MUNI ECON

Výběrové rozdělení

Povinná literatura: Mann (2016), Kapitola 7

Z čeho studovat sedmou lekci?

Povinná literatura: Mann (2016), kap. 07

Příprava na cvičení: Leaflet 08
Koncepty a procedury, cv. 08, kap. 07

Příprava na zkoušku: Mann (2016), kap. 07
Leaflet 08
Sbírka úloh, kap. 07
Koncepty a procedury, cv. 08, kap. 07

Motivační vstup

Chápání výběrových rozdělení a statistických charakteristik, jako je střední hodnota a směrodatná odchylka, nám umožňuje porozumět chování populací na základě malých vzorků dat.

Tento přístup je nepostradatelný při analýze dat v oblastech jako ekonomie, marketing, zdravotnictví nebo finance.

Pochopení těchto základních konceptů vám nejen pomůže správně interpretovat výsledky průzkumů a experimentů, ale také se vyhnout častým chybám a zkreslením v datech.

Obsah

Výběrové rozdělení
- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$
- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku
- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\hat{p}$

Pravděpodobnostní rozdělení populace a výběru

Pravděpodobnostní rozdělení populace je pravděpodobnostní rozdělení dat celé populace.

Obecně se pravděpodobnostní rozdělení statistického údaje získaného ze vzorku dat nazývá **výběrové rozdělení**.

Pravděpodobnostní rozdělení výběrového průměru $\bar{x}$ je získané ze všech možných vzorků stejné velikosti vybraných z populace. Toto pravděpodobnostní rozdělení $\bar{x}$ udává různé hodnoty, které může $\bar{x}$ nabývat a také pravděpodobnost realizace každé z těchto hodnot.

Pravděpodobnostní rozdělení populace

Ukázka: Předpokládejme, že ve třídě pokročilé statistiky je pouze pět studentů a výsledky jejich půlsemestrálních zkoušek jsou následující: 70, 78, 80, 80, 95. Nechť x označuje skóre studenta.

Četnost populace a relativní četnostní rozdělení (vlevo) a pravděpodobnostní rozdělení populace (vpravo)

x	f	Relative Frequency
70	1	$1/5 = .20$
78	1	$1/5 = .20$
80	2	$2/5 = .40$
95	1	$1/5 = .20$
	$N = 5$	$\text{Sum} = 1.00$

x	$P(x)$
70	.20
78	.20
80	.40
95	.20
	$\Sigma P(x) = 1.00$

Výběrové rozdělení

Znovu se podívejme na soubor půlsemestrálních výsledků pěti studentů. Uvažujme všechny možné vzorky tří výsledků z této skupiny, které lze vybrat bez vracení.

Celkový počet možných výběrů je

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10.$$

Předpokládejme, že přiřadíme písmena A, B, C, D a E k výsledkům pěti studentů tak, že: A = 70, B = 78, C = 80, D = 80, E = 95.

Pak deset možných výběru tří výsledků jsou:

ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE

Všechny možné výběry a jejich průměry, když je velikost vzorku 3

Sample	Scores in the Sample	$\bar{x}$
ABC	70, 78, 80	76.00
ABD	70, 78, 80	76.00
ABE	70, 78, 95	81.00
ACD	70, 80, 80	76.67
ACE	70, 80, 95	81.67
ADE	70, 80, 95	81.67
BCD	78, 80, 80	79.33
BCE	78, 80, 95	84.33
BDE	78, 80, 95	84.33
CDE	80, 80, 95	85.00

Pravděpodobnostní rozdělení výběru z populace

Četnost výběru a relativní četnostní rozdělení $\bar{x}$ (vlevo) a pravděpodobnostní rozdělení $\bar{x}$ (vpravo)

$\bar{x}$	f	Relative Frequency
76.00	2	2/10 = .20
76.67	1	1/10 = .10
79.33	1	1/10 = .10
81.00	1	1/10 = .10
81.67	2	2/10 = .20
84.33	2	2/10 = .20
85.00	1	1/10 = .10
	$\Sigma f = 10$	Sum = 1.00

$\bar{x}$	$P(\bar{x})$
76.00	.20
76.67	.10
79.33	.10
81.00	.10
81.67	.20
84.33	.20
85.00	.10
	$\Sigma P(\bar{x}) = 1.00$

Výběrová chyba a Chyby nezpůsobené výběrem

Výběrová chyba je rozdíl mezi hodnotou statistického údaje získaného ze vzorku a hodnotou odpovídajícího parametru populace. V případě průměru platí:

$$\text{Výběrová chyba} = \bar{x} - \mu,$$

za předpokladu, že výběr je náhodný a nebyla učiněna žádná chyba ve vzorkování.

Chyby, které nastanou při sběru, zaznamenávání a tabulaci dat, se nazývají **chyby nezpůsobené náhodným výběrem**.

Důvody výskytu chyb nezpůsobených náhodným výběrem

1. Pokud je výběr nenáhodný (a tudíž nejpravděpodobněji nereprezentativní), mohou být výsledky ze získaného vzorku příliš odlišné od výsledků sčítání lidu.
2. Otázky mohou být formulovány tak, že nejsou plně pochopeny členy vzorku nebo populace.
3. Respondenti mohou úmyslně poskytovat falešné informace v reakci na některé citlivé otázky.
4. Osoba provádějící průzkum může udělat chybu a zadat nesprávné číslo do záznamů nebo udělat chybu při zadávání dat do počítače.

Příklad 1

Zadání: Znovu se podívejme na soubor pěti hodnocení studentů. Předpokládejme, že z této sady byl vybrán jeden vzorek tří hodnocení, a tento vzorek obsahuje hodnocení 70, 80 a 95. Vypočtěte výběrovou chybu.

Řešení:

$$\begin{aligned}
\mu &= \frac{70+78+80+80+95}{5} &&= 80.60 \\
\bar{x} &= \frac{70+80+95}{3} &&= 81.67 \\
\textbf{Výběrová chyba} &= \bar{x}-\mu = 81.67-80.60 &&= \mathbf{1.07}
\end{aligned}$$

To znamená, že odhadnuté průměrné skóre ze vzorku je o 1.07 vyšší než průměrné skóre populace. Poznamenejme, že k tomuto rozdílu došlo náhodně, protože jsme použili vzorek místo celé populace.

Výběrová chyba a Chyby nezpůsobené výběrem

Předpokládejme nyní, že vybereme vzorek tří hodnocení a **omylem zaznamenáme druhé hodnocení jako 82 místo 80**. V důsledku toho vypočítáme průměr vzorku jako

$$\bar{x} = \frac{70 + 82 + 95}{3} = 82.33.$$

Rozdíl mezi tímto průměrem vzorku a průměrem populace je

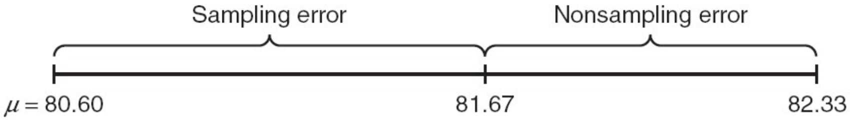
$$\bar{x} - \mu = 82.33 - 80.60 = 1.73.$$

Tento rozdíl nepředstavuje výběrovou chybu. Pouze 1.07 tohoto rozdílu je způsobeno výběrovou chybou.

Výběrová chyba a Chyby nezpůsobené výběrem

Zbývající část představuje chybu nezpůsobenou výběrem. Tato chyba je rovna 1.73 - 1.07 = **.66** a stala se kvůli chybě, kterou jsme udělali při zaznamenávání druhého skóre ve vzorku.

$$\begin{aligned}\text{Chyba nezpůsobená výběrem} &= \text{Nesprávné } \bar{x} - \text{Správné } \bar{x} \\ &= 82.33 - 81.67 = \mathbf{.66}\end{aligned}$$



Výběrové rozdělení
- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$
- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku
- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\hat{p}$

Střední hodnota a směrodatná odchylka $\bar{x}$

Střední hodnotu a směrodatnou odchylku výběrového rozdělení $\bar{x}$ budeme v tomto kurzu označovat $\mu_{\bar{x}}$ a $\sigma_{\bar{x}}$.

Střední hodnota výběrového rozdělení $\bar{x}$ je vždy rovna střední hodnotě populace. Platí:

$$\mu_{\bar{x}} = \mu.$$

Směrodatná odchylka výběrového rozdělení $\bar{x}$ je

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}},$$

kde σ je směrodatná odchylka populace a n je velikost vzorku. Tento vzorec se používá, když $\frac{n}{N} \leq .05$, kde N je velikost populace.

Vsuvka

Pokud není splněna podmínka $\frac{n}{N} \leq .05$, používáme následující vzorec pro výpočet $\sigma_{\bar{x}}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

kde faktor $\sqrt{\frac{N-n}{N-1}}$ se nazývá korekční faktor pro konečnou populaci.

Dvě důležité charakteristiky

1. Rozptyl výběrového rozdělení $\bar{x}$ je menší než rozptyl odpovídajícího rozdělení populace, tzn.

$$\sigma_{\bar{x}} < \sigma_x$$

2. Směrodatná odchylka výběrového rozdělení $\bar{x}$ klesá s rostoucí velikostí vzorku.

Příklad 2

Zadání: Střední hodnota hodinové mzdy pro všech 5000 zaměstnanců pracujících ve velké společnosti je 27.50 dolarů a směrodatná odchylka je 3.70 dolarů. Nechť $\bar{x}$ je hodnota průměrné hodinové mzdy pro vzorek určitých zaměstnanců vybraných z této společnosti náhodným výběrem.

Určete střední hodnotu a směrodatnou odchylku $\bar{x}$ pro následující velikosti vzorku: (a) 30 (b) 75 (c) 200.

Příklad 2: Řešení

$N = 5000$, $\mu = \$27.50$, $\sigma = \$3.70$.

Řešení (a) V tomto případě, $\frac{n}{N} = \frac{30}{5000} = .006 < .05$.

$$\mu_{\bar{x}} = \mu = \mathbf{\$27.50} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.70}{\sqrt{30}} = \mathbf{\$\ .676}$$

Řešení (b) V tomto případě, $\frac{n}{N} = \frac{75}{5000} = .015 < .05$.

$$\mu_{\bar{x}} = \mu = \mathbf{\$27.50} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.70}{\sqrt{75}} = \mathbf{\$\ .427}$$

Řešení (c) V tomto případě, $\frac{n}{N} = \frac{200}{5000} = .04 < .05$.

$$\mu_{\bar{x}} = \mu = \mathbf{\$27.50} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.70}{\sqrt{200}} = \mathbf{\$\ .262}$$

Obsah

Výběrové rozdělení
- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$
- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku
- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\hat{p}$

Tvar výběrového rozdělení $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$ souvisí s následujícími dvěma případy:

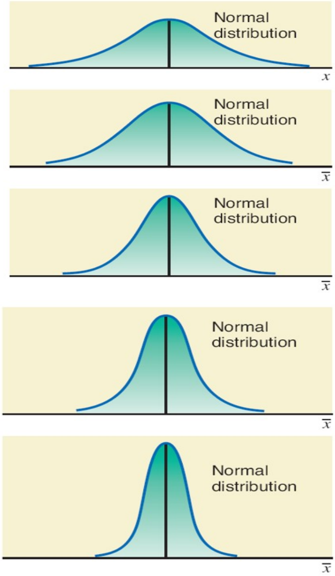
1. Populace, ze které jsou vzorky čerpány, má normální rozdělení.
2. Populace, ze které jsou vzorky čerpány, nemá normální rozdělení.

Výběr z populace, která má normální rozdělení

Pokud je populace, ze které jsou odebírány vzorky, normálně rozdělena se střední hodnotou μ a směrodatnou odchylkou σ, pak bude výběrové rozdělení průměru $\bar{x}$ získaného ze vzorku také normálně rozděleno s následující střední hodnotou a směrodatnou odchylkou (a to bez ohledu na velikost vzorku):

$$\mu_{\bar{x}} = \mu \quad \text{a} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Výběr z populace, která má normální rozdělení



(*a*) Population distribution.

(*b*) Sampling distribution of $\overline{x}$ for $n = 5$.

(*c*) Sampling distribution of $\overline{x}$ for $n = 16$.

(*d*) Sampling distribution of $\overline{x}$ for $n = 30$.

(*e*) Sampling distribution of $\overline{x}$ for $n = 100$.

Příklad 3

Podle Zprávy o odměňování lékařů za rok 2015 vydělávali v roce 2014 američtí internisté v průměru 196 000 dolarů.

Předpokládejme, že příjmy všech amerických internistů v roce 2014 jsou normálně rozděleny se střední hodnotou 196 000 dolarů a směrodatnou odchylkou 20 000 dolarů. Nechť $\bar{x}$ je průměr příjmů náhodného vzorku amerických internistů za rok 2014.

Spočtěte střední hodnotu a směrodatnou odchylku $\bar{x}$ a popište tvar jejího výběrového rozdělení pro velikost vzorku (a) 16 (b) 50 (c) 1000.

Příklad 3: Řešení

Řešení (a)

$$\mu_{\bar{x}} = \mu = \mathbf{196\,000}\,\text{USD} \quad \text{a} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20\,000\,\text{USD}}{\sqrt{16}} = \mathbf{5\,000}\,\text{USD}$$

Řešení (b)

$$\mu_{\bar{x}} = \mu = \mathbf{196\,000}\,\text{USD} \quad \text{a} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20\,000\,\text{USD}}{\sqrt{50}} = \mathbf{2\,828.427}\,\text{USD}$$

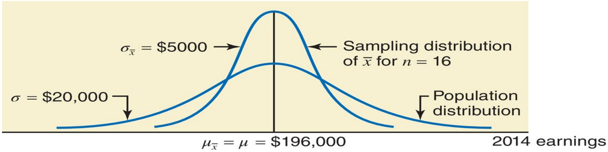
Řešení (c)

$$\mu_{\bar{x}} = \mu = \mathbf{196\,000}\,\text{USD} \quad \text{a} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$20,000}{\sqrt{1000}} = \mathbf{632.456}\,\text{USD}$$

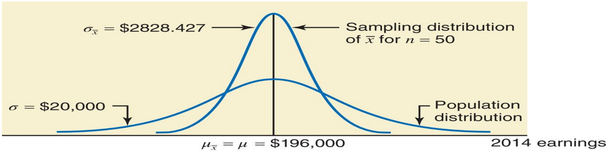
Protože roční příjmy všech amerických internistů za rok 2014 jsou normálně rozdělené, rozdělení průměru $\bar{x}$ pro vzorky 16, 50 a 1000 takových lékařů jsou také přibližně normálně rozdělené.

Příklad 3: Řešení graficky

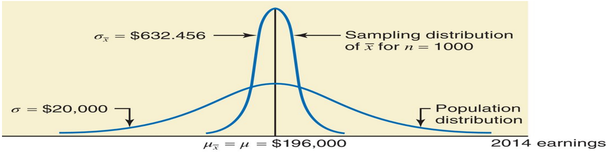
Řešení (a):



Řešení (b):



Řešení (c):



Výběr z populace, která nemá normální rozdělení

Centrální limitní věta

Podle centrální limitní věty, v případě velkého rozsahu vzorku[1], je pravděpodobnostní rozdělení $\bar{x}$ přibližně normální, a to bez ohledu na tvar rozdělení populace, na které je vzorkování provedeno. Střední hodnota a směrodatná odchylka rozdělení $\bar{x}$ jsou

$$\mu_{\bar{x}} = \mu \quad \text{a} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

[1] Velikost vzorku je obvykle považována za velkou, pokud $n \geq 30$.

Výběr z populace, která nemá normální rozdělení

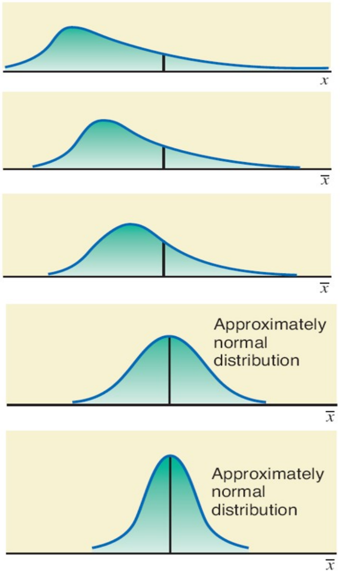
(*a*) Population distribution.

(*b*) Sampling distribution of $\bar{x}$ for $n = 4$.

(*c*) Sampling distribution of $\bar{x}$ for $n = 15$.

(*d*) Sampling distribution of $\bar{x}$ for $n = 30$.

(*e*) Sampling distribution of $\bar{x}$ for $n = 80$.



Příklad 4

Zadání: Průměrná výše nájemného, kterou platí všichni nájemci v malém městě, je 1550 dolarů a směrodatná odchylka činí 225 dolarů. Rozdělení nájemného všech nájemců v tomto městě je však zešikmené doprava.

Spočítejte střední hodnotu a směrodatnou odchylku rozdělení $\bar{x}$ a popište jeho tvar, v případech, kdy je velikost vzorku: (a) 30 (b) 100.

Příklad 4: Řešení

Řešení (a) Nechť $\bar{x}$ je střední hodnota nájemného zaplacená vzorkem 30 nájemců

$$\mu_{\bar{x}} = \mu = \mathbf{\$1550}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{225}{\sqrt{30}} = \mathbf{\$41.079}$$

Řešení (b) Nechť $\bar{x}$ je střední hodnota nájemného zaplacená vzorkem 100 nájemců

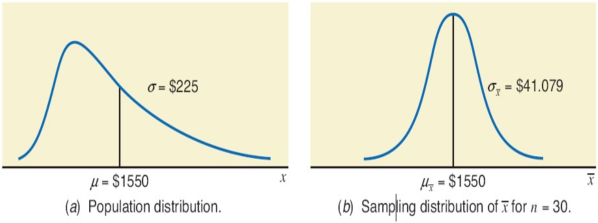
$$\mu_{\bar{x}} = \mu = \mathbf{\$1550}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{225}{\sqrt{100}} = \mathbf{\$22.500}$$

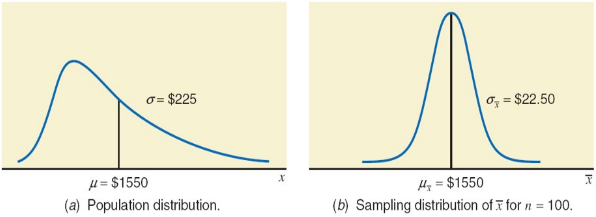
Ačkoliv rozdělení nájemného placeného všemi nájemci není normální, velikost vzorku je velká ($n \geq 30$), proto podle centrální limitní věty je tvar rozdělení $\bar{x}$ přibližně normální.

Příklad 4: Řešení graficky

Řešení (a):



Řešení (b):



Obsah

Výběrové rozdělení
- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$
- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku
- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

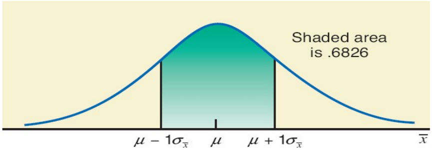
Aplikace výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\bar{x}$

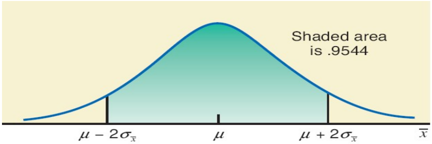
1. Pokud z populace vezmeme všechny možné vzorky stejné velikosti a vypočteme průměr pro každý z těchto vzorků, pak bude asi **68.26 %** průměrů v rozmezí jedné směrodatné odchylky od střední hodnoty (průměru) populace.
2. Pokud z populace vezmeme všechny možné vzorky stejné velikosti a vypočteme průměr pro každý z těchto vzorků, pak bude asi **95.44 %** průměrů v rozmezí dvou směrodatných odchylek od střední hodnoty (průměru) populace.
3. Pokud z populace vezmeme všechny možné vzorky stejné velikosti a vypočteme průměr pro každý z těchto vzorků, pak bude asi **99.74 %** průměrů v rozmezí tří směrodatných odchylek od střední hodnoty (průměru) populace.

Aplikace výběrového rozdělení $\bar{x}$

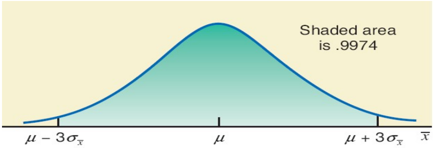
$\mu \pm 1 \cdot \sigma$



$\mu \pm 2 \cdot \sigma$



$\mu \pm 3 \cdot \sigma$



Příklad 5

Zadání: Předpokládejme, že hmotnosti všech balení určité značky sušenek jsou normálně rozdělené se střední hodnotou 32 unce a směrodatnou odchylkou 0.3 unce. Určete pravděpodobnost, že průměrná hmotnost $\bar{x}$ náhodně vybraného vzorku 20 balení těchto sušenek bude mezi 31.8 a 31.9 unce.

Řešení:

$$\mu_{\bar{x}} = \mu = 32 \text{ unce}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.3}{\sqrt{20}} \approx 0.06708204 \text{ unce}$$

!!! Pro výpočet pravděpodobnosti je nutné definovat příslušné z-skóre!

Z-skóre pro hodnotu $\bar{x}$

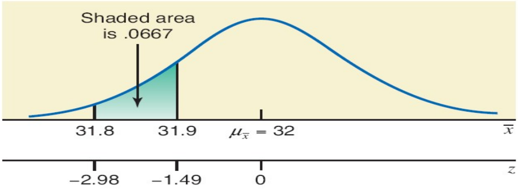
Z-skóre pro hodnotu $\bar{x}$ se vypočítá jako

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Příklad 5: Řešení

$$\text{Pro } \bar{x} = 31.8: \qquad z = \frac{31.8-32}{.06708204} = -2.98$$
$$\text{Pro } \bar{x} = 31.9: \qquad z = \frac{31.9-32}{.06708204} = -1.49$$

$$\begin{aligned} P(31.8 < \bar{x} < 31.9) &= P(-2.98 < Z < -1.49) \\ &= P(Z < -1.49) - P(Z < -2.98) \\ &= .0681 - .0014 = \mathbf{.0667} \end{aligned}$$



Příklad 6

Zadání: Podle společnosti Moebs Services Inc. stojí individuální běžný účet u amerických bank ročně mezi 350 a 450 dolary. Předpokládejme, že současná střední hodnota ročního nákladu na všechny běžné účty u amerických bank je 400 dolarů s odchylkou 30 dolarů. Nechť $\bar{x}$ je současný průměrný roční náklad na náhodně vybraný vzorek 225 individuálních běžných účtů těchto bank.

(a) Jaká je pravděpodobnost, že průměrný roční náklad na běžné účty v tomto vzorku je do 4 dolarů od střední hodnoty populace?

(b) Jaká je pravděpodobnost, že průměrný roční náklad na běžné účty v tomto vzorku je minimálně o 2.70 dolaru menší než střední hodnota populace?

Příklad 6: Řešení

Střední hodnota μ je 400 dolarů a směrodatná odchylka σ je 30 dolarů. Tvar rozdělení pravděpodobnosti populace je neznámý.

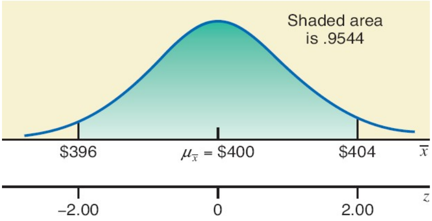
Nicméně, rozdělení vzorku $\bar{x}$ je přibližně normální, protože velikost vzorku je dostatečně velká ($n > 30$).

$$\mu_{\bar{x}} = \mu = \$400$$

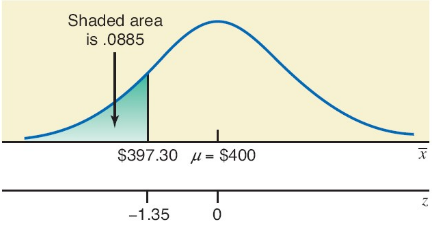
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\$30}{\sqrt{225}} = \$2.00$$

Příklad 6: Řešení Graficky

Řešení (a):



Řešení (b):



Příklad 6: Řešení výpočtem

Řešení (a)

Pro $\bar{x} = 396$ $\quad z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{396-400}{2.00} = -2.00$

Pro $\bar{x} = 404$ $\quad z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{404-400}{2.00} = 2.00$

$$P(396 \leq \bar{x} \leq 404) = P(-2.00 \leq z \leq 2.00) = .9772 - .0228 = \mathbf{.9544}$$

Pst, že průměrný roční náklad 225 běžných účtů v tomto vzorku je v rozmezí 4 dolarů od střední hodnoty populace, je .9544.

Řešení (b)

Pro $\bar{x} = 397.30$ $\quad z = \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \frac{397.30-400}{2.00} = -1.35$

$$P(\bar{x} \leq \$397.50) = P(z \leq -1.35) = \mathbf{.0885}$$

Pst, že průměrný roční náklad na běžné účty ve vzorku je minimálně o 2.70 dolarů nižší než střední hodnota populace, je .0885.

Obsah

Výběrové rozdělení
- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$
- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku
- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\hat{p}$

Podíly populace a vzorku

Podíly populace a vzorku, označované jako p a $\hat{p}$, se vypočítají jako

$$p = \frac{X}{N} \quad \text{a} \quad \hat{p} = \frac{x}{n},$$

kde

N = celkový počet prvků v populaci,

n = celkový počet prvků ve vzorku,

X = počet prvků v populaci, které vykazují určitou charakteristiku,

x = počet prvků ve vzorku, které vykazují určitou charakteristiku.

Příklad 7

Zadání: Předpokládejme, že ve městě žije celkem 789 654 rodin a 563 282 z nich vlastní domy. Z tohoto města je vybrán vzorek 240 rodin a 158 z nich vlastní domy.

Určete v populaci a ve vzorku podíly rodin, které vlastní domy.

Řešení:

$$p = \frac{X}{N} = \frac{563\,282}{789\,654} = \mathbf{.71}$$
$$\hat{p} = \frac{x}{n} = \frac{158}{240} = \mathbf{.66}$$

Výběrové rozdělení podílu $\hat{p}$

Pravděpodobnostní rozdělení podílu $\hat{p}$ získaného ze vzorku, se nazývá **výběrové rozdělení podílu**. Rozdělení poskytuje různé hodnoty, které může $\hat{p}$ nabývat, a jejich příslušné pravděpodobnosti.

Příklad 8

Zadání: Společnost Boe Consultant Associates má pět zaměstnanců. Tabulka uvádí jména těchto pěti zaměstnanců a informace týkající se jejich znalosti statistiky.

Name	Knows Statistics
Ally	Yes
John	No
Susan	No
Lee	Yes
Tom	Yes

Příklad 8: Řešení

Pokud definujeme podíl populace p jako podíl zaměstnanců, kteří rozumí statistice, pak platí:

$$p = \frac{3}{5} = 0.60.$$

Nyní předpokládejme, že vyberme všechny možné vzorky tří zaměstnanců a pro každý vzorek vypočítáme podíl zaměstnanců, kteří rozumí statistice.

Celkový počet vzorků = $C_3^5 = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 10.$

Příklad 8: Řešení

Všechny možné vzorky o velikosti 3 a hodnota $\hat{p}$ pro každý vzorek.

Sample	Proportion Who Know Statistics $\hat{p}$
Ally, John, Susan	1/3 = .33
Ally, John, Lee	2/3 = .67
Ally, John, Tom	2/3 = .67
Ally, Susan, Lee	2/3 = .67
Ally, Susan, Tom	2/3 = .67
Ally, Lee, Tom	3/3 = 1.00
John, Susan, Lee	1/3 = .33
John, Susan, Tom	1/3 = .33
John, Lee, Tom	2/3 = .67
Susan, Lee, Tom	2/3 = .67

Příklad 7-8 Řešení

Četnost výběru a relativní četnostní rozdělení $\hat{p}$ (vlevo) a pravděpodobnostní rozdělení $\hat{p}$ (vpravo)

$\hat{p}$	f	Relative Frequency
.33	3	3/10 = .30
.67	6	6/10 = .60
1.00	1	1/10 = .10
	$\Sigma f = 10$	Sum = 1.00

$\hat{p}$	$P(\hat{p})$
.33	.30
.67	.60
1.00	.10
	$\Sigma P(\hat{p}) = 1.00$

Střední hodnota s směrodatná odchylka $\hat{p}$

Střední hodnota výběrového rozdělení podílu $\hat{p}$, je označena $\mu_{\hat{p}}$ a je rovna podílu populace p. Tudíž,

$$\mu_{\hat{p}} = p$$

Směrodatná odchylka výběrového rozdělení podílu $\hat{p}$, je označena $\sigma_{\hat{p}}$ a je dána vzorcem

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

kde p je podíl populace, $q = 1 - p$, a n je velikost vzorku. Tento vzorec se používá, když $n/N \leq .05$, kde N je velikost populace.

Vsuvka

Pokud $n/N > 0.05$, pak se směrodatná odchylka $\hat{p}$ vypočítá jako:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\frac{N-n}{N-1}}$$

kde faktor $\sqrt{\frac{N-n}{N-1}}$ nazýváme korekční faktor konečné populace.

Centrální limitní věta pro výběrové rozdělení podílu

Centrální limitní věta pro výběrové rozdělení podílu

Podle centrální limitní věty je výběrové rozdělení $\hat{p}$ pro dostatečně velký vzorek přibližně normální. V případě podílu se velikost vzorku považuje za dostatečně velkou, pokud jsou np a nq obě větší než 5 - to znamená, pokud

$$np > 5 \quad \text{a} \quad nq > 5.$$

Příklad 9

Zadání: Podle průzkumu New York Times/CBS News 55 % dospělých uvedlo, že vlastnictví domu je velmi důležitou součástí amerického snu. Předpokládejme, že tento výsledek platí pro současnou populaci amerických dospělých.

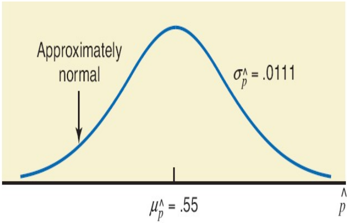
Nechť $\hat{p}$ je podíl amerických dospělých v náhodném vzorku 2000 osob, kteří řeknou, že vlastnictví domu je velmi důležitou součástí amerického snu.

Určete průměr a směrodatnou odchylku $\hat{p}$ a popište tvar jeho výběrového rozdělení.

Příklad 9: Řešení

$$\begin{aligned} p &= .55, \quad \Rightarrow \quad q = 1 - p = 1 - .55 = .45, \\ n &= 2000, \\ \mu_{\hat{p}} &= p = \mathbf{.55}, \\ \sigma_{\hat{p}} &= \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.55)(.45)}{2000}} = \mathbf{.0111}, \\ np &= 2000(.55) = 1100 \quad \text{a} \quad nq = 2000(.45) = 900. \end{aligned}$$

Výběrové rozdělení $\hat{p}$ je přibližně normální (podle *centrální limitní věty* – np a nq jsou oba větší než 5) s průměrem .55 a směrodatnou odchylkou .0111.



Obsah

Výběrové rozdělení
- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$
- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku
- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\hat{p}$

Aplikace výběrového rozdělení $\hat{p}$

Když provádíme studii, obvykle z populace odebíráme pouze jeden vzorek a všechna rozhodnutí nebo závěry činíme na základě výsledků tohoto jednoho vzorku.

K určení pravděpodobnosti, že hodnota $\hat{p}$ vypočítaná z jednoho vzorku spadne do daného intervalu, používáme koncepty průměru, směrodatné odchylky a výběrového rozdělení $\hat{p}$.

Příklad 10

Zadání: V nedávném celoamerickém telefonním průzkumu 75 % dospělých uvedlo, že vysokoškolské vzdělání je nyní pro většinu lidí příliš drahé. Předpokládejme, že tento výsledek odpovídá současné situaci mezi americkými dospělými.

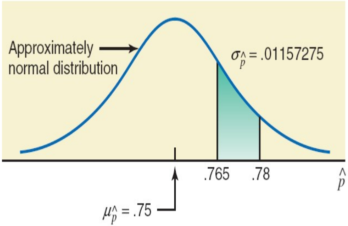
Nechť $\hat{p}$ je podíl v náhodně vybraném vzorku 1400 dospělých Američanů, kteří sdílejí tento názor.

Určete pravděpodobnost, že podíl dospělých v tomto vzorku, kteří budou tento názor zastávat je mezi mezi 76.5 % až 78 %.

Příklad 10: Řešení

$$p = .75, \quad \Rightarrow \quad q = 1 - p = 1 - .75 = .25,$$
$$n = 1400,$$
$$\mu_{\hat{p}} = p = \mathbf{.75},$$
$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.75)(0.25)}{1400}} = \mathbf{0.01157275},$$
$$np = 1400(0.75) = 1050 \quad \text{a} \quad nq = 1400(0.25) = 350.$$

Výběrové rozdělení $\hat{p}$ je přibližně normální (podle *centrální limitní věty* – np a nq jsou oba větší než 5) s průměrem .75 a směrodatnou odchylkou .0115.



Z skóre pro hodnotu $\hat{p}$

Z-skóre pro hodnotu $\hat{p}$ se vypočítá jako

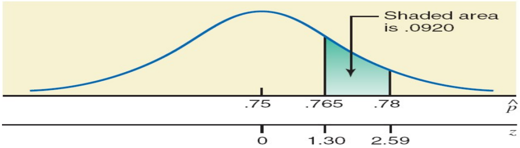
$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

Příklad 10: Řešení

$\hat{p} = .765 \qquad z = \frac{.765-.75}{.01157275} = 1.30$

$\hat{p} = .780 \qquad z = \frac{.78-.75}{.01157275} = 2.59$

$$\begin{aligned} P(.765 < \hat{p} < .78) &= P(1.30 < z < 2.59) \\ &= .9952 - .9032 \\ &= \mathbf{.0920} \end{aligned}$$



Pravděpodobnost, že 76.5 % až 78 % amerických dospělých z náhodného vzorku 1400 osob řekne, že vysokoškolské vzdělání se stalo pro většinu lidí příliš drahým, je 0.0920.

Příklad 11

Zadání: Zastánkyně práv obyvatel Pravdová, která kandiduje na post primátora ve velkém městě, tvrdí, že ji podporuje 53 % všech oprávněných voličů toho města. Předpokládejme, že toto tvrzení je pravdivé.

Jaká je pravděpodobnost, že v náhodném vzorku 400 registrovaných voličů z tohoto města bude Pravdovou podporovat méně než 49 %?

Příklad 11: Řešení

Máme $n = 400$, $p = .53$, a $q = 1 - p = 1 - .53 = .47$.
Střední hodnota a směrodatná odchylka jsou

$$\mu_{\hat{p}} = .53 \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.53)(.47)}{400}} = .02495496$$

Určíme z-skóre

$$z = \frac{0.49 - 0.53}{0.02495496} = -1.60$$

Dopočítáme pravděpodobnost

$$P(\hat{p} < .49) = P(z < -1.60) = \mathbf{.0548}$$

Pravděpodobnost, že v náhodném vzorku 400 osob bude Pravdovou preferovat méně než 49 % voličů, je .0548.

Shrnutí přednášky:

Výběrové rozdělení

- Pravděpodobnostní rozdělení populace a výběru
- Výběrová chyba a Chyby nezpůsobené výběrem

Střední hodnota a směrodatná odchylka $\bar{x}$

Tvar výběrového rozdělení $\bar{x}$

- Výběr z populace, která má normální rozdělení
- Výběr z populace, která nemá normální rozdělení

Aplikace výběrového rozdělení $\bar{x}$

Podíly populace a vzorku

- Výběrové rozdělení podílu $\hat{p}$
- Střední hodnota s směrodatná odchylka $\hat{p}$
- Tvar výběrového rozdělení $\hat{p}$

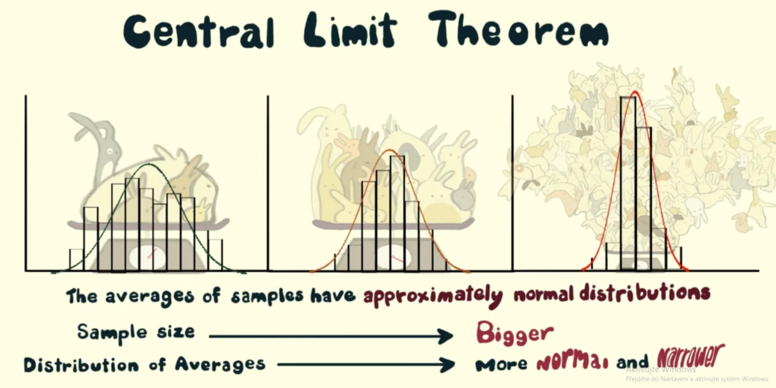
Aplikace výběrového rozdělení $\hat{p}$

Co si nastudovat na příští týden?

Příprava na cvičení: Leaflet 08
Koncepty a procedury, cv. 08, kap. 07

Povinná literatura: Mann (2016), Kapitola 8

Děkuji za pozornost!



Source: Casey Dunn & Creature Cast