

Analýza rozptylu, jednoduché třídění – řešení.

Datový soubor CHOLESTEROL.

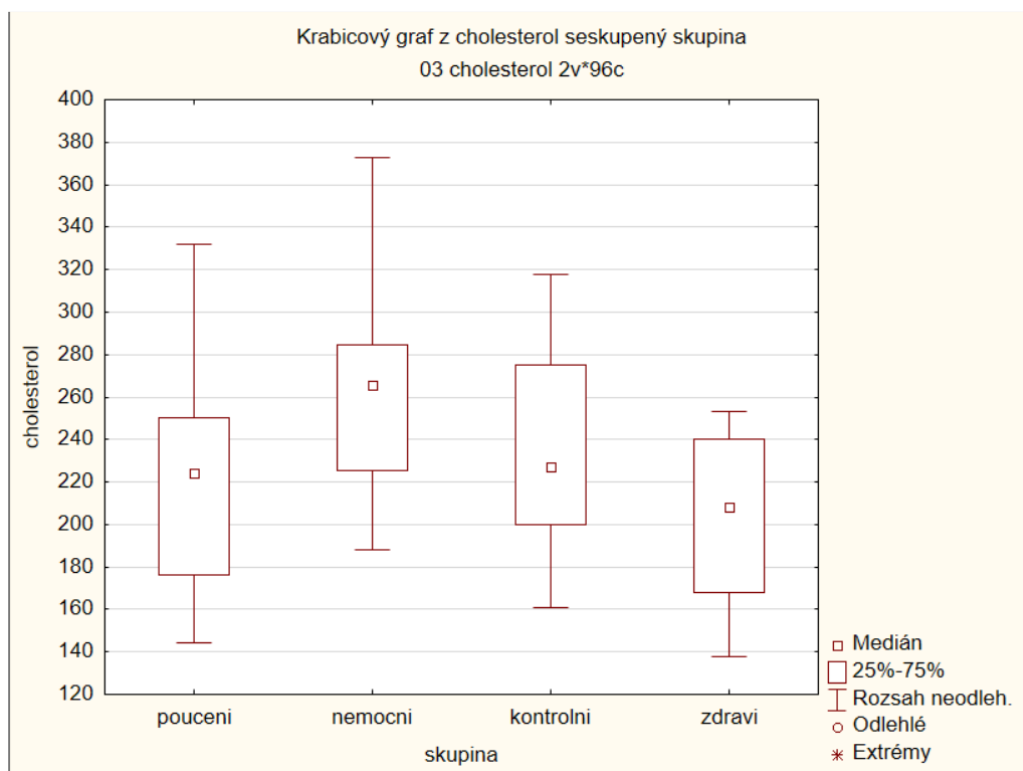
Necelé stovce mužů, zahrnutých do výzkumu kardiovaskulárních nemocí, byla měřena hladina cholesterolu v krvi. Zároveň byl každý muž označen jako „zdravý“, „nemocný“, „s rizikem onemocnění a poučený“ nebo „s rizikem onemocnění a kontrolní (= nepoučený)“. Máme posoudit, jak se tyto skupiny mužů mezi sebou liší v naměřené hladině cholesterolu a zda jsou obecně tyto rozdíly statisticky významné.

Počáteční úvaha: čtyři skupiny/výběry mužů (zdraví, poučení, kontrolní skupina a nemocní) a jejich hladina cholesterolu v krvi. Zajímá nás, jestli se skupiny od sebe průkazně liší, tedy zda populační průměry hladiny cholesterolu (μ_{skup}) jsou průkazně odlišné.

Nulová hypotéza: populační průměry hladiny cholesterolu jsou shodné pro všechny skupiny, tedy

$$\mu_{zdravi} = \mu_{pouceni} = \mu_{kontrolni} = \mu_{nemocni}$$

Jsou to kvantitativní data roztržiděná podle faktoru „skupina“, řeším analýzou rozptylu, jednoduché třídění.



Statistiky → ANOVA →

Jednofaktorová
ANOVA

Tlačítko
„Více výsledků“,

potom záložky
„Předpoklady“,
„Detaily“ a „Post-hoc“.

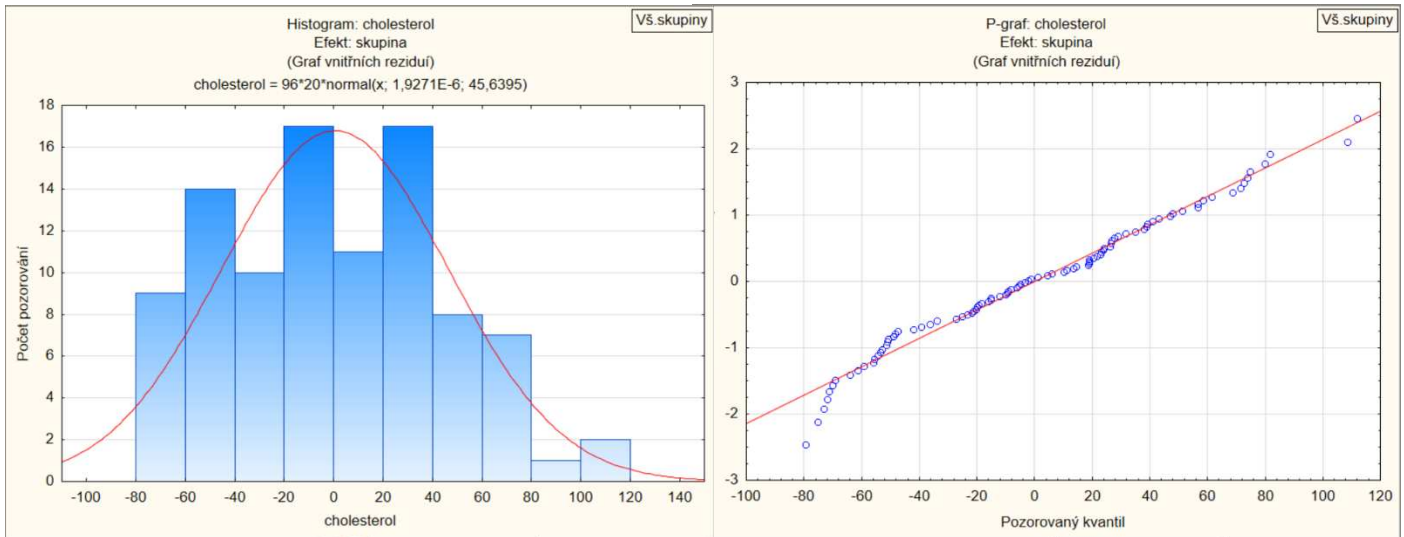
Předpoklady: nezávislost – musí být splněno při sběru dat;

Homogenita rozptylů, tj. rozptyly všech skupin jsou srovnatelné – Leveneův test: splněno, $p = 0,63$.

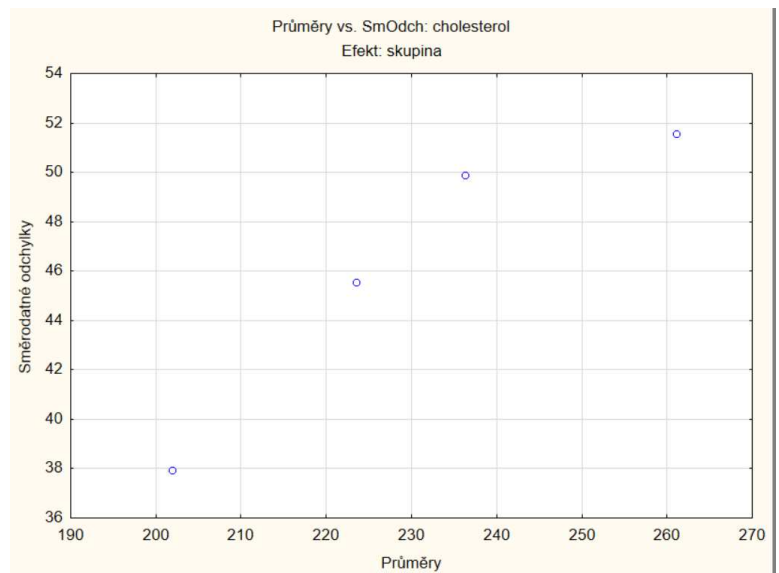
Leveneův test homogenity rozptylů (03 cholesterol)				
Efekt: skupina				
Stupně volnosti pro všechna F: 3, 92				
	PČ Efekt	PČ Chyba	F	p
cholesterol	368,7126	634,6887	0,580935	0,628998

Normální rozdělení dat v rámci skupin – jsou-li srovnatelné rozptyly, potom mohou být jen rezidua (hladina cholesterolu jednotlivce – průměrná hladina ve skupině) a kontrolovat normalitu všech reziduí dohromady (grafy níže).

→ Od pohledu zdá se splněno. Máme celkem dost pozorování, budeme pokračovat.



Další kontrolní graf: graf průměrů proti směrodatným odchýlkám pro jednotlivé skupiny. Podle něj usuzujeme, zda s rostoucím průměrem neroste také rozptyl (směrodatná odchýlka). Závislost se může objevit i v případě, že Leveneův test nezamítnul shodnost rozptylů. V tomto příkladu se tato závislost objevuje, měli bychom proto vyzkoušet logaritmickou transformaci dat, která tuto závislost dokáže přinejmenším omezit a celkové výsledky zlepšit. Po transformaci se výsledky skutečně zlepší (zde neuvedeno).

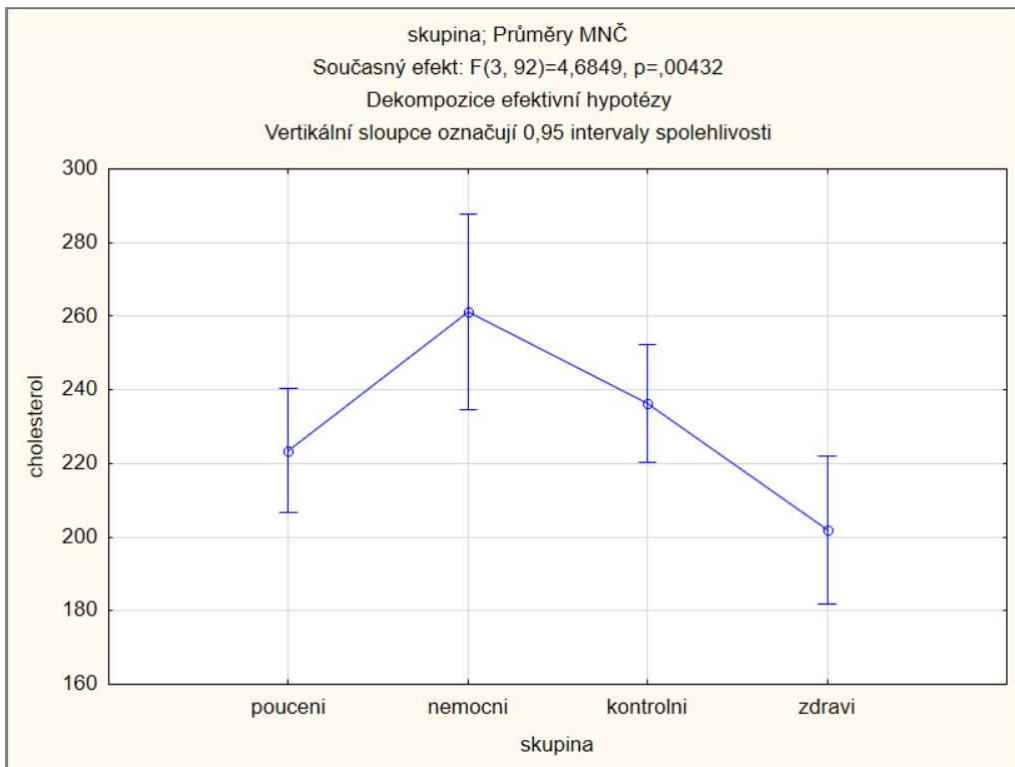


Výsledky analýzy: v řádce „skupina“ je test nulové hypotézy o rovnosti populačních průměrů → zamítám na (předem zvolené) hladině testu $\alpha = 0,05$. Testová statistika $F_{3,92} = 4,685$, p-hodnota = 0,0043.

V řádce „chyba“ jsou reziduální součty čtverců (SČ) a průměrný reziduální součet čtverců (PČ), což je zároveň odhad variance (rozptylu) σ^2 v datech.

Jednorozměrné testy významnosti pro cholesterol (03 cholesterol)					
Sigma-omezená parametrizace					
Dekompozice efektivní hypotézy					
Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	4375996	1	4375996	2034,509	0,000000
skupina	30230	3	10077	4,685	0,004325
Chyba	197881	92	2151		

Graf výběrových průměrů a 95% konfidenčních intervalů pro skutečnou polohu populačních průměrů:



Mnohonásobná porovnání: Tukeyův test pro nestejná N (HSD nestejná N).

Průkazně se liší jen skupiny zdravých a nemocných mužů (jejich neznámé populační průměry). Pro zbylé dvě skupiny nelze na základě našich dat rozhodnout, zda jejich odhadované populační průměry jsou průkazně odlišné od některého z ostatních odhadovaných populačních průměrů.

HSD při nestejných N; proměnná cholesterol (03 cholesterol)
 Přibližné pravděpodobnosti pro post hoc testy
 Chyba: meziskup. PČ = 2150,9, sv = 92,000

Č. buňky	skupina	{1}	{2}	{3}	{4}
1	poučení	223,50	0,201210	0,712196	0,438459
2	nemocní	0,201210	261,08	0,557778	0,012678
3	kontrolní	0,712196	0,557778	236,24	0,085090
4	zdraví	0,438459	0,012678	0,085090	201,95

Záložka POST-HOC (Více výsledků).

Výsledné p-hodnoty v matici:

Volba ZOBRAZIT = VÝZNAMNÉ DIFERENCE

Záložka POST-HOC (Více výsledků).

Výsledek jako seznam skupin, které od sebe nejdu rozlišit. V jednom sloupečku jsou skupiny, pro které nezamítám hypotézu o shodnosti populačních průměrů (testováno po dvojicích):

Volba ZOBRAZIT = HOMOGENNÍ SKUPINY

HSD při nestejných N; proměnná cholesterol				
Homogenní skupiny, alfa = ,05000				
Chyba: meziskup. PČ = 2150,9, sv = 92,000				
Č. buňky	skupina	cholesterol Průměr	1	2
4	zdraví	201,9524	****	
1	poučení	223,5000	****	****
3	kontrolní	236,2424	****	****
2	nemocní	261,0833		****