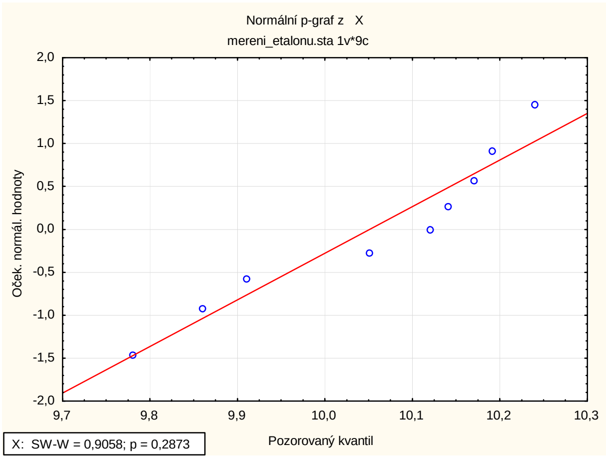
Cvičení 2.: Úlohy na t-testy

Příklad 1.: 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: Naměřené hodnoty jsou realizacemi náhodného výběru z rozložení se střední hodnotou μ. 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. Načteme datový soubor mereni_etalonu.sta. Pomocí N-P grafu a S-W testu ověříme normalitu dat.
Grafy – 2D Grafy – Normální pravděpodobnostní grafy – Proměnná X – OK - odškrtneme Neurčovat průměrnou pozici svázaných pozorování – zaškrtneme Shapiro – Wilkův test - OK.



Data lze považovat za realizace výběru z normálního rozložení.

1. způsob: V Základních statistikách a tabulkách vybereme t-test, samostatný vzorek. Do Referenční hodnoty 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 dostáváme tabulku:

	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. Odchylky od hodnoty 10 lze vysvětlit působením náhodných vlivů.
Všimněme si ještě hodnoty testového kriteria: $t_0 = 0{,}942611$. Kritický obor

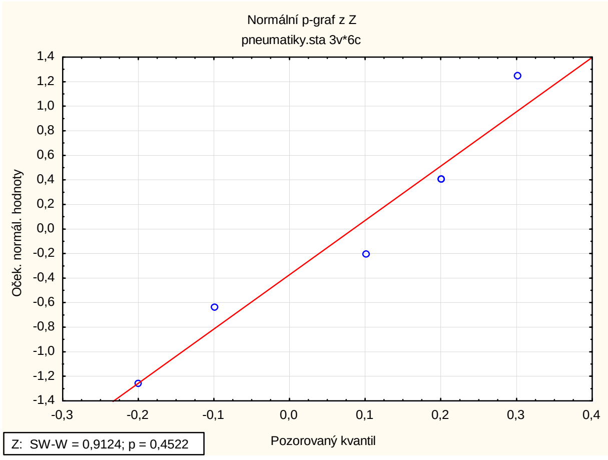
$$W = \left(-\infty, -t_{1-\alpha/2}(n-1)\right\rangle \cup \left\langle t_{1-\alpha/2}(n-1), \infty\right) = \left(-\infty, -t_{0,975}(8)\right\rangle \cup \left\langle t_{0,975}(8), \infty\right) =$$
$$= \left(-\infty, -2{,}306\right\rangle \cup \left\langle 2{,}306, \infty\right)$$

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

2. způsob: V Základních statistikách a tabulkách vypočteme průměr a směrodatnou odchylku. Pak použijeme Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma průměry (normální rozdělení) – zaškrtneme Výběrový průměr vs. Střední hodnota – do políčka Pr1 napíšeme 10,05111, do políčka SmOd1 napíšeme 0,162669, do políčka N1 napíšeme 9, do políčka Pr2 napíšeme 10 - Výpočet. Dostaneme p-hodnotu 0,3735, tedy nezamítáme nulovou hypotézu na hladině významnosti 0,05.

Příklad 2.: Bylo vybráno šest nových vozů téže značky a po určité době bylo zjištěno, o kolik mm se sjely jejich levé a pravé přední pneumatiky.
Výsledky: (1,8; 1,5), (1,0; 1,1), (2,2; 2,0), (0,9; 1,1), (1,5; 1,4), (1,6; 1,4).
Na hladině významnosti 0,05 testujte hypotézu, že obě pneumatiky se sjíždějí stejně rychle.

Návod: Naměřené hodnoty pro levé a pravé pneumatiky tvoří realizace dvourozměrného náhodného výběru s vektorem středních hodnot (μ_1, μ_2). Označme $\mu = \mu_1 - \mu_2$. Na hladině významnosti 0,05 testujeme hypotézu H_0: $\mu = 0$ proti oboustranné alternativě H_1: $\mu \neq 0$. Jde o úlohu na párový t-test.Ten je ve STATISTICE implementován. Načteme datový soubor pneumatiky.sta. Pomocí N-P grafu a S-W testu ověříme normalitu rozdílů (tj. normalitu proměnné Z).
Grafy – 2D Grafy – Normální pravděpodobnostní grafy – Proměnná Z – OK - odškrtneme Neurčovat průměrnou pozici svázaných pozorování – zaškrtneme Shapiro – Wilkův test - OK.



Data lze považovat za realizace výběru z normálního rozložení.

1. způsob: V Základních statistikách vybereme t-test, závislé vzorky. Zadáme názvy obou proměnných a ve výstupu se podíváme na hodnotu testového kritéria a na p-hodnotu.

	t-test pro závislé vzorky (Tabulka1) Označ. rozdíly jsou významné na hlad. p < ,05000							
Proměnná	Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
X	1,500000	0,489898						
Y	1,416667	0,331160	6	0,083333	0,194079	1,051758	5	0,341062

Protože p-hodnota 0,341062 > 0,05, nezamítáme na hladině významnosti 0,05 hypotézu, že obě přední pneumatiky se sjíždějí stejně rychle.
Všimněme si ještě hodnoty testového kriteria: $t_0 = 1,051758$. Kritický obor

$$W = \left(-\infty, -t_{1-\alpha/2}(n-1)\right\rangle \cup \left\langle t_{1-\alpha/2}(n-1), \infty\right) = \left(-\infty, -t_{0,975}(5)\right\rangle \cup \left\langle t_{0,975}(5), \infty\right) =$$
$$= \left(-\infty, -2{,}5706\right\rangle \cup \left\langle 2{,}5706, \infty\right)$$

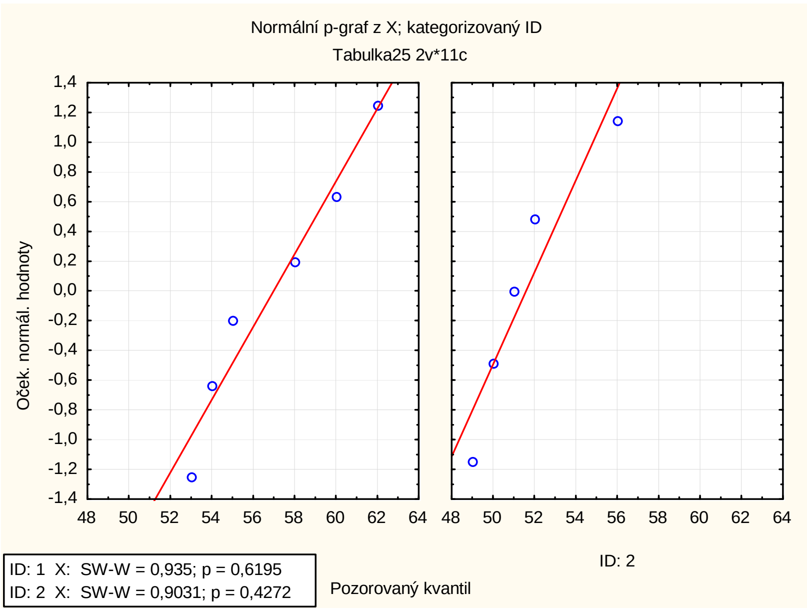
Protože $t_0 \notin W$, nezamítáme na hladině významnosti 0,05 hypotézu H_0.

Příklad 3.: Bylo vylosováno 11 stejně starých selat téhož plemene. Šesti z nich byla předepsána výkrmná dieta č. 1 a zbylým pěti výkrmná dieta č. 2. Průměrné denní přírůstky v Dg za dobu půl roku jsou následující:

dieta č. 1: 62, 54, 55, 60, 53, 58
dieta č. 2: 52, 56, 49, 50, 51.

Testujte na hladině významnosti 0,05 hypotézu, že
a) rozptyly hmotnostních přírůstků selat při obou výkrmných dietách jsou shodné
b) obě výkrmné diety mají stejný vliv na hmotnostní přírůstky selat.

Návod: Naměřené přírůstky při první dietě považujeme za realizace náhodného výběru z rozložení se střední hodnotou μ_1 a rozptylem σ_1^2, naměřené přírůstky při druhé dietě považujeme za realizace náhodného výběru z rozložení se střední hodnotou μ_2 a rozptylem σ_2^2, přičemž tyto dva výběry jsou nezávislé. Načteme datový soubor dve_diety.sta.
Nejprve pomocí N-P grafu a S-W testu ověříme normalitu 1. a 2. výběru.
Grafy – 2D Grafy – Normální pravděpodobnostní grafy – Proměnná X – OK - odškrtneme Neurčovat průměrnou pozici svázaných pozorování – zaškrtneme Shapiro – Wilkův test – na záložce Kategorizovaný zapneme Kategorie X – Změnit proměnnou - ID – OK – OK.



V obou případech lze data považovat za normálně rozložená.

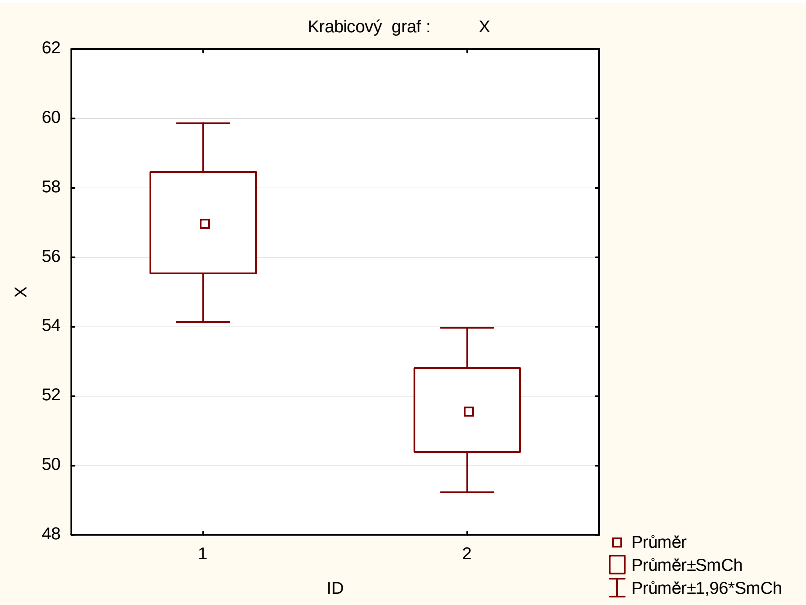
Ad a) Testujeme hypotézu H$_0$: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ proti alternativě H$_1$: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$. K tomu slouží F-test.
Statistiky – Základní statistiky /tabulky – t-test, nezávislé, dle skupin – OK, Proměnné – Závislé proměnné X, Grupovací proměnná ID – OK.
Po kliknutí na tlačítko Výpočet dostaneme tabulku:

Proměnná	t-testy; grupováno: ID (dve_diety.sta) Skup. 1: 1 Skup. 2: 2 Průměr 1	Průměr 2	t	sv	p	Poč.plat 1	Poč.plat. 2	Sm.odch. 1	Sm.odch. 2	F-poměr Rozptyly	p Rozptyly
X	57,00000	51,60000	2,771222	9	0,021710	6	5	3,577709	2,701851	1,753425	0,606345

Vidíme, že testová statistika pro test shody rozptylů se realizuje hodnotou 1,7534, odpovídající p-hodnota je 0,6063, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Ad b) Testujeme hypotézu H_0: $\mu_1 - \mu_2 = 0$ proti H_1: $\mu_1 - \mu_2 \neq 0$. K tomu slouží dvouvýběrový t-test.
1. způsob: Z výše uvedené výstupní tabulky plyne, že testová statistika pro test shody středních hodnot se realizuje hodnotou 2,7712, počet stupňů volnosti je 9, odpovídající p-hodnota 0,0217, tedy hypotézu o shodě středních hodnot hmotnostních přírůstků selat při dvou výkrmných dietách zamítáme na hladině významnosti 0,05.

Tabulku ještě doplníme krabicovými diagramy. Na záložce Detaily zaškrtneme krabicový graf a vybereme volbu Průměr/SmCh/1,96*SmCh.



2. způsob: V Základních statistikách a tabulkách vypočteme průměry ($m_1 = 57$, $m_2 = 51,6$) a směrodatné odchylky hmotnostních přírůstků při obou dietách ($s_1 = 3,5777$, $s_2 = 2,7019$). Pak použijeme Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma průměry (normální rozdělení) – do políčka Pr1 napíšeme 57, do políčka SmOd1 napíšeme 3,5777, do políčka N1 napíšeme 6, do políčka Pr2 napíšeme 51,6, do políčka SmOd2 napíšeme 2,7019, do políčka N2 napíšeme 5 – Výpočet. Dostaneme p-hodnotu 0,0217, tedy zamítáme nulovou hypotézu na hladině významnosti 0,05.

Úlohy k samostatnému řešení

Příklad 1.: U souboru náhodně vybraných pracovníků byl zjišťován počet vyrobených výrobků za směnu před provedením modernizace výrobní linky (veličina X) a po provedení modernizace (veličina Y). Zjištěné výsledky:

č. prac.	1	2	3	4	5	6	7	8	9	10	11	12
X	143	151	139	145	149	153	148	146	149	142	147	140
Y	146	152	144	144	151	156	153	147	146	145	147	139

a) Vypočtěte průměrný počet výrobků před modernizací a po modernizaci:
m_1 =
m_2 =

b) Na hladině významnosti 0,05 testujte hypotézu, že modernizace výrobní linky neměla vliv na výkon pracovníků proti alternativě, že modernizace vedla ke zvýšení výkonu pracovníků.
Zápis nulové a alternativní hypotézy:
Název použitého testu:
Realizace testové statistiky: t_0 =
Počet stupňů volnosti =
p-hodnota =
Rozhodnutí o nulové hypotéze:

c) Ukažte, že jsou splněny podmínky pro použití zvoleného testu.

Příklad 2.: Odběratel dostává vypínače od dvou různých dodavatelů, označme je A a B. Během hodnocení kvality vypínače se sleduje počet sepnutí, které vypínač snese bez poškození. Při testování bylo použito 10 vypínačů od firmy A a 8 vypínačů od firmy B. Byly získány tyto výsledky:
Dodavatel A: 6238 7153 5389 5682 5903 6690 7309 7738 5389 4890
Dodavatel B: 6739 4968 5889 5678 5290 6738 6045 6678

a) Najděte číselné charakteristiky počtů sepnutí v obou skupinách (na 1 desetinné místo).
m_1 =
s_1 =
m_2 =
s_2 =

b) S-W testem posuďte na hladině významnosti 0,05 normalitu rozložení počtu sepnutí v 1. a 2. skupině.
Hodnota testové statistiky v 1. skupině =
p-hodnota =
rozhodnutí o normalitě v 1. skupině:
Hodnota testové statistiky ve 2. skupině =
p-hodnota =
rozhodnutí o normalitě ve 2. skupině:

c) Na hladině významnosti 0,1 testujte hypotézu, že střední hodnota počtu sepnutí v 1. a 2. skupině se neliší.
Zápis nulové a alternativní hypotézy:

Název použitého testu:
Hodnota testové statistiky pro test shody středních hodnot =
počet stupňů volnosti =
p-hodnota =
rozhodnutí o nulové hypotéze:

Hodnota testové statistiky pro F- test shody rozptylů =
počty stupňů volnosti =
p-hodnota =
rozhodnutí o nulové hypotéze:

Příklad 3.: Systematická chyba měřicího přístroje se eliminuje nastavením přístroje a měřením etalonu, jehož správná hodnota je μ = 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.

a) Vypočtěte číselné charakteristiky uvedených devíti měření.
m =
s =

b) Je možné při riziku 0,05 vysvětlit odchylky od hodnoty 10,00 působením náhodných vlivů?
Zápis nulové a alternativní hypotézy:
Název použitého testu:
Realizace testové statistiky: t_0 =
Počet stupňů volnosti =
p-hodnota =
Rozhodnutí o nulové hypotéze: