Автор: Анна Смелова

Решение домашнего задания к уроку 5 “Проверка статистических гипотез. P-значения. Доверительные интервалы. A/B-тестирование”

In [1]:
```
import numpy as np
```

1. Задача

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

Решение:

Теория:

$T_{1,2} = \overline{X} \pm \frac{s_0}{\sqrt{n}} \cdot c_\gamma$

где T_1, T_2 — нижняя и верхняя границы доверительного интервала,

$\overline{X}$ — выборочное среднее арифметическое,

s_0 — среднее квадратичное отклонение по выборке (несмещенное),

n — размер выборки,

γ — доверительная вероятность.

$c_\gamma = \Phi^{-1}\frac{(1+\gamma)}{2}$ — обратное значение функции стандартного нормального распределения. То есть это количество стандартных ошибок от среднего арифметического до нижней или верхней границы.

В задаче:

$\overline{X} = M = 80$

$s_0 = \sigma = 16$

$n = 256$

$\gamma = 0.95$

95% всех выборочных средних лежат в диапазоне $\mu \pm 1.96\sigma \Rightarrow c_\gamma = 1.96$

$T_1 = 80 - 1.96 \cdot \frac{16}{\sqrt{256}} = 80 - 1.96 = 78.04$

$T_2 = 80 + 1.96 \cdot \frac{16}{\sqrt{256}} = 80 + 1.96 = 81.96$

Ответ: [78.04, 81.96].

2. Задача

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.

Решение:

Выборочное среднее $\overline{X} = \frac{6.9+6.1+6.2+6.8+7.5+6.3+6.4+6.9+6.7+6.1}{10} = 6.59$

In [2]:
```
a = [6.9, 6.1, 6.2, 6.8, 7.5, 6.3, 6.4, 6.9, 6.7, 6.1]
print(f'X={np.mean(a)}')
```

```
X=6.590000000000001
```

Среднее квадратическое отклонение σ:

In [3]:
```
print(f'sigma={np.std(a, ddof=1)}')
```

```
sigma=0.4508017549014448
```

σ генеральной совокупности неизвестна, поэтому используем t-критерий Стьюдента:

$T_1 = 6.59 - 2.262 \cdot \frac{0.4508}{\sqrt{10}} = 6.2675$

$T_2 = 6.59 + 2.262 \cdot \frac{0.4508}{\sqrt{10}} = 6.9125$

In [4]:
```
T1 = 6.59 - 2.262*0.4508 / (10)**0.5
T2 = 6.59 + 2.262*0.4508 / (10)**0.5
print(f'T1={T1}')
print(f'T2={T2}')
```

```
T1=6.267539511206077
T2=6.912460488793923
```

Ответ: истинное значение величины X с вероятностью 95% лежит в диапазоне $[6.2675, 6.9125]$.

3. Задача (решать через тестирование гипотезы)

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.

Решение:

$H_0 : \mu = 17$

$H_1 : \mu > 17$

$k_p = \frac{17.5-17}{\sqrt{4}} \cdot \sqrt{100} = \frac{5}{2} = 2.5$

$k_t = 1.645$ для $\alpha = 0.05$

Так как $k_p > k_t$, то можем отклонить нулевую гипотезу.

Ответ: исходная гипотеза неверна.

4. Задача (решать через тестирование гипотезы)

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%?

Решение:

In [5]:
```
a = [202, 203, 199, 197, 195, 201, 200, 204, 194, 190]
print(f'Выборочное среднее: {np.mean(a)}')
print(f'Несмещенное среднее квадратическое отклонение: {np.std(a, ddof=1)}')
```

```
Выборочное среднее: 198.5
Несмещенное среднее квадратическое отклонение: 4.453463071962462
```

$H_0 : \mu = 200$

$H_1 : \mu \neq 200$

$k_p = \frac{198.5-200}{4.45} \cdot \sqrt{10} = -1.0659$

По таблице критических точек распределения Стьюдента найдем критическую точку по уровню значимости $\alpha = 0,01$ и числу степеней свободы 9, откуда $k_t = 3,25$.

Так как $|k_p| < k_t$, то нулевую гипотезу можно принять.

In [6]:
```
print(f'k_p={(198.5-200)/4.45*(10)**0.5}')
```

```
k_p=-1.0659362899443976
```

Ответ: утверждение продавца верно.