Parametrické úlohy
o jednom náhodném výběru z normálního rozložení

Úkol: Testování hypotézy o parametru μ normálního rozložení

Systematická chyba měřicího přístroje se eliminuje nastavením přístroje a měřením etalonu, jehož správná hodnota je $\mu = 10{,}00$. Nezávislými měřeními za stejných podmínek byly získány hodnoty: 10,24 10,12 9,91 10,19 9,78 10,14 9,86 10,17 10,05, které považujeme za realizace náhodného výběru rozsahu 9 z rozložení $N(\mu, \sigma^2)$. Je možné při riziku 0,05 vysvětlit odchylky od hodnoty 10,00 působením náhodných vlivů?

Návod:

Na hladině významnosti 0,05 testujeme hypotézu H_0: $\mu = 10$ proti oboustranné alternativě H_1: $\mu \neq 10$. Jde o úlohu na jednovýběrový t-test. Ten je ve STATISTICE implementován. Vytvoříme datový soubor o jedné proměnné a devíti případech, kam zapíšeme naměřené hodnoty. V Základních statistikách/tabulkách vybereme t-test, samostatný vzorek. Do Referenčních hodnot zapíšeme 10. Ve výstupu se podíváme na hodnotu testového kritéria a na p-hodnotu. Pokud p-hodnota bude menší nebo rovna 0,05, zamítneme hypotézu H_0: $\mu = 10$ ve prospěch oboustranné alternativní hypotézy H_1: $\mu \neq 10$ na hladině významnosti 0,05.
V opačném případě H_0 nezamítáme. V našem případě je

	Test průměrů vůči referenční konstantě (hodnotě)							
Proměnná	Průměr	Sm.odch.	N	Sm.chyba	Referenční konstanta	t	SV	p
Prom1	10,05111	0,162669	9	0,054223	10,00000	0,942611	8	0,373470

Protože p-hodnota 0,373470 > 0,05 nulovou hypotézu nezamítáme na hladině významnosti 0,05. S rizikem omylu nejvýše 5% lze tedy odchylky od hodnoty 10 vysvětlit působením náhodných vlivů.

Všimněme si ještě hodnoty testového kriteria: $t_0 = 0{,}942611$. Kritický obor

$$W = \left(-\infty, -t_{1-\alpha/2}(n-1)\right\rangle \cup \left\langle t_{1-\alpha/2}(n-1), \infty\right) = \left(-\infty, -t_{0{,}975}(8)\right\rangle \cup \left\langle t_{0{,}975}(8), \infty\right) =$$
$$= \left(-\infty, -2{,}306\right\rangle \cup \left\langle 2{,}306, \infty\right)$$

Protože $t_0 \notin W$, nezamítáme na hladině významnosti 0,05 hypotézu H_0.

TESTOVÁNÍ NORMALITY

Kolmogorovův – Smirnovův test normality dat

Testujeme nulovou hypotézu, která tvrdí, že náhodný výběr $X_1, \ldots, X_n$ pochází z normálního rozložení s parametry μ a σ^2.

Poznámka ke K-S testu ve STATISTICE

Jestliže p-hodnota $\leq \alpha$, pak H_0 zamítáme na hladině významnosti α, je-li p-hodnota > α, pak H_0 nezamítáme na hladině významnosti α.) První p-hodnota se vztahuje k případu, kdy střední hodnotu μ a rozptyl σ^2 známe předem, druhá (ozn. Lilieforsovo p) se vztahuje k případu, kdy μ a σ^2 neznáme. Objeví-li se ve výstupu p = n.s. (tj. non significant), pak hypotézu o normalitě nezamítáme na hladině významnosti 0,05.

Shapirův – Wilkův test normality dat

Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, \ldots, X_n$ pochází z rozložení $N(\mu, \sigma^2)$. Test je založen na zjištění, zda body v Q-Q grafu jsou významně odlišné od regresní přímky proložené těmito body.

Úkol 1. : U 45 studentek VŠE v Praze byla zjišťována výška a obor studia (1 – národní hospodářství, 2 – informatika). Hodnoty jsou uloženy v souboru 06_data.sta. Pomocí Lilieforsovy modifikace K-S testu, pomocí S-W testu a pomocí testu dobré shody testujte na hladině významnosti 0,05 hypotézu, že data pocházejí z normálního rozložení. Pomocí N-P grafu posuďte vizuálně předpoklad normality.

Návod:

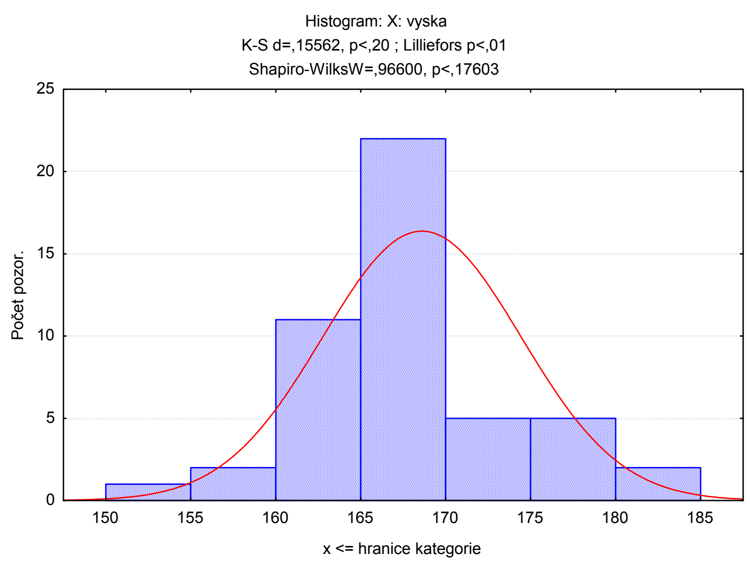
1. způsob provedení Lilieforsova a S-W testu: Statistiky – Základní statistiky/tabulky – Tabulky četností – OK – Proměnné X – OK – Normalita – zaškrtneme Lilieforsův test a S-W test – Testy normality.

	Testy normality (vyska.sta)				
Proměnná	N	max D	Lilliefors p	W	p
X: vyska	48	0,155621	p < ,01	0,965996	0,176031

Výstupní tabulka obsahuje počet pozorování, hodnotu testové statistiky Lilieforsovy modifikace K-S testu (max D = 0,155621), p-hodnotu ($p < 0{,}01$), testovou statistiku S-W testu (W = 0,965996) a odpovídající p-hodnotu (p = 0,176031). Vidíme, že Lilieforsův test zamítá hypotézu o normalitě na hladině významnosti 0,05, zatímco S-W test nikoli.

2. způsob provedení Lilieforsova a S-W testu: Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnné X – OK – Normalita – zaškrtneme K-S test & Lilieforsův test a S-W test – Tabulky četností (nebo Histogram).

	Tabulka četností:X: vyska (vyska.sta) K-S d=,15562, p<,20 ; Lilliefors p<,01 Shapiro-WilksW=,96600, p<,17603					
Kategorie	Četnost	Kumulativní četnost	Rel.četn. (platných)	Kumul. % (platných)	Rel.četn. všech	Kumul. % všech
150,0000<x<=155,0000	1	1	2,08333	2,0833	2,08333	2,0833
155,0000<x<=160,0000	2	3	4,16667	6,2500	4,16667	6,2500
160,0000<x<=165,0000	11	14	22,91667	29,1667	22,91667	29,1667
165,0000<x<=170,0000	22	36	45,83333	75,0000	45,83333	75,0000
170,0000<x<=175,0000	5	41	10,41667	85,4167	10,41667	85,4167
175,0000<x<=180,0000	5	46	10,41667	95,8333	10,41667	95,8333
180,0000<x<=185,0000	2	48	4,16667	100,0000	4,16667	100,0000
ChD	0	48	0,00000		0,00000	100,0000



V tomto případě dostaneme v záhlaví tabulky či histogramu stejné informace jako pomocí předešlého způsobu.

Příklady k samostatnému řešení

Příklad 1.: Měřením délky deseti válečků byly získány hodnoty (v mm): 5,38 5,36 5,35 5,40 5,41 5,34 5,29 5,43 5,42 5,32. Těchto deset hodnot považujeme za realizace náhodného výběru rozsahu 10 z normálního rozložení N(μ, σ^2).
Na hladině významnosti 0,01 testujte hypotézu, že střední hodnota délky válečků je 5,3 mm proti oboustranné alternativě.

$\neq$

Příklad 2.: Testy normality a grafické ověření normality proveďte jak pro výšky studentek oboru národní hospodářství, tak pro výška studentek oboru informatiky.

Proměnná	Testy normality (vyska.sta) Zhrnout podmínku: z=1 N	max D	Lilliefors p	W	p
X: vyska	28	0,167473	p < ,05	0,970969	0,606793

Proměnná	Testy normality (vyska.sta) Zhrnout podmínku: z=2 N	max D	Lilliefors p	W	p
X: vyska	20	0,172301	p < ,15	0,922747	0,111924

Příklad 3.: Bylo náhodně vybráno 15 desetiletých chlapců a byla zjištěna jejich výška (v cm). Výsledky měření 130, 140, 136, 141, 139, 133, 149, 151, 139, 136, 138, 142, 127, 139, 147 považujeme za realizace náhodného výběru rozsahu 15 z rozložení N(μ,σ^2). Podle názoru odborníků by střední hodnoty výšky desetiletých chlapců měla být 136,1 cm. Testujte tuto hypotézu na hladině významnosti 0,05.
Pomocí N-P plotu a S-W testu ověřte normalitu dat.
Výsledky: