

ЧАСТ IV СТАТИСТИЧЕСКО ОЦЕНЯВАНЕ

Учебни цели:

- Получаване на интервална оценка на средна аритметична величина – изчисляване на стандартна грешка на оценката, максимална грешка и доверителен интервал
 - ✓ при безвъзвратен подбор
 - ✓ при възвратен подбор
- Получаване на интервална оценка на относителен дял – изчисляване на стандартна грешка на оценката, максимална грешка и доверителен интервал
- Определяне на обем на репрезентативна извадка
 - ✓ за оценка на средна аритметична величина
 - ✓ за оценка на относителен дял

1. Интервална оценка на средна аритметична величина

1.1. Интервална оценка на средна аритметична при репрезентативна извадка с безвъзвратен подбор

Пример 1.1

Проведено е репрезентативно извадково изследване за оценка на средния доход на лице в трудоспособна възраст в едно населено място. В него са регистрирани 30 000 лица в тази възраст, от които е направена извадка от 300 лица чрез прост случаен подбор. Установени са:

- Средният месечен доход на лицата в извадката: 540 лв
- Стандартното отклонение по доход в извадката: 160 лв

Да се направи интервална оценка на средния доход на лице в населеното място при гаранционна вероятност 95%.

Решение:

1) Интервалната оценка представлява доверителен интервал за параметъра на генералната съвкупност – в случая средният доход на лице в населеното място – придружен с изискваната гаранционна вероятност. Доверителният се формира чрез два елемента:

- точковата оценка за средната аритметична величина;
- максималната стохастична грешка на тази оценка.

ДИ: $\bar{x}_0 \in (\bar{x} - \Delta_{\bar{x}}; \bar{x} + \Delta_{\bar{x}})$ с вероятност за сигурност 95%

2) Максималната грешка на оценката се формира чрез следните елементи:

- гаранционният множител;
- средната стохастична (стандартна) грешка на оценката.

$$\Delta_{\bar{x}} = z \cdot \mu_{\bar{x}}$$

При гаранционна вероятност 95% и нормално (Гаусово) разпределение на извадковата средна, стойността на гаранционният множител е: $z=1,96$.

3) Стандартната грешка на оценката при безвъзвратен подбор се изчислява чрез следната формула:

$$\mu_{\bar{x}} = \frac{\sigma_0}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

- Обемът на извадката е $n=300$, а обемът на генералната съвкупност е $N=30000$.
- Стандартното отклонение σ_0 по признака “доход” в генералната съвкупност (населеното място) е неизвестно. То се апроксимира чрез неговата извадкова оценка, изчислявана въз основа на стандартното отклонение от извадката:

$$\sigma_0 \approx \hat{\sigma} = \sigma \sqrt{\frac{n}{n-1}} = 160 \sqrt{\frac{300}{299}} = 160 \cdot \sqrt{1,0033} = 160,10016 = 160,26 \text{ лв}$$

- Стандартната грешка на оценката се изчислява след заместване на σ_0 с неговата извадкова оценка $\hat{\sigma}$:

$$\mu_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \frac{160,26}{\sqrt{300}} \sqrt{1 - \frac{300}{30000}} = \frac{160,26}{17,32} \sqrt{1 - 0,01}$$

$$\mu_{\bar{x}} = 9,25 \sqrt{0,99} = 9,25 \cdot 0,995 = 9,21 \text{ лв}$$

4) След получаване на стандартната грешка е възможно изчисляването и на максималната грешка:

$$\Delta_{\bar{x}} = z \cdot \mu_{\bar{x}} = 1,96 \cdot 9,21 \approx 18,05 \text{ лв}$$

5) За доверителния интервал се получава:

$$\text{ДИ: } \bar{x}_0 \in (540 - 18,05; 540 + 18,05)$$

$$\text{ДИ: } \bar{x}_0 \in (521,95; 558,05)$$

Извод: С вероятност за сигурност 95% може да се твърди, че средният доход на лице в трудоспособна възраст в населеното място е между 521,95 лв. и 558,05 лв.

 Решете самостоятелно следната задача:

По време на репрезентативното изследване на доходите в населеното място (пример 1.1) е установен и средният брой дни, през които са работили лицата от извадката през обхванатия едномесечен период. За тези лица той възлиза на 15,5 дни, при стандартно отклонение 4,5 дни. Да се направи интервална оценка на средния брой дни, отработени през месеца от лицата в трудоспособна възраст в населеното място, при гаранционна вероятност 95%.

1.2. Интервална оценка на средна аритметична при репрезентативна извадка с възвратен подбор

 Пример 1.2

Направена е случайна извадка от 100 апартамента, предлагани за продажба на пазара на недвижимите имоти в гр.София. За тези жилища е установена средна цена на квадратен метър 835 €/м² при стандартно отклонение 140 €/м². Да се направи интервална оценка на средната цена на квадратен метър при доверителна вероятност 99%.

Решение:

1) Доверителният интервал за параметъра на генералната съвкупност, в случая средната цена на кв.м. на предлаганите за продажба жилища в гр.София, ще бъде:

$$\text{ДИ: } \bar{x}_0 \in (\bar{x} - \Delta_{\bar{x}}; \bar{x} + \Delta_{\bar{x}}) \text{ с вероятност за сигурност } 99\%$$

2) Максималната грешка на оценката ще се изчисли при доверителна вероятност 99%, при което стойността на гаранционния множител е $z=2,58$.

$$\Delta_{\bar{x}} = z \cdot \mu_{\bar{x}}$$

3) Стандартната грешка на оценката при възвратен подбор се изчислява без отчитане на обема на генералната съвкупност:

$$\mu_{\bar{x}} = \frac{\sigma_0}{\sqrt{n}}$$

- Обемът на извадката е $n=100$ апартамента
- Стандартното отклонение σ_0 по признака “цена на м²” се оценява по данни от извадката:

$$\sigma_0 \approx \hat{\sigma} = \sigma \sqrt{\frac{n}{n-1}} = 140 \sqrt{\frac{100}{99}} = 140 \cdot \sqrt{1,01} = 140,1005 = 140,7 \text{ €/м}^2$$

- За стандартната грешка на оценката се получава:

$$\mu_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{140,7}{\sqrt{100}} = \frac{140,7}{10} = 14,07 \text{ €/м}^2$$

4) Тогава за максималната грешка на оценката ще се получи:

$$\Delta_{\bar{x}} = z \cdot \mu_{\bar{x}} = 2,58 \cdot 14,07 \approx 36,3 \text{ €/м}^2$$

5) Доверителният интервал се образува чрез получената максимална грешка:

$$\text{ДИ: } \bar{x}_0 \in (835 - 36,3; 835 + 36,3)$$

$$\text{ДИ: } 798,7 \leq \bar{x}_0 \leq 871,3$$

Извод: С вероятност за сигурност 99% може да се твърди, че средната цена на квадратен метър на жилищата, *предлагани за продажба в гр.София*, е между 798,7 €/м² и 871,3 €/м².

 Решете самостоятелно следната задача:

По време на репрезентативното изследване на пазара на недвижимите имоти (пример 1.2) е установена и средната площ на жилищата от извадката, възлизащ на 65 м², при стандартно отклонение 15 м². Да се направи интервална оценка на средната площ на предлаганите за продажба жилища в гр.София при гаранционна вероятност 99%.

2. Интервална оценка на относителен дял

2.1. Интервална оценка на относителен дял при репрезентативна извадка с безвъзвратен подбор

Пример 2.1

Чрез репрезентативното изследване на доходите в населеното място (пример 1.1) е установен и броят на безработните лица, възлизащ на 45 измежду 300-те лица попаднали в извадката. Да се направи интервална оценка на относителния дял на безработните лица в населеното място при гаранционна вероятност 95%.

Решение:

1) Интервалната оценка на относителен дял също представлява доверителен интервал за параметъра на генералната съвкупност, в случая относителния дял на безработните лица в населеното място. За разлика от оценката на средна аритметична величина, тук самата точкова оценка се получава като **относителна честота** на групата единици, притежаващи интересуващата ни характеристика.

- Точковата извадкова оценка за относителния дял е:

$$p = \frac{m}{n} = \frac{45}{300} = 0,15$$

- Доверителният дял се построява по аналогичен начин:
ДИ: $p_0 \in (p - \Delta_p; p + \Delta_p)$ с вероятност за сигурност 95%

2) Максималната грешка на оценката се формира чрез елементите:

- гаранционен множител (z)
- средната стохастична (стандартна) грешка на оценката μ_p

$$\Delta_p = z \cdot \mu_p$$

При гаранционна вероятност 95% и нормално разпределение на точковата оценка на относителния дял, стойността на гаранционния множител е $z=1,96$.

3) Стандартната грешка на оценката при безвъзвратен подбор се изчислява по формула:

$$\mu_p = \frac{\sigma_0}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

- Обемите на извадката и на генералната съвкупност са съответно $n=300$ и $N=30000$.
- Стандартното отклонение σ_0 по признака “заетост” в генералната съвкупност (населеното място) се апроксимира чрез неговата извадкова оценка, изчислявана въз основа на стандартното отклонение от извадката. При дихотомен признак стандартното отклонение се получава по следната формула:

$$\sigma = \sqrt{p(1-p)} = \sqrt{0,15 \cdot (1-0,15)} = \sqrt{0,15 \cdot 0,85} = \sqrt{0,1275} = 0,357$$

$$\sigma_0 \approx \hat{\sigma} = \sigma \sqrt{\frac{n}{n-1}} = 0,357 \sqrt{\frac{300}{299}} = 0,357 \cdot 1,0016 = 0,358$$

- Стандартната грешка на оценката се изчислява с помощта на извадковата оценка на стандартното отклонение $\hat{\sigma}$:

$$\mu_p = \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \frac{0,358}{\sqrt{300}} \sqrt{1 - \frac{300}{30000}} = \frac{0,358}{17,32} \sqrt{1 - 0,01}$$

$$\mu_p = 0,0207 \cdot 0,995 = 0,0206 \approx 0,021$$

4) След получаване на стандартната грешка се изчислява и максималната грешка на оценката:

$$\Delta_p = z \cdot \mu_p = 1,96 \cdot 0,021 = 0,041$$

5) Доверителният интервал за параметъра на генералната съвкупност p_0 е:

$$\text{ДИ: } p_0 \in (0,15 - 0,041; 0,15 + 0,041)$$

$$\text{ДИ: } 0,109 \leq p_0 \leq 0,191$$

$$\text{ДИ за } p_0(\%): 10,9\% \leq p_0(\%) \leq 19,1\%$$

Извод: С вероятност за сигурност 95% може да се твърди, че относителният дял на безработните лица в трудоспособна възраст в населеното място е между 10,9% и 19,1%.

 Решете самостоятелно следната задача:

По време на репрезентативното изследване на доходите в населеното място (пример 1.1) са установени 30 лица в извадката, които са наемани без писмен договор за труд от своя работодател. Да се направи интервална оценка на относителния дял на лицата в трудоспособна възраст в населеното място, работещи без писмен договор, при гаранционна вероятност 95%.

2.2. Интервална оценка на относителен дял при репрезентативна извадка с възвратен подбор

 Пример 2.2

По време на репрезентативното изследване на пазара на недвижимите имоти (пример 1.2) е установен и броят на жилищата с тухлено строителство от извадката, възлизащ на 30 бр. Да се направи интервална оценка на относителния дял на предлаганите за продажба тухлени жилища в гр.София при гаранционна вероятност 95%.

Решение:

1) Доверителният интервал за относителния дял е:

ДИ: $p_0 \in (p - \Delta_p; p + \Delta_p)$ с вероятност за сигурност 95%

- Точковата извадкова оценка за относителния дял е:

$$p = \frac{m}{n} = \frac{30}{100} = 0,30$$

2) Максималната грешка на оценката ще се изчисли при доверителна вероятност 95%, за което е необходим гаранционен множител $z=1,96$.

$$\Delta_p = z \cdot \mu_p$$

3) Стандартната грешка на оценката при възвратен подбор е:

$$\mu_p = \frac{\sigma_0}{\sqrt{n}}$$

- Обемът на извадката е $n=100$ жилища
- Стандартното отклонение по дихотомния признак “тухлен тип на строителство” се оценява по данни от извадката:

$$\sigma = \sqrt{p(1-p)} = \sqrt{0,3 \cdot 0,7} = \sqrt{0,21} = 0,458$$

$$\sigma_0 \approx \hat{\sigma} = \sigma \sqrt{\frac{n}{n-1}} = 0,458 \sqrt{\frac{100}{99}} = 0,458 \cdot 1,005 = 0,460$$

- За стандартната грешка на оценката се получава:

$$\mu_p = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{0,46}{\sqrt{100}} = \frac{0,46}{10} = 0,046$$

4) Тогава за максималната грешка на оценката ще се получи:

$$\Delta_p = z \cdot \mu_p = 1,96 \cdot 0,046 \approx 0,09$$

5) Доверителният интервал е:

$$\text{ДИ: } p_0 \in (0,30 - 0,09; 0,30 + 0,09)$$

$$\text{ДИ: } 0,21 \leq p_0 \leq 0,39$$

Извод: С вероятност за сигурност 95% може да се твърди, че относителният дял на тухлените жилища, предлагани за продажба в гр.София, е между 21% и 39%.

 Решете самостоятелно следната задача:

По време на репрезентативното изследване на пазара на недвижимите имоти (пример 1.2) е установен и броят на едностайните жилища от извадката, възлизащ на 40бр. Да се направи интервална оценка на относителния дял на предлаганите за продажба едностайни жилища в гр.София при гаранционна вероятност 95%.

3. Определяне на обем на репрезентативна извадка

3.1. Обем на извадка за оценка на средна аритметична

 Пример 3.1.1

Необходимо е да се определи обемът на представителната извадка, чрез която да се оцени средния доход на лицата в трудоспособна възраст в едно населено място, където са регистрирани 30 000 такива лица. При предходни изследвания е установена приблизително величината на стандартното отклонение по признака “доход”, възлизащо на ± 160 лв. За предстоящото изследване се предвижда:

- безвъзвратен подбор на единиците;
- максималната грешка на оценката да не надхвърля ± 10 лв;
- при гаранционна вероятност 95%.

Решение:

Обемът на извадката ще се получи по следната формула:

$$n = \frac{z^2 \cdot \sigma_0^2 \cdot N}{\Delta_{\bar{x}}^2 \cdot N + z^2 \cdot \sigma_0^2}$$

Тъй като изискваната гаранционна вероятност е 95%, стойността на гаранционния множител ще бъде $z=1,96$.

$$n = \frac{1,96^2 \cdot 160^2 \cdot 30000}{10^2 \cdot 30000 + 1,96^2 \cdot 160^2} = \frac{3,8416 \cdot 25600 \cdot 30000}{100 \cdot 30000 + 3,8416 \cdot 25600}$$

$$n = \frac{2950348800}{3000000 + 98345} = \frac{2950348800}{3098345} = 952,23 \approx 953 \text{ бр. лица}$$

Забележка: При планиране обема на извадката закръглянето се прави така, че да се обхване частично включената единица, в случая 953-та.

Пример 3.1.2

Необходимо е да се определи обемът на извадката, чрез която да се оцени средната цена на квадратен метър на жилищата, предлагани на пазара на недвижимите имоти в гр.София. От предходни изследвания е известно стандартното отклонение по признака “цена на кв.м”, възлизащо на $\pm 140 \text{ €/м}^2$. За предстоящото изследване се предвижда:

- възвратен подбор на единиците;
- максимално допустима грешка на оценката $\pm 20 \text{ €/м}^2$;
- гаранционна вероятност 99%.

Решение:

Обемът на извадката се изчислява по следната формула:

$$n = \frac{z^2 \cdot \sigma_0^2}{\Delta_{\bar{x}}^2}$$

Тъй като изискваната гаранционна вероятност е 99%, стойността на гаранционния множител ще бъде $z=2,58$.

$$n = \frac{2,58^2 \cdot 140^2}{20^2} = \frac{6,6564 \cdot 19600}{400} = \frac{130465}{400} = 326,16 \approx 327 \text{ бр. жилища}$$

Решете самостоятелно следните задачи:

А/ Да се определи необходимият обем на извадката за оценка на средния брой работни дни от месеца за лицата в трудоспособна възраст в населеното място. Като резултат от предходни изследвания е получено стандартното отклонение по признака “отработени дни” в размер на 4,5 дни. Изискванията за провеждане на репрезентативното изследване са:

- безвъзвратен подбор на единиците;
- максимално допустима грешка на оценката $\pm 0,2$ дни;
- гаранционна вероятност 95%.

Б/ Да се определи необходимият обем на извадката за оценка на средната площ на предлаганите за продажба жилища в гр.София. Като резултат от предходни изследвания е получено стандартното отклонение по признака “площ” в размер на 15 м^2 . Изискванията за провеждане на репрезентативното изследване са:

- възвратен подбор на единиците;
- максимално допустима грешка на оценката $\pm 1,5 \text{ м}^2$;
- гаранционна вероятност 99%.

3.2. Обем на извадка за оценка на относителен дял

Пример 3.2.1

Необходимо е да се определи обемът на представителната извадка, чрез която да се оцени относителния дял на безработните лица в трудоспособна възраст в едно населено място, където живеят 30000 лица от тази възрастова група. При предходни изследвания е установяван процент на безработица в размер на около 15%. За предстоящото изследване се предвижда:

- безвъзвратен подбор на единиците;
- максималната грешка на оценката да не надхвърля един процентен пункт (0,01);
- при гаранционна вероятност 95%.

Решение:

Обемът на извадката ще се получи по следната формула:

$$n = \frac{z^2 \cdot \sigma_0^2 \cdot N}{\Delta_p^2 \cdot N + z^2 \cdot \sigma_0^2}$$

Тъй като изискваната гаранционна вероятност е 95%, стойността на гаранционния множител ще бъде $z=1,96$. Дисперсията в генералната съвкупност се изчислява приблизително при предполагаема стойност на параметъра ($p_0=0,15$):

$$\sigma_0^2 = p_0(1 - p_0) = 0,15 \cdot (1 - 0,15) = 0,15 \cdot 0,85 = 0,1275$$

За обема на извадката се получава:

$$n = \frac{1,96^2 \cdot 0,1275 \cdot 30000}{0,01^2 \cdot 30000 + 1,96^2 \cdot 0,1275} = \frac{3,8416 \cdot 3825}{3 + 0,49} = \frac{14694}{3,49} \approx 4211 \text{ лица}$$

Пример 3.2.2

Необходимо е да се определи обемът на извадката, чрез която да се оцени относителния дял на тухлените жилища, предлагани на пазара на недвижимите имоти в гр.София. При предходно изследване е получен относителен дял на този тип жилища в размер на около 30%. За предстоящото изследване се предвижда:

- възвратен подбор на единиците;
- максимално допустима грешка на оценката 5 процентни пункта (0,05);
- гаранционна вероятност 95%.

Решение:

Обемът на извадката се изчислява по следната формула:

$$n = \frac{z^2 \cdot \sigma_0^2}{\Delta_p^2}$$

Тъй като изискваната гаранционна вероятност е 95%, стойността на гаранционния множител ще бъде $z=1,96$. Дисперсията в генералната съвкупност се изчислява приблизително при предполагаема стойност на параметъра ($p_0=0,30$):

$$\sigma_0^2 = p_0(1 - p_0) = 0,3 \cdot (1 - 0,3) = 0,3 \cdot 0,7 = 0,21$$

За обема на извадката се получава:

$$n = \frac{1,96^2 \cdot 0,21}{0,05^2} = \frac{3,8416 \cdot 0,21}{0,0025} = \frac{0,807}{0,0025} = 322,8 \approx 323 \text{ бр. жилища}$$

 Решете самостоятелно следните задачи:

А/ Да се определи необходимият обем на извадката за оценка на относителния дял на лицата в трудоспособна възраст в населеното място, работещи без писмен договор. Като резултат от предходни изследвания е получен относителен дял на тези лица в размер на 10%. Изискванията за провеждане на изследването са:

- безвъзвратен подбор на единиците;
- максимално допустима грешка на оценката 2 процентни пункта (0,02);
- гаранционна вероятност 95%.

Б/ Да се определи необходимият обем на извадката за оценка на относителния дял на предлаганите за продажба едностайни жилища гр.София. Като резултат от предходни изследвания е получен относителен дял на тези жилища в размер на 40%. Изискванията за провеждане на изследването са:

- възвратен подбор на единиците;
- максимално допустима грешка на оценката 3 процентни пункта (0,03);
- гаранционна вероятност 95%.

ДОПЪЛНИТЕЛНИ ЗАДАЧИ

Зад.4.1. Домакинствата от задача 1.1 (част I) са избрани по метода на простия случаен възвратен подбор измежду живеещите в столицата домакинства. Като се използват основните характеристики на честотните разпределения, при гаранционна вероятност 95% да се направят интервални оценки на:

- 1) средния брой на заетите лица в столично домакинство;
- 2) средния брой на децата в столично домакинство.

Зад.4.2. Да допуснем, че облигациите в инвестиционния портфейл на борсовия посредник са избрани от него чрез безвъзвратен лотариен подбор измежду котираните на борсата 350 вида. При гаранционна вероятност 95% да се направи интервална оценка на средната норма на възвръщаемост на облигациите, търгувани на фондовата борса.

Зад.4.3. Социологическа агенция провежда национално представително допитване сред пълнолетното население за оценка на относителния дял на поддръжниците на партия “АБВ”.

1/ Да се определи необходимия обем на извадката за оценка на този дял, ако максимално допустимата грешка е 2 %-ни пункта, при гаранционна вероятност 99%. При последните сондажи този дял е бил оценен в размер на около 6% от гласоподавателите.

2/ Направена е извадка от 1180 лица, от които 52 декларират подкрепата си за партията. Да се направи интервална оценка на относителния дял на поддръжниците при гаранционна вероятност 99%.

Зад.4.4. Маркетингова агенция провежда репрезентативно проучване на разходите за колбаси на домакинствата в един областен град с 50000 домакинства. Направена е случайна безвъзвратна извадка от 1000 домакинства, за които е установено:

- среден месечен разход за малотрайни колбаси в размер на 10 лв при оценка на стандартното отклонение 3 лв;
- среден месечен разход за трайни колбаси в размер на 15 лв при оценка на стандартното отклонение 6 лв;
- брой на домакинствата, не потребяващи колбаси – 300 бр.

1/ Да се получат интервални оценки при вероятност 99% за:

- средния месечен разход за малотрайни колбаси в града;
- средния месечен разход за трайни колбаси в града;
- делът на домакинствата, не потребяващи такива продукти.

2/ Каква по обем извадка е нужна, ако максималната грешка за оценка на средния разход за трайни колбаси трябва да се намали 2 пъти?