Урок 6. ДЗ.

1) Даны значения величины заработной платы заемщиков банка (zp) и значения их поведенческого кредитного скоринга (ks)

zp = [35, 45, 190, 200, 40, 70, 54, 150, 120, 110]

ks = [401, 574, 874, 919, 459, 739, 653, 902, 746, 832]

а) Найдите ковариацию этих двух величин

$$M(zp) = \frac{35+45+190+200+40+70+54+150+120+110}{10} = 101,4$$

$$M(ks) = \frac{401+574+874+919+459+739+653+902+746+832}{10} = 709,9$$

$$cov(zp, ks) = \frac{1}{n-1}\sum_{i=1}^{n}(zp_i - M(zp))(ks_i - M(ks)) =$$

$= (((35-101,4)(401-709,9))+((45-101,4)(574-709,9))+((190-101,4)(874-709,9))+$
$+((200-101,4)(919-709,9))+((40-101,4)(459-709,9))+((70-101,4)(739-709,9))+$
$+((54-101,4)(653-709,9))+((150-101,4)(902-709,9))+((120-101,4)(746-709,9))+$
$+((110-101,4)(832-709,9))/(10-1)) = (((-66,4)(-308,90))+((-56,40)(-135,90))+$
$+(88,6\cdot 164,10)+(98,6\cdot 209,1)+((-61,4)(-250,90))+((-31,40)\cdot 29,1)+((-47,4)(-56,90))+$
$+(48,6\cdot 192,1)+(18,6\cdot 36,1)+(8,6\cdot 122,10))/9 = (20510,96+7664,76+14539,26+$
$+20617,26+15405,26-913,74+2697,06+9336,06+671,46+1050,06)/9=$
$= 91578,40/9 = 10175,37(7)$

Ответ: cov(zp, ks = 10175,37(7)

б) Найдите коэффициент корреляции Пирсона с помощью ковариации и среднеквадратичных отклонений двух признаков

$$D(zp) = \sum_{i=1}^{n}\frac{(zp_i - M(zp))^2}{n-1} = ((35-101,4)^2+(45-101,4)^2+(190-101,4)^2+$$
$+(200-101,4)^2+(40-101,4)^2+(70-101,4)^2+(54-101,4)^2+(150-101,4)^2+$
$+(150-101,4)^2+(110-101,4)^2)/(10-1) = (4408,96+3180,96+7849,96+9721,96+$
$+3769,96+985,96+2246,76+2361,96+345,96)/9 = 34872,44/9 = 3847,71$

$$\sigma_{zp} = \sqrt{D(zp)} = \sqrt{3847,71} = 62,2472$$

$$D(ks) = \sum_{i=1}^{n}\frac{(ks_i - M(ks))^2}{n-1} = ((401-709,90)^2+(574-709,90)^2+(874-709,90)^2+$$
$+(919-709,90)^2+(459-709,90)^2+(739-709,90)^2+(653-709,90)^2+$
$+(832-709,90)^2)/(10-1) = (95419,21+$
$+(902-709,90)^2+(746-709,90)^2+$
$+18468,81+26928,81+43722,81+62980,81+846,81+3237,61+36902,41+$
$+1303,21+14908,41)/9 = 304688,90/9 = 33854,32(2)$

$\sigma_{ks} = \sqrt{D(ks)} = \sqrt{33854,32} = 183,9955$

$$r_{гр,ks} = \frac{cov_{гр,ks}}{\sigma_{гр} \cdot \sigma_{ks}} = \frac{10175,38}{62,2472 \cdot 183,9955} = \frac{10175,38}{11453,2047} = 0,8884$$

Вывод: $r_{гр,ks} = 0,8884$; $r \to 1$, $r > \frac{1}{2}$ => величины гр и ks связаны линейно, между гр и ks сильная прямая корреляционная связь, при ↑ гр ↑ ks и при ↓ гр ↓ ks.

2) Измерены значения iQ выборки студентов, обучающихся в московских технических вузах: 131, 125, 115, 122, 131, 115, 107, 99, 125, 111. Известно, что в генеральной совокупности iQ распределён нормально. Найдите доверительный интервал для математического ожидания с надёжностью 0,95.

1) Неизвестно σ => не было значения генеральной совокупности. => критерием оценки будет критерий Стьюдента t из таблицы Стьюдента

2) Тогда ошибка среднего - это $\frac{S}{\sqrt{n}}$, где S - корень из дисперсии по выборке $S = \sqrt{D_B(x)}$, а $D_B(x)$ - несмещённая оценка дисперсии

3) Генеральная совокупность распределена нормально => можем перейти от α к $\frac{\alpha}{2}$

4) $n = 10$

5) $p = 0,95 \Rightarrow \alpha = 1 - p = 1 - 0,95 = 0,05$

$\frac{\alpha}{2} = \frac{0,05}{2} = 0,025$

6) Считаем среднюю величину по выборке:

$$\overline{X_B} = \frac{131 + 125 + 115 + 122 + 131 + 115 + 107 + 99 + 125 + 111}{10} = \frac{1181}{10} = 118,1$$

7) Ищем дисперсию $D_B(x)$, а из неё среднеквадратичное отклонение σ

$D_B(x) = \frac{\sum_{i=1}^{n}(x_i - \overline{x}_B)^2}{n-1} = ((131-118,1)^2 + (125-118,1)^2 + (115-118,1)^2 + (122-118,1)^2 + (131-118,1)^2 + (115-118,1)^2 + (107-118,1)^2 + (99-118,1)^2 + (125-118,1)^2 + (111-118,1)^2) / (10-1) = (166,41 + 47,61 + 9,61 + 15,21 + 166,41 + 9,61 + 123,21 + 364,81 + 47,61 + 50,41) / 9 = 1000,9/9 = 111,21(1)$

$\sigma = \sqrt{D_B(x)} = \sqrt{111,21} = 10,5457$

8) Находим табличное значение t Стьюдента, а для него:

$k = n - 1 = 10 - 1 = 9$

$p = 1 - \frac{\alpha}{2} = 1 - 0,025 = 0,975$

$\Rightarrow t_{\frac{\alpha}{2}} = 2,262$

9) Тогда формула для нахождения доверительного интервала:

$$\left[\bar{x}_в \mp t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right] \Rightarrow \left[118,1 - 2,262 \cdot \frac{10,5457}{\sqrt{10}} ; 118,1 + 2,262 \cdot \frac{10,5457}{\sqrt{10}}\right] \Rightarrow$$

$$\Rightarrow [118,1 - 2,262 \cdot 3,3348 ; 118,1 + 2,262 \cdot 3,3348] \Rightarrow [110,5567 ; 125,6433]$$

Вывод: [110,5567; 125,6433] - этот интервал покрывает истинное значение x с доверительной вероятностью в 95%

③ Известно, что рост футболистов в сборной распределён нормально с дисперсией генеральной совокупности равной 25 кв.см. объём выборки равен 27, среднее выборочное составляет 174,2. Найдите доверительный интервал для математического ожидания с надёжностью 0,95.

1) Дана был выполнен анализ генеральной совокупности => критерием оценки будет критерий Фишера Z из таблицы распределения Лапласа для $Z < 0$ (т.к. работаем с доверительными интервалами)

2) Генеральная совокупность распределена нормально => можем перейти от α к $\frac{\alpha}{2}$

$\alpha = 1 - p = 1 - 0,95 = 0,05$

$\frac{\alpha}{2} = \frac{0,05}{2} = 0,025$

3) Т.к. был анализ генеральной совокупности, то стандартной ошибкой среднего будет $\frac{\sigma}{\sqrt{n}}$, где σ взято из генеральной совокупности

4) Находим табличное значение Z

$Z(p_{\alpha/2} = 0,025) = -1,96$

5) Тогда формула для нахождения доверительного интервала:

$$\left[\bar{X} \mp Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] \Rightarrow \left[174,2 - 1,96 \cdot \frac{5}{\sqrt{27}} ; 174,2 + 1,96 \cdot \frac{5}{\sqrt{27}}\right] \Rightarrow$$

$$\Rightarrow [172,3140 ; 176,0860]$$

Вывод: [172,3140; 176,0860] - этот интервал с вероятностью 95% покрывает истинное мат. ожидание $M(x)$ генеральной совокупности.