac{\alpha}{2}}\sqrt{S^2\Bigg(\frac{1}{n} + \underbrace{\frac{(x_* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}_{\text{notice that the further } x_* \text{ is from } \bar{x}, \text{ the larger the CI will get}}\Bigg)}$$

If I compute several pointwise CIs for varying x_*, I obtain "confidence bands"
(careful: the level $(1-\alpha)$ only holds pointwise)



These methods can be useful to formalize practical questions, for example:

· what is a reasonable set of values for Y if $x = \tilde{x}$? → compute CI for $\tilde{\mu}$

· is μ_0 a reasonable value for Y if I observe $x = \tilde{x}$? → test $H_0: \tilde{\mu} = \mu_0$ $H_1: \tilde{\mu} \neq \mu_0$