Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Фізико-технічний інститут

Статистичний аналіз результатів ЗНО 2019

предмет «Математична статистика»

2022

Зміст

Програмний етап 3
Ініціалізація даних 3
Реалізація рандомізованого формування елементів вибірок 5

1 Результати з української мови 6
Перевірка гіпотези однорідності даних 6
Таблиця спостережуваних даних 8
Висновок 8
Нормалізація даних 9
Перевірка гіпотези рівності дисперсій 10
F-статистика Фішера 11
Реалізація вибірки 13
Висновок 13
Побудова довірчого інтервалу для різниці математичних сподівань 13
Пошук центральної статистики 14
Виокремлення проміжку можливих значень параметра 15
Реалізація вибірки 16
Висновок 16
Побудова довірчого інтервалу для значення дисперсії 16
Висновок 18
Статистична похибка обчислень 18
Висновок 19

2 Результати з математики 20
Перевірка гіпотези однорідності даних 20
Таблиця спостережуваних даних 21
Висновок 21
Перевірка гіпотези рівності дисперсій 22
Висновок 23
Побудова довірчого інтервалу для різниці математичних сподівань 24
Висновок 25
Перевірка гіпотези рівності математичних сподівань 25
Висновок 26
Побудова довірчого інтервалу для значення дисперсії 26
Висновок 27
Статистична похибка обчислень 27
Висновок 28

3 Результати з англійської мови **29**
Перевірка гіпотези однорідності даних . . . 29
Таблиця спостережуваних даних . . . 30
Висновок . . . 30
Перевірка гіпотези рівності дисперсій . . . 31
Висновок . . . 32
Побудова довірчого інтервалу для різниці математичних сподівань . . . 33
Висновок . . . 34
Побудова довірчого інтервалу для значення дисперсії . . . 34
Висновок . . . 35
Статистична похибка обчислень . . . 35
Висновок . . . 36

Загальний висновок **37**

Програмний етап

Для статистичного аналізу було обрано відкриті дані із загальнодоступного освітнього сайту https://osvita.ua/news/data/, де з-поміж інших наборів даних у переліку доступні «Деперсоніфіковані дані учасників ЗНО 2019 року з кожного навчального предмета».

Ініціалізація даних

Зі сторінки сайту можна завантажити архів даних `opendatazno2019.zip`. У розархівованій папці наявні два файли з різними іменами: `opendatazno2019.xlsx` та `opendatazno2019_info.xls`. У першому містяться власне дані, а у другому – перелік атрибутів таблиці даних.

Файл `opendatazno2019.xlsx` має великий обсяг – понад 230 MB, тому жоден онлайн редактор не буде в змозі його відкрити. Рівно як і програмне забезпечення `Microsoft Excel`, `Google Sheets` чи `LibreOffice Calc`. Тож для подальшого опрацювання вхідних даних буде використано засоби мови `Python`.

Для читання файлів розширення `.xlsx` можна використати бібліотеку `xlrd` версії `1.2.0` й надалі працювати безпосередньо із рядками таблиці, пробігаючи кожну комірку так, як це показано у рядку 13 Лістингу 1:

Лістинг 1: Використання бібліотеки xlrd

```
1  import xlrd
2  
3  # open the Workbook
4  workbook = xlrd.open_workbook("opendatazno2019.xlsx")
5  
6  # open the worksheet
7  worksheet = workbook.sheet_by_index(0)
8  
9  # iterate the rows and columns
10 for i in range(0, 5):
11     for j in range(0, 3):
12         # print the cell values with a tab space
13         print(worksheet.cell_value(i, j), end="\t")
14     print("")
```

Проте у такому разі обробка файлу триватиме близько 5-ти хвилин, тож такий спосіб опрацювання великого обсягу даних є неефективним. Натомість, користуючись тією ж бібліотекою `xlrd` у додачу до засобів бібліотек `pandas` та `csv`, можна зчитати й порядково перевторити файл розширення `.xlsx` у файл `.csv`, як це наведено на Лістингу 2. Це значно зменшить тривалість виконання обробки даних. Більше про різні способи зчитування й обробки файлів розширення `.xlsx` можна довідатися за цим посиланням.

Лістинг 2: Конвертація у `.csv` файл

```python
import pandas as pd
import xlrd
import csv

# open workbook by sheet index, optional - sheet_by_index()
sheet = xlrd.open_workbook("opendatazno2019.xlsx").sheet_by_index(0)

# writer object is created
column = csv.writer(open("opendatazno2019.csv", "w", newline=""))

# write the data into csv file
for row in range(sheet.nrows):
    # row by row write operation
    column.writerow(sheet.row_values(row))

# read csv file and convert into a dataframe object
df = pd.DataFrame(pd.read_csv("opendatazno2019.csv", dtype="unicode"))
```

Як це зображено на Рис. 1, початкові дані мають рядки невідповідного формату, тобто ці дані є «брудними». На Лістингу 3 коротко вказані команди, за допомогою яких можна прибрати нульові значення чи комірки з невідповідним форматом. Як результат – матимемо готові «чисті» дані для подальшої обробки.

```
Whole initial dataframe:

        Birth SEXTYPENAME UkrBall100 mathBall100 engBall100
0        2001      жіноча      100.0         NaN        NaN
1        1985      жіноча        NaN         NaN        NaN
2        2001      жіноча      166.0         NaN        NaN
3        2000    чоловіча      127.0         0.0        NaN
4        2001      жіноча      171.0         NaN      116.0
5        2001    чоловіча        0.0         NaN        NaN
6        1999    чоловіча        0.0         NaN        NaN
...       ...         ...        ...         ...        ...
353806   2001    чоловіча      107.0         NaN        NaN
353807   2002      жіноча      127.0         NaN        NaN
353808   2001      жіноча      197.5       180.0      173.0
353809   2001      жіноча      122.0       100.0        NaN
353810   2001    чоловіча      134.0       140.0        NaN
353811   2002      жіноча      131.0         NaN        NaN
353812   2000      жіноча        NaN         NaN        NaN
```

Рис. 1: Початкові дані

Лістинг 3: Чистка даних

```python
df = pd.DataFrame(pd.read_csv("opendatazno2019.csv", dtype="unicode"))

# convert string to float
df["UkrBall100"] = df["UkrBall100"].astype("float")

# remove all rows with NULL values:
cleaned_df = df.dropna(subset=["UkrBall100"])

# resetting indexes after removing rows from dataframe
cleaned_df.reset_index(drop=True, inplace=True)

# remove all rows with "0.0" values:
cleaned_df = cleaned_df.loc[cleaned_df["UkrBall100"] != 0.0]
cleaned_df.reset_index(drop=True, inplace=True)
```

Реалізація рандомізованого формування елементів вибірок

Важливим етапом статистичного аналізу є реалізація випадкового, рандомізованого формування вибірок із усього наявного масиву даних. Програмно таку реалізацію наведено на Лістингу нижче.

Лістинг 4: Рандомізоване формування вибірок

```python
import random

sample_size = 500

# create a list of all possible indexes
index = [i for i in range(0,cleaned_df.last_valid_index()+1)]

# create an empty dictionary
random_elements = {}

# add one column to a dictionary
elements = []
random_elements.update({"UkrBall100": elements})

for j in range(sample_size):
    random_index = random.choice(index)
    elements.append(cleaned_df.loc[random_index, "UkrBall100"])
    index.remove(random_index)

random_selected_df = pd.DataFrame(random_elements)
```

Із усім використаним в ході роботи програмним кодом `Python` можна ознайомитися на `github репозиторії`.

1 Результати з української мови

Після програмного етапу завантаження й чистки даних маємо змогу безпосередньо оглянути отримані 286 413 результати, при цьому зазначимо, що жіночих виявилося на 19 341 більше за чоловічих. Зобразимо гістограми результатів ЗНО з української мови 2019 року для вибірки, наприклад, 10 000 учнів:

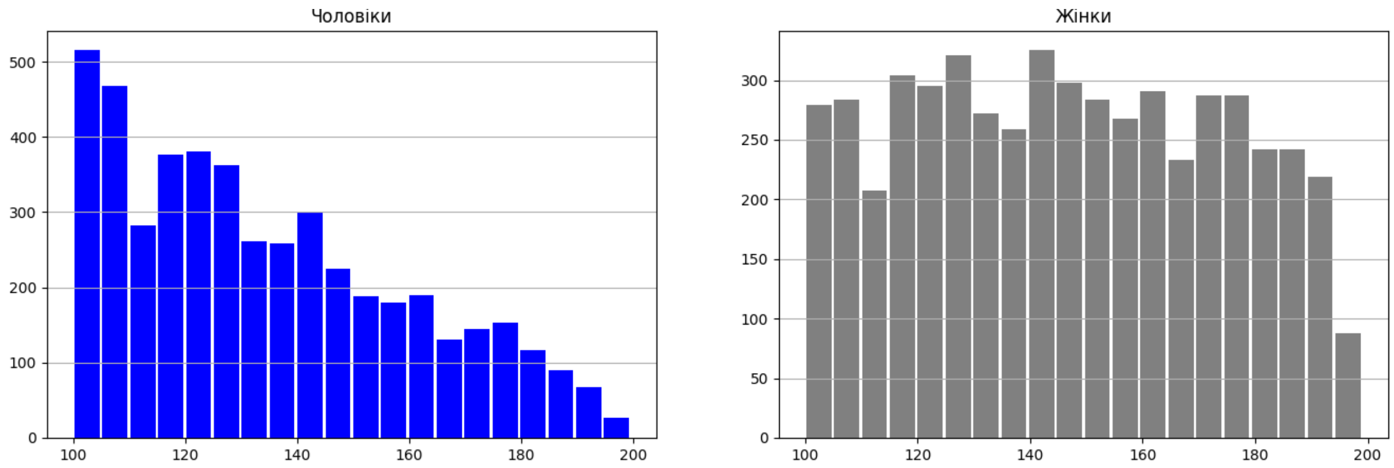


Рис. 2: Результати з української мови

На малюнку наведено розподіл балів чоловіків та жінок: неозброєним оком бачимо наявну відмінність отриманих оцінок в залежності від статі. З'ясуємо, чи ця відмінність є статистично значущою.

Перевірка гіпотези однорідності даних

Нехай маємо дві незалежні вибірки $\{X_i\}$ та $\{Y_j\}$ однаково розподілених випадкових величин, розподіли F_x й F_y яких нам невідомі:

$$\begin{aligned} &X_1 \dots X_n \sim F_x && \text{результати чоловіків} \\ &Y_1 \dots Y_m \sim F_y && \text{результати жінок} \end{aligned}$$

Перевіримо гіпотезу про однорідність статистичного матеріалу, тобто гіпотезу, що ймовірності спостереження умовно високих, помірних та низьких балів в обох вибірках є однаковими. Таким чином матимемо $k = 2$ вибірок, в яких елементи приймають $s = 3$ різних значень. Для формалізації задачі введемо позначення:

$$p_{ij} = \left\{ \begin{matrix} \text{оцінка з чоловічої}_{(i=1)} \text{ чи жіночої}_{(i=2)} \text{ вибірки належить} \\ \text{множині високих}_{(j=1)}, \text{ помірних}_{(j=2)} \text{ чи низьких}_{(j=3)} \text{ балів} \end{matrix} \right\}$$

Категорізація на множини оцінок того чи іншого рівня базувалася на прохідних балах на різні факультети КПІ у 2019 році (із переліком прохідних балів можна ознайомитися за посиланням).

При цьому «високі» оцінки обиралися із міркувань, що вступнику із таким балом доступно для вступу на бюджет більше ніж 65% усіх факультетів, натомість абітурієнту із «низькими» оцінками рівень доступності значно нижчий – лише 30%. До прикладу:

$$\begin{aligned} A_1 &= \{\text{оцінки, вищі за 178 балів}\} \\ A_2 &= \{\text{оцінки між 144 та 178 балами}\} \\ A_3 &= \{\text{оцінки, нижчі за 144 бали}\} \end{aligned}$$

Остаточно гіпотеза перевірки однорідності спосережуваних даних формулюватиметься так:

$$\begin{array}{ll} \text{нульова гіпотеза} & H_0 : p_{11} = p_{21},\ p_{12} = p_{22},\ p_{13} = p_{23} \\ \text{проти альтернативи} & H_1 : \exists i, j \quad p_{1,j} \neq p_{2,j} \end{array} \tag{1.1}$$

А тоді критерієм перевірки гіпотези слугуватиме правило

$$\delta(x_1 \ldots x_n,\ y_1 \ldots y_m) = \begin{cases} H_1, & \rho \geqslant \chi^2_{1-\alpha;\ (k-1)(s-1)} \\ H_0, & \rho < \chi^2_{1-\alpha;\ (k-1)(s-1)} \end{cases}, \tag{1.2}$$

де статистика критерію ρ обчислюється так:

$$\rho \equiv \rho(x_1 \ldots x_n,\ y_1 \ldots y_m) = (n+m) \left(\sum_{i=1}^{k} \sum_{j=1}^{s} \frac{\vartheta_{ij}^2}{\vartheta_{i\cdot} \vartheta_{\cdot j}} - 1 \right), \tag{1.3}$$

величина критичної точки $\chi^2_{1-\alpha;\ (k-1)(s-1)}$ є квантилем рівня $(1-\alpha)$ розподілу Ст'юдента із $(k-1)(s-1)$ степенями свободи, а також

$$\vartheta_{1j} = \sum_{i=1}^{n} \mathbb{1}(X_i \in A_j)$$ частота потрапляння елементів вибірки результатів чоловіків в одну із j категорій

$$\vartheta_{2j} = \sum_{i=1}^{m} \mathbb{1}(Y_i \in A_j)$$ частота потрапляння елементів вибірки результатів жінок в одну із j категорій

$$\vartheta_{i\cdot} = \sum_{j=1}^{s} \vartheta_{ij},\ \vartheta_{\cdot j} = \sum_{i=1}^{k} \vartheta_{ij}$$ суми відповідних рядків чи стовпців таблиці спостережуваних даних

Таблиця спостережуваних даних

Складемо таблицю за спостережуваними даними навмання обраних $n = 500$ та $m = 500$ елементів із вибірок результатів ЗНО з української мови для чоловіків та жінок:

	Низькі бали	Помірні бали	Високі бали	Всього
Чоловіки	343	125	32	500
Жінки	231	186	83	500
Всього	574	311	115	1000

Табл. 1: Таблиця спостережуваних значень з української мови

Обчислимо значення статистики критерію:

$$\rho = 1000 \left(\frac{343^2}{574 \cdot 500} + \frac{125^2}{311 \cdot 500} + \ldots + \frac{186^2}{311 \cdot 500} + \frac{83^2}{115 \cdot 500} - 1 \right) = 56.44$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичної точки

$$\chi^2_{1-\alpha;\ (k-1)(s-1)} = \chi^2_{0.99;\ (2-1)(3-1)} = \chi^2_{0.99;\ 2} = 9.21$$

Висновок

Оскільки значення статистики критерію перевищує значення критичної точки, то згідно критерію (1.2) гіпотеза про однорідність статистичних даних відхиляється. Отже, ймовірності спостереження оцінок різного рівня у вибірках для чоловіків та жінок, відповідно, різняться, тобто наявна ситуація неоднакового розполіду балів в залежності від статі.

Значення `p-value`

Нехай випадкова величина τ має такий самий розподіл, як і статистика критерію й при цьому не залежить від неї: $\tau \sim \chi^2(2)$. Тоді величина `p-value` обчислюється як така ймовірність:

$$P(\tau > \rho \mid H_0) = P(\tau > 56.44) = 1 - P(\tau \leqslant 56.44) = 1 - F_\tau(56.44) = 0.0001$$

Тож при значенні $\alpha <$ `p-value` гіпотеза H_0 приймалася б.

Нормалізація даних

Як уже зазначалося, маємо дві незалежні вибірки $\{X_i\}$ та $\{Y_j\}$ однаково розподілених випадкових величин, розподіли F_x й F_y яких нам невідомі:

$$\begin{aligned} &X_1 \dots X_n \sim F_x && \text{результати чоловіків} \\ &Y_1 \dots Y_m \sim F_y && \text{результати жінок} \end{aligned}$$

Тим не менш, великий обсяг початкових даних дозволяє розбити оригінальні спостереження $\{X_i\}$ й $\{Y_j\}$ на значну кількість достатньо великих випадково сформованих неперетинних множин виду

$$\left\{X_j^1\right\}_{j=\overline{1,n}} \dots \left\{X_j^N\right\}_{j=\overline{1,n}} \text{ й } \left\{Y_j^1\right\}_{j=\overline{1,n}} \dots \left\{Y_j^N\right\}_{j=\overline{1,n}},$$

та надалі розглядати ці набори як незалежні дослідження. Причому в силу центральної граничної теореми (далі – ЦГТ) із наявних наборів можна сформувати вибірки з N величин, які мають нормальний розподіл:

$$\overline{X^1} \dots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma_x^2) \qquad \text{середні результати чоловіків,} \tag{1.4}$$

$$\overline{Y^1} \dots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma_y^2) \qquad \text{середні результати жінок,} \tag{1.5}$$

де

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j^i, \ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j^i, \ i = \overline{1,N}$$

Схематично зобразимо суть процесу нормалізації початкових даних на прикладі вибірки $\{X_i\}$:

$$\begin{matrix} X_1^1 \dots X_n^1 \sim F_x & & \sqrt{n}\cdot\dfrac{\overline{X^1}-\mu_x}{\sqrt{\sigma_x^2}} \sim N(0,1) \\ \dots & \overset{\text{ЦГТ}}{\Longrightarrow} & \dots \\ X_1^N \dots X_n^N \sim F_x & & \sqrt{n}\cdot\dfrac{\overline{X^N}-\mu_x}{\sqrt{\sigma_x^2}} \sim N(0,1) \end{matrix} \tag{1.6}$$

Виконаємо ланцюжок нормалізації на прикладі вибірки результатів чоловіків. ЦГТ для достатньо великої фіксованої кількості n випадково обраних елементів послідовності незалежних однаково розподілених випадкових величин матиме вид:

$$\frac{\sum\limits_{i=1}^{n} X_i - M\sum\limits_{i=1}^{n} X_i}{\sqrt{D\sum\limits_{i=1}^{n} X_i}} \approx N(0,1)$$

Спростимо вираз, використовуючи позначення $\mu_x = MX_i,\ \sigma_x^2 = DX_i$ та скориставшись властивостями дисперсії та математичного сподівання:

$$\frac{n\overline{X} - n\mu_x}{\sqrt{n\sigma_x^2}} \approx N(0,1),\ \text{де}\ \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Залишивши n виключно у знаменнику, матимемо:

$$\frac{n\overline{X} - n\mu_x}{\sqrt{n\sigma_x^2}} \approx N(0,1)\ \Rightarrow\ \frac{\overline{X} - \mu_x}{\sqrt{\sigma_x^2/n}} \approx N(0,1)\ \Rightarrow\ \overline{X} \approx N(\mu_x, \tfrac{1}{n}\sigma_x^2)$$

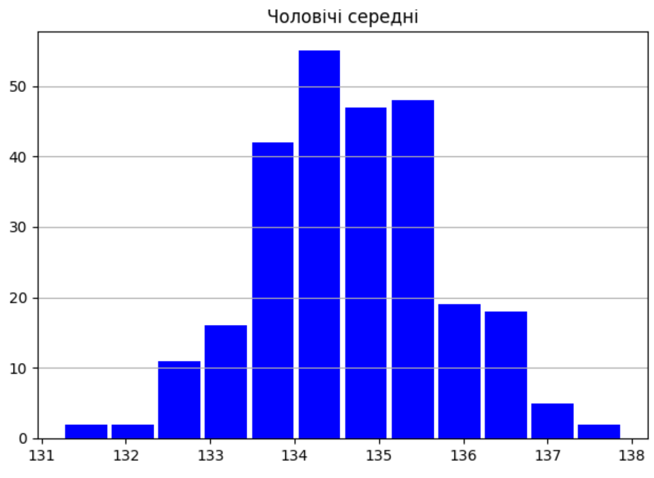
При розгляді такої ж кількості спостережень n аналогічним чином отримуємо знормоване значення результатів жінок, де $\mu_y = MY_i,\ \sigma_y^2 = DY_i$:

$$\overline{Y} \approx N(\mu_y, \tfrac{1}{n}\sigma_y^2)$$

Таким чином крок за кроком з початкових спостережуваних результатів утворюємо N випадково сформованих неперетинних вибірок, вибіркові середні яких мають нормальний розподіл. Результат показано у виразах (1.4)-(1.5).

Перевірка гіпотези рівності дисперсій

Повертаючись до отриманих наборів (1.4) та (1.5), переконаємося, що новоутворені вибірки справді мають нормальний розподіл. Для цього побудуємо гістограми, які за означенням є наближеннями істинних щільностей, а потім візуально проаналізуємо отримані криві:



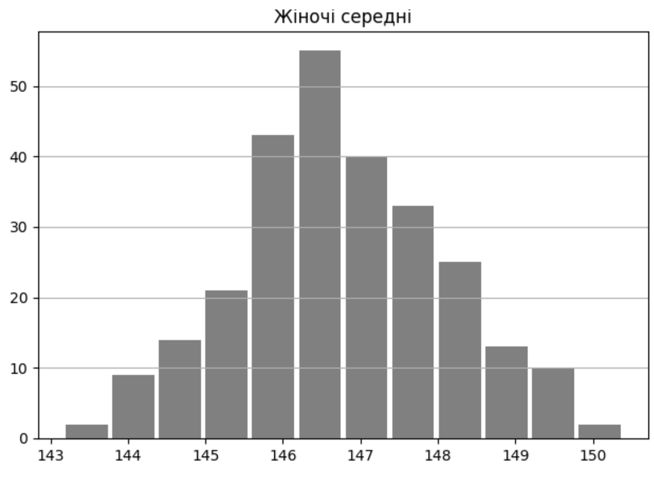


Рис. 3: Гістограми наборів усереднених результатів ЗНО з української мови

Зауважимо, що криві в обох випадках мають схожі риси (мова про «висоту» та «ширину» графіків). Це дає підстави висунути й надалі первірити гіпотезу про рівність дисперсій вибірок результатів чоловіків та жінок.

F-статистика Фішера

Нехай задано дві незалежні нормально розподілені вибірки із невідомими параметрами:

$$\begin{aligned}
&\overline{X^1}\ldots\overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma_x^2) && \text{середні результати чоловіків,}\\
&\overline{Y^1}\ldots\overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma_y^2) && \text{середні результати жінок,}
\end{aligned}$$

при цьому

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j^i,\ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j^i,\ i = \overline{1,N}$$

Перевіримо гіпотезу рівності невідомих дисперсій:

$$\begin{aligned}
&\text{нульова гіпотеза} && H_0 : \sigma_x^2 = \sigma_y^2\\
&\text{проти альтернативи} && H_1 : \sigma_x^2 \neq \sigma_y^2
\end{aligned}$$

Або в еквівалентному записі:

$$\begin{aligned}
&\text{нульова гіпотеза} && H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1 \qquad (1.7)\\
&\text{проти альтернативи} && H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1
\end{aligned}$$

Скористаємося статистикою Фішера як статистикою критерію перевірки вищевказаної гіпотези. Перш за все зазначимо, що в силу нормальної розподіленості заданих вибірок нищевказані випадкові величини матимуть розподіл хі-квадрат із $(N-1)$ степенями свободи:

$$\xi \equiv \frac{(N-1)S_X^2}{\sigma_x^2/n} \sim \chi^2(N-1),\ \eta \equiv \frac{(N-1)S_Y^2}{\sigma_y^2/n} \sim \chi^2(N-1), \qquad (1.8)$$

де значення

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{X^i} - \overline{X}) \quad \text{та} \quad S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y^i} - \overline{Y})$$

є вибірковими дисперсіями двох наборів вибірок середніх результатів, а

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N}\overline{X^i} \quad \text{й} \quad \overline{Y} = \frac{1}{N}\sum_{i=1}^{N}\overline{Y^i}$$

є вибірковими середніми відповідних вибірок.

Оскільки обидві величини (1.8) мають розподіл хі-квадрат та є незалежними, то випадкова величина

$$\zeta \overset{\text{def}}{=} \frac{\eta/(N-1)}{\xi/(N-1)} \sim F\,(N-1,\ N-1)$$

має розподіл Фішера з $(N-1,\ N-1)$ степенями свободи. Підставивши вирази (1.8), матимемо:

$$\zeta = \frac{\sigma_x^2}{\sigma_y^2} \cdot \frac{S_Y^2}{S_X^2} \sim F\,(N-1,\ N-1)$$

При істиності нульової гіпотези $H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1$ отримаємо статистику критерію Фішера, яка визначатиметься так:

$$\rho \equiv \rho(\overline{x^1} \dots \overline{x^N},\ \overline{y^1} \dots \overline{y^N}) = \frac{S_Y^2}{S_X^2} \sim F\,(N-1,\ N-1) \tag{1.9}$$

Тоді перевищення чи недобір цією статистикою певних критичних значень формуватиме критичну область нульової гіпотези. Як наслідок критерієм перевірки гіпотези слугуватиме правило

$$\delta \equiv \delta(\overline{x^1} \dots \overline{x^N},\ \overline{y^1} \dots \overline{y^N}) = \begin{cases} H_1, & \rho \leqslant k^* \text{ або } \rho \geqslant k^{**} \\ H_0, & \rho \in (k^*,\ k^{**}) \end{cases}, \tag{1.10}$$

де значення критичних точок k^* та k^{**} при заданому рівні значущості α знайдемо з означення помилки першого роду:

$$\alpha \overset{\text{def}}{=} P(\delta \in K_\alpha \mid H_0) = P(\rho \leqslant k^* \text{ або } \rho \geqslant k^{**} \mid H_0)$$

Таким чином

$$\alpha = P(\rho \leqslant k^* \mid H_0) + P(\rho \geqslant k^{**} \mid H_0)$$

За умови однакової ймовірності цих подій матимемо:

$$\begin{cases} P(\rho \leqslant k^* \mid H_0) = \frac{\alpha}{2} \\ P(\rho \geqslant k^{**} \mid H_0) = \frac{\alpha}{2} \end{cases} \Rightarrow \begin{cases} k^* = f_{\frac{\alpha}{2};\ (N-1,\ N-1)} \\ k^{**} = f_{1-\frac{\alpha}{2};\ (N-1,\ N-1)} \end{cases}, \tag{1.11}$$

де значення k^* та k^{**} є квантилями розподілу Фішера із $(N-1,\ N-1)$ степенями свободи при вказаному рівні значущості α. Отже, за допомогою критичних точок (1.11) та статистики критерію (1.9) маємо змогу перевірити гіпотезу рівності дисперсій (1.7) згідно критерія (1.10).

Реалізація вибірки

Розіб'ємо усю множину спостережуваних значень на $N = 267$ неперетинних множин по $n = 500$ елементів у кожній. У такому випадку статистична похибка обчислень є малою (детальніше – на сторінці 18). Наведемо потрібні величини для перевірки гіпотези рівності дисперсій у таблиці нижче:

n	N	α	S_X^2	S_Y^2
500	267	0.01	1.3501	1.6246

Табл. 2: Значення шуканих параметрів F-тесту

Обчислимо значення статистики критерію:

$$\rho = \frac{S_Y^2}{S_X^2} = 1.2033$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичних точок

$$k^* = f_{\frac{\alpha}{2};\ (N-1,\ N-1)} = f_{0.005;\ (266,\ 266)} = 0.7285$$
$$k^{**} = f_{1-\frac{\alpha}{2};\ (N-1,\ N-1)} = f_{0.995;\ (266,\ 266)} = 1.3728$$

Висновок

Оскільки значення статистики критерію ρ входить у проміжок $(0.7285,\ 1.3728)$, то згідно критерію (1.10) на рівні значущості $\alpha = 0.01$ нульова гіпотеза $H_0 : \sigma_x^2 = \sigma_y^2$ приймається. Для усіх подільших міркувань введемо таке позначення: $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

Побудова довірчого інтервалу для різниці мат. сподівань

В результаті нормалізації початкових даних та перевірки гіпотези рівності дисперсій отримуємо такі дві вибірки:

$$\overline{X^1} \dots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \quad \text{середні результати чоловіків,} \tag{1.12}$$
$$\overline{Y^1} \dots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \quad \text{середні результати жінок,} \tag{1.13}$$

при цьому

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j^i,\ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j^i,\ i = \overline{1, N}$$

Спробуємо оцінити, в якому інтервалі лежить значення різниці середніх балів для вибірок результатів ЗНО з української мови чоловіків та жінок.

Пошук центральної статистики

Перш за все віднайдемо для параметра $\theta = \mu_y - \mu_x$ центральну статистику

$$G \equiv G(\overline{X^1} \dots \overline{X^N},\ \overline{Y^1} \dots \overline{Y^N},\ \theta)$$

Ця функція має задовільняти двом умовам: по-перше, її щільність розподілу $f_G(x)$ не залежить від параметра θ, а по-друге, сама функція G є неперервною і монотонною за θ. Виходячи з властивостей нормального розподілу

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N}\overline{X^i} \sim N(\mu_x, \tfrac{1}{Nn}\sigma^2),\ \overline{Y} = \frac{1}{N}\sum_{i=1}^{N}\overline{Y^i} \sim N(\mu_y, \tfrac{1}{Nn}\sigma^2)$$

А отже, розподіл різниці $\overline{Y} - \overline{X}$ матиме вид

$$\overline{Y} - \overline{X} \sim N(\mu_y - \mu_x,\ \tfrac{2}{Nn}\sigma^2)$$

Тоді позначимо

$$\xi \equiv \frac{\overline{Y} - \overline{X} - (\mu_y - \mu_x)}{\sqrt{\frac{2}{Nn}\sigma^2}} \sim N(0,1) \tag{1.14}$$

В силу аналогічних до запису (1.8) міркувань нищевказані випадкові величини матимуть розподіл хі-квадрат із $(N-1)$ степенями свободи:

$$\frac{(N-1)S_X^2}{\sigma^2/n} \sim \chi^2(N-1),\ \frac{(N-1)S_Y^2}{\sigma^2/n} \sim \chi^2(N-1), \tag{1.15}$$

де

$$S_X^2 = \tfrac{1}{N-1}\sum_{i=1}^{N}(\overline{X^i} - \overline{X})$$
$$S_Y^2 = \tfrac{1}{N-1}\sum_{i=1}^{N}(\overline{Y^i} - \overline{Y})$$

вибіркові дисперсії двох наборів вибірок середніх результатів

А отже, сума випадкових величин (1.15) як сума незалежних випадкових величин матиме розподіл χ^2 із $[(N-1)+(N-1)]$ степенями свободи:

$$\eta \equiv \frac{(N-1)}{\sigma^2/n}\left(S_X^2 + S_Y^2\right) \sim \chi^2\left[2(N-1)\right] \tag{1.16}$$

Тоді випадкова величина

$$\zeta \stackrel{\text{def}}{=} \frac{\xi}{\sqrt{\frac{\eta}{2(N-1)}}} \sim t\left[2(N-1)\right] \tag{1.17}$$

має розподіл Ст'юдента із $[2(N-1)]$ степенів свободи. Підставивши вирази (1.14) й (1.16) у формулу (1.17), отримаємо шукану центральну статистику для параметра $\theta = \mu_y - \mu_x$:

$$G = \left((\overline{Y} - \overline{X}) - \theta\right) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} \sim t\left[2(N-1)\right] \tag{1.18}$$

Виокремлення проміжку можливих значень параметра

Побудуємо довірчий інтервал рівня довіри γ :

$$\gamma \stackrel{\text{def}}{=} P(g_1 < G < g_2) = \int\limits_{g_1}^{g_2} f_G(x)\, dx$$

В силу симетричності розподілу Ст'юдента, найкоротший центральний довірчий інтервал матиме вид:

$$\gamma = P(g_1 < G < g_2) = P(|\,G\,| < g) \tag{1.19}$$

На малюнку нижче схематично зображено ідею пошуку довірчого інтервалу та визначення відповідного квантиля g :

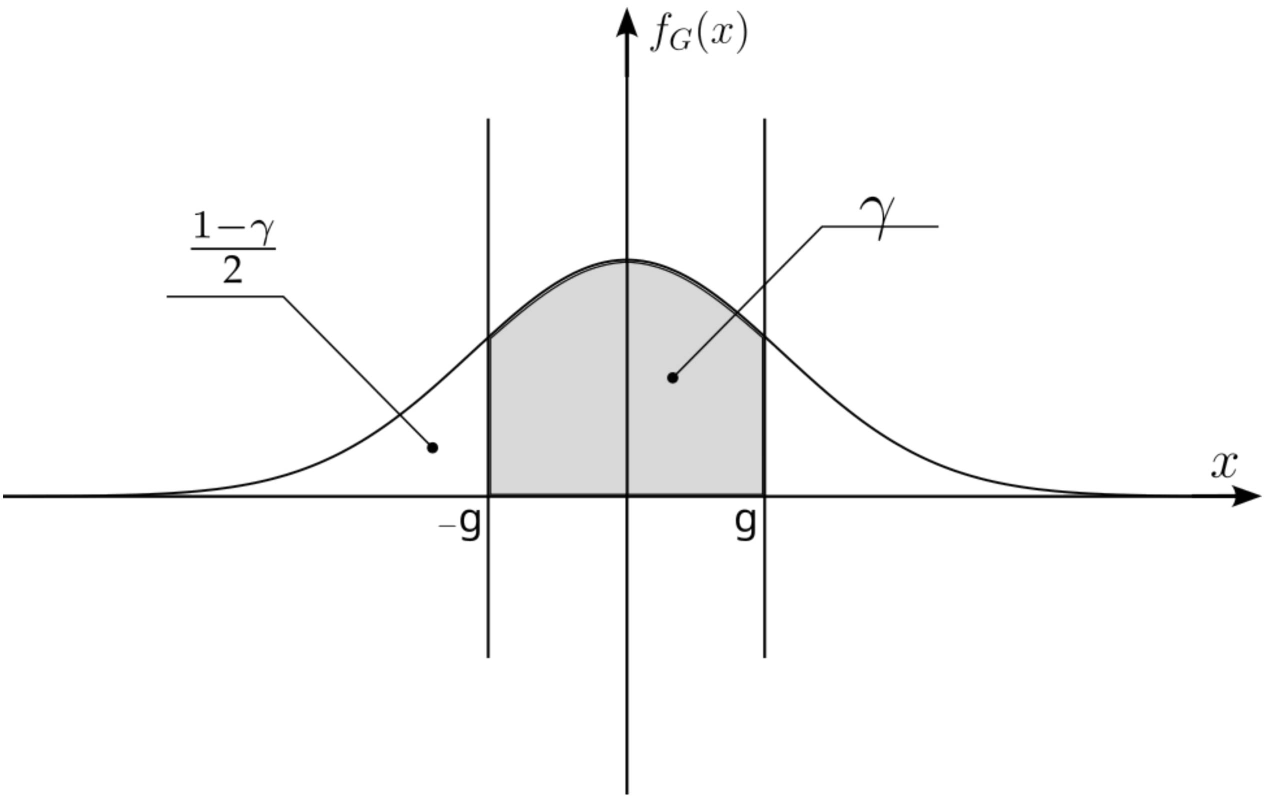


Рис. 4: Пошук квантиля розподілу Ст'юдента

Тож враховуючи, що $\int\limits_{-\infty}^{\infty} f_G(x)\, dx \stackrel{\text{def}}{=} 1$, матимемо такий ланцюжок міркувань:

$$\int\limits_{-\infty}^{g} f_G(x)\, dx = \tfrac{1-\gamma}{2} + \gamma = \tfrac{1+\gamma}{2} \;\Rightarrow\; F_t(g) = \tfrac{1+\gamma}{2} \;\Rightarrow\; g = t_{\frac{1+\gamma}{2};\; 2(N-1)}$$

Знайдено квантиль рівня $\frac{1+\gamma}{2}$ розподілу Ст'юдента із $[2(N-1)]$ степенями свободи. Підставляючи усі отримані результати у формулу (1.19), отримаємо:

$$\gamma = P\left(-t_{\frac{1+\gamma}{2};\; 2(N-1)} < \big((\overline{Y} - \overline{X}) - \theta\big) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} < t_{\frac{1+\gamma}{2};\; 2(N-1)}\right) \tag{1.20}$$

Реалізація вибірки

Останнім кроком вкажемо конкретні значення усіх необхідних величин:

n	N	γ	$g = t_{0.975;\ 532}$	S_X^2	S_Y^2	$\overline{Y} - \overline{X}$
500	267	0.95	1.96	1.3501	1.6246	12.0768

Табл. 3: Значення параметрів для довірчого інтервалу різниці середніх

Підставивши у вираз (1.20) значення, які наведені у таблиці вище, отримаємо такий довірчий інтервал для параметра $\theta = \mu_y - \mu_x$:

$$P(11.87 < \theta < 12.28) = 0.95$$

Висновок

Отримано довірчий інтервал рівня довіри $\gamma = 0.95$ для величини різниці середніх значень оцінок з української мови вибірок результатів чоловіків та жінок: $(11.87,\ 12.27) \ni \mu_y - \mu_x$. Оскільки у вказаному проміжку немає нульового значення, гіпотезу про нерозрізнюваність середніх можна відхилити.

Можемо стверджувати, що при порівнянні результатів чоловіка та жінки оцінка з української мови у жінки буде більшою за оцінку у чоловіка на бал у 12.08 ± 0.2 пункти, при цьому в середньому у п'яти зі ста таких порівнянь вказане наближення може бути хибним.

Побудова довірчого інтервалу для значення дисперсії

Наступним кроком віднайдемо проміжок можливих значень дисперсії, або інакше кажучи, розсіяння навколо математичного сподівання нормованих вибірок. В силу прийняття гіпотези рівності дисперсій (сторінка 13) із оригінальних наборів

$$\overline{X^1} \ldots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати чоловіків}$$
$$\overline{Y^1} \ldots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати жінок}$$

аналогічними кроками (як на сторінці 14) можна сформувати випадкову величину (1.16) розподілу хі-квадрат із $[2(N-1)]$ степенями свободи, при цьому для параметра дисперсії $\frac{1}{n}\sigma^2$ ця величина відіграватиме роль центральної статистики:

$$G \equiv \frac{(N-1)}{\sigma^2/n}\left(S_X^2 + S_Y^2\right) \sim \chi^2\left[2(N-1)\right] \tag{1.21}$$

А тоді довірчий інтервал рівня довіри γ матиме вид:

$$\gamma \stackrel{\text{def}}{=} P(g_1 < G < g_2) = \int\limits_{g_1}^{g_2} f_G(x)\, dx \tag{1.22}$$

На рисунку нижче схематично показано пошук величин g_1 та g_2 :

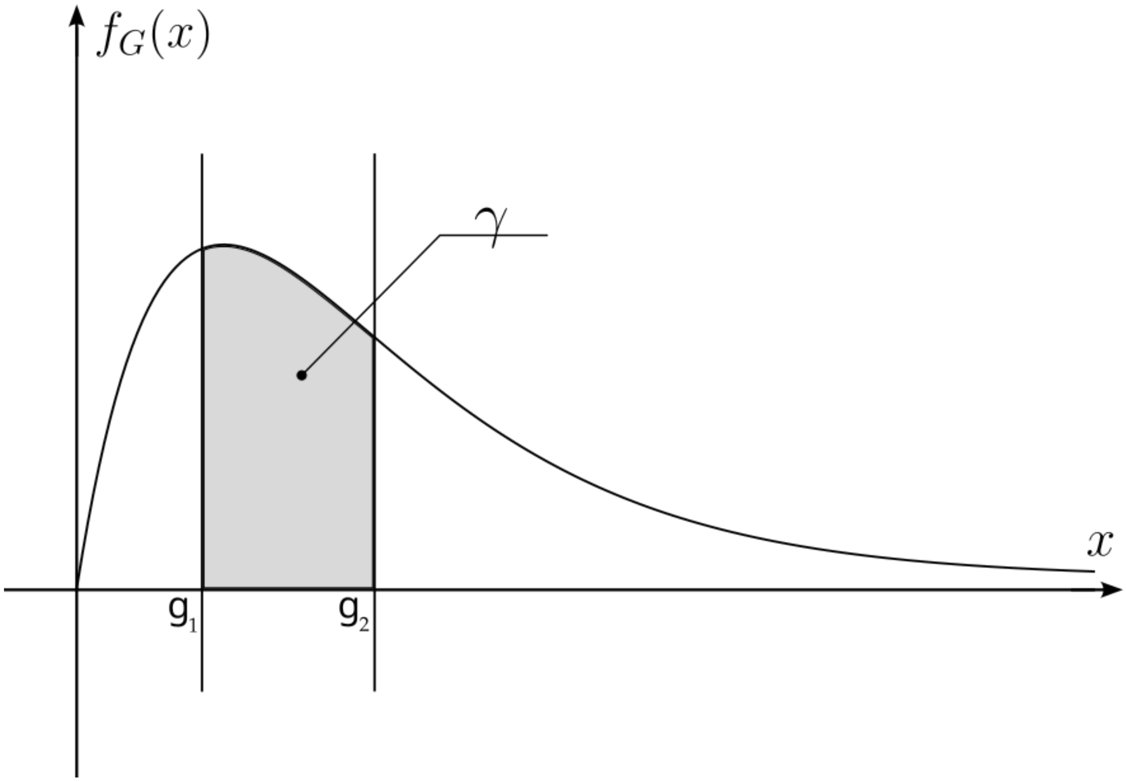


Рис. 5: Пошук квантилів розподілу хі-квадрат

Тож значення g_1 та g_2 є квантилями рівнів $\frac{1-\gamma}{2}$ та $\frac{1+\gamma}{2}$ розподілу хі-квадрат із $[2(N-1)]$ степенями свободи:

$$\int\limits_{0}^{g_1} f_G(x)\, dx = \tfrac{1-\gamma}{2} \;\Rightarrow\; F_{\chi^2}(g_1) = \tfrac{1-\gamma}{2} \;\Rightarrow\; g_1 = \chi^2_{\frac{1-\gamma}{2};\; 2(N-1)}$$

$$\int\limits_{0}^{g_2} f_G(x)\, dx = \tfrac{1+\gamma}{2} \;\Rightarrow\; F_{\chi^2}(g_2) = \tfrac{1+\gamma}{2} \;\Rightarrow\; g_2 = \chi^2_{\frac{1+\gamma}{2};\; 2(N-1)}$$

Підставляючи отримані результати у формулу (1.22), отримаємо:

$$\gamma = P\left(\chi^2_{\frac{1-\gamma}{2};\; 2(N-1)} < \frac{(N-1)}{\sigma^2/n}\left(S_X^2 + S_Y^2\right) < \chi^2_{\frac{1+\gamma}{2};\; 2(N-1)}\right) \tag{1.23}$$

Наостанок вкажемо конкретні значення усіх необхідних величин:

n	N	γ	$g_1 = \chi^2_{0.025;\ 532}$	$g_2 = \chi^2_{0.975;\ 532}$	S_X^2	S_Y^2
500	267	0.95	469.98	597.80	1.3501	1.6246

Табл. 4: Значення параметрів для довірчого інтервалу дисперсії

Підставивши у вираз (1.23) наведені вище значення, отримаємо такий довірчий інтервал для параметра $\frac{1}{n}\sigma^2$:

$$P(1.32\ < \frac{\sigma^2}{n} <\ 1.68) = 0.95$$

Висновок

Таким чином для сформованих нормально розподілених вибірок (1.12)-(1.13) значення дисперсії в середньому лише у п'яти елементів вибірки зі ста буде виходити за межі проміжку від 1.32 до 1.68 балів.

Статистична похибка обчислень

Статистичною похибкою обчислень називатимемо найменшу величину Δ відхилення різниці вибіркових середніх від істинного значення різниці математичних сподівань заданих вибірок (1.12)-(1.13) з щонайменшою ймовірністю у 95%. Тобто шукана похибка Δ обчислюватиметься як така ймовірність:

$$P\left(\ \left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \leqslant \Delta\ \right) \geqslant 0.95$$

Домножимо обидві частини нерівності всередині аргументу ймовірності на певну величину:

$$P\left(\underbrace{\left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}}} \leqslant \Delta \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}}\right) \geqslant 0.95$$

Після розкриття модуля та в ході аналогічних міркувань, які наведені для заданих вибірок на сторінці 14, з формули (1.18) випливає, що випадкова величина в лівій частини нерівності має розподіл Ст'юдента з $[2(N-1)]$ степенями свободи. А відтак шукана ймовірність згідно означення функції розподілу виглядатиме так:

$$2F_t\left(\Delta \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}}\right) - 1 \geqslant 0.95$$

У такому разі величина похибки Δ виражатиметься через квантиль рівня 0.975 розподілу Ст'юдента з $[2(N-1)]$ степенями свободи:

$$\Delta \geqslant t_{0.975;\ 2(N-1)} \cdot \sqrt{\frac{S_X^2 + S_Y^2}{N}} \tag{1.24}$$

Таблиця необхідних значень виглядатиме таким чином:

N	$t_{0.975;\ 532}$	S_X^2	S_Y^2
267	1.96	1.3501	1.6246

Табл. 5: Значення параметрів для обчислення статистичної похибки

Тож остаточно найменша величина відхилення $\Delta = 0.21$, а тоді початкова ймовірнісна нерівність матиме вид:

$$P\left(\ \left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \leqslant 0.21 \ \right) \geqslant 0.95$$

Висновок

Для встановлених значень $n = 500$ й відповідно $N = 267$ статистична похибка відхилення різниці вибіркових середніх від істинного значення різниці математичних сподівань складає $\Delta = 0.21$ бала з ймовірністю, більшою за 95%. Можемо стверджувати, що обчислення за такого значення похибки мають високу точність, оскільки величина $\Delta = 0.21$ є лише 0.1% від можливих 200 балів за результат тесту.

2 Результати з математики

Оглянемо отримані 122 026 результатів, при цьому зауважимо, що чоловічих на 11 984 більше за жіночих. Зобразимо гістограми результатів ЗНО з математики 2019 року для вибірки, наприклад, 10 000 учнів:

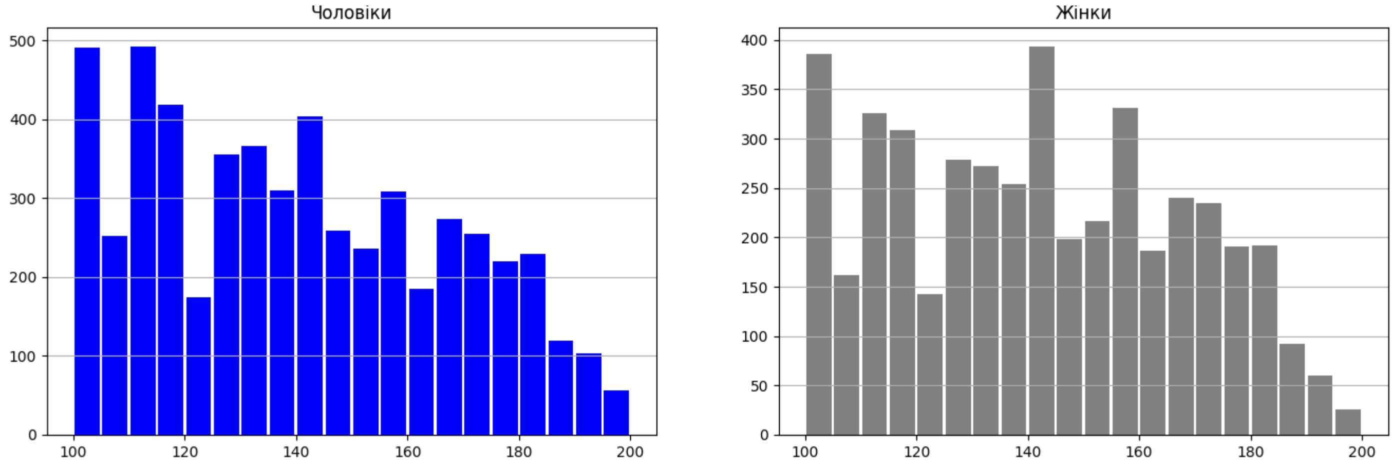


Рис. 6: Результати з математики

З'ясуємо, чи відмінність між результатами з математики для жінок та чоловіків є статистично значущою.

Перевірка гіпотези однорідності даних

Нехай маємо дві незалежні вибірки $\{X_i\}$ та $\{Y_j\}$ однаково розподілених випадкових величин, розподіли F_x й F_y яких нам невідомі:

$$
\begin{aligned}
&X_1 \ldots X_n \sim F_x && \text{результати чоловіків} \\
&Y_1 \ldots Y_m \sim F_y && \text{результати жінок}
\end{aligned}
$$

Перевіримо гіпотезу про однорідність статистичного матеріалу, тобто гіпотезу, що ймовірності спостереження умовно високих, помірних та низьких балів в обох вибірках є однаковими. Таким чином матимемо $k = 2$ вибірок, в яких елементи приймають $s = 3$ різних значень. За аналогічних позначень гіпотеза перевірки однорідності спосережуваних даних формулюватиметься так само, як і у виразах (1.1):

$$
\begin{aligned}
&\text{нульова гіпотеза} && H_0 : p_{11} = p_{21},\ p_{12} = p_{22},\ p_{13} = p_{23} \\
&\text{проти альтернативи} && H_1 : \exists i, j \quad p_{1,j} \neq p_{2,j}
\end{aligned}
$$

А тоді критерієм перевірки гіпотези рівно як і у випадку обробки результатів з української мови (1.2) слугуватиме правило

$$
\delta(x_1 \ldots x_n,\ y_1 \ldots y_m) = \begin{cases} H_1, & \rho \geqslant \chi^2_{1-\alpha;\ (k-1)(s-1)} \\ H_0, & \rho < \chi^2_{1-\alpha;\ (k-1)(s-1)} \end{cases}
$$

При цьому статистика критерію ρ обчислюється аналогічно до формули (1.3):

$$\rho \equiv \rho(x_1 \ldots x_n,\ y_1 \ldots y_m) = (n+m)\left(\sum_{i=1}^{k}\sum_{j=1}^{s}\frac{\vartheta_{ij}^2}{\vartheta_{i\cdot}\vartheta_{\cdot j}} - 1\right)$$

Таблиця спостережуваних даних

Складемо таблицю за спостережуваними даними навмання обраних $n = 500$ та $m = 500$ елементів із вибірок результатів ЗНО з математики для чоловіків та жінок:

	Низькі бали	Помірні бали	Високі бали	Всього
Чоловіки	270	165	65	500
Жінки	268	174	58	500
Всього	538	339	123	1000

Табл. 6: Таблиця спостережуваних значень з математики

Обчислимо значення статистики критерію:

$$\rho = 1000\left(\frac{270^2}{538 \cdot 500} + \frac{165^2}{339 \cdot 500} + \ldots + \frac{174^2}{339 \cdot 500} + \frac{58^2}{123 \cdot 500} - 1\right) = 0.64$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичної точки

$$\chi^2_{1-\alpha;\ (k-1)(s-1)} = \chi^2_{0.99;\ (2-1)(3-1)} = \chi^2_{0.99;\ 2} = 9.21$$

Висновок

Оскільки значення статистики критерію є меншим за значення критичної точки, то згідно критерію гіпотеза про однорідність статистичних даних приймається. Отже, ймовірності спостереження оцінок різного рівня у вибірках для чоловіків та жінок, відповідно, є однаковими. Тобто стать ніяким чином не впливає на рівень отриманого балу.

Значення `p-value`

Нехай випадкова величина τ має такий самий розподіл, як і статистика критерію й при цьому не залежить від неї: $\tau \sim \chi^2(2)$. Тоді величина `p-value` обчислюється як така ймовірність:

$$P(\tau > \rho \mid H_0) = P(\tau > 0.64) = 1 - P(\tau \leqslant 0.64) = 1 - F_\tau(0.64) = 0.7261$$

Тож для довільних значень $\alpha <$ `p-value` гіпотеза H_0 приймається.

Перевірка гіпотези рівності дисперсій

Маємо дві незалежні вибірки $\{X_i\}$ та $\{Y_j\}$ однаково розподілених випадкових величин, розподіли F_x й F_y яких нам невідомі:

$$X_1 \dots X_n \sim F_x \qquad \text{результати чоловіків}$$
$$Y_1 \dots Y_m \sim F_y \qquad \text{результати жінок}$$

Проведемо нормалізацію даних крок за кроком, як це зображено на схемі (1.6) розділу «Нормалізація даних». У результаті отримаємо такі знормовні вибірки:

$$\overline{X^1} \dots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma_x^2) \qquad \text{середні результати чоловіків,}$$
$$\overline{Y^1} \dots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma_y^2) \qquad \text{середні результати жінок,}$$

при цьому

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j, \ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j, \ i = \overline{1, N}$$

Побудуємо гістограми отриманих вибірок та переконаємося, що вони справді мають нормальний розподіл:

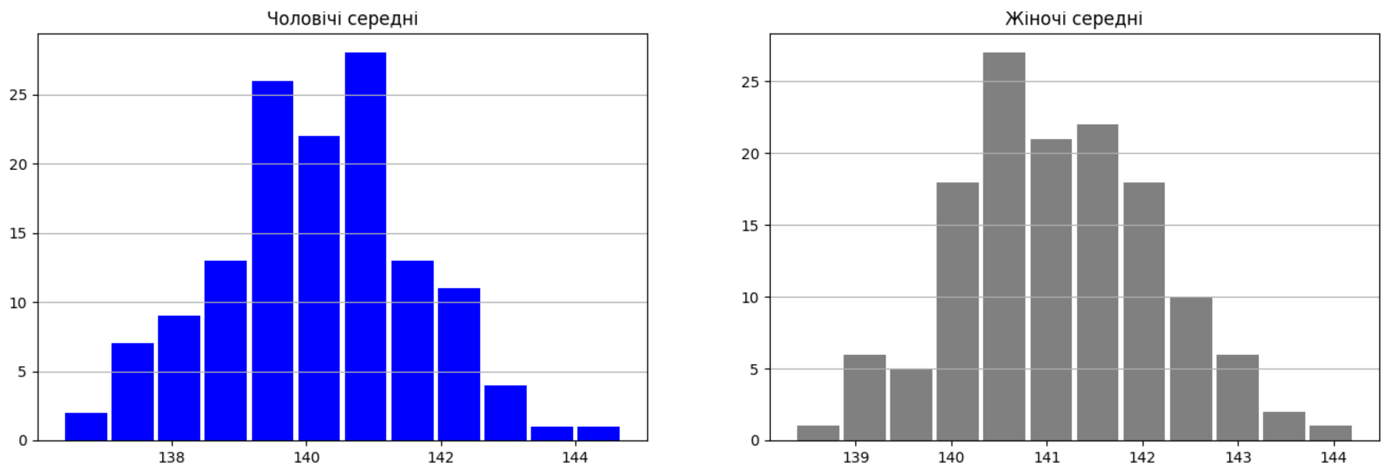


Рис. 7: Гістограми наборів усереднених результатів ЗНО з математики

Як бачимо, криві в обох випадках мають схожі риси. Тож знову висунемо гіпотезу (1.7) про рівність дисперсій вибірок результатів чоловіків та жінок:

$$\text{нульова гіпотеза} \qquad H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1$$

$$\text{проти альтернативи} \qquad H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1$$

При істиності нульової гіпотези статистика критерію Фішера (1.9) буде такою:

$$\rho \equiv \rho(\overline{x^1} \dots \overline{x^N}, \ \overline{y^1} \dots \overline{y^N}) = \frac{S_Y^2}{S_X^2} \sim F\left(N-1, \ N-1\right)$$

А критерієм перевірки гіпотези слугуватиме правило

$$\delta \equiv \delta(\overline{x^1} \ldots \overline{x^N},\ \overline{y^1} \ldots \overline{y^N}) = \begin{cases} H_1, & \rho \leqslant k^* \text{ або } \rho \geqslant k^{**} \\ H_0, & \rho \in (k^*,\ k^{**}) \end{cases},$$

де значення критичних точок k^* та k^{**} при вказаному рівні значущості α є квантилями розподілу Фішера із $(N-1,\ N-1)$ степенями свободи (1.11):

$$k^* = f_{\frac{\alpha}{2};\ (N-1,\ N-1)}$$
$$k^{**} = f_{1-\frac{\alpha}{2};\ (N-1,\ N-1)}$$

Розіб'ємо усю множину спостережуваних значень на $N = 137$ неперетинних множин по $n = 400$ елементів у кожній. У такому випадку статистична похибка обчислень є малою (детальніше – на сторінці 27). Наведемо потрібні величини для перевірки гіпотези рівності дисперсій у таблиці нижче:

n	N	α	S_X^2	S_Y^2
400	137	0.01	1.8055	1.4496

Табл. 7: Значення шуканих параметрів F-тесту

Обчислимо значення статистики критерію:

$$\rho = \frac{S_Y^2}{S_X^2} = 0.8029$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичних точок

$$k^* = f_{\frac{\alpha}{2};\ (N-1,\ N-1)} = f_{0.005;\ (136,\ 136)} = 0.6412$$
$$k^{**} = f_{1-\frac{\alpha}{2};\ (N-1,\ N-1)} = f_{0.995;\ (136,\ 136)} = 1.5595$$

Висновок

Оскільки значення статистики критерію ρ входить у проміжок $(0.6412,\ 1.5595)$, то згідно критерію на рівні значущості $\alpha = 0.01$ нульова гіпотеза $H_0 : \sigma_x^2 = \sigma_y^2$ приймається. Для подільших міркувань введемо таке позначення: $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

Побудова довірчого інтервалу для різниці мат. сподівань

В результаті нормалізації початкових даних та перевірки гіпотези рівності дисперсій отримуємо такі дві вибірки:

$$\overline{X^1}\dots\overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати чоловіків,} \tag{2.1}$$

$$\overline{Y^1}\dots\overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати жінок,} \tag{2.2}$$

при цьому

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j^i,\ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j^i,\ i = \overline{1, N}$$

Спробуємо оцінити, в якому інтервалі лежить значення різниці середніх балів для вибірок результатів ЗНО з математики чоловіків та жінок. Виконавши аналогічні перетворення, які наведенні на сторінці 14, отримаємо шукану центральну статистику G (1.18) для параметра $\theta = \mu_y - \mu_x$:

$$G = \left((\overline{Y} - \overline{X}) - \theta\right) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} \sim t\,[2(N-1)]$$

Побудуємо довірчий інтервал рівня довіри γ :

$$\gamma \overset{\text{def}}{=} P(g_1 < G < g_2) = \int\limits_{g_1}^{g_2} f_G(x)\, dx$$

В силу симетричності розподілу Ст'юдента, найкоротший центральний довірчий інтервал матиме вид:

$$\gamma = P(g_1 < G < g_2) = P(|\,G\,| < g),$$

де значення g є квантилем рівня $\frac{1+\gamma}{2}$ розподілу Ст'юдента із $2(N-1)$ степенями свободи (Рис. 4). Таким чином

$$\gamma = P\left(-t_{\frac{1+\gamma}{2};\ 2(N-1)} < \left((\overline{Y} - \overline{X}) - \theta\right) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} < t_{\frac{1+\gamma}{2};\ 2(N-1)}\right) \tag{2.3}$$

Останнім кроком вкажемо конкретні значення усіх необхідних величин:

n	N	γ	$g = t_{0.975;\ 272}$	S_X^2	S_Y^2	$\overline{Y} - \overline{X}$
400	137	0.95	1.97	1.8055	1.4496	0.9376

Табл. 8: Значення параметрів для довірчого інтервалу різниці середніх

Підставивши у вираз (2.3) значення, які наведені у таблиці вище, отримаємо такий довірчий інтервал для параметра $\theta = \mu_y - \mu_x$:

$$P(0.63 < \theta < 1.24) = 0.95 \tag{2.4}$$

Висновок

Отримано довірчий інтервал рівня довіри $\gamma = 0.95$ для величини різниці середніх значень оцінок з математики вибірок результатів чоловіків та жінок: $(0.63,\ 1.24) \ni \mu_y - \mu_x$. Хоча у вказаному проміжку немає нульового значення, тим не менш лівий кінець відрізку є близьким до $\{0\}$, що дає підстави перевірити гіпотезу про абсолютну нерозрізнюваність середніх.

Так чи інакше, наразі можемо стверджувати, що при порівнянні результатів чоловіка та жінки оцінка з математики у жінки буде більшою за оцінку у чоловіка на бал у 0.94 ± 0.3 пункти, при цьому в середньому у п'яти зі ста таких порівнянь вказане наближення може бути хибним.

Перевірка гіпотези рівності математичних сподівань

Нехай задано дві незалежні нормально розподілені вибірки (2.1)-(2.2) із невідомими математичними сподіваннями та невідомими, проте однаковими дисперсіями:

$$\overline{X^1} \dots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати чоловіків}$$
$$\overline{Y^1} \dots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати жінок}$$

Перевіримо гіпотезу рівності математичних сподівань:

нульова гіпотеза	$H_0 : \mu_y = \mu_x$
проти односторонньої альтернативи	$H_1 : \mu_y > \mu_x$

Або в еквівалентному записі:

нульова гіпотеза	$H_0 : \mu_y - \mu_x = 0$	(2.5)
проти односторонньої альтернативи	$H_1 : \mu_y - \mu_x > 0$	

Скористаємося побудованою у попередньому розділі статистикою (1.18). При істинності нульової гіпотези $H_0 : \mu_y - \mu_x = 0$ отримаємо таку статистику критерію:

$$\rho \equiv \rho(\overline{x^1} \dots \overline{x^N},\ \overline{y^1} \dots \overline{y^N}) = (\overline{Y} - \overline{X}) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} \sim t\,[2(N-1)] \tag{2.6}$$

Тоді перевищення цією статистикою певного критичного значення формуватиме критичну область нульової гіпотези. Як наслідок критерієм перевірки гіпотези слугуватиме правило

$$\delta \equiv \delta(\overline{x^1} \dots \overline{x^N},\ \overline{y^1} \dots \overline{y^N}) = \begin{cases} H_1, & \rho \geqslant k \\ H_0, & \rho < k \end{cases} \tag{2.7}$$

Значення критичної точки k знайдемо з означення помилки першого роду:

$$\alpha \stackrel{\text{def}}{=} P(\delta \in K_\alpha \mid H_0) = P(\rho \geqslant k \mid H_0) = 1 - F_t(k)$$

Таким чином k є квантилем розподілу Ст'юдента із $[2(N-1)]$ степенями свободи при вказаному рівні значущості α:

$$k = t_{1-\alpha;\ 2(N-1)} \tag{2.8}$$

Скористаємося наведеними у Табл. 8 значеннями для перевірки гіпотези рівності математичних сподівань. Обчислимо значення статистики критерію:

$$\rho = (\overline{Y} - \overline{X}) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} = 6.0827$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичної точки

$$k = t_{1-\alpha;\ 2(N-1)} = t_{0.99;\ 272} = 2.34$$

Висновок

Оскільки значення статистики критерію перевищує критичну точку, то згідно критерію (2.7) на рівні значущості $\alpha = 0.01$ нульова гіпотеза $H_0 : \mu_y = \mu_x$ відхиляється. Більш того, при спробі обчислити `p-value` натикаємося на невиправдано малі значення α, при яких гіпотеза H_0 приймалася б. Отже, навіть попри близькість лівого кінця довірчого інтервалу (2.4) до нуля, цією межою не можна знехтувати.

Побудова довірчого інтервалу для значення дисперсії

Наостанок віднайдемо проміжок можливих значень дисперсії нормованих вибірок. В силу прийняття гіпотези рівності дисперсій (сторінка 23) із оригінальних наборів

$$\overline{X^1} \ldots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати чоловіків}$$
$$\overline{Y^1} \ldots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати жінок}$$

аналогічними кроками, як це виконано на сторінці 14, можна сформувати випадкову величину (1.16) розподілу хі-квадрат із $[2(N-1)]$ степенями свободи, при цьому для параметра дисперсії $\frac{1}{n}\sigma^2$ ця величина відіграватиме роль центральної статистики:

$$G \equiv \frac{(N-1)}{\sigma^2/n} \left(S_X^2 + S_Y^2\right) \sim \chi^2 \left[2(N-1)\right] \tag{2.9}$$

А тоді довірчий інтервал рівня довіри γ матиме вид:

$$\gamma \stackrel{\text{def}}{=} P(g_1 < G < g_2) = \int\limits_{g_1}^{g_2} f_G(x)\ dx \tag{2.10}$$

Значення g_1 та g_2, як це показано на Рис. 5, є квантилями рівнів $\frac{1-\gamma}{2}$ та $\frac{1+\gamma}{2}$ розподілу хі-квадрат із $[2(N-1)]$ степенями свободи:

$$g_1 = \chi^2_{\frac{1-\gamma}{2};\ 2(N-1)},\ \ g_2 = \chi^2_{\frac{1+\gamma}{2};\ 2(N-1)}$$

Підставляючи отримані результати, матимемо:

$$\gamma = P\left(\chi^2_{\frac{1-\gamma}{2};\ 2(N-1)} < \frac{(N-1)}{\sigma^2/n}\left(S_X^2 + S_Y^2\right) < \chi^2_{\frac{1+\gamma}{2};\ 2(N-1)}\right) \tag{2.11}$$

Наостанок вкажемо конкретні значення усіх необхідних величин:

n	N	γ	$g_1 = \chi^2_{0.025;\ 272}$	$g_2 = \chi^2_{0.975;\ 272}$	S_X^2	S_Y^2
400	137	0.95	228.21	319.58	1.8055	1.4496

Табл. 9: Значення параметрів для довірчого інтервалу дисперсії

Підставивши у вираз (2.11) наведені вище значення, отримаємо такий довірчий інтервал для параметра $\frac{1}{n}\sigma^2$:

$$P(1.39\ < \frac{\sigma^2}{n} <\ 1.94) = 0.95$$

Висновок

Таким чином для сформованих нормально розподілених вибірок (2.1)-(2.2) значення дисперсії в середньому лише у п'яти елементів вибірки зі ста буде виходити за межі проміжку від 1.39 до 1.94 балів.

Статистична похибка обчислень

Статистичною похибкою обчислень називатимемо найменшу величину Δ відхилення різниці вибіркових середніх від істинного значення різниці математичних сподівань заданих вибірок (2.1)-(2.2) з щонайменшою ймовірністю у 95%. Тобто шукана похибка Δ обчислюватиметься як така ймовірність:

$$P\left(\ \left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \leqslant \Delta\ \right) \geqslant 0.95$$

Виконавши аналогічні міркування, які наведені на сторінці 18, отримаємо, що величина похибки Δ виражатиметься через квантиль рівня 0.975 розподілу Ст'юдента з $[2(N-1)]$ степенями свободи:

$$\Delta \geqslant t_{0.975;\ 2(N-1)} \cdot \sqrt{\frac{S_X^2 + S_Y^2}{N}} \tag{2.12}$$

Таблиця необхідних значень виглядатиме таким чином:

N	$t_{0.975;\ 272}$	S_X^2	S_Y^2
137	1.97	1.8055	1.4496

Табл. 10: Значення параметрів для обчислення статистичної похибки

Тож остаточно найменша величина відхилення $\Delta = 0.3$, а тоді початкова ймовірнісна нерівність матиме вид:

$$P\left(\ \left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \leqslant 0.3\ \right) \geqslant 0.95$$

Висновок

Для встановлених значень $n = 400$ й відповідно $N = 137$ статистична похибка відхилення різниці вибіркових середніх від істинного значення різниці математичних сподівань складає $\Delta = 0.3$ бала з ймовірністю, більшою за 95%. Можемо стверджувати, що обчислення за такого значення похибки мають високу точність, оскільки величина $\Delta = 0.3$ є лише 0.15% від можливих 200 балів за результат тесту.

3 Результати з англійської мови

Оглянемо отримані 78 797 результатів, де жіночих є на 8 339 більше за чоловічих. Зобразимо гістограми результатів ЗНО з англійської мови 2019 року для вибірки, наприклад, 10 000 учнів:

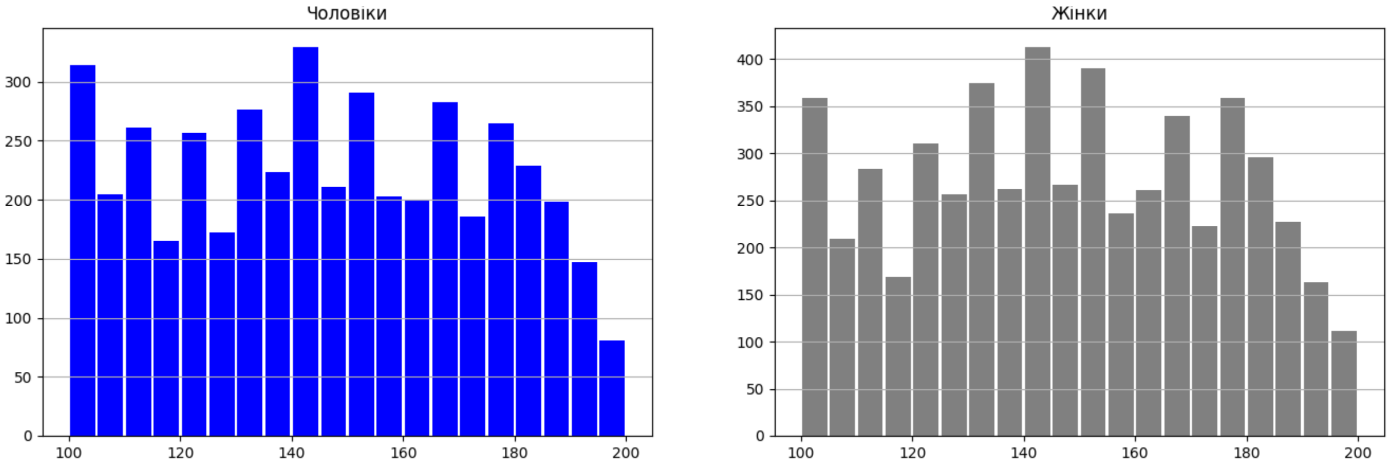


Рис. 8: Результати з англійської мови

З'ясуємо, чи відмінність між результатами для жінок та чоловіків є статистично значущою.

Перевірка гіпотези однорідності даних

Нехай маємо дві незалежні вибірки $\{X_i\}$ та $\{Y_j\}$ однаково розподілених випадкових величин, розподіли F_x й F_y яких нам невідомі:

$X_1 \dots X_n \sim F_x$ результати чоловіків

$Y_1 \dots Y_m \sim F_y$ результати жінок

Перевіримо гіпотезу про однорідність статистичного матеріалу, тобто гіпотезу, що ймовірності спостереження умовно високих, помірних та низьких балів в обох вибірках є однаковими. Таким чином матимемо $k = 2$ вибірок, в яких елементи приймають $s = 3$ різних значень. Знову за аналогічних позначень гіпотеза перевірки однорідності спосережуваних даних формулюватиметься так само як і у виразах (1.1):

нульова гіпотеза $H_0 : p_{11} = p_{21},\ p_{12} = p_{22},\ p_{13} = p_{23}$

проти альтернативи $H_1 : \exists i, j \quad p_{1,j} \neq p_{2,j}$

А тоді критерієм перевірки гіпотези рівно як і у випадку обробки результатів з української мови (1.2) слугуватиме правило

$$\delta(x_1 \dots x_n,\ y_1 \dots y_m) = \begin{cases} H_1, & \rho \geqslant \chi^2_{1-\alpha;\ (k-1)(s-1)} \\ H_0, & \rho < \chi^2_{1-\alpha;\ (k-1)(s-1)} \end{cases}$$

При цьому статистика критерію ρ обчислюється аналогічно до формули (1.3):

$$\rho \equiv \rho(x_1 \dots x_n,\ y_1 \dots y_m) = (n+m)\left(\sum_{i=1}^{k}\sum_{j=1}^{s}\frac{\vartheta_{ij}^2}{\vartheta_{i\cdot}\vartheta_{\cdot j}} - 1\right)$$

Таблиця спостережуваних даних

Зобразимо дані для навмання обраних $n = 500$ та $m = 500$ елементів із вибірок результатів ЗНО з англійської мови для чоловіків та жінок:

	Низькі бали	Помірні бали	Високі бали	Всього
Чоловіки	248	168	84	500
Жінки	224	196	80	500
Всього	472	364	164	1000

Табл. 11: Таблиця спостережуваних значень з англійської мови

Обчислимо значення статистики критерію:

$$\rho = 1000\left(\frac{228^2}{472\cdot 500} + \frac{168^2}{364\cdot 500} + \ldots + \frac{196^2}{364\cdot 500} + \frac{80^2}{164\cdot 500} - 1\right) = 3.47$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичної точки

$$\chi^2_{1-\alpha;\ (k-1)(s-1)} = \chi^2_{0.99;\ (2-1)(3-1)} = \chi^2_{0.99;\ 2} = 9.21$$

Висновок

Оскільки значення статистики критерію є меншим за значення критичної точки, то згідно критерію гіпотеза про однорідність статистичних даних приймається. Отже, ймовірності спостереження оцінок різного рівня у вибірках для чоловіків та жінок, відповідно, є однаковими. Тобто так само, як і у випадку результатів з математики, стать ніяким чином не впливає на рівень отриманого балу.

Значення `p-value`

Нехай випадкова величина τ має такий самий розподіл, як і статистика критерію й при цьому не залежить від неї: $\tau \sim \chi^2(2)$. Тоді величина `p-value` обчислюється як така ймовірність:

$$P(\tau > \rho \mid H_0) = P(\tau > 3.47) = 1 - P(\tau \leqslant 3.47) = 1 - F_\tau(3.47) = 0.1764$$

Тож для довільних значень $\alpha <$ `p-value` гіпотеза H_0 приймається.

Перевірка гіпотези рівності дисперсій

Для заданих незалежних вибірок $\{X_i\}$ та $\{Y_j\}$ однаково розподілених випадкових величин, розподіли F_x й F_y яких нам невідомі:

$$\begin{aligned} &X_1 \dots X_n \sim F_x && \text{результати чоловіків} \\ &Y_1 \dots Y_m \sim F_y && \text{результати жінок} \end{aligned}$$

Проведемо нормалізацію даних крок за кроком, як це зображено на схемі (1.6) розділу «Нормалізація даних». У результаті отримаємо такі знормовні вибірки:

$$\begin{aligned} &\overline{X^1} \dots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma_x^2) && \text{середні результати чоловіків,} \\ &\overline{Y^1} \dots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma_y^2) && \text{середні результати жінок,} \end{aligned}$$

при цьому

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j,\ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j,\ i = \overline{1, N}$$

Побудуємо гістограми отриманих вибірок та переконаємося, що вони справді мають нормальний розподіл (Рис. 9).

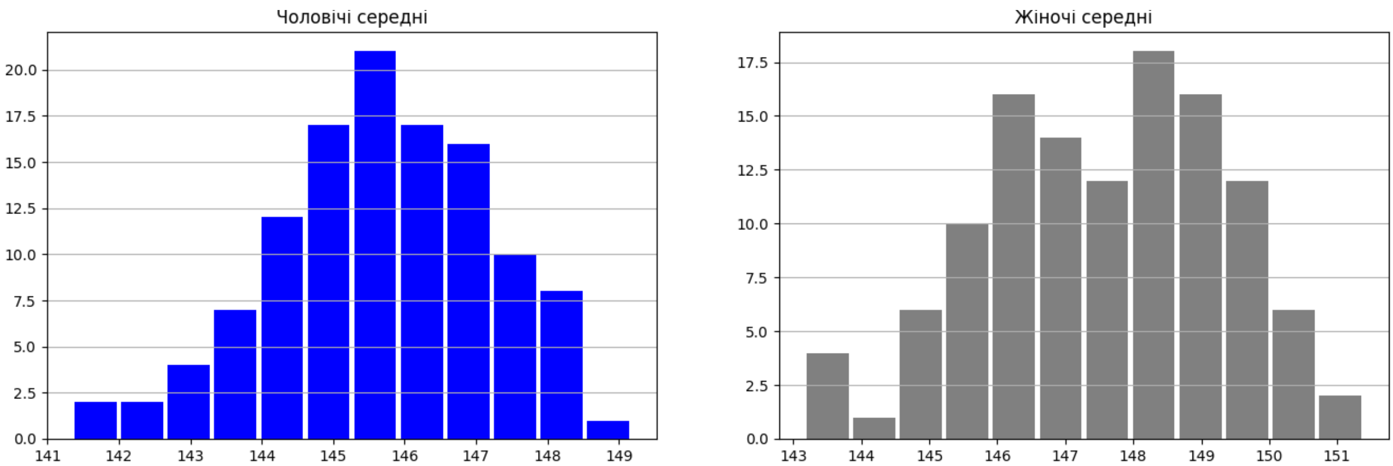


Рис. 9: Гістограми наборів усереднених результатів ЗНО з англійської мови

Знову ж таки бачимо, що криві в обох випадках мають схожі риси. Тому висунемо припущення (1.7), що дисперсії цих вибірок однакові:

$$\begin{aligned} &\text{нульова гіпотеза} && H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1 \\ &\text{проти альтернативи} && H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1 \end{aligned}$$

При істиності нульової гіпотези статистика критерію Фішера (1.9) буде такою:

$$\rho \equiv \rho(\overline{x^1} \dots \overline{x^N},\ \overline{y^1} \dots \overline{y^N}) = \frac{S_Y^2}{S_X^2} \sim F\,(N-1,\ N-1)$$

А критерієм перевірки гіпотези слугуватиме правило

$$\delta \equiv \delta(\overline{x^1} \dots \overline{x^N},\ \overline{y^1} \dots \overline{y^N}) = \begin{cases} H_1, & \rho \leqslant k^* \text{ або } \rho \geqslant k^{**} \\ H_0, & \rho \in (k^*,\ k^{**}) \end{cases},$$

де значення критичних точок k^* та k^{**} при вказаному рівні значущості α є квантилями розподілу Фішера із $(N-1,\ N-1)$ степенями свободи (1.11):

$$k^* = f_{\frac{\alpha}{2};\ (N-1,\ N-1)}$$
$$k^{**} = f_{1-\frac{\alpha}{2};\ (N-1,\ N-1)}$$

Розіб'ємо усю множину спостережуваних значень на $N = 117$ неперетинних множин по $n = 300$ елементів у кожній. У такому випадку статистична похибка обчислень є малою (детальніше – на сторінці 35). Наведемо потрібні величини для перевірки гіпотези рівності дисперсій у таблиці нижче:

n	N	α	S_X^2	S_Y^2
300	117	0.01	3.0051	2.2830

Табл. 12: Значення шуканих параметрів F-тесту

Обчислимо значення статистики критерію:

$$\rho = \frac{S_Y^2}{S_X^2} = 0.7597$$

Водночас на рівні значущості $\alpha = 0.01$ значення критичних точок

$$k^* = f_{\frac{\alpha}{2};\ (N-1,\ N-1)} = f_{0.005;\ (116,\ 116)} = 0.6178$$
$$k^{**} = f_{1-\frac{\alpha}{2};\ (N-1,\ N-1)} = f_{0.995;\ (116,\ 116)} = 1.6187$$

Висновок

Оскільки значення статистики критерію ρ входить у проміжок $(0.6178,\ 1.6187)$, то згідно критерію на рівні значущості $\alpha = 0.01$ нульова гіпотеза $H_0 : \sigma_x^2 = \sigma_y^2$ приймається. Для подільших міркувань введемо таке позначення: $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

Побудова довірчого інтервалу для різниці мат. сподівань

Аналогічним чином в результаті нормалізації початкових даних та перевірки гіпотези рівності дисперсій отримуємо такі дві вибірки:

$$\overline{X^1}\dots\overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати чоловіків,} \tag{3.1}$$

$$\overline{Y^1}\dots\overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати жінок,} \tag{3.2}$$

при цьому

$$\overline{X^i} = \frac{1}{n}\sum_{j=1}^{n} X_j^i,\ \overline{Y^i} = \frac{1}{n}\sum_{j=1}^{n} Y_j^i,\ i = \overline{1, N}$$

Спробуємо оцінити, в якому інтервалі лежить значення різниці середніх балів для вибірок результатів ЗНО з англійської мови чоловіків та жінок. Виконавши перетворення, які наведенні на сторінці 14, таким самим чином отримаємо шукану центральну статистику G (1.18) для параметра $\theta = \mu_y - \mu_x$:

$$G = \left((\overline{Y} - \overline{X}) - \theta\right) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} \sim t\,[2(N-1)]$$

Побудуємо довірчий інтервал рівня довіри γ :

$$\gamma \overset{\text{def}}{=} P(g_1 < G < g_2) = \int_{g_1}^{g_2} f_G(x)\ dx$$

В силу симетричності розподілу Ст'юдента, найкоротший центральний довірчий інтервал матиме вид:

$$\gamma = P(g_1 < G < g_2) = P(|\,G\,| < g),$$

де значення g є квантилем рівня $\frac{1+\gamma}{2}$ розподілу Ст'юдента із $2(N-1)$ степенями свободи (Рис. 4), тобто

$$\gamma = P\left(-t_{\frac{1+\gamma}{2};\ 2(N-1)} < \left((\overline{Y} - \overline{X}) - \theta\right) \cdot \sqrt{\frac{N}{S_X^2 + S_Y^2}} < t_{\frac{1+\gamma}{2};\ 2(N-1)}\right) \tag{3.3}$$

Останнім кроком вкажемо конкретні значення усіх необхідних величин:

n	N	γ	$g = t_{0.975;\ 232}$	S_X^2	S_Y^2	$\overline{Y} - \overline{X}$
300	117	0.95	1.97	3.0051	2.2830	1.9380

Табл. 13: Значення параметрів для довірчого інтервалу різниці середніх

Підставивши у вираз (3.3) значення, які наведені у таблиці вище, отримаємо такий довірчий інтервал для параметра $\theta = \mu_y - \mu_x$:

$$P(1.52 < \theta < 2.36) = 0.95 \tag{3.4}$$

Висновок

Отримано довірчий інтервал рівня довіри $\gamma = 0.95$ для величини різниці середніх значень оцінок з англійської мови вибірок результатів чоловіків та жінок: $(1.52,\ 2.36) \ni \mu_y - \mu_x$. Оскільки у вказаному проміжку немає нульового значення, гіпотезу про нерозрізнюваність середніх можна відхилити.

Можемо стверджувати, що при порівнянні результатів чоловіка та жінки оцінка з англійської мови у жінки буде більшою за оцінку у чоловіка на бал у 1.94 ± 0.42 пункти, при цьому в середньому у п'яти зі ста таких порівнянь вказане наближення може бути хибним.

Побудова довірчого інтервалу для значення дисперсії

Наостанок знову віднайдемо проміжок можливих значень дисперсії нормованих вибірок (3.1)-(3.2). В силу прийняття гіпотези рівності дисперсій із оригінальних наборів

$$\overline{X^1} \ldots \overline{X^N} \sim N(\mu_x, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати чоловіків}$$
$$\overline{Y^1} \ldots \overline{Y^N} \sim N(\mu_y, \tfrac{1}{n}\sigma^2) \qquad \text{середні результати жінок}$$

аналогічними кроками (сторінка 14) можна сформувати випадкову величину (формула 1.16) розподілу хі-квадрат із $[2(N-1)]$ степенями свободи, при цьому для параметра дисперсії $\frac{1}{n}\sigma^2$ ця величина відіграватиме роль центральної статистики:

$$G \equiv \frac{(N-1)}{\sigma^2/n}\left(S_X^2 + S_Y^2\right) \sim \chi^2\left[2(N-1)\right] \tag{3.5}$$

А тоді довірчий інтервал рівня довіри γ матиме вид:

$$\gamma \overset{\text{def}}{=} P(g_1 < G < g_2) = \int\limits_{g_1}^{g_2} f_G(x)\, dx \tag{3.6}$$

Значення g_1 та g_2 (Рис. 5) є квантилями рівнів $\frac{1-\gamma}{2}$ та $\frac{1+\gamma}{2}$ розподілу хі-квадрат із $[2(N-1)]$ степенями свободи:

$$g_1 = \chi^2_{\frac{1-\gamma}{2};\ 2(N-1)}, \quad g_2 = \chi^2_{\frac{1+\gamma}{2};\ 2(N-1)}$$

Підставляючи отримані результати, матимемо такий ймовірнісний вираз:

$$\gamma = P\left(\chi^2_{\frac{1-\gamma}{2};\ 2(N-1)} < \frac{(N-1)}{\sigma^2/n}\left(S_X^2 + S_Y^2\right) < \chi^2_{\frac{1+\gamma}{2};\ 2(N-1)}\right) \tag{3.7}$$

На завершення вкажемо конкретні значення усіх необхідних величин:

n	N	γ	$g_1 = \chi^2_{0.025;\ 232}$	$g_2 = \chi^2_{0.975;\ 232}$	S_X^2	S_Y^2
300	117	0.95	191.71	276.08	3.0051	2.2830

Табл. 14: Значення параметрів для довірчого інтервалу дисперсії

Підставивши у вираз (3.7) наведені вище значення, отримаємо такий довірчий інтервал для параметра $\frac{1}{n}\sigma^2$:

$$P(2.22\ < \frac{\sigma^2}{n} <\ 3.20) = 0.95$$

Висновок

Тож для сформованих нормально розподілених вибірок (3.1)-(3.2) значення дисперсії в середньому лише у п'яти елементів вибірки зі ста буде виходити за межі проміжку від 2.22 до 3.20 балів.

Статистична похибка обчислень

Статистичною похибкою обчислень називатимемо найменшу величину Δ відхилення різниці вибіркових середніх від істинного значення різниці математичних сподівань заданих вибірок (3.1)-(3.2) з щонайменшою ймовірністю у 95%. Тобто шукана похибка Δ обчислюватиметься як така ймовірність:

$$P\left(\ \left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \leqslant \Delta\ \right) \geqslant 0.95$$

Виконавши аналогічні міркування, які наведені на сторінці 18, отримаємо, що величина похибки Δ виражатиметься через квантиль рівня 0.975 розподілу Ст'юдента з $[2(N-1)]$ степенями свободи:

$$\Delta \geqslant t_{0.975;\ 2(N-1)} \cdot \sqrt{\frac{S_X^2 + S_Y^2}{N}} \tag{3.8}$$

Таблиця необхідних значень виглядатиме так:

N	$t_{0.975;\ 232}$	S_X^2	S_Y^2
117	1.97	3.0051	2.2830

Табл. 15: Значення параметрів для обчислення статистичної похибки

Тож остаточно найменша величина відхилення $\Delta = 0.42$, а тоді початкова ймовірнісна нерівність матиме вид:

$$P\left(\left|(\overline{Y} - \overline{X}) - (\mu_y - \mu_x)\right| \leqslant 0.42 \right) \geqslant 0.95$$

Висновок

Для встановлених значень $n = 300$ й відповідно $N = 117$ статистична похибка відхилення різниці вибіркових середніх від істинного значення різниці математичних сподівань складає $\Delta = 0.42$ бала з ймовірністю, більшою за 95%. Можемо стверджувати, що обчислення за такого значення похибки мають високу точність, оскільки величина $\Delta = 0.42$ є лише 0.2% від можливих 200 балів за результат тесту.

Загальний висновок

Метою статистичного аналізу було з'ясувати, чи є значущою відмінність між балами ЗНО для чоловіків та жінок. При цьому розглядалися результати трьох різних предметів: української мови, математики та англійської мови.

Українська мова

Шляхом перевірки гіпотези однорідності даних перш за все вдалося встановити, що на рівні значущості $\alpha = 0.01$ лише для української мови наявна залежність між рівнем набраного балу та чинником статі. При спробі побудови інтервалу можливих значень різниці середніх балів жінок та чоловіків, було отримано проміжок у 12.08 ± 0.2 пункти, що свідчить про значне перевищення оцінок жінок у порівнянні з оцінками чоловіків. Статистична похибка обчислень при цьому складає лише 0.21 пункти.

Крім того, значення вказаного довірчого інтервалу рівня довіри $\gamma = 0.95$ додатково підтверджує хибність гіпотези однорідності статистичних даних. Наостанок зауважимо, що для знормованих вибірок величина дисперсії в середньому лише у п'яти елементів вибірки зі ста виходить за межі проміжку від 1.32 до 1.68 балів.

Математика та англійська мова

У випадку аналізу балів з математики та англійської мови вже на першому кроці виявлено відсутність впливу статі на рівень отриманого балу. До того ж довірчі інтервали різниці середніх оцінок різних статей мають невеликі значення: 0.94 ± 0.3 для математики та 1.94 ± 0.42 для англійської мови. Статистичні похибки обчислень при цьому складають лише 0.3 та 0.42 пункти відповідно. А для знормованих вибірок величина дисперсії в середньому лише у п'яти елементів вибірки зі ста виходить за межі проміжку $(1.39,\ 1.94)$ балів у випадку математики й $(2.22,\ 3.20)$ балів у випадку англійської мови.

Отже, як результат статистичного аналізу можна стверджувати, що лише з української мови наявна статистична значущість відмінності результатів ЗНО в залежності від статі. Водночас на набрані бали з математики чи англійської мови чинник статі не має значного впливу.