Cvičení 4.: Základní pojmy matematické statistiky II, testy normality

Úkol: 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 s neznámou střední hodnotou μ a známou směrodatnou odchylkou $\sigma = 0,04$.
Na hladině významnosti 0,1 testujte nulovou hypotézu, že střední hodnota délky válečků je 5,35 mm. Proti nulové hypotéze postavte
a) oboustrannou alternativu
b) levostrannou alternativu
c) pravostrannou alternativu.
Test proveďte pomocí
a) kritického oboru
b) intervalu spolehlivosti
c) p-hodnoty.
Systém STATISTICA použijte jako inteligentní kalkulačku.

Návod:
Formulace nulové hypotézy: H_0: $\mu = 5,35$, formulace alternativní hypotézy:
ad a) H_1: $\mu \neq 5,35$, ad b) H_1: $\mu < 5,35$, ad c) H_1: $\mu > 5,35$
Jedná se o jednovýběrový z-test.
Testová statistika $T_0 = \dfrac{M - c}{\frac{\sigma}{\sqrt{n}}}$ bude mít rozložení N(0, 1), pokud je nulová hypotéza pravdivá.

Provedení testu:

Ad a) Pomocí kritického oboru
Kritický obor pro oboustrannou alternativu:
$W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$.
Kritický obor pro levostrannou alternativu:
$W = \left\langle -\infty, -u_{1-\alpha}\right)$.
Kritický obor pro pravostrannou alternativu:
$W = \left\langle u_{1-\alpha}, \infty\right)$.
Pokud $t_0 \in W$, H_0 zamítáme na hladině významnosti α.

Ad b) Pomocí intervalu spolehlivosti
Oboustranný interval spolehlivosti pro μ při známém σ:
$(d, h) = (m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2})$.
Pravostranný interval spolehlivosti pro μ při známém σ:
$(-\infty, h) = (-\infty, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha})$.
Levostranný interval spolehlivosti pro μ při známém σ:
$(d, \infty) = (m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, \infty)$.

Pokud číslo c (v našem případě 5,35) nepatří do 100(1-α)% intervalu spolehlivosti pro μ, H_0 zamítáme na hladině významnosti α.

Ad c) Pomocí p-hodnoty
Vzhledem k tomu, že testová statistika T_0 je spojitá náhodná veličina, můžeme použít úpravu $P(T_0 \geq t_0) = P(T_0 > t_0) = 1 - \Phi(t_0)$.
Vzorec pro výpočet p-hodnoty pro oboustrannou alternativu:
$p = 2 \min\{P(T_0 \leq t_0), P(T_0 \geq t_0)\} = 2 \min\{\Phi(t_0), 1 - \Phi(t_0)\}$.
Vzorec pro výpočet p-hodnoty pro levostrannou alternativu:
$p = P(T_0 \leq t_0) = \Phi(t_0)$.
Vzorec pro výpočet p-hodnoty pro pravostrannou alternativu:
$p = P(T_0 \geq t_0) = 1 - \Phi(t_0)$.
Pokud $p \leq \alpha$, H_0 zamítáme na hladině významnosti α.

Zjištěné hodnoty zapíšeme do nového datového souboru o 10 případech a jedné proměnné, kterou nazveme X.
Pomocí Popisných statistik spočteme realizaci výběrového průměru: m = 5,37.
Pro pomocné výpočty otevřeme nový datový soubor o jednom případu a 10 proměnných, které nazveme t0, p1, p2, p3, kv1, kv2, d, h, d1, h2.
Do proměnné t_0 uložíme realizaci testové statistiky. Do jejího Dlouhého jména napíšeme vzorec pro výpočet testové statistiky:
= (5,37-5,35)/(0,04/sqrt(10)).
Zjistíme, že $t_0 = 1,5811$.

Nyní již můžeme provést test pomocí p-hodnoty.
Do Dlouhého jména proměnné p1 napíšeme vzorec pro výpočet p-hodnoty pro oboustrannou alternativu:
=2*min(INormal(t0;0;1);1-INormal(t0;0;1))
Vypočtená p-hodnota je 0,1138, což je větší než hladina významnosti 0,1 a nulovou hypotézu nelze na této hladině významnosti zamítnout ve prospěch oboustranné alternativy.
Do Dlouhého jména proměnné p2 napíšeme vzorec pro výpočet p-hodnoty pro levostrannou alternativu:
=INormal(t0;0;1)
I tato p-hodnota (0,9431) je větší než 0,1, což znamená, že nulovou hypotézu nelze na hladině významnosti 0,1 zamítnout ve prospěch levostranné alternativy.
Do Dlouhého jména proměnné p3 napíšeme vzorec pro výpočet p-hodnoty pro pravostrannou alternativu:
=1-INormal(t0;0;1)
Vyjde nám 0,0569, tedy na hladině významnosti 0,1 zamítáme nulovou hypotézu ve prospěch pravostranné alternativy. S rizikem omylu nejvýše 10 % jsme prokázali, že střední hodnota délky válečků je větší než 5,35 mm.

Dále provedeme test pomocí kritického oboru, nejprve pro oboustrannou alternativu.
Do proměnné kv1 uložíme kvantil $u_{1-\alpha/2} = u_{0,95}$:
= VNormal(0,95;0;1).
Vyjde nám 1,6449.
Kritický obor pro oboustrannou alternativu je tedy $W = \left(-\infty, -1{,}6449\right\rangle \cup \left\langle 1{,}6449, \infty\right)$.
Vidíme, že testová statistika nepatří do W, což znamená, že H_0 nezamítáme na hladině významnosti 0,1 ve prospěch oboustranné alternativy.

Pro testování nulové hypotézy proti jednostranným alternativám musíme znát kvantil $u_{1-\alpha} = u_{0,9}$. Uložíme ho do proměnné kv2:
= VNormal(0,9;0;1).
Vyjde nám 1,2816.
Kritický obor pro levostrannou alternativu je tedy $W = \langle -\infty, -1{,}2816)$.
Vidíme, že testová statistika 1,5811 nepatří do W, což znamená, že H_0 nezamítáme na hladině významnosti 0,1 ve prospěch levostranné alternativy.
Kritický obor pro pravostrannou alternativu je tedy $W = \langle 1{,}2816, \infty)$
Vidíme, že testová statistika1,5811 patří do W, což znamená, že H_0 zamítáme na hladině významnosti 0,1 ve prospěch pravostranné alternativy.

Nakonec provedeme test pomocí intervalu spolehlivosti.
Pro oboustrannou alternativu:
Do Dlouhého jména proměnné d (resp. h) napíšeme vzorec pro dolní (resp. horní) mez oboustranného 90% intervalu spolehlivosti pro μ při známém σ:
=5,37-0,04*kv1/sqrt(10) (resp. =5,37+0,04*kv1/sqrt(10))
Zjistíme, že číslo c = 5,35 patří do intervalu (5,3492; 5,3908), tedy H_0 nezamítáme na hladině významnosti 0,1 ve prospěch oboustranné alternativy.
Pro levostrannou alternativu:
Do Dlouhého jména proměnné h2 napíšeme vzorec pro horní mez pravostranného 90% intervalu spolehlivosti pro μ při známém σ:
=5,37+0,04*kv2/sqrt(10)
Protože 5,35 patří do intervalu (-∞; 5,3862), H_0 nezamítáme na hladině významnosti 0,1 ve prospěch levostranné alternativy.
Pro pravostrannou alternativu:
Do Dlouhého jména proměnné d1 napíšeme vzorec pro dolní mez levostranného 90% intervalu spolehlivosti pro μ při známém σ:
=5,37-0,04*kv2/sqrt(10)
Protože 5,35 nepatří do intervalu (5,3538; ∞), H_0 zamítáme na hladině významnosti 0,1 ve prospěch pravostranné alternativy.

Kolmogorovův – Smirnovův test normality dat

Testujeme nulovou hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z normálního rozložení s parametry μ a σ^2. Distribuční funkci tohoto rozložení označme $\Phi_T(x)$. Nechť $F_n(x)$ je výběrová distribuční funkce. Testovou statistikou je statistika $D_n = \sup_{-\infty<x<\infty} |F_n(x) - \Phi_T(x)|$. Nulovou hypotézu zamítáme na hladině významnosti α, když $D_n \geq D_n(\alpha)$, kde $D_n(\alpha)$ je tabelovaná kritická hodnota.

V případě, že neznáme parametry μ a σ^2 normálního rozložení (což je nejčastější případ), změní se rozložení testové statistiky D_n. V takovém případě jde o Lilieforsovu modifikaci Kolmogorovova – Smirnovova testu. Příslušné modifikované kvantily byly určeny pomocí simulačních studií.

Poznámka ke K-S testu ve STATISTICE

Test normality poskytuje hodnotu testové statistiky (ozn. max D) a dvě p-hodnoty. (p-hodnota vyjadřuje pravděpodobnost, s jakou číselné realizace $x_1, ..., x_n$ náhodného výběru $X_1, ..., X_n$ podporují nulovou hypotézu, je-li pravdivá. P-hodnotu porovnáváme s námi zvolenou hladinou významnosti α. Jestliže p-hodnota $\leq \alpha$, pak H_0 zamítáme na hladině významnosti α, je-li p-hodnota $> \alpha$, 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 $X_1, ..., X_n$ 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.
(S-W test se používá především pro výběry menších rozsahů, $n < 50$, ale nyní již existuje modifikace pro velká n. V systému STATISTICA je implementováno rozšíření na n kolem 5000.)

Andersonův – Darlingův test

Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z normálního rozložení $N(\mu, \sigma^2)$.
Testová statistika má tvar:

$$AD = -\frac{1}{n}\left[\sum_{i=1}^{n}(2i-1)\left\{\ln\Phi\left(\frac{x_{(i)}-m}{s}\right)+\ln\left(1-\Phi\left(\frac{x_{n+1-(i)}-m}{s}\right)\right)\right\}\right]-n,$$

kde $x_{(i)}$ jsou vzestupně uspořádané realizace náhodného výběru, Φ je distribuční funkce rozložení N(0,1). Hypotéza H_0 se zamítá na hladině významnosti α, je-li vypočítaná hodnota testové statistiky AD větší než kritická hodnota $D_{1-\alpha}$.

Úkol 1. : U 48 studentek VŠE v Praze byla zjišťována výška a obor studia (1 – národní hospodářství, 2 – informatika). Hodnoty jsou uloženy v souboru vyska.sta. Pomocí Lilieforsovy modifikace K-S testu, pomocí S-W testu a pomocí A-D testu testujte na hladině významnosti 0,05 hypotézu, že data pocházejí z normálního rozložení. Pomocí N-P grafu posuďte vizuálně předpoklad normality.

Návod:
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.

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

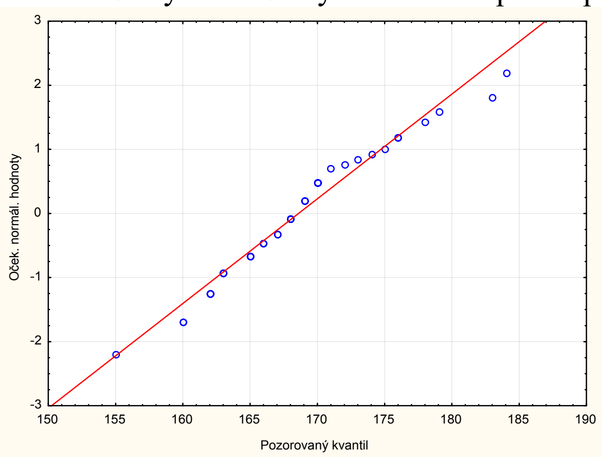
Výstupní tabulka obsahuje počet pozorování, hodnotu testové statistiky Lilieforsovy modifikace K-S testu (max D = 0,155621), p-hodnotu (p < 0,01), testovou statistiku S-W testu (W = 0,965996) a odpovídající p-hodnotu (p = 0,176031). Vidíme, že Lilieforsův test zamítá hypotézu o normalitě na hladině významnosti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

Provedení A - D testu:
Statistiky – Rozdělení & simulace – proložení dat rozděleními – OK – Proměnné Spojité: X – na záložce Spojité proměnné ponecháme zaškrtnuté pouze Normální, na záložce Možnosti vybereme Anderson – Darling – OK – Souhrnné statistiky rozdělení.

	Souhrn rozdělení for Proměnná: X (vyska.sta)							
	K-S d	K-S p-hodn.	AD stat.	AD p-hodn.	Chí-kvadrát	Chí-kvadr. p-hodn.	Chí-kvadr. SV	Posun (práh/poloha)
Normální (poloha,měřítko)	0,155621	0,175802	0,660990	0,591425	15,37500	0,017532	6,000000	

Vidíme, že Testová statistika A – D testu je 0,661, odpovídající p-hodnota je 0,5914, tedy hypotézu o normalitě nezamítáme na hladině významnosti 0,05.

Vytvoření N-P grafu:
Návod: Grafy – 2D Grafy – Normální pravděpodobnostní grafy – Proměnné X – OK.



Tečky se řadí podél ideální přímky, normalita je jen lehce porušena.

Samostatný úkol: Testy normality a grafické ověření normality proveďte jak pro výšky studentek oboru národní hospodářství, tak pro výška studentek oboru informatiky.
Pro kontrolu:
Výsledky pro obor národní hospodářství:

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

Vidíme, že Lilieforsova varianta K-S testu zamítá hypotézu o normalitě na hladině významnosti 0,05 (p-hodnota je menší než 0,05), zatímco S-W test hypotézu o normalitě nezamítá (p-hodnota je větší než 0,05).

	Souhrn rozdělení for Proměnná: X (vyska.sta) Zhrnout podmínku: z=1							
	K-S d	K-S p-hodn.	AD stat.	AD p-hodn.	Chí-kvadrát	Chí-kvadr. p-hodn.	Chí-kvadr. SV	Posun (práh/poloha)
Normální (poloha,měřítko)	0,167473	0,370570	0,419238	0,828398	2,000000	0,157299	1,000000	

A-D test poskytne hodnotu testové statistiky 0,4192, odpovídající p-hodnota je 0,8284, tedy A-D test nezamítá hypotézu o normalitě na hladině významnosti 0,05.

Výsledky pro obor informatika:

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

	Souhrn rozdělení for Proměnná: X (vyska.sta) Zhrnout podmínku: z=2							
	K-S d	K-S p-hodn.	AD stat.	AD p-hodn.	Chí-kvadrát	Chí-kvadr. p-hodn.	Chí-kvadr. SV	Posun (práh/poloha)
Normální (poloha,měřítko)	0,172301	0,536360	0,566019	0,678546				

V tomto případě ani jeden z testů hypotézu o normalitě nezamítá na hladině významnosti 0,05.

Upozornění: V archivu závěrečných prací https://is.muni.cz/auth/th/77721/prif_m/ je uložena diplomová práce Dominika Grůzy „Ověřování normality“.
Zkoumání vlastností testů normality pomocí simulací je popsáno v bakalářské práci Marka Haičmana https://is.muni.cz/auth/th/150689/prif_b_a2/