

Упражнение

Задача 1. (Задача за първата контролна работа и Задача за изпита)

Чрез случайна извадка с връщане са наблюдавани 200 работника от работещите в една фирма. Получени са следните данни за месечната им изработка (брой изделия):

Месечна изработка (бр. изделия) X_i	Брой Работници f_i
До 30	39
Над 30 До 55	50
Над 55 До 80	78
Над 80	33

1. Да се изобрази графично информацията за месечната изработка на работниците, попаднали в извадката.
2. Да се определи средната месечна изработка на работниците чрез:
 - 2.1. Средна аритметична;
 - 2.2. Медиана;
 - 2.3. Мода.

3. Да се оцени разсейването на работниците според месечната им изработка в абсолютен и в относителен размер.

4. Да се определи интервалът на доверителност, в който се намира действителната средномесечна изработка на всичките работници от фирмата при гаранционна вероятност, равна на 95%.

5. Ще се измени ли обемът на извадката и как, ако при равни други условия се постави изискване точността на оценката на средната месечна изработка да се повиши три пъти?

Решение на задачата:

1. Да се изобрази графично информацията за месечната изработка на работниците, попаднали в извадката.

Дейност 1. Ако в ИЧР има безкрайни интервали, то ние трябва да ги „затворим“, в смисъл – да обявим липсващата граница. Затваря-нето на такива интервали може да се извърши или по смисъл, или съвсем формално, като липсващите граници трябва да се изчислят така, че всички интервали да имат **еднаква ширина l** .

$l = b - a$, където: a – начало (ляв край на интервала); b – десен край на интервала.

$l = 55 - 30 = 25$ бр. изделия – ширина (дължина) на интервалите.

ИЧР($\hat{G}_{n=200}$; $X =$ мес. изработка, бр. изделия)

Месечна изработка (бр. изделия) X_i	Брой Работници f_i
От 5 До 30	39
Над 30 До 55	50
Над 55 До 80	78
Над 80 До 105	33

Началото на първия интервал ще определим, като от неговия край, в случая 30 ще извадим стойността на ширината $l = 25$. Това означава, че левият край на първия интервал ще е $30 - 25 = 5$.

Краят на последния (четвърти) интервал ще определим, като към неговото начало, в случая 80 прибавим стойността на ширината $l = 25$. Това означава, че десният край на последния интервал ще е $80 + 25 = 105$.

Така всеки един от интервалите от ИЧР има една и съща ширина $l = 25$.

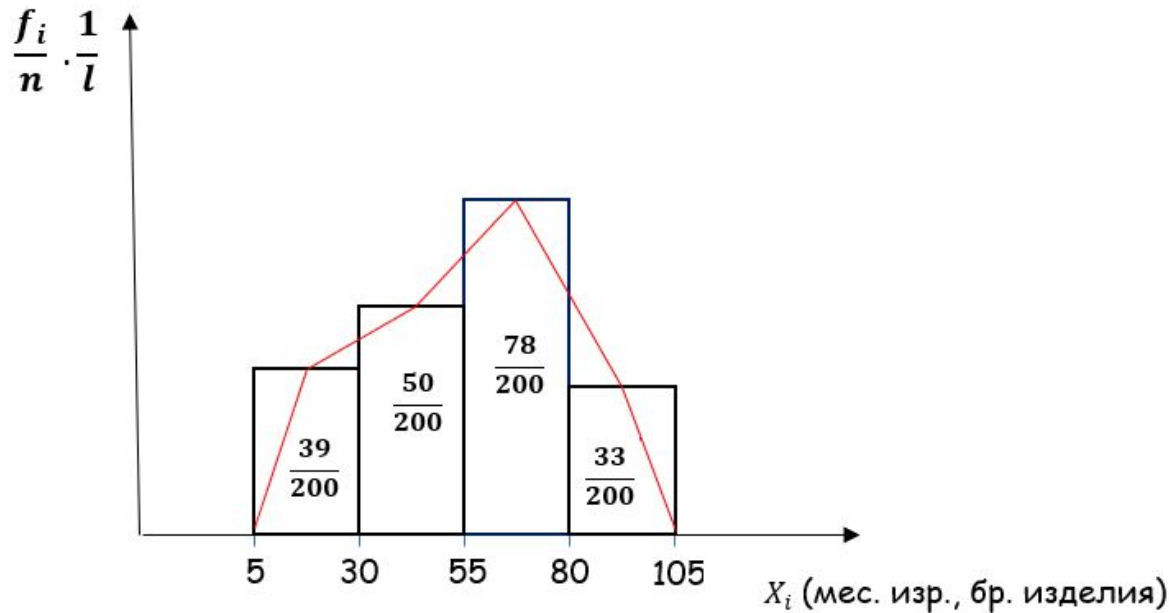
Дейност 2. Попълваме стълба на относителните честоти в таблицата на ИЧР.

ИЧР($\hat{G}_{n=200}$; $X =$ мес. изработка, бр. изделия)

Месечна изработка (бр. изделия)		
X_i		
От 5 До 30	39	39/200
Над 30 До 55	50	50/200
Над 55 До 80	78	78/200
Над 80 До 105	33	33/200
Σ	200 (n)	1

Дейност 3. Построяваме графиката на ИЧР

За целта над всеки интервал, разположен по абсцисната ос, построяваме правоъгълник с лице, равно на съответната относителна честота f_i/n . Фигурата, получена чрез така построените правоъгълници, се нарича **честотна хистограма**. Тя представлява графиката на ИЧР. Лицето на честотната хистограма е равно на 1.



$$\hat{x}'_n = 55,63 \text{ бр. изд.}$$

$$\hat{M}'_0 = 64,59 \text{ бр. изд.}$$

$$\hat{M}'_e = 58,52 \text{ бр. изд.}$$

Честотна хистограма

Начупената линия от отсечки, свързващи двата края на основата на честотната хистограма със средите на горните основи на съседните правоъгълници, се нарича **полигон на ЧР**.

2. Да се определи средната месечна изработка на работниците чрез:

2.1. Средна аритметична;

2.2. Медиана;

2.3. Мода.

Дейност 1: Изчисляваме стойността на средната аритметична по правилото:

$$\widehat{X}'_n = \frac{\sum_{i=1}^k (x'_i \cdot f_i)}{\sum_{i=1}^k f_i}, \text{ където:}$$

x'_i - среда на i -тия интервал

$$x'_i = \frac{a+b}{2}$$

$$x'_1 = \frac{5+30}{2} = \frac{35}{2} = 17,5$$

$$x'_3 = \frac{55+80}{2} = \frac{135}{2} = 67,5$$

$$x'_2 = \frac{30+55}{2} = \frac{85}{2} = 42,5$$

$$x'_4 = \frac{80+105}{2} = \frac{185}{2} = 92,5$$

ИЧР($\widehat{G}_{n=200}$; X = мес. изработка, бр. изделия)

Месечна изработка (бр. изделия)				
От 5 До 30	39	39/200	17,5	682,5
Над 30 До 55	50	50/200	42,5	2125,0
Над 55 До 80	78	78/200	67,5	5265,0
Над 80 До 105	33	33/200	92,5	3052,5
Σ	200(=n)	1	-	11 125

числитель на \widehat{x}'_n

$$\widehat{x}'_n = \frac{\sum_{i=1}^k (x'_i \cdot f_i)}{\sum_{i=1}^k f_i} = \frac{11\ 125}{200} = 55,63 \text{ бр. изд.}$$

Дейност 2: Определяме стойността на модата.

Модалният интервал е този интервал, за който абсолютната честота f_i е най-голяма.

В конкретния случай модалният интервал е:

Над 55 До 80

ИЧР ($\hat{G}_n=200$; X = мес. изработка, бр. изделия)

$$\hat{M}'_0 = L_{M_0} + \frac{(f_{M_0} - f_{M_0-1}) \cdot l_{M_0}}{(f_{M_0} - f_{M_0-1}) + (f_{M_0} - f_{M_0+1})}$$

където:

L_{M_0} - левият край (началото) на модалния интервал

$$L_{M_0} = 55$$

f_{M_0} - абсолютна честота в модалния интервал

$$f_{M_0} = 78$$

f_{M_0-1} - абсолютна честота в предмодалния интервал, ако има такъв

$$f_{M_0-1} = 50$$

f_{M_0+1} - абсолютна честота в следмодалния интервал, ако има такъв

$$f_{M_0+1} = 33$$

l_{M_0} - ширина (дължина) на модалния интервал

$$l_{M_0} = b - a = 80 - 55 = 25$$

Месечна изработка (бр. изделия) X_i		
От 5 До 30	39	39/200
Над 30 До 55	50	50/200
Над 55 До 80	78	78/200
Над 80 До 105	33	33/200
Σ	200	1

$$\widehat{M}'_0 = L_{M_0} + \frac{(f_{M_0} - f_{M_0-1}) \cdot l_{M_0}}{(f_{M_0} - f_{M_0-1}) + (f_{M_0} - f_{M_0+1})} = 55 + \frac{(78 - 50) \cdot 25}{(78 - 50) + (78 - 33)} = 55 + \frac{28 \cdot 25}{28 + 45} =$$

$$= 55 + \frac{700}{73} = 55 + 9,59 = 64,59 \text{ бр. изделия}$$

Дейност 3. Изчисляваме стойността на медианата като използваме данните от ИЧР.

1) Определяме номера на медианния случай.

$$N_{M_e} = \frac{n}{2} = \frac{200}{2} = 100$$

2) Попълваме стълба на кумулативните честоти в таблицата на ИЧР.

ИЧР($\widehat{G}_{n=200}$; X = мес. изработка, бр. изделия)

Месечна изработка (бр. изделия)					
X_i					
От 5 До 30	39	39/200	39	17,5	682,5
Над 30 До 55	50	50/200	89	42,5	2125
Над 55 До 80	78	78/200	167	67,5	5265
Над 80 До 105	33	33/200	200 (=n)	92,5	3052,5
Σ	200(=n)	1	-	-	11 125

3) Обхождаме стълба на кумулативните честоти отгоре надолу и търсим **първото** кумулативно число, което е поне равно на N_{Me} , т. е

$$c_i \geq N_{Me} ; c_i \geq 100$$

Обявяваме за медианен този интервал, който съответства на намереното кумулативно число.

Медианният интервал е Над 55 До 80

4) Записваме правилото, по което ще определим стойността на медианата.

$$\hat{M}'_e = L_{Me} + \frac{(N_{Me} - c_{Me-1}) \cdot l_{Me}}{(f_{Me})}, \text{ където:}$$

L_{Me} - левият край (началото) на медианния интервал

$$L_{Me} = 55$$

N_{Me} - номер на медианния случай

$$N_{Me} = 100$$

C_{Me-1} - кумулативна честота в предмедианния интервал, ако има такъв

$$C_{Me-1} = 89$$

f_{Me} - абсолютна честота в медианния интервал

$$f_{Me} = 78$$

l_{Me} - ширина (дължина) на медианния интервал

$$l_{Me} = b - a = 80 - 55 = 25$$

5) Изчисляваме стойността на медианата по правилото.

$$\widehat{M}'_e = L_{Me} + \frac{(N_{Me} - C_{Me-1}) \cdot l_{Me}}{f_{Me}} = 55 + \frac{(100 - 89) \cdot 25}{78} = 55 + \frac{11.25}{78} = 55 + \frac{275}{78} =$$

$$= 55 + 3,52 = 58,52 \text{ бр. изд.}$$

Изчисляваме стойността на медианата по правилото.

$$\widehat{M}'_e = L_{Me} + \frac{(N_{Me} - C_{Me-1}) \cdot l_{Me}}{f_{Me}} = 55 + \frac{(100 - 89) \cdot 25}{78} = 55 + \frac{11.25}{78} =$$

$$55 + \frac{275}{78} = 55 + 3,52 = 58,52 \text{ бр. изд.}$$

3. Да се оцени разсейването на работниците според месечната им изработка в абсолютен и в относителен размер.

Дейност 1. Изчисляваме стойностите на абсолютните мерки за статистическо разсейване по правилата:

АМСР:

1) Размах: $\hat{R}'_X = (X_{\max} - X_{\min})_{\text{м. ед.}} = (105 - 5) = 100 \text{ бр. изд.}$

2) Дисперсия: $\hat{\sigma}'^2_X = \frac{\sum_{i=1}^k (x'_i - \hat{x}'_n)^2 \cdot f_i}{\sum_{i=1}^k f_i} \text{ м. ед}^2 = \frac{121\,171,88}{200} = 605,86 \text{ бр. изд.}^2$

ИЧР ($\hat{G}_{n=200}$; $X = \text{мес. изработка, бр. изделия}$) $\hat{x}'_n = 55,63 \text{ бр. изд.}$

Месечна изработка (бр. изделия) X_i								
От 5 До 30	39	39/200	39	17,5	682,5	-38,13	1453,897	56701,9791
Над 30 До 55	50	50/200	89	42,5	2125	-13,13	172,3969	8619,845
Над 55 До 80	78	78/200	167	67,5	5265	11,87	140,8969	10989,9582
Над 80 До 105	33	33/200	200	92,5	3052,5	36,87	1359,397	44860,0977
Σ	200(=n)	1	-	-	11 125	-	-	121 171,88

числител на $\hat{\sigma}'^2_x$

3) Стандартно отклонение: $\hat{\sigma}'_X = \sqrt{\hat{\sigma}'^2_x} \text{ м. ед.} = \sqrt{605,86} = \mathbf{24,61 \text{ бр. изд.}}$

Дейност 2. Изчисляваме стойностите на относителните мерки за статистическо разсейване по правилата:

1) Коэффициент на вариация на размаха:

$$V_{\hat{R}'_x} = \left(\frac{\hat{R}'_x}{\hat{x}'_n} \cdot 100 \right) \% = \left(\frac{100}{55,63} \cdot 100 \right) \% = \mathbf{179,76\%}$$

2) Коэффициент на вариация на стандартното отклонение:

$$V_{\hat{\sigma}'_x} = \left(\frac{\hat{\sigma}'_x}{\hat{x}'_n} \cdot 100 \right) \% = \left(\frac{24,61}{55,63} \cdot 100 \right) \% = \mathbf{44,24\%}$$

4. Да се определи интервалът на доверителност, в който се намира действителната средномесечна изработка на всичките работници от фирмата при гаранционна вероятност, равна на 95%.

Алгоритъм за построяване на интервална оценка на \bar{x} , получена по данни от една случайна извадка с връщане с обем $n = 200$ от дадена РС G_N^X и при избрана от гаранционна вероятност $P(z) = 0,95$ (95%):

Стъпка 1: $\hat{x}'_n = 55,63$ бр. изд. - точкова оценка на \bar{x}

Стъпка 2: Определяме стойността на средната стохастична грешка $\mu_{\bar{x}}$ съдържаща се в точковата оценка:

$$\mu_{\bar{x}} = \frac{\hat{\sigma}'_X}{\sqrt{n}} = \frac{24,61}{\sqrt{200}} = 1,74 \text{ бр. изд. - стойност на средната стохастична}$$

грешка

Стъпка 3: При гаранционна вероятност $P(z) = 0,95$ (95%) определяме стойността на гаранционния множител z и изчисляваме стойността на максималната стохастична грешка Δ_{x_i} , съдържаща се в точковата оценка \hat{x}'_n по правилото:

При $P(z) = 0,95$ (95%) $\rightarrow z = 1,96$ (от Статистическата таблица за Стандартно нормално разпределение)

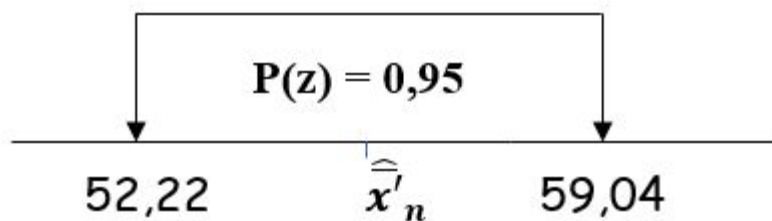
$\Delta_{x_i} = z \cdot \mu_{x_i} = 1,96 \cdot 1,74 = \mathbf{3,41 \text{ бр. изд.}}$ - стойност на максималната стохастична грешка

Стъпка 4: Обявяваме интервалната оценка за \bar{x}' :

$$\widehat{x}_n - \Delta_{\bar{x}'} \leq \bar{x} \leq \widehat{x}_n + \Delta_{\bar{x}'}$$

$$55,63 - 3,41 \leq \bar{x} \leq 55,63 + 3,41$$

52,22 ≤ \bar{x} ≤ 59,04 - Интервална оценка (доверителен интервал) на средната аритметична на реалната съвкупност \bar{x} , получена по данни от една случайна извадка с връщане с обем $n = 200$ от ГС G_N по признака мес. изработка, бр. изделия и при гаранционна вероятност $P(z) = 0,95$.



5. Ще се измени ли обемът на извадката и как, ако при равни други условия се постави изискване **ТОЧНОСТТА** на оценката на средната месечна изработка да се повиши три пъти?

При равни други условия, за да бъде три пъти по-точна интервалната оценка за \bar{x}' , стойността на максималната стохастична грешка $\Delta'_{\bar{x}'}$ на тази нова извадка трябва да бъде 3 пъти по-малка.

$$\Delta_{\bar{x}} = 3,41 \text{ бр. изд.}$$

$$\Delta'_{\bar{x}} = \frac{\Delta_{\bar{x}}}{3} = \frac{3,41}{3} = 1,14 - \text{желана стойност на максималната стохастична грешка,}$$

която ще ни осигури 3 пъти по-точна интервална оценка за \bar{x}' .

Обемът на извадката при три пъти по-точна интервална оценка ще намерим по правилото:

$$n' = \frac{\widehat{\sigma}_{\bar{X}}^2 \cdot z^2}{\Delta'_{\bar{x}}{}^2} = \frac{24,61^2 \cdot 1,96^2}{1,14^2} = \frac{24,61^2 \cdot 1,96^2}{1,14^2} = 42,31^2 = 1790 \text{ раб.}$$

Да, обемът на извадката ще се промени от $n = 200$ на $n' = 1790$, за да получим от новата извадка три пъти по-точна интервална оценка за \bar{x} .