



# Návrhy dalších možností statistického zpracování aktualizovaných dat

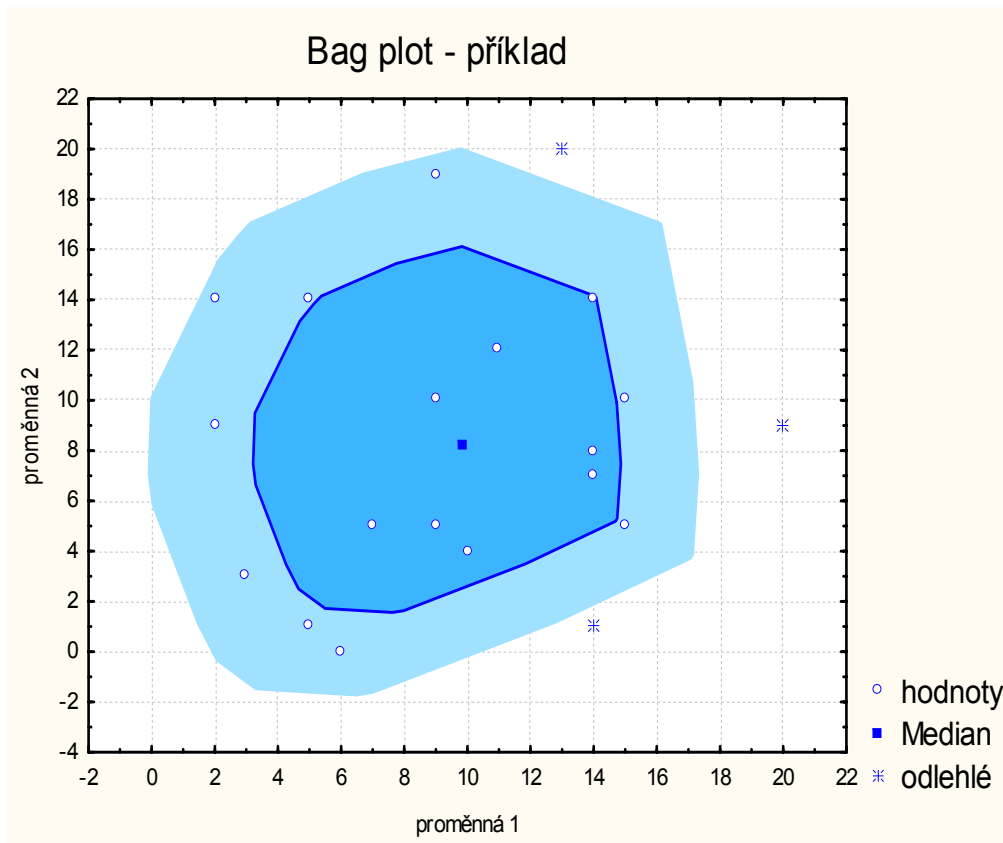
Při zjišťování disparit ve fyzické dostupnosti bydlení navrhuji použití těchto statistických metod:

- Bag plot;
- Krabicové grafy a jejich použití při zjišťování disparit ve fyzické dostupnosti bydlení;
- Grubbsův test extrémních hodnot.

# Bag plot

## Důvody pro použití:

- Bag plot je způsob grafického znázornění statistických dat.
- Jedná se o zobecnění krabicového grafu v dvourozměrném prostoru.
- Používá se tedy v případech, kdy můžeme (chceme) nějaký jev popsat pomocí dvou vysvětlujících proměnných.



## Popis:

Body v grafu vyjadřují realizace jevu, který je popsán pomocí dvou proměnných.

**Tmavě modrá oblast** (= bag) odpovídá „krabici“ klasického krabicového grafu s mediánem (tmavě modrý čtvereček) a kvartily. Uvnitř bagu leží tedy 50% pozorování (mezi prvním a třetím kvartilem).

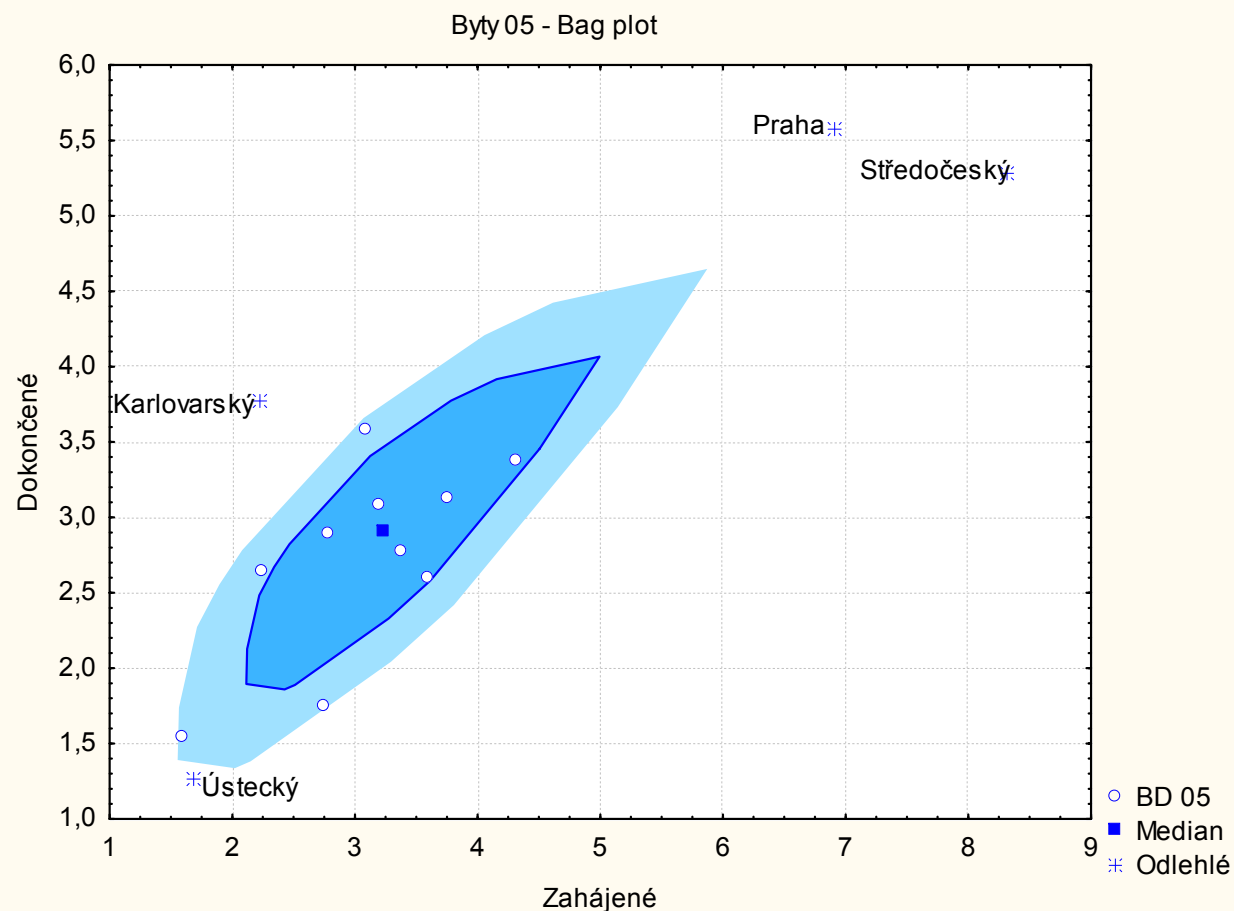
**Světle modrá oblast** reprezentuje interval vnitřních hradeb klasického krabicového grafu, uvnitř kterých leží neodlehle hodnoty. Koeficient rozsahu je roven 1,5 násobku interkvartilového rozpětí.

Mimo tuto oblast leží odlehle hodnoty, v grafu označené hvězdičkou.

# Bag plot – konkrétní příklady použití

V tabulce jsou počty zahájených a dokončených bytů na 1000 obyvatel v roce 2005 v krajích ČR. V příslušném grafu můžeme velmi dobře vidět, jaká je situace z hlediska bytové výstavby a které kraje se v tomto ohledu vymykají.

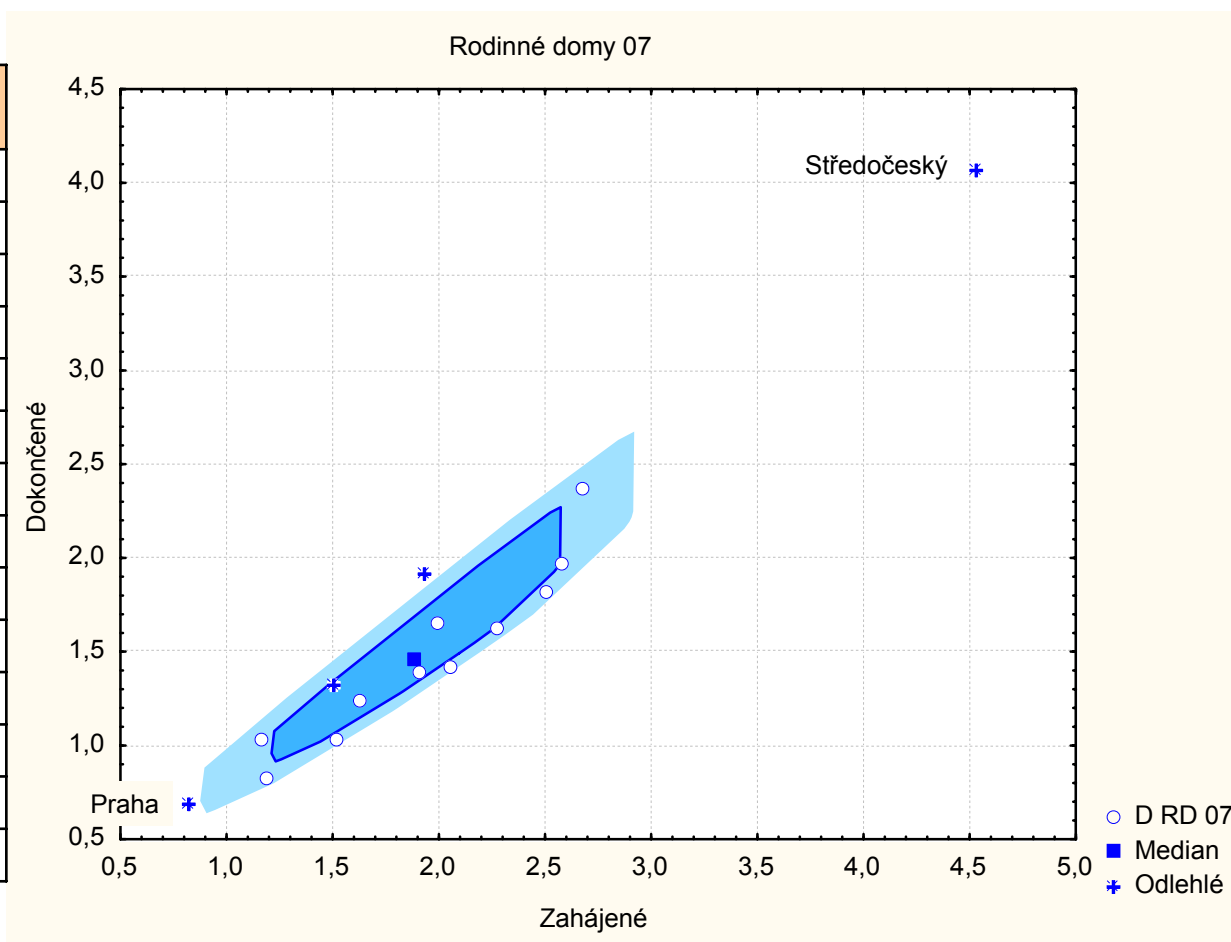
	zahájené	dokončené
Praha	6,907482	5,5810821
Středočeský	8,316466	5,289846
Jihočeský	3,763765	3,1207819
Plzeňský	3,085192	3,5884885
Karlovarský	2,222682	3,7723212
Ústecký	1,682915	1,266135
Liberecký	2,243922	2,6455397
Královéhradecký	3,595881	2,597431
Pardubický	3,384413	2,7692448
Vysočina	3,201961	3,0843137
Jihomoravský	4,3237	3,3761486
Olomoucký	2,754385	1,7496608
Zlínský	2,786025	2,8893364
Moravskoslezský	1,602535	1,5370273



# Bag plot – konkrétní příklady použití

V tabulce jsou počty zahájených a dokončených rodinných domů na 1000 obyvatel v roce 2007 v krajích ČR.

	zahájené	dokončené
Praha	0,820891	0,697139
Středočeský	4,527274	4,072966
Jihočeský	2,573966	1,972321
Plzeňský	2,272428	1,627236
Karlovarský	1,164421	1,040823
Ústecký	1,188672	0,830145
Liberecký	1,905758	1,396481
Královéhradecký	2,042694	1,43061
Pardubický	2,504889	1,822448
Vysočina	1,925334	1,917547
Jihomoravský	1,990296	1,655365
Olomoucký	1,615791	1,243395
Zlínský	1,496327	1,330444
Moravskoslezský	1,509724	1,032885





# Krabicové grafy

- Krabicové grafy můžeme použít v případech, kdy se chceme zabývat jednotlivými statistickými znaky odděleně
- Tímto způsobem opět můžeme velmi snadno určovat odlehlé hodnoty statistické proměnné.

## Grubbsův test extrémních hodnot

- V případech, kdy například pomocí krabicového grafu, určíme některou hodnotu jako odlehlou, můžeme tímto testem ověřit, zda se z hlediska statistického posuzování jedná o extrémní hodnotu.

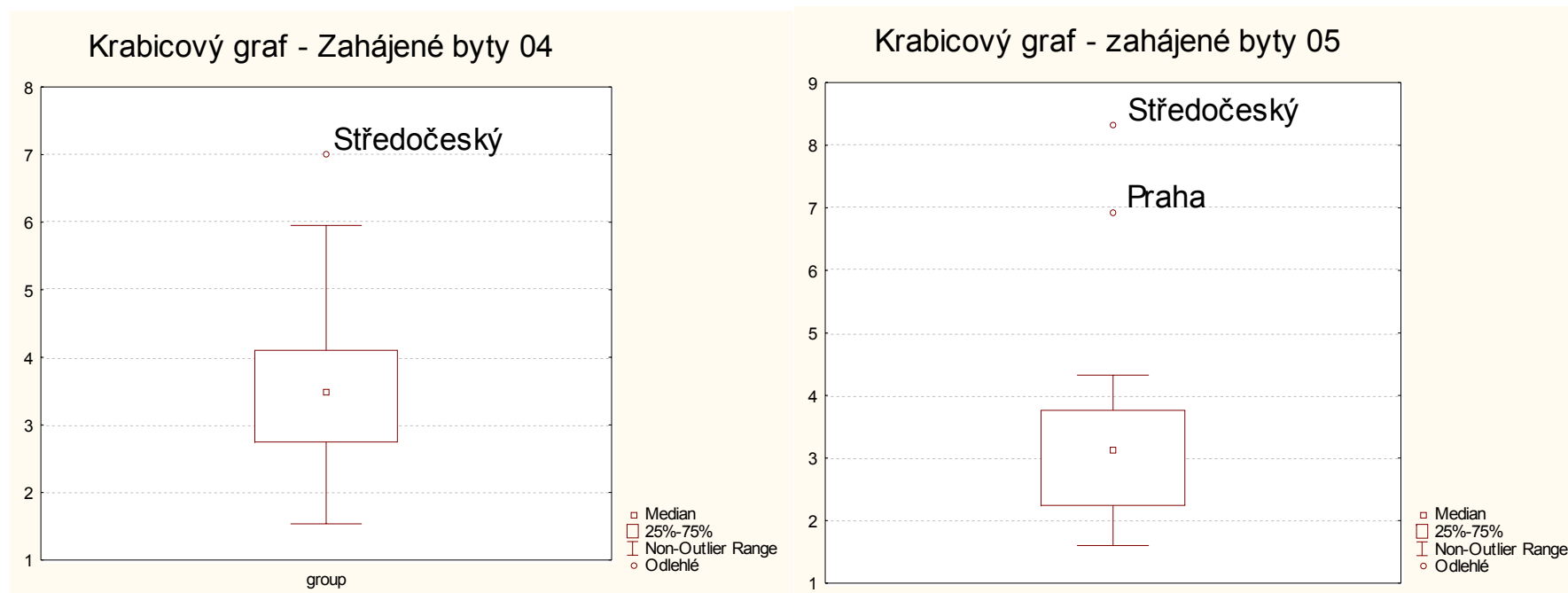
*Nulová hypotéza:* Odlehlá hodnota se významně neliší od ostatních hodnot souboru.

*Testovací kritérium:*  $T = |x_{\text{odlehlá}} - x_{\text{průměr}}|/s$  má speciální rozdělení pravděpodobnosti pro Grubbsův test.

Poznámka: V případě, že nemůžeme předpokládat normalitu dat, použijeme neparametrický Dixonův test.

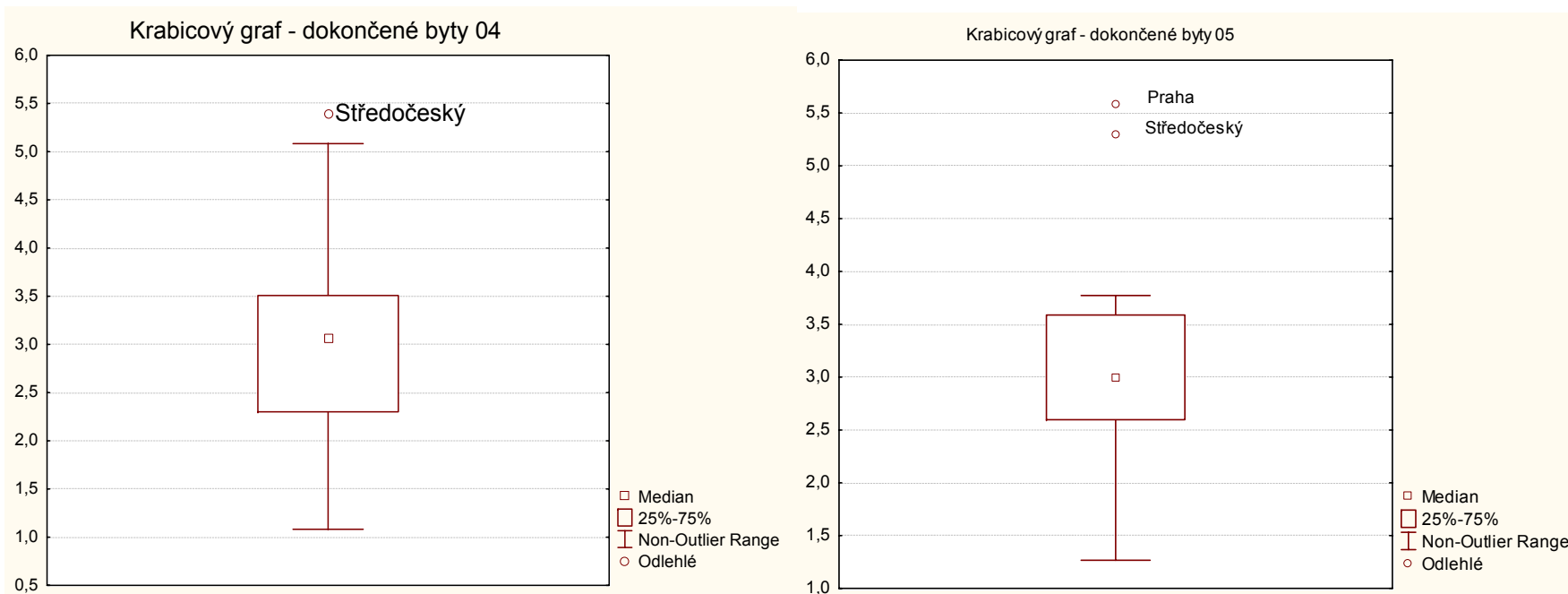
# Krabicové grafy a Grubbsův test – konkrétní příklady použití

Na následujících obrázcích vidíme krabicové grafy počtu zahájených bytů na 1000 obyvatel v jednotlivých krajích České republiky v letech 2004 a 2005. Pro lepší přehlednost jsou znázorněny pouze kraje s odlehlými hodnotami.



# Krabicové grafy a Grubbsův test

Na následujících obrázcích vidíme krabicové grafy počtu dokončených bytů na 1000 obyvatel v jednotlivých krajích České republiky v letech 2004, 2005. Pod grafy jsou výsledky Grubbsova testu extrémních hodnot.



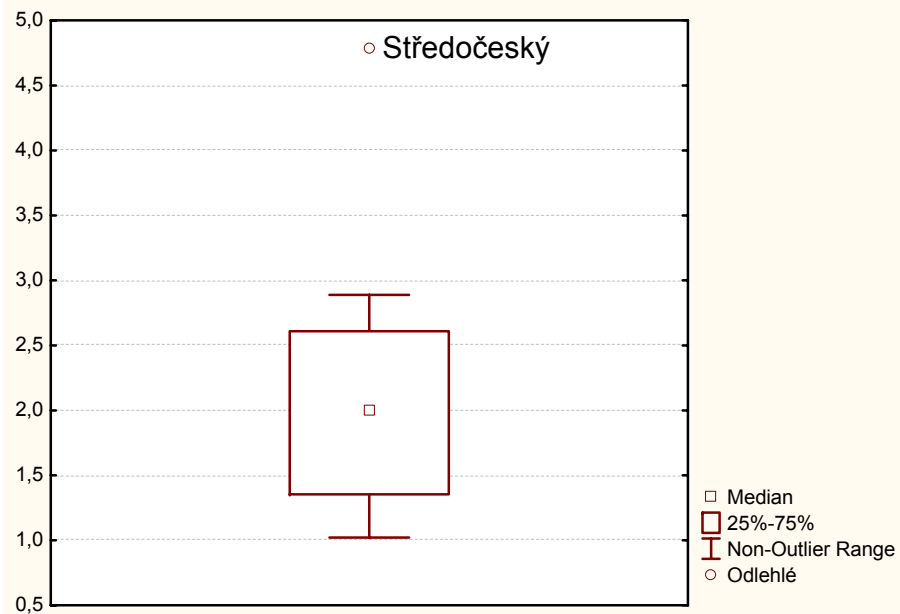
Variable	Descriptive Statistics (byty.sta)						
	Valid N	Mean	Grubbs Test Statistic	p-value	Minimum	Maximum	Std.Dev.
zahájené 04	14	3,614279	2,224198	0,382519	1,537863	7,012977	1,528055
zahájené 05	14	3,562237	2,501328	0,103347	1,602535	8,316466	1,900682
dokončené 04	14	2,995325	1,951585	1,000000	1,082110	5,385200	1,224581
dokončené 05	14	3,090526	2,009974	0,845888	1,266135	5,581082	1,239099

Jak můžeme vidět z tabulky,  $p$ -hodnoty Grubbsova testu přesáhly standardní hodnoty hladiny významnosti, tudíž odlehlé hodnoty nelze považovat za extrémně odchýlené.

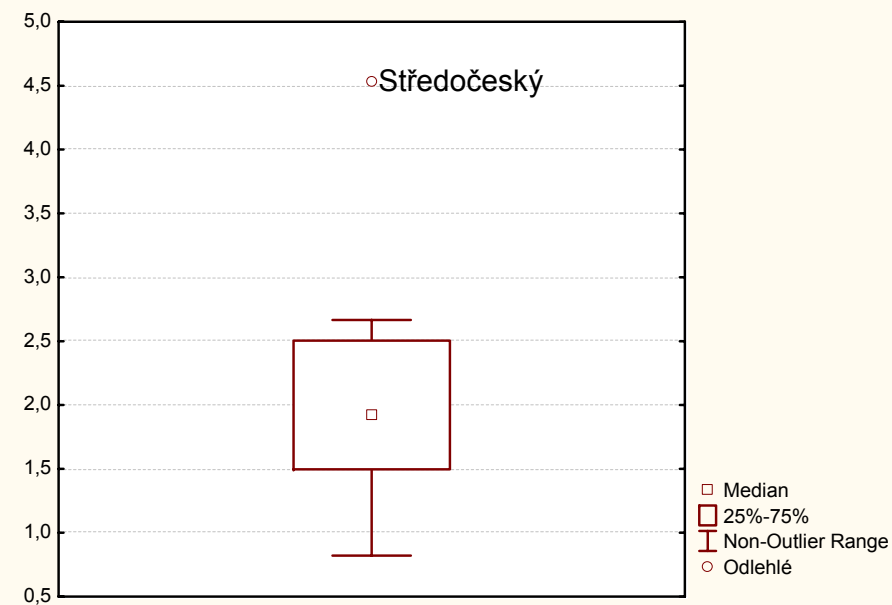
# Krabicové grafy a Grubbsův test

Na následujících obrázcích vidíme krabicové grafy počtu zahájených staveb rodinných domů na 1000 obyvatel v jednotlivých krajích České republiky v letech 2006 a 2007.

Krabicový graf - zahájené rodinné domy 06



Krabicový graf - zahájené rodinné domy 07

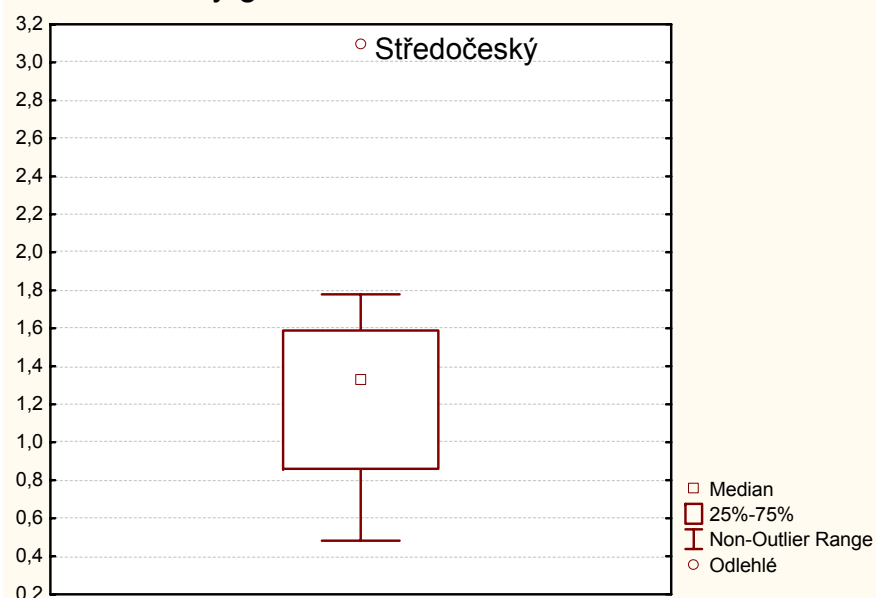




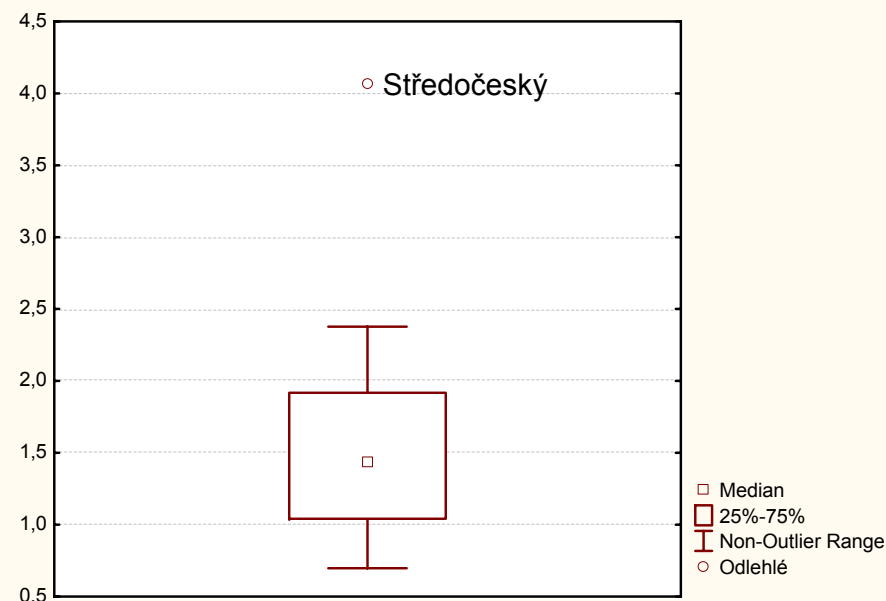
# Krabicové grafy a Grubbsův test

Na následujících obrázcích vidíme krabicové grafy počtu dokončených staveb rodinných domů na 1000 obyvatel v jednotlivých krajích České republiky v letech 2006, 2007.

Krabicový graf - dokončené RD 06



Krabicový graf - dokončené RD 07



Variable	Descriptive Statistics (rd.sta)						
	Valid N	Mean	Grubbs Test Statistic	p-value	Minimum	Maximum	Std.Dev.
zahájené RD 06	15	2,018673	2,885403	0,010826	1,021481	4,784246	0,958470
zahájené RD 07	15	2,013644	2,849847	0,014380	0,820891	4,527274	0,882023
dokončené RD 06	15	1,293031	2,833939	0,016251	0,482436	3,094630	0,635723
dokončené RD 07	15	1,629845	2,995608	0,004038	0,697139	4,072966	0,815567

Jak můžeme vidět z tabulky,  $p$ -hodnoty Grubbsova testu jsou menší než standardní hodnoty hladiny významnosti, tudíž odlehlé hodnoty můžeme považovat za extrémně odchýlené.