Розділ 3

ІНФОРМАЦІЙНО-АНАЛІТИЧНІ МЕТОДИ І МОДЕЛІ ПІДТРИМКИ ПРИЙНЯТТЯ МАРКЕТИНГОВИХ РІШЕНЬ

Сучасний розвиток обчислювальної техніки та засобів передавання інформації дає змогу автоматизувати збирання, пошук і нагромадження як внутрішньої, так і зовнішньої інформації. Ефективність маркетингових рішень, прийнятих на основі нагромаджених даних, забезпечується трьома видами засобів підтримки евристичної діяльності: засобами, що забезпечують інтерфейс кінцевого користувача в задачах пошуку деталізованої інформації; засобами, що забезпечують комплексний погляд на зібрану інформацію, її узагальнення та агрегацію; засобами інтелектуального аналізу даних.

Призначенням останніх є підтримка пошуку функціональних і логічних закономірностей у нагромадженій інформації, осмислення та прогнозування розвитку ринкових процесів, визначення основних факторів, що впливають на цей розвиток.

У попередніх розділах другої частини цього посібника розглядалися перші два типи засобів підтримки прийняття маркетингових рішень. У цьому розділі висвітлюватимуться питання реалізації інструментальних маркетингових моделей — одного з найважливіших засобів подання та оброблення знань стосовно проблемних маркетингових ситуацій.

3.1. ТИПИ ІНСТРУМЕНТАЛЬНИХ МОДЕЛЕЙ

Широке розповсюдження електронних таблиць (у першу чергу Excel) пояснюється тим, що вони надають користувачам зручне інтерактивне середовище для побудови кількісних моделей, які дають змогу автоматизувати проведення складних розрахунків і відобразити семантику певної предметної області у наочній формі. До основних властивостей середовища Excel, що роблять його високоефективним інструментом моделювання, слід віднести такі:

— можливість уводити в клітини робочого аркуша не тільки дані, а й формули для оброблення цих даних;

— перерахування формул як реакції на зміну значень у клітинах, що містять дані чи формули;

— можливість відображати поняття відповідної предметної області та зв'язки між цими поняттями у зручній для людини формі;

— наявність значної кількості засобів та інструментів для проведення аналізу даних;

— наявність системи керування моделями, що забезпечує гнучкий доступ до баз моделей;

— убудованість потужної об'єктно-орієнтованої мови програмування, поєднання якої з можливостями електронної таблиці забезпечує масштабоване середовище розроблення моделей.

Цей комплекс властивостей робить Excel, по-перше, високоефективною системою використання моделей, по-друге, високоефективною платформою створення моделей будь-якої складності і, по-третє, високоефективним засобом поповнення та модифікації існуючої бази моделей силами кінцевих користувачів.

Виходячи з методів, використовуваних для аналізу даних, і особливостей їх реалізації в Excel, інструментальні моделі можна класифікувати (з певною часткою умовності) таким чином: моделі прямих розрахунків; компаративні (порівняльні) моделі; оптимізаційні моделі; статистичні моделі; моделі прогнозування; імітаційні моделі.

Корисність наведеної класифікації полягає у тому, що вона допомагає визначити характерні особливості наявних засобів моделювання, а також способи та сферу їх застосування. Але при цьому слід розуміти, що границі між цими класами моделей не є абсолютно чіткими (наприклад, прогнозування може виконуватись із застосуванням статистичних методів аналізу даних) і що побудова моделі реальної маркетингової ситуації може потребувати комбінованого використання різних методів та інструментів аналізу.

Моделі прямих розрахунків характеризуються, як правило, трудомістким (якщо не застосовуються засоби автоматизації) обробленням даних з використанням відносно простих алгоритмів. У цих алгоритмах ураховуються основні функціональні залежності між економічними змінними, а самі алгоритми подаються як сукупність чітко визначених понять і дій з ними, заданих у вигляді формул. По суті, ці алгоритми являють собою формалізовані знання з відповідної предметної області і тому є складовою всіх інших класів моделей.

Наведені вище властивості середовища Excel роблять його майже ідеальним інструментом для реалізації моделей, основним методом подання яких є формульне відображення зв'язків між

змінними. Серед багатьох засобів Excel для роботи з формулами слід виокремити такі:

— виправлення найзагальніших помилок під час введення формул. Природно, користувач може погодитись або відкинути запропоноване коригування;

— можливість посилання у формулах на будь-які клітини робочих аркушів будь-яких робочих книг, використовуючи або їх адреси (абсолютні чи відносні), або їх імена (що надаються користувачем), або заголовки рядків і колонок таблиць, створених користувачем. Останні дві можливості зменшують кількість помилок і спрощують сприйняття формул. Для того щоб можна було використовувати заголовки таблиць як засіб посилання на клітини, необхідно командою *Сервіс*/*Параметри* на вкладці *Обчислення* включити режим *Допускати назви діапазонів*;

— ідентифікація діапазонів клітин, на які є посилання у формулі, кольоровими рамками під час редагування клітини з формулою, натисканням клавіші *F2*. За встановлення командою *Сервіс*/*Параметри* на вкладці *Правка* режиму *Виправлення прямо в комірці*, крім натискання клавіші *F2*, можна також використовувати подвійне клацання лівою кнопкою миші по відповідній клітині. Якщо цей режим вимкнуто, то подвійне клацання мишею приведе не до редагування вмісту клітини, а до виділення діапазонів клітин, на які є посилання у формулі;

— відслідковування й вилучення непотрібних циклічних посилань. Коли вводиться формула, яка змушує посилатися прямо або опосередковано на саму себе, то з'являється вікно попередження та автоматично відкривається панель інструментів *Циклічні посилання*. На цій панелі знаходиться список *Знайти циклічне посилання*, а також інструменти *Залежні комірки* та *Комірки, які впливають*, що дає змогу локалізувати клітини, які входять до циклічного посилання;

— наявність значної кількості вбудованих (стандартних) функцій різних категорій (математичні, фінансові, статистичні, логічні та ін.), з допомогою яких можна виконувати досить складні операції. Їх бажано використовувати скрізь, де це можливо, а власні формули створювати лише у тому разі, коли немає функції, яка дає змогу реалізувати потрібну операцію.

Компаративні (порівняльні) моделі призначені для виявлення відмінностей та розбіжностей. Мета застосування таких моделей — ідентифікувати ситуації, що мали або можуть мати місце, і на основі порівняння цих ситуацій або вибрати одну з можливих альтернатив, або вибрати найкращі значення керованих парамет-

рів, або виявити причинно-наслідкові взаємозв'язки, які дозволять обґрунтувати рішення. Але підхід до розв'язання цих задач з допомогою компаративних моделей однаковий — сприяти виявленню деяких специфічних відмінностей, які суттєво впливають або можуть вплинути на розвиток подій.

Порівняння альтернатив — один з основних методів подолання невизначеності (відносно можливого результату своєї діяльності), а компаративні моделі — важливий інструмент цього методу, який дає змогу спеціалістам усвідомити суть розбіжностей, з'ясувати їх причини і на основі цієї інформації, а також на основі власного досвіду та інтуїції приймати ефективні рішення. Основними підвидами порівняльних моделей є табличні (матричні) та графічні моделі, широко застосовувані у процесі управління маркетинговою діяльністю. Такі моделі відносно легко і просто реалізуються в Excel як завдяки його «табличній» природі, так і завдяки наявності досконалого інструмента побудови діаграм, який дає змогу швидко обробляти табличні дані на високому професійному рівні і надавати користувачеві найскладнішу інформацію у простому й наочному вигляді.

Оптимізаційні моделі дають можливість знайти найкраще рішення за певним критерієм. При цьому після формулювання критерію рішення розшукується, як правило, без оцінювання альтернатив людиною, тобто автоматично. Необхідність використання таких моделей пояснюється тим, що за великої кількості альтернатив метод прямого оцінювання множини можливих рішень стає непридатним. Саме у такому разі слід застосовувати оптимізаційні моделі. Але при цьому слід розуміти, що є значна кількість факторів, вплив яких неможливо визначити кількісно або з необхідною точністю. Тому отримане з допомогою оптимізаційних моделей рішення не слід сприймати як абсолютно точне, його потрібно перевіряти та у разі потреби коригувати, спираючись на досвід та інтуїцію спеціалістів.

В Excel є потужний і простий у застосуванні інструмент — надбудова *Пошук рішення*, що дає змогу розв'язувати широкий клас оптимізаційних задач: лінійного, нелінійного та цілочисельного програмування.

Варто наголосити, що спочатку необхідно забезпечити доступ до цього інструмента. Якщо в меню *Сервіс* відсутня команда *Пошук рішення,* то треба вибрати команду *Сервіс*/*Надбудова*, прокрутити список надбудов і встановити прапорець *Пошук рішення.* Якщо ця надбудова не була інстальована разом з MS Office, то її треба інсталювати окремо.

Моделі прогнозування дають змогу оцінити вплив рішень, які приймаються, на майбутнє, що в умовах ринкової економіки є нагальною потребою. Прогнози необхідні, оскільки, по-перше, майбутнє значною мірою є невизначеним і, по-друге, тому що повний ефект багатьох рішень розподілено у часі і відразу він може не відчуватися. Тому досить точне передбачення майбутнього підвищує ефективність процесу прийняття рішень, а успіх у бізнесі значною мірою залежить від уміння передбачити майбутнє та підготуватися до нього. Надійні прогнози роблять можливим прийняття вчасних рішень, що спираються на обґрунтовані плани.

Оцінювання майбутнього у формі прогнозування обсягів продажу є необхідною частиною рішень щодо фінансового планування підприємства, інвестицій в основні виробничі засоби, послідовності придбання матеріалів, установлення рівнів виробництва і запасів, потреби у персоналі та витрат на рекламу.

Моделі прогнозування поділяють на некаузальні та каузальні. *Некаузальні моделі* пропонують метод прогнозу за минулими значеннями без пояснення механізму генерації прогнозованого значення. *Каузальні моделі* використовують взаємозв'язки між змінними і намагаються пояснити їх поведінку.

Статистичні моделі базуються на використанні математичних моделей і методів, призначених для прийняття рішень під час пошуку, аналізу та оцінювання можливих варіантів рішень у ситуаціях, що характеризуються невизначеністю, пов'язаною як із статистичним характером аналізованих явищ, так і з неповнотою знань щодо значень, які можуть мати ті чи інші фактори.

Excel має значні можливості для проведення статистичного оцінювання даних. Деякі з них є вбудованими, інші доступні лише після встановлення надбудови *Пакет аналізу*. З допомогою останнього можна розв'язувати складні статистичні задачі. До складу пакета аналізу входять засоби проведення дисперсійного, кореляційного, коваріаційного, регресійного аналізу та низка інших засобів для розв'язання статистичних та інженерних задач.

Для аналізу даних з допомогою інструментів пакета аналізу достатньо вказати вхідні дані і вибрати певні параметри. Аналіз проводиться з допомогою відповідної статистичної макрофункції, а результат розміщується у вказаному вихідному діапазоні клітин робочого аркуша. Є в пакеті аналізу і засоби, що дають можливість показати результати аналізу графічно.

Оскільки пакет аналізу — це надбудова, то спочатку необхідно забезпечити доступ до нього: вибрати команду *Сервіс*/*Надбудова*,

прокрутити список надбудов і встановити прапорець *Пакет аналізу.* Після цього його можна використовувати так:

1. У меню *Сервіс* вибирається команда *Аналіз даних.*

2. У списку *Інструменти аналізу* вибирається потрібний рядок.

3. Уводиться вхідний та вихідний діапазони і встановлюються потрібні параметри.

Для успішного застосування процедур аналізу (правильного вибору інструмента та параметрів) необхідні певні знання у сфері статистичних та інженерних розрахунків, для яких ці інструменти було розроблено.

У маркетинговій діяльності часто виникають ситуації, за яких доводиться приймати рішення, виходячи з недостатньо визначених умов та оперуючи не завжди надійною інформацією. У таких випадках виникає потреба в оцінюванні ймовірності досягнення тих чи інших результатів. Основними методами, які дають змогу приймати маркетингові рішення з урахуванням фактора невизначеності і які підтримує Excel, є стандартне відхилення, довірчі інтервали та регресійний аналіз.

Імітаційні моделі є засобом вивчення поведінки об'єктів чи систем в умовах, що визначаються дослідником, шляхом нагромадження та аналізу даних, отримуваних у процесі проведення експериментів із моделлю. Метою імітаційного моделювання може бути вивчення впливу на об'єкт змін у його структурі та (або) у зовнішніх умовах його функціонування. Дані, нагромаджувані в процесі моделювання, використовуються або для проведення ретроспективного аналізу, або для визначення спектра допустимих сценаріїв майбутнього розвитку.

Для генерації випадкових чисел, необхідних для проведення імітаційного моделювання, можна використовувати декілька засобів Excel: функції *СЛЧИС(), СЛУЧМЕЖДУ()* та інструмент пакета аналізу *Генерація випадкових чисел.*

Функція *СЛЧИС()* повертає рівномірно розподілене випадкове число між 0 та 1. Нове випадкове число повертається цією функцією кожного разу, коли перераховується робочий аркуш. Якщо потрібно, щоб отримане випадкове число не змінювалося, можна ввести *=СЛЧИС()* у рядок формул і натиснути клавішу *F9.* Щоб отримати випадкове дійсне число між *a* і *b*, можна використати таку формулу: *СЛЧИС() * (b – a) + a.*

Функція *СЛУЧМЕЖДУ (нижня межа, верхня межа)* повертає розподілене ціле випадкове число між двома заданими числами. Вона має два аргументи: *нижня межа* — це найменше ціле число,

що повертається цією функцією, а *верхня межа* — це найбільше ціле число, що повертається функцією.

Інструмент *Генерація випадкових чисел* виводить діалогове вікно, в якому потрібно вказати кількість змінних (кількість стовпців значень випадкових чисел, які треба розмістити у вихідному діапазоні); кількість випадкових чисел (кількість випадкових значень, які потрібно вивести для кожної змінної); вид розподілення (рівномірне, нормальне, Бернуллі, біноміальне, Пуассона, модельне або дискретне); параметри вибраного розподілення; випадкове розсіювання та вихідний діапазон, в якому будуть розміщені згенеровані випадкові числа.

3.2. ТЕХНОЛОГІЇ ПАРАМЕТРИЧНОГО АНАЛІЗУ МЕТОДАМИ «ЩО—ЯКЩО»

Одне з типових застосувань електронних таблиць — виконання аналізу «що—якщо». Суть аналізу «що—якщо» полягає в отриманні відповідей на питання: як вестиме себе досліджуваний об'єкт у різних ситуаціях, кожна з яких визначається певними значеннями вхідних даних? Так, за оцінювання конкретної маркетингової ситуації часто потрібно отримати відповіді на питання такого типу: що буде, якщо збільшити різницю між собівартістю і продажною ціною або збільшити заробітну плату співробітникам на 1 %? Яке зростання необхідне, якщо до початку наступного року треба досягти збільшення обсягів продажу на певну суму? Як зміниться прибуток, якщо змінити різницю між собівартістю та ціною продажу? Як відіб'ється на собівартості продукції підвищення цін на певні матеріали або на електроенергію?

Для спрощення аналізу «що—якщо» в Excel є декілька інструментів, до яких належать: *Добір параметра. Таблиця підстановки даних. Диспетчер сценаріїв.*

Добір параметра

Цей інструмент дає змогу швидко й легко отримати відповідь на питання: яке значення має бути на вході моделі (у найпростішому випадку модель може складатися лише з однієї формули), щоб отримати деяке конкретне значення на її виході? За відсутності такого інструмента для відповіді на це питання необхідно або вручну добирати вхідне значення, або створювати нову модель.

Пошук значення на вході моделі виконується так:

1.У меню *Сервіс* вибирається команда *Добір параметра.*

2.У поле *Встановити в комірці* вводиться посилання на клітину, що містить формулу, результат обчислення якої повинен мати потрібне значення.

3.У поле *Значення* вводиться значення, яке має бути отримане на виході моделі, тобто в клітині, посилання на яку було введене на попередньому кроці.

4.У поле *Змінювані значення в комірці* вводиться посилання на клітину на вході моделі, значення якої потрібно добрати. Значення цієї клітини має безпосередньо або через значення інших клітин впливати на результат формули, вказаної на кроці 2.

Таблиця підстановки даних

Цей інструмент дає змогу легко проаналізувати, як зміна певних значень у формулах впливає на результати, обчислювані за цими формулами. Він є засобом швидкого обчислення декількох варіантів, а також засобом перегляду і порівняння результатів різних варіантів на одному робочому аркуші. Зазвичай таку таблицю використовують як частину деякої моделі. Існує два типи таблиць підстановки даних — з однією і з двома змінними.

Таблиця підстановки з однією змінною складається з трьох частин:

— вектора-колонки (або вектора-рядка), значення елементів якого підставляються в указану клітину (*клітину введення даних*). Кількість елементів цього вектора залежить від кількості значень, що підставляються у клітину введення даних;

— вектора-рядка (або вектора-колонки), елементи якого містять формули, що прямо або опосередковано посилаються на клітину введення даних. Цей вектор може мати й один елемент (складатися лише з однієї формули);

— векторів-колонок (або векторів-рядків) даних. Елементи кожного з цих векторів обчислюються за однією з формул, яка знаходиться над цим вектором-колонкою (або перед цим вектором-рядком). Якщо використовується лише одна формула, то буде лише один вектор з обчисленими даними.

Слід звернути увагу на те, що клітина введення даних може входити, а може і не входити у таблицю підстановки з однією змінною. Тобто ця клітина може розміщуватись у будь-якому місці робочого аркуша.

Таблиця підстановки з однією змінною створюється у такій послідовності:

1. В окрему колонку аркуша або в окремий рядок уводиться список значень, які потрібно підставляти у клітину введення даних.
2. Якщо список значень розташовано у колонку, то формули вводяться в рядок, починаючи з клітини, що знаходиться на клітину вище та правіше від колонки значень. Якщо список значень розташовано в рядок, то формули вводяться у колонку починаючи з клітини, що знаходиться на клітину нижче й лівіше списку значень.
3. Виділяється прямокутник клітин, що містить формули та значення для підстановки в ці формули.
4. У меню *Дані* вибирається команда *Таблиця підстановки.*
5. Якщо значення для підстановки розміщено у колонку (а формули — у рядок), то посилання на клітину введення вказується в полі *Підставляти значення по рядках.* Якщо ж значення для підстановки розміщено в рядок (а формули — у колонку), то посилання на клітину введення даних указується в полі *Підставляти значення по колонках.*

Таблиця підстановки з двома змінними створюється таким чином:

1. У клітину аркуша вводиться формула, результат обчислення якої залежить від двох клітин уведення даних.
2. У ту саму колонку нижче формули вводяться значення підстановки для першої змінної. Значення підстановки для другої змінної вводяться в рядок за формулою.
3. Виділяється прямокутник клітин, що містить формулу й обидва набори даних підстановки.
4. У меню *Дані* обирається команда *Таблиця*.
5. У поле *Підставляти значення по колонках* уводиться посилання на одну з клітин уведення даних.
6. У поле *Підставляти значення по рядках* уводиться посилання на другу клітину введення даних.

Слід мати на увазі, що таблиці підстановки перераховуються кожен раз при перерахунку аркуша навіть у тому разі, коли в них не було внесено ніяких змін. Тому у випадку, коли таблиця підстановки містить значну кількість елементів, для прискорення процесу перерахунку аркуша слід так змінити параметр *Обчислення* на вкладці *Обчислення* (меню *Сервіс* команда *Параметри*), щоб автоматично перераховувався аркуш, але не таблиці. Коли ж виникає потреба у перерахунку таблиці, натискується клавіша *F9.*

Слід також звернути увагу на те, що діапазон клітин з даними, що обчислюються з допомогою таблиці підстановки (з одною або двома змінними), являє собою масив, окремі частини якого не можна відредагувати і не можна вилучити. Можна лише очистити весь діапазон клітин цього масиву разом.

Диспетчер сценаріїв

Цей інструмент аналізу «що—якщо» дає змогу створювати і зберігати на робочому аркуші різні набори значень (*сценарії*) вхідних даних, а потім переключатися на будь-який з них для перегляду результатів, щоб порівняти поведінку модельованого об'єкта в різних ситуаціях. Для того щоб задати умови, які характеризують певну ситуацію, потрібно, як правило, змінити значення декількох клітин, які містять вхідні дані. Тому таблиці даних, що дають змогу варіювати не більше двох змінних, для цього не завжди підходять. Саме в таких випадках застосовуються сценарії. З їх допомогою можна проводити багатофакторний параметричний аналіз, маніпулюючи одночасно 32 змінними.

Диспетчер сценаріїв дає змогу:

— створювати сценарії, вибираючи для кожного з них до 32 клітин із вхідними змінними;

— задавати значення кожної клітини, що містить вхідну змінну;

— присвоювати імена, зберігати і переглядати створені сценарії;

— створювати звіти для огляду і порівняння значень вхідних змінних і результатів сценаріїв;

— об'єднувати декілька сценаріїв в єдину модель;

— захищати сценарії від несанкціонованих змін;

— захищати сценарії від несанкціонованого перегляду.

Створення (додавання) нового сценарію. Сценарії можна створювати на кожному аркуші робочої книги, але слід мати на увазі, що вони можуть використовувати клітини лише того аркуша, на якому створюються.

Перш ніж створювати сценарій, треба перевірити, чи надано клітинам, що використовуватимуться в сценарії (клітинам із даними, що змінюватимуться, і клітинам із результатами обчислень, що переглядатимуться) змістовні інформативні імена. Якщо цього не зробити, то за огляду звіту сценаріїв буде важко отримати інформацію про наслідки, до яких призведе зміна певних вхідних даних.

Для визначення нового сценарію виконуються такі дії:

1. Вибирається команда *Сервіс*/*Сценарії* і в діалоговому вікні *Диспетчер сценаріїв* натискується кнопка *Додати.*

2. У діалоговому вікні *Додавання сценарію* у текстовому полі *Назва сценарію* набирається ім'я сценарію. На відміну від імен клітин назва сценарію може починатися з цифри, містити пропуски і мати довжину до 255 символів.

3. Якщо перед додаванням сценарію були виділені клітини робочого аркуша, то в текстовому полі *Комірки, що змінюються* з'явиться адреса цього діапазону клітин. У разі потреби в полі *Комірки, що змінюються* можна набрати посилання на потрібні клітини або виділити з допомогою миші. Якщо треба виділити несуміжні діапазони, то тримається натиснутою клавіша *Ctrl*. Виділятися можуть лише клітини з даними, а не ті, що містять формули.

4. За умовчання в текстовому полі *Коментар* з'явиться ім'я користувача і поточна дата. Можна вилучити або відредагувати цю інформацію.

5. Режим *Заборонити зміни* вибирається за умовчання, а режим *Сховати* — ні. Активізуються встановлені режими лише у разі включення захисту робочого аркуша командою *Сервіс*/*Захист*/*Захистити аркуш.*

6. За натискання кнопки *ОК* з'явиться діалогове вікно *Значення комірок сценарію.* Якщо клітинам, адреси яких визначалися в пункті 3, були надані імена, то вони з'являться в цьому діалоговому вікні. Тільки п'ять імен (або адрес клітин) будуть видимі. Щоб побачити інші, використовується смуга прокручування.

7. У разі потреби на даному кроці можна змінити значення. Якщо це перший сценарій на цьому аркуші, то можна зберегти його з величинами, заведеними за умовчання. Після введення потрібних значень можна або натиснути кнопку *Додати* і повернутись у діалогове вікно *Додавання сценарію*, або натиснути кнопку *ОК* і повернутись у вікно *Диспетчер сценаріїв.*

Використання сценаріїв. Найпростіший спосіб перегляду значень, що відповідають певному сценарію, — це використання діалогового вікна *Диспетчер сценаріїв.* Спочатку з допомогою смуг прокручування робочого аркуша в поле зору виводиться та частина робочого аркуша, де знаходяться відповідні вихідні значення. Після виконання команди *Сервіс*/*Сценарії* у текстовому полі *Сценарії* діалогового вікна *Диспетчер сценаріїв* вибирається назва потрібного сценарію і натискується кнопка *Вивести.* Це веде до заміни вхідних даних моделі, її перерахунку та виведенню значень відповідно до вибраного сценарію (попередні вхідні дані, якщо вони не були збережені у вигляді сценарію, будуть загублені).

Але такий спосіб аналізу варіантів розвитку подій придатний лише для відносно простих моделей. Для складніших використовуються підсумкові звіти для різних сценаріїв на додаткових робочих аркушах. Щоб отримати такий звіт, треба відкрити діалогове вікно *Диспетчер сценаріїв* (команда *Сервіс*/*Сценарії*) і натиснути кнопку *Звіт.* Це приведе до відкриття діалогового вікна *Звіт за сценарієм*, в якому є текстове поле *Комірки результату* і два перемикачі (*структура* і *зведена таблиця*), що визначають тип звіту.

У разі включення у звіт декількох клітин з результатами адреси цих клітин у полі *Комірки результату* мають відокремлюватися крапкою з комою (за введення адрес з допомогою миші треба тримати натиснутою клавішу *Ctrl*).

У структурному звіті всі імена клітин, які змінюються і в яких одержуються результати, розміщені у першій колонці. Дані й результати кожного сценарію містяться в окремій колонці.

У зведеній таблиці кожний сценарій показується у вигляді окремого рядка, при цьому значення кожної комірки результату відображається в окремій колонці.

Вивчаючи засоби побудови інструментальних маркетингових моделей, особливу увагу слід звернути на способи застосування цих моделей для прийняття маркетингових рішень з урахуванням факторів невизначеності у сфері вивчення попиту, визначення найвигіднішого асортименту продукції та сегментування ринку.

3.3. СТАТИСТИЧНІ МОДЕЛІ ПРИЙНЯТТЯ МАРКЕТИНГОВИХ РІШЕНЬ З УРАХУВАННЯМ ФАКТОРА НЕВИЗНАЧЕНОСТІ

В управлінській діяльності часто виникають ситуації, за яких доводиться приймати рішення, виходячи з недостатньо чітко визначених умов і оперуючи не завжди надійною інформацією. У таких випадках виникає потреба в оцінюванні ймовірності досягнення тих чи інших результатів. Наприклад, якщо буде прийняте рішення А, то ймовірність досягнення додаткового прибутку у розмірі 10 % становитиме 95 %, а за рішення В вона зменшиться до 65 %.

Існує три основних методи, що дають змогу приймати рішення з урахуванням фактора невизначеності та підтримуються Excel: стандартні відхилення, довірчі інтервали і множинний регресійний аналіз.

Стандартне відхилення

Стандартне (середнє квадратичне) відхилення — ступінь відхилення усіх значень ознаки від свого середнього показника — є одним із найважливіших методів, що допомагають визначити, наскільки змінюється певна величина: чим більше стандартне відхилення, тим ширший діапазон змін значень цієї величини.

Припустимо, що слід проаналізувати ефективність діяльності торгових агентів і продавців. Варто наголосити, що для більшості підприємств цей показник є найвагомішим серед тих, що впливають на збут. Під діяльністю торгових агентів і продавців розуміється робота, що проводиться ними безпосередньо з потенційними покупцями: за прилавком магазину, під час відвідування клієнтів за місцем проживання або прийому попередніх замовлень телефоном. Якщо такі працівники зацікавлені у результатах своєї діяльності, добре знають товар, який продають, і мають відповідний комерційний хист, то їх внесок в успіх фірми можна вважати вирішальним.

Природно розпочати аналіз з визначення середньої суми комісійних, яку фірма сплачує цій категорії своїх працівників. Функція Excel, яка розраховує середні показники, має назву *СРЗНАЧ*. Вона підсумовує значення клітин вказаного діапазону і ділить цю суму на кількість його клітин. Проте одна ця цифра ще не дає змоги оцінити роботу торгових агентів, оскільки середня величина — це узагальнена характеристика тієї сукупності, що вивчається. Вона не показує побудови сукупності, яка є дуже суттєвою для пізнання останньої. Окремі значення можуть зосереджуватися біля середньої величини (тоді вона добре представляє всю сукупність) або значно відхилятися від неї (погано представляє сукупність). Показником надійності середньої величини є стандартне (середньоквадратичне) відхилення: близько двох третин окремих елементів сукупності знаходяться на одне стандартне відхилення нижче або вище середнього показника.

Так, якщо за середньорічного заробітку 1440 грн стандартне відхилення дорівнює 107 грн, то це означає, що заробіток двох третин торгових агентів становить (1440 ± 107) грн, звідки випливає, що ефективність їхньої роботи практично збігається. Якщо ж заробіток двох третин торгових агентів становить, наприклад, (1440 ± 645) грн, то це означає, що середнє значення (1440 грн) погано представляє сукупність, а отже, значна кількість торгових агентів працює по-різному і є сенс у тому, щоб з'ясувати причини такого становища.

Для підвищення наочності такого аналізу використовуються гістограми (рис. 2.3.1) і точкові діаграми (рис. 2.3.2).

На горизонтальній осі рис. 2.3.1 відкладено різні значення комісійних, на вертикальній — результати спостережень за кожною категорією комісійних (кількість попадань у різні категорії виплат, кожна з яких відрізняється від попередньої на 200 грн). Слід звернути увагу на те, що за меншого значення стандартного відхилення ці результати мають менший розподіл зліва направо, ніж за більшого.

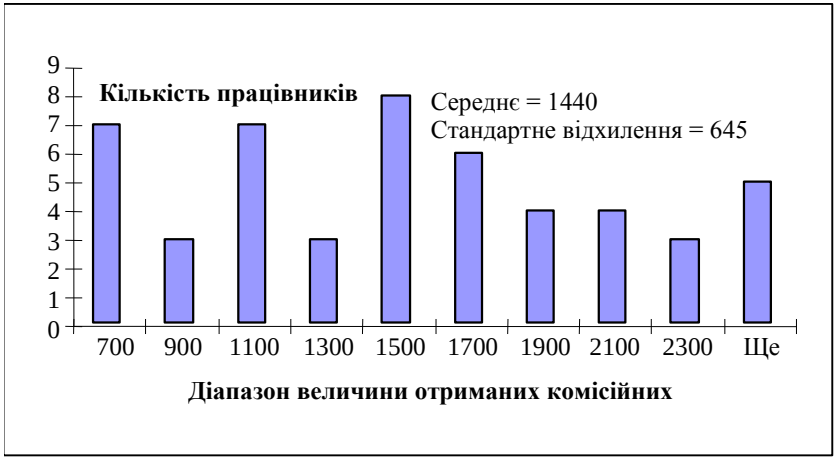


Рис. 2.3.1. Гістограма для різних значень стандартного відхилення

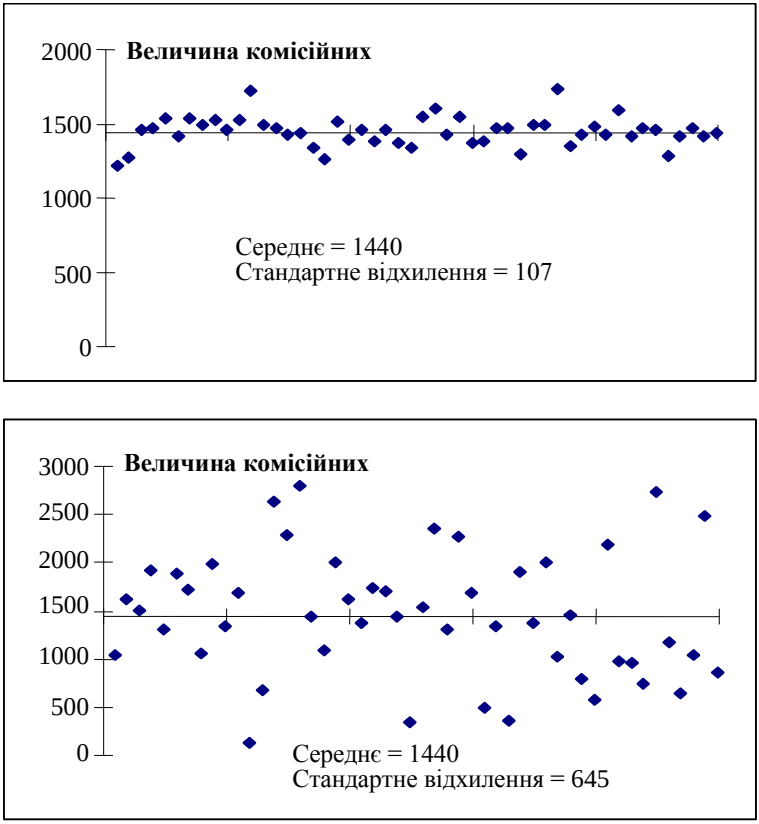


Рис. 2.3.2. Точкова діаграма за різних значень стандартного відхилення

На рис. 2.3.2 числові значення комісійних відображено на вертикальній осі, а порядковий номер спостереження — на горизонтальній. На верхній діаграмі всі значення близькі до середнього (незначне стандартне відхилення), на нижній спостерігається значне варіювання значень порівняно із середнім. Наведені діаграми показують, що знання середнього значення вибірки ще не є достатнім для прийняття рішення. Якщо ж відомі і середнє значення і стандартне відхилення, то це дає змогу досить чітко уявити, як значення згруповані біля середнього.

У разі суцільного обстеження, тобто коли до уваги беруться дані щодо кожного об’єкта сукупності, формула для обчислення стандартного відхилення має такий вигляд:

$$\sigma = \sqrt{\sum (x_i - x_{\text{сер}})^2 / n} \ ,$$

де $\sum (x_i - x_{\text{сер}})^2$ — сума квадратів усіх відхилень окремих значень від їх середньої величини;

n — кількість результатів спостережень.

На практиці до суцільного обстеження вдаються рідко. Найчастіше з усієї сукупності випадково відбирається обмежена кількість об'єктів для подальшого дослідження. Якщо в цьому разі для обчислення стандартного відхилення використати наведену вище формулу, то це призведе до систематичних похибок і даватиме зміщене (занижене) значення стандартного відхилення. Для виправлення значення вибіркового стандартного відхилення слід застосовувати формулу:

$$\sigma = \sqrt{\sum (x_i - x_{\text{сер}})^2 / (n-1)} \ .$$

Відповідно до цього в Excel є дві функції для обчислення стандартного відхилення: *СТАНДОТКЛОНП( )* і *СТАНДОТКЛОН( )*.

Функція *СТАНДОТКЛОНП( )* використовується для обчислення стандартного відхилення за суцільного обстеження, а *СТАНДОТКЛОН( )* — для обчислення стандартного відхилення за вибіркового обстеження. Буква *П* (у кінці назви першої функції) є мнемонікою слова *популяція* (генеральна сукупність).

Якщо, наприклад, є 30 результатів вибіркових спостережень, значення яких заведено у клітини *B5:B34* робочого аркуша, то для отримання стандартного відхилення всіх цих результатів треба використати формулу

= СТАНДОТКЛОН(B5:B34).

Слід бути вельми обережним у разі використання функцій *СТАНДОТКЛОНП( )* і *СТАНДОТКЛОН( )* для оброблення даних, значення яких або дуже великі (наприклад, 10^5 і більше), або дуже малі (наприклад, 10^{-5} чи менші), оскільки внаслідок впливу піднесення у квадрат різниці результатів спостережень та їх середнього значення будь-яка програма (а не тільки Excel) при округленні може видати похибку. Якщо необхідно працювати з такими даними, можна спробувати перед використанням значень цих функцій змінити масштаб чисел.

Крім стандартного відхилення, для характеристики відхилень значень ознаки сукупності від свого середнього значення

можна користуватися *дисперсією*, що дорівнює σ^2. Якщо стандартне відхилення можна подати як відстань, то дисперсію — як площину.

В Excel для обчислення дисперсії за суцільного обстеження застосовується функція *ДИСПР( )*, а за вибіркового — *ДИСП( )*.

Довірчий інтервал

У багатьох випадках отримати інформацію, яка б дала змогу приймати рішення зі 100-процентним ступенем упевненості, або зовсім неможливо, або дуже складно і потребує великих витрат. У таких випадках найкраще отримати вибірку з інформації. Використання вибірки уможливлює обчислення статистичного показника, який даватиме (можливо) досить точне наближення до результату, що міг бути отриманий за наявності доступу до повного набору даних. Точність такого результату можна вимірювати з допомогою довірчих інтервалів.

Довірчий інтервал — це інтервал, що дає змогу оцінити із заданою точністю невідоме значення генеральної сукупності. Таке невідоме значення називається довірчим, а його границі — довірчими границями (верхні та нижні границі). В їх межах можна мати деякий рівень упевненості щодо наявності конкретного значення ознаки генеральної сукупності.

Довірчі інтервали широко застосовуються для дослідження ринку. Припустимо, що розглядається питання про відкриття нового магазину. Важливим критерієм за вирішення цього питання може бути кількість потенційних покупців, що проходитимуть повз магазин. Щоб з'ясувати це, можна кожного дня протягом деякого періоду підраховувати кількість пішоходів. Одержані результати створять вибірку з генеральної сукупності всіх можливих днів, коли магазин працюватиме.

Далі можна підрахувати середній показник результатів таких спостережень й отримати певну цифру. Але відразу постає питання — наскільки точно ця цифра характеризує дійсну кількість людей, що проходитимуть повз магазин кожного дня?

Відповідь можна знайти з допомогою довірчого інтервалу середнього значення. Щоб визначити його в Excel, слід підключити пакет аналізу (якщо його не було підключено раніше):

1. Вибрати команду *Сервіс*/*Надбудова.*
2. У діалоговому вікні *Надбудова*, яке з'явиться після виконання попередньої команди, вибрати параметр *Пакет аналізу*.
3. Закрити діалогове вікно, клацнувши по кнопці *ОК*.

4. Якщо пакет аналізу підключено, довірчий інтервал можна визначити, виконавши такі дії.

5. Вибрати команду *Сервіс*/*Аналіз даних.*

6. Вибрати у списку діалогового вікна *Аналіз даних* інструмент аналізу *Описувальна статистика*.

7. Закрити список інструментів аналізу, клацнувши по кнопці *ОК*, що має призвести до появи діалогового вікна *Описувальна статистика*.

8. У поле *Вхідний інтервал* увести або виділити мишею той діапазон (колонку чи рядок) робочого аркуша, у який занесено результати спостережень.

9. Включити параметр *Мітки у першому рядку* і перевірити значення у полі *Рівень надійності*. Він звичайно має дорівнювати 95 %.

10. Включити перемикач *Вихідний інтервал* і ввести в поле, що знаходиться біля нього, адресу лівого верхнього кута області робочого аркуша, де розміщуватиметься результат аналізу (значення статистичного показника).

11. Клацнути на кнопці *ОК*.

Знаючи статистичний показник, можна легко визначити довірчий інтервал. Для цього слід додати значення статистичного показника до середнього (обчислюється з допомогою функції *СРЗНАЧ*) і відняти від середнього. Що криється за отриманими значеннями довірчого інтервалу? Якщо припустити, що дослідження буде повторене 100 разів, то буде отримано 100 двотижневих значень і 100 відповідних довірчих інтервалів; 95 з цих інтервалів включатимуть середній показник для генеральної сукупності. Тобто реальне середнє значення генеральної сукупності знаходитиметься між нижньою і верхньою межами інтервалу. І лише п’ять довірчих інтервалів не охоплюватимуть реального середнього значення сукупності. Звичайно, логічніше припустити, що проведений експеримент є одним з 95 гіпотетичних експериментів, довірчий інтервал яких включає середнє значення генеральної сукупності, ніж те, що він входить у ті п’ять, які не охоплюють цього середнього. На підставі такого припущення робиться висновок, що реальна кількість людей, які проходять повз магазин у будь-який конкретний день, знаходитиметься у межах цього довірчого інтервалу. Отже, слід лише встановити, чи відповідає така точність оцінки реальним цілям і чи достатньо велика ця кількість для того, щоб можна було прийняти позитивне рішення.

На значення довірчих інтервалів впливають декілька факторів, одним з яких є стандартне відхилення результатів спостережень.

Цей фактор не регулюється, оскільки він базується на даних, отриманих в результаті відповідних спостережень.

Інший фактор — точність (рівень надійності). Значення цього фактора можна регулювати. Чим більша ширина довірчого інтервалу, тим вища точність оцінки. Дослідник може зменшити інтервал, зменшивши рівень надійності, наприклад, до 90 %, але при цьому знизиться й ступінь довіри до отриманих оцінок. Найнадійнішим способом зменшення інтервалу із збереженням при цьому прийнятного ступеня довіри є збільшення розмірів вибірки. Проте повної гарантії точності оцінки цей спосіб також не дає. Може навіть статися, що зростання розмірів вибірки призведе до зростання стандартного відхилення. І це обов'язково станеться, якщо отримані додаткові результати спостережень досить суттєво відрізнятимуться від середнього значення попередньої вибірки.

Регресійно-кореляційний аналіз

Регресійний і кореляційний аналіз — дуже ефективні методи, які дають змогу аналізувати значні обсяги інформації з метою дослідження ймовірного взаємозв'язку двох чи більше змінних. У регресійному аналізі розглядається зв'язок між однією змінною, названою залежною змінною, або ознакою, і кількома іншими, названими незалежними змінними. Цей зв'язок подається з допомогою математичної моделі, тобто рівняння, яке зв'язує залежну змінну (y) з незалежними (x) з урахуванням множини відповідних припущень. Оскільки метою регресійного аналізу є виявлення впливу змінних x на значення змінної y, останню ще називають відгуком, або результативним фактором, а змінні x — факторами, що впливають на відгук. Регресійний аналіз використовується з двох причин. По-перше, тому що опис залежності між змінними допомагає встановити наявність можливого причинного зв'язку. По-друге, отримання аналітичної залежності між змінними дає змогу передбачати майбутні значення залежної змінної за значенням незалежних змінних.

За аналізу соціально-економічних процесів регресія застосовується водночас з кореляцією. З допомогою регресії визначаються аналітичні залежності між змінними, а через кореляційний аналіз — сила зв'язку між факторами та відгуком. Саме тому, що основні статистичні проблеми регресійного аналізу вирішуються аналізом кореляцій, методи регресійного та кореляційного аналізу тісно зв'язані між собою.

Математичний апарат кореляційно-регресійного аналізу. Кореляційно-регресійний аналіз складається з таких основних етапів: побудова системи факторів, які найсуттєвіше впливають на результативну ознаку; розроблення моделі, яка відбиває загальний зміст взаємозв'язків, що вивчаються, та кількісне оцінювання її параметрів; перевірка якості моделі; оцінювання впливу окремих факторів.

На першому етапі здійснюється відбір факторів, які найсуттєвіше впливають на результатну ознаку. Він проводиться, перш за все, виходячи із змістовного аналізу. Для отримання надійних оцінок у модель не слід включати забагато факторів, їх кількість має бути не більше однієї третини обсягу аналізованих даних. Але оскільки на початковому етапі розроблення моделі у дослідника немає однозначної відповіді на питання щодо набору суттєвих факторів, то у разі використання ЕОМ відбір факторів звичайно здійснюється безпосередньо в процесі створення моделі за методом послідовної регресії. Суть цього методу полягає у послідовному включенні додаткових факторів у модель та оцінюванні впливу доданого фактора. Використовується також підхід, за якого на фактори, включені у попередній склад моделі, не накладаються особливі обмеження і лише на наступних стадіях проводиться їх оцінювання та часткове відсіювання.

Другий етап починається з розроблення моделі, яка відбиває загальний зміст аналізованих взаємозв'язків. *Регресійна модель* — це рівняння (або система рівнянь), що показує, які фактори, на думку дослідника, мають бути залучені до взаємозв'язків, котрі підлягають аналізу. Регресійне рівняння дає також уявлення про форму зв'язку.

Регресія називається *парною*, якщо вона відбиває залежність між результатною та однією факторною ознаками. Методологія парної кореляції — найбільш розроблена в теорії статистики. Вона є фундаментом для вивчення та застосування інших методів аналізу кореляційного зв'язку.

Регресія називається *множинною*, якщо вона відбиває залежність результатної ознаки від декількох факторів.

Якщо залежності є лінійними відносно параметрів (але не обов'язково лінійною відносно незалежних змінних), то регресія називається *лінійною*. В протилежному випадку регресію називають *нелінійною.*

Під простою регресійною моделлю розуміють парну регресію. У цьому разі статистичний підхід до побудови функціональної залежності y від x базується на припущенні, що є вибірка парних спостережень (x_1, y_1), (x_2, y_2), …, (x_n, y_n) з деякої популяції. Пару значень (x_i, y_i) часто називають результатом одного вимірювання, а n — кількістю вимірювань.

Для побудови реальної регресійної моделі може використовуватися численна кількість рівнянь. Наприклад, взаємозв'язок частки ринку та відмінностей у ціні можна подати так:

$$\text{Частка ринку} = \frac{\text{Обсяг продажу конкретного підприємства}}{\text{Обсяг продажу галузі}};$$

$$\text{Різниця в цінах (у \%)} = \frac{\text{Середня ціна підприємства} - \text{Середня ціна конкурента}}{\text{Середня ціна конкурента}}.$$

Кожну з цих змінних можна подати в декілька способів, використовуючи їх зв'язки з певними факторами:

Залежна змінна (результатна ознака)	Незалежна змінна (фактор)
Обсяги продажу в натуральних одиницях	Середня ціна підприємства
Обсяги продажу в гривнях	Різниця в ціні за одиницю товару
Кількість сімей, що користуються товаром	Середня кількість продукту, куплена сім'єю за місяць

Вибір способів подання зв'язків між змінними визначає дані, які треба зібрати для побудови моделі.

Наступний крок після вибору змінних і способу їх подання – визначення форми рівняння регресії. Тут може стати у пригоді графічне зображення точок (x_1, y_1), (x_2, y_2), …, (x_n, y_n) на площині xy, назване *діаграмою розсіяння* (рис. 2.3.3).

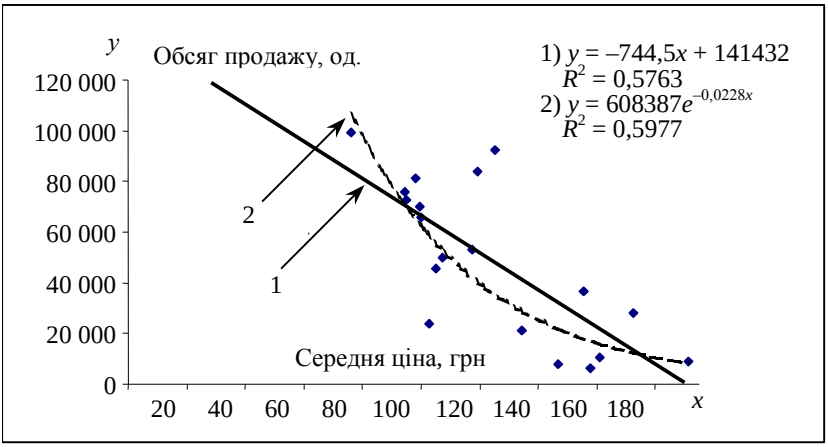


Рис. 2.3.3. Лінії еластичності попиту

Діаграма показує, що зі зростанням ціни частка ринку підприємства дійсно має тенденцію до зменшення. Але яка загальна форма взаємозв'язку? На рисунку цей взаємозв'язок має вигляд прямолінійного (лінія 1) та криволінійного (лінія 2) рівняння. Вибір потрібного виду рівняння регресії залежить від знання проблеми та досвіду.

В основу виявлення і встановлення аналітичної форми зв'язку покладено використання певних математичних функцій — лінійної, логарифмічної, степеневої, експоненційної, поліноміальної та деяких інших. У разі парної кореляції ці функції записуються так:

лінійна	$y = mx + b$;
логарифмічна	$y = mln\,(x) + b$;
степенева	$y = bx^m$;
експоненційна	$y = be^{mx}$;
поліноміальна	$y = m_6x^6 + m_5x^5 + \ldots + b$.

Практично для обчислення параметрів функцій застосовуються спеціальні комп'ютерні програми, серед яких найбільші можливості для тлумачення результатів користувачеві надають програми лінійного регресійного аналізу. Тому більшість аналітиків віддають перевагу саме йому. Але оскільки лінійні рівняння концептуально являють собою найпростіший тип взаємозв'язку, то його використання потребує особливої осторожи.

Для наведеного вище прикладу проста лінійна регресійна модель може бути подана так:

$$\begin{array}{c}\text{Обсяги}\\ \text{продажу}\\ \text{підприємства}\end{array} = \begin{array}{c}\text{Місткість}\\ \text{базового}\\ \text{ринку}\end{array} + \begin{array}{c}\text{Коефіцієнт}\\ \text{«чутливості»}\\ \text{ринку до цін}\end{array} \times \begin{array}{c}\text{Ціна}\\ \text{виробів,}\end{array}$$

або

$$y = b + m_1x_1,$$

де $m_1 < 0$.

Відомо, що на частку ринку, яка зайнята товаром, впливають, крім ціни, й інші фактори. А оскільки майже всі проблеми в маркетингу включають у себе кілька різних факторів, то у більшості випадків слід застосовувати моделі множинної регресії. Концепції і методи, використовувані у множинному регресійному аналізі, практично ті самі, що й у простому, але з деякими модифікаціями та доповненнями, пов'язаними з вивченням кількох факторів одночасно. Створення моделі множинної регресії пов'язане також із додатковими труднощами. Перш за все, неможливо дослідити взаємоз'язки з допомогою діаграми розсіяння. Її, звичайно, можна використовувати для відображення взаємозв'язку резуль-

тативної ознаки й кожного із незалежних факторів по черзі. Але не слід забувати, що отримана в такий спосіб інформація має обмежене значення, оскільки незалежні змінні часто впливають одна на одну так само, як і на результативну ознаку.

У розглядуваному прикладі природно припустити, що на обсяги продажу, крім цін, впливають й витрати на рекламу та дохід споживачів. Отже, наведене вище лінійне рівняння перетворюється на таке:

$$y = b + m_1x_1 + m_2x_2 + m_3x_3,$$

де x_2 — витрати на рекламу;

x_3 — дохід споживачів.

У разі використання степеневої функції матимемо таку модель:

$$y = bx_1^{m_1}x_2^{m_2}x_3^{m_3}.$$

Очевидно, що останнє рівняння є складнішим. Наприклад, вплив на обсяги продажу зміни ціни на 1 грн буде різним залежно від значень інших незалежних факторів. Проте це рівняння є гарним прикладом нелінійного рівняння, яке легко трансформувати у лінійну форму:

$$\log(y) = \log(b) + m_1\log(x_1) + m_2\log(x_2) + m_3\log(x_3).$$

Дані, що збираються для проведення регресійного аналізу, звичайно являють собою «історичні» відомості, тобто цифри, що показують значення кожного із факторів у кожному з попередніх періодів часу або географічних районів. Вони використовуються для отримання оцінок коефіцієнтів регресії та визначення ступеня відповідності моделі дійсним змінам результативної ознаки. Для досягнення цих цілей можна застосовувати різні методи, але найуніверсальнішим і найчастіше використовуваним є *метод найменших квадратів*. Оцінки за методом найменших квадратів — це ті величини коефіцієнтів регресійного рівняння, які мінімізують суму квадратів відхилень дійсно спостережуваних значень результатної ознаки (y_i) від тих значень, що отримуються з рівняння $y(x_i)$:

$$\sum (y_i - y(x_i))^2 = \min.$$

Метод найменших квадратів дає змогу мінімізувати дисперсію оцінок, а отже, й ступінь невизначеності, пов'язаний з оцінками. У цьому разі дисперсія фактичних значень результативної ознаки від обчислених за рівнянням визначається як

$$\sigma^2 = SS_{зал}/N,$$

де $SS_{зал} = \sum(y_i - y(x_i))^2$ — сума квадратів рівнів залишкової компоненти;

N — кількість спостережень.

Для правильного використання результатів, отримуваних на «виході» регресійного аналізу, слід розумітися на суті цих даних і проблемах, пов'язаних з їх тлумаченням. Стосовно аналізу частки ринку підприємства як функції від ціни виробів з допомогою простої лінійної моделі параметри регресії можуть тлумачитися так:

— незмінна b — це відрізок на осі координат. Він показує, в якому місці лінія тренда перетинає вісь y (вертикальну вісь). У даному прикладі це значення місткості ринку товару. Проте таке тлумачення не завжди можливе, оскільки результат може являти собою оцінку впливу факторів, не включених в аналіз;

— коефіцієнт m рівняння називається *коефіцієнтом регресії*. Він є мірою нахилу лінії регресії: чим він більший, тим крутіша лінія регресії. У наведеному прикладі коефіцієнт m — це коефіцієнт чутливості ціни, який відбиває зміну частки ринку за зміни ціни.

На етапі перевірки якості моделі її оцінюють за адекватністю і точністю. Сенс такої перевірки полягає у тому, щоб обґрунтувати застосування методу функціонального аналізу для вивчення кореляційної залежності. Це буде правомірним лише у тих випадках, коли кореляційний (співвідносний) зв'язок не дуже віддалений від функціонального (жорсткого) зв'язку.

Оскільки модель відображує вплив на результативну ознаку лише частини реальних факторів, регресійний аналіз пояснює тільки частину дисперсії відгуку (загальної дисперсії). Таким чином:

$$\text{Загальна дисперсія} = \text{Дисперсія, що пояснюється регресійним аналізом} + \text{Залишкова дисперсія}.$$

Залишкова дисперсія — це та частина варіації залежної змінної, яку не можна пояснити впливом факторів, включених у регресійне рівняння.

Для оцінювання якості моделі і повноти набору пояснювальних факторів звичайно використовують *коефіцієнт детермінованості* R^2. Його ще називають *величиною вірогідності апроксимації*, або *рівнем надійності*. Коефіцієнт детермінованості — це відношення дисперсії, що пояснюється регресійним аналізом, до загальної дисперсії. Він звичайно обчислюється за формулою

$$R^2 = SS_{\text{рег}} / (SS_{\text{рег}} + SS_{\text{зал}}),$$

де $SS_{\text{рег}} = \sum(y(x_i) - y_{\text{сер}})^2$ — сума квадратів відхилень рівнів вихідного ряду даних від його середнього значення;

$SS_{\text{зал}}$ — сума квадратів рівнів залишкової компоненти.

Коефіцієнт детермінації дає кількісну оцінку міри аналізованого зв'язку. Він показує частку варіації результативної ознаки, що знаходиться під впливом факторів, що вивчаються, тобто визначає, яка частка варіації ознаки *y* враховується у моделі й обумовлена впливом на неї незалежних факторів. Чим ближче R^2 до 1, тим у більшому ступені рівняння регресії пояснює аналізований фактор (за функціонального зв'язку R^2 дорівнює 1, а за відсутності зв'язку — 0). Якщо, наприклад, R^2 дорівнює 0,9, то можна вважати, що 90 % змін (варіацій) у відгуку обумовлюються варіаціями в урахованих факторах і лише 10 % — за рахунок впливу інших факторів. Величина *R* називається *індексом кореляції* (*множинне R*). Цей коефіцієнт, як і R^2, є універсальним, оскільки відбиває щільність зв'язку й точність моделі і може використовуватися за будь-якої форми зв'язку. За прямолінійного зв'язку індекс кореляції дорівнює *коефіцієнту кореляції* (*r*).

Для полегшення висновків щодо практичної значимості синтезованої моделі показникові щільності зв'язку дається якісна оцінка. Це здійснюється на основі шкали Чеддока:

Показник щільності зв'язку *r*	0,1—0,3	0,3—0,5	0,5—0,7	0,7—0,9	0,9—0,99
Характеристика сили зв'язку	Слабка	Помірна	Помітна	Значна	Вельми значна

Виходячи з величини індексу детермінації маємо, що у разі значної залежності результативної ознаки від факторів більше половини загальної варіації відгуку пояснюється впливом факторів, що вивчаються. Це дозволяє вважати виправданим застосування методу функціонального аналізу для вивчення кореляційного зв'язку, а синтезовані при цьому математичні моделі визнаються придатними для практичного використання. Якщо значення показника щільності зв'язку нижче 0,7, то величина індексу детермінації завжди буде менше 50 %. Це означає, що на частку варіації факторів, що вивчаються, припадає менша частина порівняно з іншими факторами, що впливають на змінну загальної дисперсії результатної ознаки. Синтезовані за таких умов математичні моделі практичного значення не мають. Але високий рівень показників щільності зв'язків не є гарантією того, що фактори, включені в модель, дійсно є основними. Може статися так, що вони лише відбивають вплив інших, глибинніших факторів. Так, чисельність населення може бути в дійсності важливішим фактором, ніж величина доходу споживачів.

Для перевірки міри точності застосовують незміщену *оцінку дисперсії* залишкової компоненти

$$MS_{зал} = SS_{зал} / df,$$

де df — ступінь вільності, що дорівнює $N - m - 1$ (N — кількість спостережень, m — кількість незалежних факторів).

Квадратний корінь з цієї величини називається *стандартною помилкою оцінки*. Для регресійного рівняння в цілому вона виступає як ступінь точності прогнозів, що базуються на рівнянні. Так, для розглядуваного прикладу з її допомогою можна визначити ймовірність того, що дійсний рівень частки ринку знаходитиметься в конкретному інтервалі близько до значення, яке виводиться з рівняння. Чим більше незалежна змінна відрізняється від середньої за спостереженнями, використовуваними для оцінки коефіцієнтів, тим більша невизначеність у прогнозі, що базується на рівнянні регресії.

Для перевірки значущості моделі регресії використовується *F-критерій* Фішера (*F-відношення*), обчислюваний за формулою

$$F_0 = MS_{рег} / MS_{зал},$$

де $MS_{рег} = SS_{рег} / m$.

Якщо обчислене значення F-критерію більше за його критичне значення, що визначається за таблицею, то значення коефіцієнта детермінованості визнається суттєвим (невипадковим), а модель — значущою. За визначення критичного значення F-критерію враховується прийнятий рівень значущості (0,05 або 0,01) і число ступенів вільності df_1 і df_2 ($df_1 = m$, $df_2 = N - m - 1$).

Слід також проаналізувати значущість окремих коефіцієнтів регресії, тобто виявити, наскільки обчислені параметри регресійного рівняння характерні для конкретного комплексу умов, чи не є отримані значення параметрів рівняння регресії дією випадкових причин. Для вивчення кореляційного зв'язку показників ринкової діяльності це особливо важливо, оскільки розглядаються сукупності, які звичайно мають порівняно невелику кількість елементів. Це здійснюється з допомогою t-статистики шляхом перевірки гіпотези про рівність нулеві відповідного параметра рівняння. Якщо обчислене значення t-критерію з ($N - m - 1$) ступенями вільності перевищує його табличне значення за заданого рівня значущості, коефіцієнт регресії вважається значущим. У противному разі фактор, що відповідає такому коефіцієнту, слід вивести з моделі (при цьому її якість не погіршиться).

Важливу роль в оцінюванні впливу факторів відіграють коефіцієнти регресійного рівняння. Проте безпосереднє їх порівняння недопустиме з таких причин:

— значення коефіцієнта регресії залежить від одиниць його вимірювання. Якщо витрати на рекламу виражено у тисячах гривень, то можна довільно збільшити коефіцієнт шляхом заміни одиниці вимірювання на десятки тисяч, сотні тисяч і т. д.;

— незалежні фактори звичайно мають різну коливність (різні варіації);

— незалежні фактори зв'язані не тільки з результатним, а й деякою мірою і між собою. Отже, навіть якщо попередні причини не заважають прямому зіставленню коефіцієнтів, кожен з них не є «чистою» мірою впливу відповідного фактора на результатну ознаку.

У загальному випадку для того щоб зробити коефіцієнти регресії порівнянними, використовують *часткові коефіцієнти еластичності, β-коефіцієнти регресії* та *коефіцієнти часткової (парної) кореляції.*

Коефіцієнт еластичності показує, на скільки процентів зміниться результативна ознака за зміни *j*-го фактора на 1 %, якщо значення решти факторів фіксується на деякому рівні. Якщо за такий рівень узяти середні значення факторів, то отримаємо середній коефіцієнт еластичності

$$E_j = m_j \cdot X_{j\,\text{сер}} / Y_{\text{сер}}.$$

Вадою коефіцієнтів еластичності є те, що вони не враховують ступінь коливності факторів.

Бета-коефіцієнт (*β-коефіцієнт*) показує величину зміни результатної ознаки в значеннях середньої квадратичної помилки за зміни *j*-го фактора на одну середньоквадратичну (стандартну) помилку у разі фіксації значень решти факторів:

$$\beta_j = m_j \cdot \sigma_j / \sigma_y.$$

Парний коефіцієнт кореляції r_j — це показник, що характеризує щільність зв'язку між результатною ознакою та *j*-м фактором за елімінації (виключення впливу) всіх інших факторів.

Зазначені коефіцієнти дають змогу зробити ранжування факторів за ступенем їх впливу на залежну змінну. Оцінити частку впливу *j*-го фактора у сумарному впливі всіх факторів, включених у регресію, можна за значенням *дельта-коефіцієнта* (Δj) цього фактора:

$$\Delta j = r_j \cdot \beta_j / (r_1 \cdot \beta_1 + r_2 \cdot \beta_2 + \ldots + r_m \cdot \beta_m).$$

Інший спосіб оцінювання важливості певного фактора полягає в оцінюванні зміни коефіцієнта детермінації при додаванні або виключенні з рівняння регресії цього фактора.

Excel забезпечує ефективну підтримку побудови та аналізу регресійних моделей: 15 функцій робочих аркушів, створених саме з цією метою, а також такі можливості, як побудова лінії тренда на графіках, та інструмент аналізу *Регресія*, з допомогою яких зручно проводити конкретні регресійні обчислення.

Найбільш наочний спосіб дослідження зв'язку між двома змінними базується на використанні точкової діаграми з лінією тренда. Крім того, Excel має набір спеціальних функцій, які за певних умов у використанні зручніші за діаграми.

Так, для обчислення значення R^2 можна використати функцію *КВПИРСОН*.

Відрізок на осі ординат можна отримати з допомогою функції *ОТРЕЗОК*. Коефіцієнт нахилу лінійної регресії — за допомогою функції *НАКЛОН*. Щоб отримати відрізок на осі координат і коефіцієнт нахилу з допомогою однієї функції, слід виділити дві клітини, натиснути на панелі інструментів кнопку *Вставка функції*, у діалоговому вікні вибрати функцію *ЛИНЕЙН*, вказати перші два параметри (діапазон клітин, де знаходяться значення відгуку, та діапазон клітин зі значеннями незалежної змінної) і, тримаючи натиснутими клавіші *Ctr* і *Shift*, натиснути клавішу *Enter*.

Одночасно з обчисленням параметрів лінійного рівняння регресії (у тому числі й множинної) функція *ЛИНЕЙН* може повертати додаткову регресійну статистику. До цієї статистики входять:

se_1, …, se_n — стандартні значення помилок для коефіцієнтів m_1, …, m_n;

se_b — стандартне значення помилки для постійної b;

R^2 — величина вірогідності апроксимації (коефіцієнт детермінації);

se_y — стандартна помилка для оцінки y;

F — F-статистика, або F-відношення;

df — кількість ступенів вільності ($N - m - 1$);

$SS_{рег}$ — регресійна сума квадратів;

$SS_{зал}$ — залишкова сума квадратів.

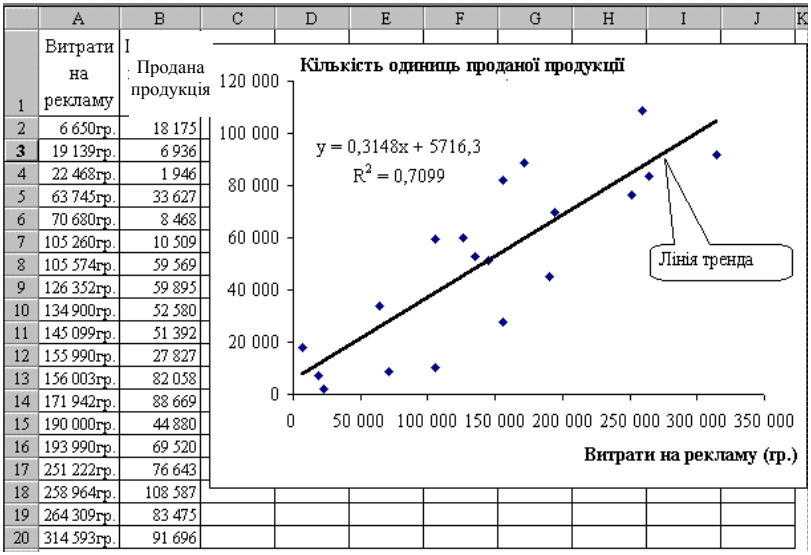
Розглянемо технологію проведення регресійно-кореляційного аналізу з допомогою Excel на конкретному прикладі. Припустимо, що треба дослідити результати збільшення витрат на рекламу деякої продукції і зниження ціни на одиницю цієї продукції з метою збільшення обсягів продажу. Почнемо з перевірки пропозиції про збільшення витрат на рекламу. Звичайно, не можна не враховува-

ти того, що реальний продаж додаткових обсягів продукції може навіть не виправдати витрат на рекламну кампанію. Для з'ясування цього питання слід виявити зв'язок між витратами на рекламу для кожного виду продукції і кількістю одиниць продаваної продукції. Якщо є необхідні дані для проведення регресійного аналізу, то цей зв'язок можна оцінити у кількісній формі.

На рис. 2.3.4 наведено таблицю, що має дві колонки — витрати на рекламу і кількість одиниць проданої продукції. Зв'язок між цими змінними можна легко (хоча і в дещо спрощеній формі) оцінити з допомогою точкової діаграми.

Для побудови цієї діаграми слід виділити дані в діапазоні A2:B20, вибрати команду *Вставка*/*Діаграма* (або клацнути на кнопці *Майстер діаграм*) і ввести потрібну інформацію на кожному з чотирьох кроків побудови діаграми. Зауважимо, що на першому кроці вибирається тип діаграми, яка має назву *Точкова*.

Коли діаграма з'явиться у робочому аркуші, треба клацнути по ній лівою кнопкою миші і вибрати команду *Діаграма*/*Додати лінію тренда*. У діалоговому вікні *Лінія тренда* на вкладці *Тип* вибирається тип апроксимації *Лінійна*, а на вкладці *Параметри* встановлюються режими: *Показувати рівняння на діаграмі* та *Розмістити на діаграмі величину вірогідності апроксимації (R^2)*.



	A	B
1	Витрати на рекламу	Продана продукція
2	6 650гр.	18 175
3	19 139гр.	6 936
4	22 468гр.	1 946
5	63 745гр.	33 627
6	70 680гр.	8 468
7	105 260гр.	10 509
8	105 574гр.	59 569
9	126 352гр.	59 895
10	134 900гр.	52 580
11	145 099гр.	51 392
12	155 990гр.	27 827
13	156 003гр.	82 058
14	171 942гр.	88 669
15	190 000гр.	44 880
16	193 990гр.	69 520
17	251 222гр.	76 643
18	258 964гр.	108 587
19	264 309гр.	83 475
20	314 593гр.	91 696

Рис. 2.3.4. Зв'язок між обсягами продажу (в одиницях продукції) і витратами на рекламу

На рис. 2.3.4 лінію тренда подано прямою, яка йде з нижнього лівого кута у правий верхній. Це говорить про те, що за збільшення витрат на рекламу пропорційно збільшуються й обсяги продажу (в одиницях продукції).

Діаграма також вміщує рівняння

$$y = 0{,}3148x + 5716{,}3.$$

Це *рівняння регресії*. Воно найкраще апроксимує дані, наведені у робочому аркуші, у вигляді прямої ($y = mx + b$). Для даного прикладу таке рівняння показує залежність між сумою грошей, витрачених на рекламу певної продукції, і обсягами продажу в одиницях цієї продукції (y — це обсяги продажу, x — витрати на рекламу в гривнях). Це не означає, що за відомих витрат на рекламу можна точно визначити обсяги продажу. Наприклад, у разі підстановки в рівняння замість x значення 171 942 грн (сума витрат на рекламу) буде отримано результат — 59843,64 одиниць продукції. У таблиці з даними значення 171 942 грн наведено як один з фактичних результатів спостереження, але цьому значенню відповідає зовсім інша величина обсягів продажу — 88 669 одиниць продукції. Отже, регресія на основі тих даних, які ми маємо, дає найбільш точну оцінку, але не абсолютно точний прогноз. Точність апроксимації з допомогою прямої залежить від ступеня розкиду даних. Чим ближче дані до прямої, тим точнішою є лінійна регресійна модель.

Розглядаючи принципи регресійного аналізу, дуже важливо зрозуміти, що регресія виражає зв'язок між змінними, а це не те саме, що причинна обумовленість, яка означає, що маніпуляції з однією змінною обов'язково приведуть до певних змін іншої.

Якщо в дане рівняння регресії підставити значення 400 000 грн, то отримаємо, що приблизний обсяг продажу в одиницях продукції дорівнюватиме 531636. Це зовсім не означає, що, витративши на рекламу 400000 грн, підприємство продасть 531 636 одиниць цієї продукції (хоча, звичайно, цього і не можна виключати). Є велика кількість факторів, крім витрат на рекламу, які впливають на зміни в обсягах продажу, і ці фактори (наприклад, продажна ціна одиниці продукції) у цьому рівнянні регресії ніяк не відображені. Навіть якщо зв'язок між змінними має причинний характер, треба ще знати напрям цієї причинності. Цілком вірогідно, що відділ маркетингу збільшував витрати на рекламу продукції після того, як збільшувались обсяги продажу. У цьому разі ми можемо бути впевненими лише у впливі обсягів продажу на розміри витрат на рекламу, а не навпаки.

Коефіцієнт регресії *m* є мірою нахилу лінії тренда: чим він більший, тим крутіша лінія тренда. У даному разі — це число 0,3148. Воно інтерпретується так: «Якщо між обсягами продажу та витратами на рекламу є зв'язок, то, за попередніми оцінками, за збільшення витрат на рекламу на 10 000 грн буде продано додатково приблизно 3148 одиниць продукції».

Незмінна *b* — це відрізок на осі ординат. Вона вказує, в якому місці лінія тренда перетинає вісь *y* (вертикальну вісь). У даному разі — це число 5716,3. Воно інтерпретується так: «Якщо між обсягами продажу та витратами на рекламу є зв'язок і якщо не виділятимуться гроші на проведення реклами, то, за попередніми оцінками, буде продано приблизно 5716 одиниць продукції».

Значення коефіцієнта детермінації на діаграмі дорівнює 0,7099. Це означає, що приблизно 71 % міри мінливості обсягу продажу одиниць продукції зв'язано з мірою мінливості витрат на рекламу.

На рис. 2.3.5 у діапазоні *A2:C20* знаходяться дані спостережень про обсяги продажу ($y_ф$), які включають, крім витрат на рекламу (x_1), ціни (x_2), за якими продавалися рекламовані товари.

	A	B	C	
1	Витрати на рекламу (x_1), грн	Ціна (x_2), грн	Продана продукція ($y_ф$), од.	
2	6 650	147,2	7 175	
3	19 139	158,5	5 836	
4	22 468	161,5	9 946	
5	63 745	103,2	23 627	
6	70 680	191,9	8 468	
7	105 60	134,9	20 509	
8	105 574	107,8	49 569	
9	126 352	155,8	35 895	
10	134 900	117,8	52 580	
11	145 099	100,7	65 392	
12	155 990	172,9	27 827	
13	156 003	95,6	72 058	
14	171 942	98,8	80 669	
15	190 000	105,5	44 880	
16	193 990	99,9	69 520	
17	251 222	76,8	98 643	
18	258 964	95,2	75 587	
19	264 309	119,7	83 475	
20	314 593	125,5	91 696	

Рис. 2.3.5. Дані про обсяги продажу, витрати на рекламу та ціни на товари

Моделі парної кореляції між обсягами продажу цих товарів і їх цінами наведено на рис. 2.3.3. При виборі лінійної моделі маємо регресійне рівняння $y = –744{,}5x + 141432$ з коефіцієнтом детермінації 0,5763. Слід звернути увагу на те, що коефіцієнт регресії від'ємний, а лінія тренда йде з верхнього лівого у нижній правий кут діаграми. Тобто маємо звичайну залежність обсягів продажу від ціни продукції: чим вища ціна, тим менша кількість продукції продається.

Найпростіший спосіб визначити в Excel залежність результатної ознаки від декількох факторів — використати інструмент *Регресія*, який повертає всю потрібну інформацію, згруповану у декілька таблиць.

Для отримання цієї інформації слід активізувати робочий аркуш з даними і виконати такі дії:

1. Виконати команду *Сервіс*/*Аналіз даних*.
2. У діалоговому вікні *Аналіз даних* зі списку *Інструменти аналізу* вибрати інструмент *Регресія* і натиснути кнопку *ОК*.
3. Після появи діалогового вікна *Регресія* потрібно:

1) у текстовому полі *Вхідний інтервал Y* встановити діапазон *C2:C20* (увести з клавіатури або виділити мишею ці клітини у робочому аркуші);

2) у текстовому полі *Вхідний інтервал X* встановити діапазон *A2:A20*;

3) у поле *Рівень надійності* ввести число 95 (якщо воно там не стоїть);

4) перемикач *Параметр виведення* встановити в положення *Новий робочий аркуш*;

5) клацнути по кнопці *ОК*.

Результати (рис. 2.3.6), отримані з допомогою інструмента *Регресія*, містять всю потрібну інформацію. Так, у клітині *B5* знаходиться значення параметра R^2, а в *B4* — значення множинного *R*, яке являє собою квадратний корінь з дисперсії (R^2). Це значення (0,9085) є коефіцієнтом кореляції і виражає кореляцію між кількістю проданої продукції і отриманою комбінацією незалежних змінних. Воно означає, що приблизно 91 % міри мінливості кількості продажу продукції зв'язано з мірою мінливості ціни, за якою продається ця продукція, і розмірами витрат на її рекламу.

Значення R^2 двофакторної моделі суттєво більше, ніж відповідних однофакторних. Це свідчить про те, що остання модель набагато краще за попередні пояснює зміни результативної ознаки

(обсяги продажу). Проте ще треба з'ясувати, чи не є цей результат випадковим. Припустимо, що в дійсності ніякого взаємозв'язку змінної *y* та змінних *x* немає. Величину ймовірності помилковості твердження про те, що є значний взаємозв'язок між змінними, приймемо рівною 0,05. Для ступенів вільності маємо: $df_1 = 2$ (кількість факторів), $df_2 = 16$ (значення клітини *B13*). У будь-якому статистичному довіднику можна знайти, що *F*-критичне (для вказаних величин) дорівнює 3,36. Спостережуване *F*-значення більше 79 (клітина *E12*), що значно більше за *F*-критичне значення 3,36. Отже, припущення про відсутність взаємозв'язку залежної та незалежних змінних не підтверджується.

	A	B	C	D	E	F	G
1	**ВЫВОД ИТОГОВ**						
2							
3	Регрессионная статистика						
4	Множественный R	0,95317					
5	R-квадрат	0,90854					
6	Нормированный R-квадрат	0,8971					
7	Стандартная ошибка	9941,79					
8	Наблюдения	19					
9							
10	***Дисперсионный анализ***						
11		*df*	*SS*	*MS*	*F*	Значимость *F*	
12	Регрессия	2	1,57E+10	7,85E+09	79,4666	4,898E-09	
13	Остаток	16	1,58E+09	98839123			
14	Итого	18	1,73E+10				
15							
16		Коэффициенты	Стандартная ошибка	*t*-статистика	*P*-значение	Нижние 95 %	Верхние 95 %
17	*Y*-пересечение	61304,1	14182,96	4,355678	0,000490366	31237,591	69924,2
18	Переменная *X* 1	0,24181	0,031717	7,623928	1,0298E-06	0,1745723	0,33337
19	Переменная *X* 2	–383,28	87,99497	–4,3224	0,00052558	–569,8207	–52,2293

Рис. 2.3.6. Інформація, видана інструментом *Регресія*

Нормований R^2, що знаходиться у клітині *B6*, ураховує кількість результатів спостережень і незалежних змінних. Якщо кількість

спостережень відносно кількості незалежних змінних не досить велика, R^2 має тенденцію відхилятися в бік підвищення. Нормований R^2 забезпечує інформацією про те, яке значення могло б бути отримано в іншому наборі даних, значно більшому за аналізований. Якщо б розглядуваний приклад базувався на значно більшій кількості спостережень, то нормований R^2 і фактичний R^2не дуже різнилися б.

Діапазон *A17:C19* містить детальну інформацію щодо членів регресії — постійної *b* (*Y-пересечение*) та коефіцієнтів регресії — та їх стандартних похибок.

У колонці *t-статистика* знаходяться стандартизовані (нормовані) змінні, які представляють частку кожного члена рівняння в його стандартній похибці.

У колонці *P-значення* розташовано результати обчислень, які дають змогу перевірити, чи є отримані значення коефіцієнтів регресії дійсно корисними у разі оцінювання з їх допомогою кількості продажу. Ці результати уможливлюють висновок, що у даному разі за умови, що відповідний коефіцієнт реально має нульове значення, ймовірність отримати значення

$|m_2|$ = 383,28 не більша 0,0006 (або 6 шансів з 10 000),

m_1 = 0,2418 приблизно 0,000001 (або 1 шанс з 1 000 000),

b = 61 304,11 не більша 0,0005 (або 5 шансів з 10 000).

А це підтверджує статистичну значущість отриманих коефіцієнтів регресії.

Останні колонки третього розділу результатів вміщують нижню і верхню границі 95-процентного рівня надійності як для постійної, так і для кожного коефіцієнта регресії. Тут треба звернути увагу на те, що жоден з трьох довірчих інтервалів не охоплює нульове значення. Це саме той результат, який і треба було очікувати, оскільки всі *P*-значення, що знаходяться вище 5-процентного рівня, є значущими. Якщо б *P*-значення дорівнювало 0,05 або більше, довірчий інтервал цього показника включав би нуль.

Отже, можна з 95-процентною впевненістю стверджувати, що всі показники регресії не є нульовими. З цього випливає, що незалежні змінні додають до рівняння регресії значущу інформацію і на основі даних про витрати на рекламу продукції та ціни, за якими вона продається, можна досить точно прогнозувати обсяги продажу.

3.4. ВИБІРКОВИЙ МЕТОД ЗА ВИЗНАЧЕННЯ ПОПИТУ

Опитування та інші види спостережень, застосовувані у маркетинговому дослідженні, можуть бути *суцільними* або *вибірко-*

вими. Суцільне спостереження переважно обмежується рамками фірми і використовується порівняно нечасто. Основним способом отримання даних, особливо про споживачів, є вибірковий метод. *Вибірковий метод* — це метод статистичного дослідження, за якого узагальнюючі показники сукупності, що вивчається, встановлюються за деякою її частиною, як правило, на основі положень випадкового відбору. За такого методу обстеженню підлягає порівняно невелика частина всієї досліджуваної сукупності (до 5—10 %, зрідка до 15—25 %). При цьому статистична сукупність, яка підлягає вивченню і з якої здійснюється відбір частини одиниць, називається генеральною сукупністю. Відібрана з генеральної сукупності деяка частина одиниць, що підлягає обстеженню, називається вибірковою сукупністю, або вибіркою. Прикладом використання вибіркових спостережень є оцінка можливості проникнення на ринок з допомогою вибірки з фірм-виробників.

Переваги вибіркового методу полягають у тому, що його використання дає змогу краще організувати спостереження, забезпечує проведення дослідження у стисліші терміни, з мінімальними затратами праці і витратами коштів. За дотримування умов випадковості і досить великої кількості спостережень він дає змогу, використовуючи отримані дані, з достатньою для практики точністю робити висновки про характеристики генеральної сукупності. Проте отримані з матеріалів вибіркового спостереження статистичні показники звичайно не збігаються з відповідними характеристиками генеральної сукупності. Це відхилення називається *помилкою спостереження* і складається з двох частин: помилки репрезентативності (представництва) та помилки вимірювання (реєстрації).

Помилка репрезентативності характеризує розмір розбіжностей між величинами показників, отриманих у вибірковій і генеральній сукупності за умови однакової точності поодиноких спостережень. Вона властива саме вибірковим спостереженням і виникає, коли вибірка нерепрезентативна, тобто не представляє генеральну сукупність у потрібному аспекті. Така помилка має три складові. Це:

— систематична (тенденційна) помилка — виникає через недосконалість або порушення правил формування вибіркової сукупності і призводить до зсуву результатів обстеження (незсуненість — одна з вимог вибіркового обстеження, яка досягається правильною організацією його);

—випадкова помилка — пов'язана з недостатньо рівномірним представленням у вибірковій сукупності всіх категорій елементів генеральної сукупності. Уникнути таких похибок за вибіркового обстеження принципово неможливо, але теорія вибіркового методу базується на математичній основі, яка дає змогу обчислити й регулювати їх розмір;

—помилка, що виникає через недоступність окремих елементів вибірки або у разі відмови відповідати на запитання, яке треба враховувати під час спостережень.

Виникнення помилок вимірювання пояснюється тим, що вимірюється не зовсім те, що було потрібно (так, людина може збрехати, відповідаючи на певне запитання). Такі помилки властиві як суцільному, так і вибірковому спостереженню. Вони пов'язані з неправильною організацією спостереження, неправильно обраною технікою вимірювання, недосконалістю вимірювальних приладів, недостатньою кваліфікацією спостерігачів, неточністю підрахунків тощо. Слід зазначити, що за вибіркового спостереження помилка вимірювання звичайно значно менша, ніж за суцільного, оскільки техніка вимірювання може бути розроблена і здійснена ретельніше за рахунок використання більш кваліфікованих і підготовлених кадрів. Тому навіть коли час і можливості дозволяють виконати повне обстеження всіх елементів сукупності, перевагу може бути віддано отриманню інформації з допомогою вибірки саме з метою підвищення точності результатів. Але цього можна досягти лише у разі додержання правил наукової організації вибіркового дослідження.

Найбільш поширений спосіб відбору одиниць сукупності для дослідження базується на принципі однакових можливостей потрапляння у вибірку кожної одиниці генеральної сукупності. Завдяки цьому виключається створення вибіркової сукупності тільки за рахунок одного типу елементів. Це дає змогу уникнути систематичних помилок і здійснювати кількісне оцінювання помилки репрезентативності.

За вибіркового методу використовуються переважно два основних види узагальнюючих показників: відносна величина альтернативної ознаки та середня величина кількісної ознаки.

Відносна величина альтернативної ознаки характеризує частку (питому вагу) одиниць у статистичній сукупності, які відрізняються від усіх інших одиниць цієї сукупності тільки наявністю досліджуваної ознаки.

Середня величина кількісної ознаки — це узагальнена характеристика ознаки, яка має різні значення в окремих одиницях статистичної сукупності.

У генеральній сукупності частка одиниць, що мають досліджувану ознаку, називається *генеральною часткою*, а середня величина цієї ознаки — *генеральною середньою*. У вибірковій сукупності частку досліджуваної ознаки називають *вибірковою часткою*, або *частістю*, а її середню величину — вибірковою середньою.

Основне завдання вибіркового дослідження полягає у тому, щоб на основі характеристик вибіркової сукупності (частості або середньої) отримати з певною вірогідністю висновки про частку або середню генеральної сукупності.

Як правило, організація вибіркового обстеження складається з таких елементів:

— визначення цільової величини у вигляді запланованого вимірювання цільової частини генеральної сукупності (наприклад, частки домогосподарок певного регіону, що користуються пральними машинами);

— вибір інформаційної основи вибіркового спостереження (наприклад, певні статистичні матеріали), визначення структури вибірки (наприклад, кількість людей певного віку з певним рівнем прибутку) та способу відбору одиниць з генеральної сукупності;

— визначення способів (одного чи більше) отримання інформації для визначення цільової величини (наприклад, спостереження або відповіді на запитання);

— визначення методу аналізу результатів вибіркового спостереження та оцінювання точності дослідження.

Необхідною умовою організації вибіркового спостереження є попереднє вивчення генеральної сукупності, оцінювання її однорідності, поділ за головними ознаками та визначення необхідної кількості спостережень. Результати вибіркового обстеження відображаються у термінах імовірності настання деякої події із зазначенням (оскільки усі вибіркові методи пов’язані з похибками) деякого рівня вірогідності того, що отриманий результат є правильним і знаходиться у прийнятних межах.

Способи відбору одиниць з генеральної сукупності

У статистиці залежно від завдань дослідження та специфіки об'єкта, що вивчається, застосовуються різні способи формування вибіркових сукупностей.

Головною умовою здійснення вибіркового обстеження є уникнення систематичних (тенденційних) похибок. Вони виникають у разі невиконання принципів рівних можливостей потрапляння у вибірку для кожної одиниці генеральної сукупності.

Способи відбору визначаються правилами формування вибіркової сукупності. Найчастіше використовуються такі вибірки: власне-випадкові, або прості випадкові; механічні; типові (розшаровані, районовані); територіальні та цільові.

Власне-випадкова вибірка полягає у тому, що вибіркова сукупність створюється в результаті випадкового відбору окремих одиниць з генеральної сукупності. Для добору елементів сукупності проводять жеребкування або використовують псевдовипадкові числа. Реалізація цього способу потребує попередньої підготовки до формування вибірки (наприклад, нумерації кожної одиниці генеральної сукупності). У разі великих за обсягом сукупностей ручне її проведення може бути досить трудомістким.

Власне-випадкова вибірка може здійснюватися за схемою повторного або безповторного відбору. Вибір схеми відбору залежить від характеру досліджуваного об'єкта. У разі *безповторного відбору* чисельність генеральної сукупності у процесі вибірки скорочується. За *повторного відбору* кожна одиниця, яка потрапила у вибірку, після її дослідження має повернутися у генеральну сукупність, де їй надається така сама можливість знову потрапити у вибірку. Так, у разі вивчення споживацького попиту населення не виключена повторна реєстрація незадоволеного попиту того самого покупця у декількох магазинах міста.

За *механічної вибірки* генеральна сукупність механічно поділяється на рівновеликі групи і з кожної з них лише один елемент потрапляє у вибірку. Кількість елементів у кожній групі дорівнює n/N, де n — обсяг вибірки, а N — обсяг генеральної сукупності. Якщо елементи генеральної сукупності впорядковано за суттєвою ознакою, тобто ознакою, яка повністю визначає поведінку досліджуваного показника, то у вибірку має відбиратись елемент, який знаходиться всередині кожної групи (це дає змогу уникнути систематичної помилки вибірки). Якщо ж елементи генеральної сукупності впорядковано за нейтральною ознакою, то з першої групи можна взяти будь-який елемент, а з інших добираються ті, що відповідають порядковому номеру елемента, відібраного з першої групи. Доведено, що механічна вибірка за точністю ре-

зультату близька до власне-випадкової, а здійснити її (у разі неавтоматизованого добору елементів) простіше.

Якщо стикаються з досить неоднорідною інформацією (прикладом такої неоднорідності може слугувати неоднорідність населення), то використовують *типову вибірку*, яка передбачає попередню структуризацію генеральної сукупності. За такої вибірки генеральна сукупність спочатку поділяється на однорідні типові групи, а потім з кожної групи проводиться незалежний індивідуальний відбір елементів у вибіркову сукупність.

Важливою особливістю типової вибірки є те, що вона може дати точніші порівняно з іншими способами відбору одиниць у вибіркову сукупність результати. Оскільки похибка типової вибірки визначається середньою з групових дисперсій, то репрезентативність такої вибірки забезпечується поділом генеральної сукупності на якісно однорідні групи. Якщо групи об'єднують однорідні елементи, а групові середні помітно різні, варіація ознаки в групах буде значно меншою, ніж в цілому по сукупності. У такому разі середня з групових дисперсій буде меншою за дисперсію по сукупності, а отже, й похибка типової вибірки порівняно з власне-випадковою буде менша. Забезпечити більшу точність типової вибірки можна обґрунтованим вибором ознаки поділення генеральної сукупності, кількості груп, обсягів кожної з них і способів відбору. Зменшення варіації ознаки за поділу сукупності можливе лише у тому разі, коли ознака поділення корелює з ознакою, характеристики якої оцінюються. Чим щільніший зв'язок між ознаками, тим помітніше зменшення похибки.

Якісно однорідні групи за типової вибірки можуть утворюватися як в результаті спеціально проведеного типового групування одиниць генеральної сукупності, так і в результаті використання тих, що вже є, у тому числі й тих, що склалися природно. Так, у разі вивчення споживацького попиту на певній території магазини, що продають товар, попит на який досліджується, можуть групуватися за їх типом (універмаги, магазини культтоварів та ін.).

У більшості випадків використовуються типові вибірки з неоднаковою кількістю елементів. Проте з кожної типової групи можна відібрати кількість одиниць, пропорційну їх чисельності, тобто використовувати пропорційний типовий відбір.

За визначення статистичних показників типової вибірки не можна застосовувати відповідні стандартні функції Excel. Це пов'язане з тим, що такі функції призначені для обчислення показників вибірки, всі елементи якої входять до однієї групи, а в

типовій вибірці треба обчислювати статистичні показники по варіаційному ряду, в якому дані об'єднано (згруповано) за значенням ознаки та підраховано кількість випадків повторення кожного з них. Тому середня вибірки та дисперсія середньої розраховуються як зважені показники за такими формулами:

$$x_{\text{сер}} = \Sigma\, x_i\, f_i / \Sigma\, f_i\,,$$

$$\sigma^2 = \Sigma\, (x_i - x_{\text{сер}})^2\, f_i / \Sigma\, f_i\,,$$

де x_i — значення ознаки в i-й групі;

f_i — кількість елементів, що входять до цієї групи.

Якщо в рамках виділеного бюджету неможливо точно визначити склад певної групи (наприклад, у випадку, коли це потребує проведення суцільної вибірки), то використовують *територіальну вибірку*. Основною ідеєю її є те, що елементи вибірки можуть бути ідентифіковані у межах певного району й можна скласти список цих районів. У маркетингових дослідженнях методи територіальної вибірки найчастіше застосовуються в опитуванні домогосподарств. Часто така вибірка є єдиним способом отримання ймовірної вибірки на великій території з недостатньо визначеними елементами. Може застосовуватися й техніка *«зосередження»*, що полягає у створенні невеликих осередків проведення вибіркових досліджень. Це має місце у пробному маркетингу (наприклад, коли необхідно оцінити можливі обсяги продажу у регіональному масштабі за запуску у виробництво нового продукту або нової маркетингової програми). Вплив техніки «зосередження» на похибку вибірки можна оцінити лише тоді, коли відомі кореляційні залежності між елементами кожного осередку. Оскільки у пробному маркетингу дуже складно провести кореляцію всередині осередку, то й неможливо визначити ступінь точності проведених досліджень.

Цільова вибірка полягає у систематичному відборі елементів з метою залучення до дослідження достатньої кількості елементів кожного основного типу. Але використання результатів такої вибірки обмежується неможливістю оцінити помилку вибірки в якийсь об'єктивний спосіб. До неї вдаються за вивчення реакції ринку на новий виріб або на модернізацію старого, коли ймовірнісна вибірка потребує великих витрат. При цьому робиться припущення, що смаки споживачів більш-менш ідентичні, принаймні, всередині однієї групи.

Помилка вибірки

Після проведення певної кількості спостережень отримують розподілення результатів (вибіркових оцінок) того самого істинного рівня (наприклад, низки характеристик населення). Це вибіркове розподілення підлягає законові нормального розподілення, якщо вибірка достатньо велика. Оскільки істинний рівень може не збігатися з рівнем вибіркових характеристик, необхідно брати до уваги похибку вибірки. У цьому разі можна знайти ступінь вірогідності вибіркових характеристик.

У математичній статистиці значення *середньої похибки* визначається за формулою

$$\mu \approx \sqrt{\frac{k \cdot \sigma^2}{n}},$$

де σ^2 — дисперсія вибіркової сукупності;

n — чисельність одиниць вибіркової сукупності;

k — коефіцієнт, який для повторного відбору дорівнює одиниці, а для безповторного — $1 - n/N$, де N — чисельність генеральної сукупності.

Середня похибка вибірки використовується для визначення межі відхилень характеристик вибірки від характеристик генеральної сукупності. Суттєвим є твердження, що ці відхилення не будуть більші за значення, яке в статистиці називається *граничною помилкою вибірки*, лише з певним ступенем імовірності.

Гранична помилка вибірки пов'язана із середньою похибкою вибірки співвідношенням

$$\Delta = \mu \cdot t,$$

де t — коефіцієнт кратності помилки.

Значення коефіцієнта кратності помилки залежить від того, з якою довірчою ймовірністю (надійністю) слід гарантувати результати вибіркового обстеження. Для його визначення користуються таблицею значень інтеграла ймовірностей нормального закону розподілення. В економічних дослідженнях звичайно обмежуються значеннями t, що не перевищують двох-трьох одиниць:

Кратність помилки	Імовірність (надійність)	Кратність помилки	Імовірність (надійність)	Кратність помилки	Імовірність (надійність)

0,1	0,0797	1,5	0,8664	2,6	0,9907
0,5	0,3829	2,0	0,9545	3,0	0,9973
1,0	0,6827	2,5	0,9876	4,0	0,999937

При цьому вибір тієї чи іншої довірчої імовірності залежить від того, з яким ступенем вірогідності слід гарантувати результати вибіркового обстеження (найчастіше спираються на ймовірність 0,9545, за якої t дорівнює 2).

Якщо в формулу для визначення Δ підставити конкретний вміст μ, то для обчислення граничної помилки можна буде використати такі вирази:

— у разі альтернативної ознаки

$$\Delta_w \approx t\sqrt{\frac{k \cdot w \cdot (1-w)}{n}},$$

де w — вибіркова частка, яка визначається з відношення одиниць, які мають досліджувану ознаку, до загальної чисельності одиниць вибіркової сукупності;

— у разі кількісної ознаки

$$\Delta_x \approx t\sqrt{\frac{k \cdot \sigma^2}{n}},$$

де σ^2 — дисперсія кількісної ознаки у вибірці.

Визначення розміру вибірки

У разі організації вибіркових досліджень важливо визначити, наскільки великим має бути обсяг вибірки. Для загальної відповіді на це питання слід знати:

— витрати на проведення вибіркового дослідження;

— витрати на отримання наближених оцінок;

— ступінь мінливості процесу;

— ступінь надійності результатів, необхідний для прийняття подальших рішень.

Обминаючи вартісні фактори, розмір оптимальної вибірки можна визначити, базуючись на формулі граничної похибки. Приміром, за безповторного відбору для середньої кількісної ознаки необхідна чисельність обчислюється так:

$$n = \frac{t^2 \sigma^2 N}{\Delta^2 N + t^2 \sigma^2},$$

де n — чисельність одиниць вибірки;

N — обсяг генеральної сукупності;

t — коефіцієнт кратності помилки (або коефіцієнт довіри);

σ^2 — дисперсія;

Δ — гранична (задана) помилка середньої (звичайно вибирається рівною 10 % від значення середньої).

Нехай для обстеження, що має на меті виявити потреби у певному товарі тривалого використання в регіоні, де мешкає 10 тис. сімей, необхідно провести анкетування.

Умовно приймаємо, що в кожній квартирі проживає одна сім'я і на неї виділяється одна анкета. Припустимо, що попередніми дослідженнями встановлено, що середній розмір покупки та дисперсія становлять відповідно 17 і 150 грн. Виходячи з того, що гранична помилка не повинна перевищувати 10 % від середньої і що результати обстеження необхідно гарантувати з довірчою імовірністю, не меншою 0,954, чисельність вибірки має становити

$$n = \frac{2^2 \times 150 \times 10\,000}{(0{,}1 \times 17)^2 \times 10\,000 + 2^2 \times 150} = 203{,}4.$$

Ясна річ, деяка частина анкет не повертається (припустимо, практика показує, що приблизно кожна п'ята), тому треба збільшити кількість анкет до 255. Отже, можемо зробити висновок, що необхідно включити у вибірку щонайменше кожну 40-у квартиру.

Технологію визначення розміру вибірки розглянемо на прикладі вибору магазинів для вивчення на деякій території споживацького попиту на певний товар.

Для цього на окремому робочому листі Excel слід створити список усіх магазинів, що торгують товаром, попит на який вивчається (рис. 2.3.7).

Заголовок списку мусить мати такі поля: номер магазину, тип (наприклад, універмаг і культтовари), місцезнаходження, загальний товарообіг, товарообіг по товару, частка продажу товару в товарообігу магазину.

	A	B	C	D	E	F
1	ВИЗНАЧЕННЯ ОБСЯГУ ВИБІРКИ					

2	Номер магазину	Тип магазину	Адреса	Загальний товарообіг	Товарообіг по товару	Частка продажу товару (%)
3						

Рис. 2.3.7. Оформлення заголовка списку магазинів

У цей список щодо кожного магазину заноситься інформація у перші п'ять стовпців (*A, B, C, D, E*). У клітину *F3* заводиться формула *=D3/E3*100* і копіюється на всі рядки списку.

Оскільки типова вибірка дає найточніші, порівняно з іншими способами відбору одиниць у вибіркову сукупність, результати, то відбір конкретних магазинів бажано проводити окремо для кожного типу магазинів. Тому список магазинів треба відразу впорядкувати за типом магазинів, а всередині цієї впорядкованості — за часткою продажу товару (команда *Дані*/*Сортування*).

Нехай повний список магазинів, що торгують товаром, попит на який вивчається, займає 52 рядки робочого аркуша. У такому разі в клітину *C53* треба завести формулу *=СЧЕТ(C3:C52)*, а у клітину *F53* — формулу *=ДИСПР(F3:F52)*. Перша формула дасть змогу обчислити загальну кількість магазинів, а друга — дисперсію розподілення частки продажу товару в цій сукупності магазинів. Слід звернути увагу на те, що використання функції *ДИСПР* передбачає, що її параметри представляють усю генеральну сукупність. Якщо дані представляють тільки вибірку з генеральної сукупності, то дисперсію слід обчислювати, використовуючи функцію *ДИСП*.

Для полегшення подальшого використання у формулах значень клітин *C53* і *E53* (кількості магазинів і дисперсії), цим клітинам і робочому аркушу слід надати імена (наприклад, *N*, *D* і *СписокМ* відповідно). Щоб дати ім'я клітині, можна завести курсор у цю клітину, клацнути мишею в полі імені, набрати там нове ім'я й обов'язково натиснути клавішу *Enter*. Можна також надавати імена клітинам, використовуючи діалогове вікно *Надати ім'я*. Для цього необхідно виконати таку послідовність дій:

— розташувати курсор в клітині, якій треба надати ім'я;

— вибрати команду *Вставка*/*Ім'я*/*Надати;*

— у діалоговому вікні *Надати ім'я*, що з'явиться після цього, набрати нове ім'я у текстовому полі *Ім'я*;

— натиснути на кнопку *Додати*, а потім — на кнопку *ОК*.

Для того щоб надати нове ім'я робочому аркушу, треба клацнути правою кнопкою миші на ярлику відповідного робочого ар-

куша, у контекстному меню вибрати команду *Перейменувати*, ввести потрібне ім'я і натиснути клавішу *Enter*.

Для проведення розрахунків на окремому робочому аркуші створюється таблиця, яка заповнюється відповідно до рис. 2.3.8.

Коефіцієнт граничної помилки вибираємо виходячи з 10 % рівня помилки середньої від її значення. Сьомий та восьмий рядки робочої таблиці показують межі вибірки, дотримання яких з вибраною ймовірністю гарантуватиме вірогідність результатів вибіркового обстеження. Останній показник у цій таблиці — це частка магазинів (*n*/*N*), що мають потрапити у вибірку.

	A	B	C	D
1	ВИЗНАЧЕННЯ ОБСЯГУ ВИБІРКИ			
2	Середня вибірки (середня частка продажу)	= СРЗНАЧ (СписокМ! F3:F52)		
3	Гранична помилка середньої	= 0,1*B2		
4	Довірча ймовірність (надійність)	0,9545	0,9876	0,9973
5	Коефіцієнт кратності помилки (*t*)	2	2,5	3
6	Обсяг вибірки (*n*)	= B4^2*D*N / ($B3^2 *N + B4^2*D)	= C4^2*D*N / ($B3^2*N + C4^2*D)	= D4^2*D*N / ($B3^2*N + D4^2*D)
7	Нижня межа вибірки	= B6-$B3	= C6-$B3	=D6-$B3
8	Верхня межа вибірки	= B6+$B3	= C6+$B3	=D6+$B3
9	Частка вибірки	= B6/N	= C4/N	=D6/N

Рис. 2.3.8. Формули у робочому аркуші для обчислення обсягу вибірки

На рис. 2.3.9 наведено результати обчислень за середньої 16 і трьох значень коефіцієнта довірчої ймовірності (0,9545; 0,9876; 0,9973). Після аналізу отриманих результатів слід вибрати значення довірчої ймовірності та надати відповідній клітині дев'ятого рядка робочого аркуша ім'я *ЧасткаВ*.

	A	B	C	D
1	ВИЗНАЧЕННЯ ОБСЯГУ ВИБІРКИ			
2	Середня вибірки (середня частка продажу)	16		

3	Гранична помилка середньої	1,6		
4	Довірча ймовірність	0,9545	0,9876	0,9973
5	Коефіцієнт кратності помилки (*t*)	2	2,5	3
6	Обсяг вибірки (*n*)	33,74	38,21	41,18
7	Нижня межа вибірки	32,14	36,61	39,58
8	Верхня межа вибірки	35,34	39,81	42,78
9	Частка `вибірки`	0,67	0,76	0,82

Рис. 2.3.9. Результати обчислення обсягу вибірки

Найпростіший спосіб підрахувати кількість магазинів кожного типу полягає у створенні на новому робочому аркуші зведеної таблиці, для чого слід виконати такі дії:

1. Установити курсор на будь-якій клітині списку магазинів.
2. Вибрати з меню команду *Дані*/*Зведена таблиця*.
3. На першому кроці *Майстра зведених таблиць* вибрати режим «*у списку або базі даних Microsoft Excel*».
4. На другому кроці просто клацнути по кнопці *Далі*.
5. На третьому кроці:

— перетягнути поле *Тип магазину* в область діаграми *Колонка*;

— ще раз перетягнути поле *Тип магазину* в область діаграми *Дані* і подвійно клацнути по ньому лівою кнопкою миші;

— у діалоговому вікні *Обчислення поля зведеної таблиці* вибрати операцію *Кількість значень* і клацнути по кнопці *ОК*;

— завершити формування зведеної таблиці натисканням кнопки *Готово*.

Таблиця, створювана *Майстром зведених таблиць* (рис. 2.3.10), займатиме лише перші три рядки (на рисунку їх подано жирним курсивом).

У четвертий рядок, безпосередньо за останнім рядком зведеної таблиці з метою отримання для кожного типу магазину такого розміру вибірки, щоб вона була пропорційна чисельності даного типу магазину, в клітини *B4* та *C4* вводяться формули, що обчислюють добуток частки вибірки (*ЧасткаВ*), яка відповідає вибраній довірчій імовірності (обчислюється у клітині дев'ятого рядка таблиці на рис. 2.3.8) і чисельності відповідного типу магазину.

	A	B	C	D
1	**Кількість значень по полю Тип магазину**	Тип магазину		

2		**Культтовари**	**Універмаг**	**Підсумок**
3	**Всього**	**41**	**11**	**52**
4	Кількість магазинів у вибірці	= ОКРУГЛ (ЧасткаВ*В3)	= ОКРУГЛ (ЧасткаВ*С3)	
5	Генератори псевдовипадкових чисел	= 1 + ЦІЛЕ (В3* СЛЧИСЛ())		
6				

Рис. 2.3.10. Вигляд зведеної таблиці (перші три рядки) списку магазинів

Найпростіший спосіб відбору одиниць у вибіркову сукупність — з допомогою псевдовипадкових чисел. Саме цей спосіб доцільно застосувати для включення у вибірку конкретних магазинів, тобто для визначення опорних магазинів з вивчення споживацького попиту.

Техніка використання псевдовипадкових чисел може бути такою:

— рядки таблиці (рис. 2.3.10) закріплюються на екрані. Для цього табличний курсор розміщується у клітині *D6* і виконується команда *Вікно*/*Закріпити ділянки*;

— у клітину *B5* уводиться формула, яка завдяки використанню функції *СЛЧИС( )* виконуватиме роль генератора випадкових чисел у діапазоні від 1 до кількості магазинів відповідного типу, тобто роль генератора випадкових порядкових номерів магазинів відповідного типу у їх списку. Функція *СЛЧИС( )* повертає рівномірно розподілене випадкове число, що більше або дорівнює 0 і менше 1. Нове випадкове число повертається цією функцією кожного разу, коли перераховується робочий аркуш;

— після введення формули і натискання клавіші *Enter* у клітині *B5* буде відображено порядковий номер магазину, який може виконувати роль опорного магазину типу «Культтовари». Далі слід змістити табличний курсор у клітину *B6*, занести туди цей номер і натиснути клавішу *Enter* (а краще клавішу ↓). У клітині *B5* з'явиться порядковий номер наступного магазину. Кількість повторів цієї операції визначається значенням клітини *B4*;

— після копіювання формули з клітини *B5* у клітину *C5* аналогічно визначаються опорні магазини типу «Універмаг».

Оброблення результатів опитування

Припустимо, що проведено вибіркове анкетування 10 000 сімей з метою отримати відповіді на такі запитання:

— Наскільки населення регіону забезпечене певним виробом?

— Який середній вік експлуатації цього виробу?

— Яке зношення (середній вік) виробів, що знаходяться в експлуатації?

Нехай з 255 розісланих анкет повернулося 208, з яких виявлено, що 183 сім'ї вже мають зазначений виріб. На викладене в анкеті прохання вказати термін експлуатації виробу отримано такі відповіді: вироби, що використовувалися до 3 років, має 21 сім'я, від 4 до 6 років — мають 47 сімей, від 7 до 9 років — 96 сімей, від 10 років і більше — 19 сімей. Бажання про заміну виробу на сучасніший висловили 137 сімей (рис. 2.3.11).

Ступінь забезпеченості сімей товаром визначається як відношення кількості сімей, що використовують виріб, до загальної кількості отриманих відповідей.

Для визначення середнього «віку» виробів, що має населення, і дисперсії середньої використовуються такі формули:

$$x_{\text{сер}} = \sum xf / \sum f,$$

$$\sigma^2 = \sum (x - x_{\text{сер}})^2 f / \sum f.$$

Термін експлуатації виробу	Середнє значення терміну експлуатації (x)	Кількість сімей, що мають виріб (f)	Загальний час експлуатації ($x \cdot f$)	$(x - x_{\text{сер}})^2 \cdot f$
До 3 років	2	21	42	516
Від 4 до 6 років	5	47	235	180
Від 7 до 9 років	8	96	768	105
Від 10 років і більше	12	19	228	483
Разом		183	1273	1283,65
Кількість сімей, N	10 000			
Відправлено анкет	255			
Отримано відповідей, n	208			
Кількість сімей, що мають виріб	183			

Кількість сімей, що планують заміну виробу	137			
Ступінь забезпеченості товаром, w	88 %			
Середній «вік» виробів, що знаходяться в експлуатації, $x_{сер}$	7			
Середній «вік» спрацьованості виробів, що знаходяться в експлуатації	9,29			
Дисперсія середнього віку виробів, що знаходяться в експлуатації	7,01			
Коефіцієнт кратності помилки, t	2			
Гранична помилка забезпеченості товаром	4,5 %			
Гранична помилка середнього віку виробів, що знаходяться в експлуатації	0,36			

Рис. 2.3.11. Оброблення результатів опитування

Гранична помилка ступеня забезпеченості сімей товаром обчислюється за формулою для альтернативної ознаки:

$$\Delta_w = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)},$$

де w — ступінь забезпеченості сімей товаром (клітина *C14*);

n — обсяг вибірки.

Гранична помилка середнього «віку» товару обчислюється (у клітині *C18*) за формулою для середнього значення кількісної ознаки:

$$\Delta_x = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}.$$

Отримані результати дозволяють стверджувати:

— забезпеченість населення товаром, що вивчається, з імовірністю 0,954 у межах інтервалу від 83,5 % до 92,5 % (88 ± ± 4,5);

— середній «вік» експлуатованих населенням виробів в інтервалі від 6,64 до 7,36 року (7 ± 0,36);

— середній термін зношення виробу в інтервалі від 8,34 до 9,06 років (8,7 ± 0,36).

3.5. РЕАЛІЗАЦІЯ ОПТИМІЗАЦІЙНИХ МОДЕЛЕЙ

Оптимізаційні моделі дають змогу вирішувати широке коло питань планування економічних процесів, знаходити найкращі (оптимальні) рішення, що відповідають певним обмеженням. Методи вибору з низки альтернативних рішень найсприятливішого (з найменшими витратами, максимальним прибутком і т. ін. за однакових інших умов) використовуються для вирішення багатьох маркетингових проблем: пошуку найвигіднішого асортименту за обмежених ресурсів; розрахунку оптимальної величини товарних запасів; планування маршрутів руху агентів зі збуту та ін. Саме тому програма *Пошук рішення*, яка є в Excel, може стати потужним допоміжним інструментом у маркетинговій діяльності. Користувач може задати режим, за якого отримані значення змінних автоматично заноситимуться у таблицю і, крім того, представити результати роботи програми у вигляді звітів.

Етапи роботи з прийняття оптимальних рішень

Розв'язання задачі з допомогою програми *Пошук рішення* складається з таких етапів:

— постановка задачі і створення математичної моделі;

— запис задачі у табличній формі, придатній для введення даних;

— уведення даних у комп'ютер і розв'язання задачі;

— економіко-математичний аналіз отриманого рішення.

За моделювання реальної задачі робота з моделлю на цьому не закінчується. Економіко-математична модель лише з деяким наближенням відбиває реальні обставини. Тому після аналізу отриманого рішення ці обставини уточнюються, постановка задачі і модель належним чином коригуються, введені дані змінюються й програма знову запускається на виконання. Цей ітераційний процес триває, допоки не буде отримано модель, найбільш адекватну для вирішення поставленої проблеми.

Задача, розв'язувана програмою *Пошук рішення*, у найзагальнішому вигляді формулюється так:

знайти вектор $\mathbf{x} = (x_1, x_2, ..., x_n)$, який
мінімізує (або максимізує) функцію $\varphi(\mathbf{x})$ (1)
за обмежень
$X_{k\,\min} \le x_k \ge X_{k\,\max}$, $k = 1, ..., n$; (2)
$F_{i\,\min} \le g_i(\boldsymbol{x}) \le F_{i\,\max}$, $i = 1, ..., m$; (3)
$h_i(\boldsymbol{x}) = F_j$, $j = 1, ..., p$. (4)

Функція φ (***x***) називається цільовою функцією, умови (2) — прямими, а умови (3) та (4) — функціональними обмеженнями.

Якщо функції φ, g_j, h_j лінійні, то задача називається задачею лінійного програмування, якщо хоча б одна з них нелінійна — задачею нелінійного програмування. Моделі, які використовують тільки складання, віднімання або функцію *СУММ*, є лінійними. Якщо використовується множення, ділення або такі функції, як експоненційна, модель є нелінійною. Приклади нелінійних задач:

— розсилання рекламних листівок звичайно веде до збільшення обсягів продажу, але у разі значного зростання їх настає насичення, за якого процентне співвідношення відповідей на відправлені листівки зменшуватиметься;

— збільшення чисельності службовців сприяє кращому обслуговуванню клієнтів і збільшенню прибутку. Проте за досягнення певного рівня кількості працівників залучення додаткових призведе до зменшення прибутку.

Якщо поряд з обмеженнями (2), (3), (4) накладається умова, щоб x_k були цілими числами, то задача називається задачею цілочисельного програмування; якщо накладається умова, щоб x_k дорівнювали 1 або 0, — то задачею двоїстого програмування.

Для застосування програми *Пошук рішення* необхідно побудувати модель аналізованого об'єкта чи процесу у вигляді таблиці з формулами, які відбивають задані обмеження і цільову функцію. Від форми подання цих даних і структури табличної моделі значною мірою залежать трудомісткість і безпомилковість уведення даних у програму, а також придатність моделі для обґрунтування рішень, що приймаються на підставі отриманого результату. З огляду на це часто (але не завжди) є сенс подавати модель у вигляді, принаймні, трьох секцій: секції керованих змінних; секції цільової функції; секції функціональних обмежень (якщо вони є). Орієнтовний вигляд цих секцій наведено на рис. 2.3.12.

A	B	C	D	E

Цільова функція

Назва	Поточне значення

Змінні

Назва	Поточні значення	Межі	
		макс.	мін.

Функціональні обмеження

Назва	Поточні значення	Межі		
		макс.	мін.	=

Рис. 2.3.12. Основні секції моделі

Клітина колонки *B* секції цільової функції називається *цільовою клітиною*. Це клітина, в якій відбивається значення оптимізованого показника. Цільова клітина має містити формулу (або адресу клітини з формулою), результат якої змінюється залежно від значення клітин колонки *B* секції змінних. Програма *Пошук рішення* змінює значення клітин секції змінних, допоки у цільовій клітині не з'явиться результат, який потрібно отримати (якщо, звичайно, його можна отримати). У клітини колонок *C* і *D* секції змінних вносяться допустимі граничні значення змінюваних клітин (прямі обмеження).

У клітини колонки *B* секції функціональних обмежень уводяться формули для обчислення значення функцій g_i та h_j або адреси їх розташування. У колонки *C*, *D* та *E* цієї секції даних уводяться відповідно значення $F_{i\ \max}$, $F_{i\ \min}$ та F_j.

Для полегшення введення обмежень (як прямих, так і функціональних) бажано, щоб рядки з обмеженнями одного типу створювали безперервну групу.

Крім цих секцій в моделі можуть бути (а у більшості задач обов'язково мають бути) й інші блоки, кількість яких залежить від складності модельованих процесів і вміння розробника моделі подати їх у наочному вигляді.

Для запуску програми *Пошук рішення* виконується команда *Сервіс*/*Пошук рішення*. Після цього з'явиться вікно діалогу *Пошук рішення*. У його полі *Змінюючи комірки* необхідно вказати діапазон, що містить клітини, значення яких програма повинна змінити для отримання оптимального значення (клітини колонки *B* секції *Змінні*). Після активізації цього поля з допомогою миші треба виділити ці клітини або з клавіатури ввести адресу діапазону. За натискання кнопки *Припустити* програма виділяє клітини, на які прямо або опосередковано посилається цільова клітина.

Для того щоб задати обмеження (як прямі, так і функціональні), треба натиснути на кнопку *Додати*. В результаті з'явиться наступне діалогове вікно — *Додавання обмежень*.

У лівому полі *Посилання на клітину* вказується адреса клітини або діапазону клітин, вміст яких має відповідати одному типу обмежень.

Праве поле (*Обмеження*) цього вікна призначене для введення значення обмеження або у вигляді константи, або у вигляді адреси клітини, яка містить це значення. Кнопка *Додати* використовується для переходу до введення наступного обмеження (або групи обмежень).

По закінченні введення всіх параметрів натискається кнопка *ОК*.

Для внесення змін та вилучення обмежень використовуються відповідно кнопки *Змінити* та *Вилучити*.

Для того щоб користувач мав можливість, змінюючи параметри, кілька разів послідовно повторювати пошук оптимального значення, введені параметри зберігаються у робочому аркуші так само, як і інші дані робочої книги.

У разі натискання у вікні *Пошук рішення* кнопки *Параметри* з'явиться вікно *Параметри пошуку рішення*, в якому можна задати додаткові параметри.

Слід звернути увагу на те, що в разі використання лінійної моделі треба включити параметр *Лінійна модель.* Результат буде отримано швидше, і він матимете більшу точність, а звіт на стійкість міститиме більше інформації, яка може бути використана для прийняття більш слушних рішень.

Запуск програми на виконання обчислень здійснюється натисканням кнопки *Виконати***.** Залежно від складності задачі і швидкості процесора пошук рішення може відібрати певний час. Окремі кроки цього процесу відображаються у нижній частині вікна Excel у рядку стану. Якщо оптимальне рішення буде знайдено, відповідні значення вставляться в таблицю і на екрані з'явиться діалогове вікно *Результати пошуку рішення* з інформацією про закінчення цього процесу.

Однак не завжди задача має рішення. У такому разі в діалоговому вікні замість «*Рішення знайдено. Всі обмеження та умова оптимальності виконані*» буде повідомлення «*Пошук не може знайти слушне рішення*».

Якщо рішення знайдено, то користувач повинен вибрати один з режимів (збереження отриманого рішення або поновлення початкових значень) і матиме змогу задати видачу звітів, потрібних для проведення аналізу оптимального рішення. Такий аналіз — дуже важливий етап використання *Програми пошуку*. Передусім це пояснюється тим, що значна частина маркетингових даних є неточною — на практиці здебільшого використовуються набли-

жені значення. Саме тому користувач повинен з'ясувати, як зміни у даних позначаться на оптимальному рішенні. Загалом можна задати видачу звітів трьох типів: *Результати*, *Стійкість*, *Межі*.

Аналіз рентабельності виробництва сукупності товарів

Визначення найвигіднішого асортименту — досить складна задача, для розв'язання якої у загальному випадку слід ураховувати як ринкові умови, так і технологічні можливості виробництва.

Розглянемо модель підприємства, що виробляє два види продуктів — П1 і П2. Передбачається, що у разі встановлення ціни в 3,80 грн/т останній може бути проданий в будь-якій кількості. Існує контракт, за яким треба виробляти не менше як 40 тис. т/день продукту П1 за ціною 5,50 грн/т. У разі вироблення додаткової кількості продукту його можна продати за ціною 5,50 грн/т (не більше 5 тис. т/день), використати для збільшення запасів (не більше 4 тис. т/день), які з урахуванням витрат на збереження оцінюються за подальшої реалізації в 5,20 грн/т, або продати за зниженою ціною (5 грн/т) у необмеженій кількості. У разі потреби обидва продукти можна докупити: закупівельна ціна продукту П1 становить 5,75 грн/т, а продукту П2 — 4 грн/т.

Для виробництва цих продуктів підприємство може отримати два види сировини: сировину С1 — до 100 тис. т/день за ціною 3,25 грн/т і сировину С2 більш високої якості — до 30 тис. т/день за ціною 3,40 грн/т.

Загальна потужність основного процесу переробки сировини — 100 тис. т/день за витрат на переробку 0,35 грн/т. Він дає змогу одержати 0,15 т продукту П1 і 0,85 т продукту П2 з 1 т сировини виду С1. У разі використання 1 т сировини С2 вихід продукції становитиме 0,25 т продукту П1 і 0,75 т продукту П2. Продукт П1 потребує додаткового очищення, витрати на яке — 0,10 грн/т сировини.

Для переробки продукту П2 в більш цінний продукт П1 можна застосовувати агрегат (конвертор), який дає змогу з 1 т продукту П2 одержати 0,5 т продукту П1 і 0,5 т продукту, що може бути проданий як продукт П2, але який не можна повторно переробляти конвертором. Потужність конвертора — 50 тис. т сировини (продукту П2) на день за витрат на цю обробку 0,25 грн/т сировини.

Модель підприємства складається з декількох таблиць, які розміщуються на трьох робочих аркушах з назвами *Асортимент*, *Виробництво*, *Витрати* (рис. 2.3.13—2.3.15).

Починати будувати модель слід з таблиці «Регульовані фактори (змінні)». Після цього створюються модель виробництва (на аркуші *Виробництво*) і калькуляція витрат (на аркуші *Витрати*), в яких використовуються адреси клітин з поточними значеннями планованих змінних.

За калькулювання витрат використовується метод попроцесного обліку, який акумулює виробничі витрати по однакових продуктах виробництва. Слід звернути увагу на те, що постійні витрати не враховуються, оскільки вони не залежать від асортименту продукції, що випускається. Після калькулювання витрат може бути сформована цільова функція, значення якої обчислюється в клітині *H6* аркуша *Асортимент*, а після побудови моделі виробничого процесу можна записати функціональні обмеження (рядки 24–27 аркуша *Асортимент*). Ці обмеження враховують максимально допустиме значення загальної потужності основного процесу переробки сировини і можливі величини наявної кількості (залишків) продуктів.

	A	B	C	D	E	F	G	H	I
1		**АНАЛІЗ РЕНТАБЕЛЬНОСТІ ГАМИ ТОВАРІВ**							
2									
3		***1. Цільова функція***							
4		Виручка від реалізації							=СУММ(I16:I19)
5		Маржинальна собівартість реалізованої продукції							=Витрати!D17
6		Маржинальний прибуток (тис. грн)							=H4-H5
7									
8		***2. Регульовані фактори (змінні)***							

9	Змінна	Планована кількість (тис. т)	Прямі обмеження		Ціна	Вартість
10			Мінім. кількість	Максим. кількість		
11	Кількість переробленої сировини С1	0	0	100	3,25	=E11*H11
12	Кількість переробленої сировини С2	0	0	30	3,40	=E12*H12
13	Кількість докупленого продукту П1	0	0	без обмежень	5,75	=E13*H13
14	Кількість докупленого продукту П2	0	0	без обмежень	4,00	=E14*H14
15	Кількість продукту П2, що конвертується	0	0	50	0,25	=E15*H15
16	Кількість продукту П1 на складі	0	0	4	5,20	=E16*H16

17	Кількість продукту П1, проданого за макс. ціною		43	45	5,50	=E17*H17
18	Кількість продукту П1, проданого за зниж. ціною	0	0	без обмежень	5,00	=E18*H18
19	Кількість проданого продукту П2	7	0	без обмежень	3,80	=E11*H19

20
21
22 ***3. Функціональні обмеження:***
23 Баланс з отримання і витрачання кожного продукту:

24 Залишок П1	=Виробництво!	> =	0
25 Залишок П2	=Виробництво!	> =	0

26 Загальна потужність основного процесу переробки сировини:

27 Поточна потужність	=E11+E12	< =	100

Рис. 2.3.13. Частина моделі, що знаходиться на аркуші *Асортимент*

	A	B	C	D
1		**ВИРОБНИЦТВО ПРОДУКЦІЇ**		
2				
3		***1. Норми отримання продукції***		

4 Об'єкт переробки	Норма виходу продукту (т)	
5	П1	П2
6 С1 (1 т)	0,15	0,85
7 С2 (1 т)	0,25	0,75
8 П2 (1 т)	0,50	0,50

9
10 ***2. Максимальні виробничі потужності (тис. т)***

11 Процес	Максим. потужність
12	
13 Переробка сировини	100
14 Конвертація продукту П2	50

15		
16	***3. Плановані обсяги виробництва (тис. т)***	
17	Спосіб виробництва продукту	Кількість (тис. т)
18	Виробництво продукту П1 із сировини С1 і С2	=СУММ(G19:G20)
19	у тому числі: із сировини С1 ……	=D6*Асортимент!E11
20	із сировини С2 ……	=D6*Асортимент!E12
21	Виробництво продукту П2 із сировини С1 і С2	=СУММ(G19:G20)
22	у тому числі: із сировини С1 ……	=E6*Асортимент!E11
23	із сировини С2 ……	=E6*Асортимент!E12
24	Виробництво продукту П1 конвертацією П2	=D8*Асортимент!E15
25	Неконвертований залишок продукту П2	=E8*Асортимент!E15

Рис. 2.3.14. Частина моделі, що знаходиться на аркуші *Виробництво*

	A	B	C	D	E
1		**ОБЛІК МАРЖИНАЛЬНОЇ СОБІВАРТОСТІ ПРОДУКЦІЇ**			
2					
3		***1. Вартість виробничих операцій***			
4		Операція			Вартість (грн)
5		Переробка 1 т сировини			0,35
6		Вартість очищення продукту П1, виробленого з 1 т сировини			0,10
7		Конверторна обробка 1 т продукту П2			0,25
8					
9		***2. Прямі змінні витрати***			
10		Вид витрат		Сума (тис. грн)	
11		На сировину		=СУММ(Асортимент!I11:I12)	
12		Переробка сировини		=(Асортимент!E11+Асортимент!E11)*F5	

13	Очищення продукту П 1	=Виробництво!G18*F6
14	Конверторна обробка П2	=Асортимент!E15*F7
15	Докупівля продукту П 1	=Асортимент!I13
16	Докупівля продукту П 2	=Асортимент!I14
17	Підсумок	СУММ(D11:D16)

Рис. 2.3.15. Частина моделі, що знаходиться на аркуші *Витрати*

Наявна кількість продукту П1 (клітина *E24*) обчислюється за формулою

$$= \text{Виробництво!}G18 + \text{Виробництво!}G24 + \\ + E13 - E16 - E17 - \text{E18},$$

а наявна кількість продукту П2 (клітина *E25*) за формулою

$$= \text{Виробництво!}G21 + E14 + \\ + \text{Виробництво!}G25 - E15 - E19.$$