

Vzorová prezentace do předmětu Výpočetní statistika

Popis situace: U 32 náhodně vybraných osob byly zjišťovány hodnoty těchto proměnných:

SEX ... 1 - muž, 2 – žena

PUVOD ... 1 – Skandinávie, 2 – Středomoří, 3 – západní Evropa

IQ ... hodnota inteligenčního kvocientu v bodech

VYSKA ... tělesná výška v cm

HMOTNOST ... tělesná hmotnost v kg

BMI ... Body Mass Index,
$$\text{BMI} = \frac{\text{hmotnost (v kg)}}{\text{vyska}^2 \text{ (v m)}}$$

PRIJEM ... roční příjem (v tisících euro)

Cílem výzkumu je porovnat, zda v proměnných IQ, BMI a PRIJEM existují rozdíly mezi pohlavími a mezi oblastmi původu.

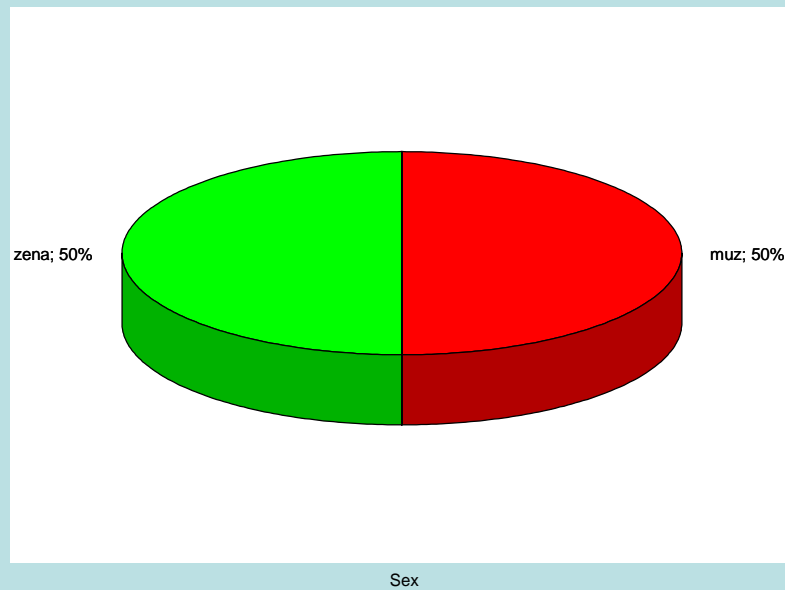
Základní charakteristiky datového souboru

Tabulka četností variant proměnné SEX:

Kategorie	Četnost	Rel.četnost
muz	16	50,0
zena	16	50,0

V souboru je stejný počet mužů i žen.

Výsečový diagram proměnné SEX

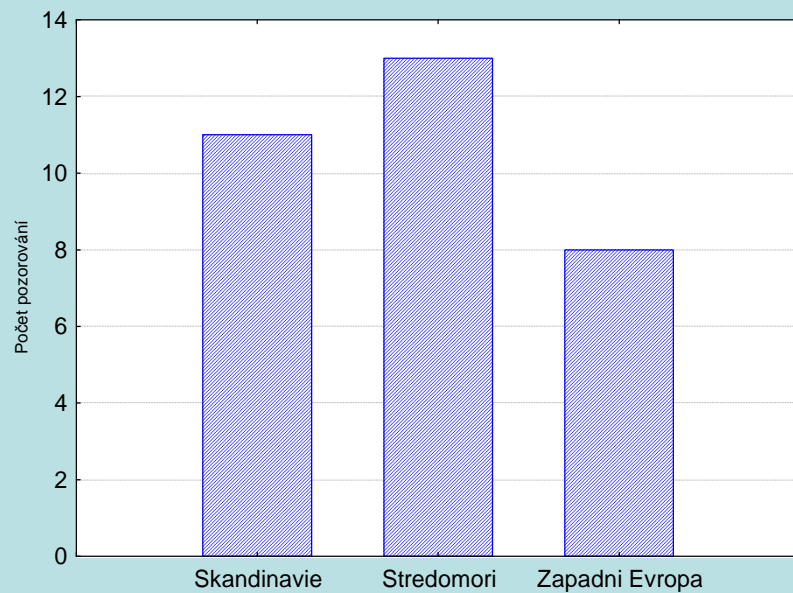


Tabulka četností proměnné PUVOD

Kategorie	Četnost	Rel.četnost
Skandinávie	11	34,4
Středomoří	13	40,6
Západní Evropa	8	25,0

Nevíce jsou v našem souboru zastoupeni lidé ze Středomoří (40,6%), méně často ze Skandinávie (34,4%) a nejméně ze západní Evropy (25%).

Sloupkový diagram proměnné PUVOD



Číselné charakteristiky proměnných IQ, BMI, PRIJEM

Proměnná	N platných	Průměr	Minimum	Maximum	Sm.odch.	Koef.prom.
IQ	32	115,1	96,0	140	12,2	10,6
BMI	32	21,2	17,1	25	2,7	13,0
Prijem	32	27,4	11,0	45	8,9	32,5

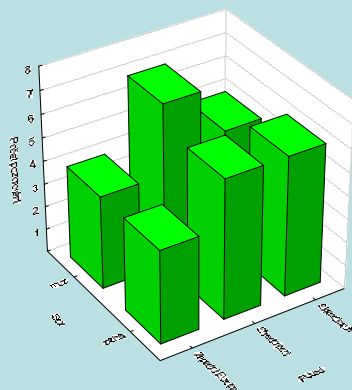
Průměrné IQ činí 115,1, průměrné BMI 21,2 a průměrný příjem 27 000 euro ročně. Největší proměnlivost vykazuje příjem (koeficient variace je 32,5%), nejmenší IQ (koeficient variace je 10,6%).

Kontingenční tabulka absolutních a relativních četností proměnných SEX a PUVOD

	Sex	Puvod Skandinavie	Puvod Stredomori	Puvod Zapadni Evropa	Řádk. součty
Četnost	muz	5	7	4	16
Celková četn.		15,63%	21,88%	12,50%	50,00%
Četnost	zena	6	6	4	16
Celková četn.		18,75%	18,75%	12,50%	50,00%
Četnost	Vš.skup.	11	13	8	32
Celková četn.		34,38%	40,63%	25,00%	

V našem souboru jsou nejvíce zastoupeni muži ze Středomoří (7 osob, tj. 21,9%), nejméně muži ze západní Evropy (4 osoby, tj. 12,5%) a ženy ze západní Evropy (4 osoby, tj. 12,5%).

Grafické znázornění absolutních četností proměnných SEX a PUVOD



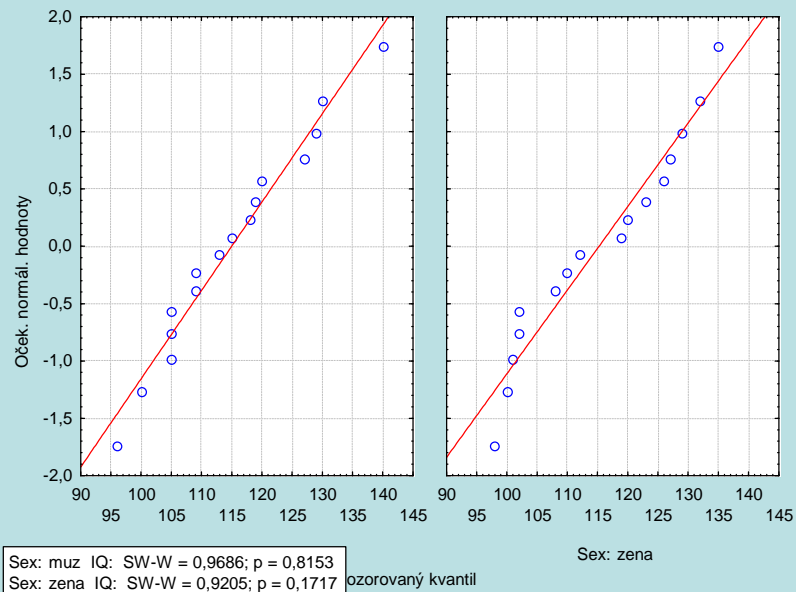
Porovnání proměnné IQ z hlediska pohlaví

Nulová hypotéza tvrdí, že střední hodnoty proměnné IQ jsou stejné pro muže a ženy, tj.

$$H_0 : \mu_1 = \mu_2 \text{ proti } H_1 : \mu_1 \neq \mu_2$$

K testování použijeme dvouvýběrový t-test.

Nejprve ověříme normalitu proměnné IQ ve skupině mužů a ve skupině žen pomocí S-W testu a pomoci normálního pravděpodobnostního grafu:



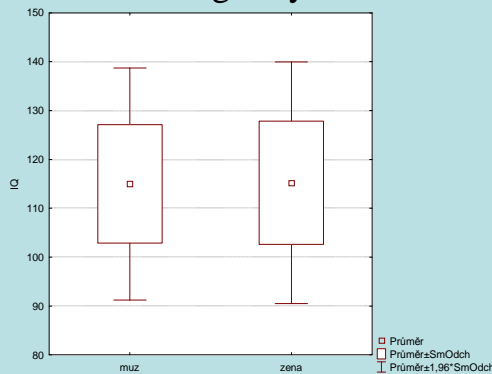
Hypotézu o normalitě proměnné IQ nelze na hladině významnosti 0,05 zamítnout ani pro muže, ani pro ženy.

Vypočítáme číselné charakteristiky proměnné IQ ve skupinách mužů a žen:

Sex	IQ průměr	IQ N	IQ Sm.odch.
muz	115,00	16	12,12
zena	115,25	16	12,61
Vš.skup.	115,13	32	12,16

Vidíme, že rozdíl v průměrném IQ je velmi malý, pouhá čtvrtina bodu.

Krabicové diagramy:



Hypotézu o shodě rozptylů proměnné IQ v daných dvou skupinách ověříme pomocí F-testu:

Proměnná	Průměr žena	Průměr muz	t	sv	p	Poč.plat žena	Poč.plat. muz	Sm.odch. žena	Sm.odch. muz	F-poměr Rozptyly	p Rozptyly
IQ	115,2500	115,0000	0,057185	30	0,954777	16	16	12,60952	12,11610	1,083106	0,879169

Testová statistika F-testu nabývá hodnoty 1,0831, odpovídající p-hodnota je 0,8792, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Současně z tabulky plyne, že testová statistika dvouvýběrového t-testu se realizuje hodnotou 0,0572, odpovídající p-hodnota je 0,9548, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě středních hodnot.

Neprokázali jsme tedy, že by se lišily střední hodnoty IQ mužů a žen.

Cohenův koeficient věcného účinku je 0,02, tedy vliv proměnné SEX na IQ je zcela zanedbatelný.

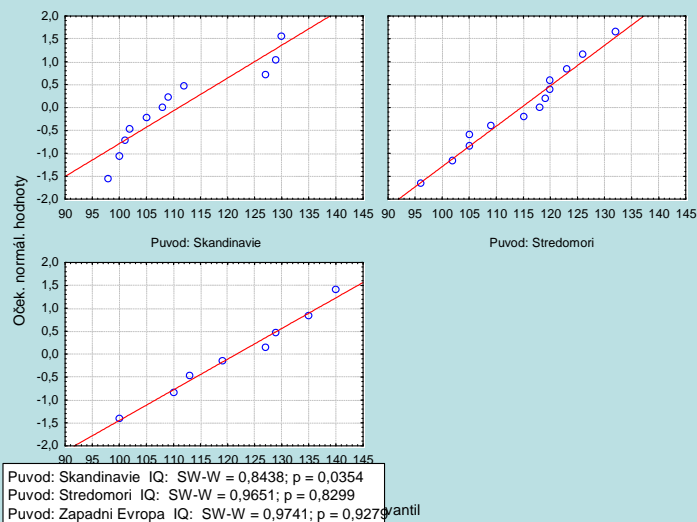
Porovnání proměnné IQ z hlediska původu

Nulová hypotéza tvrdí, že střední hodnoty proměnné IQ jsou stejné pro obyvatele Skandinávie, Středomoří a západní Evropy, tj.

$H_0 : \mu_1 = \mu_2 = \mu_3$ proti H_1 : aspoň jedna dvojice středních hodnot se liší

K testování použijeme jednofaktorovou analýzu rozptylu.

Nejprve ověříme normalitu proměnné IQ ve skupinách obyvatel Skandinávie, Středomoří a západní Evropy pomocí S-W testu a pomocí normálního pravděpodobnostního grafu:



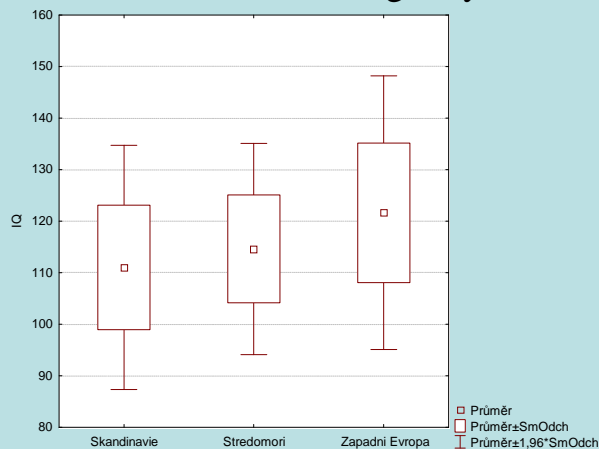
Hypotézu o normalitě proměnné IQ zamítáme na hladině významnosti 0,05 pro obyvatele ze Skandinávie, zde je p-hodnota S-W testu 0,0354. Porušení však není příliš výrazné, proměnnou IQ budeme považovat za normálně rozloženou i ve skupině obyvatel Skandinávie.

Spočteme číselné charakteristiky proměnné IQ v daných třech skupinách:

Puvod	IQ průměr	IQ N	IQ Sm.odch.
Skandinávie	111,0	11	12,09
Stredomori	114,6	13	10,46
Zapadni Evropa	121,6	8	13,54
Vš.skup.	115,1	32	12,16

Průměrné nejvyšší IQ mají obyvatelé západní Evropy (a současně vykazují největší variabilitu), nejnižší obyvatelé Skandinávie.

Nakreslíme krabicové diagramy:



Hypotézu o shodě rozptylů proměnné IQ v daných třech skupinách ověříme pomocí Brownova – Forsytheova testu:

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
IQ	39,59441	2	19,79720	1507,281	29	51,97519	0,380897	0,686616

Testová statistika Brownova – Forsytheova testu nabývá hodnoty 0,3809, odpovídající p-hodnota je 0,6866, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Nyní provedeme test hypotézy o shodě středních hodnot.

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
IQ	528,5481	2	264,2740	4058,952	29	139,9639	1,888159	0,169492

Testová statistika jednofaktorové analýzy rozptylu se realizuje hodnotou 1,8882, odpovídající p-hodnota je 0,1695, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě středních hodnot.

Neprokázali jsme tedy, že by se lišily střední hodnoty IQ obyvatel Skandinávie, Středomoří a západní Evropy.

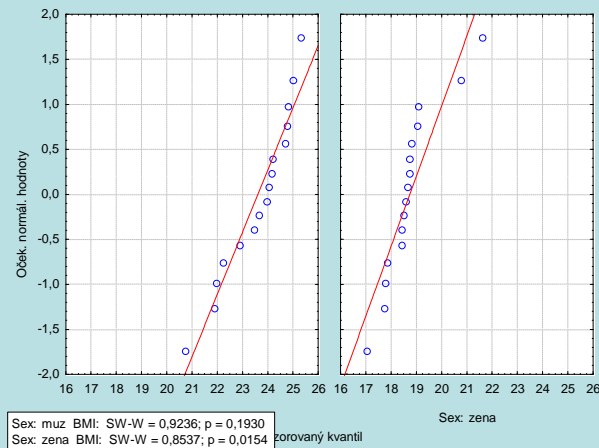
Porovnání proměnné BMI z hlediska pohlaví

Nulová hypotéza tvrdí, že střední hodnoty proměnné BMI jsou stejné pro muže a ženy.

$$H_0 : \mu_1 = \mu_2 \text{ proti } H_1 : \mu_1 \neq \mu_2$$

K testování bychom rádi použili dvouvýběrový t-test.

Nejprve ověříme normalitu proměnné BMI ve skupině mužů a ve skupině žen pomocí S-W testu a pomoci normálního pravděpodobnostního grafu:



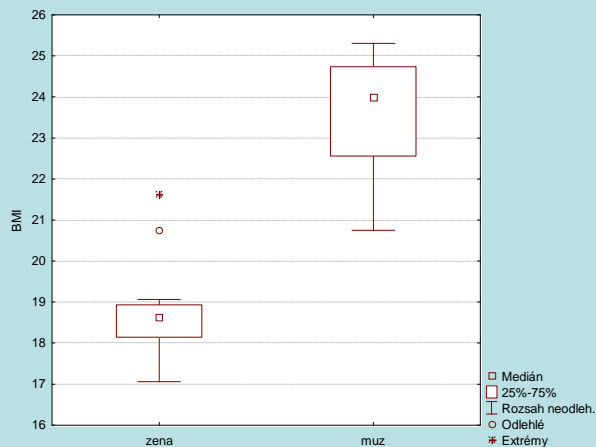
Hypotézu o normalitě proměnné BMI zamítáme na hladině významnosti 0,05 pro ženy. Porušení normality je výraznější, proto použijeme neparametrický test, konkrétně dvouvýběrový Wilcoxonův test.

Vypočítáme číselné charakteristiky proměnné BMI pro muže a pro ženy:

Sex	BMI průměr	BMI N	BMI Sm.odch.
muz	23,6	16	1,32
zena	18,7	16	1,11
Vš.skup.	21,2	32	2,75

Je patrný výrazný rozdíl v průměrném BMI mužů a žen.

Data ještě znázorníme graficky pomocí krabicových diagramů:



Je patrný značný rozdíl v mediánech BMI mužů a žen. U žen se vyskytuje jedna odlehlá a jedna extrémní hodnota BMI.

Nyní provedeme dvouvýběrový Wilcoxonův test:

Proměnná	Sčt poč. muz	Sčt poč. žena	U	Z	p-hodn.	Z upravené	p-hodn.	N platn. muz	N platn. žena	2*1str. přesné p
BMI	390,0000	138,0000	2,000000	4,729959	0,000002	4,729959	0,000002	16	16	0,000000

Vidíme, že p-hodnota je velice blízká 0, tedy na hladině významnosti 0,05 **zamítáme hypotézu, že BMI mužů a žen se neliší.**

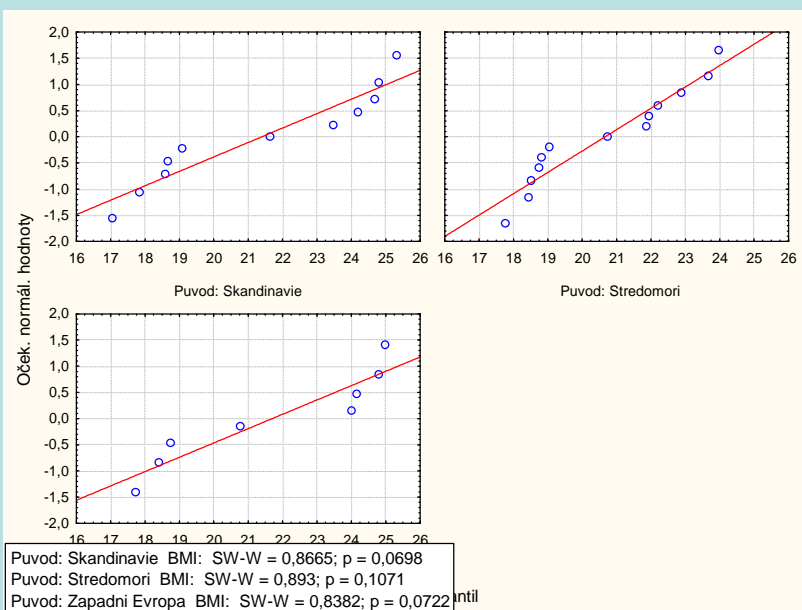
Porovnání proměnné BMI z hlediska původu

Nulová hypotéza tvrdí, že střední hodnoty proměnné BMI jsou stejné pro obyvatele Skandinávie, Středomoří a západní Evropy.

$H_0 : \mu_1 = \mu_2 = \mu_3$ proti H_1 : aspoň jedna dvojice středních hodnot se liší

K testování použijeme jednofaktorovou analýzu rozptylu.

Nejprve ověříme normalitu proměnné BMI ve skupinách obyvatel Skandinávie, Středomoří a západní Evropy pomocí S-W testu a pomoci normálního pravděpodobnostního grafu:



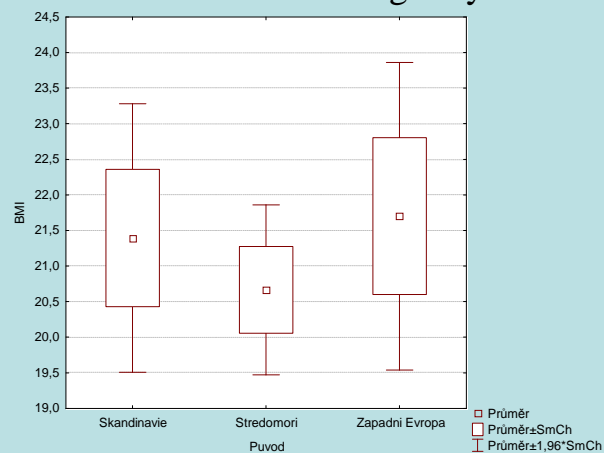
Hypotézu o normalitě proměnné BMI nezamítáme na hladině významnosti 0,05 ani v jednom případě.

Spočteme číselné charakteristiky proměnné IQ v daných třech skupinách:

Puvod	BMI průměr	BMI N	BMI Sm.odch.
Skandinavie	21,4	11	3,19
Stredomori	20,7	13	2,20
Zapadni Evropa	21,7	8	3,12
Vš.skup.	21,2	32	2,75

Průměrné nejvyšší BMI mají obyvatelé západní Evropy, nejnižší obyvatelé Středomoří, u nichž je současně nejnižší variabilita.

Nakreslíme krabicové diagramy:



Hypotézu o shodě rozptylů proměnné BMI v daných třech skupinách ověříme pomocí Brownova – Forsytheova testu:

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
BMI	5,926672	2	2,963336	32,80450	29	1,131190	2,619664	0,089982

Testová statistika Brownova – Forsytheova testu nabývá hodnoty 2,6297, odpovídající p-hodnota je 0,09, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Nyní provedeme test hypotézy o shodě středních hodnot.

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
BMI	6,116344	2	3,058172	228,1304	29	7,866565	0,388756	0,681380

Testová statistika jednofaktorové analýzy rozptylu se realizuje hodnotou 0,3888, odpovídající p-hodnota je 0,6814, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě středních hodnot.

Neprokázali jsme, že by se lišily střední hodnoty BMI obyvatel Skandinávie, Středomoří a západní Evropy.

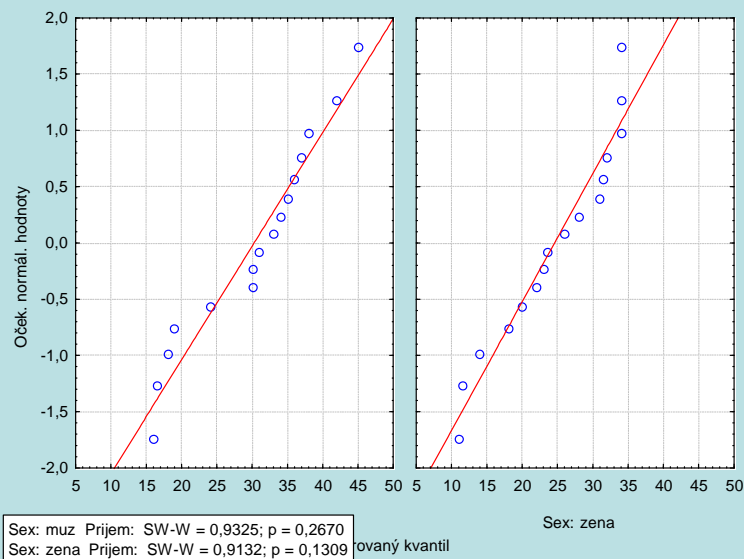
Porovnání proměnné PRIJEM z hlediska pohlaví

Nulová hypotéza tvrdí, že střední hodnoty proměnné PRIJEM jsou stejné pro muže a ženy.

$$H_0 : \mu_1 = \mu_2 \text{ proti } H_1 : \mu_1 \neq \mu_2$$

K testování použijeme dvouvýběrový t-test.

Nejprve ověříme normalitu proměnné PRIJEM ve skupině mužů a ve skupině žen pomocí S-W testu a pomocí normálního pravděpodobnostního grafu:



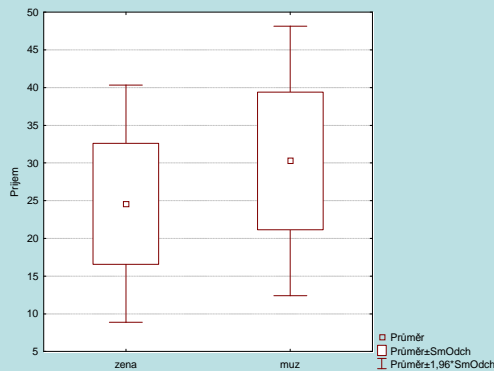
Hypotézu o normalitě proměnné PRIJEM nelze na hladině významnosti 0,05 zamítnout ani pro muže, ani pro ženy.

Vypočítáme číselné charakteristiky proměnné PRIJEM ve skupinách mužů a žen:

Sex	Prijem průměr	Prijem N	Prijem Sm.odch.
muz	30,28	16	9,12
zena	24,59	16	8,03
Vš.skup.	27,44	32	8,93

Vidíme, že rozdíl v průměrném ročním příjmu mužů a žen činí téměř 6 000 euro.

Krabicové diagramy:



Hypotézu o shodě rozptylů proměnné PRIJEM v daných dvou skupinách ověříme pomocí F-testu:

Proměnná	Průměr muz	Průměr žena	t	sv	p	Poč.plat muz	Poč.plat. žena	Sm.odch. muz	Sm.odch. žena	F-poměr Rozptyly	p Rozptyly
Prijem	30,28125	24,59375	1,872954	30	0,070849	16	16	9,117691	8,025415	1,290728	0,627395

Testová statistika F-testu nabývá hodnoty 1,2907, odpovídající p-hodnota je 0,6274, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Současně z tabulky plyne, že testová statistika dvouvýběrového t-testu se realizuje hodnotou 1,873, odpovídající p-hodnota je 0,0708, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě středních hodnot.

Neprokázali jsme, že by se lišily střední hodnoty proměnné PRIJEM mužů a žen.

Cohenův koeficient věcného účinku je 0,31, tedy vliv proměnné SEX na příjem je pouze malý.

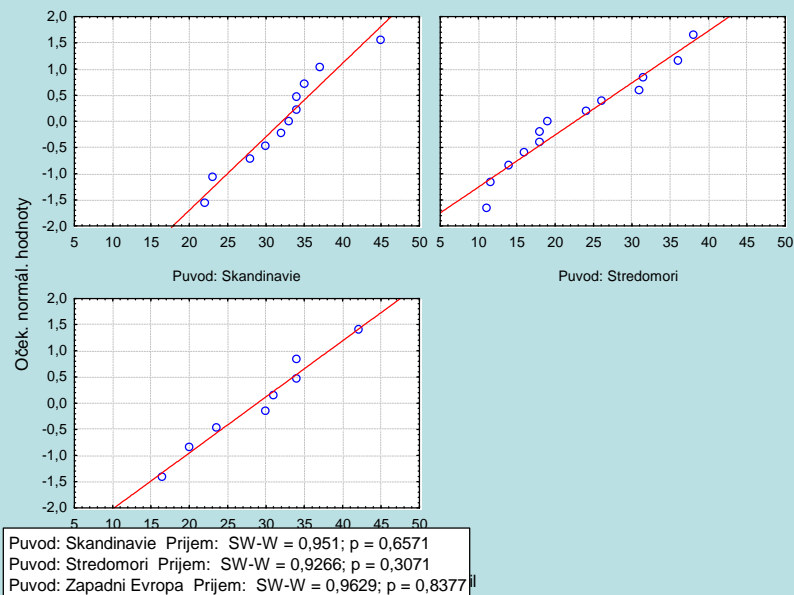
Porovnání proměnné PRIJEM z hlediska původu

Nulová hypotéza tvrdí, že střední hodnoty proměnné PRIJEM jsou stejné pro obyvatele Skandinávie, Středomoří a západní Evropy.

$H_0 : \mu_1 = \mu_2 = \mu_3$ proti H_1 : aspoň jedna dvojice středních hodnot se liší

K testování použijeme jednofaktorovou analýzu rozptylu.

Nejprve ověříme normalitu proměnné PRIJEM ve skupinách obyvatel Skandinávie, Středomoří a západní Evropy pomocí S-W testu a pomocí normálního pravděpodobnostního grafu:



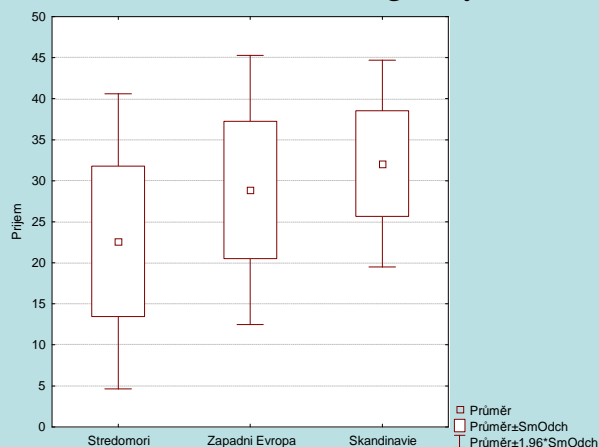
Hypotézu o normalitě proměnné PRIJEM nezamítáme na hladině významnosti 0,05 ani v jednom případě.

Spočteme číselné charakteristiky proměnné PRIJEM v daných třech skupinách:

Puvod	Prijem průměr	Prijem N	Prijem Sm.odch.
Skandinávie	32,09	11	6,43
Stredomori	22,62	13	9,18
Zapadni Evropa	28,88	8	8,38
Vš.skup.	27,44	32	8,93

Průměrný nejvyšší příjem mají obyvatelé Skandinávie, nejnižší obyvatelé Středomoří. Nejnižší variabilitu příjmů vykazují obyvatelé Skandinávie.

Nakreslíme krabicové diagramy:



Hypotézu o shodě rozptylů proměnné PRIJEM v daných třech skupinách ověříme pomocí Brownova – Forsytheova testu:

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
Prijem	53,63571	2	26,81785	823,8330	29	28,40804	0,944024	0,400691

Testová statistika Brownova – Forsytheova testu nabývá hodnoty 0,944, odpovídající p-hodnota je 0,4, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů.

Nyní provedeme test hypotézy o shodě středních hodnot.

Proměnná	SČ efekt	SV efekt	PČ efekt	SČ chyba	SV chyba	PČ chyba	F	p
Prijem	557,0140	2	278,5070	1914,861	29	66,02969	4,217905	0,024666

Testová statistika jednofaktorové analýzy rozptylu se realizuje hodnotou 4,218, odpovídající p-hodnota je 0,0247, tedy na hladině významnosti 0,05 **zamítáme hypotézu o shodě středních hodnot příjmů obyvatel Skandinávie, Středomoří a západní Evropy.**

Použijeme Scheffého metodu mnohonásobného porovnávání, abychom našli dvojice zemí s odlišnou střední hodnotou příjmů

Puvod	{1}	{2}	{3}
	M=32,091	M=22,615	M=28,875
Skandinávie {1}		0,028087	0,698895
Středomoří {2}	0,028087		0,246690
Západní Evropa {3}	0,698895	0,246690	

Na hladině významnosti 0,05 se liší střední hodnota příjmů obyvatel Skandinávie a Středomoří.

Závěr

Při analýze datového souboru jsme došli k těmto závěrům:

V souboru 32 osob je stejný počet mužů a žen.

34,4% osob pochází ze Skandinávie, 40,6% ze Středomoří a 25% ze západní Evropy.

Na hladině významnosti 0,05 jsme neprokázali, že by se lišily střední hodnoty IQ mužů ($m_1 = 115$) a žen ($m_2 = 115,25$).

Na hladině významnosti 0,05 jsme neprokázali, že by se lišily střední hodnoty IQ obyvatel Skandinávie ($m_1 = 111$), Středomoří (114,6) a západní Evropy (121,6).

Na hladině významnosti 0,05 jsme prokázali, že mediány BMI se liší pro muže ($x_{0,50} = 23,99$) a pro ženy ($y_{0,50} = 18,63$).

Na hladině významnosti 0,05 jsme neprokázali, že by se lišily střední hodnoty BMI obyvatel Skandinávie ($m_1 = 21,4$), Středomoří ($m_2 = 20,7$) a západní Evropy ($m_3 = 21,7$).

Na hladině významnosti 0,05 jsme neprokázali, že by se lišily střední hodnoty příjmů mužů ($m_1 = 30\,280$ euro) a žen ($m_2 = 24\,594$ euro).

Na hladině významnosti 0,05 jsme prokázali, že se liší střední hodnoty příjmů obyvatel Skandinávie ($m_1 = 32\,091$ euro), Středomoří ($m_2 = 22\,615$ euro) a západní Evropy ($m_3 = 28\,875$ euro).