

Практическое задание №6

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].

Найдите ковариацию этих двух величин с помощью элементарных действий, а затем с помощью функции cov из numpy. Полученные значения должны быть равны.

Найдите коэффициент корреляции Пирсона с помощью ковариации и среднеквадратичных отклонений двух признаков, а затем с использованием функций из библиотек numpy и pandas.

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = M(XY) - M(X)M(Y)$$

```
In [40]: import numpy as np
import pandas as pd
```

```
In [41]: zp = [35, 45, 190, 200, 40, 70, 54, 150, 120, 110]
ks = [401, 574, 874, 919, 459, 739, 653, 902, 746, 832]
```

```
In [42]: z_mean = np.mean(zp)
k_mean = np.mean(ks)
print('Среднее M(zp) =', z_mean, '\nСреднее M(ks) =', k_mean)
```

Среднее M(zp) = 101.4
Среднее M(ks) = 709.9

```
In [43]: zp_ks_mean = np.mean(np.multiply(zp, ks))
print('Среднее произведения M(zp*ks) =', zp_ks_mean)
```

Среднее произведения M(zp*ks) = 81141.7

```
In [44]: cov_0 = zp_ks_mean - z_mean*k_mean
print('Ковариация, ручной расчёт = ', cov_0)
```

Ковариация, ручной расчёт = 9157.839999999997

```
In [45]: cov_1 = np.cov(zp, ks, ddof=0)[0,1]
print('Ковариация, расчёт с помощью np.cov', cov_1)
```

Ковариация, расчёт с помощью np.cov 9157.84

$$r_{XY} = \frac{\text{cov}_{XY}}{\sigma_X * \sigma_Y}, \quad \sigma_X = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

```
In [46]: std_zp=np.sqrt(np.sum((zp-z_mean)**2)/len(zp))
print('Стандартное отклонение std(zp) =', std_zp)
```

Стандартное отклонение std(zp) = 59.115480206118605

```
In [47]: std_ks=np.sqrt(np.sum((ks-k_mean)**2)/len(ks))
print('Стандартное отклонение std(ks) =', std_ks)
```

Стандартное отклонение std(ks) = 174.55340157098058

```
In [48]: r_0 = cov_1/(std_zp*std_ks)
print('Коэффициент корреляции r =', r_0)
```

Коэффициент корреляции r = 0.8874900920739162

```
In [49]: r_1 = np.corrcoef(zp, ks)[0,1]
print('Коэффициент корреляции (вычисленный функцией np.corrcoef) r =', r_1)
```

Коэффициент корреляции (вычисленный функцией np.corrcoef) r = 0.8874900920739162

```
In [50]: r_2 = pd.Series(zp).corr(pd.Series(ks))
print('Коэффициент корреляции (вычисленный функцией pd.corr) r =', r_2)
```

Коэффициент корреляции (вычисленный функцией pd.corr) r = 0.8874900920739162

2. Измерены значения IQ выборки студентов, обучающихся в местных технических вузах: 131, 125, 115, 122, 131, 115, 107, 99, 125, 111. Известно, что в генеральной совокупности IQ распределен нормально. Найдите доверительный интервал для математического ожидания с надежностью 0.95.

$M(X) \pm t_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ – доверительный интервал. Используем распределение Стьюдента т. к. $\sigma_{Г.С.}$ не определена

$$M(X) = \frac{131 + 125 + 115 + 122 + 131 + 115 + 107 + 99 + 125 + 111}{10} = 118.1$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - M(X))^2}{n - 1}} \approx 10.55$$

$$\frac{\alpha}{2} = 2.5\%, \Rightarrow t_{\frac{\alpha}{2}} = 2.262$$

$$118.1 \pm 2.262 * \frac{10.55}{\sqrt{10}} \approx 118.1 \pm 7.55$$

[110.55, 125.65] – доверительный интервал

3. Известно, что рост футболистов в сборной распределен нормально с дисперсией генеральной совокупности, равной 25 кв.см. Объем выборки равен 27, среднее выборочное составляет 174.2. Найдите доверительный интервал для математического ожидания с надежностью 0.95.

$M(X) \pm Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ – доверительный интервал. Используем Z распределение т. к. $\sigma_{Г.С.}$ определена

$$\alpha = 100\% - 95\% = 5\%, \quad \frac{\alpha}{2} = 2.5\%$$

$$f(Z_{\frac{\alpha}{2}}) = 0.975, \Rightarrow Z_{\frac{\alpha}{2}} = 1.96$$

$$174.2 \pm 1.96 * \frac{5}{\sqrt{27}} = 174.2 \pm 1.89$$

[172.31, 176.09] – доверительный интервал