

4. Úkol na předmět Mnohorozměrná analýza

Jaromír Macoun

Březen 2022

1 Úvod

Pro řešení úkolu jsem si vybral data o akcích amerických firem. Vybral jsem si stejný datový soubor jako při prvním úkolu, kde jsme prováděli explorativní analýzu. Data obsahují informaci o denních výsledcích jednotlivých akcií. Opět jako v prvním úkole budeme uvažovat, že denní výsledky jsou vzájemně nezávislé.

Pro testování jsem si konkrétně vybral akcie PanAm a Delta, to jsou akcie dvou leteckých společností a akcie DEC a Texaco, akcie dvou ropných společností. A chceme vědět zdali se očekávané denní výnosy PanAM a DEC liší a zároveň jestli se liší očekávané denní výnosy Delta a Texaco. Budeme tedy využívat dvouvýběrový Hottelingův T^2 test. Prvně se podíváme zdali je rozumné tento test využít.

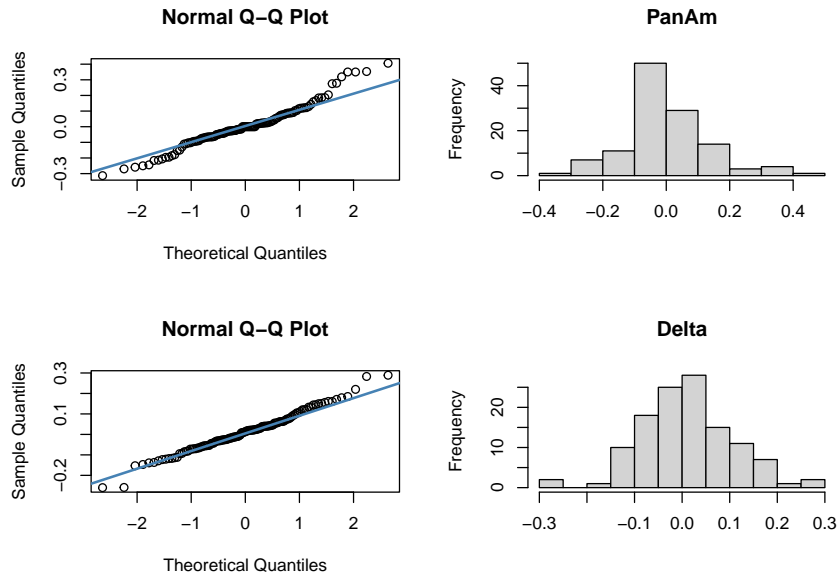
Marginální rozdělení: Na obrázku č. 1 vidíme, že pro Delta asi normální rozdělení bude mít. Jak histogram tak QQ graf se zdají v pořádku.

Pro PanAm tyto data asi normální nebudou, nicméně na základě obrázků se zdá, že normalita není nějak extrémně porušena.

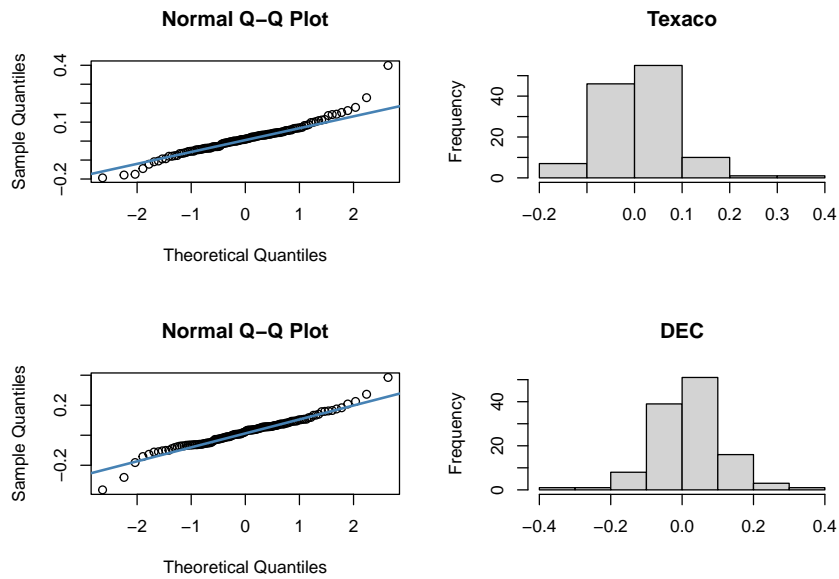
Stejně závěry by byly pro zbylé akcie.

Sdružená normalita: Jak jsme si uvědomili v minulém cvičení, tak marginální normalita neimplikuje sdruženou normalitu. Zatím jsme si nepředvedli žádné sofistikovanější metody, jak říci jestli data mají mnohorozměrné normální rozdělení. Uvedeme tedy 3D odhad hustoty.

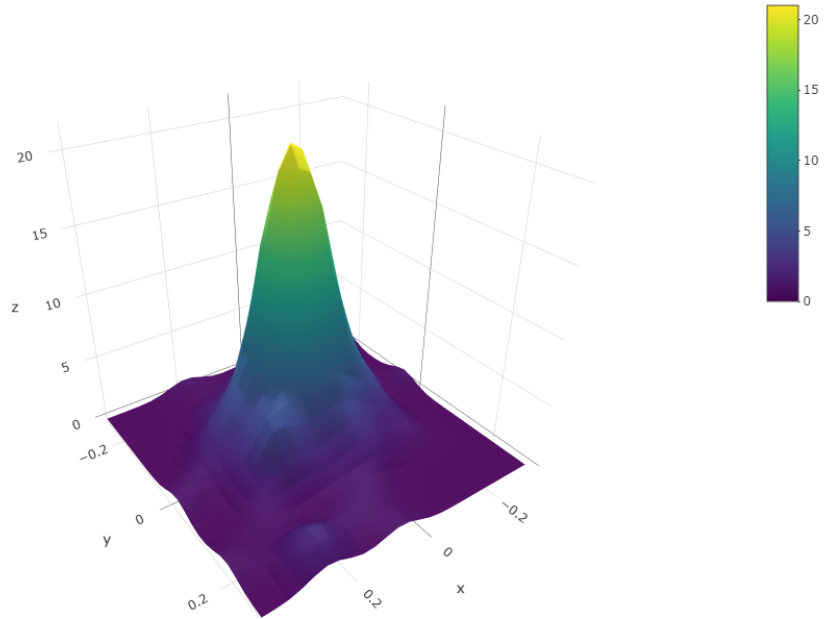
Jak je vidět na obrázcích č. 3 a 4, tak hustota se zdá být symetrická. Například na obrázku č.3 je vidět v bodě $(0, 2; 0, 2)$ je vidět nějaký shluk pozorování, což by asi normální rozdělení nemělo. Nicméně těchto pozorování není moc. Sdruženě normální rozdělení to sice není nicméně se od něho zas tolik neliší.



Obrázek 1: Marginální rozdělení aerolinií



Obrázek 2: Marginální rozdělení aerolinií



Obrázek 3: Sdružená hustota aerolinií

2 Hotellingův test

Máme tedy dva náhodné výběry $\mathbf{X}_1, \dots, \mathbf{X}_{120}$ a $\mathbf{Y}_1, \dots, \mathbf{Y}_{120}$ obě se sdruženým normálním rozdělením s vektory středních hodnot μ_x a μ_y . Protože neznáme varianční matice budeme muset předpokládat homoskedasticitu těchto dvou náhodných výběrů.

Hypotéza:

$$H_0 : \mu_x = \mu_y \quad H_1 : \mu_x \neq \mu_y$$

Testová statistika:

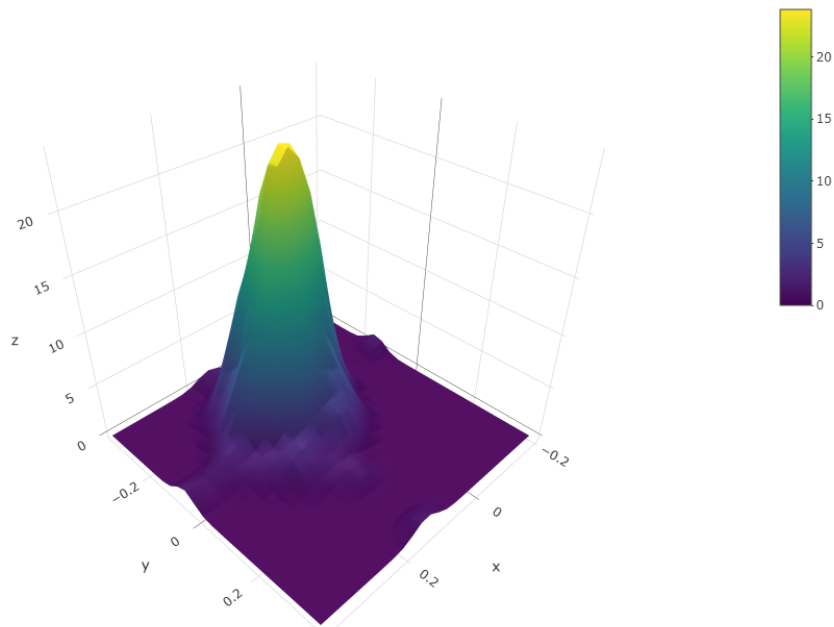
$$\frac{nm(n+m-p-1)}{p(n+m)^2} (\bar{\mathbf{X}}_{120} - \bar{\mathbf{Y}}_{120})^T \mathcal{S}^{-1} (\bar{\mathbf{X}}_{120} - \bar{\mathbf{Y}}_{120})$$

Kde $p = 2$; $n = m = 120$; \mathcal{S} je odhad společné kovarianční matice

$$\mathcal{S} = \frac{1}{n+m} (n\mathcal{S}_X + m\mathcal{S}_Y).$$

\mathcal{S}_Y a \mathcal{S}_X jsou empirické kovarianční matice výběrů.

Celkově po dosazení získáváme hodnotu testové statistiky 0,64306. P-hodnota vychází vysoká 0,5266 a proto na 5% hladině nezamítáme. Nelze prokázat jestli



Obrázek 4: Sdružená hustota ropných akcií

jsou očekávané hodnoty výnosu dvou akcií aerolinek rozdílné oproti dvojici ropným akciím.

Na základě teorie probrané na cvičení v tabulce 1 jsou shrnuté výsledky 95% simultánních intervalů spolehlivosti. V prvním je to interval pro očekávaný rozdíl denního výnosu akcie PanAm a akcie DEC.

	spodní hranice	horní hranice	odhad
PanAm-DEC	-0.0461	0.0136	-0.0162
Delta-Texaco	-0.022	0.0215	-0.0002

Tabulka 1: Simultánní intervaly spolehlivosti