Analýza malých výběrů

- Závěry jsou vždy zatíženy značnou mírou nejistoty.
- Malých rozsahů jen tam, kde není možné zvýšit počet.

n = 2: 100(1 - α)%ní konfidenční interval střední hodnoty

$$\frac{x_1 + x_2}{2} - T_\alpha \frac{|x_1 - x_2|}{2} \leq \mu \leq \frac{x_1 + x_2}{2} + T_\alpha \frac{|x_1 - x_2|}{2}$$

- pro normální rozdělení $T_\alpha = \mathrm{cotg}(\alpha\, \pi / 2)$, $T_{0.05} = 12.71$,
- pro rovnoměrné rozdělení $T_\alpha = 1/\alpha - 1$, tj. $T'_{0.05} = 19$.

n = 3: 100(1 - α)%ní konfidenční interval střední hodnoty

$$\bar{x} - T'_\alpha \frac{s}{\sqrt{3}} \leq \mu \leq \bar{x} + T'_\alpha \frac{s}{\sqrt{3}}$$

- pro normální rozdělení je $T'_\alpha \approx 1/\sqrt{\alpha} - 3\sqrt{\alpha}/4$ ..., $T'_\alpha = 4.30$.
- pro rovnoměrné rozdělení je $T_{0.05} = 5.74$,

$4 \leq n \leq 20$, (Hornův postup):

je založený na pořádkových statistikách.

Hloubka pivotu je $H = (\mathrm{int}((n + 1)/2))/2$

nebo $H = (\mathrm{int}((n + 1)/2 + 1)/2,$

Dolní pivot je $x_D = x_{(H)}$ a **horní pivot** $x_H = x_{(n+1-H)}$.

Odhadem parametru polohy je **pivotová polosuma**

$$P_L = \frac{x_D + x_H}{2}$$

odhadem parametru rozptýlení **pivotové rozpětí**

$$R_L = x_H - x_D$$

Náhodná veličina k testování

$$T_L = \frac{P_L}{R_L} = \frac{x_D + x_H}{2\,(x_H - x_D)}$$

má přibližně symetrické rozdělení, jehož vybrané kvantily jsou v tabulce.

95%ní interval spolehlivosti střední hodnoty se vypočte

$$P_L - R_L\, t_{L,0.975}(n) \le \mu \le P_L + R_L\, t_{L,0.975}(n)$$

Tabulka 3.11 Kvantily $t_{L,\ 1-\alpha}(n)$ rozdělení T_L

n \ 1 - α	0.9	0.95	0.975	0.99	0.995
4	0.477	0.555	0.738	1.040	1.331
5	0.869	1.370	2.094	3.715	5.805
6	0.531	0.759	1.035	1.505	1.968
7	0.451	0.550	0.720	0.978	1.211
8	0.393	0.469	0.564	0.741	0.890
9	0.484	0.688	0.915	1.265	1.575
10	0.400	0.523	0.668	0.878	1.051
11	0.363	0.452	0.545	0.714	0.859
12	0.344	0.423	0.483	0.593	0.697
13	0.389	0.497	0.608	0.792	0.945
14	0.348	0.437	0.525	0.661	0.776
15	0.318	0.399	0.466	0.586	0.685
16	0.299	0.374	0.435	0.507	0.591
17	0.331	0.421	0.502	0.637	0.774
18	0.300	0.380	0.451	0.555	0.650
19	0.288	0.361	0.423	0.502	0.575
20	0.266	0.337	0.397	0.464	0.519

Úloha C3.11 *Test správnosti koncentrace tenzidů (Horn)*
Standardní vzorek obsahuje 2.5 mg/l anionaktivních tenzidů. Aplikujte i Hornův postup. Testujte, zda výsledky koncentrace standardu jsou správné. Jde o symetrické rozdělení ?
Data: Koncentrace tenzidů [mg/l]: 2.36 2.40 2.48 2.50 2.57 2.62 2.68
[Výsledky: Gauss. rozd., $\bar{x} = 2.52, \bar{x}_R = 2.51, \tilde{x}_{0.5} = 2.50, s = 0.12,$
$\hat{g}_1 = 0.04, \hat{g}_2 = 1.78, \quad 2.41 < \bar{x} < 2.62$*]*

Hornův postup:

1. Pořádkové statistiky:

i	1	2	3	4	5	6	7
$x_{(1)}$	2.36	2.40	2.48	2.50	2.57	2.62	2.68

2. Hloubka pivotu: n = 7, liché

$$H = integer\frac{\frac{n+1}{2}}{2} = \qquad \text{int}(2.0) \approx 2$$

3. Pivoty:

Dolní pivot $x_D = x_{(H)} =$ $x_{(2)}$ = **2.40**
Horní pivot $x_H = x_{(n+1-H)} =$ $x_{(6)}$ = **2.62**

4. Pivotová polosuma $P_L = \dfrac{x_D + x_H}{2} =$ $= \mathbf{2.51}$

5. Pivotové rozpětí $R_L = x_H - x_D =$ $2.62 - 2.40 = \mathbf{0.22}$

6. 95%ní interval spolehlivosti střední hodnoty μ: $t_{L,\ 1-\alpha/2}(7) = 0.720$

$$P_L - R_L\, t_{L,\ 1-\alpha/2}(n) \leq \mu \leq P_L + R_L\, t_{L,\ 1-\alpha/2}(n)$$

$$2.51 - 0.22 \times 0.72 \leq \mu \leq 2.51 + 0.22 \times 0.72$$

$$\mathbf{2.35 \leq \mu \leq 2.67}$$

Úloha B3.01 *Střední hodnota haptoglobinu v lidském krevním séru (Horn)*
Bylo provedeno měření koncentrace haptoglobinu v lidském krevním séru od osmi dospělých jedinců. Vypočtěte střední hodnotu, parametr rozptýlení a 95%ní interval spolehlivosti střední hodnoty. Aplikujte i Hornův postup.
Data: Koncentrace haptoglobinu [g l^{-1}]: 1.82 3.32 1.07 1.27 0.49 3.79 0.15 1.98

[Výsledky: Gauss. rozd., $\bar{x} = 1.74, \bar{x}_R = 1.51, \tilde{x}_{0.5} = 1.55, s = 1.28,$ $\hat{g}_1 = 0.46, \hat{g}_2 = 1.99,$ $0.66 < \bar{x} < 2.81$*]*

Hornův postup:

1. Pořádkové statistiky:

i	1	2	3	4	5	6	7	8
$x_{(1)}$	0.15	0.49	1.07	1.27	1.82	1.98	3.32	3.79

2. Hloubka pivotu: n = 8, sudé

$$H = integer\frac{\frac{n+1}{2}+1}{2} = \qquad \text{int}(2.75) \approx 2$$

3. Pivoty:

Dolní pivot $x_D = x_{(H)} =$ $x_{(2)}$ = **0.49**

Horní pivot $x_H = x_{(n+1-H)} =$ $x_{(7)}$ = **3.32**

4. Pivotová polosuma $P_L = \dfrac{x_D + x_H}{2} =$ $= \mathbf{1.905}$

5. Pivotové rozpětí $R_L = x_H - x_D =$ $3.32 - 0.49 = \mathbf{2.83}$

6. 95%ní interval spolehlivosti střední hodnoty μ: $t_{L,\ 1-\alpha/2} = 0.564$

$$P_L - R_L\, t_{L,\ 1-\alpha/2}(n) \leq \mu \leq P_L + R_L\, t_{L,\ 1-\alpha/2}(n)$$

$$1.905 - 2.83 \times 0.564 \leq \mu \leq 1.905 + 2.83 \times 0.564$$

$$\mathbf{0.31 \leq \mu \leq 3.50}$$

Úloha B3.09 *Test správnosti koncentrace cyclosporinu metodou HPLC (Horn)*
Pro studie biologické dostupnosti cyclosporinu A byl zakoupen roztok této látky v metanolu. Deklarovaná koncentrace cyclosporinu A byla 20 ng/ml. Při HPLC analýzách byly naměřeny následující koncentrace. Test správnosti je třeba provést na hladině významnosti $\alpha = 0.05$. Obsahuje intervalový odhad hodnotu 20 ng/ml ?

Data: Koncentrace cyclosporinu A [ng/ml]: 19.96 20.05 20.00 19.99 20.01 19.98 20.00 20.02 20.01 19.93

[Výsledky: Gauss. rozd., $\bar{x} = 20.00$, $\tilde{x}_{0.5} = 20.00$, $s = 0.03$, $\hat{g}_1 = -0.43$, $\hat{g}_2 = 2.99$, $19.97 < \bar{x} < 20.02$*]*

Hornův postup:

1. Pořádkové statistiky:

i	1	2	3	4	5	6	7	8	9	10
$x_{(1)}$	19.93	19.96	19.98	19.99	20.00	20.00	20.01	20.01	20.02	20.05

2. Hloubka pivotu: n = 10, sudé

$$H = integer\frac{\frac{n+1}{2}+1}{2} = \qquad 3.25 \approx 3$$

3. Pivoty:

Dolní pivot $x_D = x_{(H)} =$ $x_{(3)} = \mathbf{19.98}$

Horní pivot $x_H = x_{(n+1-H)} =$ $x_{(8)} = \mathbf{20.01}$

4. Pivotová polosuma $P_L = \dfrac{x_D + x_H}{2} = \dfrac{19.98 + 20.01}{2} = \mathbf{19.995}$

5. Pivotové rozpětí $R_L = x_H - x_D = 20.01 - 19.98 = \mathbf{0.03}$

6. 95%ní interval spolehlivosti střední hodnoty μ: $t_{L,\,1-\alpha/2}(10) = 0.668$

$$P_L - R_L\, t_{L,\,1-\alpha/2}(n) \leq \mu \leq P_L + R_L\, t_{L,\,1-\alpha/2}(n)$$

$$19.995 - 0.03 \times 0.668 \leq \mu \leq 19.995 + 0.03 \times 0.668$$

$$\mathbf{19.98 \leq \mu \leq 20.02}$$

Počítačová analýza malého výběru

Analýza jednorozměrných dat na počítači (výstup programu ADSTAT)

(a) TEST NORMALITY:

Tabulkový kvantil Chi^2(1-alfa,2):	5.9915
Chi^2-statistika:	0.80885
Závěr: *Předpoklad normality přijat*	
Vypočtená hladina významnosti:	0.66736

(b) TEST NEZÁVISLOSTI:

Tabulkový kvantil t(1-alfa/2,n+1):	2.2622
Test autokorelace:	1.0768
Závěr: *Předpoklad nezávislosti přijat*	
Vypočtená hladina významnosti:	0.15479

(c) DETEKCE ODLEHLÝCH BODŮ:

Závěr: *Ve výběru nejsou odlehlé body*

(d) PROSTÁ MOCNINNÁ TRANSFORMACE, (e) BOX-COXOVA TRANSFORMACE:

Zvolená mocnina:	0.53
Opravený průměr:	1.5102

(f) PARAMETRY TVARU:

Šikmost:	0.4574
Špičatost:	1.9926

(g) KLASICKÉ ODHADY PARAMETRŮ:

Průměr:		1.7363		
Směr. odchylka:		1.2830		
95.0% spolehlivost:	Mez spodní:	0.6636	horní:	2.8089

(h) ROBUSTNÍ ODHADY PARAMETRŮ:

Medián:		1.5450		
Směr. odchylka mediánu:		1.4215		
95.0% spolehlivost:	Mez spodní:	0.05249	horní:	3.0375
Uřezání 10% (pro P=0.10):				
Průměr:		1.6778		
Směr. odchylka:		1.4696		
95.0% spolehlivost:	Mez spodní:	0.3489	horní:	3.0067
Biweight:				
Průměr:		1.6791		
Směr. odchylka:		1.2340		
95.0% spolehlivost:	Mez spodní:	0.6022	horní:	2.7560

(i) ADAPTIVNÍ ODHADY PARAMETRŮ, Hoggovy odhady:

Průměr:		1.7363		
Směr. odchylka:		1.2830		
95.0% spolehlivost:	Mez spodní:	0.6636	horní:	2.8089

Poznatky z výstupu EDA:

Z průzkumové analýzy dat EDA a ověření předpokladů o výběru plyne, že rozdělení výběru pochází z Gaussova rozdělení, prvky výběru jsou nezávislé a ve výběru nejsou odlehlé body.

Lze proto užít také klasické odhady parametrů 1.74 g/l a 0.66 g/l $\leq \mu \leq$ 2.81 g/l.