

**МІНІСТЕРСТВО ВНУТРІШНІХ СПРАВ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ВНУТРІШНІХ СПРАВ**

*Факультет № 6
Кафедра соціології та психології*

ТЕКСТ ЛЕКЦІЇ

з навчальної дисципліни

«Комп'ютерні методи практичної психології»
обов'язкових компонент освітньої програми
першого (бакалаврського) рівня вищої освіти

053 Психологія (практична психологія)

**Тема №9. Аналіз психологічних даних в SPSS:
регресійний аналіз**

Харків 2023

ЗАТВЕРДЖЕНО

Науково-методичною радою
Харківського національного
університету внутрішніх справ
Протокол від 30.08.2023 р. №7

СХВАЛЕНО

Вченою радою факультету №6
Протокол від 25.08.2023 р. №7

ПОГОДЖЕНО

Секцією Науково-методичної
ради ХНУВС з гуманітарних та
соціально-економічних дисциплін
Протокол від 29.08.2023 р. №7

Розглянуто на засіданні кафедри соціології та психології
Протокол від 15.08.2023 р. №8

Розробник:

Професор кафедри соціології та психології факультету №6
д-р соціол. н., професор Нечитайло Ірина Сергіївна

Рецензенти:

1. Керівник психологічної служби Харківського гуманітарного університету «Народна українська академія», доцент кафедри соціології та гуманітарних дисциплін, к. психол. н., Гога Н. П.;
2. Доцент кафедри соціології та психології факультету №6, к. психол. н., доцент Філоненко В. М.

ТЕМА №9. АНАЛІЗ ПСИХОЛОГІЧНИХ ДАНИХ В SPSS: РЕГРЕСІЙНИЙ АНАЛІЗ

План лекції

9.1. Поняття та практичне значення регресії.

9.2. Регресійний аналіз у SPSS.

Рекомендована література

Основна

1. Нечітайло І., Бірюкова М. Математичні методи в соціології : підручник для студентів ВНЗ / Нар. укр. акад., [каф. соціології]. Харків : Вид-во НУА, 2012. 243 с.

2. Татьянчиков А. О. Математичні методи в психології: навчально-методичні рекомендації (в допомогу до самостійної роботи для здобувачів вищої освіти ступеня бакалавра факультету психології, політології та соціології) ; кафедра психології НУ «Одеська юридична академія». Одеса : Фенікс, 2021. 48 с.

Допоміжна

3. Катаєв Є.С. Використання статистичних методів обробки даних у дослідженнях “я-концепції” особистості. Вісник Національного університету оборони України. 2012. №2 (27) /2012. С. 171-176.

4. Салюк М. А Статистична обробка даних експериментального дослідження. Методичний посібник з курсу «Експериментальна психологія» / за ред. Е.Л. Носенко. Дніпропетровськ: Інновація, 2010. 26 с.

5. Старушенко Г. А. Статистична обробка даних в системі публічного управління : навч. посіб. Дніпро : ГРАНІ, 2018. 144 с.

6. Татьянчиков А.О. Методичні рекомендації до виконання лабораторних робіт з курсу «Методи психологічного дослідження: математичні методи в психології». Одеса : Вид-во Університету Ушинського, 2019. 38 с.

ТЕКСТ ЛЕКЦІЇ

9.1. Поняття та практичне значення регресії

Регресія (англ. *regression*, нім. *regression*) – форма зв’язку між випадковими величинами, закон зміни математичного сподівання однієї випадкової величини залежно від значень іншої.

Розрізняють прямолінійну, криволінійну, ортогональну, параболічну та інші регресії, а також лінію та площість регресії.

Розглянемо дві безперервні змінні $X = (x_1, x_2, \dots, x_n)$, $Y = (y_1, y_2, \dots, y_n)$. Якщо ми вважаємо, що Y залежить від X , причому зміни Y викликаються саме змінами X , ми можемо визначити лінію регресії (регресія Y на X), яка найкраще описує прямолінійне співвідношення між цими двома змінними.

Статистичне використання терміну «регресія» виходить із явища, відомого як

«регресія до середнього», винахід кого приписується Френсісу Гальтону (1889 р.). Він показав, що, хоча всі високі батьки мають тенденцію мати високих синів, середній зріст їх синів є меншим, ніж у їхніх високих батьків. Середній зріст синів «регресував» і «рухався назад» до середнього зросту всіх батьків у популяції. Таким чином, у середньому високі батьки мають нижчих (але все-таки високих) синів, а низькі батьки мають синів більш високих (але все-таки досить низьких).

Лінія регресії. Математичне рівняння, яке оцінює лінію простої (парної) лінійної регресії: $Y=a+bx$. X називається незалежною змінною чи предиктором. Y – залежна змінна або змінна-відгук. Це значення, яке ми очікуємо для y (у середньому), якщо знаємо величину X .

Парну лінійну регресію можна розширити, включивши до неї більше однієї незалежної змінної. У цьому випадку вона відома як множинна регресія.

Метод найменших квадратів. Ми виконуємо регресійний аналіз, використовуючи вибірку спостережень, де a та b – вибіркові оцінки істинних (генеральних) параметрів. Найбільш простим методом визначення коефіцієнтів a та b є метод найменших квадратів (МНК). Цей метод базується на розгляді залишків (вертикальна відстань кожної точки від лінії, яку визначають так, щоб сума квадратів залишків була мінімальною).

Припущення лінійної регресії. Кожен залишок може бути позитивним або негативним. Можна використовувати залишки для перевірки наступних припущень, що лежать в основі лінійної регресії:

1. Існує лінійне співвідношення: для будь-яких пар дані повинні апроксимувати пряму лінію. Якщо нанести на двовимірний графік залишки, ми повинні спостерігати випадкове розсіювання точок, а не якусь систематичну картину;

2. Залишки нормально розподілені з нульовим середнім значенням;

3. Залишки мають однакову варіабельність (постійну дисперсію) для всіх передбачуваних величин. Якщо нанести залишки проти передбачуваних величин, уде спостерігатися випадкове розсіювання точок. Якщо графік розсіювання залишків збільшується або зменшується зі збільшенням, то це припущення не виконується;

4. Якщо припущення лінійності, нормальності та/або постійної дисперсії є сумнівними, видається можливим перетворення або розрахунок нової лінії регресії, для якої ці припущення задовольняються (наприклад, використовуючи логарифмічне перетворення або ін.).

Аномальні значення (викиди) та точки впливу можуть змінювати одну або більше оцінок параметрів моделі (зокрема, кутовий коефіцієнт). Викид (спостереження, яке суперечить більшості значень у наборі даних) може бути «впливовим» спостереженням і може добре виявлятися візуально при огляді двовимірної діаграми розсіювання або графіку залишків. І для викидів, і для «впливових» спостережень (крапок) використовують моделі, як із їх включенням, так і без них, звертаючи увагу на зміну оцінки (коефіцієнтів регресії). При проведенні аналізу не варто відкидати викиди або точки впливу автоматично, оскільки звичайне ігнорування може вплинути на отримані результати. Завжди вивчайте причини появи цих викидів та аналізуйте їх.

Застосування лінії регресії для прогнозу. Можна застосовувати регресійну лінію

для прогнозування значення за межею спостережуваного діапазону (ніколи не екстраполуйте поза цими межами). Ми передбачаємо середню величину для спостережень, які мають певне значення шляхом підстановки цього значення рівняння лінії регресії. Отже, якщо прогнозуємо, то використовуємо цю передбачену величину та її стандартну помилку, щоб оцінити довірчий інтервал для середньої величини в популяції. Повторення цієї процедури для різних величин дозволяє побудувати довірчі межі цієї лінії. Це смуга або область, яка містить справжню лінію, наприклад, з 95% вірогідністю. Подібним чином можна розрахувати ширшу область, усередині якої, як очікуємо, лежить найбільше (зазвичай 95%) спостережень.

9.2. Регресійний аналіз у SPSS

Проста лінійна регресія

Щоб викликати регресійний аналіз в SPSS, виберіть у меню Analyze (Аналіз) > Regression (Регресія). Відкриється відповідне підменю (Див. Рис. 9.1.).

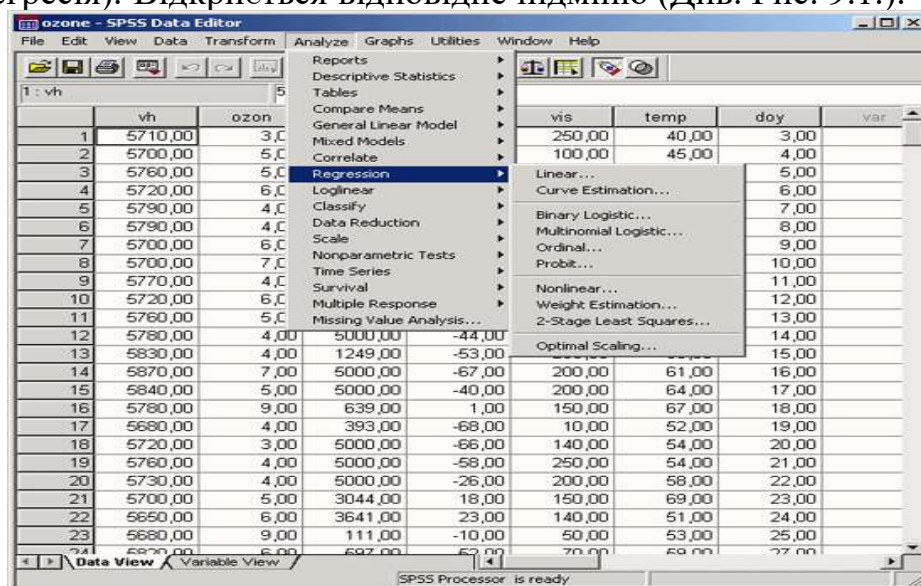


Рисунок 9.1. Допоміжне меню – Regression

При вивченні лінійного регресійного аналізу необхідно провести відмінності між простим аналізом (одна незалежна змінна) і множинним аналізом (кілька незалежних змінних). Ніяких принципових відмінностей між цими видами регресії немає, однак проста лінійна регресія є більш простою і застосовується частіше за всі інші види. Цей вид регресії найкраще підходить для того, щоб продемонструвати основоположні принципи регресійного аналізу.

Розглянемо приклад із залежністю показника холестерину через один місяць після початку лікування від вихідного показника. Тут наявний очевидний зв'язок: обидві змінні розвиваються в одному напрямку і безліч точок (відповідних спостережуваних значень показників) явно концентрується (за деякими винятками) поблизу лінії регресії. У такому разі говорять про лінійний зв'язок. $Y = b \cdot x + a$, де b – регресійні коефіцієнти, a – зміщення по осі ординат (ОУ).

Зсув по осі ординат відповідає точці на осі Y (вертикальній осі), де лінія регресії перетинає цю вісь. Коефіцієнт регресії b через співвідношення $b = \tan(\alpha)$ вказує на кут

нахилу інїї регресії.

При проведенні простої лінійної регресії основним завданням є визначення параметрів b і a . Оптимальним рішенням цього завдання є така пряма, для якої сума квадратів вертикальних відстаней до окремих точок даних є мінімальною (див. Рис. 9.2.)

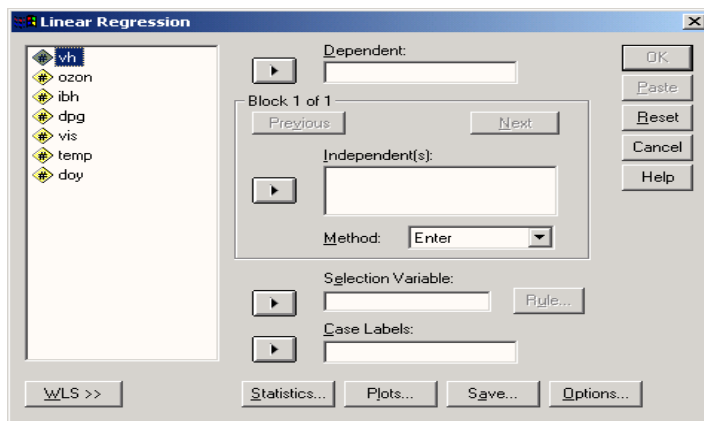


Рисунок 9.2. Діалогове вікно - Лінійна регресія

Висновок основних результатів виглядає наступним чином:

Model Summary (Зводна таблиця по моделі)

Model (Модель)	R	R Square (R-квадрат)	Adjusted R Square (Скоректований R-квадрат)	Std. Error of the Estimate (Стандартна похибка оцінки)
1	,861 ^a	,741	,740	25,26

a. Predictors: (Constant), Cholesterin, Ausgangswert (Змінні, що впливають: (константи), холестерін, вихідна величина)

ANOVA ^b

Model (Модель)		Sum of Squares (Сума Квадратів)	df	Mean Square (Середнє значення квадрата)	F	Sig. (Значу- щість)
1	Regression (Регресія)	314337,948	1	314337,9	492,722	,000 ^a
	Residual (Залишки)	109729,408	172	637,962		
	Total (Сума)	424067,356	173			

a. Predictors: (Constant), Cholesterin, Ausgangswert (Змінні, що впливають: (константи), холестерін, вихідна величина)

b. Dependent Variable: Cholesterin, nach 1 Monat (Залежна змінна холестерін через 1 місяць)

Coefficients (Коефіцієнти) a

Model (Модель)		Unstandardized Coefficients (Нестандартизовані коефіцієнти)		Standardized Coefficients (Стандартизовані коефіцієнти)	t	Sig. (Значу- щість)
		B	Std. Error (Станд. похибка)	β (Beta)		
1	(Constant) (Константа)	34,546	9,416		3,669	,000
	Cholesterin, Ausgangswert	,863	,039	,861	22,197	,000

a. Dependent Variable (Залежна змінна)

Розглянемо спочатку нижню частину результатів розрахунків. Тут виводяться коефіцієнт регресії b і зсув по осі ординат a під іменем «константа». Тобто, рівняння регресії виглядає наступним чином: $chol1 = 0,863 \cdot chol0 + 34,546$.

Якщо значення вихідного показника холестерину становить, наприклад, 280, то через один місяць можна очікувати показник рівний 276.

Частки розрахованих коефіцієнтів і їх стандартна помилка дають контрольну величину T . Відповідний рівень значимості відноситься до існування ненульових коефіцієнтів регресії. Значення коефіцієнта β буде розглянуто при вивченні багатовимірного аналізу.

Середня частина розрахунків відображає два джерела дисперсії: дисперсію, яка описується рівнянням регресії (сума квадратів, обумовлена регресією) і дисперсію, яка не враховується при запису рівняння (залишкова сума квадратів). Частка від суми квадратів, обумовлених регресією і залишкової суми квадратів називається «коефіцієнтом детермінації». В таблиці результатів ця частка виводиться під ім'ям « R -квадрат». У нашому прикладі міра визначеності дорівнює: $314337,948 / 424067,356 = 0,741$. Ця величина характеризує якість регресійної прямої, тобто ступінь відповідності між регресійною моделлю і вихідними даними.

Міра визначеності завжди лежить в діапазоні від 0 до 1. Існування ненульових коефіцієнтів регресії перевіряється за допомогою обчислення контрольної величини F , до якої належить відповідний рівень значущості.

У простому лінійному регресійному аналізі квадратний корінь з коефіцієнта детермінації, що позначається R , дорівнює кореляційному коефіцієнту Пірсона. При множинному аналізі ця величина менш наочна, ніж сам коефіцієнт детермінації. Величина «Зміщений R -квадрат» завжди є меншим, ніж незміщений. При наявності великої кількості незалежних змінних, міра визначеності коригується в бік зменшення. Принципове питання про те, чи може взагалі наявний зв'язок між змінними розглядатися як лінійна регресія, простіше і наочніше всього вирішувати, дивлячись на відповідну діаграму розсіювання. Крім того, на користь гіпотези про лінійний зв'язок свідчить також високий рівень дисперсії, описуваної рівнянням регресії.

І, нарешті, стандартизовані прогнозовані значення і стандартизовані залишки

можна надати у вигляді графіку. Ви отримаєте цей графік, якщо через кнопку Plots (Графіки) зайдете у відповідне діалогове вікно і поставите в ньому параметри *ZRESID і *ZPRED в якості змінних, які відображаються по осях Y і X відповідно. У випадку лінійної регресії залишки розподіляються випадково по обидві сторони від нульової горизонтальної лінії.

Збереження нових змінних. Численні допоміжні значення, що розраховуються в ході побудови рівняння регресії, можна зберегти як змінні і використовувати в подальших розрахунках.

Для цього в діалоговому вікні Linear Regression (Лінійна регресія) клацніть на кнопці Save (Зберегти). Відкриється діалогове вікно Linear Regression: Save (Лінійна регресія: Збереження) як зображено на рисунку 9.3.

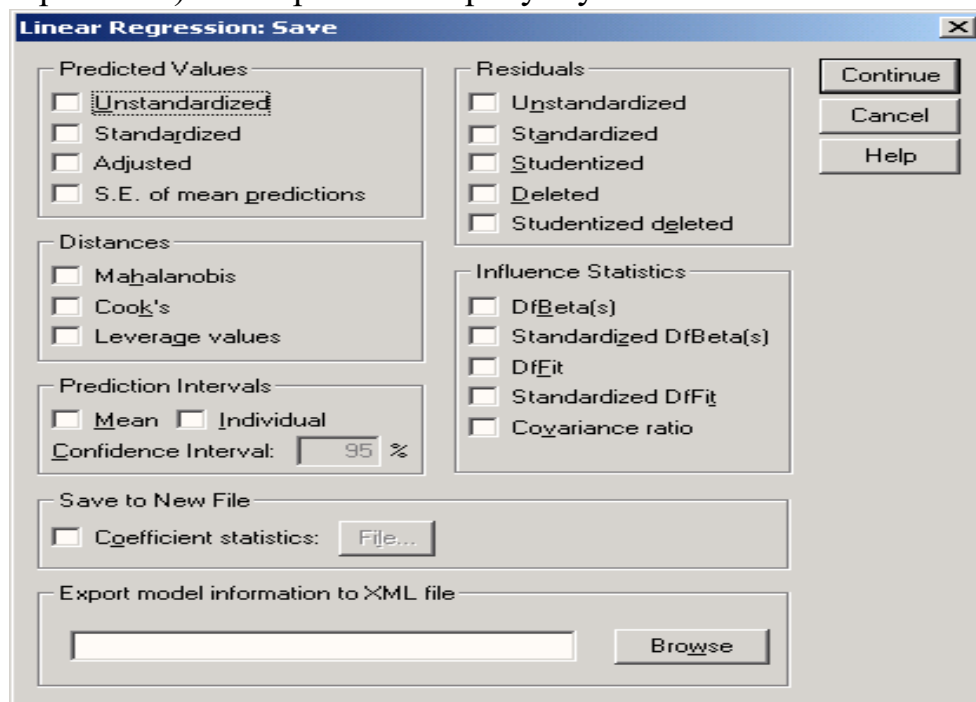


Рисунок 9.3: Діалогове вікно – Лінійна регресія: Збереження

Цікавими тут видаються опції Standardized (Стандартизовані значення) і Unstandardized (Нестандартизовані значення), які знаходяться під рубрикою Predicted values (Прогнозовані величини). При виборі опції Нестандартизовані значення будуть розраховуватися значення, що відповідають рівнянню регресії. При виборі опції «Стандартизовані значення» прогнозована величина нормалізується. SPSS автоматично присвоює нове ім'я кожної новоутвореної змінної, незалежно від того, чи розраховуються прогнозовані значення, відстані, прогнозовані інтервали, залишки або які-небудь інші важливі статистичні характеристики. Нестандартизованим значень SPSS присвоює імена pre_1 (predicted value), pre_2 тощо, а стандартизованим zpr_1.

Покроковий алгоритм

1. Клацніть у діалоговому вікні Linear Regression: Save (Лінійна регресія: Збереження) в поле Predicted values (Прогнозовані значення) на опції Unstandardized (Нестандартизовані значення);
2. Підтвердіть натисканням Continue (Далі) і висновок OK;

3. В редакторі даних буде утворена нова змінна під ім'ям `rge_1` і додано її в кінець списку змінних у файлі.

4. Для пояснення значень змінної `rge_1`, візьмемо випадок 5. Для випадку 5 мінлива `rge_1` містить нестандартизоване прогнозоване значення 263,11289. Це прогнозоване значення злегка відрізняється в сторону збільшення від реального показника вмісту холестерину, взятого через один місяць (`chol1`) і рівного 260. Нестандартизоване прогнозоване значення для змінної `chol1`, так само як і інші значення змінної `rge_1`, було обчислено виходячи з відповідного рівняння регресії.

Якщо ми в рівняння регресії: $chol1 = 0,863 * chol0 + 34,546$ підставимо вихідне значення для `chol0` (265), то отримаємо: $chol1 = 0,863 * 265 + 34,546 = 263,241$.

Невелике відхилення від значення, що зберігається у змінній `rge_1` пояснюється тим, що SPSS застосовую у розрахунках більш точні значення, ніж ті, які виводяться у вікні перегляду результатів.

Для цього додайте в кінець файлу `hyper.sav`, ще два випадки, використовуючи фіктивні значення для змінної `chol0`. Нехай приміром будуть значення 282 і 314.

Ми виходимо з того, що нам не відомі значення показника холестерину через місяць після початку лікування, і ми хочемо спрогнозувати значення змінної `chol1`.

Залиште попередні установки без змін і проведіть новий розрахунок рівняння регресії.

В кінці списку змінних додається мінлива `rge_2`. Для нового доданого випадку (№175) для змінної `chol1` буде передбачено значення 277,77567, а для випадку №176 – значення 305,37620.

Побудова регресійної прямої. Щоб на діаграмі розсіювання зобразити регресійну пряму, потрібно виконати наступні дії:

1. Виберіть у меню наступні опції Graphs (Графіки) / Scatter plots / Діаграми розсіювання. Відкриється діалогове вікно Scatter plots (Діаграма розсіювання) (Див. рис. 9.4).

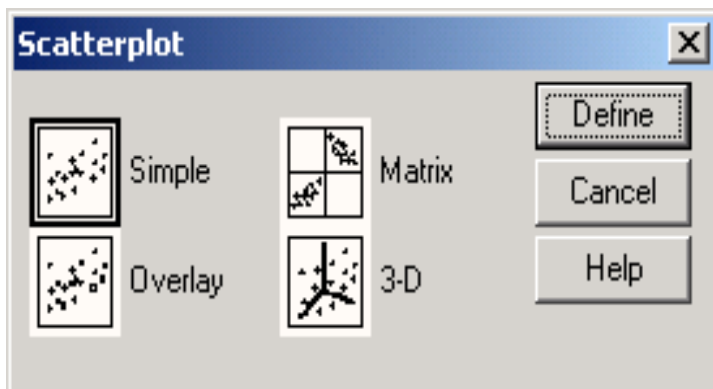


Рисунок 9.4: Діалогове вікно Scatter plots (Діаграма розсіювання)

2. У діалоговому вікні Scatter plots (Діаграма розсіювання) залиште попередню установку Simple (Проста) і клацніть на кнопці Define (Визначити). Відкриється діалогове вікно Simple Scatter plot (Проста діаграма розсіювання) (Див. рис. 9.5).

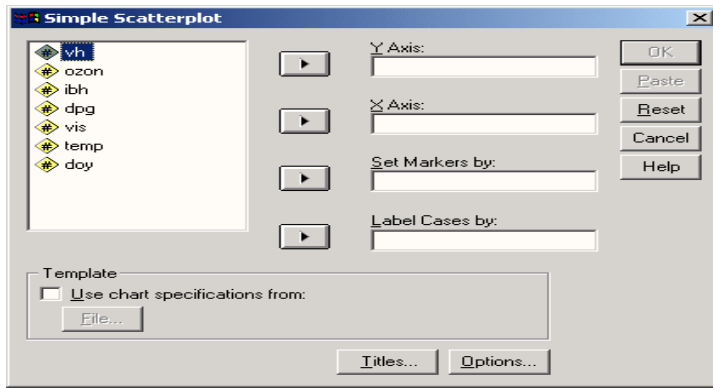


Рисунок 9.5. Діалогове вікно *Simple Scatter plot* (Проста діаграма розсіювання)

У діалоговому вікні *Scatter plots* (Діаграма розсіювання) залиште попередня установку *Simple* (Проста) и клацніть на кнопці *Define* (Визначити). Відкриється діалогове вікно *Simple Scatter plot* (Див. Рис. 9.6)

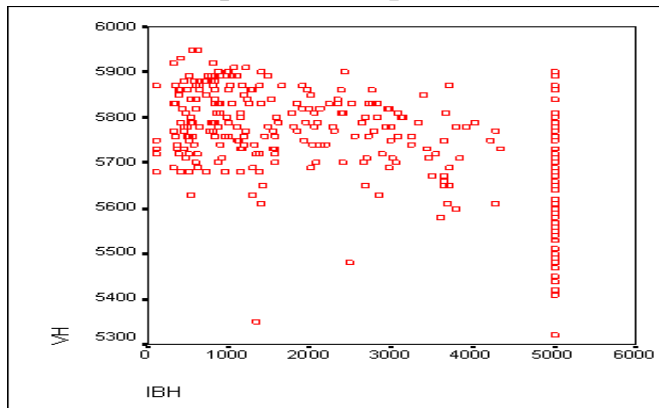


Рисунок 9.5. Проста Діаграма розсіювання

Клацніть двічі на цьому графіку, щоб перенести його в редактор діаграм.

Виберіть у редакторі діаграм меню Chart (Діаграма) / Options (Параметри). Відкриється діалогове вікно Scatterplot Options (Опції для діаграми розсіювання) (Див. Рис. 9.7).

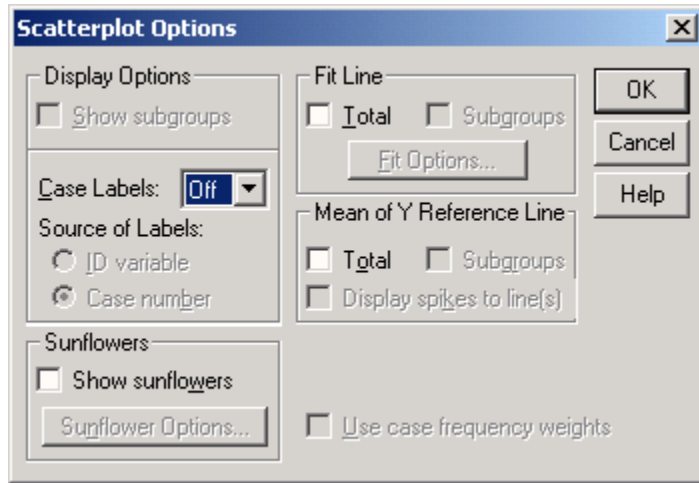


Рисунок 9.7. Діалогове вікно Scatterplot Options (Опції для діаграми розсіювання)

У рубриці Fit Line (Наближена крива) потрібно поставити прапорець навпроти опції Total (Цілком для всього файлу даних) і клацнути на кнопці Fit Options (Опції для наближення). Відкриється діалогове вікно Scatterplot Options: Fit Line (Опції для діаграми розсіювання: наближена крива) (див. Рис. 9.8).

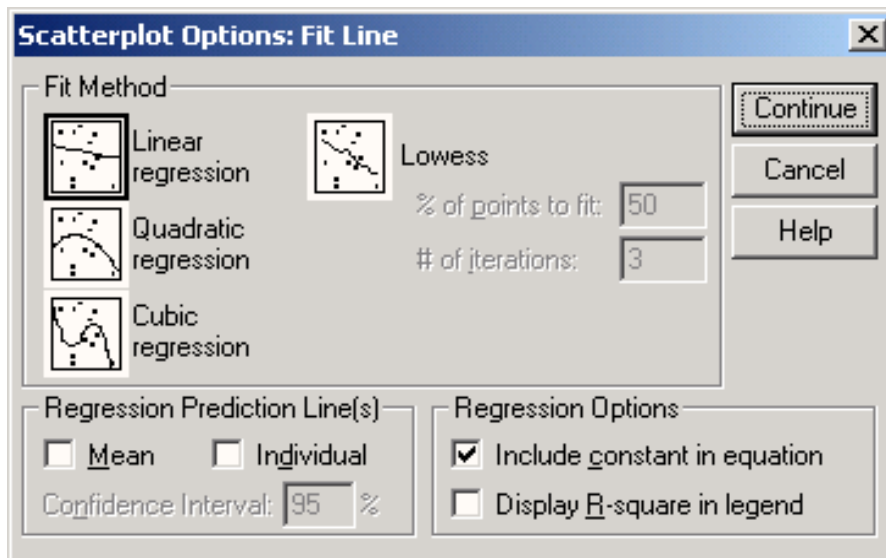


Рисунок 9.8: Діалогове вікно Scatterplot Options: Fit Line (Опції для діаграми розсіювання)

Потрібно підтвердити попередню установку Linear Regression (Лінійна регресія) клацанням Continue (Далі) і потім на OK.

Закрийте редактор діаграм і клацніть один раз десь поза графіком. Тепер в діаграмі розсіювання відображається регресійна пряма (Див. Рис. 9.9).

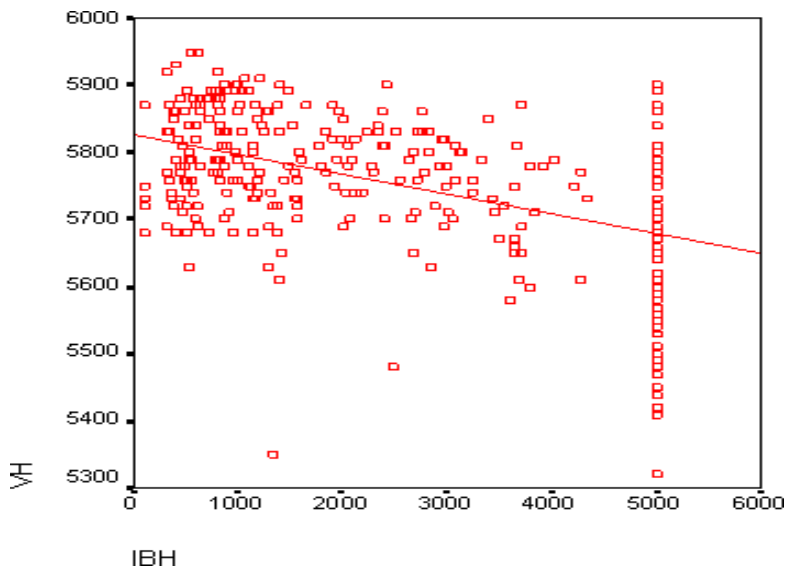


Рисунок 9.9. Відображення у діаграмі розсіювання регресійної прямої

Вибір вісей. Для діаграм розсіювання часто виявляється необхідною додаткова коригування осей. Продемонструємо таку корекцію за допомогою одного прикладу. У файлі `raucher.sav` знаходяться десять фіктивних наборів даних. Мінлива `konsum` вказує на кількість сигарет, які викурює одна людина в день, а змінна `puls` на кількість часу, необхідне кожному випробуваному для відновлення пульсу до нормальної частоти після двадцяти присідань. Як було показано раніше, побудуйте діаграму розсіювання з впровадженої регресійної прямої.

У діалоговому вікні Simple Scatterplot (Проста діаграма розсіювання) перенесіть змінну `puls` в поле осі Y, а змінну `konsum` – в поле осі X. Після відповідної обробки даних у вікні перегляду з'явиться діаграма розсіювання, зображена на рисунку 9.10.

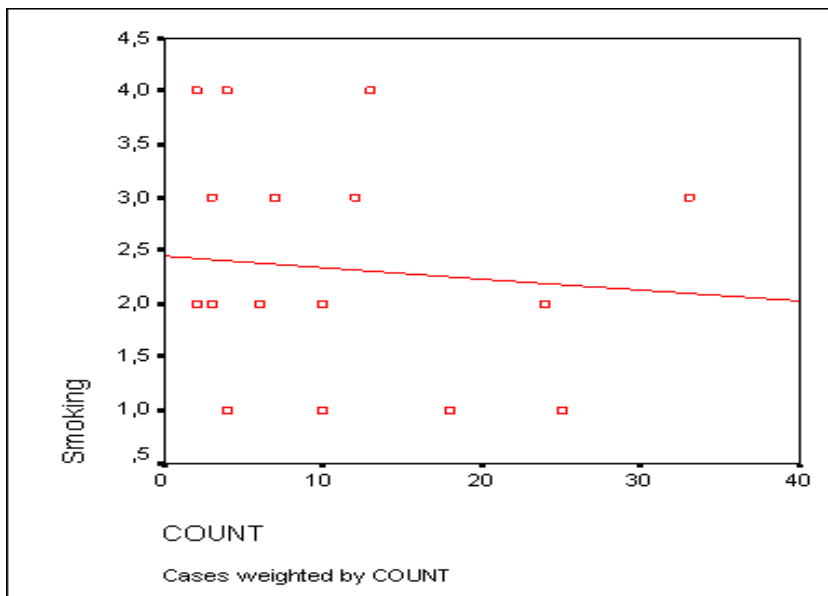


Рисунок 9.10. Діаграма розсіювання

Оскільки ніхто не викурює мінус 10 сигарет в день, точка початку відліку осі X є не зовсім коректною. Тому цю вісь необхідно відкоригувати.

Потрібно двічі клацнути на графіку і в меню редактора діаграм обрати опції Chart (Діаграма) / Axis (Осі). Відкриється діалогове вікно Axis Selection (Вибір осі) (Див. Рис. 9.11).

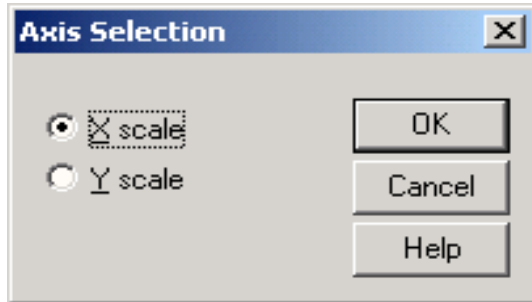


Рисунок 9.11: Діалогове вікно Axis Selection (Вибір осі)

Підтвердіть попередній вибір осі X натисканням кнопки ОК. Відкриється діалогове вікно X-Scale Axis (Вісь X) (Див. рис. 9.12).

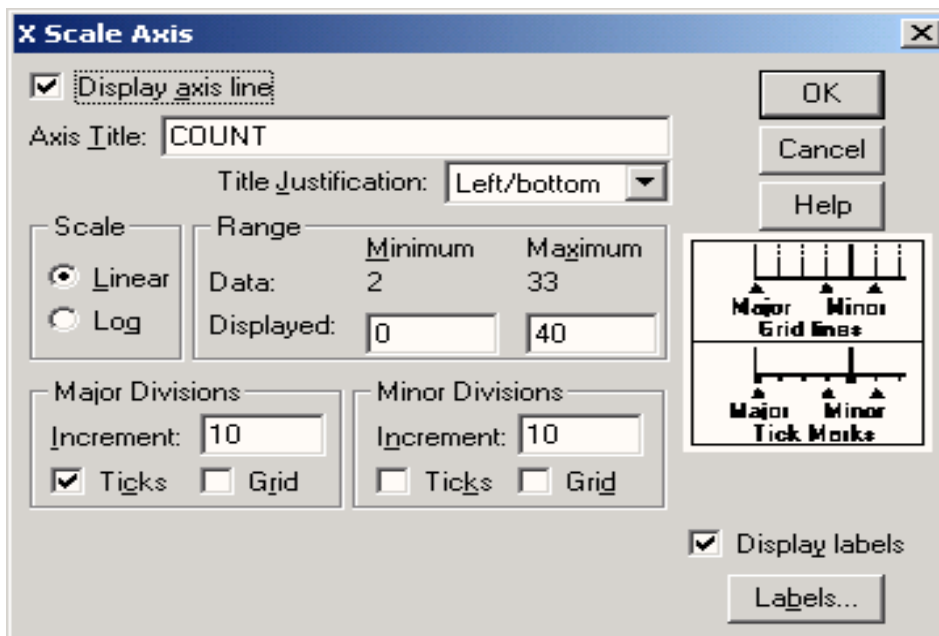


Рисунок 9.12. Попередній вибір осі X. Діалогове вікно X-Scale Axis

Покроковий алгоритм

1. У редагованому полі Displayed (відображається) в рубриці Range (Діапазон) змінюємо мінімальне значення на 0;
2. Підтверджуємо натисканням на ОК;
3. Обираємо знову в меню редактора діаграм опції Chart (Діаграма) * Axis (Осі);
4. Для початку процесу в діалоговому вікні Axis Selection (Вибір осі) обираємо опцію Y Scale (Ось Y);
5. Відкриється діалогове вікно Y-Scale Axis (Вісь Y);
6. В рубриці Range (Діапазон) в редагованому полі Displayed (відображається) змініть мінімальне значення на 0;

7. Підтвердіть натисканням на ОК.

У вікні перегляду можна тепер побачити відкориговану діаграму розсіювання (див. Рис. 9.13).

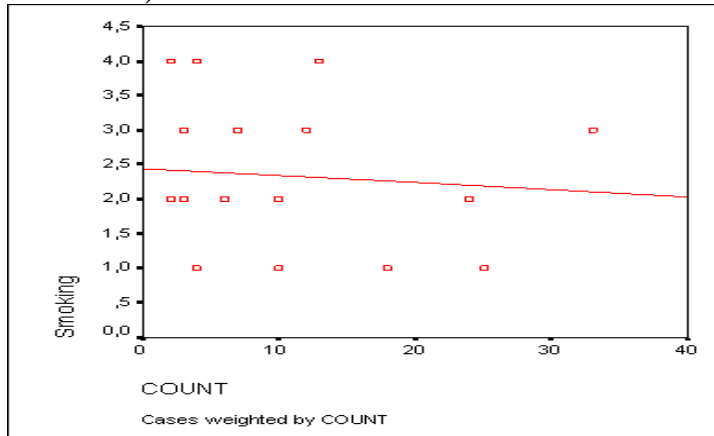


Рисунок 9.13. Відкоригована діаграма розсіювання

На відкоригованій діаграмі розсіювання тепер стало простіше розпізнати початкову точку на осі Y, яка утворюється при перетині з регресійною прямою. Значення цієї точки приблизно дорівнює 2,9.

Порівняємо це значення з рівнянням регресії для змінних puls (залежна змінна) і konsum (незалежна змінна). В результаті розрахунку рівняння регресії у вікні відображення результатів з'являться наступні значення:

Model (Модель)		Unstandardized Coefficients (Нестандартизовані коефіцієнти)		Standardized Coefficients (Стандартизовані коефіцієнти)	t	Sig. (Значимість)
		B	Std. Error (Станд. ошибка)	β (Beta)		
1	(Constant) (Константа)	2,871	,639		4,492	,002
	tgl. Zigarettenkonsum	,145	,038	,804	3,829	,005

a. Dependent Variable: Pulsfrequenz unter 80 (Залежна змінна: частота пульсу нижче 80)

Що дає наступне рівняння регресії: $puls = 0,145 * konsum + 2,871$

Константа у вищенаведеному рівнянні регресії (2,871) відповідає точці на осі Y, яка утворюється в точці перетину з регресійній прямий.

Множинна регресія є розширенням простої лінійної регресії. За допомогою простої регресії оцінювалася ступінь впливу однієї незалежної змінної (предиктора) на залежну змінну (критерій). На відміну від простої регресії множинна регресія досліджує вплив двох і більше предикторів на критерій.

Аналіз регресії можна звести до геометричної інтерпретації. Коли обчислена проста кореляція між двома змінними, можна побудувати лінію регресії (лінію «найкращої відповідності»). Ця лінія будується на основі рівняння регресії, а її кутовий коефіцієнт дорівнює коефіцієнту при незалежній змінній, а зрушення вгору по вертикальній осі визначається константою.

Далі розглянемо множинну регресію як послідовне ускладнення простого регресійного рівняння. Робочий файл містить дані психологічного дослідження схильності людей надавати допомогу своїм знайомим. Хоча дані є вигаданими, результати їх обробки близькі до результатів одного з реальних досліджень.

Рівняння множинної регресії. Змінна «допомога» являє час (у секундах), витрачений людиною на надання допомоги своєму партнеру, і її значення мають нормальний розподіл (середнє дорівнює 30, стандартне відхилення – 10). Змінна «симпатія» відображає оцінку симпатії до партнера в балах від 1

до 20. На прикладі цих двох змінних ми продемонструємо просту регресію. В якості залежної виступить змінна «допомога», а в якості незалежної – змінна «симпатія» (передбачається, що симпатія і співчуття змушують людину надавати допомогу, а не навпаки). Як показав аналіз, коефіцієнт кореляції між змінними «допомога» і «симпатія» становить 0,416 при значущості $p=0,004$, що говорить про значний зв'язок між цими змінними. Константа і коефіцієнт регресії склали, відповідно, 14,739 і 1,547.

Якщо для деякого випробуваного значення змінної «симпатія» складе 16, то на основі регресійного рівняння ми можемо прогнозувати, що змінна допомогу прийме таке значення: $14,739 + 1,547 (16) = 39,5$.

Значення 16 є порівняно великим у відповідності з прийнятою шкалою; те ж саме стосується і прогнозованого значення «допомога», що перевищує середнє значення цієї змінної майже на одне стандартне відхилення.

Аналогічні розрахунки можна виконувати і при множинному регресійному аналізі: Відмінність полягає лише в тому, що при множинному аналізі рівняння регресії включає більш ніж одну залежну змінну.

Крім змінної «симпатія» зі змінною «допомога» корелюють й інші змінні робочого файлу. Зокрема, це змінні «агресія» (агресивність людини по відношенню до партнера, виміряна в балах від 1 до 20) та «користь» (самооцінка власної корисності в балах від 1 до 20).

Візьмемо об'єкт з номером 7 і розрахуємо для нього прогнозоване значення змінної «допомога»: $\text{допомога} = -5,3147 + 1,0328 (2) + 1,1676 (10) + 1,2569 (9) - 19,74$.

Таким чином, людина, яка має низький показник симпатії і середні показники агресивності і самооцінки корисності, повинен, згідно з прогнозом, надавати незначну допомогу. Фактичне значення змінної допомогу для об'єкта 7 склало 21, що свідчить про високу точність нашого прогнозу.

Коефіцієнти регресії. Позитивний коефіцієнт при незалежній змінній говорить про те, що із зростанням останньої значення залежної змінної також зростає. Вірно і протилежне твердження: при від'ємному коефіцієнті із зростанням значення незалежної змінної значення залежної змінної убуває.

Тим не менш, в більшості досліджень співвідношення коефіцієнтів не дозволяє робити висновок про вплив того або іншого фактору на залежну змінну, оскільки незалежні змінні, як правило, вимірюються в різних шкалах і мають різний масштаб. Щоб вирішити цю проблему, був введений коефіцієнт регресії r , приймає значення від -1 до 1. Він відображає приватну кореляцію незалежної і залежної змінних.

Під приватною кореляцією розуміється вплив, який чиниться на залежну змінну з боку однієї незалежної змінної при фіксованих значеннях інших незалежних змінних (з урахуванням впливу останніх). Чим більше ця незалежна змінна корелює з іншими незалежними змінними, тим менше абсолютна величина її коефіцієнта r .

Так, для нашого прикладу з трьома незалежними змінними, коефіцієнт r змінної «агресія» – це її кореляція з залежної змінної після виключення впливу змінних симпатія і користь. Для простої регресії з однією незалежною змінною

коефіцієнт r дорівнює коефіцієнту парної кореляції залежної і незалежної змінних. Таким чином, коефіцієнт r є універсальною мірою впливу незалежної змінної. Його часто називають стандартним коефіцієнтом регресії. І саме стандартні коефіцієнти регресії дозволяють співвідносити незалежні змінні за їх значенням для оцінки залежної змінної.

Звернемося до записаного вище рівняння регресії. Як ми можемо бачити, всі три незалежні змінні входять до нього з позитивними коефіцієнтами. Такий результат є цілком логічним для змінних «симпатія» і «користь», однак викликає подив те, що зростання агресивності суб'єкта тягне за собою збільшення його прагнення надавати допомогу. Причина появи такого дивного результату – гарний привід для дискусії. Кожен раз, коли ви будете потрапляти в подібні «непередбачені» ситуації, перевіряйте правильність кодування і введення даних. Пам'ятайте, що програма здатна лише генерувати результати аналізу, але щодо їх інтерпретації вона практично безпорадна.

Якщо ми розглянемо показник холестерину через один місяць (змінна chol1) як залежну змінну (Y), а вихідну величину як незалежну змінну (X), то тоді для проведення регресійного аналізу потрібно буде визначити параметри співвідношення: $\text{chol1} = b * \text{chol0} + a$.

Після визначення цих параметрів, знаючи вихідний показник холестерину, можна спрогнозувати показник, який буде через один місяць.

Розрахунок рівняння регресії

Покроковий алгоритм

1. Відкрийте робочий файл;
2. Виберіть у меню Analyze (Аналіз) > Regression (Регресія) > Linear (Лінійна);
3. З'явиться діалогове вікно Linear Regression (Лінійна регресія);
4. Перенесіть змінну chol1 в поле для залежних змінних і надайте змінній chol0 статус незалежної змінної;
5. Нічого не змінюючи, почніть розрахунок натисненням ОК.