Практическое задание 5

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import warnings
import scipy
from scipy import stats
warnings.filterwarnings('ignore')
```

1. Известно, что генеральная совокупность распределена нормально со средним квадратическим отклонением, равным 16 Найти доверительный интервал для оценки математического ожидания с надежностью 0.95, если выборочная средняя M = 80, а объем выборки n = 256.

Поскольку мы имеем дело с нормальным распределением с известным среднеквадратичным отклонением, то воспользуемся формулой:

$$\overline{X} \pm Z_{\frac{a}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Все данные у нас есть, кроме Z-значения. Чтобы долго не искать его в таблице, найдем его используя средства питона.

In [2]:

```
z_value = stats.norm.ppf(q = 0.975)
z_value
```

Out[2]:

```
1.959963984540054
```

In [3]:

```
interval = z_value*(16/np.sqrt(256))
interval
```

Out[3]:

```
1.959963984540054
```

In [4]:

```
print("Доверительный интервал: ", (80 - interval, 80 + interval))
```

```
Доверительный интервал:  (78.04003601545995, 81.95996398454005)
```

2. В результате 10 независимых измерений некоторой величины X, выполненных с одинаковой точностью, получены опытные данные: 6.9, 6.1, 6.2, 6.8, 7.5, 6.3, 6.4, 6.9, 6.7, 6.1 Предполагая, что

результаты измерений подчинены нормальному закону распределения вероятностей, оценить истинное значение величины X при помощи доверительного интервала, покрывающего это значение с доверительной вероятностью 0,95.

Поскольку СКО генеральной свокупности нам не известно, воспользуемся критерием Стьюдента. Формула поиска ловерительного интервала немного изменится:

$$\overline{X} \pm t_{\frac{a}{2}} \cdot \frac{S}{\sqrt{n}}$$

Найдем $\overline{X}$, S и n из выборки

In [5]:

```
selection = pd.DataFrame([6.9, 6.1, 6.2, 6.8, 7.5, 6.3, 6.4, 6.9, 6.7, 6.1])
```

In [6]:

```
S = np.sqrt(((selection[0] - selection[0].mean())**2).sum() / (selection[0].count() - 1
S
```

Out[6]:

```
0.4508017549014448
```

In [7]:

```
n = selection[0].count()
n
```

Out[7]:

```
10
```

In [8]:

```
X = selection[0].mean()
X
```

Out[8]:

```
6.589999999999999
```

Все данные у нас есть, кроме t-критерия. Чтобы долго не искать его в таблице, найдем его используя средства питона.

In [9]:

```
t_value = scipy.stats.t.ppf(0.975, 10-1)
t_value
```

Out[9]:

```
2.2621571627409915
```

In [10]:

```
interval = t_value*(S/np.sqrt(10))
interval
```

Out[10]:

```
0.3224841485842873
```

In [11]:

```
print("Доверительный интервал: ", (X - interval, X + interval))
```

```
Доверительный интервал:  (6.267515851415712, 6.912484148584286)
```

3. Утверждается, что шарики для подшипников, изготовленные автоматическим станком, имеют средний диаметр 17 мм. Используя односторонний критерий с α=0,05, проверить эту гипотезу, если в выборке из n=100 шариков средний диаметр оказался равным 17.5 мм, а дисперсия известна и равна 4 кв.мм.

3.1. Формулировка основной и альтернативной гипотез:

H_0 - Диаметр шариков в выборке = 17 мм
H_1 - Диаметр шариков в выборке > 17 мм

3.2. Выбор статистического критерия, на основе которого будет проводиться проверка

Поскольку нам известно СКО генеральной совокупности, в качестве критерия выбираем Z-критерий

3.3. Выбор уровня значимости α

Для нашей задачи вполне достаточно занчение уровня значимости равного 0,05

3.4 Определение границ области данной гипотезы

In [12]:

```
Z = stats.norm.ppf(q = 0.95)
Z
```

Out[12]:

```
1.6448536269514722
```

In [13]:

```
Z_n = (17.5-17)/2*10
Z_n
```

Out[13]:

```
2.5
```

5.1 Подведение итогов и формулировка вывода

Z_h > Z, следовательно мы отвергаем нулевую гипотезу. Т.е. в выборке диаметр шариков достоверно, на уровне значимости 95%, больше 17 мм.

4. Продавец утверждает, что средний вес пачки печенья составляет 200 г. Из партии извлечена выборка из 10 пачек. Вес каждой пачки составляет: 202, 203, 199, 197, 195, 201, 200, 204, 194, 190. Известно, что их веса распределены нормально. Верно ли утверждение продавца, если учитывать, что доверительная вероятность равна 99%?

Найдем $\overline{X}$, S и n из выборки

In [14]:

```
selection = pd.DataFrame([202, 203, 199, 197, 195, 201, 200, 204, 194, 190])
```

In [15]:

```
S = np.sqrt(((selection[0] - selection[0].mean())**2).sum() / (selection[0].count() - 1
S
```

Out[15]:

```
4.453463071962462
```

In [16]:

```
n = selection[0].count()
n
```

Out[16]:

```
10
```

In [17]:

```
X = selection[0].mean()
X
```

Out[17]:

```
198.5
```

4.1. Формулировка основной и альтернативной гипотез:

H_0 - Вес пачки = 200 г
H_1 - Вес пачки не равен 200 г (больше или меньше)

4.2. Выбор статистического критерия, на основе которого будет проводиться проверка

Поскольку нам не известно СКО генеральной совокупности, в качестве критерия выбираем t-критерий

4.3. Выбор уровня значимости α

Для нашей задачи вполне достаточно занчение уровня значимости равного 0,01 (по 0,005 с каждой стороны)

4.4 Определение границ области данной гипотезы

In [18]:

```
t_value = scipy.stats.t.ppf(0.995, 10-1)
t_value
```

Out[18]:

```
3.2498355440153697
```

In [19]:

```
t_value_n = (200-X)/S*np.sqrt(n)
t_value_n
```

Out[19]:

```
1.0651074037450896
```

In [20]:

```
t_value_2 = scipy.stats.t.ppf(1 - 0.995, 10-1)
t_value_2
```

Out[20]:

```
-3.2498355440153697
```

In [21]:

```
t_value_n_2 = (X-200)/S*np.sqrt(n)
t_value_n_2
```

Out[21]:

```
-1.0651074037450896
```

4.5 Подведение итогов и формулировка вывода

t_value_n < t_value, а t_value_n_2 > t_value_2, следовательно мы подтверждаем нулевую гипотезу. Т.е. средний вес пачки печенья составляет 200 г.