Теория вероятностей и математическая статистика

Урок 3. Описательная статистика. Качественные и количественные характеристики популяции. Графическое представление данных

Урок 3.

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.
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?
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.
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?
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?

В [1]:
```
from math import factorial
import pandas as pd
import numpy as np
```

Число сочетаний

$$C_n^k = \frac{n!}{k!(n-k)!}$$

В [2]:
```
def combinations(n, k):
    # Число сочетаний из  n  элементов по  k  элементов в каждом (в сочетаниях порядок не важен):
    return int(factorial(n) / (factorial(k) * factorial(n - k)))
```

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.

В [3]:
```
X = [100, 80, 75, 77, 89, 33, 45, 25, 65, 17, 30, 24, 57, 55, 70, 75, 65, 84, 90, 150]

N = len(X)  # Количество элементов в выборке
X_cp = np.sum(X)/N  # Среднее арифмитическое для данной выборки
```

В [4]:
```
# 1 вариант
sum = 0
for x in X:
    sum += (x - X_cp)**2

print(sum)

# 2-й вариант
arr = np.array(X) - X_cp
sum = np.sum(arr**2)
print(sum)

# 3-й вариант
sum = np.dot(arr, arr)
print(sum)

S_cm = sum/N  # Смещенная оценка дисперсий для данной выборки.
S_ncm = sum/(N-1)  # Несмещенная оценка дисперсий для данной выборки.
```

```
19002.2
19002.2
19002.2
```

В [5]:
```
print(f'Сумма значений элементов в выборке = {np.sum(X)}')
print(f"Количество элементов в выборке = {N}\n")

print("Ответ:\n")
print(f"1. Среднее арифмитическое для данной выборки = {X_cp}")
print(f"2. Смещенная оценка дисперсий для данной выборки = {S_cm}")
print(f"3. Несмещенная оценка дисперсий для данной выборки = {S_ncm}")
print(f"4. Смещенная оценка стандартного отклонения (СТО) для данной выборки = {np.sqrt(S_cm)}")
print(f"5. Несмещенная оценка стандартного отклонения (СТО) для данной выборки = {np.sqrt(S_ncm)}")
```

```
Сумма значений элементов в выборке = 1306
Количество элементов в выборке = 20

Ответ:

1. Среднее арифмитическое для данной выборки = 65.3
2. Смещенная оценка дисперсий для данной выборки = 950.11
3. Несмещенная оценка дисперсий для данной выборки = 1000.1157894736842
4. Смещенная оценка стандартного отклонения (СТО) для данной выборки = 30.823854398825596
5. Несмещенная оценка стандартного отклонения (СТО) для данной выборки = 31.624607341019814
```

В [6]:
```
# Проверка
print('Проверка:\n')
print(f"1. Среднее арифмитическое з/п = {np.mean(X)}")
print(f"2. Смещенная оценка дисперсий для данной выборки = {np.var(X)}")
print(f"3. Несмещенная оценка дисперсий для данной выборки = {np.var(X, ddof=1)}")
print(f"4. Смещенная оценка стандартного отклонения (СТО) для данной выборки = {np.std(X)}")
print(f"5. Несмещенная оценка стандартного отклонения (СТО) для данной выборки = {np.std(X, ddof=1)}")
```

```
Проверка:

1. Среднее арифмитическое з/п = 65.3
2. Смещенная оценка дисперсий для данной выборки = 950.11
3. Несмещенная оценка дисперсий для данной выборки = 1000.1157894736842
4. Смещенная оценка стандартного отклонения (СТО) для данной выборки = 30.823854398825596
5. Несмещенная оценка стандартного отклонения (СТО) для данной выборки = 31.624607341019814
```

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?

A11- из первого ящика вынули два белых мяча

A12 - из второго ящика вынули один белый мяч

A21 - из первого ящика вынули один белый мяч

A22 - из второго ящика вынули два белых мяча

A31 - из первого ящика вынули только черные мячи

A32 - из второго ящика вынули три белый мячи

A: A11 и A12 или A21 и A22 или A31 и A32 (вынули ровно три белых мяча)

$$P(A11) = \frac{C_5^2}{C_8^2}$$

$$P(A12) = \frac{C_5^1}{C_{12}^4}$$

В [7]:
```
P_A11 = combinations(5, 2)/combinations(8, 2)
print(f"P_A11 = {P_A11*100:.1f} % ({P_A11})")
P_A12 = combinations(5, 1)/combinations(12, 4)
print(f"P_A2 = {P_A12*100:.1f} %  ({P_A12})")
```

```
P_A11 = 35.7 % (0.35714285714285715)
P_A2 = 1.0 %  (0.010101010101010102)
```

$$P(A21) = \frac{C_5^1}{C_8^2}$$

$$P(A22) = \frac{C_5^2}{C_{12}^4}$$

В [8]:
```
P_A21 = combinations(5, 1)/combinations(8, 2)
print(f"P_A21 = {P_A21*100:.1f} %  ({P_A21})")
P_A22 = combinations(5, 2)/combinations(12, 4)
print(f"P_A22 = {P_A22*100:.1f} %  ({P_A22})")
```

```
P_A21 = 17.9 %  (0.17857142857142858)
P_A22 = 2.0 %  (0.020202020202020204)
```

$$P(A31) = \frac{C_3^2}{C_8^2}$$

$$P(A32) = \frac{C_5^3}{C_{12}^4}$$

В [9]:
```python
P_A31 = combinations(3, 2)/combinations(8, 2)
print(f"P_A31 = {P_A31*100:.1f} %  ({P_A31})")
P_A32 = combinations(5, 3)/combinations(12, 4)
print(f"P_A32 = {P_A32*100:.1f} %  ({P_A32})")
```

```
P_A31 = 10.7 %  (0.10714285714285714)
P_A32 = 2.0 %  (0.020202020202020204)
```

В [10]:
```python
PA = P_A11*P_A12 + P_A21*P_A22 + P_A31*P_A32
print(f'Вероятность того, что вынуть ровно три белых мяча = {PA*100 :.3f} %  ({PA})')
```

```
Вероятность того, что вынуть ровно три белых мяча = 0.938 %  (0.00937950937950938)
```

Ответ: Вероятность того, что вынут ровно два белых мяча P(A)= 0.94%

Задача 3

На соревновании по биатлону один из трех спортсменов стреляет и попадает в мишень.

Вероятность попадания:

- для первого спортсмена равна P(A|B1)=0.9,
- для второго — P(A|B2)=0.8,
- для третьего — P(A|B3)=0.6.

Найти вероятность того, что выстрел произведен:

1. первым спортсменом
2. вторым спортсменом
3. третьим спортсменом.

- Событие A - попадание в мишень.
- Событие B1 - выстрел сделан 1-м спортсменом. $P(B1) = \frac{1}{3}$
- Событие B2 - выстрел сделан 2-м спортсменом. $P(B2) = \frac{1}{3}$
- Событие B3 - выстрел сделан 3-м спортсменом. $P(B3) = \frac{1}{3}$

Вероятность того, что выстрел произведен i-м спортсменом

$$P(Bi|A) = \frac{P(Bi)P(A|Bi)}{P(A)}$$

Вероятность попадания в мишень находим по формуле полной вероятности:

$$P(A) = \frac{1}{3}P(A|B1) + \frac{1}{3}P(A|B2) + \frac{1}{3}P(A|B3)$$

В [11]:
```python
P_A_B1 = 0.9
P_A_B2 = 0.8
P_A_B3 = 0.6
P_A = (0.9 + 0.8 + 0.6)/3
P_B1 = 1/3
P_B2 = 1/3
P_B3 = 1/3
P_A
```

Out[11]:
```
0.7666666666666667
```

В [12]:
```python
P_B1_A = (P_B1*P_A_B1)/P_A
print(f'Вероятность того, что выстрел произведен 1-м спортсменом = {P_B1_A*100:.1f} % ({P_B1_A})')
```

```
Вероятность того, что выстрел произведен 1-м спортсменом = 39.1 % (0.3913043478260869)
```

В [13]:
```python
P_B2_A = (P_B2*P_A_B2)/P_A
print(f'Вероятность того, что выстрел произведен 2-м спортсменом = {P_B2_A*100:.1f} % ({P_B2_A})')
```

```
Вероятность того, что выстрел произведен 2-м спортсменом = 34.8 % (0.34782608695652173)
```

В [14]:
```python
P_B3_A = (P_B3*P_A_B3)/P_A
print(f'Вероятность того, что выстрел произведен 2-м спортсменом = {P_B3_A*100:.1f} % ({P_B3_A})')
```

```
Вероятность того, что выстрел произведен 2-м спортсменом = 26.1 % (0.26086956521739124)
```

В [15]:
```python
# print(P_B1_A + P_B2_A + P_B3_A)
```

Ответ:

- Вероятность того, что выстрел произведен 1-м спортсменом = 39.1%
- Вероятность того, что выстрел произведен 2-м спортсменом = 34.8%
- Вероятность того, что выстрел произведен 2-м спортсменом = 26.1%

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 $P(A|B1) = 0.8$.

- Для студента факультета B эта вероятность равна 0.7 $P(A|B2) = 0.7$,
- а для студента факультета C - 0.9 $P(A|B3) = 0.9$.

Студент сдал первую сессию. Какова вероятность, что он учится:

- (а) на факультете A?
- (б) на факультете B?
- (в) на факультете C?

- Событие A - сдал сессию.
- Событие B1 - сессию сдавал студент факультета A.
$$P(B1) = \frac{1}{4}$$
- Событие B2 - сессию сдавал студент факультета B.
$$P(B2) = \frac{1}{4}$$
- Событие B3 - сессию сдавал студент факультета C.
$$P(B3) = \frac{2}{4} = \frac{1}{2}$$

В [16]:
```
P_A_B1 = 0.8
P_A_B2 = 0.7
P_A_B3 = 0.9
P_A = (0.8 + 0.7 + 2*0.9)/4
P_B1 = 1/4
P_B2 = 1/4
P_B3 = 1/2
P_A
```

Out[16]:
```
0.825
```

В [17]:
```
P_B1_A = (P_B1*P_A_B1)/P_A
print(f'Вероятность того, что сессию сдал студент факультета A = {P_B1_A*100:.2f} % ({P_B1_A})')
```

```
Вероятность того, что сессию сдал студент факультета A = 24.24 % (0.24242424242424246)
```

В [18]:
```
P_B2_A = (P_B2*P_A_B2)/P_A
print(f'Вероятность того, что сессию сдал студент факультета B = {P_B2_A*100:.2f} % ({P_B2_A})')
```

```
Вероятность того, что сессию сдал студент факультета B = 21.21 % (0.21212121212121213)
```

В [19]:
```
P_B3_A = (P_B3*P_A_B3)/P_A
print(f'Вероятность того, что сессию сдал студент факультета C = {P_B3_A*100:.2f} % ({P_B3_A})')
```

```
Вероятность того, что сессию сдал студент факультета C = 54.55 % (0.5454545454545455)
```

В [20]:
```
print(P_B1_A + P_B2_A + P_B3_A)
```

```
1.0
```

Ответ:

- Вероятность того, что сессию сдал студент факультета A = 24.24%
- Вероятность того, что сессию сдал студент факультета B = 21.21%
- Вероятность того, что сессию сдал студент факультета C = 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?

- Событие B1 - из строя вышла 1-я деталь. $P(B_1) = 0.1$
- Событие B2 - из строя вышла 2-я деталь. $P(B_2) = 0.2$
- Событие B3 - из строя вышла 3-я деталь. $P(B_3) = 0.25$

A1: не B1 и не B2 и не B3 - **ни одна деталь не вышла из строя**

$$P(A_1) = (1 - P(B_1))(1 - P(B_2))(1 - P(B_3)) = 0.9 * 0.8 * 0.75 = 0.54 = 54\%$$

A2: (B1 и не B2 и не B3) или (B2 и не B1 и не B3) или (B3 и не B1 и не B2) - **из строя вышла только одна деталь**

$$P(A_2) = P(B_1)(1 - P(B_2))(1 - P(B_3)) + P(B_2)(1 - P(B_1))(1 - P(B_3)) + P(B_3)(1 - P(B_1))(1 - P(B_2)) = 0.1 * 0.8 * 0.75 + 0.2 * 0.9 * 0.75 + 0.25 * 0.9 * 0.8 = 0.375 = 37.5\%$$

A3: (B1 и B2 и не B3) или (B1 и B3 и не B2) или (B2 и B3 и не B1) - **из строя вышли две детали**

$$P(A_3) = P(B_1)P(B_2)(1 - P(B_3)) + P(B_1)P(B_3)(1 - P(B_2)) + P(B_2)P(B_3)(1 - P(B_1)) = 0.1 * 0.2 * 0.75 + 0.1 * 0.25 * 0.8 + 0.2 * 0.25 * 0.9 = 0.08 = 8.\%$$

A4: B1 и B2 и B3 - **из строя вышли 3 детали**

$$P(A_4) = P(B_1)P(B_2)P(B_3) = 0.1 * 0.2 * 0.25 = 0.005 = 0.5\%$$

A5: A3 или A4 - **из строя вышли 1 или 2 детали (от одной до двух деталей)**

$$P(A_5) = P(A_3) + P(A_4) = 0.375 + 0.08 = 0.455 = 45.5\%$$

A6: A2 или A3 или A4 - **из строя вышли 1 или 2 или 3 детали (из строя вышла хотя бы одна деталь)**

$$P(A_6) = P(A_2) + P(A_3) + P(A_4) = 0.375 + 0.08 + 0.05 = 0.46 = 46\%$$

В [21]:
```
# A1: ни одна деталь не вышла из строя
P_A1=(1-0.1)*(1-0.2)*(1-0.25)
print(f'Вероятность того, что в первый месяц ни одна деталь не вышла из строя = {P_A1*100: 0.1f}% ({P_A1})')
```

```
Вероятность того, что в первый месяц ни одна деталь не вышла из строя =  54.0% (0.54)
```

В [22]:
```
# A2: из строя вышла только одна деталь
P_A2=0.1*0.8*0.75 + 0.2*0.9*0.75 + 0.25*0.9*0.8
print(f'Вероятность того, что в первый месяц из строя вышла только одна деталь = {P_A2*100: 0.1f}% ({P_A2})')
```

```
Вероятность того, что в первый месяц из строя вышла только одна деталь =  37.5% (0.375)
```

В [23]:
```
# A3: из строя вышли две детали
P_A3=0.1*0.2*0.75+0.1*0.25*0.8+0.2*0.25*0.9
print(f'Вероятность того, что в первый месяц из строя вышли 2 детали = {P_A3*100: 0.1f}% ({P_A3})')
```

```
Вероятность того, что в первый месяц из строя вышли 2 детали =  8.0% (0.08000000000000002)
```

В [24]:
```
# A4: из строя вышли 3 детали
P_A4=0.1*0.2*0.25
print(f'Вероятность того, что в первый месяц из строя вышли 3 детали = {P_A4*100: 0.1f}% ({P_A4})')
```

```
Вероятность того, что в первый месяц из строя вышли 3 детали =  0.5% (0.005000000000000001)
```

В [25]:
```
# A5: из строя вышли 1 или 2 детали (от одной до двух деталей)
P_A5=P_A2 + P_A3
print(f'Вероятность того, что в первый месяц из строя вышли 1 или 2 детали = {P_A5*100: 0.1f}% ({P_A5})')
```

```
Вероятность того, что в первый месяц из строя вышли 1 или 2 детали =  45.5% (0.455)
```

В [26]:
```
# A6: из строя вышла хотя бы одна деталь
P = P_A2+P_A3+P_A4
# или P = 1 - P_A1
print(f'Вероятность того, что в первый месяц из строя вышла хотя бы одна деталь = {P*100: 0.1f}% ({P})')
```

```
Вероятность того, что в первый месяц из строя вышла хотя бы одна деталь =  46.0% (0.46)
```

В [27]:
```
P = P_A1 + P_A2 + P_A3 + P_A4
print(f'Полная вероятность = {P*100: 0.1f}% ({P})')
```

```
Полная вероятность =  100.0% (1.0)
```