

# Лабораторна робота №5 з дисципліни "регресійний аналіз" Варіант №4

Горбунова Даніела Денисовича  
4 курс бакалаврату  
група "комп'ютерна статистика"

26 листопада 2020 р.

## 1 Вступ.

У даній роботі перевірили залежність між змінною *dauer* та змінною *alter* (гіпотеза про однаковість середніх та дисперсій незалежно від рівня фактора).

## 2 Обробка даних.

Дані за змінною *dauer* прологарифмовано, внаслідок чого отримали більш чітке уявлення про розподіл. Маємо справду з нормально розподіленими величинами із зрізанням від'ємних значень. Загальна кількість спостережень велика, спробуємо продовжити роботу.

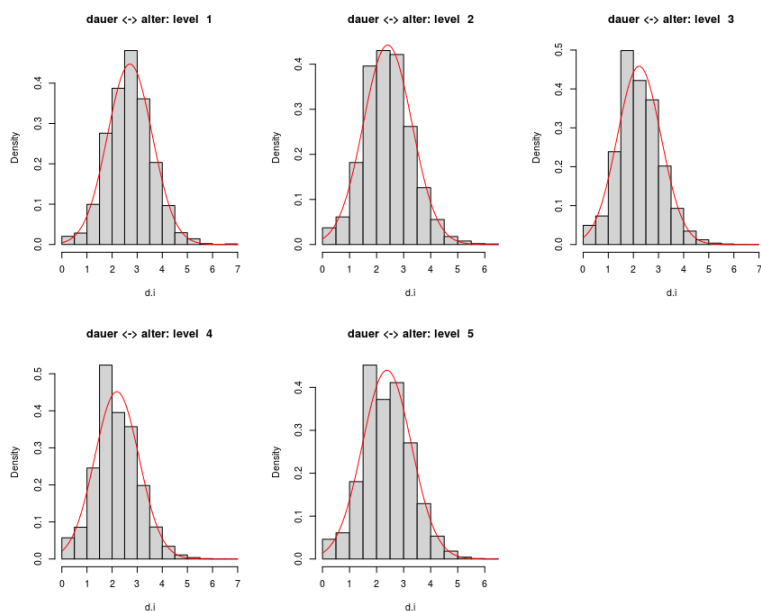


Рис. 1: Гістограма з щільністю нормального розподілу, параметри якого оцінені за методом моментів, для спостережень за змінною *dauer* в залежності від факторного рівня.

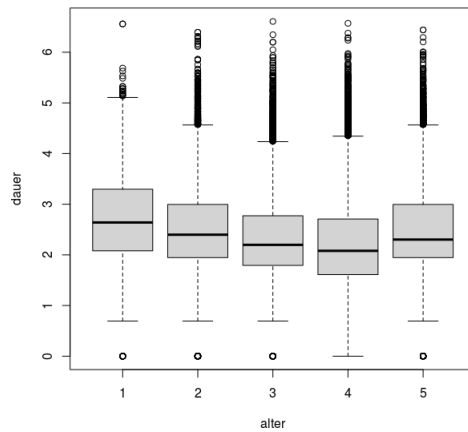


Рис. 2: Скриньки з вусами для спостережень за змінною *dauer* в залежності від факторного рівня. Хвости на малюнку спричинені зрізанням розподілу.

Судячи з останнього рисунку, на якому відображені скриньки з вусами для спостережень в залежності від рівня фактора, є припущення, що середні кожного рівня неоднорідні, однак з дисперсіями можлива інакша ситуація. Але таке вгадування не принесе результатів, тому почнемо із перевірки середніх на однорідність.

### 3 Перевірка гіпотези про однорідність середніх у моделі.

Застосуємо тест Фішера для однофакторного дисперсійного аналізу на перетворених даних.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
alter	1	36	35.57	44.27	2.87e-11 ***
Residuals	130984	105255	0.80		

З отриманої таблиці бачимо, що досягнутий рівень значущості у даному тесті близький до нуля. Тому фіксуємо довільний адекватний рівень значущості (нехай  $\alpha = 0.05$ ), основну гіпотезу про однорідність середніх слід відхилити та прийняти альтернативу про відмінність середніх значень. Конкретно виділимо ті рівні, середні значення яких суттєво відрізняються. Тут ми побудуємо одночасні довірчі інтервали рівня  $\gamma = 0.05$  (звідси номінальний рівень значущості дорівнює  $\gamma_0 = 1 - \sqrt[5]{1 - \gamma} = 1 - \sqrt[5]{0.95} \approx 0.01020622$ ). З рисунку видно (див. наступну сторінку), що суттєві відмінності помітні для середніх 1,3 та 4 рівнів. Для підмножик даних за рівнями 2 та 5 гіпотезу про однорідність слід приймати, якщо брати зовсім малий рівень значущості. Якщо узагальнити, то середні значення в залежності від рівня є неоднорідними.

```
> summary(aov(dauer ~ alter, data = d[(d$alter == 2)|(d$alter == 5),]))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
alter	1	10	9.894	12.09	0.000507 ***
Residuals	54959	44975	0.818		

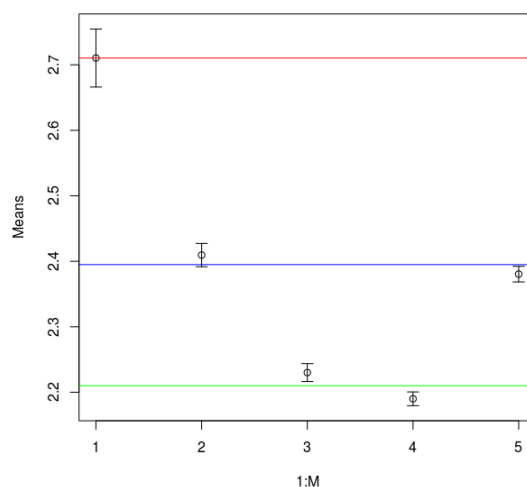


Рис. 3: Одночасні довірчі інтервали для математичних сподівань в залежності від рівня фактора.

## 4 Перевірка гіпотези про однорідність дисперсій у моделі.

Для перевірки основної гіпотези про однорідність дисперсій підмножик в залежності від рівня, спочатку використаємо графічний тест, а потім критерій Левена. Графічний метод полягає у побудові одночасних інтервалів номінального рівня  $\gamma_0$  (тобто підхід аналогічний до того, що розглядали для середніх). Після побудови таких інтервалів, ми бачимо, що перетин таких є непорожньою множиною. Тому з точки зору графічного тесту ми не можемо відхилити основну гіпотезу.

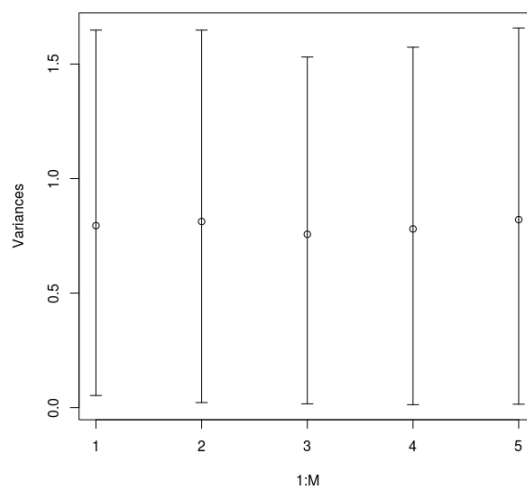


Рис. 4: Одночасні довірчі інтервали для дисперсій в залежності від рівня фактора.

Можна було б на цьому завершити написання звіту, якщо б не одна екзотична ситуація з числовим тестом.

Досягнутий рівень значущості критерію Левена малий. Тому за цим тестом альтернативну гіпотезу слід прийняти. Довіримося прийнятому рішення за першим критерієм хоча б з тієї причини, що воно дещо узгоджується з початковими припущеннями автора під час початкового аналізу спостережень за змінною `dauer`.

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group   4  31.297 < 2.2e-16 ***
130981
```

## 5 Висновки.

За результатами однофакторного дисперсійного аналізу вдалося показати, що математичні сподівання підмножин, в залежності від факторного рівня, є неоднорідними, тобто залежність від фактора наявна. На думку автора, однорідність дисперсій у даній моделі має місце.