**Домашнее задание 3. EDA (exploratory data analysis) или Разведочный анализ**

**Задача 1.**

*Даны значения зарплат из выборки выпускников: 100, 80, 75, 77, 89, 33, 45, 25, 65, 17, 30, 24, 57, 55, 70, 75, 65, 84, 90, 150. Посчитать (желательно без использования статистических методов наподобие std, var, mean):*

*- среднее арифметическое,*

*- среднее квадратичное отклонение,*

*- смещенную и несмещенную оценки дисперсий для данной выборки.*

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**from** math **import** factorial

arr**=**np**.**array([100, 80, 75, 77, 89, 33, 45, 25, 65, 17, 30, 24, 57, 55, 70, 75, 65, 84, 90, 150])

*а) Найти среднее арифметическое, для этого используем формулу:*

**def** mean\_value(array):

**return** sum(array) **/** len(array)

print(f'Среднее арифметическое = {round(mean\_value(arr),2)}')

a**=**round(mean\_value(arr),2)

Среднее арифметическое = 65.3

np**.**mean(arr)

65.3

*б) Рассчитать среднее квадратичное отклонение, для этого используем формулу:*

**def** mean\_square\_deviation(array):

square\_deviation **=** (array **-** mean\_value(array)) **\*\*** 2

**return** (sum(square\_deviation) **/** (len(square\_deviation) **-** 1)) **\*\*** (1**/**2)

print(f'Среднее квадратичное отклонение = {round(mean\_square\_deviation(arr),2)}')

Среднее квадратичное отклонение = 31.62

np**.**std(arr, ddof**=**1)

31.624607341019814

*в) Рассчитать смещенную и несмещенную оценки дисперсий, для этого используем формулы:*

смещенной дисперсии:

несмещенной дисперсии:

**def** dispersion(array, offset):

square\_deviation **=** (array **-** mean\_value(array)) **\*\***2

**return** sum(square\_deviation) **/** (len(square\_deviation) **-** 1) **if** offset **else** sum(square\_deviation) **/** len(square\_deviation)

print(f'Смещенная оценка дисперсии = {round(dispersion(arr, **False**),2)}\n'

f'Немещенная оценка дисперсии = {round(dispersion(arr, **True**),2)}')

Смещенная оценка дисперсии = 950.11

Несмещенная оценка дисперсии = 1000.12

np**.**var(arr, ddof**=**0)

950.11

np**.**var(arr, ddof**=**1)

1000.1157894736842

**Задача 2.**

*В первом ящике находится 8 мячей, из которых 5 - белые. Во втором ящике - 12 мячей, из которых 5 белых. Из первого ящика вытаскивают случайным образом два мяча, из второго - 4. Какова вероятность того, что 3 мяча белые?*

Это событие может произойти следующими способами:

1. из первого ящика вытащили 0 белых мячей и из второго вытащили 3 белых мяча;

2. из первого ящика вытащили 1 белый мяч и из второго вытащили 2 белых мяча;

3. из первого ящика вытащили 2 белых мяча и из второго вытащили 1 белый мяч.

: количество извлечения из первого ящика 2-х мячей из 8:

: количество извлечения из второго ящика 4-х мячей из 12:

Вероятность события, что 3 мяча будут белыми, определяется, как сумма вероятностей этих событий:

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**from** math **import** factorial

**def** combinations(n,k):

**return** factorial(n) **/** (factorial(k) **\*** factorial(n **-** k))

1. Определим вероятность события:

P1 **=** (combinations(5,0) **\*** combinations(3,2) **/** combinations(8,2)) **\*** (combinations(5,3) **\*** combinations(7,1) **/** combinations(12,4))

print(f'Из первого ящика вытащили 0 белых мячей И из второго вытащили 3 белых мяча = {round(P1, 4)}')

Из первого ящика вытащили 0 белых мячей И из второго вытащили 3 белых мяча = 0.0152

2. Определим вероятность события:

P2 **=** (combinations(5,1) **\*** combinations(3,1) **/** combinations(8,2)) **\*** (combinations(5,2) **\*** combinations(7,2) **/** combinations(12,4))

print(f'Из первого ящика вытащили 1 белый мяч И из второго вытащили 2 белых мяча = {round(P2, 4)}')

Из первого ящика вытащили 1 белый мяч И из второго вытащили 2 белых мяча = 0.2273

3. Определим вероятность события:

P3**=**(combinations(5,2) **\*** combinations(3,0) **/** combinations(8,2)) **\*** (combinations(5,1) **\*** combinations(7,3) **/** combinations(12,4))

print(f'Из первого ящика вытащили 2 белых мяча И из второго вытащили 1 белый мяч = {round(P3, 4)}')

Из первого ящика вытащили 2 белых мяча И из второго вытащили 1 белый мяч = 0.1263

P **=** P1 **+** P2 **+** P3

print(f'Вероятность того, что 3 мяча белые Р = {round(P, 4)} = {round(P **\*** 100, 2)}%')

Вероятность того, что 3 мяча белые Р = 0.3687 = 36.87%

**Задача 3.**

*На соревновании по биатлону один из трех спортсменов стреляет и попадает в мишень. Вероятность попадания для первого спортсмена равна 0.9, для второго — 0.8, для третьего — 0.6.*

*Найти вероятность того, что выстрел произведен:*

*a) первым спортсменом*

*б) вторым спортсменом*

*в) третьим спортсменом.*

Это событие может произойти следующими способами:

1. : в результате выстрела цель поражена, выстрел произведен первым спортсменом , 1/3 так как выстрел произведен определенным спортсменом равны и вероятность попадания = 0.9;

2. : в результате выстрела цель поражена, выстрел произведен первым спортсменом , 1/3 так как выстрел произведен определенным спортсменом равны и вероятность попадания = 0.8;

3. : в результате выстрела цель поражена, выстрел произведен первым спортсменом , 1/3 так как выстрел произведен определенным спортсменом равны и вероятность попадания = 0.6;

Вероятность того, что цель поражена, определяется, как сумма вероятностей этих событий:

PBA **=** 1**/**3

PB **=** PBA **\*** 0.9 **+** PBA **\*** 0.8 **+** PBA **\*** 0.6

print(f'Вероятность того, что цель поражена P(B) = {round(PB, 4)}')

Вероятность того, что цель поражена P(B) = 0.7667

С помощью формулы Байеса оценим вероятность того, что выстрел произведен каждым спортсменом:

PAB1 = PBA \* 0.9 / PB

PAB2 = PBA \* 0.8 / PB

PAB3 = PBA \* 0.6 / PB

print(f'Вероятность того, что выстрел произвёл первый спортсмен: PAB1 = {round(PAB1, 4)} = {round(PAB1 \* 100, 2)}%;\n'

f'Вероятность того, что выстрел произвёл второй спортсмен: PAB1 = {round(PAB2, 4)} = {round(PAB2 \* 100, 2)}%;\n'

f'Вероятность того, что выстрел произвёл третий спортсмен: PAB1 = {round(PAB3, 4)} = {round(PAB3 \* 100, 2)}%.')

Вероятность того, что выстрел произвёл первый спортсмен:

PAB1 = 0.3913 = 39.13%;

Вероятность того, что выстрел произвёл второй спортсмен:

PAB1 = 0.3478 = 34.78%;

Вероятность того, что выстрел произвёл третий спортсмен:

PAB1 = 0.2609 = 26.09%.

**Задача 4.**

*В университет на факультеты A и B поступило равное количество студентов, а на факультет C студентов поступило столько же, сколько на A и B вместе. Вероятность того, что студент факультета A сдаст первую сессию, равна 0.8. Для студента факультета B эта вероятность равна 0.7, а для студента факультета C - 0.9. Студент сдал первую сессию.*

*Какова вероятность, что он учится:*

*a) на факультете A*

*б) на факультете B*

*в) на факультете C?*

Это событие может произойти следующими способами:

1. : студент поступил на факультет A=n и вероятность сдачи первой сессии = 0.8;

2. : студент поступил на факультет B=n и вероятность сдачи первой сессии = 0.7;

3. : студент поступил на факультет C=2n и вероятность сдачи первой сессии = 0.9;

События , и образуют полную вероятность сдачи первой сессии = 1, значит:

,

,

.

Вероятность того, что студент учится, определяется, как сумма вероятностей этих событий:

PB = 0.25 \* 0.8 + 0.25 \* 0.7 + 0.5 \* 0.9

print(f'Полная вероятность наступления события PВ = {round(PB, 4)}.')

Полная вероятность наступления события P(B) = 0.825

С помощью формулы Байеса оценим вероятность того, что выстрел произведен каждым спортсменом:

PAB1 = 0.25 \* 0.8 / PB

PAB2 = 0.25 \* 0.7 / PB

PAB3 = 0.5 \* 0.9 / PB

print(f'Вероятность того, что студент учится на факультете А: {round(PAB1, 4)} = {round(PAB1 \* 100, 2)}%;\n'

f'Вероятность того, что студент учится на факультете B: {round(PAB1, 4)} = {round(PAB1 \* 100, 2)}%;\n'

f'Вероятность того, что студент учится на факультете C: {round(PAB3, 4)} = {round(PAB3 \* 100, 2)}%.')

Вероятность того, что студент учится на факультете А: 0.2424 = 24.24%;

Вероятность того, что студент учится на факультете B: 0.2424 = 24.24%;

Вероятность того, что студент учится на факультете C: 0.5455 = 54.55%.

**Задача 5.**

*Устройство состоит из трех деталей. Для первой детали вероятность выйти из строя в первый месяц равна 0.1, для второй - 0.2, для третьей - 0.25.*

*Какова вероятность того, что в первый месяц выйдут из строя:*

*а) все детали*

*б) только две детали*

*в) хотя бы одна деталь*

*г) от одной до двух деталей?*

Это событие может произойти следующими способами:

1. : вышла из строя первая деталь, вероятность этого события: ;

: противоположное событию , первая деталь не вышла из строя, вероятность этого события:

;

2. : вышла из строя первая деталь, вероятность этого события: ;

: противоположное событию , вторая деталь не вышла из строя, вероятность этого события:

;

3. : вышла из строя первая деталь, вероятность этого события: ;

: противоположное событию , третья деталь не вышла из строя, вероятность этого события:

;

*а) Вероятность того, что вышли из строя все детали, произойдет если события*

*произойдут одновременно:*

PA = 0.1 \* 0.2 \* 0.25

print(f'Вероятность того, что из строя выйдут все детали Р(A) = {round(PA, 4)} = {round(PA \* 100, 2)}%')

Вероятность того, что из строя выйдут все детали Р(A) = 0.005 = 0.5%

*б) Вероятность того, что вышли из строя только две детали, если:*

Произошли события ;

Произошли события ;

Произошли события .

А вероятностью этого события будет суммой вероятностей этих комбинаций событий.

PB **=** 0.1 **\*** 0.2 **\*** 0.75 **+** 0.1 **\*** 0.25 **\*** 0.8 **+** 0.2 **\*** 0.25 **\*** 0.9

print(f'Вероятность того, что из строя выйдут только 2 детали Р(B) = {round(PB, 4)} = {round(PB **\*** 100, 2)}%')

Вероятность того, что из строя выйдут только 2 детали Р(B) = 0.08 = 8.0%

*в) Вероятность того, что вышли из строя хотя бы одна деталь, как противоположное событию не вышло из строя ни одной детали:*

Вероятность того, что не вышло из строя ни одной детали . Произойдут одновременно:

PC = 0.9 \* 0.8 \* 0.75

print(f'Вероятность того, что из строя не выйдет ни одной детали Р(C) = {round(PC, 4)} = {round(PC \* 100, 2)}%')

Вероятность того, что из строя не выйдет ни одной детали Р(C) = 0.54 = 54.0%

Вероятность того, что вышли из строя хотя бы одна деталь:

PD = 1 - PC

print(f'Вероятность того, что вышли из строя хотя бы одна деталь PD = {round(PD, 4)} = {round(PD \* 100, 2)}%')

Вероятность того, что вышли из строя хотя бы одна деталь PD = 0.46 = 46.0%

*г) Вероятность того, что вышли из строя выйдут от одной до двух деталей, если:*

Произошли события ;

Произошли события ;

Произошли события .

А вероятностью этого события будет **суммой** вероятностей этих комбинаций событий.

PE = 0.1 \* 0.8 \* 0.75 + 0.2 \* 0.9 \* 0.75 + 0.25 \* 0.9 \* 0.8

PF = PE + PB

print(f'Вероятность того, что выйдет из строя одна деталь Р(E) = {round(PF, 4)} = {round(PF \* 100, 2)}%')

Вероятность того, что выйдет из строя одна деталь Р(E) = 0.455 = 45.5%