CAPITOLUL 3. MODELUL DE REGRESIE LINIARĂ MULTIPLĂ

<table>
<tr><th colspan="2">ESTIMAREA</th><th>INTERPRETARE</th></tr>
<tr><td rowspan="2">punctuală a parametrilor $\boldsymbol{\beta_0}$ și $\boldsymbol{\beta_j}$ $(\boldsymbol{j = \overline{1,p}})$</td><td>$\boldsymbol{b_0}$</td><td>$\boldsymbol{b_0}$: nivelul mediu estimat al variabilei dependente (Y) atunci când variabilele independente $(X_1, X_2\ și\ X_3)$ iau simultan valoarea zero</td></tr>
<tr><td>$\boldsymbol{b_j}\ (\boldsymbol{j = \overline{1,p}})$

$\boldsymbol{b_j = \frac{\Delta Y}{\Delta X_j}}$</td><td>Exemplu: $p = 3 \Longrightarrow$ sunt 3 variabile independente $(X_1, X_2\ și\ X_3)$
$\boldsymbol{b_1}$: la o creștere a variabilei independente X_1 cu 1 unitate, variabila dependentă (Y) variază (scade sau crește în funcție de semn), în medie, cu $\boldsymbol{b_1}$, în condițiile în care influența celorlalte variabile independente $(X_2\ și\ X_3)$ se menține constantă
$\boldsymbol{b_2}$: la o creștere a variabilei independente X_2 cu 1 unitate, variabila dependentă (Y) variază (scade sau crește în funcție de semn), în medie, cu $\boldsymbol{b_2}$, în condițiile în care influența celorlalte variabile independente $(X_1\ și\ X_3)$ se menține constantă
$\boldsymbol{b_3}$: la o creștere a variabilei independente X_3 cu 1 unitate, variabila dependentă (Y) variază (scade sau crește în funcție de semn), în medie, cu $\boldsymbol{b_3}$, în condițiile în care influența celorlalte variabile independente $(X_1\ și\ X_2)$ se menține constantă</td></tr>
<tr><td rowspan="2">prin interval de încredere a parametrilor $\boldsymbol{\beta_0}$ și $\boldsymbol{\beta_j}$</td><td>$\mathrm{IC}(\beta_0)$: $\left[\boldsymbol{b_0 \pm t_{\alpha/2;n-k} s_{\hat{\beta}_0}}\right]$</td><td>Cu o probabilitate de $(1-\alpha)$, se poate garanta că parametrul β_0 este acoperit de intervalul $\left[b_0 \pm t_{\alpha/2;n-k} s_{\hat{\beta}_0}\right]$</td></tr>
<tr><td>$\mathrm{IC}(\beta_j)$: $\left[\boldsymbol{b_j \pm t_{\alpha/2;n-k} s_{\hat{\beta}_j}}\right]$
$(\boldsymbol{j = \overline{1,p}})$</td><td>Cu o probabilitate de $(1-\alpha)$, se poate garanta că parametrul β_j este acoperit de intervalul $\left[b_j \pm t_{\alpha/2;n-k} s_{\hat{\beta}_j}\right]$</td></tr>
<tr><td>coeficientului de corelaţie bivariată</td><td>între Y şi X_1: $\boldsymbol{r_{y1}}$
între Y şi X_2: $\boldsymbol{r_{y2}}$
între Y şi X_3: $\boldsymbol{r_{y3}}$</td><td>$\boldsymbol{r}$: indică sensul (după semn) și măsoară intensitatea (după valoarea în modul) legăturii dintre două variabile.
<table>
<tr><th>$\boldsymbol{|r| = 0}$</th><th>$\boldsymbol{0 \leftarrow |r|}$</th><th>$\boldsymbol{|r| \rightarrow\ 0,5\ \leftarrow |r|}$</th><th>$\boldsymbol{|r| \rightarrow 1}$</th><th>$\boldsymbol{|r| = 1}$</th></tr>
<tr><td>nu există o leg. liniară între Y și X</td><td>leg. liniară de intensitate slabă între Y și X</td><td>leg. liniară de intensitate moderată între Y și X</td><td>leg. liniară de intensitate puternică între Y și X</td><td>leg. liniară perfectă între Y și X</td></tr>
</table>
Exemplu:
$\boldsymbol{r_{y1} = 0,25}$: între Y și X_1 există o legătură liniară directă (după semn) și de intensitate slabă (după valoarea în modul $|r|$)</td></tr>
</table>

<table>
<tr><td></td><td></td><td>$\boldsymbol{r_{y2} = -0,57}$: între Y şi X_2 există o legătură liniară inversă (după semn) și de intensitate moderată (după valoarea în modul $|\boldsymbol{r}|$)
$\boldsymbol{r_{y3} = -0,87}$: între Y şi X_3 există o legătură liniară inversă (după semn) și de intensitate puternică (după valoarea în modul $|\boldsymbol{r}|$)</td></tr>
<tr><td>coeficientului de corelaţie parţială</td><td>între Y şi X_1: $\boldsymbol{r_{y1.23}}$
între Y şi X_2: $\boldsymbol{r_{y2.13}}$
între Y şi X_3: $\boldsymbol{r_{y3.12}}$</td><td>$\boldsymbol{r}$: indică sensul (după semn) și măsoară intensitatea (după valoarea în modul) legăturii dintre două variabile
<table>
<tr><th>$|\boldsymbol{r}| = \boldsymbol{0}$</th><th>$\boldsymbol{0} \leftarrow |\boldsymbol{r}|$</th><th>$|\boldsymbol{r}| \rightarrow \ \boldsymbol{0,5} \ \leftarrow |\boldsymbol{r}|$</th><th>$|\boldsymbol{r}| \rightarrow \boldsymbol{1}$</th><th>$|\boldsymbol{r}| = \boldsymbol{1}$</th></tr>
<tr><td>nu există o leg. liniară între Y și X</td><td>leg. liniară de intensitate slabă între Y și X</td><td>leg. liniară de intensitate moderată între Y și X</td><td>leg. liniară de intensitate puternică între Y și X</td><td>leg. liniară perfectă între Y și X</td></tr>
</table>
Exemplu:
$\boldsymbol{r_{y1.23} = 0,25}$: între Y şi X_1 există o legătură liniară directă (după semn) și de intensitate slabă (după valoarea în modul $|\boldsymbol{r}|$), în condițiile în care influența celorlalte variabile independente (X_2 și X_3) se menține constantă
$\boldsymbol{r_{y2.13} = -0,57}$: între Y şi X_2 există o legătură liniară inversă (după semn) și de intensitate moderată (după valoarea în modul $|\boldsymbol{r}|$), în condițiile în care influența celorlalte variabile independente (X_1 și X_3) se menține constantă
$\boldsymbol{r_{y3.12} = -0,87}$: între Y şi X_3 există o legătură liniară inversă (după semn) și de intensitate puternică (după valoarea în modul $|\boldsymbol{r}|$), în condițiile în care influența celorlalte variabile independente (X_1 și X_2) se menține constantă</td></tr>
<tr><td>raportului de determinaţie multiplă $\boldsymbol{\eta^2}$</td><td>$\boldsymbol{R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}}$</td><td>$\boldsymbol{R^2}$: măsoară cât din variația totală a variabilei dependente (Y) este explicată de modelul de regresie SAU de <u>variația simultană</u> a variabilelor independente
Exemplu:
$\boldsymbol{R^2 = 0,311}$ ($\boldsymbol{R^2}$% $= \boldsymbol{31,1}$%)
Variația variabilei dependente (Y) este explicată în proporție de 31,1% de <u>variația simultană</u> a variabilelor independente (X_1, X_2 și X_3). Restul de 68,9% reprezintă influența factori nespecificați în model (factorii reziduali).</td></tr>
<tr><td>raportului de corelaţie multiplă $\boldsymbol{\eta}$</td><td>$\boldsymbol{R = \sqrt{R^2}}$</td><td>$\boldsymbol{R}$: măsoară intensitatea legăturii liniare dintre variabile</td></tr>
</table>

<table>
<tr>
<td></td>
<td></td>
<td>
<table>
<tr><th>$R = 0$</th><th>$0 \leftarrow R$</th><th>$R \rightarrow\ \ 0,5\ \ \leftarrow R$</th><th>$R \rightarrow 1$</th><th>$R = 1$</th></tr>
<tr><td>nu există o leg. liniară între variabile</td><td>leg. liniară de intensitate slabă între variabile</td><td>leg. liniară de intensitate moderată între variabile</td><td>leg. liniară de intensitate puternică între variabile</td><td>leg. liniară perfectă între variabile între variabile</td></tr>
</table>
Exemplu:
$\boldsymbol{R = 0,558}$
Între variabila dependentă (Y) și variabilelor independente $(X_1, X_2\ și\ X_3)$ există o legătură liniară de intensitate moderată.
</td>
</tr>
<tr>
<td>raportului de determinaţie multiplă ajustat $\bar{\eta}^2$</td>
<td>$\bar{R}^2 = 1 - (1 - R^2)\dfrac{n-1}{n-k}$
SAU
$\bar{R}^2 = 1 - \dfrac{RSS}{TSS}\dfrac{n-1}{n-k}$</td>
<td>$\bar{R}^2$: măsoară mai precis cât din variația totală a variabilei dependente este explicat de modelul de regresie SAU de <u>variația simultană</u> a variabilelor independente.
Exemplu:
$\bar{R}^2 = 0,276$ $(\bar{R}^2\% = 27,6\%)$: Se interpretează la fel ca R^2, dar oferă un rezultat mai precis având în vedere că ține cont de numărul de variabile independente incluse în model. Așadar, variația variabilei dependente (Y) este explicată în proporție de 27,6% de <u>variația simultană</u> a variabilelor independente $(X_1, X_2\ și\ X_3)$. Restul de 62,4% reprezintă influența altor factori neincluși în model (factorii aleatori sau reziduali).</td>
</tr>
<tr>
<td>coeficienților standardizați</td>
<td>$\tilde{b}_j\ (j = \overline{1,p})$</td>
<td>Exemplu: $p = 3 \Longrightarrow$ sunt 3 variabile independente $(X_1, X_2\ și\ X_3)$
$\tilde{b}_1$: la o creștere a variabilei independente X_1 cu 1 <u>abatere standard</u>, variabila dependentă (Y) variază (scade sau crește în funcție de semn), în medie, cu b_1 <u>abateri standard</u>, în condițiile în care influența celorlalte variabile independente $(X_2\ și\ X_3)$ se menține constantă
$\tilde{b}_2$: la o creștere a variabilei independente X_2 cu 1 <u>abatere standard</u>, variabila dependentă (Y) variază (scade sau crește în funcție de semn), în medie, cu b_2 <u>abateri standard</u>, în condițiile în care influența celorlalte variabile independente $(X_1\ și\ X_3)$ se menține constantă
$\tilde{b}_3$: la o creștere a variabilei independente X_3 cu 1 <u>abatere standard</u>, variabila dependentă (Y) variază (scade sau crește în funcție de semn), în medie, cu b_3 <u>abateri standard</u>, în condițiile în care influența celorlalte variabile independente $(X_1\ și\ X_2)$ se menține constantă</td>
</tr>
</table>

<table>
<tr><th colspan="4">TESTAREA</th></tr>
<tr><td rowspan="4">Testarea constantei $\boldsymbol{\beta_0}$ și a coeficienților de regresie parțiali $\boldsymbol{\beta_j}$ $(\boldsymbol{j = \overline{1,p}})$</td><td>Ipoteze:</td><td>$H_0: \beta_0 = 0$ (parametrul β_0 nu diferă semnificativ de 0 SAU constanta modelului nu este semnificativă statistic)
$H_1: \beta_0 \neq 0$ (parametrul β_0 diferă semnificativ de 0 SAU constanta modelului este semnificativă statistic)</td><td>$H_0: \beta_j = 0$ (coeficientul de regresie parțial β_j nu diferă semnificativ de 0, ceea ce înseamnă că variabila independentă X_j nu are o influență parțială semnificativă asupra variabilei dependente Y SAU X_j nu explică semnificativ variația lui Y, în condițiile în care celelalte variabile independente se mențin constante)
$H_1: \beta_j \neq 0$ (coeficientul de regresie parțial β_j diferă semnificativ de 0, ceea ce înseamnă că variabila independentă X_j are o influență parțială semnificativă asupra variabilei dependente Y SAU X_j explică semnificativ variația lui Y, în condițiile în care celelalte variabile independente se mențin constante)</td></tr>
<tr><td>Valoarea teoretică a statisticii test Student:</td><td colspan="2">$\boldsymbol{t_{teoretic} = t_{\alpha/2;n-k}}$</td></tr>
<tr><td>Valoarea calculată a statisticii test Student:</td><td>$(\beta_0): \boldsymbol{t_{calc}} = \dfrac{\boldsymbol{b_0}}{\boldsymbol{s_{\hat{\beta}_0}}}$</td><td>$(\beta_j): \boldsymbol{t_{calc}} = \dfrac{\boldsymbol{b_j}}{\boldsymbol{s_{\hat{\beta}_j}}}$</td></tr>
<tr><td>Decizia:</td><td colspan="2">Dacă se ţine cont de valoarea calculată a testului, regula de decizie este următoarea:
- dacă $|\boldsymbol{t_{calc}}| \leq \boldsymbol{t_{\alpha/2;\,n-k}}$, nu se respinge ipoteza nulă (H_0)
- dacă $|\boldsymbol{t_{calc}}| > \boldsymbol{t_{\alpha/2;\,n-k}}$, se respinge ipoteza nulă (H_0), în condițiile unui risc α
Dacă se ţine cont de semnificaţia testului, regula de decizie este următoarea:
- dacă $\boldsymbol{Sigt \geq \alpha}$, nu se respinge ipoteza nulă (H_0)
- dacă $\boldsymbol{Sigt < \alpha}$, se respinge H_0, în condițiile unui risc α</td></tr>
<tr><td>Testarea coeficienților de corelație bivariați (de ex. $\boldsymbol{\rho_{y1}}$) și a coeficienților de corelație parțiali (de ex. $\boldsymbol{\rho_{y1.23}}$)</td><td>Ipoteze:</td><td>$H_0: \rho_{y1} = 0$ (coeficientul de corelație bivariată ρ_{y1} nu diferă semnificativ de 0, ceea ce înseamnă că între cele două variabile (Y și X_1) nu există o legătură liniară semnificativă SAU cele două variabile (Y și X_1) nu sunt corelate semnificativ)
$H_1: \rho_{y1} \neq 0$ (coeficientul de corelație bivariată ρ_{y1} diferă semnificativ de 0, ceea ce înseamnă că între cele două variabile (Y și</td><td>$H_0: \rho_{y1.2} = 0$ (coeficientul de corelație parțială $\rho_{y1.23}$ nu diferă semnificativ de 0, ceea ce înseamnă că între cele două variabile (Y și X_1) nu există o legătură liniară semnificativă, în condițiile în care celelalte variabile independente (X_2 și X_3) se menține constant SAU cele două variabile (Y și X_1) nu sunt corelate semnificativ, în condițiile în care celelalte variabile independente (X_2 și X_3) se menține constant)</td></tr>
</table>

		X_1) există o legătură liniară semnificativă **SAU** cele două variabile (Y și X_1) sunt corelate semnificativ)	$H_1: \rho_{y1.2} \neq 0$ (coeficientul de corelație parțială $\rho_{y1.23}$ diferă semnificativ de 0, **ceea ce înseamnă** că între cele două variabile (Y și X_1) există o legătură liniară semnificativă, în condițiile în care celelalte variabile independente (X_2 și X_3) se menține constant **SAU** cele două variabile (Y și X_1) sunt corelate semnificativ, în condițiile în care celelalte variabile independente (X_2 și X_3) se menține constant)
	Valoarea teoretică a statisticii test ***Student***:	$\boldsymbol{t_{teoretic} = t_{\alpha/2;n-2}}$	$\boldsymbol{t_{teoretic} = t_{\alpha/2;n-k}}$
	Valoarea calculată a statisticii test ***Student***:	$\boldsymbol{t_{calc} = \dfrac{r_{y1}}{\sqrt{\dfrac{1-r_{y1}^2}{n-2}}}}$	$\boldsymbol{t_{calc} = \dfrac{r_{y1.23}}{\sqrt{\dfrac{1-r_{y1.23}^2}{n-k}}}}$
	Decizia:	Dacă se ține cont de valoarea calculată a testului, regula de decizie este următoarea: - dacă $\lvert \boldsymbol{t_{calc}} \rvert \leq \boldsymbol{t_{\alpha/2;\,n-k}}$, nu se respinge ipoteza nulă (H_0) - dacă $\lvert \boldsymbol{t_{calc}} \rvert > \boldsymbol{t_{\alpha/2;\,n-k}}$, se respinge ipoteza nulă (H_0), în condițiile unui risc α Dacă se ține cont de semnificația testului, regula de decizie este următoarea: - dacă $\boldsymbol{Sigt} \geq \boldsymbol{\alpha}$, nu se respinge ipoteza nulă (H_0) - dacă $\boldsymbol{Sigt} < \boldsymbol{\alpha}$, se respinge H_0, în condițiile unui risc α	
Testarea modelului de regresie și a raportului de determinație $\boldsymbol{\eta^2}$ (sau raportului de corelație $\boldsymbol{\eta}$)	Ipoteze:	$H_0: \beta_0 = 0 \text{ și } \beta_1 = 0 \text{ și } \beta_2 = 0$ (toți coeficienții de regresie sunt simultan 0, **ceea ce înseamnă** că modelul de regresie nu explică semnificativ dependența liniară dintre variabile **SAU** modelul de regresie construit nu este corect specificat) $H_1: \beta_1 \neq 0 \text{ sau } \beta_2 \neq 0$ (există cel puțin un coeficient de regresie semnificativ statistic, **ceea ce înseamnă** că există cel puțin un model de regresie semnificativ statistic **SAU** modelul de regresie explică semnificativ dependența liniară dintre variabile)	$H_0: \eta = 0$ (raportul de determinație η^2 sau raportul de corelația η nu diferă semnificativ de 0, **ceea ce înseamnă** că între variabile nu există o legătură liniară semnificativă) $H_1: \eta > 0$ (raportul de determinație η^2 sau raportul de corelația η este semnificativ mai mare decât 0, **ceea ce înseamnă** că între variabile există o legătură liniară semnificativă)
	Valoarea teoretică a statisticii test ***Fisher***:	$\boldsymbol{F_{teoretic} = F_{\alpha;k-1;n-k}}$	

	Valoarea calculată a statisticii test ***Fisher***:	$F_{calc} = \frac{\boldsymbol{ESS}}{\boldsymbol{RSS}}\frac{\boldsymbol{n-k}}{\boldsymbol{k-1}}$	$F_{calc} = \frac{\boldsymbol{R^2}}{\boldsymbol{1-R^2}}\frac{\boldsymbol{n-k}}{\boldsymbol{k-1}}$
	Decizia:	Dacă se ţine cont de valoarea calculată a testului, regula de decizie este următoarea: - dacă $\boldsymbol{F_{calc} \leq F_{\alpha;\, k-1;\, n-k}}$, nu se respinge ipoteza nulă (H_0) - dacă $\boldsymbol{F_{calc} > F_{\alpha;\, k-1;\, n-k}}$, se respinge ipoteza nulă (H_0), cu în condițiile unui risc α Dacă se ţine cont de semnificaţia testului, regula de decizie este următoarea: - dacă $\boldsymbol{SigF \geq \alpha}$, nu se respinge ipoteza nulă (H_0) - dacă $\boldsymbol{SigF < \alpha}$, se respinge H_0, în condițiile unui risc α	
Testarea influenţei marginale a unei **variabile incluse** în model şi a influenţei marginele a unei **variabile excluse** din model	Ipoteze:	H_0: variabila introdusă în model nu are o influenţă marginală semnificativă asupra variaţiei variabilei dependente **SAU** includerea variabilei nu produce o modificare semnificativă asupra raportului de determinație **SAU** includerea variabilei nu are un efect semnificativ asupra capacităţii explicative a modelului **SAU** modelul nou (în care este inclusă noua variabilă) nu este mai bun decât modelul vechi (modelul inițial), **ceea ce înseamnă** că se ia decizia de a nu include această variabilă independentă în model H_1: variabila introdusă în model are o influenţă marginală semnificativă asupra variaţiei variabilei dependente **SAU** includerea variabilei produce o modificare semnificativă asupra raportului de determinație **SAU** includerea variabilei are un efect semnificativ asupra capacităţii explicative a modelului **SAU** modelul nou (în care este inclusă noua variabilă) este mai bun decât modelul vechi (modelul inițial), **ceea ce înseamnă** că se ia decizia de a include această variabilă independentă în model	H_0: variabila exclusă din model nu are o influenţă marginală semnificativă asupra variaţiei variabilei dependente **SAU** excluderea variabilei nu produce o modificare semnificativă asupra raportului de determinație **SAU** excluderea variabilei nu are un efect semnificativ asupra capacităţii explicative a modelului **SAU** modelul nou (din care este exclusă variabila) este mai bun decât modelul vechi (modelul inițial), **ceea ce înseamnă** că se ia decizia de a exclude această variabilă independentă din model inițial H_1: variabila exclusă din model are o influenţă marginală semnificativă asupra variaţiei variabilei dependente **SAU** excluderea variabilei produce o modificare semnificativă asupra raportului de determinație **SAU** excluderea variabilei are un efect semnificativ asupra capacităţii explicative a modelului **SAU** modelul nou (din care este exclusă variabila) nu este mai bun decât modelul vechi (modelul inițial), **ceea ce înseamnă** că se ia decizia de a nu exclude această variabilă independentă din modelul inițial
	Decizia:	Dacă se ţine cont de semnificaţia testului, regula de decizie este următoarea: - dacă $\boldsymbol{SigFChange \geq \alpha}$, nu se respinge ipoteza nulă (H_0) - dacă $\boldsymbol{SigFChange < \alpha}$, se respinge H_0, în condițiile unui risc α unde $SigFChange$ aparţine modelului nou	