Теория вероятностей и математическая статистика

Задача 1.

Проведите тест 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.

Подсказка № 1

Определите нулевую гипотезу H0 и альтернативную гипотезу HA. В этом случае нулевая гипотеза H0 утверждает, что средний диаметр μ равен 17 мм (то есть μ=17). Альтернативная гипотеза HA утверждает, что средний диаметр отличается от 17 мм (или больше 17 мм, в зависимости от теста).

Подсказка № 2

Z-статистика используется для проверки гипотезы, когда дисперсия известна. Формула для Z-статистики: Z= $\frac{\bar{x}-\mu 0}{\sigma/\sqrt{n}}$ $\bar{x}$ — среднее значение выборки, μ0 — среднее значение в нулевой гипотезе, σ — стандартное отклонение (квадратный корень из дисперсии), и n — размер выборки.

Подсказка № 3

Для одностороннего теста с уровнем значимости α=0.05, критическое значение Z можно найти в таблице стандартного нормального распределения или использовать функции в Python. Для уровня значимости 0.05 в одностороннем тесте критическое значение Z примерно равно 1.645.

Подсказка № 4

Сравните рассчитанную Z-статистику с критическим значением. Если Z-статистика больше критического значения, это означает, что результат статистически значим и нулевая гипотеза отвергается. В противном случае, недостаточно оснований для её отвергнуть.

Эталонное решение:

```
mu_0 = 17

x_bar = 17.5
```

```
sigma = math.sqrt(4)

n = 100

alpha = 0.05


# Z-статистика

Z = (x_bar - mu_0) / (sigma / math.sqrt(n))


# Критическое значение для одностороннего теста с α = 0.05

Z_crit = 1.645  # Для одностороннего теста при α = 0.05


print("Задача 1: Проверка гипотезы с использованием Z-критерия")

print(f"Z-статистика: {Z}")

print(f"Критическое значение: {Z_crit}")


# Проверка гипотезы

if Z > Z_crit:

    print("Отвергаем нулевую гипотезу в пользу альтернативной гипотезы.")

else:

    print("Нет оснований отвергать нулевую гипотезу.")

print()
```

Задача 2.

Проведите тест 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%? (Провести двусторонний тест.)

Подсказка № 1

Определите нулевую гипотезу H0 и альтернативную гипотезу HA. В этом случае нулевая гипотеза утверждает, что средний вес пачки печенья равен 200 г (то есть μ=200). Альтернативная гипотеза утверждает, что средний вес отличается от 200 г (то есть μ≠200).

Подсказка № 2

Рассчитайте среднее значение выборки и выборочное стандартное отклонение. Среднее значение можно найти как сумму всех значений в выборке, деленную на количество значений. Выборочное стандартное отклонение sss можно найти как квадратный корень из выборочной дисперсии, где выборочная дисперсия рассчитывается как сумма квадратов отклонений от среднего, деленная на n−1, где n — размер выборки.

Подсказка № 3

Рассчитайте t-статистику. Используйте следующую формулу для расчета t-статистики: $t=\frac{\bar{x}-\mu 0}{s/\sqrt{n}}$, где $\bar{x}$ — среднее значение выборки, μ0 — среднее значение в нулевой гипотезе (200 г), s — выборочное стандартное отклонение, и n — размер выборки.

Подсказка № 4

Для двустороннего теста с уровнем значимости α=0.01 и n−1 степенями свободы найдите критическое значение t. В Python можно использовать функцию `scipy.stats.t.ppf` для вычисления критического значения. Для уровня значимости 0.01 и 9 степеней свободы критическое значение t примерно равно 3.2498.

Эталонное решение:

```
weights = [202, 203, 199, 197, 195, 201, 200, 204, 194, 190]

mu_0 = 200

alpha = 0.01

n = len(weights)

mean_weight = sum(weights) / n

s = math.sqrt(sum((x - mean_weight) ** 2 for x in weights) / (n -
1))


# t-статистика

t_stat = (mean_weight - mu_0) / (s / math.sqrt(n))
```

```
# Критическое значение для t-распределения с n-1 степенью свободы
t_crit = 3.249835  # Для двустороннего теста при α = 0.01 и df = 9

print("Задача 2: Проверка гипотезы с использованием t-теста")
print(f"t-статистика: {t_stat}")
print(f"Критическое значение: {t_crit}")

# Проверка гипотезы
if abs(t_stat) > t_crit:
    print("Отвергаем нулевую гипотезу в пользу альтернативной гипотезы.")
else:
    print("Нет оснований отвергать нулевую гипотезу.")
print()
```

Задача 3.

Есть ли статистически значимые различия в росте дочерей? Рост матерей 172, 177, 158, 170, 178,175, 164, 160, 169, 165 Рост взрослых дочерей: 173, 175, 162, 174, 175, 168, 155, 170, 160

Подсказка № 1

Сформулируйте нулевую гипотезу H0 и альтернативную гипотезу HA. В данном случае нулевая гипотеза утверждает, что средний рост матерей и дочерей не отличается (то есть $\mu_1=\mu_2$). Альтернативная гипотеза утверждает, что средний рост отличается (то есть $\mu_1\neq\mu_2$).

Подсказка № 2

Найдите среднее значение и дисперсию для роста матерей и дочерей. Среднее значение можно найти как сумму всех значений в выборке, деленную на количество значений. Дисперсию можно рассчитать как сумму квадратов отклонений от среднего, деленную на n−1, где n — размер выборки.

Подсказка № 3

Используйте формулу для объединённой дисперсии:

pooled_variance=$\frac{(n1-1)\cdot var1+(n2-1)\cdot var2}{n1+n2-2}$
где n1 и n2 — размеры выборок, а var1 и var2 — дисперсии в группах.

Подсказка № 4

Используйте формулу для t-статистики для независимых выборок:
t=$\frac{\bar{x}1-\bar{x}2}{\sqrt{pooled\_variance*(\frac{1}{n1}+\frac{1}{n2})}}$
где x̄1 и x̄2 — средние значения в группах, и n1 и n2 — размеры выборок.

Подсказка № 5

Для уровня значимости α=0.05\alpha = 0.05α=0.05 и степени свободы n1+n2−2 найдите критическое значение t. В Python можно использовать функцию `scipy.stats.t.ppf` для вычисления критического значения. Сравните модуль t-статистики с критическим значением. Если модуль t-статистики больше критического значения, это указывает на статистически значимые различия.

Эталонное решение:

```
mothers_height = [172, 177, 158, 170, 178, 175, 164, 160, 169, 165]

daughters_height = [173, 175, 162, 174, 175, 168, 155, 170, 160]



n1 = len(mothers_height)

n2 = len(daughters_height)

mean_mothers = sum(mothers_height) / n1

mean_daughters = sum(daughters_height) / n2

var_mothers = sum((x - mean_mothers) ** 2 for x in mothers_height) /
(n1 - 1)

var_daughters = sum((x - mean_daughters) ** 2 for x in
daughters_height) / (n2 - 1)



# t-статистика для независимых выборок
```

```
pooled_variance = ((n1 - 1) * var_mothers + (n2 - 1) *
var_daughters) / (n1 + n2 - 2)

t_stat_ind = (mean_mothers - mean_daughters) /
math.sqrt(pooled_variance * (1 / n1 + 1 / n2))


# Критическое значение для двустороннего теста при α = 0.05 и df =
18

t_crit_ind = 2.100922


print("Задача 3: Статистически значимые различия в росте матерей и
дочерей")

print(f"t-статистика: {t_stat_ind}")

print(f"Критическое значение: {t_crit_ind}")


if abs(t_stat_ind) > t_crit_ind:

    print("Есть статистически значимые различия в росте матерей и
дочерей.")

else:

    print("Нет статистически значимых различий в росте матерей и
дочерей.")

print()
```