**Лабораторна робота**

**Тема:** Перевірка статистичних гіпотез

**Мета:** використовуючи емпіричні данні наукових досліджень навчитись перевіряти статистичні гіпотези на підпорядкованість нормальному закону розподілу вибірок, рівність середніх, наявність кореляційного зв’язку.

**Матеріали та обладнання**: комп’ютери із доступом до мережі Інтернет, із офісним додатком Microsoft Excel та спеціалізованим пакетом для аналізу біологічних даних PAST version 4. PAST – це вільний для поширення програмний пакет, що можна безкоштовно завантажити та встановити на власний комп’ютер використовуючи наступне посилання: <https://folk.uio.no/ohammer/past/>. На цій же сторінці знаходиться посібник для користування в форматі PDF.

ХІД РОБОТИ

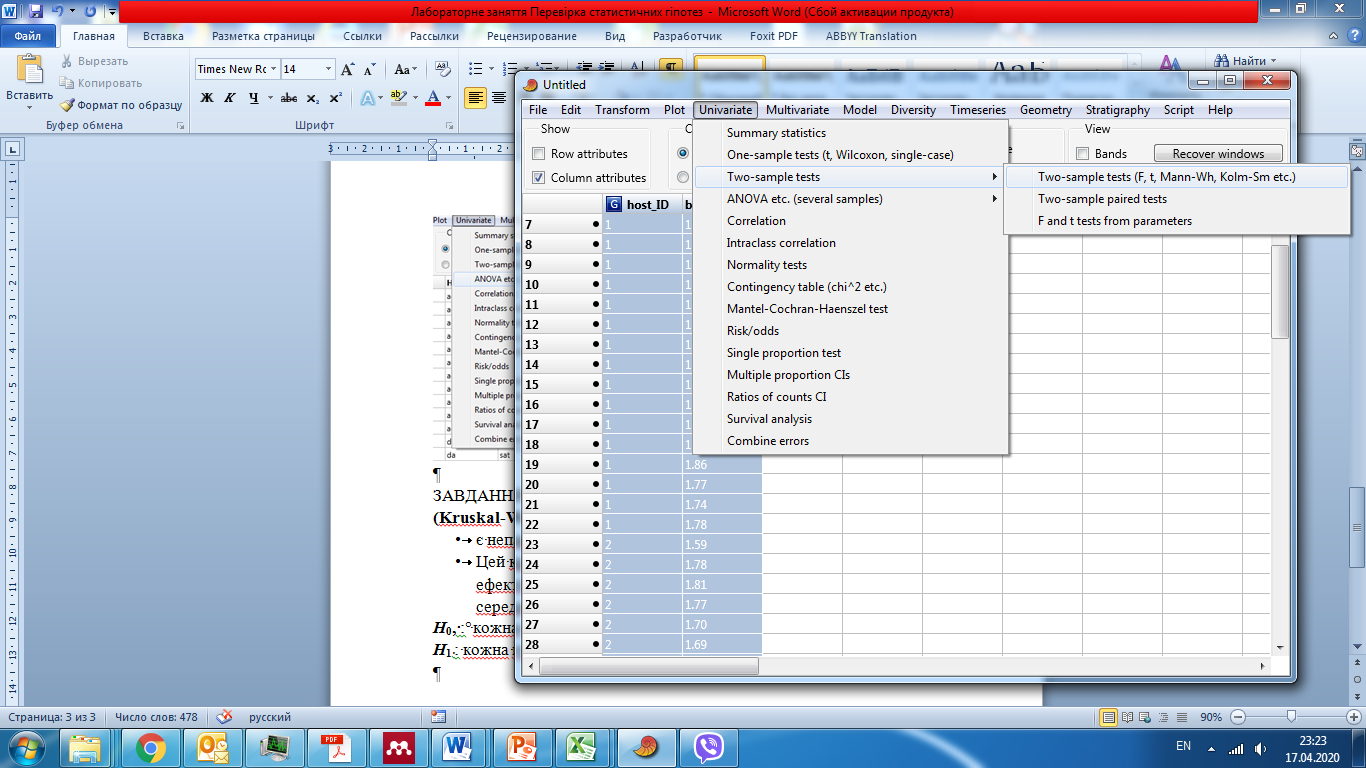
ЗАВДАННЯ 1. Перевірка вибірки на підпорядкованість закону нормального розподілу використовуючи критерій Lilliefors в PAST. За результатами перевірки зробити висновок про підпорядкованість вибірки нормальному розподілу.

* Н0 – вибірка підпорядковується нормальному розподілу, p-значення (англ. p-value) або p-рівень значущості > 0.05
* Н1 - вибірка не підпорядковується нормальному розподілу, P < 0.05

P-рівень = 0,05 (тобто 1/20) показує, що існує 5% імовірність того, що знайдений у вибірці зв'язок між змінними є лише випадковою особливістю даної вибірки. Іншими словами, якщо дана залежність у популяції відсутня, а ви багаторазово проводили б подібні експерименти, то приблизно в одному з двадцяти повторень експерименту можна було б очікувати таку саме або сильнішу залежність між змінними.

ЗАВДАННЯ 2. Перевірка гіпотези про рівність генеральних середніх за допомогою Критерія Стьюдента використовуючи пакет аналізу даних в Excel та PAST.

* Критерій Стьюдента використовується для перевірки гіпотез про рівність генеральних середніх, якщо статистичні дані розподілені за нормальним законом (для цього виконується завдання 1). Формулюються гіпотези:
* Н0 - середні двох генеральних сукупностей рівні
* Н1 - середні двох генеральних сукупностей не рівні
* Перевірка виконується за даними двох вибірок об'ємом n1 та n2.
* При цьому можливо два випадки, коли дисперсії порівнюваних вибірок рівні або ні.

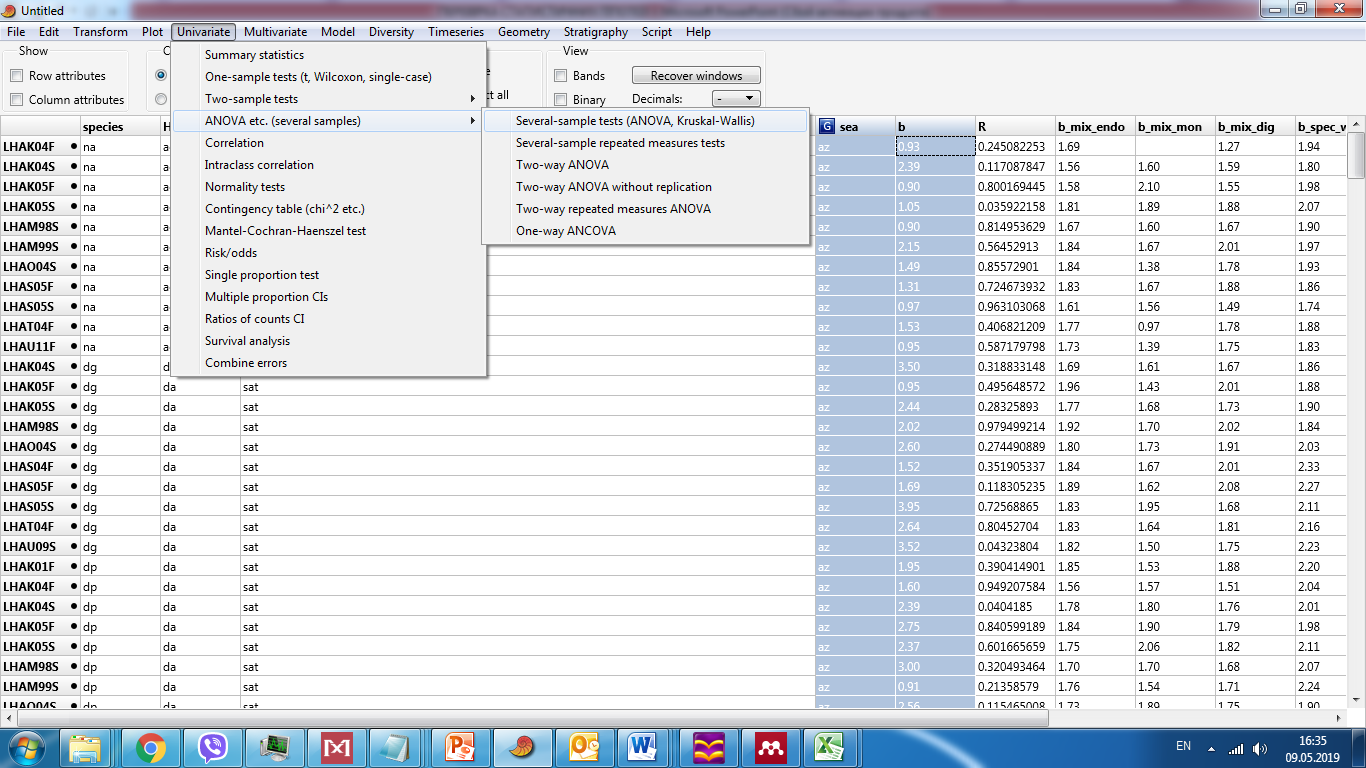


ЗАВДАННЯ 3. Провести порівняння вибірок використовуючи U-критерій Манна-Уітні (доступний у вікні результатів t-тесту в PAST). Цей тест являє собою непараметричну альтернативу t-критерію для незалежних вибірок. Перевага його полягає в тому, що ми відмовляємося від припущення нормальності розподілу і однакових дисперсій. Він дозволяє виявити відмінності в значенні параметра між малими вибірками. Цей метод визначає, чи досить мала зона значень, що перехрещуються, між двома рядами (ранжованим рядом значень параметра в першій вибірці і таким же в другій вибірці). Чим менше значення критерію, тим вірогідніше, що відмінності між значеннями параметра у вибірках достовірні

* Н0 – зона значень, що перехрещуються, між двома рядами велика
* Н1 - зона значень, що перехрещуються, між двома рядами мала

ЗАВДАННЯ 4. Використовуючи метод дисперсійного аналізу провести перевірку статистичних гіпотез відносно середніх значень в кількох генеральних сукупностях, які мають нормальний розподіл (провести попередню перевірку на нормальність розподілу)**.** Завдання виконується у програмному пакеті PAST

* Н0 - середні двох або більше генеральних сукупностей рівні
* Н1 - середні двох або більше генеральних сукупностей не рівні



ЗАВДАННЯ 5. Виконати ранговий однофакторний аналіз Краскела– Уолліса (Kruskal-Wallis test) для перевірки нульової гіпотези про рівність медіан декількох вибірок. Цей критерій є багатовимірним (може порівнювати більше ніж дві вибірки) узагальненням критерію Манна-Уітні (див вище) та вважається непараметричним аналогом однофакторного дисперсійного аналізу. Тест Краскела– Уолліса доступний у вікні результатів ANOVA в PAST.

***Н*0,** :  кожна група має однаковий розподіл величин в популяції;

***Н*1** : кожна група не має однакового розподілу величин в популяції.

ЗАВДАННЯ 6. Побудувати кореляційне поле та провести кореляційний аналіз для двох вибірок.

* У випадку, коли між *Х* та *Y* існує лінійний зв'язок та вибіркові дані розподілені за нормальним законом, використовується **коефіцієнт кореляції Пірсона,** який ще називається параметричним коефіцієнтом кореляції
* коли між *Х* та *Y* існує нелінійний зв'язок або вибіркові дані не розподілені за нормальним законом, варто використовувати коефіцієнт кореляції **Спірмена**
* Коефіцієнт кореляції приймає значення на проміжку [—1;1], тобто —1 < *r* < 1.
* Якщо 0,1 < |r| < 0,5, то зв'язок вважається **слабким**; якщо 0,5 < |r| < 0,7, то зв'язок вважається **середнім**; 0,7 < |r| < 1, то зв'язок вважається **сильним**.
* Якщо *r* > 0, то зв'язок називається додатнім, тобто зі збільшенням значень *Х* значення *Y* також збільшуються. Якщо *r* < 0, то зв'язок називається від'ємним, тобто зі збільшенням значень *Х* значення *Y* зменшуються
* Зв'язок вважається достовірним при р<0.05

Кореляцію за Пірсоном та Спірменом та супровідну статистику розраховуємо в PAST Univariate -> Correlation

* Linear r - кореляцію за Пірсоном
* Spearman’s rs - кореляцію за Спірменом

Результати кореляційного аналізу:

