RECAPITULARE BAZELE STATISTICII

1. Probabilităţi

1. Ştiind că variabila $X \sim N(10, 4)$, calculaţi probabilităţile: $P(8 < X < 12)$ şi $P(X > 12)$.

Rezolvare:

$$Z = \frac{X - \mu}{\sigma}$$

$$z_1 = \frac{8 - 10}{2} = -1$$

$$z_2 = \frac{12 - 10}{2} = 1$$

$$P(8 \leq X \leq 12) = P(-1 \leq Z \leq 1) = P(-1 \leq Z \leq 0) + P(0 \leq Z \leq 1) =$$
$$= \phi(-1) + \phi(1) = 2 \cdot \phi(1) = 2 \cdot 0{,}3413 = 0{,}6826$$

$$P(X > 12) = P(Z > 1) = 0{,}5 - P(0 \leq Z \leq 1) = 0{,}5 - \phi(1) = 0{,}5 - 0{,}3413 = 0{,}1587$$

Observație: Valoarea lui $\phi(1)$ se citește din tabela Laplace.

2. Estimarea punctuală a parametrilor unei populații

Parametrul (la nivelul populației)		**Estimarea punctuală a parametrului folosind datele de la nivelul eșantionului**	
Media (μ)	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	Estimația punctuală a mediei μ ($\bar{x}$)	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Proporția (π)	$\pi = \frac{N_a}{N}$	Estimația punctuală a proporției π (p)	$p = \frac{n_a}{n}$
Varianța (σ^2)	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	Estimația punctuală a varianței σ^2 (s^2)	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_j)^2}{n}$
		Estimația punctuală a varianței modificate σ^2 (s'^2)	$s'^2 = \frac{n}{n-1} s^2$ **sau** $s'^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_j)^2}{n-1}$
Abaterea standard (σ)	$\sigma = \sqrt{\sigma^2}$	Estimația punctuală a abaterii standard σ (s)	$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_j)^2}{n}}$
		Estimația punctuală a abaterii standard modificate σ (s')	$s' = \sqrt{s'^2} = \sqrt{\frac{n}{n-1} s^2}$ **sau** $s' = \sqrt{s'^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_j)^2}{n-1}}$
Observații:	N: volumul populației N_a: volumul (sub)populației care îndeplinește caracteristica a	Observații:	n: volumul eșantionului n_a: volumul (sub)eșantionului care îndeplinește caracteristica a

3. Estimarea prin interval de încredere a parametrilor unei populații

Interval de încredere	Medie (μ)	Proporție (π)
se cunoaște σ	$IC(\mu)$: $\left[\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}};\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$	$IC(\pi)$: $\left[p-z_{\alpha/2}\frac{\sigma}{\sqrt{n}};p+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$
	$\phi(z_{\alpha/2})=\frac{1-\alpha}{2}\Rightarrow z_{\alpha/2}=linie+coloana\ (Tabela\ Laplace)$	
	$IC(\mu)$: $[\bar{x}-z_{\alpha/2}\sigma_{\hat{\mu}};\bar{x}+z_{\alpha/2}\sigma_{\hat{\mu}}]$ $\sigma_{\hat{\mu}}=\frac{\sigma}{\sqrt{n}}$ $\sigma_{\hat{\mu}}$: eroarea medie de reprezentativitate a mediei	$IC(\pi)$: $[p-z_{\alpha/2}\sigma_{\hat{\pi}};p+z_{\alpha/2}\sigma_{\hat{\pi}}]$ $\sigma_{\hat{\pi}}=\frac{\sigma}{\sqrt{n}}$ $\sigma_{\hat{\pi}}$: eroarea medie de reprezentativitate a proporției
	$IC(\mu)$: $[\bar{x}-\Delta_{\hat{\mu}};\bar{x}+\Delta_{\hat{\mu}}]$ $\Delta_{\hat{\mu}}=\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}=\pm z_{\alpha/2}\sigma_{\hat{\mu}}$ $\Delta_{\hat{\mu}}$: eroarea maximă admisibilă a mediei	$IC(\pi)$: $[p-\Delta_{\hat{\pi}};p+\Delta_{\hat{\pi}}]$ $\Delta_{\hat{\pi}}=\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}=\pm z_{\alpha/2}\sigma_{\hat{\pi}}$ $\Delta_{\hat{\pi}}$: eroarea maximă admisibilă a proporției
nu se cunoaște σ	$IC(\mu)$: $\left[\bar{x}-t_{\alpha/2;n-1}\frac{s'}{\sqrt{n}};\bar{x}+t_{\alpha/2}\frac{s'}{\sqrt{n}}\right]$	$IC(\pi)$: $\left[p-t_{\alpha/2;n-1}\frac{s}{\sqrt{n}};p+t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$
	$s'=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$	$s=\sqrt{p(1-p)}$
	$IC(\mu)$: $[\bar{x}-t_{\alpha/2;n-1}s_{\hat{\mu}};\bar{x}+t_{\alpha/2;n-1}s_{\hat{\mu}}]$ $s_{\hat{\mu}}=\frac{s'}{\sqrt{n}}$ $s_{\hat{\mu}}$: eroarea medie de reprezentativitate a mediei	$IC(\pi)$: $[p-t_{\alpha/2;n-1}s_{\hat{\pi}};p+t_{\alpha/2;n-1}s_{\hat{\pi}}]$ $s_{\hat{\pi}}=\frac{s}{\sqrt{n}}=\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ $s_{\hat{\pi}}$: eroarea medie de reprezentativitate a proporției
	$IC(\mu)$: $[\bar{x}-\Delta_{\hat{\mu}};\bar{x}+\Delta_{\hat{\mu}}]$ $\Delta_{\hat{\mu}}=\pm t_{\alpha/2;n-1}\frac{s'}{\sqrt{n}}=\pm t_{\alpha/2;n-1}s_{\hat{\mu}}$ $\Delta_{\hat{\mu}}$: eroarea maximă admisibilă a mediei	$IC(\pi)$: $[p-\Delta_{\hat{\pi}};p+\Delta_{\hat{\pi}}]$ $\Delta_{\hat{\pi}}=\pm t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}=\pm t_{\alpha/2;n-1}s_{\hat{\pi}}$ $\Delta_{\hat{\mu}}$: eroarea maximă admisibilă a proporției

Aplicații

1. Din totalul judeţelor României se extrage aleator repetat un eşantion de 5 judeţe, care au fost observate după câştigul salarial net. La nivelul eşantionului s-au obţinut: $\sum x_i = 100$, $\sum (x_i - \bar{x})^2 = 8$. Se cere să se estimeze prin interval de încredere câştigul mediu la nivelul tuturor judeţelor, considerând o probabilitate de 95% ($\alpha = 0.05$).

Rezolvare:

$$IC(\mu): \left[\bar{x} - t_{\alpha/2;n-1} \cdot \frac{s'}{\sqrt{n}}; \bar{x} + t_{\alpha/2;n-1} \cdot \frac{s'}{\sqrt{n}}\right]$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{100}{5} = 20$$

$$s' = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{8}{4}} = 1{,}41$$

$$t_{\alpha/2;n-1} = t_{0{,}025;4} = 2{,}776$$

Observație: Valoarea lui $t_{0{,}025;4}$ se citește din tabela Student, se găsește la intersecția coloanei 0,025 cu linia 4.

$$IC(\mu): \left[20 - 2{,}776 \cdot \frac{1{,}41}{\sqrt{5}}; 20 + 2{,}776 \cdot \frac{1{,}41}{\sqrt{5}}\right]$$

$$IC(\mu): [18{,}24; 21{,}75]$$

Interpretare: Cu o probabilitate de 95%, se garantează că media populaţiei, μ, (din care a fost extras eşantionul) este acoperită de intervalul $[18{,}24; 21{,}75]$.

2. Dintr-un lot de 1000 de piese se extrage aleator repetat un eşantion de 50 de piese, care au fost observate după calitate. În urma observării a rezultat un număr de 7 piese defecte. Se cere să se estimeze prin interval de încredere proporţia pieselor defecte la nivelul întregului lot, probabilitatea cu care se garantează rezultatul fiind de 99%.

Rezolvare:

$$IC(\pi): \left[p - t_{\alpha/2;n-1} \cdot \frac{s}{\sqrt{n}}; p + t_{\alpha/2;n-1} \cdot \frac{s}{\sqrt{n}}\right]$$

$$p = \frac{7}{50} = 0{,}14$$

$$s = \sqrt{p(1-p)} = \sqrt{0{,}14(1 - 014)} = 0{,}346$$

$$t_{\alpha/2;n-1} = t_{0{,}005;49} = 2{,}576$$

$$IC(\pi): \left[0{,}14 - 2{,}576 \cdot \frac{0{,}346}{\sqrt{50}}; 0{,}14 + 2{,}576 \cdot \frac{0{,}346}{\sqrt{50}}\right]$$

$$IC(\pi): [0{,}013; 0{,}266]$$

Interpretare: Cu o probabilitate de 99%, se garantează că ponderea pieselor defecte, π, este acoperită de intervalul $[1{,}3\%; 26{,}6\%]$.

4. Testare parametrilor unei populații

1. Din totalul autoturismelor vândute într-un an de o firmă a fost extras un eşantion de 1000 de autoturisme pentru care a fost înregistrat preţul (mii euro). Datele cunoscute sunt: $\bar{x} = 28, \sigma = 4$. Pentru o probabilitate de 90%, se cere să se testeze dacă există diferenţe semnificative între preţul mediu al lotului de autoturisme vândute în anul curent şi preţul mediu înregistrat anul trecut, de 30 mii euro (μ_0).

Paşii testării statistice	**Cazul în care se cunoaşte σ**
1. Formularea ipotezelor	$\boldsymbol{H_0: \mu = 30\ (\mu - 30 = 0)}$: între preţul mediu al lotului de autoturisme vândute în anul curent şi preţul mediu înregistrat anul trecut **nu există** diferențe semnificative $\boldsymbol{H_1: \mu \neq 30\ (\mu - 30 \neq 0)}$: între preţul mediu al lotului de autoturisme vândute în anul curent şi preţul mediu înregistrat anul trecut **există** diferențe semnificative
2. Alegerea riscului asumat	$\alpha = 0{,}10$
3. Alegerea testului statistic	$Z = \dfrac{\hat{\mu} - \mu_0}{\sigma/\sqrt{n}} \sim N(0{,}1)$
4. Determinarea valorii teoretice a testului statistic	$\phi(z_{\alpha/2}) = \dfrac{1-\alpha}{2} = \dfrac{1-0{,}10}{2} = 0{,}45 \Rightarrow$ $\Rightarrow z_{teoretic} = z_{\alpha/2} = 1{,}6 + 0{,}05 = 1{,}65$
5. Determinarea valorii statisticii test	$z_{calc} = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{28 - 30}{4/\sqrt{1000}} = -15{,}87$
6. Regula de decizie	$\lvert z_{calc} \rvert \leq z_{\alpha/2} \Rightarrow A.H_0(1-\alpha)$ $\lvert z_{calc} \rvert > z_{\alpha/2} \Rightarrow R.H_0$
7. Luarea deciziei	$\lvert z_{calc} \rvert = 15{,}87 > z_{\alpha/2} = 1{,}65 \Rightarrow R.H_0\ (90\%)$
8. Interpretare	Cu o probabilitate de 90%, se poate garanta că între preţul mediu al lotului de autoturisme vândute în anul curent şi preţul mediu înregistrat anul trecut există diferențe semnificative

2. Pentru o cercetare de sondaj asupra consumatorilor unui produs se cunoaşte că media de vârstă a consumatorilor din eşantion este de 38 de ani. Pentru un risc asumat de 5% şi o valoare Sig. de 0,008 se cere să se verifice dacă există diferenţe semnificative de vârstă între consumatorii produsului şi o valoare stabilită prin strategia de marketing, de 40 de ani.

Paşii testării statistice	**Cazul în care nu se cunoaşte σ**
1. Formularea ipotezelor	$\boldsymbol{H_0: \mu = 38\ (\mu - 38 = 0)}$: între vârsta medie a populaţiei μ şi valoarea medie de referință de 38 de ani **nu există** diferențe semnificative sau vârsta medie a populaţiei μ **nu diferă** semnificativ de 38 de ani

	$\boldsymbol{H_1}$: $\boldsymbol{\mu \neq 38}$ $(\boldsymbol{\mu - 38 \neq 0})$: între vârsta medie a populației μ și valoarea medie de referință de 38 de ani **există** diferențe semnificative sau vârsta medie a populației μ **diferă** semnificativ de 38 de ani
2. Alegerea riscului asumat	$\alpha = 0,05$
6. Regula de decizie	$\lvert t_{calc} \rvert \leq t_{\alpha/2;n-1} \Leftrightarrow Sig \geq \alpha \Rightarrow A.H_0(1-\alpha)$ $\lvert t_{calc} \rvert > t_{\alpha/2;n-1} \Leftrightarrow Sig < \alpha \Rightarrow R.H_0$
7. Luarea deciziei	$Sig = 0.008 < \alpha = 0.05 > \Rightarrow R.H_0(95\%)$
8. Interpretare	Cu o probabilitate de 95% se poate garanta că între vârsta medie a populației μ și valoarea medie de referință de 38 de ani există diferențe semnificative.

3. Pentru două eşantioane extrase din două populaţii cu varianţe diferite, se cunosc $n_1 = n_2 = 100$, $\bar{x}_1 = 42, \bar{x}_2 = 37, s'_1 = 4, s'_2 = 7$. Pentru o probabilitate de 95%, se cere să se verifice dacă există diferenţe semnificative între mediile celor două populaţii.

Pașii testării statistice	**Cazul în care nu se cunosc $\boldsymbol{\sigma_1}$ și $\boldsymbol{\sigma_2}$**	
1. Formularea ipotezelor	$\boldsymbol{H_0}$: $\boldsymbol{\mu_1 = \mu_2}$ $(\boldsymbol{\mu_1 - \mu_2 = 0})$**:** între mediile de la nivelor celor două populații **nu există** diferențe semnificative sau μ_1 **nu diferă** semnificativ de μ_2 $\boldsymbol{H_1}$: $\boldsymbol{\mu_1 \neq \mu_2}$ $(\boldsymbol{\mu_1 - \mu_2 \neq 0})$**:** între mediile de la nivelor celor două populații **există** diferențe semnificative sau μ_1 **diferă** semnificativ de μ_2	
2. Alegerea riscului asumat	$\alpha = 0,05$	
3. Alegerea testului statistic	$(\sigma_1 \neq \sigma_2)$	$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1'^2}{n_1} + \frac{\hat{\sigma}_2'^2}{n_2}}} \sim t(n_1 + n_2 - 2)$
4. Determinarea valorii teoretice a testului statistic	$t_{teoretic} = t_{\alpha/2;n_1+n_2-2} = t_{0,025;198} = 1,96$	
5. Determinarea valorii statisticii test	$(\sigma_1 \neq \sigma_2)$	$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1'^2}{n_1} + \frac{s_2'^2}{n_2}}} = \frac{42 - 37}{\sqrt{\frac{16}{100} + \frac{49}{100}}} = \frac{5}{\sqrt{\frac{16}{100} + \frac{49}{100}}} = 6,2$
6. Regula de decizie	$\lvert t_{calc} \rvert \leq t_{\alpha/2;n_1+n_2-2} \Rightarrow A.H_0(1-\alpha)$ $\lvert t_{calc} \rvert > t_{\alpha/2;n_1+n_2-2} \Rightarrow R.H_0$	
7. Luarea deciziei	$\lvert t_{calc} \rvert = 6,2 > t_{\alpha/2;n_1+n_2-2} = 1,96 \Rightarrow R.H_0(95\%)$	
8. Interpretare	Cu o probabilitate de 95%, se poate garanta că între mediile de la nivelor celor două populații există diferențe semnificative.	