Лабораторная работа №1(исправленная)

Выполнил: Величко Максим Иванович, М32061

Варианты для задания: 3,2.

№1

Пусть выборка 𝑋1, . . ., 𝑋𝑛 соответствует классу распределений 𝐹𝜃,

𝜃 ∈ 𝐸 ⊂ R. При каком минимальном объеме выборки 𝑛 равномерно для 𝜃 ∈ 𝐸 выборочное

среднее отличается от математического ожидания 𝜇𝜃 не более чем на 𝜀 > 0 с вероятностью,

не меньшей 1 − 𝛿, 𝛿 ∈ (0, 1)?

Пусть 𝑋1, . . ., 𝑋𝑛 — выборка из распределения 𝐹𝜃, 𝜃 ∈ 𝐸 ⊂ R, с математическим ожиданием 𝜇𝜃 и дисперсией 𝜎^2(𝜃). Обозначим выборочное среднее как 𝑋̄ = (𝑋1 + ⋯ + 𝑋𝑛)/𝑛.

Используя центральную предельную теорему, можно оценить вероятность того, что выборочное среднее отклоняется от математического ожидания не более чем на 𝜀 > 0:

P(|𝑋̄ − 𝜇𝜃| ≤ 𝜀) = P(|(𝑋̄ − 𝜇𝜃)/𝜎(𝜃)𝑛^0.5| ≤ 𝜀/𝜎(𝜃)𝑛^0.5) ≈ Φ(𝜀/𝜎(𝜃)𝑛^0.5) − Φ(−𝜀/𝜎(𝜃)𝑛^0.5)

где Φ(·) обозначает функцию распределения стандартного нормального распределения.

Для достижения заданной вероятности 1 − 𝛿, нужно выбрать минимальный объем выборки 𝑛 такой, чтобы выполнялось:

Φ(𝜀/𝜎(𝜃)𝑛^0.5) − Φ(−𝜀/𝜎(𝜃)𝑛^0.5) ≥ 1 − 𝛿

Так как функция распределения стандартного нормального распределения является симметричной относительно нуля, можно записать:

2Φ(𝜀/𝜎(𝜃)𝑛^0.5) − 1 ≥ 1 − 𝛿

Φ(𝜀/𝜎(𝜃)𝑛^0.5) ≥ (1 + 𝛿)/2

𝜀/𝜎(𝜃)𝑛^0.5 ≥ Φ^−1((1 + 𝛿)/2)

𝑛 ≥ (𝜎(𝜃)𝜀/Φ^−1((1 + 𝛿)/2)) ^2

Таким образом, чтобы выборочное среднее отличалось от математического ожидания не более чем на 𝜀 с вероятностью не меньшей 1 − 𝛿, необходимо выбрать выборку размера не менее (𝜎(𝜃)𝜀/Φ^−1((1 + 𝛿)/2)) ^2.

Сгенерировать 500 выборок найденного объема при 𝜀 = 0.01 и

𝛿 = 0.05 из указанного распределения 𝐹𝜃 при конкретном параметре 𝜃 и посчитать, сколько

раз выборочное среднее отличается от математического ожидания 𝜇𝜃 более чем на 𝜀. Geom(𝑝) (указать вид используемой параметризации), 𝑝 ∈ (1/4, 1), 𝑝 = 4/5.

Вот переписанный код, решающий данную задачу через ЦПТ:

import numpy as np  
from scipy.stats import norm, geom  
  
# Задаем параметры распределения Geom(p)  
p = 4/5  
  
# Рассчитываем объем выборки  
epsilon = 0.01  
delta = 0.05  
n = np.ceil((norm.ppf(1 - delta/2)\*\*2 \* p \* (1-p)) / epsilon\*\*2).astype(int)  
  
# Генерируем 500 выборок и вычисляем выборочное среднее  
samples = geom.rvs(p, size=(500, n))  
sample\_means = np.mean(samples, axis=1)  
  
# Вычисляем математическое ожидание  
mu = geom.mean(p)  
  
# Определяем, сколько раз выборочное среднее отличается от математического ожидания на более чем epsilon  
count = np.sum(np.abs(sample\_means - mu) > epsilon)  
  
# Выводим результат  
print("Количество раз, когда выборочное среднее отличается от математического ожидания на более чем", epsilon, ":", count)

Результат нашей программы:



№2

В файле sex\_bmi\_smokers.csv приведены данные (пол, ИМТ, курит/не курит) о более

1000 испытуемых. Сравните количество курящих мужчин и некурящих женщин. Рас-

считайте выборочное среднее, выборочную дисперсию, выборочную медиану и выборочную квантиль порядка 3/5 ИМТ всех наблюдателей и отдельно для каждой возможной

комбинации пол-курение. Построить график эмпирической функции распределения, гистограмму и box-plot ИМТ для всех наблюдателей.

import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
# Чтение данных из файла  
df = pd.read\_csv('D:\MatStat\_lab1\sex\_bmi\_smokers.csv', sep=';', header=1)  
  
# Сравнение количества курящих мужчин и некурящих женщин  
smokers = df[df['smoker'] == 'yes']  
male\_smokers = smokers[smokers['sex'] == 'male']  
female\_non\_smokers = df[(df['smoker'] == 'no') & (df['sex'] == 'female')]  
  
print('Количество курящих мужчин:', len(male\_smokers))  
print('Количество некурящих женщин:', len(female\_non\_smokers))  
  
# Расчет статистик для ИМТ  
bmi\_mean = df['bmi'].mean()  
bmi\_var = df['bmi'].var()  
bmi\_median = df['bmi'].median()  
bmi\_q35 = df['bmi'].quantile(0.6)  
  
male\_smokers\_bmi = male\_smokers['bmi']  
male\_non\_smokers\_bmi = df[(df['smoker'] == 'no') & (df['sex'] == 'male')]['bmi']  
female\_smokers\_bmi = smokers[smokers['sex'] == 'female']['bmi']  
female\_non\_smokers\_bmi = female\_non\_smokers['bmi']  
  
# Построение графиков  
fig, axs = plt.subplots(2, 2, figsize=(10, 8))  
plt.subplots\_adjust(hspace=0.5)  
  
# Эмпирическая функция распределения  
axs[0, 0].hist(df['bmi'], cumulative=True, density=True, bins=30)  
axs[0, 0].set\_title('Эмпирическая функция распределения ИМТ')  
  
# Гистограмма ИМТ  
axs[0, 1].hist(df['bmi'], density=True, bins=30)  
axs[0, 1].set\_title('Гистограмма ИМТ')  
  
# Box-plot ИМТ  
axs[1, 0].boxplot([male\_smokers\_bmi, male\_non\_smokers\_bmi, female\_smokers\_bmi, female\_non\_smokers\_bmi])  
axs[1, 0].set\_xticklabels(['Курящие мужчины', 'Некурящие мужчины', 'Курящие женщины', 'Некурящие женщины'])  
axs[1, 0].set\_title('Box-plot ИМТ')  
  
# Графики ИМТ для каждой комбинации пол-курение  
axs[1, 1].hist([male\_smokers\_bmi, male\_non\_smokers\_bmi, female\_smokers\_bmi, female\_non\_smokers\_bmi], density=True, bins=30, label=['Курящие мужчины', 'Некурящие мужчины', 'Курящие женщины', 'Некурящие женщины'])  
axs[1, 1].legend(loc='upper right')  
axs[1, 1].set\_title('Графики ИМТ для каждой комбинации пол-ку')  
  
# Выборочное среднее, дисперсия, медиана и квантиль порядка 3/5  
print("Выборочное среднее:", df["bmi"].mean())  
print("Выборочная дисперсия:", df["bmi"].var())  
print("Выборочная медиана:", df["bmi"].median())  
print("Выборочная квантиль порядка 3/5:", df["bmi"].quantile(3/5))  
  
# Выборочные характеристики по полу и курению  
grouped = df.groupby(["sex", "smoker"])["bmi"]  
for group, values in grouped:  
 print(f"\nГруппа {group}")  
 print("Выборочное среднее:", values.mean())  
 print("Выборочная дисперсия:", values.var())  
 print("Выборочная медиана:", values.median())  
 print("Выборочная квантиль порядка 3/5:", values.quantile(3/5))  
  
# График эмпирической функции распределения  
plt.figure(figsize=(8, 6))  
plt.step(sorted(df["bmi"]), np.arange(len(df))/len(df))  
plt.xlabel("ИМТ")  
plt.ylabel("Эмпирическая функция распределения")  
plt.show()  
  
# Гистограмма ИМТ для всех наблюдателей  
plt.figure(figsize=(8, 6))  
plt.hist(df["bmi"], bins=20)  
plt.xlabel("ИМТ")  
plt.ylabel("Частота")  
plt.show()  
  
# Box-plot ИМТ для всех наблюдателей  
plt.figure(figsize=(8, 6))  
plt.boxplot(df["bmi"])  
plt.xlabel("Все наблюдатели")  
plt.ylabel("ИМТ")  
plt.show()

Вот результаты:

Изображение выглядит как текст

Автоматически созданное описание

И графики:

Изображение выглядит как диаграмма

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описание