Odhady středních hodnot

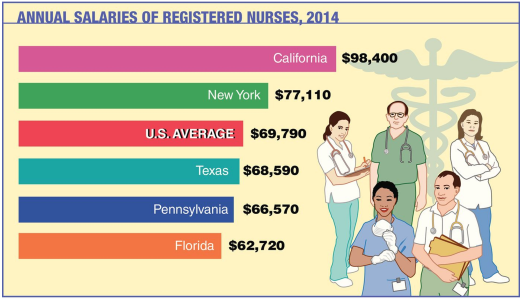
Povinná literatura: Mann (2016), Kapitola 8

Z čeho studovat osmou lekci?

Povinná literatura:	Mann (2016), kap. 08
Příprava na cvičení:	Leaflet 09 Koncepty a procedury, cv. 09, kap. 08
Příprava na zkoušku:	Mann (2016), kap. 08 Leaflet 09 Sbírka úloh, kap. 08 Koncepty a procedury, cv. 09, kap. 08

Obsah

TÉMA: Jaké jsou roční platy zdravotních sester v USA?



Data source: U.S. Bureau of Labor Statistics, May 2014

Průměrné výdělky zdravotních sester v USA

Podle zprávy *Occupational Employment and Wages* za květen 2014 od amerického úřadu pro statistiku práce činil průměrný roční výdělek registrovaných zdravotních sester ve Spojených státech v roce 2014 **69 790 dolarů**.

Průměrné výdělky se výrazně lišily mezi jednotlivými státy.

- V Kalifornii činil průměrný výdělek **98 400 dolarů**.
- V Jižní Dakotě to bylo pouze **53 970 dolarů** (not in graf).

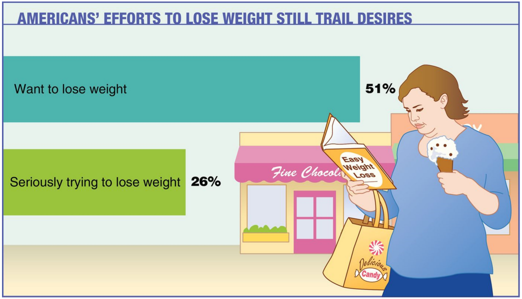
Regionální rozdíly

Průměrné výdělky registrovaných sester se také značně lišily mezi různými metropolitními oblastmi:

- V oblasti San Francisco-San Mateo-Redwood City (Kalifornie) činil průměrný výdělek **128 190 dolarů**.
- V oblasti Los Angeles-Long Beach-Glendale to bylo **94 140 dolarů**.
- V oblasti Greenville, Severní Karolína, to bylo **63 130 dolarů**.

Otázka: Jak mohou statistické metody, jako je výpočet intervalu spolehlivosti, pomoci odhadnout skutečný průměrný výdělek registrovaných zdravotních sester v různých státech USA, pokud máme k dispozici pouze vzorková data?

TÉMA: Zaostává úsilí Američanů o hubnutí za jejich přáními zhubnout?



Data source: www.gallup.com

Průzkum agentury Gallup

Agentura Gallup provedla průzkum mezi 828 americkými dospělými ve věku 18 a více let na téma hmotnosti. Tito dospělí byli dotázáni:

- zda chtějí zhubnout,
- zda vážně usilují o hubnutí.

Výsledky:

- 51 % dospělých uvedlo, že chtějí zhubnout,
- 26 % uvedlo, že se vážně snaží zhubnout.

Průzkum byl proveden od 6. do 9. listopadu 2014.

Otázka: Jak nám výpočet intervalu spolehlivosti může pomoci lépe porozumět rozdílu mezi tím, kolik lidí chce zhubnout, a kolik lidí se skutečně snaží vážně zhubnout?

Případové studie

Odhad, bodový a intervalový odhad

Odhad střední hodnoty populace: σ je známo

Odhad střední hodnoty populace: σ je neznámo

Odhad podílu populace

Odhad

Přiřazení hodnot(y) parametru populace na základě hodnoty odpovídající statistiky vzorku se nazývá **odhad**.

Statistická veličina použitá k odhadu parametru populace se nazývá **estimátor**.

Postup odhadu zahrnuje následující kroky:

1. Vyberte vzorek.
2. Sbírejte potřebné informace od členů vzorku.
3. Vypočítejte hodnotu statistiky vzorku.
4. Přiřaďte hodnoty odpovídajícímu parametru populace.

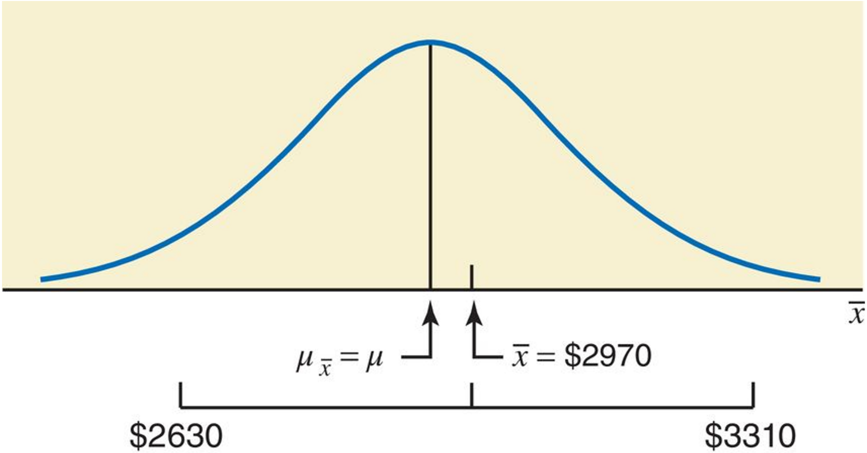
Bodový a intervalový odhad

Hodnota statistiky vzorku, která se používá k odhadu parametru populace, se nazývá **bodový odhad**.

V **intervalovém odhadu** je kolem bodového odhadu konstruován interval, a je deklarováno, že tento interval obsahuje odpovídající parametr populace s určitou mírou spolehlivosti.

Číslo, které přičteme a odečteme od bodového odhadu, se nazývá **chybové rozpětí** (margin of error) nebo maximální chyba odhadu. Následující obrázek ilustruje koncept intervalového odhadu.

Intervalový odhad



Intervalový odhad

Každý interval je konstruován s ohledem na určitou **hladinu spolehlivosti** a je nazýván **interval spolehlivosti**.

Interval spolehlivosti je vyjádřen jako:
Bodový odhad $\pm$ **Chybové rozpětí**

Hladina spolehlivosti je spojená s intervalovým odhadem a udává, jak velkou máme jistotu, že tento interval obsahuje skutečný parametr populace. Hladina spolehlivosti je vyjádřena $(1-\alpha)100\%$.

Obsah

Odhad střední hodnoty populace: σ je známo

Tři možné případy

Případ I. Pokud jsou splněny následující tři podmínky:

1. Směrodatná odchylka populace σ je známa.
2. Velikost vzorku je malá (tj. $n < 30$).
3. Populace, ze které byl vzorek vybrán, je přibližně normálně rozdělená.

Případ II. Pokud jsou splněny následující dvě podmínky:

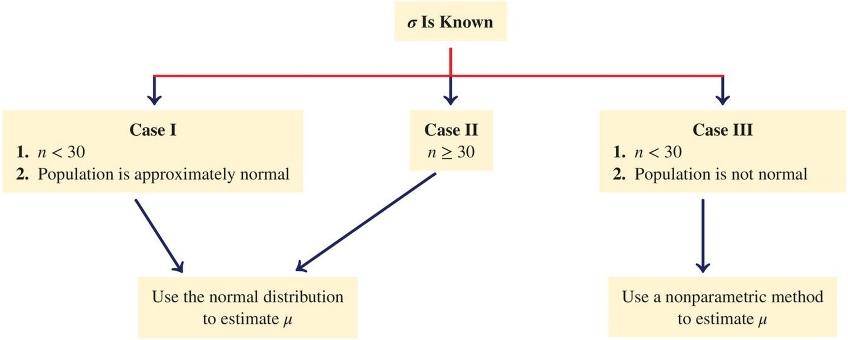
1. Směrodatná odchylka populace σ je známa.
2. Velikost vzorku je velká (tj. $n \geq 30$).

Případ III. Pokud jsou splněny následující tři podmínky:

1. Směrodatná odchylka populace σ je známa.
2. Velikost vzorku je malá (tj. $n < 30$).
3. Populace, ze které byl vzorek vybrán, není normálně rozdělená (nebo její rozdělení není známo).

Odhad střední hodnoty populace: σ je známo

Tři možné případy



Odhad střední hodnoty populace: σ je známo

Interval spolehlivosti pro μ

Interval spolehlivosti $(1-\alpha)100\%$ pro μ v případech I a II je

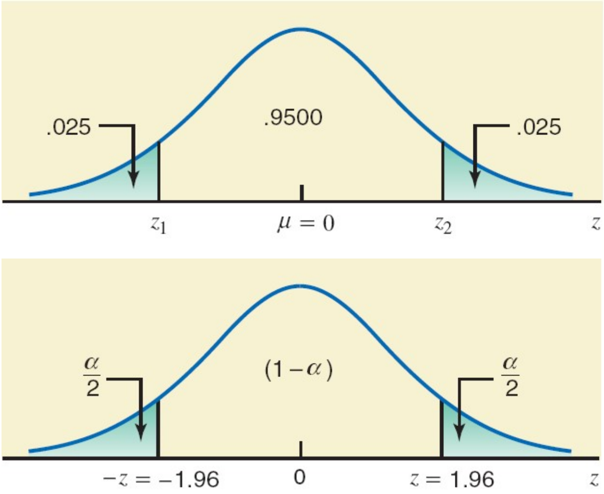
$$\bar{x} \pm z\sigma_{\bar{x}}, \quad \text{kde} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Použitá hodnota z je získána z tabulky standardizovaného normálního rozdělení (viz například Tabulka IV z Přílohy B) pro danou hladinu spolehlivosti, případně využití vhodných funkcí v Excelu.

Chybové rozpětí odhadu pro μ, označované E, je veličina, která je odečítána a přičítána k hodnotě $\bar{x}$ k získání intervalu spolehlivosti pro μ. Tedy,

$$E = z\sigma_{\bar{x}}$$

z-hodnoty pro 95% interval spolehlivosti a plocha v chvostech



z-hodnoty (kvantily) pro běžně používané úrovně spolehlivosti

Confidence Level	Areas to Look for in Table IV	z Value
90%	.0500 and .9500	1.64 or 1.65
95%	.0250 and .9750	1.96
96%	.0200 and .9800	2.05
97%	.0150 and .9850	2.17
98%	.0100 and .9900	2.33
99%	.0050 and .9950	2.57 or 2.58

Příklad 1

Zadání: Vydavatelství právě publikovalo novou vysokoškolskou učebnici. Než se společnost rozhodne za jakou cenu učebnici prodávat, chce znát průměrnou cenu všech obdobných učebnic na trhu.

Výzkumné oddělení společnosti vzalo vzorek 25 srovnatelných učebnic a shromáždilo informace o jejich cenách. Tyto informace poskytly průměrnou cenu 145 dolarů pro tento vzorek. Je známo, že směrodatná odchylka cen všech takových učebnic je 35 dolarů a populace těchto cen je normálně rozdělena.

a) Jaký je bodový odhad průměrné ceny všech takových učebnic?
b) Sestavte 90% interval spolehlivosti pro průměrnou cenu všech takových vysokoškolských učebnic.

Příklad 1: Řešení

Řešení (a): $n = 25, \bar{x} = \$145,$ a $\sigma = \$35$

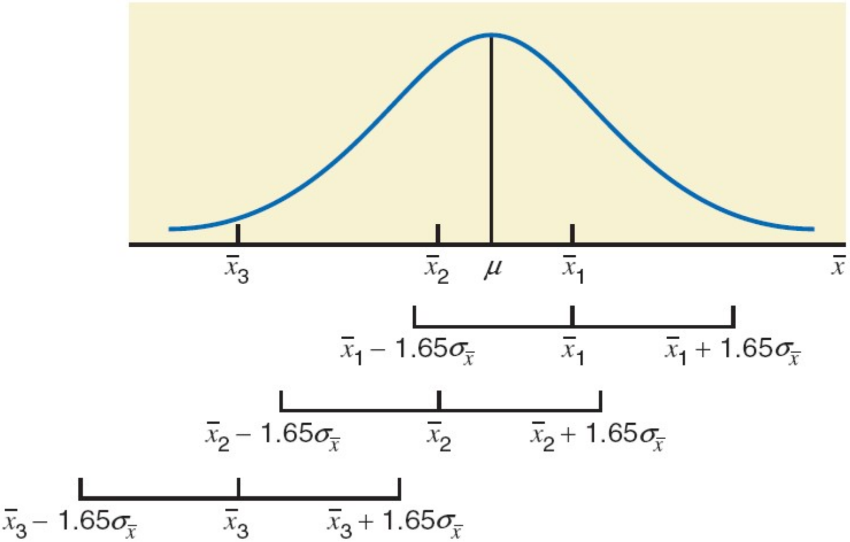
Bodový odhad pro μ je $\bar{x} = \$145$

Řešení (b): Hladina spolehlivosti je 90 % nebo 0.90.
Díky tomu víme, že oblast v každém konci normální distribuční křivky je $\alpha/2 = (1 - 0.90)/2 = 0.05$. Prvně nalezneme kvantil $z = 1.65$ a dále dopočítáme $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{25}} = \$7.00$.

$$\begin{aligned}
\bar{x} \pm z\sigma_{\bar{x}} &= 145 \pm 1.65(7.00) \\
&= 145 \pm 11.55 \\
&= (145 - 11.55) \text{ do } (145 + 11.55) \\
&= \mathbf{\$133.45} \text{ do } \mathbf{\$156.55}
\end{aligned}$$

Můžeme říci, že máme 90% jistotu, že průměrná cena všech takových vysokoškolských učebnic je mezi 133.45 a 156.55 dolarů.

Intervaly spolehlivosti získané ze tří různých vzorků



Příklad 2

Zadání: Podle studie společnosti Moebs Services Inc. z roku 2013 náklady na individuální běžné účty u předních amerických bank převyšují 380 dolarů na účet ročně.

Nedávný náhodně provedený výběr 600 takových běžných účtů ukázal průměrnou roční nákladovost 500 dolarů pro tyto banky.

Za předpokladu, že směrodatná odchylka ročních nákladů pro tyto banky ze všech běžných účtů je 40 dolarů, sestavte 99% interval spolehlivosti pro aktuální průměrné roční náklady předních amerických bank na tyto běžné účty.

Příklad 2: Řešení

- Míra spolehlivosti 99 % nebo 0.99
- Velikost vzorku je velká ($n \geq 30$)
- Proto používáme normální rozdělení $\Rightarrow$ z = 2.58

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{600}} = 1.63299316$$

$$\begin{aligned} \bar{x} \pm z\sigma_{\bar{x}} &= 500 \pm 2.58(1.63299316) \\ &= 500 \pm 4.21 \\ &= \mathbf{\$495.79} \text{ až } \mathbf{\$504.21} \end{aligned}$$

Můžeme tedy s 99% jistotou říci, že současné průměrné roční náklady hlavních amerických bank na všechny individuální běžné účty jsou mezi \$495.79 a \$504.21.

Šířka intervalu spolehlivosti

Šířka intervalu spolehlivosti závisí na velikosti **chybového rozpětí** $Z\sigma_{\bar{x}}$, které závisí na hodnotách z, σ a n, protože $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Hodnota σ však **není** pod kontrolou vyšetřovatele.

Proto lze šířku intervalu spolehlivosti kontrolovat pomocí

1. hodnoty z, která závisí na hladině spolehlivosti,
2. velikosti vzorku, n.

Určení velikosti vzorku pro odhad μ

Vzhledem k hladině spolehlivosti a směrodatné odchylce populace je velikost vzorku, která vyprodukuje předem stanovenou chybu rozpětí E u konfidenčního intervalu odhadu μ, dána vzorcem:

$$n = \frac{z^2 \sigma^2}{E^2},$$

kde z je hodnota kvantilu z odpovídající zadané hladině spolehlivosti, σ je směrodatná odchylka populace a E je požadovaná chyba rozpětí.

Příklad 3

Zadání: Asociace absolventů chce odhadnout průměrný dluh letošních vysokoškolských absolventů. Je známo, že směrodatná odchylka dluhů letošních absolventů je 11 800 dolarů. Jak velký vzorek by měl být vybrán, aby odhad s 99% hladinou spolehlivosti byl v rozmezí 800 dolarů od průměru populace?

Řešení:

- Maximální velikost chybového rozpětí odhadu má být \$800; to znamená, $E = \pm\$800$.
- Hodnota kvantilu z pro 99% hladinu spolehlivosti je $z = 2.58$.
- Hodnota σ je udána jako \$11 800.

$$n = \frac{z^2\sigma^2}{E^2} = \frac{(2.58)^2(11800)^2}{(800)^2} = 1448.18 \approx 1449$$

- Tedy minimální požadovaná velikost vzorku je 1449.

Obsah

Odhad střední hodnoty populace: σ je neznámo

Tři možné případy

Případ I. Pokud jsou splněny následující tři podmínky:

1. Směrodatná odchylka populace σ je neznáma.
2. Velikost vzorku je malá (tj. $n < 30$).
3. Populace, ze které byl vzorek vybrán, je přibližně normálně rozdělená.

Případ II. Pokud jsou splněny následující dvě podmínky:

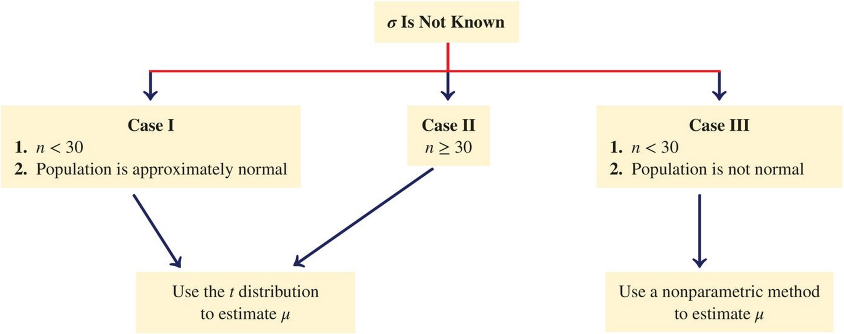
1. Směrodatná odchylka populace σ je neznáma.
2. Velikost vzorku je velká (tj. $n \geq 30$).

Případ III. Pokud jsou splněny následující tři podmínky:

1. Směrodatná odchylka populace σ je neznáma.
2. Velikost vzorku je malá (tj. $n < 30$).
3. Populace, ze které byl vzorek vybrán, není normálně rozdělená (nebo její rozdělení není známo).

Odhad střední hodnoty populace: σ je neznámo

Tři možné případy



Studentovo t-rozdělení

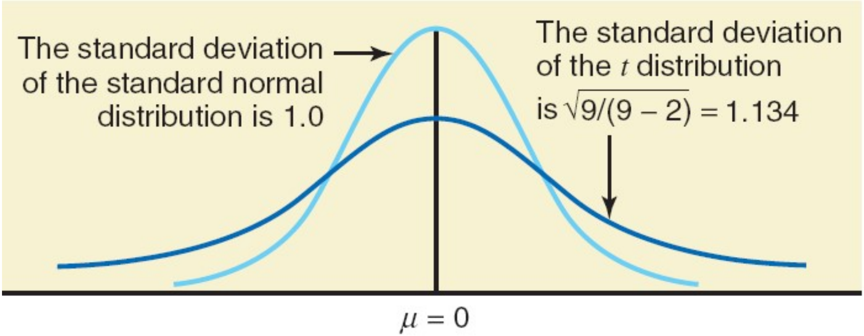
Studentovo t-rozdělení je specifický typ zvonovitého rozdělení s nižší výškou a větším rozpětím než standardní normální rozdělení. Jak velikost vzorku narůstá, t-rozdělení se blíží standardizovanému normálnímu rozdělení. Studentovo t-rozdělení má pouze jeden parametr, nazývaný stupně volnosti (df).

$$df = n - 1$$

Střední hodnota t-rozdělení je rovna 0 a jeho směrodatná odchylka je

$$\sqrt{\frac{df}{(df - 2)}}.$$

Studentovo t-rozdělení



Příklad 4

Zadání: Najděte hodnotu t pro 16 stupňů volnosti a plochu 0.05 v pravé části t-distribuční křivky.

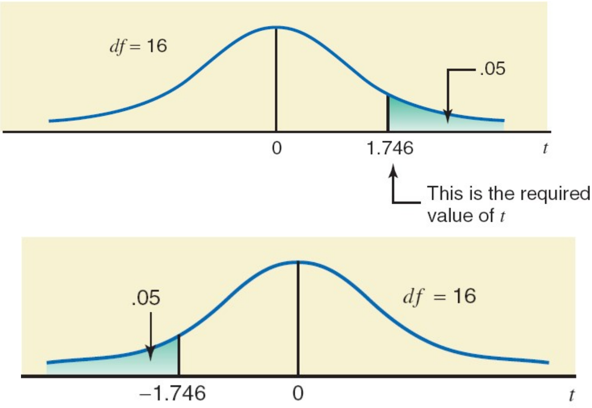
Řešení:

Area in the right tail

df	Area in the Right Tail Under the t Distribution Curve				
	.10	$\boxed{.05}$	.025	...	.001
1	3.078	6.314	12.706	...	318.309
2	1.886	2.920	4.303	...	22.327
3	1.638	2.353	3.182	...	10.215
.	...	...	...	...	...
.	...	...	...	...	...
.	...	...	...	...	...
$df \rightarrow$ $\boxed{16}$	1.337	$\boxed{1.746}$	2.120	...	3.686
.	...	...	...	...	...
.	...	...	...	...	...
.	...	...	...	...	...
75	1.293	1.665	1.992	...	3.202
∞	1.282	1.645	1.960	...	3.090

The required value of t for 16 df and .05 area in the right tail

Hodnotu t pro 16 stupňů volnosti a plochu 0.05 v pravé/levé části t-distribuční křivky



Interval spolehlivosti pro μ s využitím studentova t-rozdělení

$(1-\alpha)100\%$ **interval spolehlivosti pro μ** je

$$\bar{x} \pm t s_{\bar{x}}, \qquad \text{kde} \qquad s_{\bar{x}} = \frac{s}{\sqrt{n}}.$$

Hodnota t je získána z t-distribuční tabulky pro $n-1$ stupňů volnosti a danou hladinu spolehlivosti. Zde $t s_{\bar{x}}$ je chybové rozpětí odhadu; to znamená, že v tomto případě

$$E = t s_{\bar{x}}.$$

Příklad 5

Zadání: Náhodný vzorek 25 zaměstnanců z New Yorku, kteří měli zdravotní pojištění poskytnuté zaměstnavatelem, platil průměrný příspěvek 6600 dolarů za rodinné zdravotní pojištění se směrodatnou odchylkou 800 dolarů.

Sestrojte 95% interval spolehlivosti pro aktuální průměrnou pojistnou částku platby za rodinné zdravotní pojištění, kterou platí všichni pracovníci v New York City. Předpokládejte, že rozdělení pojistných částek zaplacených za rodinné zdravotní pojištění všemi pracovníky v New York City je přibližně normálně rozděleno.

Příklad 5: Řešení

- Směrodatná odchylka populace není známá, $n < 30$, a populace je normálně rozdělená (Případ I). Použijeme studentovo t-rozdělení k vytvoření intervalu spolehlivosti pro μ
- $n = 25$, $\bar{x} = \$6600$, $s = \$800$, hladina spolehlivosti = 95%
- Stupně volnosti $df = n - 1 = 25 - 1 = 24$
- Plocha v každém konci = .5 - (.95/2) = .5 - .4750 = .025
- Hodnota t v pravém konci je 2.064
- $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{800}{\sqrt{25}} = \$160$
- $\bar{x} \pm t s_{\bar{x}} = 6600 \pm 2.064(160) = 6600 \pm 330.24 =$ **\$6269.76** až **\$6930.24**

Můžeme s 95% jistotou říci, že aktuální průměrná pojistná částka platby za rodinné zdravotní pojištění, kterou platí všichni pracovníci v New York City, je mezi 6269.76 dolaru a 6930.24 dolaru.

Příklad 6

Zadání: Šedesát čtyři náhodně vybraných dospělých, kteří si kupují knihy ke čtení, bylo dotázáno, kolik obvykle ročně utratí za knihy.

Vzorek ukázal průměrné roční výdaje 1450 dolarů a směrodatnou odchylku 300 dolarů.

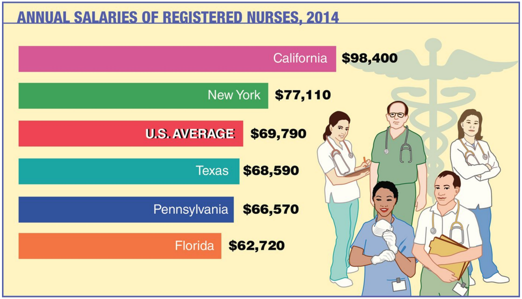
Určete 99% interval spolehlivosti pro průměr výdajů v odpovídající populaci.

Příklad 6: Řešení

- Směrodatná odchylka σ není známá, ale velikost vzorku je $n > 30$ (Případ II). Použijeme studentovo t-rozdělení k vytvoření intervalu spolehlivosti pro μ
- $n = 64$, $\bar{x} = 1450$, $s = 300$, hladina spolehlivosti je 99 %
- Stupně volnosti $df = n - 1 = 64 - 1 = 63$
- Plocha v každém konci = .5 - (.99/2) = .5 - .4950 = .005
- Hodnota t v pravém konci je 2.656
- $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{300}{\sqrt{64}} = 37,50$
- $\bar{x} \pm ts_{\bar{x}} = 1450 \pm 2.656(37.50) = 1450 \pm 99.60 =$ **\$1350.40** až **\$1549.60**

Můžeme s 99% jistotou tvrdit, že na základě tohoto vzorku je průměrný roční výdaj na knihy od všech dospělých, kteří si kupují knihy, mezi 1350.40 dolarů a 1549.60 dolarů.

TÉMA: Jaké jsou roční platy registrovaných zdravotních sester?



Data source: U.S. Bureau of Labor Statistics, May 2014

TÉMA: Jaké jsou roční platy registrovaných zdravotních sester?

Výpočet intervalu spolehlivosti
Pokud známe velikost vzorku a směrodatnou odchylku populace, můžeme spočítat interval spolehlivosti pro průměrné výdělky.

Chceme najít 98% interval spolehlivosti pro průměrné výdělky všech registrovaných sester v **Texasu.**

Pokud jsou průměrné výdělky v Texasu založeny na vzorku 1600 sester a směrodatná odchylka populace je 6 240 dolarů, pak je interval vypočítán takto:

$$X \pm z\frac{\sigma}{\sqrt{n}} = 68590 \pm 2.33 \times 156.00 = \mathbf{68\,226.52} \text{ až } \mathbf{68\,953.48}$$

S 98% spolehlivostí můžeme říci, že průměrný výdělek všech sester v Texasu v roce 2014 byl mezi **68 226.52** a **68 953.48 dolary**.

Interval spolehlivosti pro μ s využitím t-rozdělení

Co když je velikost vzorku příliš velká?

1. Použijte hodnotu t z posledního řádku (řádek ∞) v Tabulce V.
2. Použijte normální distribuci jako aproximaci t-distribuce.

Obsah

Podíl populace $\hat{p}$ a estimátor jeho směrodatné odchylky

Odhad podílu populace p získaný ze vzorku značíme $\hat{p}$.

Hodnota $s_{\hat{p}}$, která poskytuje bodový odhad $\sigma_{\hat{p}}$, je vypočítána následovně

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Zde $s_{\hat{p}}$ je **estimátorem** $\sigma_{\hat{p}}$.

Interval spolehlivosti pro podíl populace, p,

Pro velký vzorek je $(1-\alpha)100\%$ **interval spolehlivosti pro podíl populace** p:

$$\hat{p} \pm z s_{\hat{p}}.$$

Hodnota z je získána z tabulky standardizovaného normálního rozdělení pro danou hladinu spolehlivosti, a

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Chybové rozpětí $z s_{\hat{p}}$ opět označujeme E.

Příklad 7

Zadání: Organizace PolicyInteractive z Eugene v Oregonu provedla v dubnu 2014 studii zkoumající naplnění amerického snu.

Studie zahrnovala vzorek 1821 amerických dospělých. Sedmdesát pět procent lidí zahrnutých do této studie uvedlo, že mít pokryté základní potřeby je velmi nebo extrémně důležité pro jejich vizi amerického snu.

(a) Jaký je bodový odhad podílu populace?

(b) Stanovte s 99% hladinou spolehlivosti procento všech amerických dospělých, kteří uvedou, že mít pokryté základní potřeby je pro jejich vizi amerického snu velmi nebo extrémně důležité. Jaké je chybové rozpětí tohoto odhadu?

Příklad 7: Řešení

- Vzorek má velikost 1821, odhadovaný podíl $\hat{p}$ je 0.75
- Směrodatná odchylka odhadovaného podílu $\hat{p}$, je vypočítána jako

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(.75)(.25)}{1821}} = .010147187$$

- Poznamenejme, že $n\hat{p}$ a $n(1-\hat{p})$ jsou obě větší než 5.
- **Řešení (a):** Bodový odhad $p = \hat{p} = .75$
- **Řešení (b):** Hladina spolehlivosti je 99 %, nebo .99, $z = 2.58$.

$$\begin{aligned} \mathrm{IS} &= \hat{p} \pm z s_{\hat{p}} = .75 \pm 2.58(.010147187) = .75 \pm .026 \\ &= \mathbf{.724} \text{ až } \mathbf{.776} \text{ nebo } \mathbf{72.4\,\%} \text{ až } \mathbf{77.6\,\%} \end{aligned}$$

$$\begin{aligned} \text{Chybové rozpětí} &= \pm 2.58 s_{\hat{p}} = \pm 2.58(.010147187) \\ &= \pm\, \mathbf{.026} \text{ nebo } \pm\, \mathbf{2.6\,\%} \end{aligned}$$

Příklad 8

Zadání: Podle studie o absolventech vysokých škol, provedené od 4. února do 7. března 2014, uvedlo 63 % dotázaných absolventů vysokých škol, že měli alespoň jednoho profesora, který je nadchnul pro studium.

Předpokládejme, že tato studie byla založena na náhodném vzorku 2000 absolventů.

Sestavte 97% interval spolehlivosti pro odpovídající podíl populace.

Příklad 8: Řešení

- Hladina spolehlivosti = 97 %
- Směrodatná chyba $s_{\hat{p}}$ se rovná

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(.63)(.37)}{2000}} = .01079583$$

- Příslušná hodnota z je 2.17.

$$\begin{aligned}\hat{p} \pm zS_{\hat{p}} &= .63 \pm 2.17(.01079583) \\ &= .63 \pm .023 \\ &= .607 \text{ do } .653 \text{ nebo } 60.7\,\% \text{ do } 65.3\,\%\end{aligned}$$

Můžeme s 97% jistotou tvrdit, že podíl všech absolventů vysokých škol, kteří by řekli, že měli alespoň jednoho profesora, který je nadchnul pro studium je mezi **.607** a **.653**. Tento interval spolehlivosti lze převést na procentní interval jako **60.7 %** až **65.3 %**.

Stanovení velikosti vzorku pro odhad podílu

Vzhledem k hladině spolehlivosti a hodnotám $\hat{p}$ a $\hat{q} = 1 - \hat{p}$, velikost vzorku, která vyprodukuje předem stanovenou maximální chybu E intervalu spolehlivosti pro **odhad podílu** p, je

$$n = \frac{z^2 \hat{p}\hat{q}}{E^2}$$

V případě, že hodnoty $\hat{p}$ a $\hat{q}$ nejsou známy:

1. Použijeme nejkonzervativnější odhad velikosti vzorku n s použitím $\hat{p} = .5$ a $\hat{q} = .5$.
2. Vezmeme předběžný vzorek (o libovolně určené velikosti) a spočítáme $\hat{p}$ a $\hat{q}$ z tohoto vzorku. Pak použijeme tyto hodnoty k nalezení n.

Příklad 9

Zadání: Společnost právě nainstalovala nový stroj, který vyrábí součástky používané v hodinách. Společnost chce odhadnout podíl součástek vyrobených tímto strojem, které jsou vadné.

(a) Manažer společnosti chce, aby byl odhad v rozmezí .02 od podílu populace s 95% hladinou spolehlivosti. Jaký je nejkonzervativnější odhad velikosti vzorku, který omezí maximální chybu na rozmezí ± 0.02 od podílu populace?

(b) Předpokládejme, že předběžný vzorek 200 dílů vyrobených tímto strojem ukázal, že 7 % z nich je vadných. Jak velký vzorek by měla společnost vybrat, aby byl 95% interval spolehlivosti pro p v rozmezí ± 0.02 populace podílu?

Příklad 9

Řešení (a): Hodnota z je pro 95% hladinu spolehlivosti 1.96.

$$\hat{p} = .50 \quad \text{a} \quad \hat{q} = .50$$

$$n = \frac{z^2 \hat{p}\hat{q}}{E^2} = \frac{(1.96)^2(.50)(.50)}{(.02)^2} = 2401$$

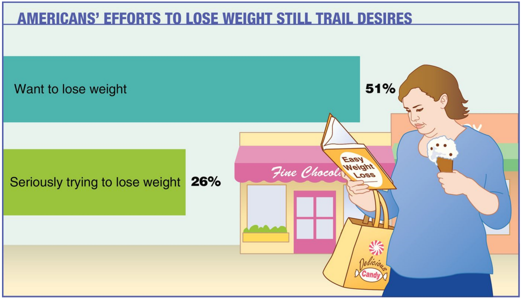
Pokud společnost vezme náhodný vzorek 2401 dílů, existuje 95% šance, že odhad p bude v rozmezí .02 od podílu populace.

Řešení (b):

$$\hat{p} = 0.07 \quad \text{a} \quad \hat{q} = 0.93$$

$$n = \frac{z^2 \hat{p}\hat{q}}{E^2} = \frac{(1.96)^2(0.07)(0.93)}{(0.02)^2} \approx 625.22 \approx 626$$

TÉMA: Zaostává úsilí Američanů o hubnutí za jejich přáními zhubnout?



Data source: www.gallup.com

Intervaly spolehlivosti

Procenta z průzkumu jsou založena na vzorku. Pomocí metody naučené v této sekci můžeme vypočítat intervaly spolehlivosti pro podíly populace:

Kategorie: Dospělí,	**Podíl vzorku**	**Interval spolehlivosti**
kteří chtějí zhubnout	0.51	$0.51 \pm zs_p$
kteří vážně usilují o hubnutí	0.26	$0.26 \pm zs_p$

Příklad: Chceme najít 96% interval spolehlivosti pro podíl dospělých, kteří chtějí zhubnout.

Výpočet intervalu spolehlivosti

Výpočet:

$$Sp = \sqrt{\frac{(0.51)(0.49)}{828}} = 0.01737$$

$$p \pm zSp = 0.51 \pm 2.05 \times 0.01737 = 0.51 \pm 0.036$$

$$\text{Interval} = 0.474 \text{ až } 0.546$$

S 96% spolehlivostí můžeme říci, že 47.4 % až 54.6 % všech dospělých uvede, že chtějí zhubnout. Chyba odhadu je 3.6 %.

Stejným způsobem lze spočítat interval pro dospělé, kteří vážně usilují o hubnutí.

Shrnutí přednášky:

Případové studie

Odhad, bodový a intervalový odhad

Odhad střední hodnoty populace: σ je známo

Odhad střední hodnoty populace: σ je neznámo

Odhad podílu populace

Co si nastudovat na příští týden?

Příprava na cvičení: Leaflet 09
Koncepty a procedury, cv. 09, kap. 08

Povinná literatura: Mann (2016), Kapitola 9

Děkuji za pozornost!

Source: Pinterest