

UNIVERZITA PARDUBICE
CHEMICKO-TECHNOLOGICKÁ FAKULTA
KATEDRA ANALYTICKÉ CHEMIE

STATISTICKÁ ANALÝZA JEDNOROZMĚRNÝCH DAT

V OSTRAVĚ 20.3.2006

MAREK MOČKOR

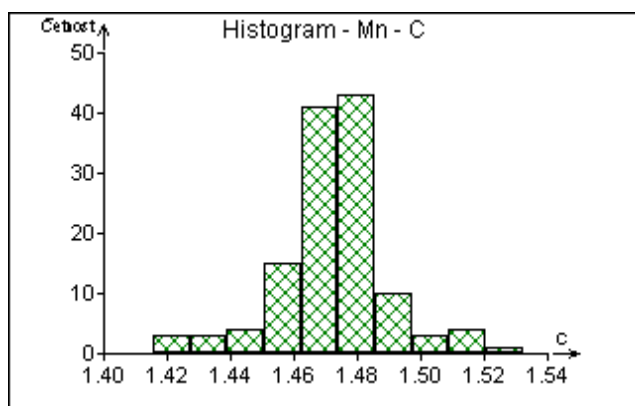
PŘÍKLAD Č.1 : ANALÝZA VELKÝCH VÝBĚRŮ

Zadání: Pro kontrolu stability měření Mn v oceli se v naší laboratoři používají kontrolní vzorky, připravené z části ocelové bramy o určité jakosti. Tato část bramy se překovává do tvaru válcové tyče a tato tyč se poté rozřeže na vzorky o stejných velikostech. Každý z těchto vzorků se následně 8 krát změří po celé ploše pomocí opticko emisního spektrometru s jiskrovým výbojem. Cílem tohoto experimentu je zjistit, zda jsou vzorky homogenní na celém svém povrchu a zda je možno je použít ke kontrole stability měření Mn. Dále je třeba udělat odhady střední hodnoty a stanovit interval spolehlivosti tohoto odhadu.

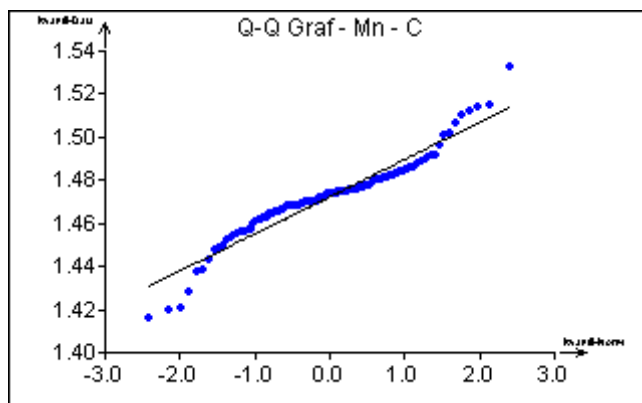
Naměřená data:

Mn	OZNAČENÍ	Mn	OZNAČENÍ	Mn	OZNAČENÍ	Mn	OZNAČENÍ
1,5099	CHK3/1	1,46609	CHK3/5	1,48335	CHK3/9	1,46793	CHK3/13
1,42791	CHK3/1	1,47368	CHK3/5	1,48539	CHK3/9	1,47265	CHK3/13
1,42009	CHK3/1	1,46411	CHK3/5	1,47466	CHK3/9	1,46834	CHK3/13
1,48126	CHK3/1	1,4752	CHK3/5	1,46984	CHK3/9	1,4697	CHK3/13
1,44798	CHK3/1	1,47529	CHK3/5	1,46581	CHK3/9	1,47585	CHK3/13
1,41566	CHK3/1	1,47556	CHK3/5	1,46689	CHK3/9	1,46271	CHK3/13
1,5324	CHK3/1	1,47392	CHK3/5	1,47489	CHK3/9	1,47166	CHK3/13
1,47513	CHK3/1	1,46851	CHK3/5	1,46526	CHK3/9	1,46129	CHK3/13
1,46838	CHK3/2	1,45221	CHK3/6	1,47412	CHK3/10	1,46548	CHK3/14
1,47684	CHK3/2	1,41976	CHK3/6	1,45644	CHK3/10	1,47911	CHK3/14
1,47192	CHK3/2	1,45716	CHK3/6	1,50192	CHK3/10	1,50099	CHK3/14
1,48504	CHK3/2	1,51357	CHK3/6	1,46226	CHK3/10	1,4433	CHK3/14
1,47173	CHK3/2	1,48982	CHK3/6	1,48668	CHK3/10	1,45562	CHK3/14
1,48577	CHK3/2	1,46427	CHK3/6	1,47995	CHK3/10	1,4381	CHK3/14
1,46163	CHK3/2	1,4752	CHK3/6	1,49182	CHK3/10	1,43736	CHK3/14
1,46799	CHK3/2	1,48144	CHK3/6	1,49014	CHK3/10	1,44821	CHK3/14
1,46519	CHK3/3	1,47396	CHK3/7	1,4538	CHK3/11	1,47491	CHK3/15
1,46227	CHK3/3	1,44957	CHK3/7	1,46019	CHK3/11	1,47747	CHK3/15
1,46829	CHK3/3	1,48252	CHK3/7	1,51154	CHK3/11	1,47027	CHK3/15
1,47763	CHK3/3	1,47046	CHK3/7	1,51441	CHK3/11	1,47475	CHK3/15
1,46932	CHK3/3	1,47038	CHK3/7	1,482	CHK3/11	1,47497	CHK3/15
1,48129	CHK3/3	1,46987	CHK3/7	1,48396	CHK3/11	1,47657	CHK3/15
1,47952	CHK3/3	1,49616	CHK3/7	1,48396	CHK3/11	1,46081	CHK3/15
1,46985	CHK3/3	1,48906	CHK3/7	1,46775	CHK3/11	1,47752	CHK3/15
1,49193	CHK3/4	1,47576	CHK3/8	1,48599	CHK3/12	1,45602	CHK3/16
1,47524	CHK3/4	1,48019	CHK3/8	1,47998	CHK3/12	1,50621	CHK3/16
1,45355	CHK3/4	1,47527	CHK3/8	1,47406	CHK3/12	1,46903	CHK3/16
1,47042	CHK3/4	1,46918	CHK3/8	1,47157	CHK3/12	1,48789	CHK3/16
1,47432	CHK3/4	1,47404	CHK3/8	1,47817	CHK3/12	1,48319	CHK3/16
1,45548	CHK3/4	1,46853	CHK3/8	1,46697	CHK3/12	1,47035	CHK3/16
1,47259	CHK3/4	1,45691	CHK3/8	1,47342	CHK3/12	1,47614	CHK3/16
1,4805	CHK3/4	1,46464	CHK3/8	1,47618	CHK3/12	1,48233	CHK3/16

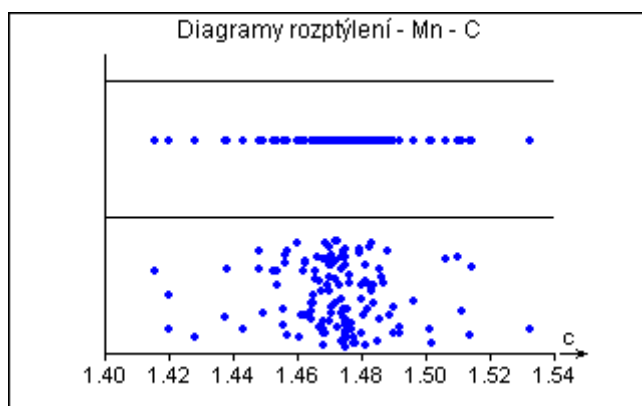
1.1 Exploratorní analýza dat (QC Expert)



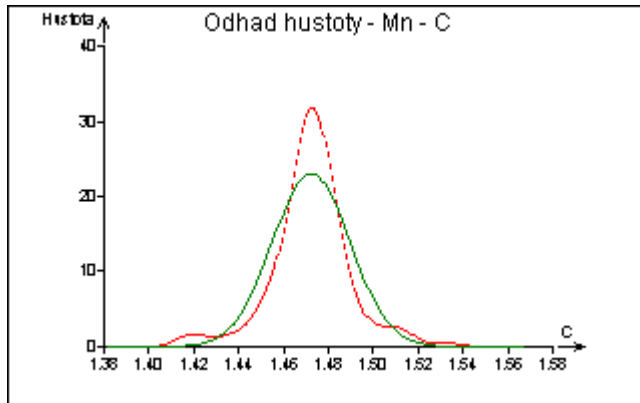
Obr. 1a Histogram



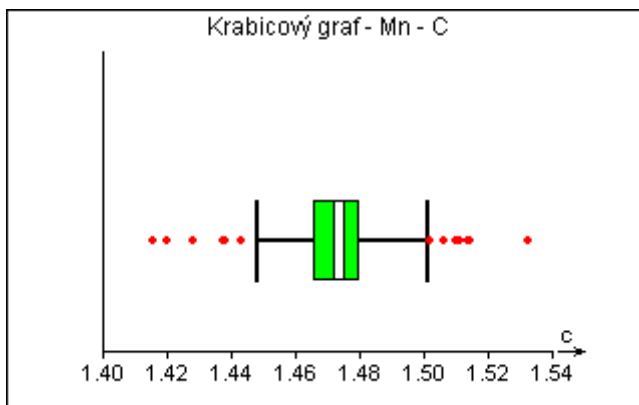
Obr. 1b Kvantil-Kvantilový (Q-Q) graf



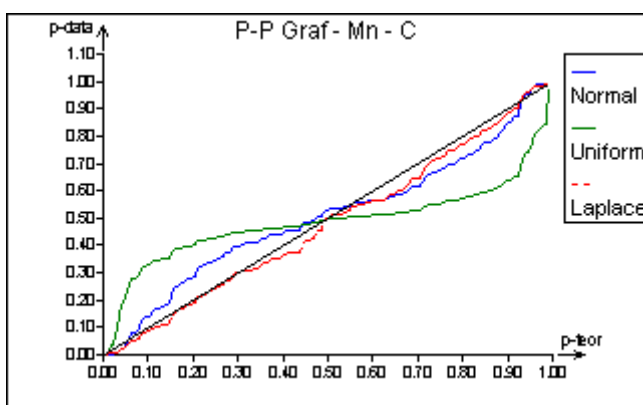
Obr. 1c Diagram rozptýlení a rozmítnutý diagram rozptýlení



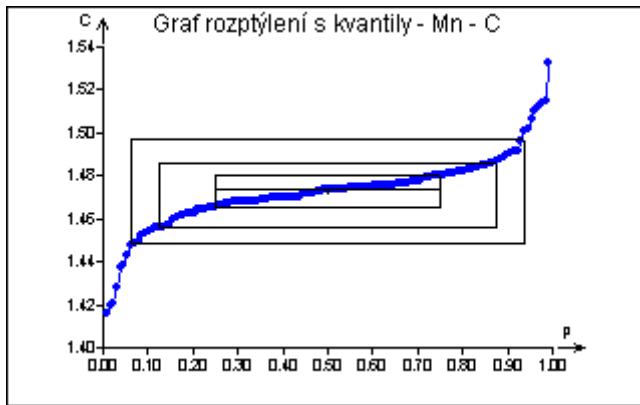
Obr. 1d Graf hustoty pravděpodobnosti



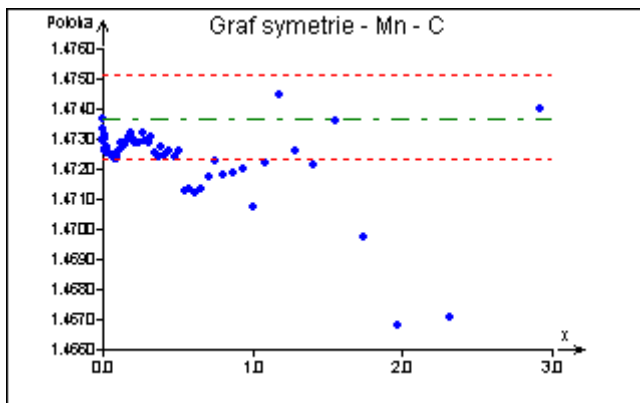
Obr. 1e Krabicový graf



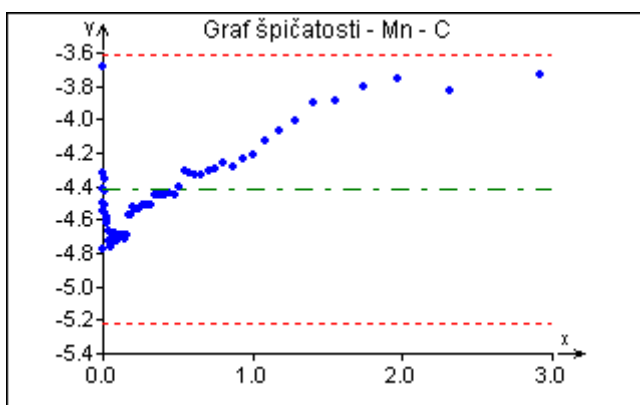
Obr. 1f P-P graf



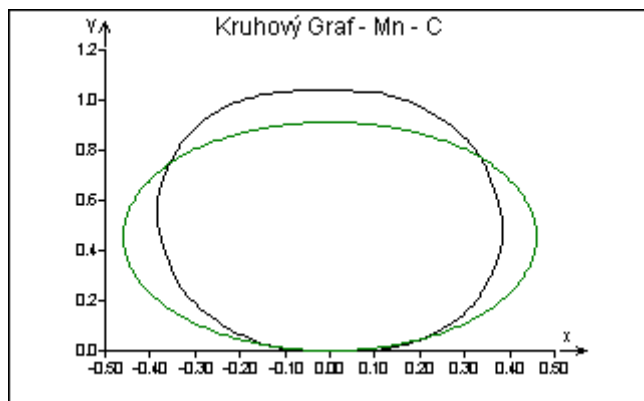
Obr. 1g Graf rozptýlení s kvantily



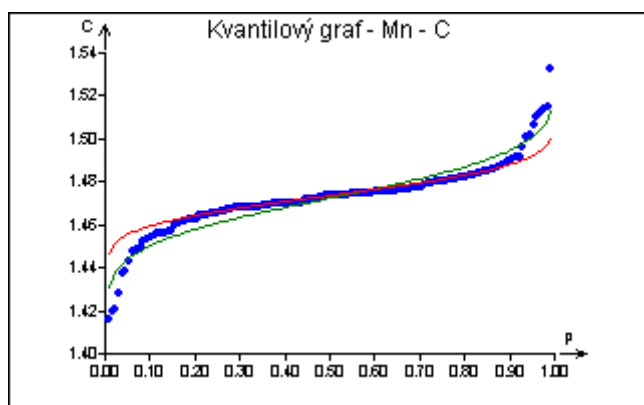
Obr. 1h Graf symetrie



Obr. 1i Graf špičatosti



Obr. 1j Kruhový graf



Obr. 1k Kvantilový graf

1. Histogram (obr. 1a) ukazuje na symetrické silně špičaté rozdělení
2. Kvantil- Kvantilový graf (obr. 1b) většina bodů leží na přímce. Což ukazuje na symetrické rozdělení, je patrných 5 odlehlých bodů dole a 7 odlehlých bodů nahoře.
3. Diagram rozptýlení (Obr. 1c) ukazuje 7 odlehlých hodnot nahoře a 6 dole
4. Graf hustoty pravděpodobnosti (Obr. 1d) svědčí o symetrickém, špičatém rozdělení.
5. Krabicový graf (obr. 1e) v horní části detekováno 7 odlehlých bodů, v dolní části 7 odlehlých bodů, krabicový graf je uprostřed rozdělen mediánem.
6. P-P graf (obr.1f) Empirická křivka se nejvíce blíží Laplaceovu rozdělení
7. Graf rozptýlení s kvantily (obr. 1g) symetrie kvantilových obdelníků ukazuje na symetrické rozdělení, body mimo sedecilový obdelník jsou indikovány jako odlehlé. Nahoře 7 odlehlých bodů, dole 7 odlehlých bodů
8. Graf symetrie (obr. 1h) indikuje, že valná část bodů nevybočuje ze symetrického rozdělení, není patrný jednoznačný trend.
9. Graf špičatosti (obr. 1i) ukazuje výrazný trend- data jsou špičatá
10. Kruhový graf (obr.1j) ukazuje na rovnoměrné rozdělení.
11. Kvantilový graf (obr. 1k) data lépe kopírují robustní křivku (červená přerušovaná) je však patrné, dle kolmice vztyčené z osy x v bodě 0,5, že medián se prakticky

shoduje s aritmetickým průměrem, což indikuje symetrické rozdělení. V horní části patrných 8 odlehlých bodů v dolní části 7 odlehlých bodů.

Závěr exploratorní analýzy: Jedná se o symetrická data s významnou špičatostí. Bylo indikováno 7 odlehlých hodnot nahoře a 7 odlehlých hodnot dole, což může znamenat nehomogenitu v některých vzorcích.

1.2 Ověření předpokladů o datech (QC Expert)

Řád trendu :	4	Test normality :	
Testovaná hodnota :	0	Název sloupce :	Mn
Vyhlazení hustoty :	0,5	Průměr :	1,472395354
Hladina významnosti :	0,05	Rozptyl :	0,000299378
Název sloupce :	Mn	Šikmost :	-0,20549611
Počet platných dat :	127	Špičatost :	5,55238498
		Normalita :	Přijata
Klasické parametry :		Vypočtený :	1,170519497
Název sloupce :	Mn	Teoretický :	5,991464547
Průměr :	1,472395354	Pravděpodobnost :	0,556961173
Spodní mez :	1,469356939	Vybočující body :	
Horní mez :	1,47543377	Název sloupce :	Mn
Rozptyl :	0,000299378	Homogenita :	Zamítnuta
Směr. odchylka :	0,017302534	Počet vybočujících bodů :	7
Šikmost :	-0,20549611	Spodní mez :	1,43255726
Odchylka od 0 :	Nevýznamná	Horní mez :	1,51268274
Špičatost :	5,55238498	Autokorelace :	
Odchylka od 3 :	Významná	Řád autokorelace :	4
Polosuma :	1,47403	Název sloupce :	Mn
Modus :	1,476209146	Počet :	-0,035383058
Robustní parametry :			
Název sloupce :	Mn	Řád autokorelace 1 :	
Medián :	1,47368	Korelační koeficient :	0,085633392
IS spodní :	1,471226435	Pravděpodobnost :	0,170192376
IS horní :	1,476133565	Závěr :	Nevýznamný
Medianová směr. odchylka :	0,001239819	Řád autokorelace 2 :	
Medianový rozptyl :	1,54E-06	Korelační koeficient :	-0,087463968
10% Průměr :	1,472584779	Pravděpodobnost :	0,166044898
10% IS spodní :	1,469994612	Závěr :	Nevýznamný
10% IS horní :	1,475174946	Řád autokorelace 3 :	
10% Směr. odchylka :	0,009658022	Korelační koeficient :	0,127009888
10% Rozptyl :	9,33E-05	Pravděpodobnost :	0,079906892
20% Průměr :	1,472632673	Závěr :	Nevýznamný
20% IS spodní :	1,470338238	Řád autokorelace 4 :	
20% IS horní :	1,474927108	Korelační koeficient :	-0,035383058
20% Směr. odchylka :	0,007001515	Pravděpodobnost :	0,348811171
20% Rozptyl :	4,90E-05	Závěr :	Nevýznamný
40% Průměr :	1,4728264	Test významnosti trendu :	
40% IS spodní :	1,470784738	Název sloupce :	Mn
40% IS horní :	1,474868062	Směrnice :	3,14E-05
40% Směr. odchylka :	0,003635554	Významnost :	Nevýznamný
40% Rozptyl :	1,32E-05	Pravděpodobnost :	0,772479329

Závěr ověření předpokladů o datech:

Každý reprezentativní náhodný výběr by měl být charakterizován třemi základními předpoklady jsou to : nezávislost, homogenita a normalita.

V našem případě byla nezávislost prokázána podle autokorelace 1.,2.,3.a 4. řádu jsou prvky nezávislé.

Hypotéza o normalitě dat byla přijata: Tabulkový kvantil $\chi^2(1-\alpha,2)$: 5,991464547
Vypočtený χ^2 : 1,170519497

Homogenita byla zamítnuta, test homogenity našel celkem 7 odlehlých hodnot. Pomocí grafických testů bylo však nalezeno 14 odlehlých hodnot, z toho 7 nahoře a 7 dole. Grafické testy jsou obvykle citlivější než testy početní.

Vzhledem k tomu, že měřené vzorky mají být použity jako vzorky pro kontrolu nastavení stanovení Mn v oceli, není žádoucí, aby tyto vzorky obsahovali jakékoliv nehomogenity, které by v případě provedení kontroly nastavení mohly vést ke špatnému rozhodnutí laboranta o dalším postupu případně nápravných opatřeních. Naše vzorky vykazují celkem 14 odlehlých hodnot, které mohou být způsobeny nehomogenitami ve vzorcích. Proto byly tyto hodnoty pomocí QC Expertu označeny v krabicovém grafu a následně ve výběru dat (tab.3) a vzorky, ve kterých se odlehlé hodnoty vyskytly byly vyloučeny a navraceny do konvertoru jako šrot. Vyloučeny byly vzorky: CHK3/1, CHK3/6, CHK3/10, CHK3/11, CHK3/14, CHK3/16. Zbylá data (tab.4) byla znovu analyzována.

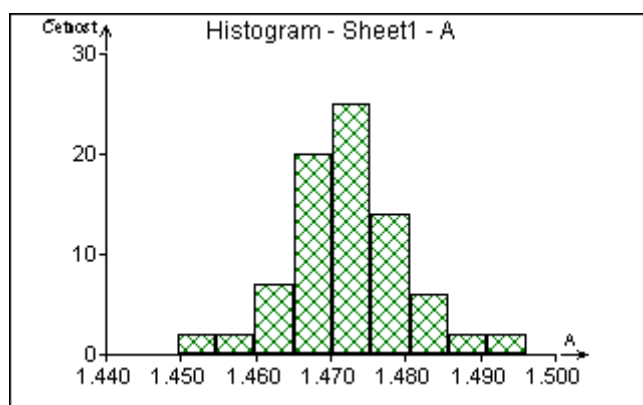
Mn	OZNAČENÍ	Mn	OZNAČENÍ	Mn	OZNAČENÍ	Mn	OZNAČENÍ
1,5099	CHK3/1	1,46609	CHK3/5	1,48335	CHK3/9	1,46793	CHK3/13
1,42791	CHK3/1	1,47368	CHK3/5	1,48539	CHK3/9	1,47265	CHK3/13
1,42009	CHK3/1	1,46411	CHK3/5	1,47466	CHK3/9	1,46834	CHK3/13
1,48126	CHK3/1	1,4752	CHK3/5	1,46984	CHK3/9	1,4697	CHK3/13
1,44798	CHK3/1	1,47529	CHK3/5	1,46581	CHK3/9	1,47585	CHK3/13
1,41566	CHK3/1	1,47556	CHK3/5	1,46689	CHK3/9	1,46271	CHK3/13
1,5324	CHK3/1	1,47392	CHK3/5	1,47489	CHK3/9	1,47166	CHK3/13
1,47513	CHK3/1	1,46851	CHK3/5	1,46526	CHK3/9	1,46129	CHK3/13
1,46838	CHK3/2	1,45221	CHK3/6	1,47412	CHK3/10	1,46548	CHK3/14
1,47684	CHK3/2	1,41976	CHK3/6	1,45644	CHK3/10	1,47911	CHK3/14
1,47192	CHK3/2	1,45716	CHK3/6	1,50192	CHK3/10	1,50099	CHK3/14
1,48504	CHK3/2	1,51357	CHK3/6	1,46226	CHK3/10	1,4433	CHK3/14
1,47173	CHK3/2	1,48982	CHK3/6	1,48668	CHK3/10	1,45562	CHK3/14
1,48577	CHK3/2	1,46427	CHK3/6	1,47995	CHK3/10	1,4381	CHK3/14
1,46163	CHK3/2	1,4752	CHK3/6	1,49182	CHK3/10	1,43736	CHK3/14
1,46799	CHK3/2	1,48144	CHK3/6	1,49014	CHK3/10	1,44821	CHK3/14
1,46519	CHK3/3	1,47396	CHK3/7	1,4538	CHK3/11	1,47491	CHK3/15
1,46227	CHK3/3	1,44957	CHK3/7	1,46019	CHK3/11	1,47747	CHK3/15
1,46829	CHK3/3	1,48252	CHK3/7	1,51154	CHK3/11	1,47027	CHK3/15
1,47763	CHK3/3	1,47046	CHK3/7	1,51441	CHK3/11	1,47475	CHK3/15
1,46932	CHK3/3	1,47038	CHK3/7	1,482	CHK3/11	1,47497	CHK3/15
1,48129	CHK3/3	1,46987	CHK3/7	1,48396	CHK3/11	1,47657	CHK3/15
1,47952	CHK3/3	1,49616	CHK3/7	1,48396	CHK3/11	1,46081	CHK3/15
1,46985	CHK3/3	1,48906	CHK3/7	1,46775	CHK3/11	1,47752	CHK3/15
1,49193	CHK3/4	1,47576	CHK3/8	1,48599	CHK3/12	1,45602	CHK3/16
1,47524	CHK3/4	1,48019	CHK3/8	1,47998	CHK3/12	1,50621	CHK3/16
1,45355	CHK3/4	1,47527	CHK3/8	1,47406	CHK3/12	1,46903	CHK3/16
1,47042	CHK3/4	1,46918	CHK3/8	1,47157	CHK3/12	1,48789	CHK3/16
1,47432	CHK3/4	1,47404	CHK3/8	1,47817	CHK3/12	1,48319	CHK3/16
1,45548	CHK3/4	1,46853	CHK3/8	1,46697	CHK3/12	1,47035	CHK3/16
1,47259	CHK3/4	1,45691	CHK3/8	1,47342	CHK3/12	1,47614	CHK3/16
1,4805	CHK3/4	1,46464	CHK3/8	1,47618	CHK3/12	1,48233	CHK3/16

Tab.3: data s označenými odlehlými hod

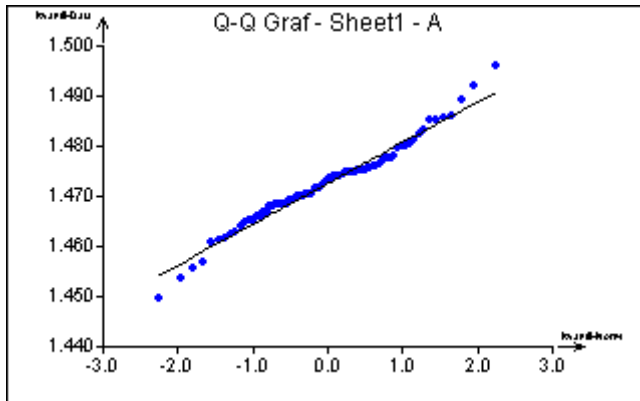
2.1 Exploratorní analýza dat bez vyloučených vzorků (QC Expert)

Mn	OZNAČENÍ	Mn	OZNAČENÍ	Mn	OZNAČENÍ	Mn	OZNAČENÍ
1,46838	CHK3/2	1,46609	CHK3/5	1,48335	CHK3/9	1,47491	CHK3/15
1,47684	CHK3/2	1,47368	CHK3/5	1,48539	CHK3/9	1,47747	CHK3/15
1,47192	CHK3/2	1,46411	CHK3/5	1,47466	CHK3/9	1,47027	CHK3/15
1,48504	CHK3/2	1,4752	CHK3/5	1,46984	CHK3/9	1,47475	CHK3/15
1,47173	CHK3/2	1,47529	CHK3/5	1,46581	CHK3/9	1,47497	CHK3/15
1,48577	CHK3/2	1,47556	CHK3/5	1,46689	CHK3/9	1,47657	CHK3/15
1,46163	CHK3/2	1,47392	CHK3/5	1,47489	CHK3/9	1,46081	CHK3/15
1,46799	CHK3/2	1,46851	CHK3/5	1,46526	CHK3/9	1,47752	CHK3/15
1,46519	CHK3/3	1,47396	CHK3/7	1,48599	CHK3/12		
1,46227	CHK3/3	1,44957	CHK3/7	1,47998	CHK3/12		
1,46829	CHK3/3	1,48252	CHK3/7	1,47406	CHK3/12		
1,47763	CHK3/3	1,47046	CHK3/7	1,47157	CHK3/12		
1,46932	CHK3/3	1,47038	CHK3/7	1,47817	CHK3/12		
1,48129	CHK3/3	1,46987	CHK3/7	1,46697	CHK3/12		
1,47952	CHK3/3	1,49616	CHK3/7	1,47342	CHK3/12		
1,46985	CHK3/3	1,48906	CHK3/7	1,47618	CHK3/12		
1,49193	CHK3/4	1,47576	CHK3/8	1,46793	CHK3/13		
1,47524	CHK3/4	1,48019	CHK3/8	1,47265	CHK3/13		
1,45355	CHK3/4	1,47527	CHK3/8	1,46834	CHK3/13		
1,47042	CHK3/4	1,46918	CHK3/8	1,4697	CHK3/13		
1,47432	CHK3/4	1,47404	CHK3/8	1,47585	CHK3/13		
1,45548	CHK3/4	1,46853	CHK3/8	1,46271	CHK3/13		
1,47259	CHK3/4	1,45691	CHK3/8	1,47166	CHK3/13		
1,4805	CHK3/4	1,46464	CHK3/8	1,46129	CHK3/13		

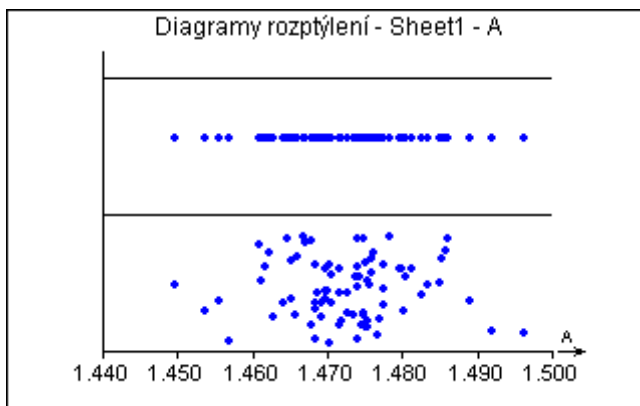
Tab. 4 Tabulka dat po vyloučení „nehomogenních“ vzorků



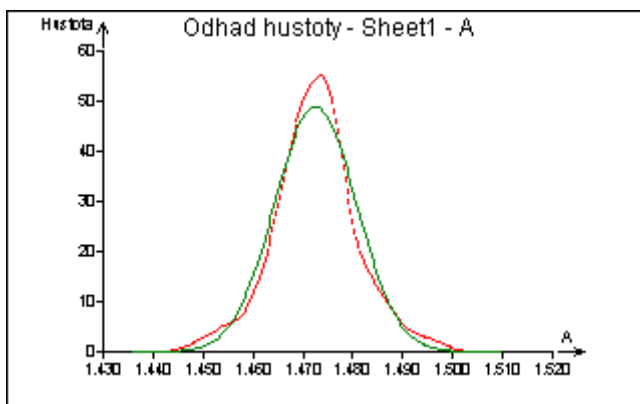
Obr. 2a Histogram



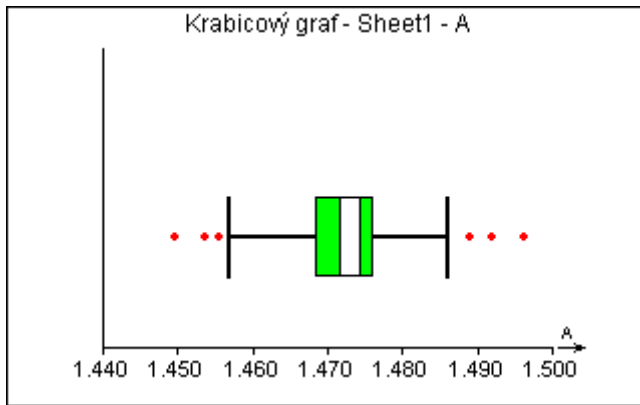
Obr. 2b Kvantil-Kvantilový (Q-Q) graf



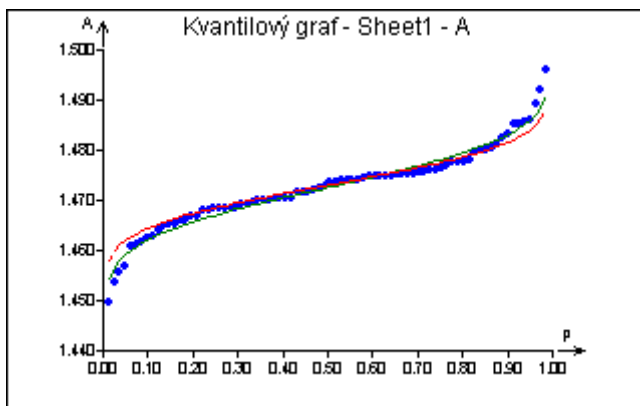
Obr. 2c Diagram rozptýlení a rozmítnutý diagram rozptýlení



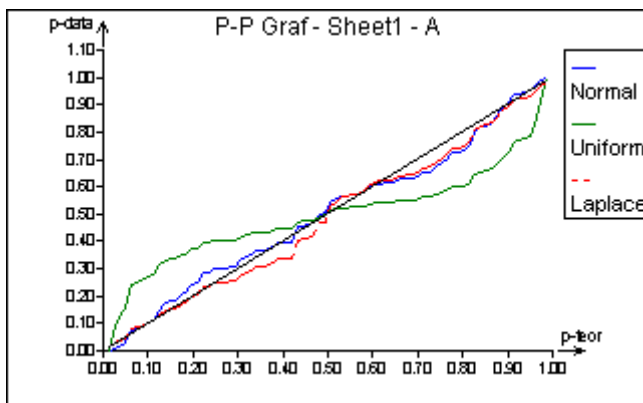
Obr. 2d Graf hustoty pravděpodobnosti



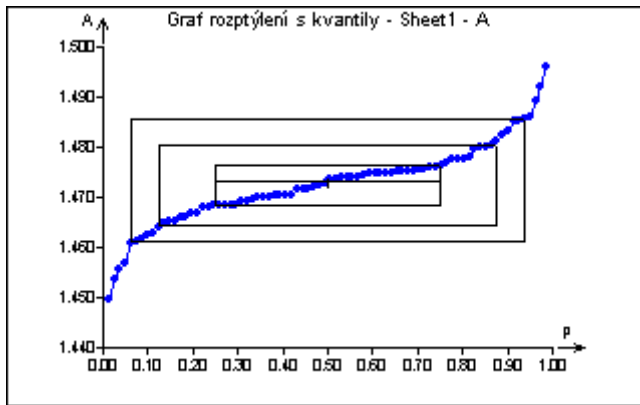
Obr. 2e Krabicový graf



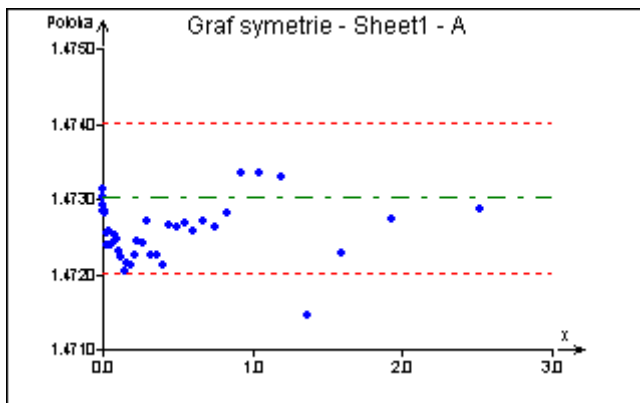
Obr. 2f Kvantilový graf



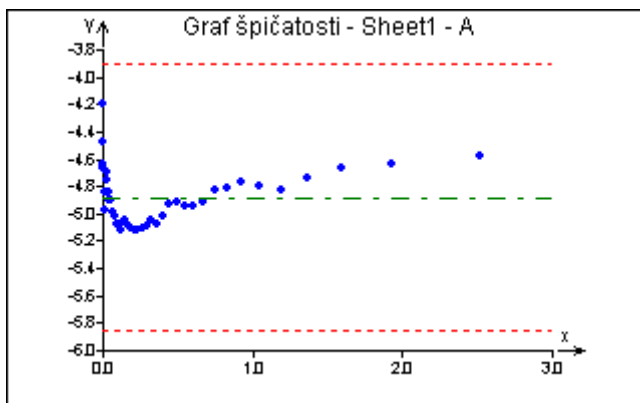
Obr. 2g P-P graf



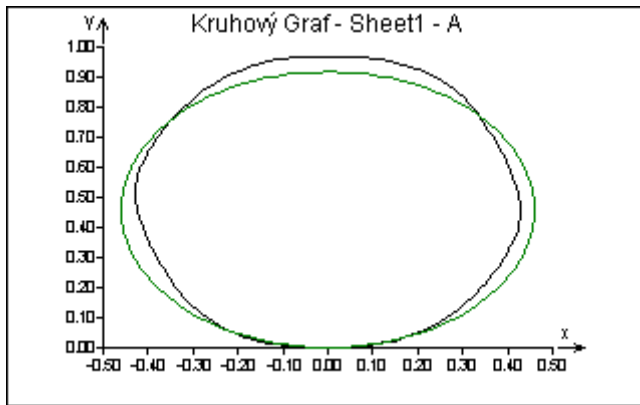
Obr. 2h Graf rozptýlení s kvantily



Obr. 2i Graf symetrie



Obr. 2j Graf špičatosti



Obr. 2k Kruhový graf

1. Histogram (obr. 2a) ukazuje na symetrické, mírně zešikmené rozdělení
2. Kvantil-kvantilový graf (obr. 2b) detekuje normální rozdělení se třemi odlehlými body nahoře a čtyřmi dole
3. Diagram rozptýlení (obr. 2c) ukazuje tři odlehlé body nahoře a čtyři odlehlé body dole
4. Graf hustoty pravděpodobnosti (obr. 2d) naznačuje, že rozdělení se velmi blíží normálnímu Gaussovskému rozdělení
5. Krabicový graf (obr. 2e) detekuje symetrické rozdělení se třemi odlehlými body nahoře a třemi odlehlými body dole
6. Kvantilový graf (obr. 2f) robustní křivka se téměř kryje s křivkou empirickou. Kolmice vztyčená z bodu 0,5 na ose x protíná robustní i empirickou křivku v tomtéž místě, průměr je totožný s mediánem. Nahoře detekovány tři odlehlé body, dole čtyři odlehlé body.
7. P-P graf (obr. 2g) empirická křivka nejlépe kopíruje křivku normálního rozdělení
8. Graf rozptýlení s kvantily (obr. 2h) ukazuje symetrické rozdělení se třemi odlehlými body nahoře a čtyřmi odlehlými body dole
9. Graf symetrie (obr. 2i) indikuje, že většina bodů nevybočuje ze symetrického rozdělení, není patrný jednoznačný trend
10. Graf špičatosti (obr. 2j) indikuje velmi mírný trend, data jsou mírně špičatá
11. Kruhový graf (obr. 2k) ukazuje, že jde o normální rozdělení

Závěr exploratorní analýzy: Jedná se o data s normálním rozdělením, byly detekovány tři odlehlé body nahoře a čtyři odlehlé body dole. Tyto odlehlé body však nebudeme vylučovat, protože odchylky již nejsou z technologického hlediska tak výrazné, jako v předchozím případě.

2.2 Ověření předpokladů o datech (QC Expert)

Řád trendu :	4	Test normality :	
Testovaná hodnota :	0	Název sloupce :	Mn
Vyhlazení hustoty :	0,5	Průměr :	1,47256725
Hladina významnosti :	0,05	Rozptyl :	6,65E-05
Název sloupce :	Mn	Šikmost :	0,031543772
Počet platných dat :	80	Špičatost :	3,862729898
		Normalita :	Přijata
Klasické parametry :		Vypočtený :	0,059440816
Název sloupce :	Mn	Teoretický :	5,991464547
Průměr :	1,47256725	Pravděpodobnost :	0,9707169
Spodní mez :	1,470751902		
Horní mez :	1,474382598	Vybočující body :	
Rozptyl :	6,65E-05	Název sloupce :	Mn
Směr. odchylka :	0,008157433	Homogenita :	Zamítnuta
Šikmost :	0,031543772	Počet vybočujících bodů :	2
Odchylka od 0 :	Nevýznamná	Spodní mez :	1,45089255
Špičatost :	3,862729898	Horní mez :	1,49357745
Odchylka od 3 :	Nevýznamná		
Polosuma :	1,472865	Autokorelace :	
Modus :	1,473947401	Řád autokorelace :	4
Robustní parametry :		Název sloupce :	Mn
Název sloupce :	Mn	Počet :	-0,108609709
Medián :	1,473035		
IS spodní :	1,470689069	Řád autokorelace 1 :	
IS horní :	1,475380931	Korelační koeficient :	0,019014284
Medianová směr. Odch. :	0,001178593	Pravděpodobnost :	0,433951033
Medianový rozptyl :	1,39E-06	Závěr :	Nevýznamný
10% Průměr :	1,472593472	Řád autokorelace 2 :	
10% IS spodní :	1,470898251	Korelační koeficient :	0,004215976
10% IS horní :	1,474288694	Pravděpodobnost :	0,48538862
10% Směr. odchylka :	0,005564222	Závěr :	Nevýznamný
10% Rozptyl :	3,10E-05	Řád autokorelace 3 :	
20% Průměr :	1,472518437	Korelační koeficient :	0,026628841
20% IS spodní :	1,470819691	Pravděpodobnost :	0,409090204
20% IS horní :	1,474217184	Závěr :	Nevýznamný
20% Směr. odchylka :	0,004124361	Řád autokorelace 4 :	
20% Rozptyl :	1,70E-05	Korelační koeficient :	-0,108609709
40% Průměr :	1,472514167	Pravděpodobnost :	0,175174684
40% IS spodní :	1,47097538	Závěr :	Nevýznamný
40% IS horní :	1,474052954		
40% Směr. odchylka :	0,002329447	Test významnosti trendu :	
40% Rozptyl :	5,43E-06	Název sloupce :	Mn
		Směrnice :	-4,65E-06
Znaménkový test :		Významnost :	Nevýznamný
Závěr :	Data jsou nezávislá	Pravděpodobnost :	0,546384453

Závěr ověření předpokladů o datech:

Každý reprezentativní náhodný výběr by měl být charakterizován třemi základními předpoklady jsou to : nezávislost, homogenita a normalita.

V našem případě byla nezávislost prokázána podle autokorelace 1.,2.,3.a 4. řádu jsou prvky nezávislé.

Hypotéza o normalitě dat byla přijata: Tabulkový kvantil $\chi^2(1-\alpha,2)$: 5,991464547
Vypočtený χ^2 :0,059440816

Homogenita byla zamítnuta, test našel 2 odlehlé hodnoty. Pomocí grafických testů bylo však nalezeno 7 odlehlých hodnot, z toho 3 nahoře a 4 dole. Grafické testy jsou obvykle citlivější než testy početní.

Vzhledem k tomu, že se jedná o normální data s téměř Gaussovským rozdělením, není třeba data transformovat a k odhadu střední hodnoty je možno použít klasický aritmetický průměr.

Vzorky budou použity pro kontrolu nastavení Mn na optickém emisním spektrometru. 95% naměřených výsledků by mělo ležet v intervalu **(1,470751902; 1,474382598)** . Jako odhad střední hodnoty byl použit aritmetický průměr a střední hodnota byla odhadnuta na: **1,47256725**.

PŘÍKLAD Č. 3: ANALÝZA MALÝCH VÝBĚRŮ + TEST SPRÁVNOSTI

Zadání: V naší laboratoři provádíme stanovení obsahu síry v motorových palivech pomocí rentgenové energodisperzní spektrometrie. Tuto metodu máme řádně zvalidovanou a akreditovanou ČIA. Pro ověření stability křivky jsme používaly certifikované referenční materiály s obsahem síry 5 ppm, 25 ppm, 50 ppm, 100 ppm.... Vzhledem k tomu, že životnost těchto CRM je maximálně jeden rok, byli jsme nuceni zakoupit CRM nové u jiného výrobce. Cílem tohoto testu je zjistit, jestli obsah síry v novém CRM o obsahu 25 ppm síry odpovídá certifikovanému obsahu.

Naměřená data:

23,9
24,2
24,3
25,1
24,9
23,8
25,3
24,7
24,9
23,4

Vzhledem k malému počtu dat je vhodné použít Hornův postup testování malých výběrů, který je založen na pořádkových statistikách.

3.1 Hornův postup testování malých výběrů:

1. Pořádkové statistiky

i	1	2	3	4	5	6	7	8	9	10
x_i	23,4	23,8	23,9	24,2	24,3	24,7	24,9	24,9	25,1	25,3

2. Hloubka pivotu

$$H = \text{int}((n+1)/2+1)/2 = 3$$

3. Dolní a horní pivot

$$X_d = X_H = X_3 = 23.9$$

$$X_h = X_{(n+1-H)} = X_8 = 24.9$$

4. Pivotová polosuma (míra polohy)

$$PI = (X_d + X_h)/2 = 24,4$$

5. Pivotové rozpětí (míra rozptýlení)

$$RI = X_h - X_d = 1$$

6. Kvantil $t_{L,0,975}(10) = 0,668$ byl určen odečtem z tabulky

7. Pro 95% interval spolehlivosti platí:

$$PI - RI * t_{L,0,975}(10) \leq \mu \leq PI + RI * t_{L,0,975}(10)$$

$$23.732 \leq \mu \leq 25.068$$

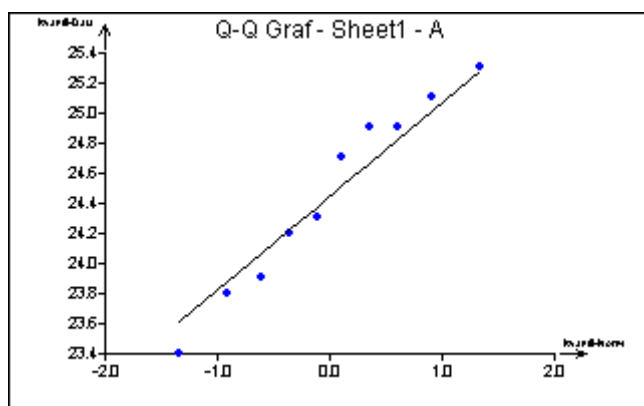
Závěr analýzy provedené Hornovým postupem

S 95% statistickou jistotou můžeme tvrdit, že odhad střední hodnoty μ naměřené koncentrace S v CRM se nalézají v interval **$23.732 \leq \mu \leq 25.068$** .

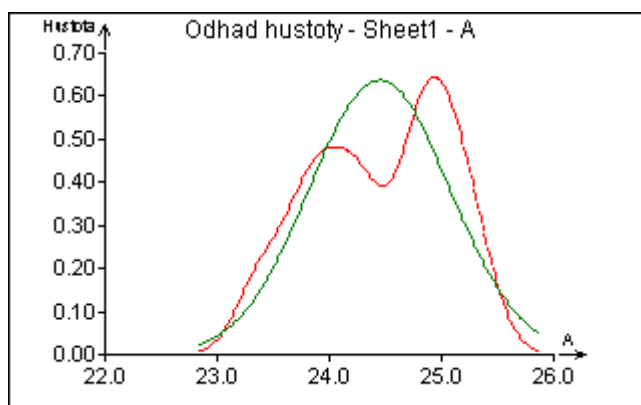
3.2 Test správnosti

Vzhledem k tomu, že stanovený interval spolehlivosti obsahuje hodnotu **25 ppm**, která je uvedena v certifikátu, můžeme učinit závěr, že obsah S v CRM odpovídá certifikované hodnotě a CRM může být použito pro kontrolu nastavení metody.

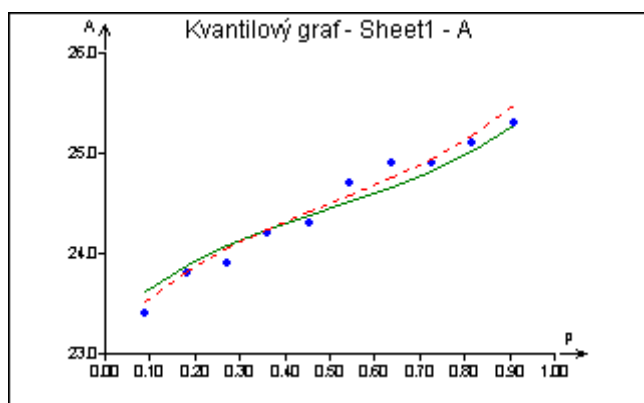
3.2 Průzkumová analýza (EDA)



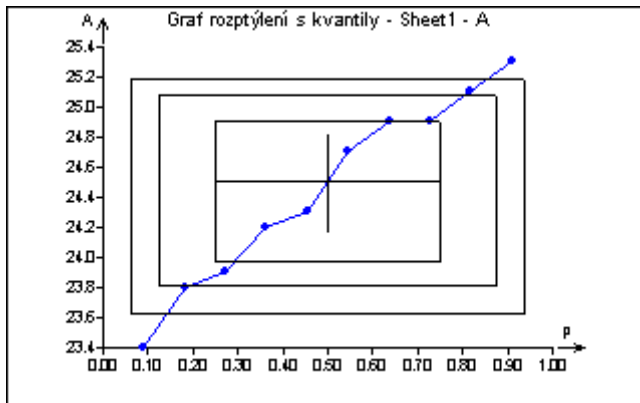
obr. 3a Kvantil-Kvantilový (Q-Q) graf



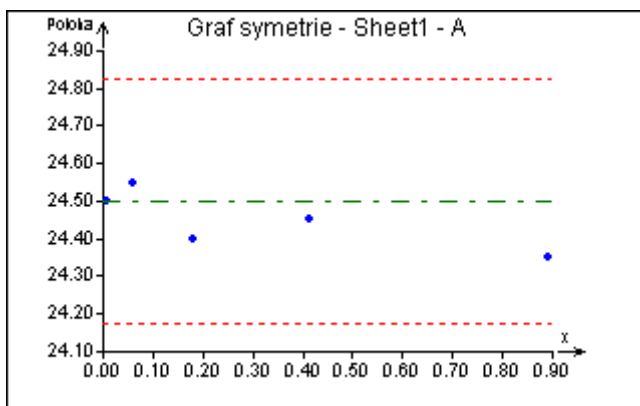
obr. 3b Graf hustoty pravděpodobnosti



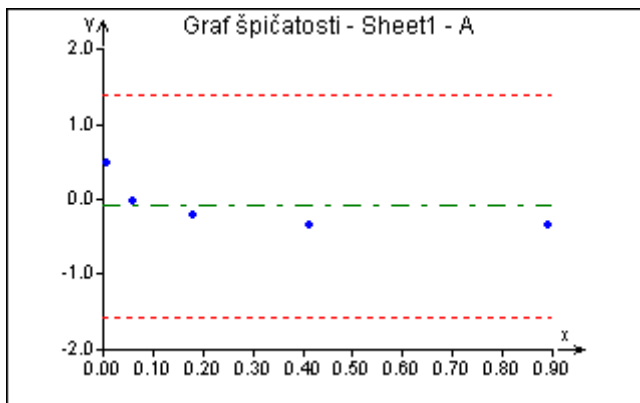
obr. 3c Kvantilový graf



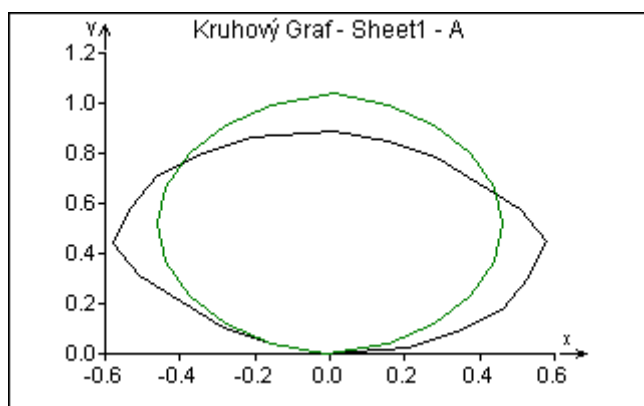
obr. 3d Graf rozptýlení s kvantily



obr. 3e Graf symetrie



obr. 3f Graf špičatosti



Obr.3g Kruhový graf

1. Kvantil – kvantilový graf (obr. 3a) detekuje přibližně normální rozdělení (malý počet dat) s jedním odlehlým bodem dole
2. Graf hustoty pravděpodobnosti ukazuje možnou bimodálnost dat, což je však nejspíš způsobeno malým množstvím dat. Rozdělení se blíží rozdělení Gaussovskému.
3. Kvantilový graf (obr. 3c) data lépe kopírují robustní křivku. Dole jeden odlehlý bod.
4. Graf rozptýlení s kvantily (obr. 3d) ukazuje na symetrické rozdělení s jedním odlehlým bodem nahoře a jedním odlehlým bodem dole.
5. Z grafu symetrie (obr.3e) vyplývá, že se jedná o symetrická data.
6. Graf špičatosti (obr. 3f) detekuje nevýznamnou špičatost dat.
7. Kruhový graf (obr. 3g) osa elipsy je rovnoběžná s osou x , což ukazuje na symetrické rozdělení.

Závěr EDA:

Jedná se o symetrické rozdělení s jedním odlehlým bodem nahoře a s podezřením na bimodálnost dat. Vzhledem k malému počtu dat nebudeme odlehlý bod vylučovat.

3.3 OVĚŘENÍ PŘEDPOKLADŮ O DATECH

Řád trendu :	4		
Testovaná hodnota :	0	Analýza malých výběrů	
Vyhlazení hustoty :	0,5	N :	10
Hladina významnosti :	0,05	Střední hodnota :	24,4
Název sloupce :	S	Spodní mez (5%) :	23,877
Počet platných dat :	10	Horní mez (95%) :	24,923
		Spodní mez (2.5%) :	23,732
Klasické parametry :		Horní mez (97.5%) :	25,068
Název sloupce :	S	Pivotové rozpětí :	1
Průměr :	24,45		
Spodní mez :	24,0023062	Test normality :	
Horní mez :	24,8976938	Název sloupce :	S
Rozptyl :	0,391666667	Průměr :	24,45
Směr. odchylka :	0,625832779	Rozptyl :	0,391666667
Šikmost :	-0,253720664	Šikmost :	-0,253720664
Odchylka od 0 :	Nevýznamná	Špičatost :	1,831990343
Špičatost :	1,831990343	Normalita :	Přijata
Odchylka od 3 :	Nevýznamná	Vypočtený :	0,299056204
Polosuma :	24,35	Teoretický :	5,991464547
Modus :	24,58181818	Pravděpodobnost :	0,861114239
t-test		Vybočující body :	
Testovaná hodnota :	0		S
Rozdíl :	Významný	Homogenita :	Přijata
Vypočtený :	123,5436868	Počet vybočujících bodů :	0
Teoretický :	2,262157163	Spodní mez :	21,721
Pravděpodobnost :	3,79E-16	Horní mez :	26,979
Konfidenční interval levý:	24,0872165		
Konfidenční interval pravý:	24,8127835	Autokorelace :	
		Řád autokorelace :	4
Robustní parametry :		Název sloupce :	S
Název sloupce :	S	Počet :	0,236879433
Medián :	24,5		
IS spodní :	23,74978104	Řád autokorelace 1	
IS horní :	25,25021896	Korelační koeficient :	-0,176595745
Medianová směr. odchylka :	0,331638745	Pravděpodobnost :	0,324730454
Medianový rozptyl :	0,109984257	Závěr :	Nevýznamný
10% Průměr :	24,475	Řád autokorelace 2	
10% IS spodní :	23,97130173	Korelační koeficient :	-0,065248227
10% IS horní :	24,97869827	Pravděpodobnost :	0,43900318
10% Směr. odchylka :	0,433999488	Závěr :	Nevýznamný
10% Rozptyl :	0,188355556	Řád autokorelace 3	
20% Průměr :	24,475	Korelační koeficient :	-0,253191489
20% IS spodní :	23,97130173	Pravděpodobnost :	0,291906315
20% IS horní :	24,97869827	Závěr :	Nevýznamný
20% Směr. odchylka :	0,433999488	Řád autokorelace 4	
20% Rozptyl :	0,188355556	Korelační koeficient :	0,236879433
40% Průměr :	24,48333333	Pravděpodobnost :	0,325663362
40% IS spodní :	23,84240389	Závěr :	Nevýznamný
40% IS horní :	25,12426278		

Závěr ověření předpokladů o datech:

Jedná se o normální, nezávislá data. Interval spolehlivosti aritmetického průměru neobsahuje hodnotu 25 ppm. Tato skutečnost je však způsobena malým počtem dat a proto budeme za směrodatné pokládat hodnoty stanovené pomocí Hornova postupu. Střední hodnoty stanovené klasickými i robustními metodami se velmi blíží střední hodnotě stanovené dle Hornova postupu.

Příklad č.4 TEST SHODNOSTI

V naší laboratoři provádím stanovení Cr₂O₃ v ocelářenských struskách pomocí rentgenové vlnově disperzní spektrometrie. Cílem tohoto úkolu je určit, zda dva laboranti, z nichž každý provedl 8 měření totožného vzorku, včetně přípravy tohoto vzorku pracují stejně.

Naměřená data:

Lab.1	Lab.2
11,65	11,47
11,74	11,52
11,84	11,39
11,79	11,44
11,72	11,44
11,84	11,4
11,83	11,36
11,74	11,41

1. Ověření normality dat (QC Expert)

Lab.1:

Test normality :	
Název sloupce	Lab.1
Průměr :	11,76875
Rozptyl :	0,004641071
Šikmost	-0,422723268
Špičatost :	2,059351415
Normalita :	Přijata
Vypočtený :	0,639913014
Teoretický :	5,991464547
Pravděpodobnost :	0,72618062

Lab.2:

Test normality :	
Název sloupce :	lab.2
Průměr :	11,42875
Rozptyl :	0,002526786
Šikmost	0,504214683
Špičatost :	2,497340209
Normalita :	Přijata
Vypočtený :	0,882110371
Teoretický :	5,991464547
Pravděpodobnost :	0,643357202

Normalita dat byla v obou případech přijata.

2. Test shodnosti (QC Expert- porovnání dvou výběrů)

a) test shody rozptylů:

vzhledem k tomu, že u obou výběrů byla přijata normalita použijeme klasický test shody rozptylů. Hypotéza $H_0: \sigma_1 = \sigma_2$ proti $H_a: \sigma_1 \neq \sigma_2$

Test shody rozptylů	
Poměr rozptylů :	1,836749117
Počet stupňů volnosti :	7
Kritická hodnota :	3,787043538
Závěr :	Rozptyly jsou SHODNÉ
Pravděpodobnost :	0,22048867

b) test shody průměrů :

vzhledem k tomu, že oba výběry mají shodné rozptyly, použijeme test shody průměrů pro shodné rozptyly. Hypotéza $H_0: \mu_1 = \mu_2$ proti $H_a: \mu_1 \neq \mu_2$

Test shody průměrů pro SHODNÉ rozptyly	
t-statistika :	11,35871597
Počet stupňů volnosti :	14
Kritická hodnota :	2,144786688
Závěr :	Průměry jsou ROZDÍLNÉ
Pravděpodobnost :	1,89E-08

Závěr: Test shody rozptylů ukázal, že oba laboranti pracují se stejným rozptylem. Test shody průměrů ukázal, že laboranti mají odlišné výsledky, tzn. nepracují stejně.