9 Analýza rozptylu jednoduchého třídění - Domácí procvičování

Příklad 9.1. Pan Novák může cestovat z místa bydliště do místa pracoviště třemi různými způsoby: tramvají, autobusem a metrem s následným přestupem na tramvaj. Máme k dispozici jeho naměřené časy cestování do práce v době ranní špičky (včetně čekání na příslušný spoj) v minutách:

autobus:	32	39	42	37	34	38	
tramvaj:	30	34	28	26	32		
metro:	40	37	31	39	38	33	34

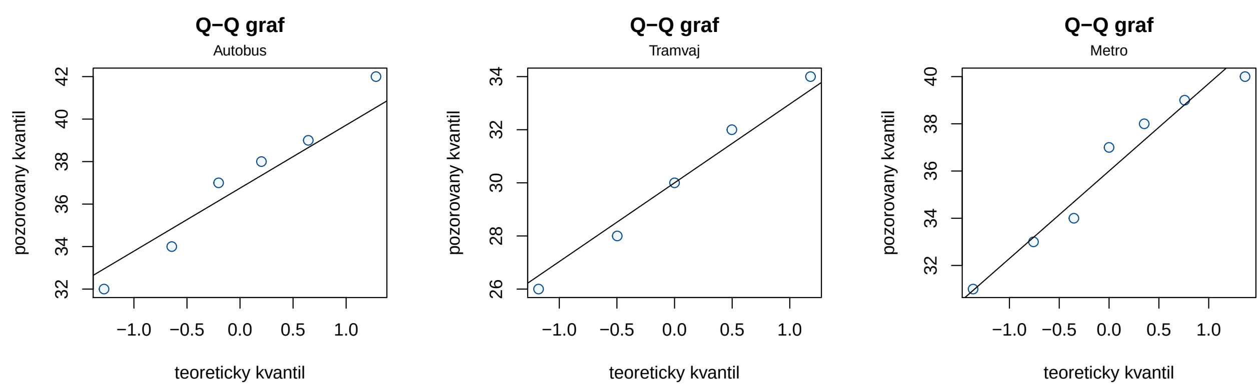
Pro všechny tři způsoby dopravy vypočtěte průměrné časy cestování. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že doba cestování do práce nezávisí na způsobu dopravy. V případě zamítnutí nulové hypotézy zjistěte, které způsoby dopravy do práce se od sebe liší na hladině významnosti $\alpha = 0.05$.

Testování normality

1. H_0 : ....
2. H_1 : ....

Na testování normality všech tří výběrů použijeme kvůli jejich malým rozsahům .... test.

```
## [1] "Autobus: 0.9539"
## [1] "Tramvaj: 0.9672"
## [1] "Metro:   0.6294"
```



Protože ve všech případech je p-hodnota testu .... než $\alpha =$ ...., nulovou hypotézu o normalitě dat .... na hladině významnosti .... Všechny tři výběry tedy .... z normálního rozdělení.

Test homogenity rozptylů

1. H_0 : ....
2. H_1 : ....

Jelikož náhodné výběry pochází z normálního rozdělení, na testování hypotézy o shodě rozptylů použijeme .... test.

```
## Test Statistic
##      0.1053556
## [1] 0.9006645
```

Testovací statistika .... testu nabývá hodnoty ...., odpovídající p-hodnota = ... je .... než α = ..., tedy na hladině významnosti .... hypotézu o shodě rozptylů ....

Test o shodě středních hodnot:

1. H_0 : ....
2. H_1 : ....

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## factor(ID)   2    154   77.00   6.715 0.00827 **
## Residuals   15    172   11.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] 3.68232
```

Skupinový součet čtverců S_A = ..., počet stupňů volnosti f_A = ..., reziduální součet čtverců S_E = ..., počet stupňů volnosti f_E = ..., testovací statistika F_A = ... p-hodnota = ..., nulovou hypotézu o shodě středních hodnot tedy .... na hladině významnosti ....

Kritický obor má tvar W=.. Protože ...., hypotézu H_0 o shodě středních hodnot .... na hladině významnosti ....

S rizikem omylu nejvýše 5 % se prokázal/neprokázal rozdíl v dobách cestování pana Nováka do zaměstnání autobusem, tramvají a metrem.

Metoda mnohonásobného porovnávání

Jelikož jsme nulovou hypotézu o shodě středních hodnot ...., chceme nyní zjistit, které dvojice středních hodnot se od sebe významně liší. Stanovíme nulové a alternativní hypotézy pro dvojice středních hodnot

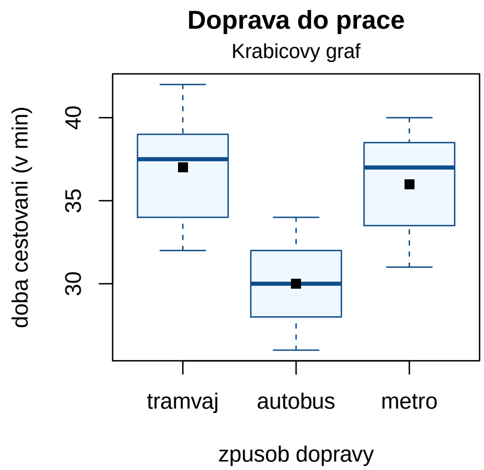
- H_{01} : .... oproti H_{11} : ....
- H_{02} : .... oproti H_{12} : ....
- H_{03} : .... oproti H_{13} : ....

Protože v každé skupině máme různý počet pozorování, použijeme na mnohonásobné porovnávání .... metodu.

```
## $L
##         tramvaj autobus metro
## tramvaj       0       7     1
## autobus       7       0     6
## metro         1       6     0
##
## $R
##          tramvaj  autobus    metro
## tramvaj 5.305591 5.564551 5.112595
## autobus 5.564551 5.811983 5.380851
## metro   5.112595 5.380851 4.912023
```

Porovnáním pravé a levé strany .... metody vidíme, že na hladině významnosti .... zamítáme nulovou hypotézu o shodě středních hodnot $\mu_{........}$ a $\mu_{........}$ a středních hodnot $\mu_{........}$ a $\mu_{........}$. Z tabulky vyplývá, že s rizikem omylu nejvýše 5 % se liší cestování .... a .... a dále cestování .... a ....

Krabicový graf



Příklad 9.2. V rámci studie bylo získáno pět nezávislých náhodných výběrů o rozsazích 13, 18, 69, 19, 44 přičemž i-tý výběr pochází z rozdělení $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, 5$. Skupinový součet čtverců vyšel 80 a celkový součet čtverců vyšel 3441. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu H_0 o shodě středních hodnot. Testování proveďte pomocí kriického oboru a pomocí p-hodnoty.

```
## [1] 0.9401964
## [1] 0.4423163
## [1] 2.428885
```

Testovací statistika F nabývá hodnoty .... Kritický obor má tvar W=..................................... . Protože ...., hypotézu H_0 o shodě středních hodnot .... na hladině významnosti ....

Testovací statistika F nabývá hodnoty ...., p-hodnota = ... Protože p-hodnota= .... je .... než $\alpha =$ ...., hypotézu H_0 o shodě středních hodnot .... na hladině významnosti ....

Příklad 9.3. Na střední škole byl uskutečněn experminet zjišťující efektvitu jednotlivých pedagogických metod. Studenti byli rozděleni do pěti skupin a každá skupina byla vyučována pomocí jedné z pedagogických metod: tradiční způsob, programová výuka, audiotechnika, audiovizuální technika a vizuální technika. Z každé skupiny byl potom vybrán náhodný vzorek studentů a všichni byli podrobeni témuž písemnému testu. Výsledky testu jsou uvedeny v následující tabulce a v souboru pet_metod.txt:

metoda	počet bodů							
tradicni	76.2	48.3	85.1	63.7	91.6	87.2		
programova	85.2	74.3	76.5	80.3	67.4	67.9	72.1	60.4
audio	67.3	60.1	55.4	72.3	40.0			
audiovizualni	75.8	81.6	90.3	78.0	67.8	57.6		
vizualni	50.5	70.2	88.8	67.1	77.7	73.9		

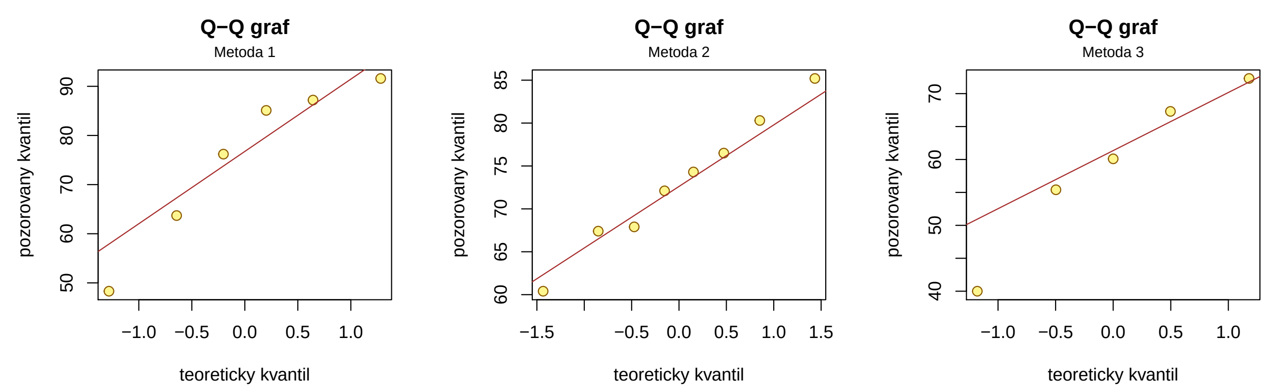
Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že znalosti všech studentů jsou stejné a nezávisí na použité pedagogické metodě. V případě zamítnutí hypotézy zjistěte, které výběry se liší na hladině významnosti 0.05.

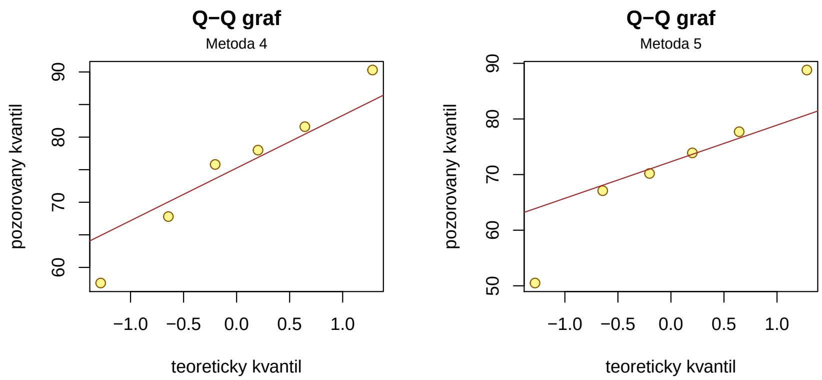
Testování normality

1. H_0 : ....
2. H_1 : ....

Na testování normality všech tří výběrů použijeme kvůli jejich malým rozsahům .... test.

```
## [1] "Tradicni m.:      0.4177"
## [1] "Programova m.:    0.9966"
## [1] "Audio m.:         0.7663"
## [1] "Audiovizualni m.: 0.9577"
## [1] "Viualni m.:       0.8814"
```





Protože ve všech pěti případech je p-hodnota testu .... než $\alpha =$ ...., nulovou hypotézu o normalitě dat u každé metody .... na hladině významnosti .... Všech pět výběrů tedy .... z normálního rozdělení.

Test homogenity rozptylů

1. H_0 : ....
2. H_1 : ....

Jelikož náhodné výběry pochází z normálního rozdělení, na testování hypotézy o shodě rozptylů všech pěti výběrů použijeme .... test.

```
## Test Statistic
##      0.8190294
## [1] 0.5247907
```

Testovací statistika .... testu nabývá hodnoty ...., odpovídající p-hodnota = ... je .... než $\alpha =$ ...., tedy na hladině významnosti .... hypotézu o shodě rozptylů ....

Test o shodě středních hodnot:

1. H_0 : ....
2. H_1 : ....

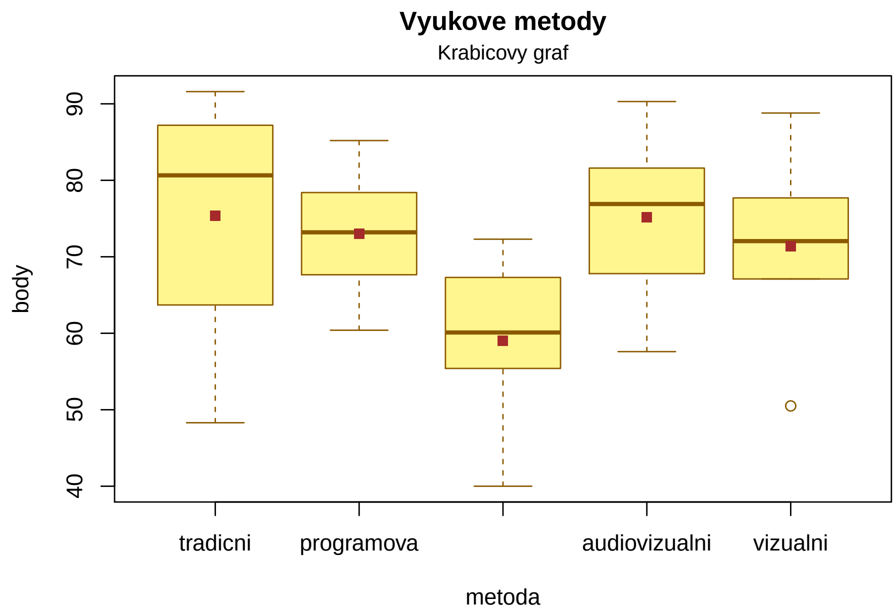
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## factor(ID)   4    966   241.6   1.624  0.198
## Residuals   26   3869   148.8
## [1] 2.742594
```

Skupinový součet čtverců $S_A =$ ...., počet stupňů volnosti $f_A =$ ...., reziduální součet čtverců $S_E =$ ...., počet stupňů volnosti $f_E =$ ...., testovací statistika $F_A =$ .... p-hodnota = ..., nulovou hypotézu o shodě středních hodnot tedy .... na hladině významnosti ....

Kritický obor má tvar W=.. Protože ...., hypotézu H_0 o shodě středních hodnot .... na hladině významnosti ....

V účinnosti jednotlivých pedagogických metod tedy .... významný rozdíl.

Krabicový graf



Příklad 9.4. Je dána neúplná tabulka analýzy rozptylu jednoduchého třídění. Na volná místa doplňte chybějící hodnoty a na hladině významnosti 0.05 testujte hypotézu o shodě středních hodnot. Stanovte, jaký je celkový počet pozorování n a kolik úrovní r má faktor A?

zdroj variability	součet čtverců	stupně volnosti	podíl	statistika F
skupinový			3	
reziduální	3	26		-
celkový		30	-	-

```
## [1] 4
## [1] 12
## [1] 15
## [1] 0.1153846
## [1] 26
## [1] 2.742594
## [1] 9.33888e-09
```

Testovací statistika F nabývá hodnoty .... Kritický obor má tvar W=...................................... . Protože ...., hypotézu H_0 o shodě středních hodnot .... na hladině významnosti ....

Testovací statistika F nabývá hodnoty ...., p-hodnota = ... Protože p-hodnota= .... je .... než $\alpha =$ ...., hypotézu H_0 o shodě středních hodnot .... na hladině významnosti ....

Celkový počet pozorování je ....; faktor A má celkem .... úrovní.