**Analýza rozptylu, jednoduché třídění – řešení.**

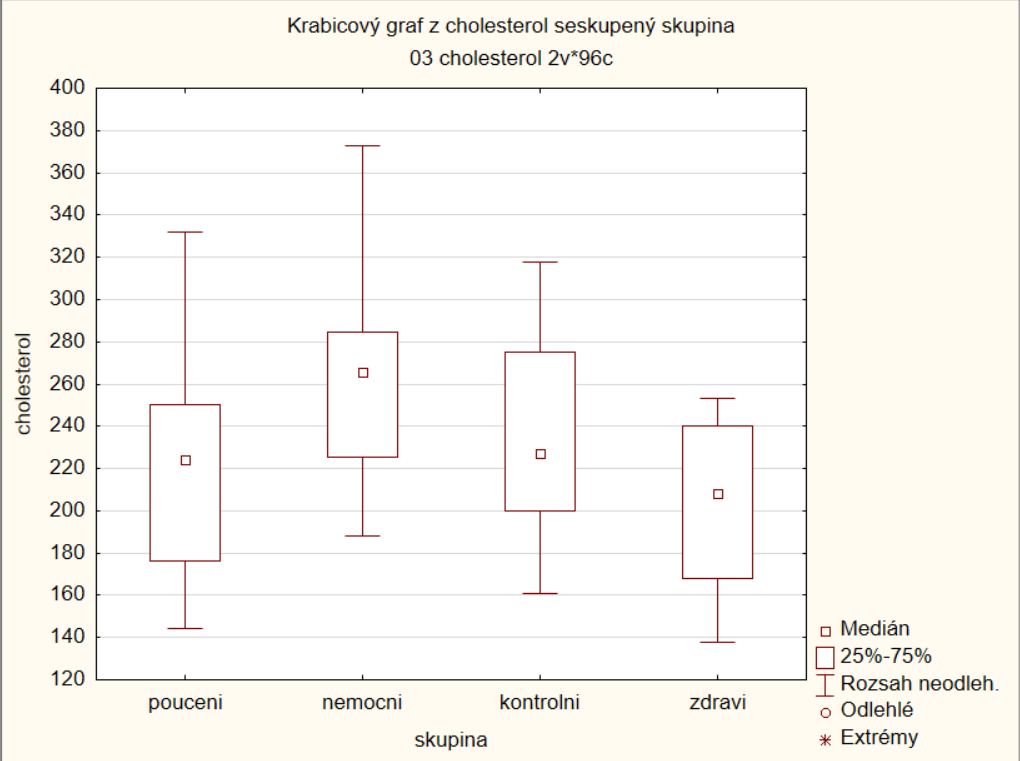
**Datový soubor CHOLESTEROL.**

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 (***μskup***) jsou průkazně odlišné.

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

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

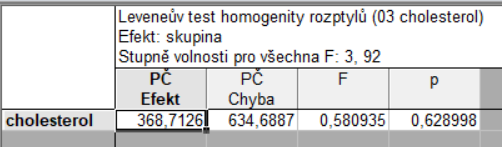


Statistiky 🡪 ANOVA 🡪 Jednofaktorová ANOVA

Tlačítko „Více výsledků“,

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

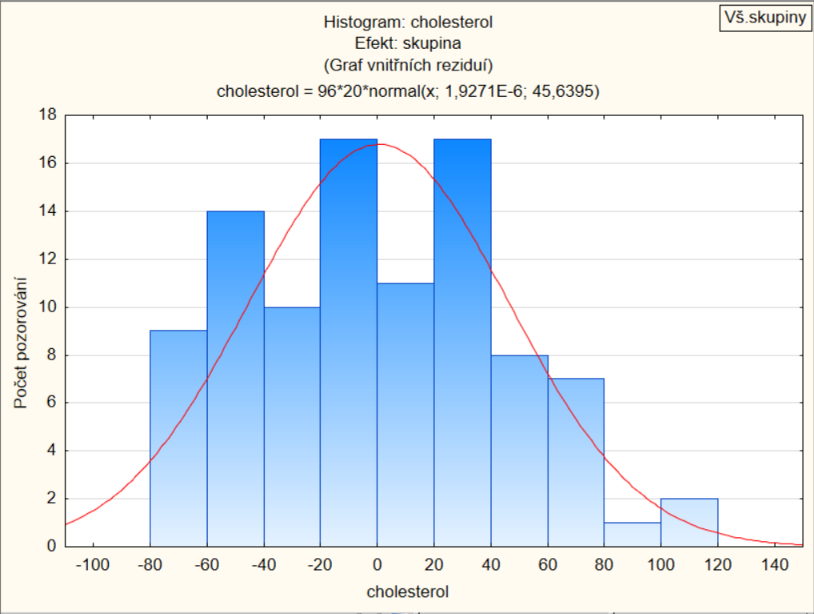
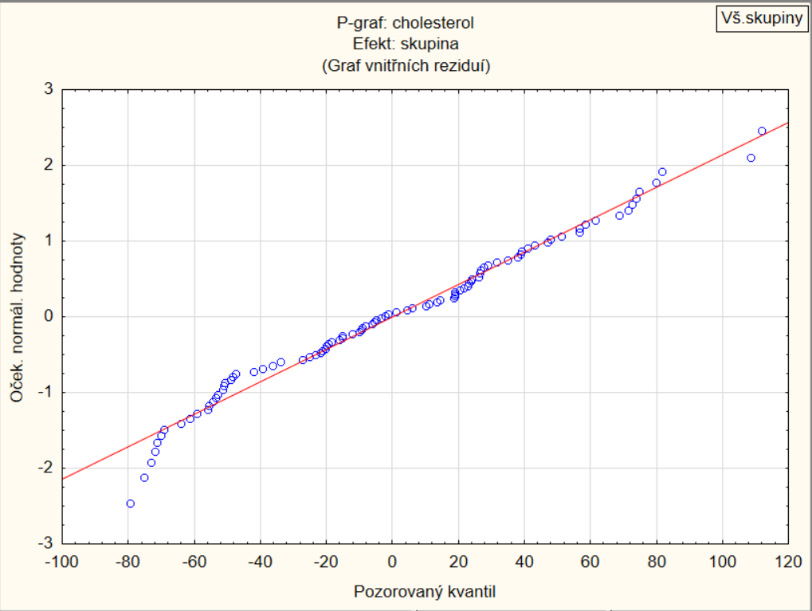
**Předpoklady:** **nezávislost** – musí být splněno při sběru dat;

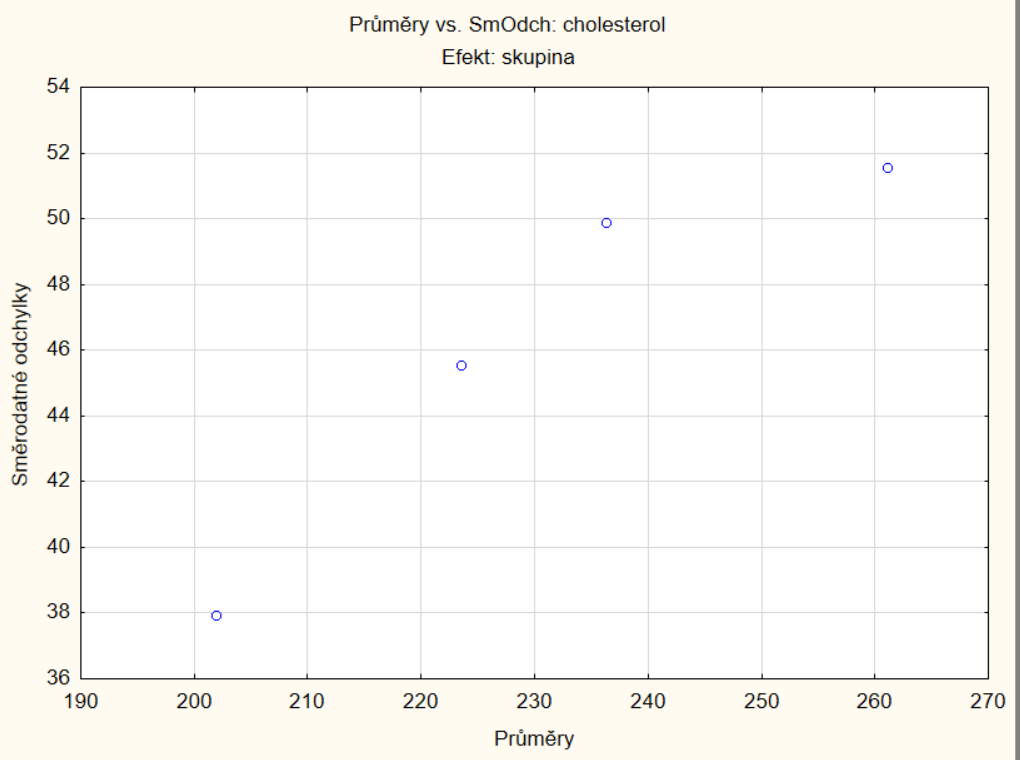


**Homogenita rozptylů**, tj. rozptyly všech skupin jsou srovnatelné – Leveneův test: splněno, p = 0,63.

**Normální rozdělení dat** v rámci skupin – jsou-li srovnatelné rozptyly, potom mohu vzít 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.

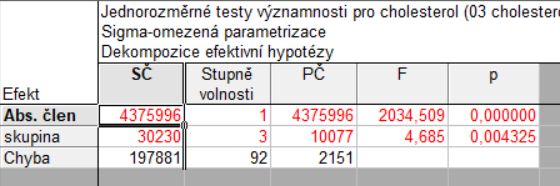


**Další kontrolní graf**: graf průměrů proti směrodatným odchylkám pro jednotlivé skupiny. Podle něj usuzujeme, zda s rostoucím průměrem neroste také rozptyl (směrodatná odchylka). Závislost se může objevit i v případě, že Leveneův test nezamítnul shodnost rozptylů. V tomto příkladu se tato závislost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

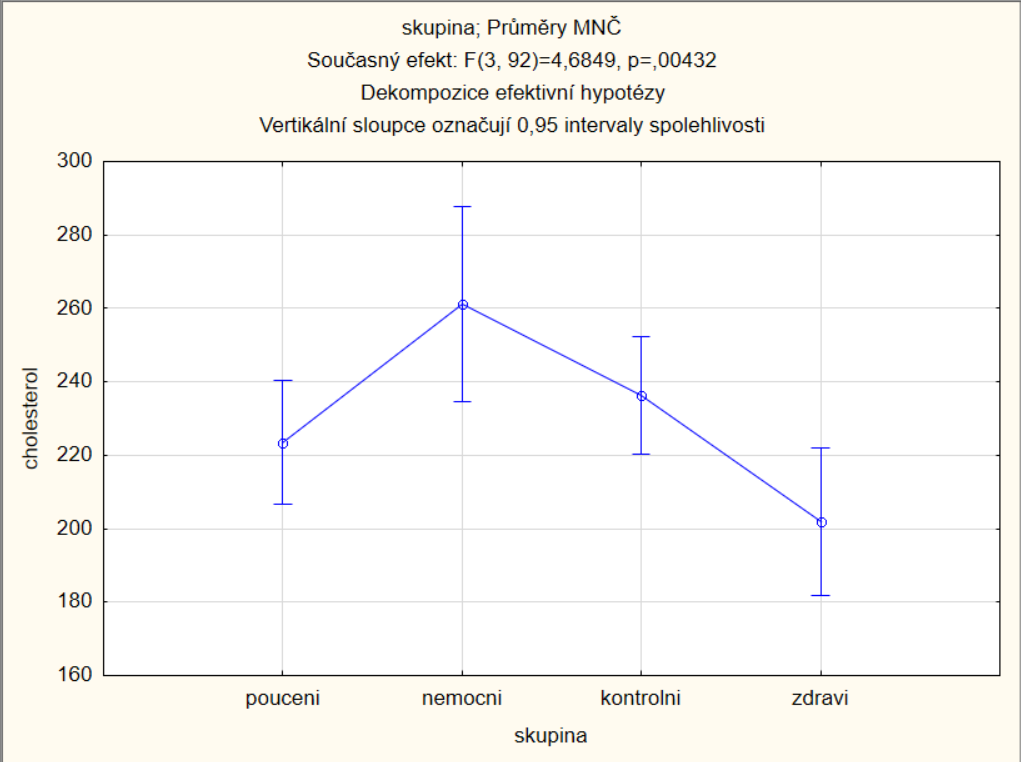
Po transformaci se výsledky skutečně zlepší (zde neuvedeno).

**Výsledky analýzy:** v řádku „skupina“ je test nulové hypotézy o rovnosti populačních průměrů 🡪 zamítám na (předem zvolené) hladině testu α = 0,05. Testová statistika F3,92 = 4,685, p-hodnota = 0,0043.

V řádku „chyba“ jsou reziduální součty čtverců (SČ) a průměrný reziduální součet čtverců (PČ), což je zároveň odhad variance (rozptylu) σ2 v datech.

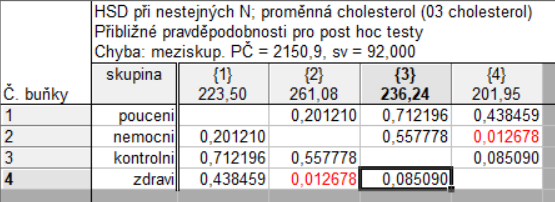
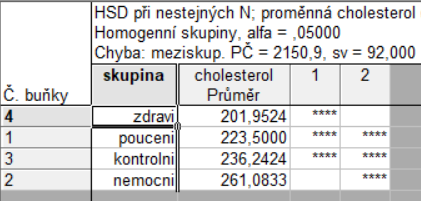


Graf výběrových průměrů a 95% konfidenčních intervalů pro skutečnou polohu populačních průměrů:



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

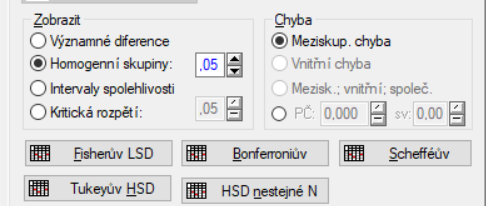
Průkazně se liší jen skupiny zdravých a nemocných mužů (jejich neznámé populační průměry). Pro zbylé dvě skupiny nelze na základě našich dat rozhodnout, zda jejich odhadované populační průměry jsou průkazně odlišné od některého z ostatních odhadovaných populačních průměrů.

Záložka POST-HOC (Více výsledků).

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

Volba ZOBRAZIT = HOMOGENNÍ SKUPINY



Záložka POST-HOC (Více výsledků).

Výsledné p-hodnoty v matici:

Volba ZOBRAZIT = VÝZNAMNÉ DIFERENCE

