Katedra aplikovanej matematiky, Strojnícka fakulta, UNIZA

Popisná štatistika a teoria odhadu

Pavol ORŠANSKÝ

17. októbra 2023

Popisná štatistika

V tejto kapitole budeme uvažovať o štatistickom súbore rozsahu N štatistických jednotiek (náhodných pokusov), pre ktorý sme zistili hodnoty skúmanej náhodnej premennej ξ pre každú štatistickú jednotku (náhodný pokus), čo znamená, že poznáme hodnoty x_i (pre $i = 1, 2, \ldots, N$) skúmanej premennej ξ.

[1] skupina entít so spoločnými znakmi

Popisná štatistika

V tejto kapitole budeme uvažovať o štatistickom súbore rozsahu N štatistických jednotiek (náhodných pokusov), pre ktorý sme zistili hodnoty skúmanej náhodnej premennej ξ pre každú štatistickú jednotku (náhodný pokus), čo znamená, že poznáme hodnoty x_i (pre $i = 1, 2, \ldots, N$) skúmanej premennej ξ.

Rozdelenie pravdepodobností hodnôt náhodnej premennej ξ dostaneme tzv. **triedením**, t. j. vytváraním *tried*[1] podobných štatistických jednotiek (pokusov).

[1]skupina entít so spoločnými znakmi

Popisná štatistika

V tejto kapitole budeme uvažovať o štatistickom súbore rozsahu N štatistických jednotiek (náhodných pokusov), pre ktorý sme zistili hodnoty skúmanej náhodnej premennej ξ pre každú štatistickú jednotku (náhodný pokus), čo znamená, že poznáme hodnoty x_i (pre $i = 1, 2, \ldots, N$) skúmanej premennej ξ.

Rozdelenie pravdepodobností hodnôt náhodnej premennej ξ dostaneme tzv. **triedením**, t. j. vytváraním *tried*[1] podobných štatistických jednotiek (pokusov).

Pri triedení je potrebné dodržať **zásadu úplnosti**, t. j. každý prvok musí byť zaradený v nejakej triede a **zásadu jednoznačnosti**, t. j. každý prvok musí byť zaradený iba v jednej triede.

[1] skupina entít so spoločnými znakmi

Popisná štatistika - **Základné pojmy a označenie**

hodnoty náhodnej premennej ξ	$x_i \dots$ konkrétna hodnota ξ pri i-to pokuse,
absolútna početnosť	$n_i \dots$ koľkokrát sa znak vo výbere nachádza,
relatívna početnosť	$p_i = \dfrac{n_i}{N}$,
kumulatívna početnosť	$N_i = \sum_{j=1}^{i} n_j$,
kumulatívna relatívna početnosť	$M_i = \dfrac{N_i}{N}$,

kde $i = 1, 2, \dots, N$.

Príklad (1.a - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1				
2				
3				
4				
5				
6				

Príklad (Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2				
3				
4				
5				
6				

Popisná štatistika - príklad

Príklad (1.c - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2	6			
3				
4				
5				
6				

Popisná štatistika - príklad

Príklad (1.d - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2	6			
3	7			
4				
5				
6				

Popisná štatistika - príklad

Príklad (1.e - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2	6			
3	7			
4	8			
5				
6				

Popisná štatistika - príklad

Príklad (1.f - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2	6			
3	7			
4	8			
5	4			
6				

Príklad (1.g - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2	6			
3	7			
4	8			
5	4			
6	2			

Príklad (1.h - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3			
2	6			
3	7			
4	8			
5	4			
6	2			
	$\sum = 30$			

Príklad (1.ch - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$		
2	6	$6/30$		
3	7	$7/30$		
4	8	$8/30$		
5	4	$4/30$		
6	2	$2/30$		
	$\sum = 30$	$\sum = 1$		

Príklad (1.i - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$	3	
2	6	$6/30$		
3	7	$7/30$		
4	8	$8/30$		
5	4	$4/30$		
6	2	$2/30$		
	$\sum = 30$	$\sum = 1$		

Príklad (1.i - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$	3	
2	6	$6/30$	9	
3	7	$7/30$		
4	8	$8/30$		
5	4	$4/30$		
6	2	$2/30$		
	$\sum = 30$	$\sum = 1$		

Popisná štatistika - príklad

Príklad (1.i - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	3/30	3	
2	6	6/30	9	
3	7	7/30	16	
4	8	8/30		
5	4	4/30		
6	2	2/30		
	$\sum = 30$	$\sum = 1$		

Popisná štatistika - príklad

Príklad (1.i - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$	3	
2	6	$6/30$	9	
3	7	$7/30$	16	
4	8	$8/30$	24	
5	4	$4/30$		
6	2	$2/30$		
	$\sum = 30$	$\sum = 1$		

Popisná štatistika - príklad

Príklad (1.i - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$	3	
2	6	$6/30$	9	
3	7	$7/30$	16	
4	8	$8/30$	24	
5	4	$4/30$	28	
6	2	$2/30$		
	$\sum = 30$	$\sum = 1$		

Popisná štatistika - príklad

Príklad (1.i - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$	3	
2	6	$6/30$	9	
3	7	$7/30$	16	
4	8	$8/30$	24	
5	4	$4/30$	28	
6	2	$2/30$	30	
	$\sum = 30$	$\sum = 1$		

Popisná štatistika - príklad

Príklad (1.j - Popisná štatistika - tabuľka rozdelenie pravdepodobnosti)

Zisťoval sa počet obyvateľov v tridsiatich bytoch. Hodnoty sú:

$$3, 2, 4, 5, 2, 2, 3, 2, 4, 5, 1, 3, 4, 4, 5, 4, 1, 3, 6, 2, 3, 4, 6, 2, 3, 4, 1, 3, 5, 4.$$

Zostavme tabuľku rozdelenia početnosti.

Riešenie:

x_i :	n_i :	p_i :	N_i :	M_i :
1	3	$3/30$	3	$3/30$
2	6	$6/30$	9	$9/30$
3	7	$7/30$	16	$16/30$
4	8	$8/30$	24	$24/30$
5	4	$4/30$	28	$28/30$
6	2	$2/30$	30	$30/30=1$
	$\sum = 30$	$\sum = 1$		

Príklad (2.a - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

[2] S ohľadom na **zásadu úplnosti vždy zaokrúhľujeme dĺžku intervalu nahor**, aby sme zaistili, dostatočný rozsah (minimálne rámec variačného rozpätia) a i krajné znaky boli obsiahnuté vo svojich triedach.

Príklad (2.a - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie: Počet tried je zjavne menší ako počet hodnôt znaku, budeme preto musieť vytvoriť **intervalové rozdelenie pravdepodobnosti**.
Najskôr určíme na základe **variačného rozpätia**

$$R = x_{\text{max}} - x_{\text{min}} = 116 - 34.8 = 81.2.$$

[2] S ohľadom na **zásadu úplnosti vždy zaokrúhľujeme dĺžku intervalu nahor**, aby sme zaistili, dostatočný rozsah (minimálne rámec variačného rozpätia) a i krajné znaky boli obsiahnuté vo svojich triedach.

Príklad (2.a - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie: Počet tried je zjavne menší ako počet hodnôt znaku, budeme preto musieť vytvoriť **intervalové rozdelenie pravdepodobnosti**.
Najskôr určíme na základe **variačného rozpätia**

$$R = x_{\text{max}} - x_{\text{min}} = 116 - 34.8 = 81.2.$$

A potom **šírku triedy**[2], reprezentovanú intervalom dlžky, dostávame ako

$$h = \frac{R}{k} = \frac{\text{variačného rozpätia}}{\text{počet tried}} = \frac{81.2}{9} = 9.022\,2 \doteq 10.$$

[2] S ohľadom na **zásadu úplnosti vždy zaokrúhľujeme dĺžku intervalu nahor**, aby sme zaistili, dostatočný rozsah (minimálne rámec variačného rozpätia) a i krajné znaky boli obsiahnuté vo svojich triedach.

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$					
$[40, 50)$					
$[50, 60)$					
$[60, 70)$					
$[70, 80)$					
$[80, 90)$					
$[90, 100)$					
$[100, 110)$					
$[110, 120)$					

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35				
$[40, 50)$	45				
$[50, 60)$	55				
$[60, 70)$	65				
$[70, 80)$	75				
$[80, 90)$	85				
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45				
$[50, 60)$	55				
$[60, 70)$	65				
$[70, 80)$	75				
$[80, 90)$	85				
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55				
$[60, 70)$	65				
$[70, 80)$	75				
$[80, 90)$	85				
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65				
$[70, 80)$	75				
$[80, 90)$	85				
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65	7			
$[70, 80)$	75				
$[80, 90)$	85				
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65	7			
$[70, 80)$	75	7			
$[80, 90)$	85				
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65	7			
$[70, 80)$	75	7			
$[80, 90)$	85	4			
$[90, 100)$	95				
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65	7			
$[70, 80)$	75	7			
$[80, 90)$	85	4			
$[90, 100)$	95	2			
$[100, 110)$	105				
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65	7			
$[70, 80)$	75	7			
$[80, 90)$	85	4			
$[90, 100)$	95	2			
$[100, 110)$	105	1			
$[110, 120)$	115				

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1			
$[40, 50)$	45	2			
$[50, 60)$	55	5			
$[60, 70)$	65	7			
$[70, 80)$	75	7			
$[80, 90)$	85	4			
$[90, 100)$	95	2			
$[100, 110)$	105	1			
$[110, 120)$	115	1			

Príklad (2.b - Popisná štatistika - intervalové rozdelenie pravdepodobnosti)

Nájdime skupinové rozdelenie početnosti (počet tried $k = 9$) pre rozlohy týchto bytov:

$$82.6, 57.3, 70.4, 65, 48.4, 103.8, 73.6, 43.5, 66.1, 93, 52.6, 70, 84.2, 55, 81.3,$$
$$61.5, 75.1, 34.8, 62.4, 116, 70.1, 63.6, 93, 59.2, 65.9, 77.2, 52.8, 68.7, 79.2, 87.4.$$

Riešenie:

	x_j :	n_j :	p_j :	N_j :	M_j :
$[30, 40)$	35	1	$1/30$	1	$1/30$
$[40, 50)$	45	2	$2/30$	3	$3/30$
$[50, 60)$	55	5	$5/30$	8	$8/30$
$[60, 70)$	65	7	$7/30$	15	$15/30$
$[70, 80)$	75	7	$7/30$	22	$22/30$
$[80, 90)$	85	4	$4/30$	26	$26/30$
$[90, 100)$	95	2	$2/30$	28	$28/30$
$[100, 110)$	105	1	$1/30$	29	$29/30$
$[110, 120)$	115	1	$1/30$	30	$30/30 = 1$

Odhady parametrov

Odhadom máme na mysli štatistickú metódu, pomocou ktorej sa približne určujú (odhadujú) neznáme parametre štatistických súborov (ich číselné charakteristiky)[3].

[3]Odhady, pri ktorých hľadáme určitý parameter, nazývame *parametrické odhady*. *Neparametrickými odhadmi* nazývame odhady, pri ktorých nie je požadovaná parametrická špecifikácia typu pravdepodobnostného rozdelenia

[4]Θ je veľké grécke písmeno „théta"

Odhady parametrov

Odhadom máme na mysli štatistickú metódu, pomocou ktorej sa približne určujú (odhadujú) neznáme parametre štatistických súborov (ich číselné charakteristiky)[3].

Majme náhodný výber $\xi_1, \xi_2, \ldots, \xi_n$ z určitého rozdelenia, ktoré závisí na neznámom parametri Θ[4], potom parameter Θ môže nadobúdať iba určité hodnoty z priestoru Ω. Prostredníctvom teórie odhadu sa snažíme vytvoriť štatistiku $T(\xi_1, \xi_2, \ldots, \xi_n)$, ktorej rozdelenie sa najviac blíži danému parametru $\Theta \subset \Omega$.

---

[3] Odhady, pri ktorých hľadáme určitý parameter, nazývame *parametrické odhady*. *Neparametrickými odhadmi* nazývame odhady, pri ktorých nie je požadovaná parametrická špecifikácia typu pravdepodobnostného rozdelenia

[4] Θ je veľké grécke písmeno „théta"

Odhady parametrov

Odhadom máme na mysli štatistickú metódu, pomocou ktorej sa približne určujú (odhadujú) neznáme parametre štatistických súborov (ich číselné charakteristiky)[3].

Majme náhodný výber $\xi_1, \xi_2, \ldots, \xi_n$ z určitého rozdelenia, ktoré závisí na neznámom parametri Θ[4], potom parameter Θ môže nadobúdať iba určité hodnoty z priestoru Ω. Prostredníctvom teórie odhadu sa snažíme vytvoriť štatistiku $T(\xi_1, \xi_2, \ldots, \xi_n)$, ktorej rozdelenie sa najviac blíži danému parametru $\Theta \subset \Omega$.

Rozlišujeme dva typy odhadov, a to **bodové odhady**, kedy parameter odhadneme konkrétnou hodnotou s prípustnou chybou a **intervalové odhady,** kedy určíme interval, ktorý odhadovaný parameter obsahuje s vopred stanovenou pravdepodobnosťou.

---

[3] Odhady, pri ktorých hľadáme určitý parameter, nazývame *parametrické odhady*. *Neparametrickými odhadmi* nazývame odhady, pri ktorých nie je požadovaná parametrická špecifikácia typu pravdepodobnostného rozdelenia

[4] Θ je veľké grécke písmeno „théta"

Odhady parametrov - Bodový odhad

Bodový odhad spočíva v nahradení neznámej hodnoty parametra základného súboru, alebo jeho funkcie, hodnotou výberovej charakteristiky.
Kladieme na určité nároky, čo sa týka jeho konzistentnosti a nevychýlenosti.

Odhady parametrov - Bodový odhad

Bodový odhad spočíva v nahradení neznámej hodnoty parametra základného súboru, alebo jeho funkcie, hodnotou výberovej charakteristiky.
Kladieme na určité nároky, čo sa týka jeho konzistentnosti a nevychýlenosti.

Konzistentným (nesporným) bodovým odhadom parametra Θ základného súboru nazývame takú štatistiku $T_n = T(\xi_1, \xi_2, \ldots, \xi_n)$, ktorá pre dostatočne veľké hodnoty indexu n spĺňa podmienku

$$P(|T_n - \Theta| \leq \varepsilon) < 1 - \eta,$$

pre ľubovoľné $\varepsilon > 0$ a $\eta > 0$. Inak povedané, konzistentným bude bodový odhad vtedy, ak bude ležať v čo najmenšom intervale s čo najväčšou pravdepodobnosťou.

Odhady parametrov - Bodový odhad

Bodový odhad spočíva v nahradení neznámej hodnoty parametra základného súboru, alebo jeho funkcie, hodnotou výberovej charakteristiky.
Kladieme na určité nároky, čo sa týka jeho konzistentnosti a nevychýlenosti.

Konzistentným (nesporným) bodovým odhadom parametra Θ základného súboru nazývame takú štatistiku $T_n = T(\xi_1, \xi_2, \ldots, \xi_n)$, ktorá pre dostatočne veľké hodnoty indexu n spĺňa podmienku

$$P(|T_n - \Theta| \leq \varepsilon) < 1 - \eta,$$

pre ľubovoľné $\varepsilon > 0$ a $\eta > 0$. Inak povedané, konzistentným bude bodový odhad vtedy, ak bude ležať v čo najmenšom intervale s čo najväčšou pravdepodobnosťou.

Nevychýlený (nestranný) bodový odhad parametra Θ základného súboru nazývame štatistiku $T_n = T(\xi_1, \xi_2, \ldots, \xi_n)$, pre ktorej strednú hodnotu platí $E(T_n) = \Theta$.

Odhady parametrov - Bodový odhad

Najlepším neskresleným **bodovým odhadom strednej hodnoty** základného súboru

$$E(\xi) = \mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

je **výberový priemer**

$$\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i .$$

Odhady parametrov - Bodový odhad

Najlepším neskresleným **bodovým odhadom strednej hodnoty** základného súboru

$$E(\xi) = \mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

je **výberový priemer**

$$\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i.$$

Najlepším neskresleným **bodovým odhadom rozptylu** základného súboru

$$D(\xi) = \sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$$

je **výberový rozptyl**

$$S_x^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2.$$

Odhady parametrov - Bodový odhad

V prípade **skupinového rozdelenia pravdepodobnosti**, kde sú získané hodnoty reprezentované stredmi intervalov a ich početnosťami, t. j. ak nemáme k dispozícii priamo hodnoty ale iba tabuľku skupinového rozdelenia, používame nasledujúce vzťahy. Najlepším neskresleným bodovým **odhadom strednej hodnoty** základného súboru je **skupinový výberový priemer**

$$\overline{x} = \frac{1}{N} \cdot \sum_{j=1}^{k} n_j \cdot x_j.$$

V prípade **skupinového rozdelenia pravdepodobnosti**, kde sú získané hodnoty reprezentované stredmi intervalov a ich početnosťami, t. j. ak nemáme k dispozícii priamo hodnoty ale iba tabuľku skupinového rozdelenia, používame nasledujúce vzťahy. Najlepším neskresleným bodovým **odhadom strednej hodnoty** základného súboru je **skupinový výberový priemer**

$$\overline{x} = \frac{1}{N} \cdot \sum_{j=1}^{k} n_j \cdot x_j.$$

Najlepším neskresleným bodovým **odhadom rozptylu** základného súboru je **skupinový výberový rozptyl**

$$S_x^2 = \frac{1}{N-1} \cdot \sum_{j=1}^{k} n_j \cdot (x_j - \overline{x})^2.$$

Príklad (3. - Bodový odhad skupinového rozdelenia pravdepodobnosti)

Z predošlého príkladu vezmime iba hodnoty stredov intervalov a ich početností a bodovo odhadnime výberový priemer a výberový rozptyl.

x_j :	35	45	55	65	75	85	95	105	115
n_j :	1	2	5	7	7	4	2	1	1

Príklad (3. - Bodový odhad skupinového rozdelenia pravdepodobnosti)

Z predošlého príkladu vezmime iba hodnoty stredov intervalov a ich početností a bodovo odhadnime výberový priemer a výberový rozptyl.

x_j :	35	45	55	65	75	85	95	105	115
n_j :	1	2	5	7	7	4	2	1	1

Riešenie: Výberový priemer

$$\overline{x} = \frac{1}{N}\cdot\sum_{j=1}^{k} n_j\cdot x_j = \frac{1}{30}\cdot\sum_{j=1}^{9} n_j\cdot x_j = \frac{1}{30}\cdot\Big(1\cdot 35 + 2\cdot 45$$

$$+5\cdot 55 + 7\cdot 65 + 7\cdot 75 + 4\cdot 85 + 2\cdot 95 + 1\cdot 105 + 1\cdot 115\Big) = 71.$$

Odhady parametrov - Bodový odhad

Riešenie: Výberový rozptyl

$$
\begin{aligned}
S_x^2 &= \frac{1}{N-1}\cdot\sum_{j=1}^{k} n_{j\cdot}\,(x_j-\overline{x})^2 = \frac{1}{30-1}\cdot\sum_{j=1}^{9} n_j\cdot(x_j-71)^2\\
&= \frac{1}{30-1}\cdot\Big(1\cdot(35-71)^2+2\cdot(45-71)^2+5\cdot(55-71)^2\\
&\quad +7\cdot(65-71)^2+7\cdot(75-71)^2+4\cdot(85-71)^2\\
&\quad +2\cdot(95-71)^2+1\cdot(105-71)^2+1\cdot(115-71)^2\Big)\\
&= \frac{1}{30-1}\cdot\Big(1296+1352+1280+252+112+784+1152+1156+1936\Big)\\
&= 321.38.
\end{aligned}
$$

Intervalový odhad nazývame interval spoľahlivosti (T_a, T_b) a neznámy parameter Θ bude tento interval obsahovať s pravdepodobnosťou $(1-\alpha)\%$, kde číslo α nazývame **hladina významnosti**[5].

[5] vopred zvolená „mieru presnosti"s akou interval, neznámy parameter obsahuje

Intervalový odhad nazývame interval spoľahlivosti (T_a, T_b) a neznámy parameter Θ bude tento interval obsahovať s pravdepodobnosťou $(1-\alpha)\%$, kde číslo α nazývame **hladina významnosti**[5].

Ak výber opakujeme mnohokrát (vzhľadom na tendenciu stability náhodných procesov), potom neznámy parameter Θ „padne"do intervalu spoľahlivosti (T_a, T_b) približne v $100 \cdot (1-\alpha)\%$ prípadoch, t. j. pravdepodobnosť, že daný parameter sa nachádza v intervale (T_a, T_b) je rovná práve číslu $1-\alpha$.

[5] vopred zvolená „mieru presnosti"s akou interval, neznámy parameter obsahuje

Odhady parametrov - Intervalový odhad parametrov

Intervalový odhad nazývame interval spoľahlivosti (T_a, T_b) a neznámy parameter Θ bude tento interval obsahovať s pravdepodobnosťou $(1-\alpha)\%$, kde číslo α nazývame **hladina významnosti**[5].

Ak výber opakujeme mnohokrát (vzhľadom na tendenciu stability náhodných procesov), potom neznámy parameter Θ „padne"do intervalu spoľahlivosti (T_a, T_b) približne v $100 \cdot (1-\alpha)\%$ prípadoch, t. j. pravdepodobnosť, že daný parameter sa nachádza v intervale (T_a, T_b) je rovná práve číslu $1-\alpha$.
Hovoríme o tzv. $100 \cdot (1-\alpha)\%$-om intervale spoľahlivosti a zapisujeme

$$P(T_a \leq \Theta \leq T_b) = 1-\alpha$$

V prípade, že hranice T_a, T_b sú konečné, hovoríme o **obojstrannom intervale spoľahlivosti**.

[5] vopred zvolená „mieru presnosti"s akou interval, neznámy parameter obsahuje

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Výberový priemer má mať normálne rozdelenie $\overline{x} \sim N\left(\mu, \frac{S_x^2}{n}\right)$ jej zodpovedá hodnotu distribučnej funkcie $\Phi\left(\frac{\overline{x}-\mu}{S_x} \cdot \sqrt{n}\right)$ pre normálne normované rozdelenie $N(0,1)$.

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Výberový priemer má mať normálne rozdelenie $\overline{x} \sim N\left(\mu, \frac{S_x^2}{n}\right)$ jej zodpovedá hodnotu distribučnej funkcie $\Phi\left(\frac{\overline{x}-\mu}{S_x}\cdot\sqrt{n}\right)$ pre normálne normované rozdelenie $N(0,1)$.

Pre zvolené presnosť α nájdeme **kvantily** $N(0,1)$ rozdelenia $u_{\left(1-\frac{\alpha}{2}\right)}$ a $-u_{\left(1-\frac{\alpha}{2}\right)}$ a platí

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Výberový priemer má mať normálne rozdelenie $\overline{x} \sim N\left(\mu, \frac{S_x^2}{n}\right)$ jej zodpovedá hodnotu distribučnej funkcie $\Phi\left(\frac{\overline{x}-\mu}{S_x} \cdot \sqrt{n}\right)$ pre normálne normované rozdelenie $N(0,1)$.

Pre zvolené presnosť α nájdeme **kvantily** $N(0,1)$ rozdelenia $u_{\left(1-\frac{\alpha}{2}\right)}$ a $-u_{\left(1-\frac{\alpha}{2}\right)}$ a platí

$$P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\overline{x}-\mu}{S_x} \cdot \sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Výberový priemer má mať normálne rozdelenie $\overline{x} \sim N\left(\mu, \dfrac{S_x^2}{n}\right)$ jej zodpovedá hodnotu distribučnej funkcie $\Phi\left(\dfrac{\overline{x}-\mu}{S_x} \cdot \sqrt{n}\right)$ pre normálne normované rozdelenie $N(0,1)$.

Pre zvolené presnosť α nájdeme **kvantily** $N(0,1)$ rozdelenia $u_{\left(1-\frac{\alpha}{2}\right)}$ a $-u_{\left(1-\frac{\alpha}{2}\right)}$ a platí

$$P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\overline{x}-\mu}{S_x} \cdot \sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

$$\begin{aligned} -\frac{S_x}{\sqrt{n}} \cdot u_{\left(1-\frac{\alpha}{2}\right)} \quad &\leq \quad \overline{x}-\mu \leq \frac{S_x}{\sqrt{n}} \cdot u_{\left(1-\frac{\alpha}{2}\right)}, \qquad && /-\overline{x} \\ -\overline{x} - \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \quad &\leq \quad -\mu \leq -\overline{x} + \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, \qquad && /\cdot(-1) \end{aligned}$$

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Výberový priemer má mať normálne rozdelenie $\overline{x} \sim N\left(\mu, \frac{S_x^2}{n}\right)$ jej zodpovedá hodnotu distribučnej funkcie $\Phi\left(\frac{\overline{x}-\mu}{S_x} \cdot \sqrt{n}\right)$ pre normálne normované rozdelenie $N(0,1)$.

Pre zvolené presnosť α nájdeme **kvantily** $N(0,1)$ rozdelenia $u_{\left(1-\frac{\alpha}{2}\right)}$ a $-u_{\left(1-\frac{\alpha}{2}\right)}$ a platí

$$P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\overline{x}-\mu}{S_x} \cdot \sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

$$\begin{aligned} -\frac{S_x}{\sqrt{n}} \cdot u_{\left(1-\frac{\alpha}{2}\right)} &\leq \overline{x}-\mu \leq \frac{S_x}{\sqrt{n}} \cdot u_{\left(1-\frac{\alpha}{2}\right)}, && /-\overline{x} \\ -\overline{x}-\frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} &\leq -\mu \leq -\overline{x}+\frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, && /\cdot(-1) \end{aligned}$$

$$\overline{x}-\frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{x}+\frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}.$$

Studentovo rozdelenie

Predpokladajme štatistiku $\Theta \backsim \dfrac{\overline{x} - \mu}{S_x} \cdot \sqrt{n}$, v ktorej smerodajnú odchýlku σ nahradí výberovou štandardou odchýlkou S_x, podhodnotí sa variabilita. Studentovo rozdelenie s n stupňami voľnosti, ktoré označujeme $t^{(n)}$, je rozdelenie náhodnej veličiny $\xi = \dfrac{X}{\sqrt{\dfrac{Y}{n}}}$,

kde X a Y sú navzájom nezávislé náhodné veličiny, pričom X má normované normálne rozdělenie $X \sim N(0, 1)$ a Y má chi-kvadrát rozdelenie s n stupňami voľnosti $\chi^2(n)$.

Studentovo rozdelenie

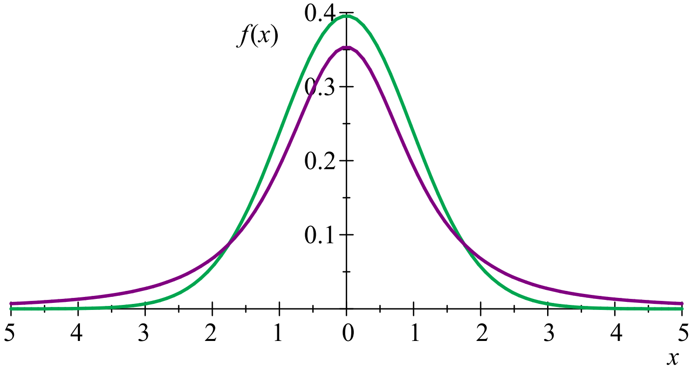
Predpokladajme štatistiku $\Theta \backsim \frac{\overline{x} - \mu}{S_x} \cdot \sqrt{n}$, v ktorej smerodajnú odchýlku σ nahradí výberovou štandardou odchýlkou S_x, podhodnotí sa variabilita. Studentovo rozdelenie s n stupňami voľnosti, ktoré označujeme $t^{(n)}$, je rozdelenie náhodnej veličiny $\xi = \frac{X}{\sqrt{\frac{Y}{n}}}$,

kde X a Y sú navzájom nezávislé náhodné veličiny, pričom X má normované normálne rozdělenie $X \sim N(0,1)$ a Y má chi-kvadrát rozdelenie s n stupňami voľnosti $\chi^2(n)$. Rozdelenie $t^{(n)}$ má pre $\infty < x < \infty$ a stupne voľnosti $n = 1, 2, 3, \ldots$ hustotu pravdepodobnosti určenú fukciou

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{1}{n} \cdot x^2\right)^{-\frac{n+1}{2}},$$

a strednú hodnotu a rozptyl sú pre $n > 1$

$$E(\xi) = 0, \qquad D(\xi) = \frac{n}{n-2}.$$



Obr.: Graf hustoty pravdepodobnosti Studentovho rozdelenia pre 30 stupňov voľnosti a $N(0,1)$

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Pre súbory **veľkého ($n \geq 30$) rozsahu** použijeme pre intervalový odhad strednej hodnoty μ na hladine významnosti α hodnoty kvantilov normálneho normovaného rozdelenia $u_{1-\frac{\alpha}{2}}$, a teda platí:

$$\boxed{\overline{x} - \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}}.$$

Odhady parametrov - Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

Pre súbory **veľkého ($n \geq 30$) rozsahu** použijeme pre intervalový odhad strednej hodnoty μ na hladine významnosti α hodnoty kvantilov normálneho normovaného rozdelenia $u_{1-\frac{\alpha}{2}}$, a teda platí:

$$\boxed{\overline{x} - \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}}.$$

Pre súbory **malého ($n < 30$) rozsahu** použijeme pre intervalový odhad strednej hodnoty μ na hladine významnosti α hodnoty kvantilov Studentovho t-rozdelenia $t_{1-\frac{\alpha}{2}}^{(n-1)}$ s $(n-1)$ stupňami voľnosti, a teda platí:

$$\boxed{\overline{x} - \frac{S_x}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}^{(n-1)} \leq \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}^{(n-1)}}.$$

Príklad (4.)

Letecká spoločnosť odhaduje priemerný počet cestujúcich. V priebehu 20 dní bol priemerný počet cestujúcich 112 s výberovým rozptylom 25.
Nájdime 95% obojstranný interval spoľahlivosti pre priemerný počet cestujúcich μ.

Príklad (4.)

Letecká spoločnosť odhaduje priemerný počet cestujúcich. V priebehu 20 dní bol priemerný počet cestujúcich 112 s výberovým rozptylom 25.
Nájdime 95% obojstranný interval spoľahlivosti pre priemerný počet cestujúcich μ.

Riešenie: $\overline{x} = 112$; $S_x^2 = 25$ (t. j.: $S_x = \sqrt{25} = 5$); $n = 20$.

Príklad (4.)

Letecká spoločnosť odhaduje priemerný počet cestujúcich. V priebehu 20 dní bol priemerný počet cestujúcich 112 s výberovým rozptylom 25.
Nájdime 95% obojstranný interval spoľahlivosti pre priemerný počet cestujúcich μ.

Riešenie: $\overline{x} = 112$; $\quad S_x^2 = 25$ (t. j.: $S_x = \sqrt{25} = 5$); $\quad n = 20$.

Súbor malý $(n = 20 < 30)$ $\Rightarrow$ Studentovo rozdelenie, kde $t_{1-\frac{5\%}{2}}^{(20-1)} = t_{0.975}^{(19)} = 2.1$, a platí

Príklad (4.)

Letecká spoločnosť odhaduje priemerný počet cestujúcich. V priebehu 20 dní bol priemerný počet cestujúcich 112 s výberovým rozptylom 25.
Nájdime 95% obojstranný interval spoľahlivosti pre priemerný počet cestujúcich μ.

Riešenie: $\overline{x} = 112$; $S_x^2 = 25$ (t. j.: $S_x = \sqrt{25} = 5$); $n = 20$.

Súbor malý ($n = 20 < 30$) $\Rightarrow$ Studentovo rozdelenie, kde $t_{1-\frac{5\%}{2}}^{(20-1)} = t_{0.975}^{(19)} = 2.1$, a platí

$$
\begin{aligned}
\overline{x} - \frac{S_x}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}^{(n-1)} &\leq \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}^{(n-1)}, \\
112 - \frac{5}{\sqrt{20}} \cdot 2.1 &\leq \mu \leq 112 + \frac{5}{\sqrt{20}} \cdot 2.1, \\
112 - 2.35 &\leq \mu \leq 112 + 2.35, \\
109.65 &\leq \mu \leq 114.35. \spadesuit
\end{aligned}
$$

Príklad (5.)

Z produkcie automatickej linky vyrábajúcej krúžky guľôčkových ložísk bola odobraná vzorka 50 kusov a zistený polomer krúžkov. Z nameraných hodnôt sme získali realizáciu priemeru $\overline{x} = 70.012\text{mm}$. *Nájdime 99-percentný interval spoľahlivosti pre strednú hodnotu polomeru vyrábaných krúžkov, ak z nameraných hodnôt sa realizácia štatistiky rovná* $S_x^2 = 0.00723$.

Príklad (5.)

Z produkcie automatickej linky vyrábajúcej krúžky guľôčkových ložísk bola odobraná vzorka 50 kusov a zistený polomer krúžkov. Z nameraných hodnôt sme získali realizáciu priemeru $\overline{x} = 70.012\text{mm}$. *Nájdime 99-percentný interval spoľahlivosti pre strednú hodnotu polomeru vyrábaných krúžkov, ak z nameraných hodnôt sa realizácia štatistiky rovná* $S_x^2 = 0.00723$.

Riešenie: $\overline{x} = 70.012$, $S_x^2 = 0.00723$ (t. j.: $S_x = \sqrt{0.00723} = 0.08503$), $n = 50$.

Príklad (5.)

Z produkcie automatickej linky vyrábajúcej krúžky guľôčkových ložísk bola odobraná vzorka 50 kusov a zistený polomer krúžkov. Z nameraných hodnôt sme získali realizáciu priemeru $\overline{x} = 70.012$mm. *Nájdime 99-percentný interval spoľahlivosti pre strednú hodnotu polomeru vyrábaných krúžkov, ak z nameraných hodnôt sa realizácia štatistiky rovná* $S_x^2 = 0.00723$.

Riešenie: $\overline{x} = 70.012$, $S_x^2 = 0.00723$ (t. j.: $S_x = \sqrt{0.00723} = 0.08503$), $n = 50$.
Nakoľko je súbor veľký $(n = 50 > 30)$ použijeme normálne normované rozdelenie, kde $u_{1-\frac{\alpha}{2}} = u =_{0.995} = 2.58$, a interval spoľahlivosti bude mať tvar

Príklad (5.)

Z produkcie automatickej linky vyrábajúcej krúžky guľôčkových ložísk bola odobraná vzorka 50 kusov a zistený polomer krúžkov. Z nameraných hodnôt sme získali realizáciu priemeru $\overline{x} = 70.012$mm*. Nájdime 99-percentný interval spoľahlivosti pre strednú hodnotu polomeru vyrábaných krúžkov, ak z nameraných hodnôt sa realizácia štatistiky rovná* $S_x^2 = 0.00723$.

Riešenie: $\overline{x} = 70.012$, $S_x^2 = 0.00723$ (t. j.: $S_x = \sqrt{0.00723} = 0.08503$), $n = 50$.
Nakoľko je súbor veľký $(n = 50 > 30)$ použijeme normálne normované rozdelenie, kde $u_{1-\frac{\alpha}{2}} = u =_{0.995}= 2.58$, a interval spoľahlivosti bude mať tvar

$$
\begin{aligned}
\overline{x} - \frac{S_x}{\sqrt{n}} \cdot u_{0.995} &\leq \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot u_{0.995}, \\
70.012 - \frac{0.08503}{\sqrt{50}} \cdot 2.58 &\leq \mu \leq 70.012 + \frac{0.08503}{\sqrt{50}} \cdot 2.58, \\
69.981 &\leq \mu \leq 70.043. \spadesuit
\end{aligned}
$$

χ^2-rozdelenie

Nech má náhodná premenná n náhodných premenných $x_1, x_2, \ldots, x_n$ a každá má normálne rozdelenie s priemerom μ a štandardnou odchýlkou σ. Týmto náhodným premenným zodpovedajú normované náhodné premenné:

$$Z_1 = \frac{x_1 - \mu}{\sigma}, \ldots, Z_n = \frac{x_n - \mu}{\sigma}.$$

Štatistika

$$W = Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_n^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^2}$$

má χ^2-rozdelenie s počtom stupňov voľnosti n.

χ^2-rozdelenie

χ^2-rozdelenie môžeme uplatniť i pre rozdelenie výberového rozptylu S_x^2. Ak náhodným výberom zo základného súboru s normálnym rozdelením vytvoríme výberové súbory veľkosti n, potom náhodná premenná

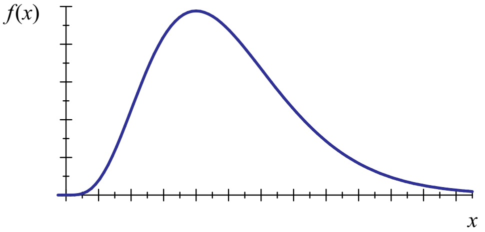
$$\chi^2 = \frac{(n-1)\cdot S_x^2}{\sigma^2}$$

má χ^2-rozdelenie s $n-1$ stupňami voľnosti.

Pre náhodnú premennú $\chi^2 = \dfrac{(n-1)\cdot S_x^2}{\sigma^2}$, modelujúcu rozptyl σ^2, môžeme so spoľahlivosťou $(1-\alpha)$ určiť interval spoľahlivosti:

$$\chi^2_{1-\frac{\alpha}{2}} \leq \chi^2 \leq \chi^2_{\frac{\alpha}{2}},$$

kde $\chi^2_{1-\frac{\alpha}{2}}$ a χ^2_1 , sú $\frac{\alpha}{2}$ a $\left(1-\frac{\alpha}{2}\right)$ **kritické hodnoty** χ^2-rozdelenie.



Je dôležité si uvedomiť, že χ^2-rozdelenie, naproti normálnemu rozdeleniu, je nesymetrické rozdelenie (viď. obrázok), a teda $\chi^2_{1-\frac{\alpha}{2}} \neq \chi^2_{\frac{\alpha}{2}}$.

Odhady parametrov - Interval spoľahlivosti pre rozptyl $D(\xi) = \sigma^2$

Hodnotu náhodnej premennej χ^2 teda nahradíme:

$$\chi^2_{1-\frac{\alpha}{2}} \leq \frac{(n-1)\, S_x^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}},$$

Hodnotu náhodnej premennej χ^2 teda nahradíme:

$$\chi^2_{1-\frac{\alpha}{2}} \leq \frac{(n-1)\,S_x^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}},$$

úpravou dostaneme **interval spoľahlivosti pre rozptyl** σ^2

$$\frac{(n-1)\,S_x^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)\,S_x^2}{\chi^2_{1-\frac{\alpha}{2}}},$$

Hodnotu náhodnej premennej χ^2 teda nahradíme:

$$\chi^2_{1-\frac{\alpha}{2}} \leq \frac{(n-1)\,S_x^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}},$$

úpravou dostaneme **interval spoľahlivosti pre rozptyl** σ^2

$$\frac{(n-1)\,S_x^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)\,S_x^2}{\chi^2_{1-\frac{\alpha}{2}}},$$

po odmocnení dostaneme **interval spoľahlivosti pre štandardnú odchýlku** σ

$$\sqrt{\frac{(n-1)\,S_x^2}{\chi^2_{\frac{\alpha}{2}}}} \leq \sigma \leq \sqrt{\frac{(n-1)\,S_x^2}{\chi^2_{1-\frac{\alpha}{2}}}}.$$

Príklad (6.)

Nech systematická chyba meracieho prístroja je nulová. Za rovnakých podmienok sa vykonalo desať nezávislých meraní jednej a tej istej veličiny μ, kde $\mu = 1000$m.

i :	1	2	3	4	5	6	7	8	9	10
x_i :	992	1010	1005	994	998	1000	1002	999	1000	997

Nájdime 90%-ný interval spoľahlivosti pre smerodajnú odchýlku σ.

Odhady parametrov - Interval spoľahlivosti pre rozptyl $D(\xi) = \sigma^2$

Príklad (6.)

Nech systematická chyba meracieho prístroja je nulová. Za rovnakých podmienok sa vykonalo desať nezávislých meraní jednej a tej istej veličiny μ, kde $\mu = 1000$m.

i :	1	2	3	4	5	6	7	8	9	10
x_i :	992	1010	1005	994	998	1000	1002	999	1000	997

Nájdime 90%-ný interval spoľahlivosti pre smerodajnú odchýlku σ.

Riešenie: Potrebujeme vypočítať hodnotu výberového rozptylu S_x^2

$$
\begin{aligned}
S_x^2 \quad = \quad & \frac{1}{9} \cdot \Big((992 - 1000)^2 + (1010 - 1000)^2 + (1005 - 1000)^2 \\
& + (998 - 1000)^2 + (1000 - 1000)^2 + (1002 - 1000)^2 \\
& + (999 - 1000)^2 + (1006 - 1000)^2 + (997 - 1000)^2 \Big) = 27.
\end{aligned}
$$

Odhady parametrov - Interval spoľahlivosti pre rozptyl $D(\xi) = \sigma^2$

Z tabuliek kvantilov χ^2-rozdelenie dostávame pre $\alpha = 0.1$ a $n = 10$ hodnotu

$$\chi^2_{1-\frac{\alpha}{2}}(9) = \chi^2_{0.95}(9) = 3.325, \qquad \chi^2_{\frac{\alpha}{2}}(9) = \chi^2_{0.05}(9) = 16.919.$$

Odhady parametrov - Interval spoľahlivosti pre rozptyl $D(\xi) = \sigma^2$

Z tabuliek kvantilov χ^2-rozdelenie dostávame pre $\alpha = 0.1$ a $n = 10$ hodnotu

$$\chi^2_{1-\frac{\alpha}{2}}(9) = \chi^2_{0.95}(9) = 3.325, \qquad \chi^2_{\frac{\alpha}{2}}(9) = \chi^2_{0.05}(9) = 16.919.$$

90%-ný interval spoľahlivosti pre smerodajnú odchýlku σ bude teda

$$\begin{aligned}
\sqrt{\frac{(n-1)\,S_x^2}{\chi^2_{\frac{\alpha}{2}}}} &\leq \sigma \leq \sqrt{\frac{(n-1)\,S_x^2}{\chi^2_{1-\frac{\alpha}{2}}}}, \\
\sqrt{\frac{(10-1)\,S_x^2}{\chi^2_{0.05}}} &\leq \sigma \leq \sqrt{\frac{(10-1)\,S_x^2}{\chi^2_{0.95}}}, \\
\sqrt{\frac{9 \cdot 27}{16.919}} &\leq \sigma \leq \sqrt{\frac{9 \cdot 27}{3.325}}, \\
\sqrt{14.363} &\leq \sigma \leq \sqrt{73.08}, \\
3.790 &\leq \sigma \leq 8.549. \spadesuit
\end{aligned}$$

Zhrnutie I.

Najlepším neskresleným *bodovým odhadom str. hodnoty* $E(\xi)$ je **výberový priemer**

$$\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i,$$

najlepším neskresleným *bodovým odhadom rozptylu* $D(\xi)$ je **výberový rozptyl**

$$S_x^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2 .$$

resp. pre skupinové rozdelenie pravdepodobnosti sú tieto odhady tvaru

$$\overline{x} = \frac{1}{N} \cdot \sum_{j=1}^{k} n_j \cdot x_j,$$

$$S_x^2 = \frac{1}{N-1} \cdot \sum_{j=1}^{k} n_j \cdot (x_j - \overline{x})^2 .$$

Zhrnutie II.

Interval spoľahlivosti pre strednú hodnotu $E(\xi) = \mu$

$$\overline{x} - \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}} \quad \leq \quad \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot u_{1-\frac{\alpha}{2}}, \qquad \text{pre } n \geq 30,$$

$$\overline{x} - \frac{S_x}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}^{(n-1)} \quad \leq \quad \mu \leq \overline{x} + \frac{S_x}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}^{(n-1)}, \qquad \text{pre } n < 30,$$

kde $u_{1-\frac{\alpha}{2}}$ sú kvantily $N(0,1)$ rozdelenia a $t_{1-\frac{\alpha}{2}}^{(n-1)}$ kvantily Studentovho t-rozdelenia.

Interval spoľahlivosti pre štandardný rozptyl $D(\xi) = \sigma^2$

$$\frac{(n-1)\, S_x^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)\, S_x^2}{\chi^2_{1-\frac{\alpha}{2}}},$$

kde $\chi^2_{\frac{\alpha}{2}}$ resp. $\chi^2_{1-\frac{\alpha}{2}}$ sú kritické hodnoty χ^2-rozdelenia.