

Hodnocení sebeúcty

Popis situace:

28 náhodně vybraných plnoletých osob mladších 35 let české národnosti s trvalým bydlištěm v Brně vyplnilo dotazník týkající se jejich sebeúcty.

Dotazník se skládá z 10 tvrzení, s nimiž může respondent:

- výrazně nesouhlasit (hodnota 1),
- mírně nesouhlasit (hodnota 2),
- mírně souhlasit (hodnota 3),
- výrazně souhlasit (hodnota 4).

Celkový rozsah hodnocení sebeúcty se tedy nachází v intervalu [10, 40]. Čím vyšší hodnota, tím vyšší sebeúcta konkrétní osoby.

U zkoumaných osob se také zjišťovalo pohlaví a věk.

Datový soubor:

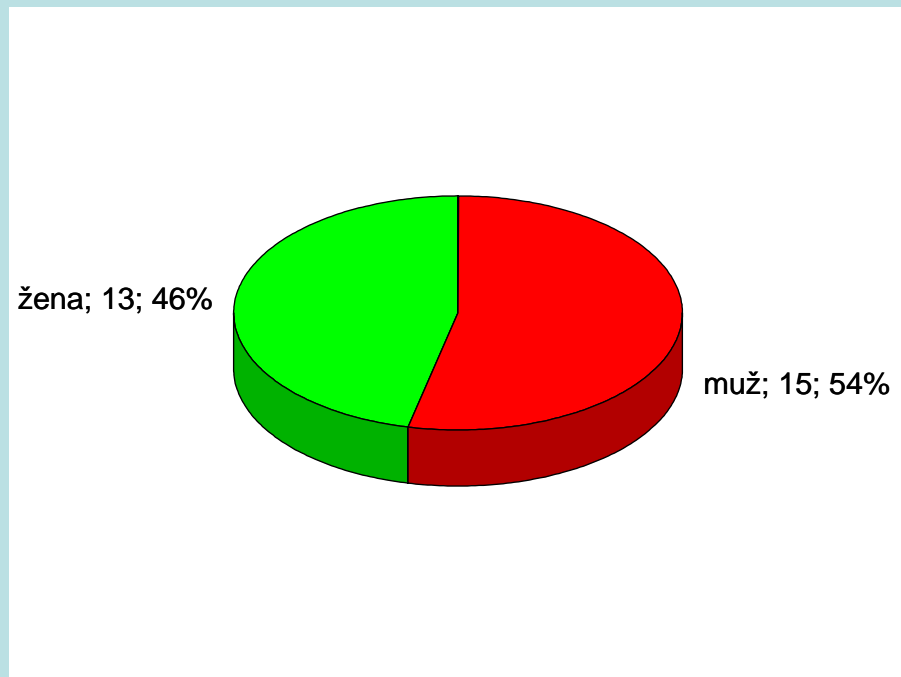
| č. respondenta | pohlaví | sebeúcta | věk |
|----------------|---------|----------|-----|
| 1 | žena | 22 | 24 |
| 2 | žena | 26 | 32 |
| 3 | muž | 24 | 24 |
| 4 | žena | 28 | 28 |
| 5 | muž | 36 | 33 |
| 6 | muž | 37 | 32 |
| 7 | muž | 36 | 28 |
| 8 | žena | 26 | 26 |
| 9 | žena | 21 | 30 |
| 10 | žena | 26 | 29 |
| 11 | muž | 29 | 26 |
| 12 | žena | 30 | 27 |
| 13 | muž | 38 | 30 |
| 14 | muž | 31 | 25 |
| 15 | žena | 36 | 20 |
| 16 | žena | 37 | 21 |
| 17 | žena | 30 | 24 |
| 18 | muž | 31 | 25 |
| 19 | muž | 26 | 32 |
| 20 | muž | 36 | 23 |
| 21 | muž | 29 | 21 |
| 22 | žena | 30 | 19 |
| 23 | muž | 34 | 21 |
| 24 | muž | 37 | 29 |
| 25 | žena | 28 | 21 |
| 26 | žena | 29 | 22 |
| 27 | muž | 31 | 26 |
| 28 | muž | 35 | 22 |

Nejprve popíšeme datový soubor pomocí tabulek, grafů a číselných charakteristik.

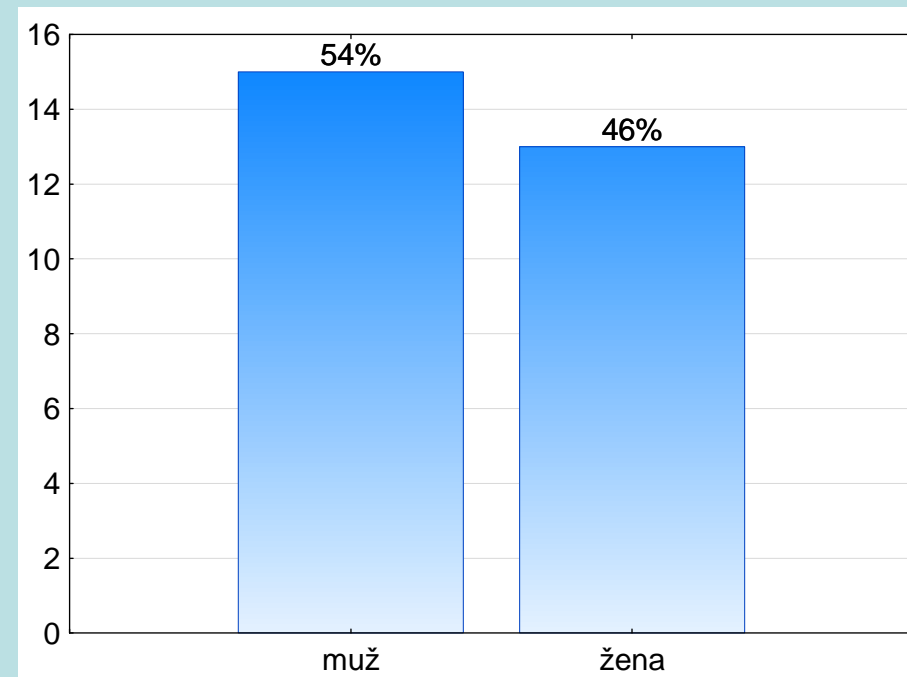
Tabulka četností proměnné pohlaví:

| Kategorie | Četnost | Rel.četnost |
|-----------|---------|-------------|
| muž | 15 | 53,6 |
| žena | 13 | 46,4 |

Výsečový diagram proměnné pohlaví:



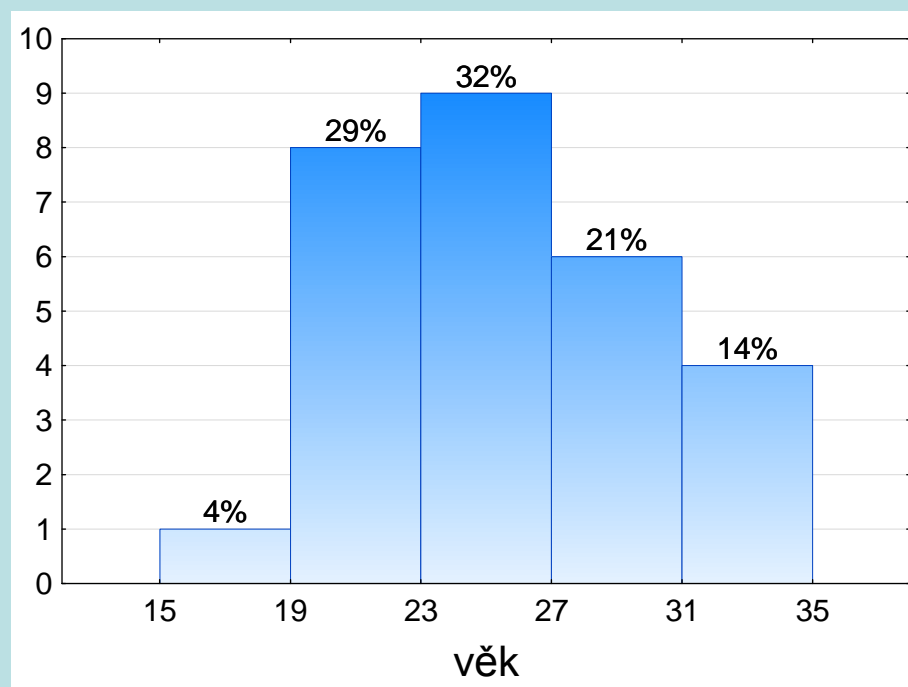
Sloupkový diagram proměnné pohlaví:



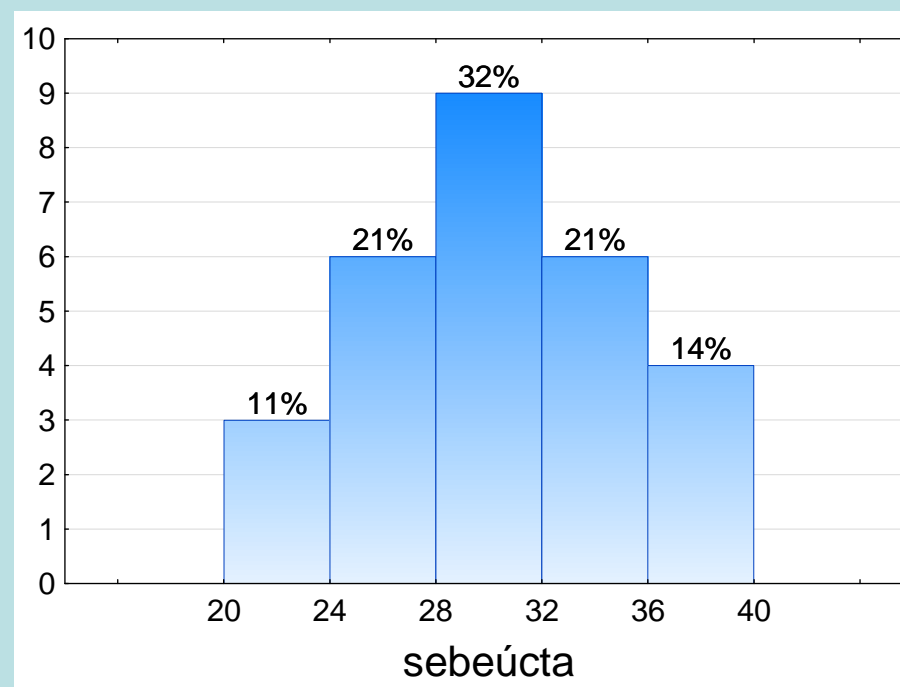
Číselné charakteristiky proměnných věk a sebeúcta:

| Proměnná | N platných | Průměr | Minimum | Maximum | Sm.odch. |
|----------|------------|--------|---------|---------|----------|
| věk | 28 | 25,7 | 19 | 33 | 4,10 |
| sebeúcta | 28 | 30,7 | 21 | 38 | 4,88 |

Histogram proměnné věk:



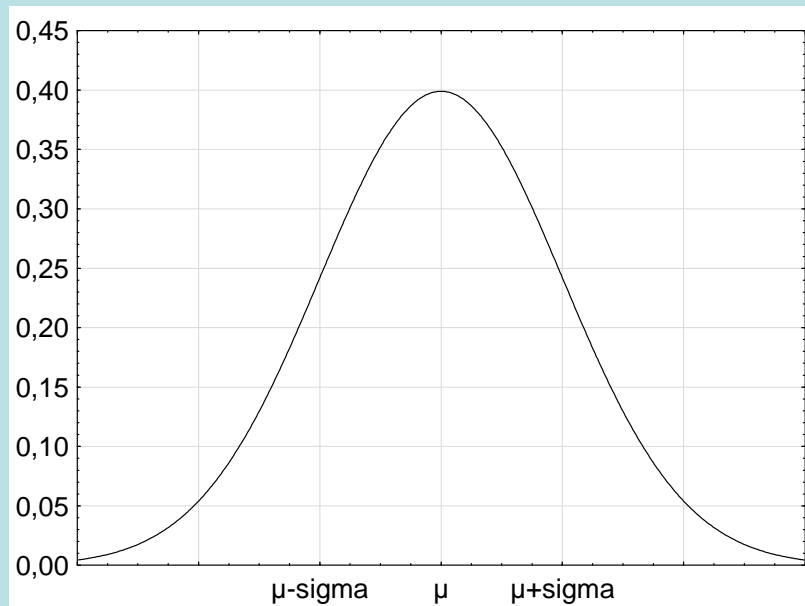
Histogram proměnné sebeúcta:



Výzkumná otázka č. 1: Liší se muži a ženy z hlediska věku?

Základní soubor je tvořen všemi plnoletými osobami mladšími 35 let, které jsou české národnosti a mají trvalé bydliště v Brně. Střední hodnotu věku těchto osob mužského pohlaví označíme μ_1 a ženského pohlaví μ_2 . Předpokládáme, že směrodatná odchylka σ_1 věku mužů je stejná jako směrodatná odchylka σ_2 věku žen a rozložení věku v obou skupinách je normální.

Graf hustoty normálního rozložení s parametry μ a σ :



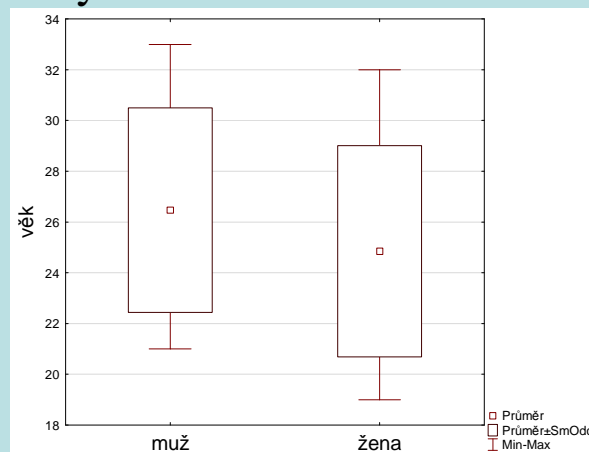
Máme k dispozici nikoliv celý základní soubor, ale dva nezávislé výběrové soubory, první obsahuje 15 mužů a druhý 13 žen.

Na základě vypočtených výběrových průměrů m_1 , m_2 (považujeme je za odhady neznámých středních hodnot μ_1 , μ_2) a vypočtených výběrových směrodatných odchylek s_1 , s_2 (považujeme je za odhady neznámých výběrových odchylek σ_1 , σ_2) budeme na dané hladině významnosti α (zpravidla 0,05) usuzovat na existenci či neexistenci rozdílu mezi neznámými středními hodnotami μ_1 , μ_2 .

Číselné charakteristiky věku pro muže a pro ženy:

| pohlaví | věk N | věk průměr | věk Sm.odch. | věk Minimum | věk Maximum |
|----------|----------|---------------|-----------------|----------------|----------------|
| muž | 15 | 26,5 | 4,03 | 21 | 33 |
| žena | 13 | 24,8 | 4,16 | 19 | 32 |
| Vš.skup. | 28 | 25,7 | 4,10 | 19 | 33 |

Krabicový diagram věku pro muže a pro ženy:



Odpověď na výzkumnou otázku č. 1 budeme hledat pomocí **dvouvýběrového t-testu**.

Nulová hypotéza: Střední hodnota věku mužů je stejná jako střední hodnota věku žen.

Statistický zápis nulové hypotézy: $H_0: \mu_1 = \mu_2$

Alternativní hypotéza: Střední hodnota věku mužů se liší od střední hodnoty věku žen.

Statistický zápis alternativní hypotézy: $H_1: \mu_1 \neq \mu_2$

Dvouvýběrový t-test je založen na porovnání výběrových průměrů v obou skupinách při zohlednění vlivu odhadnuté společné směrodatné odchylky.

V našem případě nás tedy zajímá, zda rozdíl mezi průměrným věkem mužů 26,5 let a průměrným věkem žen 24,8 let (s výběrovými směrodatnými odchylkami 4,03 let a 4,16 let a rozsahy 15 a 13) je způsoben pouze náhodnými vlivy nebo je prokazatelný na hladině významnosti 0,05.

Statistický software vypočte podle určitého vzorce tzv. **p-hodnotu**, kterou porovnáme s námi zvolenou hladinou významnosti α .

Pokud $p \leq \alpha$, nulovou hypotézu zamítáme na hladině významnosti α a přijímáme alternativní hypotézu. Znamená to, že s rizikem omylu nejvýše $100 \alpha \%$ jsme prokázali pravdivost alternativní hypotézy.

V opačném případě nulovou hypotézu nezamítáme. Neznamená to však, že jsme prokázali její pravdivost. Lze pouze říci, že naše data nejsou natolik průkazná, aby umožnila zamítnutí nulové hypotézy.

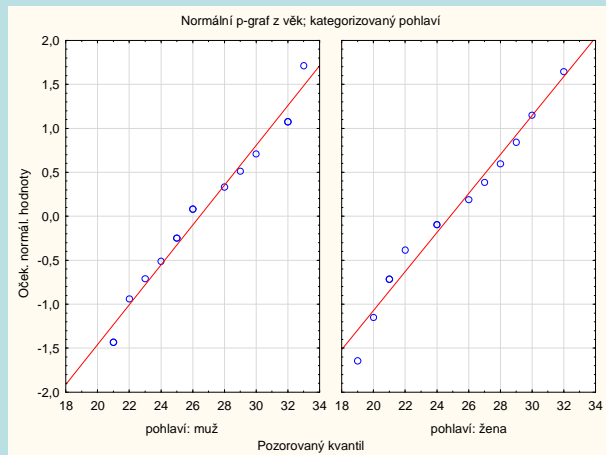
Na vypočtenou p-hodnotu můžeme také pohlížet jako na pravděpodobnost, s jakou naše data podporují nulovou hypotézu, je-li pravdivá.

Než provedeme samotný dvouvýběrový t-test, musíme ověřit jeho **předpoklady**.

1. Nezávislost výběrových souborů: splněno, plyne přímo ze způsobu získání dat.

2. Rozložení věku mužů a žen je normální.

Grafické ověření předpokladu normality – např. pomocí **normálního pravděpodobnostního grafu**. Pokud se tečky v grafu řadí blízko ideální přímky, lze usuzovat na normalitu.



Grafická metoda je však subjektivní. Objektivní metodou jsou **testy normality**. Ukážeme výsledek Shapirova – Wilkova testu:

| Proměnná | N | W | p |
|------------|----|----------|----------|
| věk - muž | 15 | 0,937276 | 0,349355 |
| věk - žena | 13 | 0,953659 | 0,654608 |

V obou případech je p-hodnota větší než zvolená hladina významnosti 0,05, tedy hypotézu o normalitě věku nezamítáme ani pro muže, ani pro ženy.

3. Směrodatné odchylky (tedy i rozptyly) věku jsou v obou skupinách stejné.

Pro testování této hypotézy použijeme **F-test**.

Testujeme $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ proti $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$.

Výsledek F-testu:

| Proměnná | Poč.plat muž | Poč.plat. žena | Sm.odch. muž | Sm.odch. žena | F-poměr Rozptyly | p Rozptyly |
|----------|--------------|----------------|--------------|---------------|------------------|------------|
| věk | 15 | 13 | 4,033196 | 4,160251 | 1,063997 | 0,901422 |

Vidíme, že p-hodnota (0,9014) je větší než hladina významnosti 0,05, tudíž na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Nyní již přistoupíme k provedení dvouvýběrového t-testu:

| Proměnná | Průměr muž | Průměr žena | t | sv | p | Poč.plat muž | Poč.plat. žena |
|----------|------------|-------------|----------|----|----------|--------------|----------------|
| věk | 26,46667 | 24,84615 | 1,045010 | 26 | 0,305637 | 15 | 13 |

Jelikož p-hodnota je 0,3056, nelze na hladině významnosti 0,05 zamítnout hypotézu, že střední hodnoty věku mužů a žen jsou shodné.

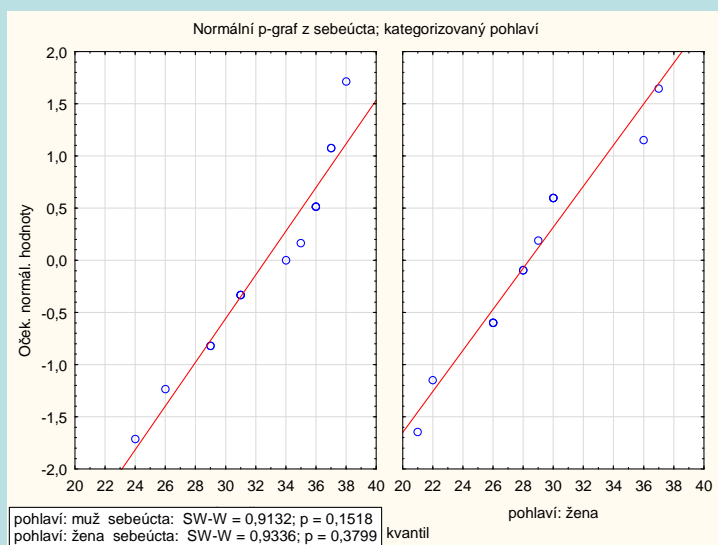
Výzkumná otázka č. 2: Liší se muži a ženy z hlediska sebeúcty?

Na tuto otázku opět odpovíme za pomoci dvouvýběrového t-testu.

Ověření předpokladů:

- 1. Nezávislost výběrových souborů:** splněno, plyne přímo ze způsobu získání dat.
- 2. Rozložení hodnocení sebeúcty mužů a žen je normální.**

Normální pravděpodobnostní graf společně se Shapirovým – Wilkovým testem normality:



Na hladině významnosti 0,05 nelze zamítnout hypotézu o normalitě hodnocení sebeúcty pro muže a pro ženy.

3. Směrodatné odchylky (tedy i rozptyly) hodnocení sebeúcty jsou v obou skupinách stejné.

| Proměnná | Poč.plat muž | Poč.plat. žena | Sm.odch. muž | Sm.odch. žena | F-poměr Rozptyly | p Rozptyly |
|----------|--------------|----------------|--------------|---------------|------------------|------------|
| sebeúcta | 15 | 13 | 4,320494 | 4,592357 | 1,129808 | 0,818192 |

Na hladině významnosti 0,05 nezamítáme F-testem hypotézu o shodě rozptylů, neboť p-hodnota F-testu je větší než hladina významnosti 0,05.

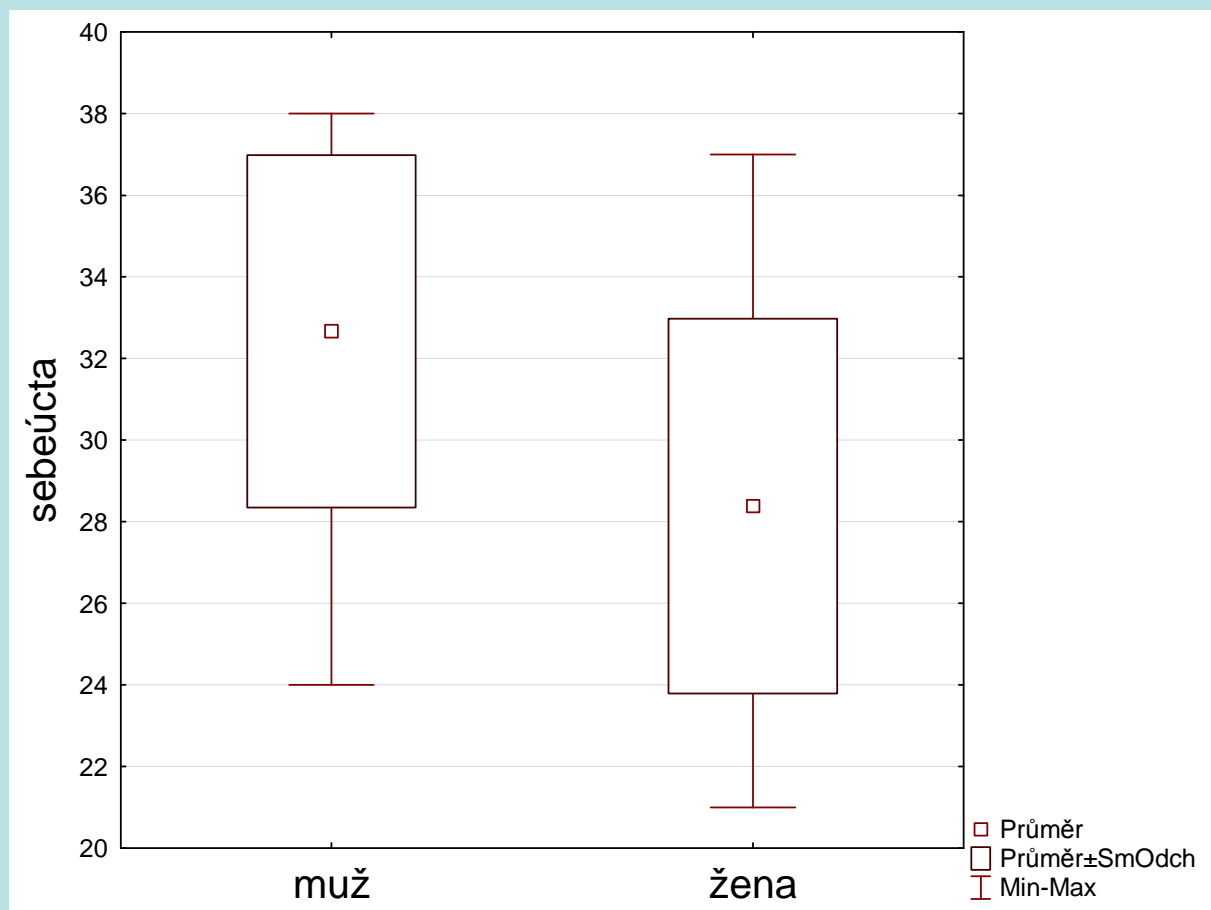
Výsledky dvouvýběrového t-testu:

| Proměnná | Průměr muž | Průměr žena | t | sv | p |
|----------|------------|-------------|----------|----|----------|
| sebeúcta | 32,66667 | 28,38462 | 2,540516 | 26 | 0,017382 |

Vidíme, že p-hodnota je 0,0174, což je menší než 0,05, tudíž na hladině významnosti 0,05 zamítáme hypotézu o shodě středních hodnot sebeúcty u mužů a žen.

Průměrná hodnota pro muže je 32,7, pro ženy 28,4. Rozdíl mezi těmito hodnotami je průkazný s rizikem omylu nejvýše 5 %.

Krabicový graf:



Zajímá nás však, zda je tento rozdíl nejenom statisticky, ale i věcně významný. Pro hodnocení věcné významnosti rozdílu dvou průměrů slouží **Cohenův koeficient věcného účinku**. Počítá se podle vzorce $d = \frac{|m_1 - m_2|}{s_*}$, kde m_1 , m_2 jsou průměry a s_* je odhad společné neznámé směrodatné odchylky. Velikost účinku hodnotíme podle následující tabulky:

| hodnota d | účinek |
|-----------------|--------------|
| aspoň 0,8 | velký |
| mezi 0,5 až 0,8 | střední |
| mezi 0,2 až 0,5 | malý |
| pod 0,2 | zanedbatelný |

(Uvedené hodnoty nemají samozřejmě absolutní platnost, posouzení, jaký účinek považujeme za velký či malý, závisí na kontextu.)

V našem případě $d = 0,963$, tedy z věcného hlediska je rozdíl v sebehodnocení mužů a žen velký.

Výzkumná otázka č. 3: Existuje závislost mezi věkem respondenta a hodnocením sebeúcty?

Omezíme se pouze na **lineární závislost**. Budeme ji posuzovat pomocí **korelačního koeficientu**. V základním souboru ho označíme ρ , ve výběrovém souboru ho budeme nazývat výběrový koeficient korelace a označíme ho r .

Koeficient korelace nabývá hodnot od -1 do 1. Čím je jeho hodnota bližší 1, tím je silnější přímá lineární závislost mezi sledovanými dvěma veličinami a čím je jeho hodnota bližší -1, tím je silnější nepřímá lineární závislost mezi sledovanými dvěma veličinami. Hodnoty blízké nule svědčí o neexistenci lineárního vztahu.

Význam absolutní hodnoty koeficientu korelace:

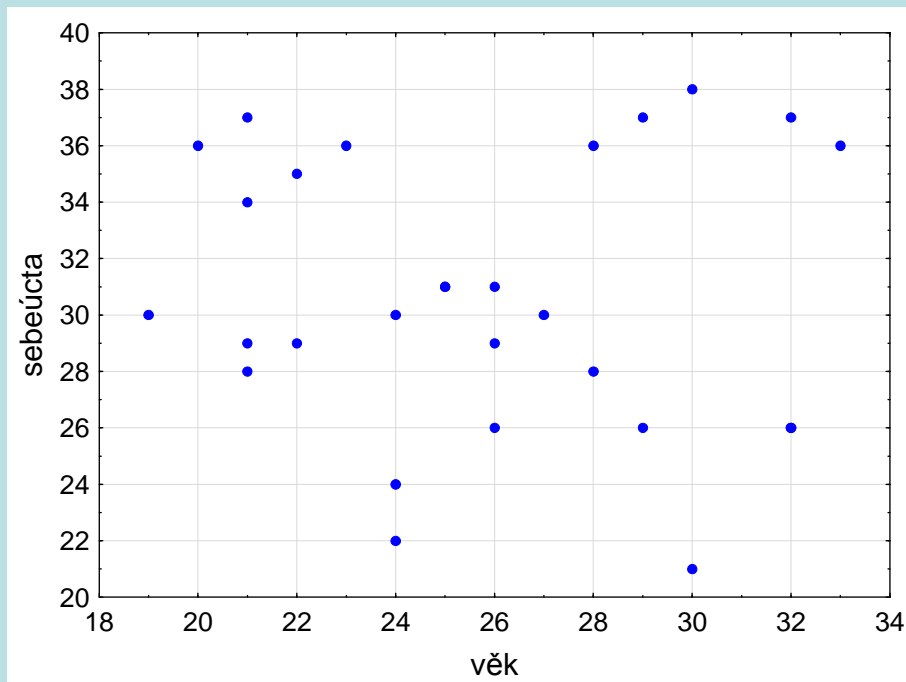
mezi 0 až 0,1 ... zanedbatelná lineární závislost,

mezi 0,1 až 0,3 ... slabá lineární závislost,

mezi 0,3 až 0,7 ... střední lineární závislost,

mezi 0,7 až 1 ... silná lineární závislost.

Pomocí **dvourozměrného tečkového diagramu** orientačně posoudíme, zda mezi věkem a sebeúctou je patrná nějaká lineární závislost.



Pokud by mezi danými dvěma proměnnými byla lineární závislost, tečky v diagramu by se řadily podél přímky. V našem případě tomu tak není, lze tedy soudit, že výběrový koeficient korelace bude blízký 0 a závislost mezi věkem a sebeúctou nebude prokazatelná na hladině významnosti 0,05.

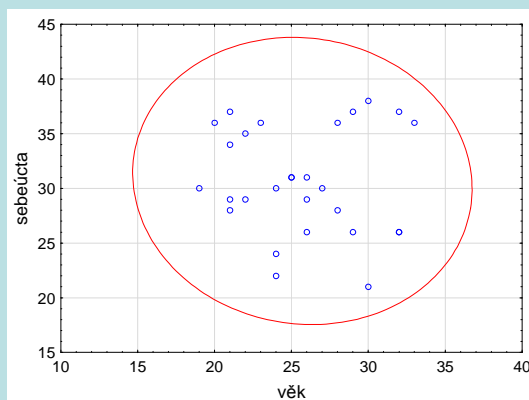
Nulová hypotéza: Mezi věkem respondenta a sebeúctou neexistuje žádná lineární závislost.

Statistický zápis nulové hypotézy: $H_0: \rho = 0$.

Alternativní hypotéza: Mezi věkem respondenta a sebeúctou existuje lineární závislost .

Statistický zápis alternativní hypotézy: $H_1: \rho \neq 0$.

Provedeme tedy **test nekorelovanosti** (tj. lineární nezávislosti), který je založen na výběrovém koeficientu korelace. Předpokládá se, že sledované proměnné mají dvourozměrné normální rozložení, což orientačně ověříme tak, že do dvourozměrného tečkového diagramu zakreslíme **elipsu konstantní 95% hustoty pravděpodobnosti**. Pokud uvnitř této elipsy leží aspoň 95 % teček, můžeme data považovat za dvourozměrně normální:



Vypočteme výběrový koeficient korelace mezi věkem a sebeúctou:

| Proměnná | sebeúcta |
|----------|----------|
| věk | -0,0548 |

Vidíme, že výběrový koeficient korelace nabývá hodnoty $-0,0548$, což svědčí o zanedbatelně malé nepřímé lineární závislosti mezi věkem a sebeúctou.

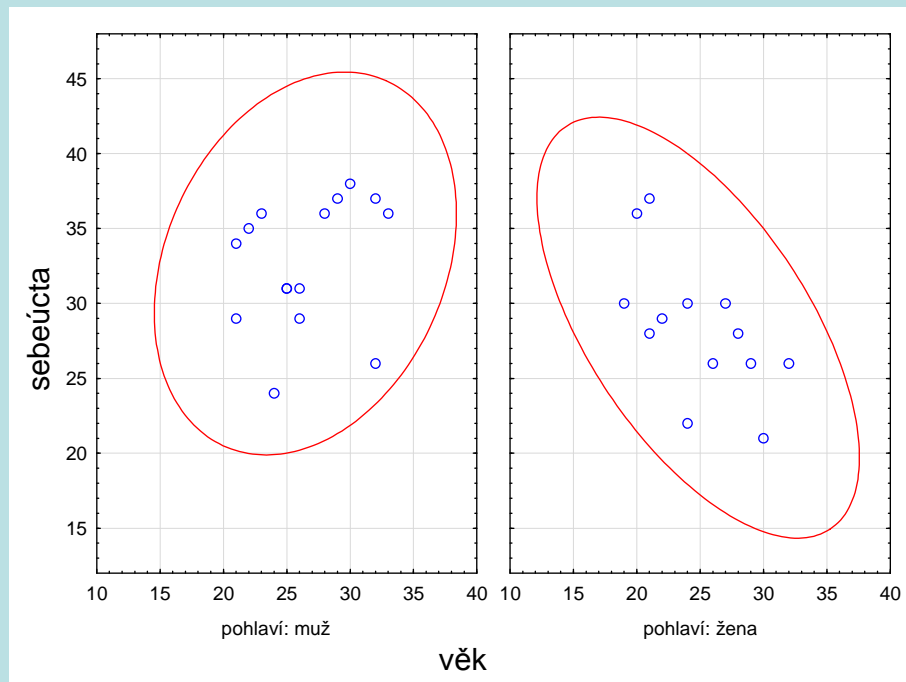
Výsledek testu nekorelovanosti:

| Proměnná | sebeúcta |
|----------|----------|
| věk | -,0548 |
| | p=,782 |

Jelikož p-hodnota je $0,782$, což je větší než hladina významnosti $0,05$, nulovou hypotézu nelze zamítnout. Neprokázalo se tedy, že by mezi věkem respondenta a jeho sebeúctou existovala lineární závislost.

Nyní provedeme podrobnější rozbor závislosti věku a sebeúcty, a to zvlášť pro muže a zvlášť pro ženy.

Ověření předpokladu dvourozměrné normality:



Vidíme, že jak u mužů tak u žen lze data považovat za dvourozměrně normální.

Výběrové koeficienty korelace společně s p-hodnotami pro test nekorelovanosti:

Výsledky pro muže

| | |
|------|----------|
| muži | sebeúcta |
| věk | ,2596 |
| | p=,350 |

U mužů je výběrový koeficient korelace kladný a jeho hodnota svědčí o tom, že s rostoucím věkem slabě narůstá sebeúcta. Závislost však není prokazatelná na hladině významnosti 0,05, neboť p-hodnota je 0,35.

Výsledky pro ženy

| | |
|------|----------|
| ženy | sebeúcta |
| věk | -,6160 |
| | p=,025 |

U žen existuje mezi věkem a sebehodnocením statisticky prokazatelná středně silná nepřímá lineární závislost. Čím starší žena, tím nižší sebehodnocení.