Összefoglaló óra

9.27

① üdítőfogyasztás 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 4 4

a) $N = 20$

$Me = (Y^*_{10} + Y^*_{11})/2 = 2$

$\text{Variance} = (\text{Std. Dev})^2 = 0{,}9206^2 = 0{,}8475$

$\text{Kurtosis} = \frac{M_4'}{M_2'^2} - 3 = \frac{M_4(\bar{Y})}{\sigma^4} - 3$

lapultság

$M_4^c = M_4(\underset{2{,}05}{\bar{x}}) = \frac{6\cdot(1-2{,}05)^4 + 9\cdot(2-2{,}05)^4 + 3\cdot(3-2{,}05)^4 + 2\cdot(4-2{,}05)^4}{20} = \frac{38{,}654625}{20}$

$= 1{,}93273125$

$M_2^c = M_2(\bar{x}) = 16{,}95/20 = 0{,}8475 \rightarrow \sigma = 0{,}9205976$

$\text{Kurtosis} = \frac{1{,}9327}{0{,}8475^2} - 3 = 2{,}6908 - 3 = -0{,}3092$

az azonos paraméterű norm. elo-nál kissé laposabb

$Max = 4$

$Q_1 = Y_{\frac{1}{4}} = \quad \frac{1}{4}\cdot(20+1) = 5\frac{1}{4} < {5 \atop 1/4} = Y^*_5 + \frac{1}{4}(Y^*_6 - Y^*_5) = 1 + \frac{1}{4}(1-1) = 1$

$Q_2 = Me = 2$

$Q_3 = Y_{\frac{3}{4}} = \quad \frac{3}{4}(20+1) = 15\frac{3}{4} < {15 \atop 3/4} = Y^*_{15} + \frac{3}{4}(Y^*_{16} - Y^*_{15}) = 2 + \frac{3}{4}(3-2) = 2{,}75$

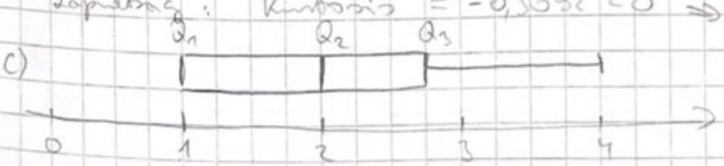
$f_2 = 20-6-3-2 = 9$

$g_4 = \frac{2}{20} = 0{,}1 = 10\%$

$g_2' = 30+45 = 75\%$, $\quad g_3' = 75+15 = 90\%$, $\quad g_4' = 90+10 = 100\%$

b) Ferdeség: Skewness $= 0{,}726 > 0 \Rightarrow$ jobbra hosszan elnyúló (balra ferdült)
Lapultság: Kurtosis $= -0{,}3092 < 0 \Rightarrow$ kissé lapultabb, mint az azonos paraméterű norm. elo.

c)



(2.) Cserebogarak

a) =

Cserebogarak száma (Milliárd) $Y_{i0} - Y_{i1}$		Megyék száma f_i	osztály közép $\overline{Y}_i$	f_i'	rel. gyakoriság g_i
1 – 5	4	8	3	8	8/23 = 0,3478 = 34,78%
5 – 9	4	9	7	17	9/23 = 0,3913 = 39,13%
9 – 13	4	3	11	20	3/23 = 0,1304 = 13,04%
13 – 17	4	3	15	23	3/23 = 0,1304 = 13,04%
		23			

$$\overline{Y} = \frac{8\cdot 3 + 9\cdot 7 + 3\cdot 11 + 3\cdot 15}{23} = \frac{165}{23} = 7,1739$$

$$Mo = Y_{mo,0} + \frac{d_a}{d_a + d_f}\cdot h_{mo} = 5 + \frac{9-8}{(9-8)+(9-3)}\cdot(9-5) = 5 + \frac{1}{7}\cdot 4 = 5\tfrac{4}{7} = 5,5714$$

$$Me = Y_{\frac{1}{2}} = \qquad = Y_{2,0} + \frac{\frac{1}{2}\cdot 23 - f_1'}{f_2}\cdot h_2 = 5 + \frac{11,5-8}{9}\cdot(9-5) =$$

$$f_1' < \tfrac{1}{2}\cdot 23 = 11,5 \leq f'_{②}$$

$$= 5 + \frac{3,5}{9}\cdot 4 = 5 + \frac{14}{9} = 6,5556$$

Ferdeség: Pearson-féle: $P = 3\cdot\dfrac{\overline{Y} - Me}{\sigma} = 3\cdot\dfrac{7,1739 - 6,5556}{\sigma} > 0 \Rightarrow$ jobbra hosszan elnyúló

pontosabban:

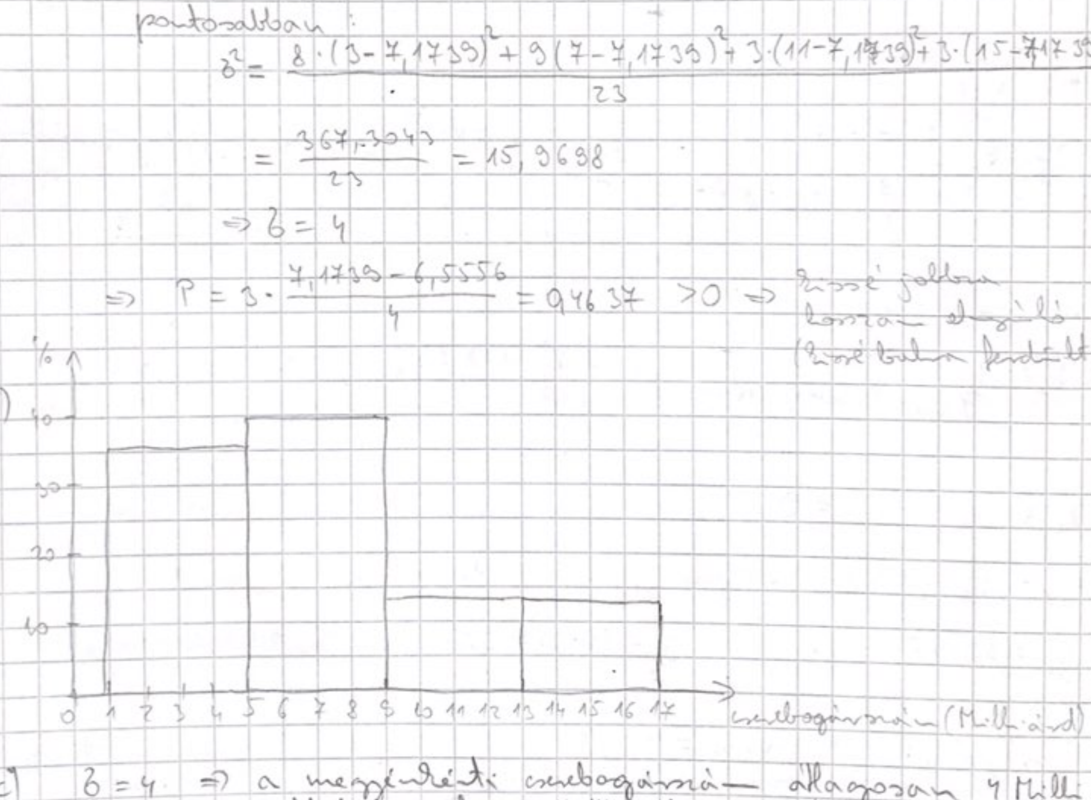
$$\sigma^2 = \frac{8\cdot(3-7,1739)^2 + 9(7-7,1739)^2 + 3\cdot(11-7,1739)^2 + 3\cdot(15-7,1739)^2}{23}$$

$$= \frac{367,3043}{23} = 15,9698$$

$$\Rightarrow \sigma = 4$$

$\Rightarrow P = 3\cdot\dfrac{7,1739 - 6,5556}{4} = 0,4637 > 0 \Rightarrow$ kissé jobbra hosszan elnyúló (kissé balra ferdült)

b)



c) $\sigma = 4 \Rightarrow$ a megyénkénti cserebogárszám átlagosan 4 Milliárddal tér el a 7,17 milliárdos átlagtól

$$Q_1 = Y_{\frac{1}{4}} = \qquad = Y_{1,0} + \frac{\frac{1}{4}\cdot 23 - 0}{f_1}\cdot h_1 = 1 + \frac{5,75}{8}\cdot(5-1) = 3,875$$

$$\tfrac{1}{4}\cdot 23 = 5,75 \leq f'_{①}$$

A megyék legfeljebb 1/4-ében kevesebb, mint 3,875 milliárd cserebogár van és legfeljebb 3/4-ben több.

(3.) Sablon

a) $N_5 = 500 - (122 + 139 + 107 + 63) = 69$

$$\bar{Y} = \frac{N_1 \cdot \bar{Y}_1 + N_2 \bar{Y}_2 + N_3 \bar{Y}_3 + N_4 \cdot \bar{Y}_4 + N_5 \cdot \bar{Y}_5}{N} = \frac{122 \cdot 41{,}84 + 139 \cdot 41{,}90 + 107 \cdot 41{,}51 + 63 \cdot 41{,}06 + 69 \cdot 41{,}93}{500}$$

$$= \frac{20850{,}1}{500} = 41{,}7002$$

← számok NÉGYZETEKre hasonlóból ⇒ $\sigma^2_{belső}$

$Max_3 = 68$

$$SSK = \sum_{j=1}^{5} N_j (\bar{Y}_j - \bar{Y})^2 =$$

$$= 122 \cdot (41{,}84 - 41{,}70)^2 + 139 \cdot (41{,}90 - 41{,}70)^2 + 107 \cdot (41{,}51 - 41{,}70)^2 + 63 \cdot (41{,}06 - 41{,}70)^2 +$$

$$+ 69 \cdot (41{,}93 - 41{,}70)^2 =$$

$$= 122 \cdot 0{,}14^2 + 139 \cdot 0{,}2^2 + 107 \cdot 0{,}19^2 + 63 \cdot 0{,}64^2 + 69 \cdot 0{,}23^2 = 41{,}2688$$

$SST = SSK + SSB = 41{,}2688 + 42\,801{,}744 = 42\,843{,}0128$

! vagy $SST = N \cdot \sigma_T^2 = 500 \cdot 9{,}2566^2 = 42\,842{,}32178 \Rightarrow SSK = SST - SSB = 40{,}57778$

b) Legfiatalabb: 23 éves
Legidősebb: 64 éves
Legbalesetmentesebbek átlagéletkora: 41,93 év

c) $H^2 = \frac{SSK}{SST} = \frac{41{,}2688}{42\,843{,}0128} = 0{,}000963256$ – egy biztosított által okozott balesetek számának ismerete 0,09%-át magyarázza az életkor szórásának

$H = 0{,}0310$ ← gyenge a kapcsolat az életkor és az okozott balesetek száma között.

tény\vált

(4.)

Haj \ Szem	kék	zöld	sötét	
szőke	5 (2,64)	9 (5,28)	8 (14,08)	22
sötét	7 (9,36)	15 (18,72)	56 (49,92)	78
	12	24	64	100

$$\chi^2 = \frac{(5-2,64)^2}{2,64} + \frac{(9-5,28)^2}{5,28} + \frac{(8-14,08)^2}{14,08} + \frac{(7-9,36)^2}{9,36} + \frac{(15-18,72)^2}{18,72} + \frac{(56-49,92)^2}{49,92} =$$

$$= 9,4308$$

Cramer-féle assoc. e.h.

$$C^2 = \frac{\chi^2}{N \cdot \min\{\underbrace{r-1}_{1}, \underbrace{c-1}_{2}\}} = \frac{9,4308}{100} = 0,0943 \Rightarrow C = 0,3071$$

gyenge-közepes a kapcsolat a szem és hajszín közt

(5.) Sör - kísérlet → korreláció

Pearson féle korreláció

$$r_{X,Y} = \frac{\sum d_{x_i} \cdot d_{y_i}}{\sqrt{\sum d_{x_i}^2 \cdot \sum d_{y_i}^2}}$$

$$\sigma_Y^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} d_{Y_i}^2 \Rightarrow \sum d_{Y_i}^2 = N \cdot \sigma_Y^2 = 25 \cdot 410^2 = 4\,202\,500$$

Lin. korr. e.h.

$$r_{XY} = \frac{25\,380}{\sqrt{170 \cdot 4\,202\,500}} = 0,9495$$

Determinációs e.h.

$$r^2 = 0,9016$$

← PRE-mutató

- a középkísérlet a ...tás szórásának 90,16%-át magyarázza
- nagyon lineáris! erős a kapcsolat a két vált. közt pozitív irányú

(6.) szállodák

Spearman féle rangkorreláció

$$\rho = 1 - \frac{6}{N(N^2-1)} \sum_{i=1}^{n} (R_{X_i} - R_{Y_i})^2$$

$$\sum_{i=1}^{N} (R_{X_i} - R_{Y_i})^2 = 1^2 + 1^2 + 2^2 + 0^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2 + 2^2 =$$

$$= 1 + 1 + 4 + 1 + 1 + 1 + 1 + 4 = 14$$

$$\rho = 1 - \frac{6}{10(10^2-1)} \cdot 14 = 0,9152$$

→ erős a kapcsolat a két rangsorolás között

$$\rho^2 = 0,8375$$

– A földrajzi helyzet a vendégszám szórásának kb 84%-át magyarázza (szimmetrikus a vált-ban)

7. Autó teljesítmény

a) $Q_1 = Y_{\frac{1}{4}} = \underset{\frac{1}{4}\cdot(7+1)=2}{} = Y^*_2 = 123,23$

118,43 ; 123,23 ; 124,12 ; 129,23 ; 129,61 ; 131,18 ; 132,30

$Q_3 = Y_{\frac{3}{4}} = \underset{\frac{3}{4}\cdot(7+1)=6}{} = Y^*_6 = 131,18$

$\text{Std. Deviation}_{91\text{-es}} = \sqrt{\text{Variance}_{91}} = \sqrt{2,055} = 1,4335$

$\text{Min}_{91\text{-es}} = \text{Max} - \text{Range} = 126,31 - 4,58 = 121,73$

$\text{Median}_{95\text{-}59} = Q_2 = 129,2300$

$\text{Range}_{95\text{-}59} = \text{Max} - \text{Min} = 132,30 - 118,43 = 13,87$

$\text{Interquartile Range} = Q_3 - Q_1 = 131,18 - 123,23 = 7,95$

csoporton belüli szóródás: csoportátlagtól való átlagos eltérés egyéb, nem vizsgált tényezők hatása

$SSB = \sum N_j \cdot \sigma_j^2 = 7\cdot 2,055 + 7\cdot 25,557 + 7\cdot 24,835 = 367,129$

$SSK = SST - SSB = 52,955$ – csoportképző ismérvnek tulajdonítható: a részátlagoknak a főátlagtól való átlagos eltérése

b) Variancia hányados

$$H^2 = \frac{\sigma_K^2}{\sigma_T^2} = \frac{SSK}{SST} = \frac{52,955}{420,084} = 0,1261$$

$H = 0,3550$

c) A max forgatónyomaték ingadozásának 12,61%-a használt benzinnek tulajdonítható, a maradék 87,39%-a pedig egyéb, itt külön nem vizsgált tényezőknek tulajdonítható.

d) A választott benzin a maximális forgatónyomaték szórásnégyzetének 12,61%-át magyarázza.

8.

	Családok száma	Átlagos $\bar{Y}$ gyermekszám
Megyei jogú város	120	1,5
Nem megyei jogú város	130	2
Község	250	$\bar{Y}_3$
	500	$\frac{1200}{500} = 2,4$

A lakóhely az átlagos gyermekszám ingadozásának kb 10,40%-át magyarázza.

$$M_2 = \frac{\sum Y_i^2}{N} = 10 \quad (\overline{Y^2})$$

Vegyes kapcsolat: $\sigma^2 = \frac{\sum Y_i^2}{N} - (\bar{Y})^2 = 10 - 2,4^2 = 4,24$

$$\frac{120\cdot 1,5 + 130\cdot 2 + 250\,\bar{Y}_3}{500} = \frac{1200}{500} \Rightarrow \bar{Y}_3 = 3,04$$

$$\sigma_K^2 = \frac{120\cdot(1,5-2,4)^2 + 130(2-2,4)^2 + 250(3,04-2,4)^2}{500} = \frac{220,4}{500} = 0,4408$$

$$H^2 = \frac{\sigma_K^2}{\sigma_T^2} = \frac{0,4408}{4,24} = 0,10396 \Rightarrow H = 0,3224$$ – közepes kapcsolat