**НАЦІОНАЛЬНА АКАДЕМІЯ СТАТИСТИКИ, ОБЛІКУ ТА АУДИТУ**

**КАФЕДРА  СТАТИСТИКИ**

# Методичні рекомендації до вивчення курсу “Критерії статистики ” (для студентів денної та заочної форм навчання)

спеціальності 051 “Економіка”
освітньої програми  “Прикладна статистика та бізнес аналітика”

**Київ - 2019**

Герасименко С.С. Методичні рекомендації до вивчення курсу “Критерії статистики ”. – К.: Національна академія статистики, обліку та аудиту, 2019. – 45 с.

Рецензент:
Гончар І.А., канд. екон. наук, доцент

Затверджено на засіданні кафедри статистики (протокол №1 від 13 вересня 2019 р.) та схвалено Вченою радою Національної академії статистики, обліку та аудиту (протокол №1 від 24 вересня 2019 р.).

# ВСТУП

Навчальна дисципліна “Критерії статистики” вивчається згідно навчального плану Національної академії статистики, обліку та аудиту щодо підготовки фахівців освітньо-кваліфікаційного рівня “магістр” галузі знань 05 “Соціальні та поведінкові науки” спеціальності 051 “Економіка” освітньої програми “Прикладна статистика та бізнес аналітика” і є важливою складовою системи статистичних наук, оскільки управління соціально-економічним розвитком вимагає використання відповідних критеріїв для перевірки статистичних гіпотез з метою доведення надійності статистичних висновків .

У процесі підготовки економістів засвоєння навчальної дисципліни “Критерії статистики ” має важливе місце, оскільки цей курс містить основні методологічні засади статистичного вивчення соціально-економічних явищ та процесів в умовах функціонування ринкової економіки. Набуття знань та навичок за дисципліною “Критерії статистики” потребує попереднього вивчення низки економічних та статистичних дисциплін, зокрема, таких як

- Теорія ймовірності і математична статистика,
- Економіко-математичні методи та моделі,
- Економетрика,
- Статистика,
- Кореляційний та регресійний аналіз

Поняття та методи, що розглядаються в цих предметах, є базовими в ході вивчення дисципліни “Критерії статистики”.

**Метою вивчення дисципліни** “Критерії статистики ” є набуття студентами поглиблених знань з теорії та практики використання критеріїв статистики в аналізі соціально-економічних явищ та процесів та ознайомлення з методами їх застосування в управлінській діяльності.

**Завданнями дисципліни є**:

вивчення сучасних концепцій та проблематики використання критеріїв статистики, що забезпечують обґрунтованість розроблення управлінських рішень на різних рівнях керівництва;

аналіз найбільш відомих методів, за допомогою яких виконують перевірку статистичних гіпотез;

засвоєння технологій використання критеріїв статистики фінансовому аналізі, бюджетуванні, управлінні ефективністю і бізнес процесами;

дослідження можливостей застосування у сфері управлінні бізнесом засвоєння підходів та правил використання критеріїв статистики, а також прийняття з їх використанням рішень у бізнесі.

Зміст дисципліни реалізується шляхом опануванням трьох блоків – теоретичного, практичного та самостійної роботи, що забезпечує закріплення теоретичних знань, сприяє отриманню практичних навичок та розвитку самостійного наукового мислення.

“Методичні рекомендації” містять:

- тематичний план курсу;
- зміст тем курсу;
- вказівки до виконання контрольних робіт;
- перелік рекомендованої літератури.

# ТЕМАТИЧНИЙ ПЛАН КУРСУ

## Тема 1.  Статистичні гіпотези та статистичний критерій

1. Гіпотеза, наукова гіпотеза та статистична гіпотеза.
2. Мета перевірки статистичної гіпотези
3. Визначення статистичного критерію
4. Основні складові статистичного критерію

## Тема 2. Похибки перевірки статистичних гіпотез

1. Перевірка нульової гіпотези.
2. Необхідність перевірки альтернативної гіпотези
3. Потужність критерію.
4. Критерій згоди Пірсона

## Тема 3. Застосування критеріїв в соціально-економічних дослідженнях

1. Етапи перевірки статистичної гіпотези
2. Перевірка гіпотези про рівність середніх
3. Перевірка гіпотези про рівність часток
4. Перевірка гіпотези про рівність дисперсій

## Тема 4.  Статистичні критерії в моделюванні соціально-економічних явищ та процесів

1. Критерії перевірки значущості регресійної моделі
2. Перевірка значимості коефіцієнтів регресії
3. Критерії перевірки адекватності прогнозних моделей
4. Оцінювання похибок прогнозування

## ЗМІСТ ТЕМ КУРСУ

### Тема 1 Статистичні гіпотези та статистичний критерій

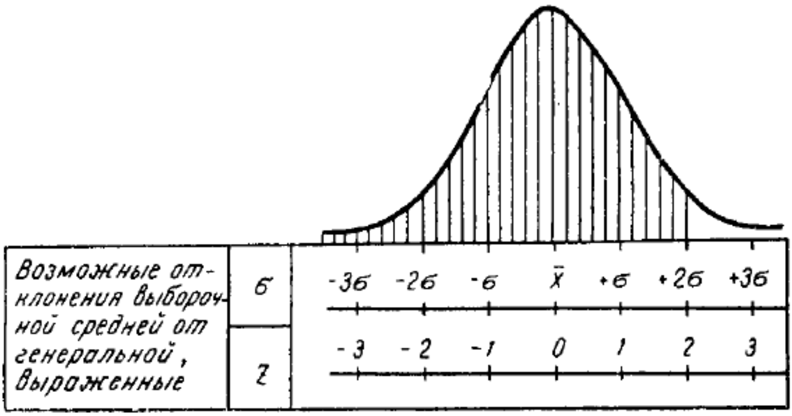
Математична статистика розробляє методи систематизації і аналізу даних, отриманих в процесі спостереження масових явищ. Як правило, у дослідників немає повних даних про сукупність, що вивчається. Методи математичної статистики дозволяють на основі наявної інформації зробити висновки про явище, що вивчається. Одним з основних розділів математичної статистики є теорія перевірки гіпотез.

Гіпотеза - це припущення про деякі властивості явищ, що вивчаються. Статистичними гіпотезами називаються припущення про властивості генеральної сукупності, які можна перевірити, спираючись на дані випадкової виборки. Сенс перевірки статистичних гіпотез полягає в тому, щоб за даними випадкової вибірки прийняти найбільш обгрунтоване рішення про вид або параметри генеральної сукупності, тобто прийняти або відхилити гіпотезу з мінімальним ризиком помилки.

Параметри вибіркової сукупності (частка, середня і дисперсія), як правило, відхиляються від параметрів генеральної сукупності. Можна зробити припущення про величину того або іншого параметра генеральної сукупності і перевірити його на основі вибірки. Такого роду припущення і є статистичними гіпотезами. Перевіримо, наприклад, гіпотезу про рівність генеральної середньою Х деякому постійному числу. Припустимо, що в групі цехів норма часу на складання вузла машини - 89 хв. Пропонується ввести технічно обгрунтовану галузеву норму 81 хв. Перевіримо гіпотезу про те, що витрати часу па складання цього вузла відповідають технічно обгрунтованій галузевій нормі. Позначимо випробовувану гіпотезу через $H_0$, але, її можна записати так: $H_0: \overline{\tilde{O}} = 81$. Це означає наступне: гіпотеза полягає в тому, що генеральна середня $\overline{\tilde{O}}$ дорівнює 81 хв.

Нам відома лише вибіркова середня. Як правило, вибіркова середня відрізняється від генеральної середньої. Усі можливі значення вибіркових середніх (їх сукупність називається вибірковою множиною) визначаються складом генеральної сукупності. Знаючи усі варіанти генеральної сукупності, можна було б, перебравши їх поєднання, визначити сукупність можливих вибіркових середніх. Передбачити ж заздалегідь вибіркову середню неможливо, оскільки вибірка робиться випадково. Тому вибіркова середня є випадковою величиною. Об'єктивна можливість появи значень випадкової величини називається вірогідністю (Р). Співвідношення окремих варіант випадкової величини і вірогідності, що відповідає їм, називається законом расподілу випадкової величини. Якщо випадковою величиною є параметр вибірки, наприклад вибіркова середня, частість або дисперсія, то розподіл їх у випадкових вибірках називається вибірковим.

Розподіл вибіркових середніх підкоряється нормальному закону розподілу. Це закон розподілу безперервної випадкової величини. Розподіл таких випадкових величин характеризується щільністю вірогідності, яка являється похідною функцією розподілу F (х). Щільність вірогідності f (х) показує вірогідність попадання випадкової величини *x* на ділянку від *x* до x + Δx за умови, що приріст Δx наближається до нуля.




В графіку нормального розподілу площа між кривою і віссю абсцис - це вірогідність появи усіх можливих варіант середньої. Ця вирогідгість дорівнює одиниці. Форма нормальної кривої і положення її на осі абсцис залежать від середньої $\overline{\tilde{O}}$ і середнього квадратичного відхилення *б* генеральній сукупності. Функція розподілу, тобто вірогідність того, що випадкова величина x, що підкоряється нормальному розподілу, буде менше деякого числа X, - це плтцадь під кривою зліва від точки X. Вона виражається інтегралом з межами інтеграції від - ∞ до X. Його не можна виразити через елементарні функції. Тому функція нормального розподілу табульована, тобто складені готові таблиці вірогідності P {x<X}. Для зручності табулювання в якості випадкової змінної використано нормоване відхилення z. Змінна z може бути позитивною і негативною. Конкретні значення, що приймаються цією змінною, позначимо

буквою Z. Її середня дорівнює нулю, а середнє квадратичне відхилення - одиниці.

## Статистичний критерій.

Враховуючи, що розподіл вибіркових середніх асимптотично нормальний, вичислимо функцію його розподілу. Для цього потрібно оцінити параметри генеральної сукупності: математичне очікування і дисперсію. Математичне очікування вибірковою середньою дорівнює середній генеральній сукупності: $\mathrm{M}\tilde{\tilde{o}} = \overline{\tilde{O}}$. Дисперсія вибіркових середніх ($\mu^2$) в $n$ разів менше дисперсії генеральної сукупності. Корінь квадратний з дисперсії вибіркових середніх називається середньою похибкою вибірки . Для нашого прикладу перевірки гіпотези про середні витрати часу на складання вузла генеральна середня передбачається рівною 81 хв. Розрахуємо $\mu^2$. Оскільки генеральна дисперсія витрат часу не відома, використовуємо для розрахунку середньої помилки вибіркову дисперсію $\hat{\sigma}^2 = 261{,}11$. Звідси:

$$\mu^2 = \frac{261{,}11}{36} = 7{,}25, çâ³äêè\, \mu = \sqrt{7{,}25} = 2{,}7$$

Нагадаємо, що $\tilde{\tilde{o}} = 85$ і відхиляється від передбачуваної генеральної середньої на 4 хв. Вичислимо вірогідність отримати в окремій вибірці відхилення, що не перевищує за абсолютною величиною, 4 хв. Визначимо |Z|:

$$|Z| = \left|\frac{\tilde{x} - \bar{X}}{\mu}\right| = \frac{4}{2,7} = 1,48$$

Використовуючи функцію нормального розподілу, знаходимо, що у 862 вибірках з 1000 вибіркова середня може відхилитися від генеральної менше ніж на 4 хв. Тоді вірогідність отримати відхилення більше 4 хв = 0,138. Припустимо, що вірогідність 0,138 досить велика і подія, що відповідає цій вірогідності, практично можлива. Отже, припущення про те, що вибірка з $\tilde{o}$ = 85 могла бути зроблена з генеральної сукупності, де $\bar{\tilde{O}}$ = 81 припустимо. У такому разі гіпотеза про те, що середні витрати часу на складання вузла відповідають технічно обгрунтованій нормі ($H_0$: $\bar{\tilde{O}}$ = 81), не відхиляється. Якщо ж вважати, що вірогідність 0,138 мала і що вибіркова середня $\tilde{o}$ практично не може бути рівною 85, то гіпотеза відхиляється.

Мінімальна вірогідність, починаючи з якої ми визнаємо подію практично неможливим, називається рівнем значущості α. Так, якщо в нашому прикладі α =0,05, то гіпотеза не відхиляється, оскільки 0,138 > 0,05. Квантиль $|Z|_{1-\alpha}$, що відповідає вибраному рівню значущості, ділить усю безліч можливих вибіркових значень z на дві області. Перша з них (для $z < |Z|_{1-\alpha}$) називається областю допустимих значень, друга - критичною областю. Між випадковою величиною z, критичною областю і обраним рівнем значущості існує наступне співвідношення: вірогідність того, що z знаходиться в критичній області (за умови, що вірна випробовувана гіпотеза), рівна обраному рівню значущості α. Значення z, що відповідають рівню значущості α, отримати практично неможливо. Отже, якщо значення z знаходяться в критичній області, гіпотезу можна відхилити. В іншому випадку подія визнається практично можливою (допустимою) і гіпотеза не відхиляється. Тому критична область називається областю відхилення гіпотези, а область допустимих значень - областю прийняття її.

Функція результатів вибірки, на підставі якої гіпотеза відкидається або визнається допустимою, називається статистичною характеристикою гіпотези. При випробуванні гіпотези про середню генеральну сукупність статистичною характеристикою є нормоване відхилення z. Квантиль $|Z|_{1-\alpha}$ , що ділить вибіркову множину середніх на дві області, називається критичним значенням статистичної характеристики.

Статистична гіпотеза перевіряється в такій послідовності:

а) обирається рівень значущості α;

б) визначається критична область, що йому відповідає

в) за результатами вибірки розраховується фактичне значення статистичної характеристики;

г) якщо фактичне значення статистичної характеристики потрапляє в критичну область, то гіпотеза відхиляється, якщо в область допустимих значень, то гіпотеза не відхиляється.

У нашому прикладі (α = 0,05) критичним значенням є модуль квантиля чи $|Z|_{0,95}$ = 1,96. Якщо вірна випробовувана гіпотеза, то в 95 вибірках з 100 отримаємо значення z, що не перевищує 1,96 за абсолютною величиною). Фактичне значення |Zl, рівне 1,48, менше критичного і знаходиться в області допустимих значень. Отже, гіпотеза про те, що середні витрати часу на складання вузла усіма слюсарями-збирачами обстежуваних цехів складають 81 хв., при $\alpha$ = 0,05 не відхиляється.

Це, проте, не дозволяє стверджувати, що генеральна середня Х дорівнює 81 хв. Ми можемо тільки зробити висновок, що результати досвіду не суперечать випробовуваній гіпотезі, узгоджуються з нею. Але вони одночасно можуть не суперечити і іншим гіпотезам. Наприклад, якщо випробувати гіпотезу $H_0$: $\overline{\tilde{O}}$ = 82, то очевидно, що вибіркова середня $\tilde{\tilde{o}}$, рівна 85, також знаходитиметься в області допустимих значень цієї гіпотези.

Тобто, якщо статистична характеристика знаходиться в області допустимих значень, то випробовувана гипотеза не відхиляється. Але не означає, що вона вірна: подальші дослідження можуть привести до протилежного висновку і відхилення гіпотези.

Рішення про відхилення гіпотези залежить від прийнятого рівня значущості. Так, при $\alpha$ = 0,05 гіпотеза про те, що середні витрати часу на складання вузла в обстежених цехах складають 81 хв., не відхиляється. Якщо ж $\alpha$ = 0,20, то критичне значення $|Z|_{0,80}$=1,28. Тобто критична область ширша за критичну область для $\alpha$ = 0,05. При $\alpha$ = 0,20 фактичне значення Z = 1,48 належить критичній області; отже, гіпотеза $H_0$: $\overline{\tilde{O}}$ = 81 відхиляється.

Який результат випробування гіпотези слід вважати більше обгрунтованим? Чи не суперечать вони один одному? У обох випадках рішення про відхилення гіпотези приймається з деякою вірогідністю, що не виключає помилки. При $\alpha$ = 0,05, якщо випробовувана гіпотеза вірна, в 5 вибірках з 100 значення z знаходитиметься в критичній області і гіпотеза буде помилково відхилена. При $\alpha = 0{,}20$ така помилка робитиметься в 20 вибірках з 100. Вірогідність помилкового відхилення гіпотези при $\alpha = 0{,}05$ меньше. Здавалося б, для отримання більше обгрунтованого висновку рівень значущості має бути якомога менше. Наприклад, при $\alpha = 0{,}001$ гіпотеза буде помилково відхилена в одній вибірці з 1000. Проте вибір такого низького рівня значимости- зазвичай не виправданий.

Зі зменшенням рівня значущості $\alpha$ звужуватиметься критична область і відповідно розширюватися область допустимих значень. Якщо середня генеральній сукупності дещо відрізняється від передбачуваного значення α, то очевидне, що з розширенням області допустимих значень статистична характеристика все частіше потраплятиме в цю область. У таких випадках гіпотеза $H_0$: $\overline{\tilde{O}} = \alpha$ не відхилятиметься, хоча насправді вона невірна. Інакше кажучи, зі зменшенням рівня значущості, з одного боку, все рідше помилково відхилятиметься випробовувана гіпотеза, але з іншого боку, критерій не зможе уловити відмінностей між дійсною і передбачуваною середніми. Гіпотезу треба перевіряти так, щоб не відхилити її помилково і в той же час не прийняти її, коли вона невірна.

Для такої перевірки формулюють альтернативну гіпотезу. Альтернативна гіпотеза може бути сформульована по-різному залежно від того, які відхилення генеральної середньої від α (позитивні, негативні або ті і інші) нас цікавлять. Ці гіпотези можна записати так: $H_0$: $\overline{\tilde{O}} \neq \alpha$; $H_0$: $\overline{\tilde{O}} \neq > \alpha$; $H_0$: $\overline{\tilde{O}} < \alpha$. Такі гіпотези називаються складними; вони є об'єднанням простих гіпотез, кожна з яких затверджує рівність генеральної середньої числу, відмінному від α (наприклад, $\overline{\tilde{O}}$ = 82, $\overline{\tilde{O}}$ =86) і т.д. Залежно від вибору альтернативної гіпотези будується критична область випробовуваної гіпотези, що відповідає їй. При цьому доцільно побудувати критичну область так, щоб вона якнайкраще розрізняла ці гіпотези. Тому в критичну область повинні включатися можливі значення статистичної характеристики, в найбільшій мірі що відповідають альтернативній гіпотезі. Так, альтернативній гіпотезі $H_0$: $\overline{\tilde{O}} \neq$ α в рівній мірі відповідають як позитивні, так і негативні відхилення вибірковою середньою від α.

У нашому прикладі чинна норма часу на складання вузла складає 89 хв., тобто вона вище пропонованою технічно обгрунтовної галузевої норми (81 хв.). За результатами хронометражу слід вирішити, чи понизити норму часу на складання вузла до рівня пропонованої галузевої або залишити її колишньою. Тому випробовуваній гіпотезі $H_0$: $\overline{\tilde{O}}$ = 81 протиставляється альтернативна гіпотеза На : X > 81. В цьому випадку фактичне значення Z = 1,48 не належить критичній області. Отже, критерій не відхиляє гіпотезу $H_0$: $\overline{\tilde{O}}$ = 81.

## Тема 2 Похибки перевірки статистичних гіпотез

Одночасна перевірка $H_0$ та $H_\alpha$ дозволяє визначити, наскільки гарно критерій розрізнює ці гіпотези; обчислити, з якою ймовірністю застосовуючи критерій робимо правильний висновок (наприклад, відхиляє

***Помилка першого роду*** полягає в тому, що буде відкинута правильна гіпотеза.

***Помилка другого роду*** полягає в тому, що буде не відкинута неправильна гіпотеза.

Ймовірність *a* допустити мо $H_0$, коли вона невірна), а з якою – помилковий (не відхиляємо $H_0$, хоча вірною є $H_\alpha$).

Тобто, при прийнятті рішень за допомогою гіпотез можуть статися помилки двох родів:

помилку 1-го роду, тобто відкинути гіпотезу $H_0$, коли вона вірна, називається ***рівнем значущості***, або розміром критерію. Найчастіше рівень значущості приймають рівним 0,05 або 0,01. Якщо прийнято рівень значущості рівним 0,05, то це означає, що в п'яти випадках із

100 ми ризикуємо одержати похибку першого роду (відкинути правильну гіпотезу). Ймовірність ($1 - a$) не допустити помилку 1-го роду, тобто не відхилити гіпотезу $H_0$, коли вона вірна, інколи називають оперативною характеристикою критерію.

Ймовірність допустити помилку 2-го роду, тобто не відхилити гіпотезу $H_0$, коли вона невірна, зазвичай позначають $b$. Ймовірність ($1 - b$) не допустити помилку 2-го роду, тобто відкинути гіпотезу $H_0$, коли вона невірна, називається ***потужністю*** (або функцією потужності) ***критерію***.

Варто віддати перевагу тій критичній області, при якій потужність критерію буде найбільшою. Тобто, критичну область варто обирати так, щоб ймовірність потрапляння в неї статистичної характеристики була мінімальною і дорівнювала $a$, якщо нульова гіпотеза $H_0$ вірна, і максимальною в протилежному випадку. Критична область повинна бути такою, щоб при заданому рівні значущості $a$ потужність критерію ($1–b$) була максимальною.

***Критерієм згоди*** називають статистичний критерій перевірки гіпотези про закон розподілу ймовірностей випадкової величини (ознаки генеральної сукупності). Є кілька критеріїв згоди: критерій Колмогорова, критерій Смірнова, критерій Пірсона та ін.

Найбільш розповсюдженим критерієм перевірки вірогідності про закон розподілу ознаки генеральної сукупності є критерій згоди Пірсона, який ґрунтується на порівнянні емпіричних і теоретичних частот та визначається Критерій згоди Пірсона дає відповідь на питання, чи розбіжність між емпіричними і теоретичними частотами зумовлена випадковістю, чи вона є значущою. Як і будь-який інший критерій він не доводить справедливості гіпотези $H_0$, а лише дозволяє встановити на прийнятному рівні значущості *узгодженість чи неузгодженість* гіпотези $H_0$, з даними спостережень

Гіпотезу, що підлягає перевірці, називають основною. Оскільки ця гіпотеза припускає відсутність систематичних розбіжностей (нульові розбіжності) між невідомим параметром генеральної сукупності і величиною, що одержана внаслідок обробки вибірки, то її називають нульовою гіпотезою і позначають $H_0$. Проста гіпотеза, як правило, належить до параметра ознак генеральної сукупності і є однозначною.

*Наприклад*, згідно з простою гіпотезою параметр генеральної сукупності дорівнює конкретному числу. Нульова гіпотеза може стверджувати як про значення одного параметра генеральної сукупності, так і про значення кількох параметрів, а також про закон розподілу ознаки генеральної сукупності.

В основу перевірки покладено наступний принцип: ймовірність того, що статистичний критерій потрапляє в критичну область , дорівнює малій імовірності. Якщо ж виявиться, що хоча ця подія малоймовірна і все ж відбулася, все одно немає підстав приймати нульову гіпотезу.

Кожній нульовій гіпотезі можна протиставити кілька альтернативних (конкуруючих) гіпотез, які позначають символом $H_\alpha$, що заперечують твердження нульової.

Мінімальна ймовірність, починаючи з якої подія вважається за практично неможливу, називається ***рівнем значущості α.***

Квантіль $| Z |_{1-\alpha}$ , що відповідає обраному рівню значущості, поділяє множину вибіркових значень на дві підмножини, які не перетинаються:

- за яких нульова гіпотеза не відхиляється (називають областю допустимих значень),
- за яких нульова гіпотеза відхиляється (називають критичною областю),

і називається ***критичним значенням статистичної характеристики***

Статистична гіпотеза $H_0$ перевіряється у такі послідовності:

а) обирається рівень значущості α.

б) визначається відповідна йому критична область,

в) обчислюється фактична вибіркова статистична характеристика,

г) якщо фактична статистична характеристика попадає в критичну область, то гіпотеза відхиляється, якщо в область допустимих значень – перевірка продовжується з використанням $H_\alpha$.

***Вибіркове спостереження*** — такий вид несуцільного спостереження, при якому обстежуються не всі елементи сукупності, що вивчається, а лише певним чином дібрана їх частина. Сукупність, з якої вибирають елементи для обстеження, називається ***генеральною***, а сукупність, яку безпосередньо обстежують, — ***вибірковою***. Статистичні характеристики вибіркової сукупності розглядаються як оцінки відповідних характеристик генеральної сукупності.

Оскільки вибіркова сукупність не точно відтворює склад генеральної сукупності, то й вибіркові оцінки не збігаються з відповідними характеристиками генеральної сукупності. Розбіжності між ними називають ***похибками репрезентативності***: для середньої — це різниця між генеральною $\bar{x}_0$ та вибірковою $\bar{x}$ середніми, для частки — різниця між генеральною $d_0$ і вибірковою $p$ частками, для дисперсії — відношення генеральної $\sigma_0^2$ та вибіркової $\sigma^2$ дисперсій тощо.

Формування вибірки виконується за певними правилами. Передусім визначається основа вибірки. ***Механічний добір***. Основа вибірки — упорядкована множина елементів сукупності. Добір елементів здійснюється через рівні інтервали. Крок інтервалу обчислюється діленням обсягу сукупності $N$ на передбачений обсяг вибірки $n$. Початковий елемент вибірки визначається як випадкове число всередині першого інтервалу, другий елемент залежить від початкового числа й кроку інтервалу. Так, для частки вибірки $\frac{n}{N} = 0{,}05$ кроком інтервалу є число $\frac{n}{N} = \frac{1}{0{,}05} = 20$, тобто у вибірку має потрапити кожний двадцятий елемент. Якщо початковий елемент — випадкове число 7, то другим елементом буде 7 + 20 = 27, третім — 27 + 20 = 47 і т. д.

У статистиці використовують два типи оцінок параметрів генеральної сукупності — точкові та інтервальні. ***Точкова оцінка*** — це значення параметра за даними вибірки: вибіркова середня $\bar{x}$ та вибіркова частка *p*. ***Інтервальною оцінкою*** називають інтервал значень параметра, розрахований за даними вибірки для певної ймовірності, тобто ***довірчий інтервал***. Чим менший довірчий інтервал, тим точніша вибіркова оцінка.

Межі довірчого інтервалу визначаються на основі точкової оцінки та граничної похибки вибірки $\Delta = t\mu$:

для середньої

$$\bar{x} - t\mu \le \bar{x}_0 \le \bar{x} + t\mu \ ;$$

для частки

$$p - t\mu \le d_0 \le p + t\mu \ ,$$

де $\mu$ — стандартна (середня) похибка вибірки; *t* — квантиль розподілу ймовірностей (довірче число).

***Стандартна похибка вибірки*** $\mu$ є середнім квадратичним відхиленням вибіркових оцінок від значення параметра в генеральній сукупності. Як доведено в теорії вибіркового методу, дисперсія вибіркових середніх у *n* раз менша від дисперсії ознаки в генеральній сукупності, тобто $\mu^2 = \sigma_0^2 / n$. Оскільки на практиці генеральна дисперсія ознаки $\sigma_0^2$ невідома, у розрахунках можна використати вибіркову незсунену оцінку дисперсії: для повторної вибірки $\sigma^2 \frac{n}{n-1}$, для безповторної $\sigma^2 \frac{N-n}{N-1}$. Отже, формули стандартної похибки:

для повторної вибірки

$$\mu = \sqrt{\frac{\sigma^2}{n-1}},$$

для безповторної вибірки

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

***Гранична похибка вибірки*** $\Delta = t\mu$ — це максимально можлива похибка для взятої ймовірності $F(x)$. Довірче число $t$ показує, як співвідносяться гранична та стандартна похибки. Як бачимо з рис. 6.1, з імовірністю 0,683 гранична похибка не вийде за межі стандартної $\Delta = \pm 1\mu$, з імовірністю 0,954 вона не перевищить $\pm 2\mu$, з імовірністю 0,997 — $\pm 3\mu$. На практиці найчастіше застосовують імовірність 0,954

З урахуванням сказаного формули граничних похибок середньої та частки записують так:

| | **<u>Повторна вибірка</u>** | **<u>Безповторна вибірка</u>** |
|---|---|---|

Для середньої $\Delta_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}}$ ; $\Delta_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$ ;

Для частки $\Delta p = t\sqrt{\frac{pq}{n}}$ ; $\Delta p = t\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)}$ .

## Тема 3 Застосування критеріїв в соціально-економічних дослідженнях

Перевірка статистичних гіпотез при дослідженні соціально-економічних явищ складається з наступних етапів:

1. Формулюється завдання дослідження у вигляді статистичної гіпотези
2. Обирається статистична характеристика гіпотези вибірковий розподіл якої відомих (частіш за все – це нормальний розподіл).
3. Порівнюються можливі помилкові рішення і оцінюються їх наслідки.
4. Обираються нульова та альтернативна гіпотези .
5. Задається рівень значущості та визначається критичне значення статистичної характеристики.

Наступні етапи перевірки гіпотези залежать від того, чи вже виконане вибіркове дослідження, чи тільки готується.

Якщо вибіркове дослідження тільки готується, то для передбачуваного обсягу сукупності визначається потужність критерію чи відповідно до обраної потужності максимальна відмінність між параметрами. Якщо потужність недостатня, визначається обсяг вибіркової сукупності, що відповідає заданній потужності, та – про можливості – обстежується сукупність встановленого обсягу.

Якщо вибіркове обстеження вже здійснене, то обчислюється фактичне значення статистичної характеристики, перевіряється нульова гіпотеза, Якщо вона не відхиляється, то обчислюють потужність критерію. Якщо потужність критерію достатня, то не відхилення нульової гіпотези рівноважне тому, що вона вірна.

***Перевірка гіпотези про рівність часток ознаки двох сукупностей***

Задача порівняння часток (відносних частот) ознаки в двох сукупностях досить часто зустрічається на практиці. Наприклад, якщо вибіркова частка ознаки однієї сукупності відрізняється від такої ж частки в другій сукупності, чи вказує це на те, що наявність ознаки в одній сукупності дійсно ймовірніше, чи ця різниця часток є випадковою? Вибір виду критичної області і перевірка гіпотези здійснюється таким же чином, як і вище, при перевірці гіпотези про рівність середніх.

***Перевірка гіпотези про рівність дисперсій двох сукупностей***

Гіпотези про дисперсії виникають доволі часто, оскільки дисперсія характеризує такі виключно важливі показники, як точність машини, приладу, технологічних процесів, ступінь однорідностей сукупностей і т.і. Досить часто йдеться про перевірку гіпотези щодо відмінностей генеральної дисперсії $\sigma^2$ та певної постійної величини $\alpha$. Для цього використовується співвідношення

$$\lambda^2 = \frac{\sigma^2}{\alpha}$$

Таке саме співвідношення використовується і при порівнянні двох дисперсій. В якості статистичної характеристики використовується величина $\lambda^2$ множена на число ступенів свободи *k (* загальному випадку *число ступенів свободи* – це кількість випадкових величин зменшена на кількість лінійних співвідношень між ними), розподіл якої називається розподілом *хі-квадрат* ($\chi^2$).

Для перевірки гіпотези про значення $\lambda^2$ потрібно обрати нульову та альтернативну гіпотези. Нульовими можуть бути: 1) дисперсія не змінилась; 2) дисперсія збільшилась; 3) дисперсія зменшилась; 4) дисперсія змінилася.

Дисперсія посідає особливе місце у статистичному аналізі соціально-економічних явищ. На відміну від інших характеристик варіації завдяки своїм математичним властивостям вона є невіддільним і важливим елементом інших статистичних методів, зокрема дисперсійного аналізу.

Для ознак метричної шкали дисперсія — це середній квадрат відхилень індивідуальних значень ознаки від середньої:

$$\sigma^2 = \frac{\sum_{1}^{m}(x_j - \bar{x})^2 f_j}{\sum_{1}^{m} f_j}.$$

Якщо сукупність розбито на групи за певною ознакою *x*, то для будь-якої іншої ознаки *y* можна обчислити дисперсію як у цілому по сукупності, так і в кожній групі. Центром розподілу сукупності в цілому є загальна середня $\bar{y} = \frac{\sum_{1}^{n} y}{n}$, центром розподілу в *j*-й групі — групова середня $\bar{y}_j = \frac{\sum_{1}^{f_j} y}{f_j}$. Відхилення індивідуальних значень ознаки *y* від загальної середньої $\bar{y}$ можна подати як дві складові: $(y - \bar{y}) = (y - \bar{y}_j) + (\bar{y}_j - \bar{y})$. Узагальнюючими характеристиками цих відхилень є дисперсії: загальна, групова та міжгрупова.

Взаємозв’язок факторної та залишкової варіацій описується правилом декомпозиції варіації:

$$\sigma^2 = \delta^2 + \sigma_e^2,$$

де : $\sigma^2 = \frac{\sum_{1}^{m_y}(y_i - \bar{y})^2 f_i}{\sum_{1}^{m_y} f_i}$ — загальна дисперсія ознаки $y$, яка за спрощеним підходом може бути обчислена як $\sigma^2 = \overline{y^2} - (\bar{y})^2$

$$\delta^2 = \frac{\sum_1^{m_x} (\bar{y}_j - \bar{y})^2 f_j}{\sum_1^{m_x} f_j}$$ — факторна (міжгрупова дисперсія, яка вимірює варіацію ознаки $y$ під впливом фактора $x$) дисперсія;

$\sigma_e^2$ — залишкова дисперсія.

Мірою щільності зв'язку за наявності аналітичної групувальної таблиці є ***кореляційне відношення***

$$\eta^2 = \frac{\delta^2}{\sigma^2},$$

## Тема 4 Статистичні критерії в моделюванні соціально-економічних явищ та процесів

Важливою характеристикою кореляційного зв'язку є ***лінія регресії*** — емпірична в моделі аналітичного групування і теоретична в моделі регресійного аналізу. ***Емпірична лінія регресії*** представлена груповими середніми результативної ознаки $\bar{y}_j$, кожна з яких належить до відповідного інтервалу значень групувального фактора $x_j$. ***Теоретична лінія регресії*** описується певною функцією $Y = f(x)$, яку називають ***рівнянням регресії***, а $Y$ — ***теоретичним рівнем результативної ознаки***.

Якщо зі зміною фактора $x$ результат $y$ змінюється більш-менш рівномірно, такий зв'язок описується лінійною функцією $Y = a + bx$. Коли йдеться про нерівномірне співвідношення варіацій взаємозв'язаних ознак (наприклад, коли прирости значень $y$ зі зміною $x$ прискорені чи сповільнені або напрям зв'язку змінюється), застосовують нелінійні регресії, зокрема:

степеневу $Y = ax^b$;

гіперболічну $Y = a + \frac{b}{x}$;

параболічну $Y = a + bx + cx^2$ тощо.

Вибір та обґрунтування функціонального виду регресії ґрунтується на теоретичному аналізі суті зв'язку.

Зауважимо, що теоретичний аналіз суті зв'язку, хоча й дуже важливий, лише окреслює особливості форми регресії і не може точно визначити її функціонального виду. До того ж у конкретних умовах простору і часу межі варіації взаємозв'язаних ознак $x$ і $y$ значно вужчі за теоретично можливі. І якщо кривина регресії невелика, то в межах фактичної варіації ознак зв'язок між ними досить точно описується лінійною функцією. Цим значною мірою пояснюється широке застосування лінійних рівнянь регресії:

$$Y = a + bx .$$

Параметр $b$ (***коефіцієнт регресії***) — величина іменована, має розмірність результативної ознаки і розглядається як ***ефект впливу*** $x$ на $y$. Параметр $a$ — вільний член рівняння регресії, це значення $y$ при $x = 0$. Якщо межі варіації $x$ не містять нуля, то цей параметр має лише розрахункове значення.

Будуємо регресійне рівняння виду $Y = a + bx$ .

Для отримання значень коефіцієнтів цього рівняння використовуються формули

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x},$$

$$a = \bar{y} - b\bar{x}.$$

Важливою характеристикою регресійної моделі є відносний ефект впливу фактора $x$ на результат $y$ — ***коефіцієнт еластичності***:

$$\gamma = b\frac{\bar{x}}{\bar{y}}.$$

Про якість моделей регресії можна судити також за значеннями коефіцієнта кореляції (індексу кореляції) і коефіцієнта детермінації для однофакторної моделі і за значеннями коефіцієнта множинної кореляції та сукупного коефіцієнта детермінації для моделей множинної регресії. Формули розрахунку цих коефіцієнтів наведені в параграфі 7.2. Чим ближче абсолютні величини зазначених коефіцієнтів до 1, тим тісніше зв'язок між досліджуваним ознакою і вибраними факторами і, отже, з тим більшою впевненістю можна судити про адекватність побудованої моделі, що включає в себе найбільш впливові фактори.

Для оцінки точності регресійних моделей зазвичай використовуються ті ж статистичні критерії точності, що і для трендових моделей, зокрема, середня відносна помилка апроксимації (див. формулу (5.14)). Перевірка ***значущості моделі*** регресії проводиться з використанням $F$-критерії Фішера, розрахункове значення якого знаходиться як відношення дисперсії вихідного ряду спостережень досліджуваного показника і незміщеної оцінки дисперсії залишкової послідовності для даної моделі. Якщо розрахункове значення цього критерію зі ступенями свободи $\boldsymbol{v}_1 = \boldsymbol{n} - 1$ і $\boldsymbol{v}_2 = \boldsymbol{n} - m - 1$, де $\boldsymbol{n}$ - кількість спостережень і $\boldsymbol{m}$ - кількість включених у модель факторів, що більше табличного значення критерію Фішера при заданому рівні значимості, то модель визнається статистично значущою.

При перевірці якості регресійної моделі доцільно також оцінити ***значимість коефіцієнтів регресії.*** Ця оцінка проводиться за і-статистики Стьюдента шляхом перевірки гіпотези про рівність нулю $\boldsymbol{k}$-го коефіцієнта регресії $(\boldsymbol{k} = 1, 2,..., \boldsymbol{m})$. Розрахункове значення $\boldsymbol{t}$ критерію з числом ступенів свободи $(\boldsymbol{n} - \boldsymbol{m} - 1)$ знаходять шляхом ділення $\boldsymbol{k}$-го коефіцієнта регресії на середньоквадратичне відхилення цього коефіцієнта, яке в свою чергу обчислюється як квадратний корінь з добутку незміщеної оцінки дисперсії залишкової компоненти і $\boldsymbol{k}$-го діагонального елемента матриці, оберненої до матриці системи нормальних рівнянь щодо параметрів моделі. Це розрахункове значення порівнюється з табличним значенням критерію Стьюдента при заданому рівні значущості, і якщо воно більше табличного значення, коефіцієнт регресії вважається статистично значимим. В іншому випадку відповідний даного коефіцієнту регресії фактор слід виключити з моделі, при цьому якість моделі не погіршиться.

Перейдемо до питання економічного прогнозування на основі моделі регресії, при цьому будемо припускати, що модель, побудована на базі тимчасових рядів досліджуваного показника і включених у модель факторів, є адекватною і досить точною. При використанні побудованої моделі для прогнозування робиться припущення про збереження існуючих взаємозв'язків змінних і на період попередження.

Для прогнозування залежної перемінної (результативної ознаки) на ***L*** кроків вперед необхідно знати прогнозні значення всіх вхідних у модель факторів. Ці значення можуть бути отримані на основі екстраполяційних методів, наприклад, з використанням середніх абсолютних приростів факторних ознак; вони можуть бути також визначені методами експертних оцінок або безпосередньо задані дослідником економічного процесу. Прогнозні значення факторів підставляють в модель і отримують точкові прогнозні оцінки досліджуваного показника.

Для визначення області можливих значень результативного показника при відомих значеннях факторів, тобто довірчого інтервалу прогнозу, необхідно враховувати два можливі джерела помилок. Помилки першого роду викликаються розсіюванням спостережень щодо лінії регресії, і їх можна врахувати, зокрема, величиною середньоквадратичної помилки апроксимації досліджуваного показника з допомогою регресійної моделі. Позначимо цю величину $s_{\hat{y}}$ і обчислимо її за формулою, аналогічною (5.17).

Помилки другого роду обумовлені тим, що в дійсності жорстко задані в моделі коефіцієнти регресії є випадковими величинами, розподіленими за нормальним законом. Ці помилки враховуються введенням поправочного коефіцієнта при розрахунку ширини довірчого інтервалу; формула для його розрахунку включає табличне значення *t*-статистики при заданому рівні значущості і залежить від виду регресійної моделі.

Для лінійної однофакторної моделі, загальний вигляд якої має структуру, аналогічну (7.1), величина відхилення від лінії регресії задається виразом (позначимо його ***R***):

$$R(n,L,\alpha)=S_{\hat{y}}t_{\alpha}\sqrt{1+1/n+(x_{n+L}-\bar{x})^2\Big/\sum_{t=1}^{n}(x_t-\bar{x})^2}.$$

Тут ***n*** - число спостережень, ***L*** - кількість кроків вперед, ***a*** - рівень значимості прогнозу, $x_t$ - спостережуване значення факторної ознаки в момент ***t***, $\bar{x}$ - середнє значення спостережуваного фактора, $x_{n+L}$ - прогнозне значення фактора на ***L*** кроків вперед.

Таким чином, для розглянутої моделі формула розрахунку нижньої і верхньої межі довірчого інтервалу прогнозу має вигляд

$$U_y=\hat{y}_{n+L}\pm R(n,L,\alpha),$$

де $\hat{y}_{n+L}$ означає точкову прогнозну оцінку досліджуваного результативного показника з моделі на ***L*** кроків вперед.

## ВКАЗІВКИ ДО ВИКОНАННЯ КОНТРОЛЬНИХ РОБІТ.

Виконання контрольної роботи є одним із заключних етапів вивчення курсу, її підготовка сприяє поглибленню та закріпленню теоретичних знань, одержаних студентами при самостійному вивченні окремих розділів дисципліни, набутті навичок роботи над навчальним матеріалом. Перед виконанням контрольної роботи студент зобов’язаний вивчити рекомендовану літературу з дисципліни, зібрати, проаналізувати та узагальнити практичний матеріал. Із списку рекомендованої літератури, що приведений в кінці методичних вказівок, в контрольній роботі слід навести тільки ті джерела, які використовуються студентом при виконанні роботи. Відповіді на завдання повинні бути повними і обґрунтованими, повинні містити статистичні формули, а висновки мають бути економічно обґрунтованими. В кінці роботи студент зобов'язаний навести перелік використаної літератури з зазначенням автора, повного найменування книги (посібника, брошури, статті тощо), міста, видавництва, року видання. Контрольну роботу слід представити на кафедру статистики у встановлений графіком строк, але не пізніше 15 діб до початку екзаменаційної сесії.

Під час виконання контрольної роботи слід додержуватись таких правил:

1. вказати варіант у зошиті і на лицьовому бланку роботи;

2. перед розв’язком задачі привести її зміст, надати докладне рішення задачі з відповідними статистичними формулами, повними розрахунками і короткими поясненнямиодержаних результатів;

3. супроводжувати результати розв’язку задач статистичними таблицями, які слід грамотно оформити;

4. перевірити правильність застосування методу розв’язку задач;

5. подати коментар до отриманих результатів з урахуванням економічного змісту показників.

## ОФОРМЛЕННЯ КОНТРОЛЬНОЇ РОБОТИ

Контрольна робота виконується українською мовою і повинна бути стилістично, граматично та технічно оформлена. Композиція роботи складається у наступній послідовності:➢титульний аркуш;➢зміст;➢основна частина;➢список літератури.

Перша сторінка контрольної сторінки –титульний аркуш. На ньому слід вказати назву вищого навчального закладу та кафедри; прізвище, ім'я та по батькові автора, рік написання роботи та інші відомості, які ідентифікують контрольну роботу.

Текст контрольної роботи друкують на аркушах з одною боку білого паперу формату А4 (210x297 мм) на комп'ютері через 1,5 інтервали або машинописним способом через 2 інтервали, або може бути написаним від руки чорнилом (до тридцяти рядків на сторінці). Текст контрольної роботи друкують, залишаючи береги таких розмірів: лівий –25 мм, правий –10 мм, верхній та нижній –20 мм. Шрифт друку повинен бути чітким з однаковою щільністю тексту, стрічка –чорного кольору середньої жирності. Висота шрифту 1,8 мм (14 комп'ютерний, Times New Roman). Нумерацію сторінок контрольної роботи подають арабськими цифрами у правому верхньому куті сторінки без крапки в кінці.

Першою сторінкою роботи є титульний аркуш, який включається до загальної нумерації сторінок, однак номер сторінки на ньому не проставляється. Контрольна робота завершується останньою сторінкою – списком літератури. Список використаної літератури має суцільну нумерацію.

## ЗАВДАННЯ КОНТРОЛЬНОЇ РОБОТИ

Для виконання контрольної роботи треба внести до таблиці 1 дані про результати діяльності фірми у 2019 році. Для цього треба скористатися з індивідуального N, що надається кожному студенту викладачем. Отримавши свій N необхідно:

- розрахувати товарооборот фірми у 2019 році як величину, що на +4 N% відрізняється від 2018 року, і отриманий результат вписати до відповідної клітинки табл.1,
- розподілити (на власний розсуд) товарооборот фірми за 2018 рік по 20 магазинах та вписати числа у відповідні клітини табл.2,

Табл.1

Розподіл товарообороту фірми за спеціалізацією магазинів в 2018 - 2019 рр.

<table>
<tr><td rowspan="2">Магазини</td><td colspan="2">Товарооборот, млн.грн.</td><td colspan="2">Зміна тавро-обороту</td><td colspan="2">Структура товарообороту, %</td><td colspan="2">Зміна структури товарообороту</td></tr>
<tr><td>2018</td><td>2019</td><td>абс. млн.грн</td><td>відн. %</td><td>2018</td><td>2019</td><td>абс. відсот. пункти</td><td>відн. %</td></tr>
<tr><td>Спеціалізовані</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Універсальні</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Фірма в цілому</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

- розрахувати прибуток фірми у 2019 році як величину, що на -2 N% відрізняється від 2018 року, і отриманий результат вписати до відповідної клітинки табл.2,
- розподілити (на власний розсуд) прибуток фірми за 2019 рік по 20 магазинах та вписати числа у відповідні клітини табл.2.

Табл. 2

Дані про результати діяльності магазинів фірми в 2018-2019 рр.

| Магазини | Товарооборот, млн.грн. | | Прибуток, млн.грн | |
|---|---|---|---|---|
| | 2018 | 2019 | 2018 | 2019 |
| 1 | 20 | | | |
| 2 | 23 | | | |
| 3 | 25 | | | |
| 4 | 26 | | | |
| 5 | 28 | | | |
| 6 | 30 | | | |
| 7 | 27 | | | |
| 8 | 28 | | | |
| 9 | 29 | | | |
| 10 | 19 | | | |
| 11 | 31 | | | |
| 12 | 32 | | | |
| 13 | 33 | | | |
| 14 | 34 | | | |
| 15 | 30 | | | |
| 16 | 36 | | | |
| 17 | 37 | | | |
| 18 | 38 | | | |
| 19 | 34 | | | |
| 20 | 40 | | | |
| Фірма | 600 | | | |

1. За даними табл.2 скласти **структурну групувальну табл.3** (поділивши сукупність на 3 групи – великі, середні та малі магазини, визначивши інтервали групування за даними 2019 р.).

Табл. 3

Розподіл магазинів фірми за товарооборотом у 2018 - 2019 р.

| Групи магазинів за товарообортом | Кількість магазинів | | Структура магазинів | | Зміна структури | |
|---|---|---|---|---|---|---|
| | 2018 | 2019 | 2018 | 2019 | абс. відсот. пункти | відн. % |
| Малі | | | | | | |
| Середні | | | | | | |
| Великі | | | | | | |
| Фірма в цілому | | | | | | |

2. Використовуючи розподіл магазинів за товарооборотом з таб.3 скласти аналітичну групувальну табл..4

Таблиця 4

Розподіл магазинів фірми у 2019 році за товарооборотом та прибутком

| Групи магазинів за товарооборотом ($x$) | Кількість магазинів ($f$) | Середній прибуток одного магазина ($y$) |
|---|---|---|
| Малі | | |
| Середні | | |
| Великі | | |
| Фірма | | |

3. За даними табл.2 про товарооборот магазинів :

- сформувати вибірку з 5 магазинів 2019 р. використовуючи механічний добір;
- обчислити похибку репрезентативності для середньої;

- визначити стандартну, граничну та відносну похибки вибірки для середньої.

4. Використовуючи інформацію табл.4 проаналізувати взаємозв'язок товарообороту та прибутку фірми використавши дисперсійний аналіз. Подати оцінки отриманих результатів за допомогою відповідних критеріїв.

Зразок виконання завд.4

Таблиця 5

Розподіл магазинів фірми у 2017 році за товарооборотом та прибутком

| Групи магазинів за товарооборотом (*x*) | Кількість магазинів (*f*) | Середній прибуток одного магазина (*y*) |
|---|---|---|
| Малі | 5 | 1,12 |
| Середні | 8 | 2,33 |
| Великі | 7 | 3,24 |
| Фірма | 20 | 2,35 |

$$\text{Середній прибуток одного магазину} = \frac{\text{Сума прибутку магазинів групи}}{\text{Кількість магазинів групи}}$$

Взаємозв'язок факторної та залишкової варіацій описується правилом декомпозиції варіації:

$$\sigma^2 = \delta^2 + \sigma_e^2,$$

де : $\sigma^2 = \dfrac{\sum_1^{m_y} (y_i - \bar{y})^2 \; f_i}{\sum_1^{m_y} f_i}$ — загальна дисперсія ознаки $y$,

яка за спрощеним підходом може бути обчислена як $\sigma^2 = \overline{y^2} - \left(\bar{y}\right)^2$

$$\delta^2 = \frac{\sum_{1}^{m_x}\left(\bar{y}_j - \bar{y}\right)^2 f_j}{\sum_{1}^{m_x} f_j}$$ — факторна (міжгрупова дисперсія, яка вимірює варіацію ознаки $y$ під впливом фактора $x$) дисперсія;

$\sigma_e^2$ — залишкова дисперсія.

Мірою щільності зв'язку за наявності аналітичної групувальної таблиці є ***кореляційне відношення***

$$\eta^2 = \frac{\delta^2}{\sigma^2},$$

За даними табл.2:

$\sigma^2$ = Σ прибутків$^2$ кожного з 20 магазинів фірми/ 20 - $2{,}35^2$ =

6,32 – 5,52 = 0,80

$$\delta^2 = \frac{(1{,}12 - 2{,}35)^2 * 5 + (2{,}33 - 2{,}35)^2 * 8 + (3{,}24 - 2{,}35)^2 * 7}{20} = 0{,}655$$

$$\eta^2 - \frac{0{,}655}{0{,}80} = 0{,}818, \text{ або } 81{,}8\%$$

Висновок: варіація прибутку на 81,8% обумовлюється варіацією товарообороту.

Ця гіпотеза потребує перевірки з використанням відповідного критерію – критерію Фішера, яка дозволяє довести випадковість або не випадковість виявленого зв'язку. Але критерій Фішера функціонально пов'язаний з кореляційним відношенням

$$F = \frac{\eta^2}{1-\eta^2} * \frac{k_2}{k_1},$$

де: $k_1$ – число ступенів свободи факторної дисперсії = m – 1

(m – кількість груп)

$k_2$ – число ступенів свободи залишкової дисперсії = n – m

(n – обсяг сукупності).

А тому достатньо перевірити гіпотезу $H_0 : \eta^2 = 0$. Критичні значення (для істотності α = 0,05), які могли б виникнути тільки за рахунок інших (випадкових) чинників, а не того, за яким побудована групувала таблиця і вплив якого оцінюється, наведені в табл.3.

Таблиця 3

Критичні значення коефіцієнта детермінації та кореляційного відношення
(для рівня істотності α = 0,05)

| $k_2/ k_1$ | 1 | 2 | 3 |
|---|---|---|---|
| 10 | 0,332 | 0.451 | 0.527 |
| 12 | 0,283 | 0,394 | 0,466 |
| 14 | 0,247 | 0,348 | 0,417 |
| 16 | 0,219 | 0,312 | 0,378 |
| 18 | 0,197 | 0,283 | 0,345 |
| 20 | 0,179 | 0,259 | 0,318 |

В нашому випадку $k_1$=3-1=2, $k_2$ =20-3=17. Оскільки значення для $k_2$=17 в табл.3 відсутнє, використовуємо найближче значення для $k_2$=16. Критичне значення $\eta^2_{0,05}$ (2,16) = 0,312. Емпіричне кореляційне відношення $\eta^2$ = 0,818 значно перевищує критичне, а отже гіпотеза про випадковий характер відхилень групових середніх відхиляється з ймовірністю 0,95. Тобто, припущення про те, що товарооборот впливає на прибуток не відхиляється з тою ж самою ймовірністю.

5. Використовуючи інформацію табл.2 за 2019 рік проаналізувати взаємозв'язок товарообороту та прибутку фірми використавши кореляційно-регресійний аналіз, побудувавши лінійне рівняння регресії. Подати оцінки отриманих результатів за допомогою відповідних критеріїв.

Зразок виконання завд.5

Будуємо регресійне рівняння виду $Y = a + bx$.

Для отримання значень коефіцієнтів цього рівняння використовуються формули

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x},$$

$$a = \bar{y} - b\bar{x}.$$

Розрахунок величини складових вказаних формул дав такі результати:

$n$ = 20, $\Sigma x$ = 672, $y$ = 47,0, $\bar{y} = 2{,}35$, $\bar{x} = \frac{672}{20} = 33{,}6$, $\Sigma x^2$=22580, $\Sigma xy$ = 1579,3

$$b = \frac{20 * 1579{,}3 - 672 * 47{,}0}{20 * 22580 - 672 * 672} = 0{,}075,$$

Тобто, із збільшенням товарообороту на 1 грн. прибуток може зрости, в середньому, на 0,03 грн., «за інших однакових умов».

Звідти $a$ = 2,35 – 0,075∗33,6 = - 0,17.

Гіпотеза про значення коефіцієнту регресії $b$ = 0,03 потребує перевірки за допомогою відповідного критерію – критерію $t$ - Стьюдента, щоб довести випадковість або не випадковість виявленого зв'язку

$t = \frac{b}{\mu_b}$, де $\mu$ - стандартна похибка коефіцієнта регресії. Стандартна похибка коефіцієнта регресії залежить від варіації факторної ознаки $\sigma_x^2$, залишкової дисперсії $\sigma_e^2$ і числа ступенів свободи $k = n - m$, де $m$ — кількість параметрів рівняння регресії:

$$\mu_b = \sqrt{\frac{\sigma_e^2}{\sigma_x^2(n-m)}}.$$

Для лінійної функції $m$ = 2. $\sigma_x^2 = \frac{22580}{20} - 33{,}6^2 = 0{,}03$

Для розрахунку $\sigma_e^2$ використовується формула $\sigma_e^2 = \frac{1}{n}\sum_1^n (y - Y)^2$, в якій $Y$ – теоретичні значення прибутку, обчислені за побудованим рівнянням регресії наступним чином:

$y$ = - 0,17 + 0,06$x$

Тоді, для першого магазину, де товарооборот у 2017 році становив 24 млн. грн. (див. табл.1): $Y_{24}$ = - 0,17 + 0,075*24 = 1,63, для другого $Y_{25}$ = - 0,17 + 0,075*25 = 1,70. Аналогічно – для всіх інших магазинів.

Тоді $\sigma_e^2$ = $(0{,}9 - 1{,}63)^2 + (1{,}0 - 1{,}70)^2 + \ldots$(по інших 18 магазинах)/20 = 1,64

Звідки $\mu_b = \sqrt{\dfrac{1,64}{0,03(20-2)}} = 2,38$, що перевищує критичне значення $t_{0,95(20)}=2,04$ (див. табл. 4). Тобто, гіпотеза про випадкових характер коефіцієнта регресії відхиляється, а тому припущення про вплив товарообороту на прибуток з ймовірністю 0,95 не відхиляється.

Таблиця 4

Значення квантілів *t* розподілу Стьюдента для α = 0,05

| Число ступенів свободи | 4 | 5 | 6 | 7 | 8 | 10 | 15 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|---|
| Квантілі | 2,78 | 2,57 | 2,45 | 2,38 | 2,31 | 2,23 | 2,13 | 2,00 | 2,04 |

Для оцінювання щільності виявленого зв'язку використовують коефіцієнт детермінації $R^2 = \dfrac{\delta_Y^2}{\sigma_y^2}$, де $\delta_Y^2 = \dfrac{1}{n}\sum_1^n\left(Y - \bar{y}\right)$ - факторна дисперсія,

$\sigma_y^2 = \dfrac{1}{n}\sum_1^n\left(y - \bar{y}\right)^2$ - загальна дисперсія прибутку.

Виходячи із взаємозалежності дисперсій $\sigma_y^2 = \delta_Y^2 + \sigma_e^2$, знаючи з попередніх розрахунків, що $\sigma_e^2$= 1,64 та розрахувавши за формулою $\sigma^2_y$ = 8,48, отримуємо $\delta_Y^2$ = 8,48 – 1,64 = 6,84. Звідки

$$R^2 = \frac{6,84}{8,48} = 0,81.$$

Можна припустити, що в такому випадку 81% відсоток варіації прибутку обумовлений варіацією товарообороту. Але цей висновок лише гіпотеза, яка потребує перевірки шляхом порівняння розрахованого (емпіричного) коефіцієнта детермінації з критичним (див.табл.3). Критичне значення для наступних рівнів ступенів свободи $k_1 = m – 1 = 2 – 1=1$ та $k_2=m-n=20-2=18$ дорівнює $R^2_{0,95(1,18)}$= 0,197. Таким чином, обчислений коефіцієнт детермінації значно перевищує критичне значення, а тому припущення про вплив товарообороту на прибуток не відхиляється з ймовірністю 0,95.

## ПЕРЕЛІК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ


1. Головач А.В., Ерина А.М., Трофимов В. П. Критерии математической статистики в экономических исследованиях. – М:Статистика, 1973. –135 с.
2. Єріна А.М. Статистичне моделювання та прогнозування: Навч.посібник. – Київ: КНЕУ, 2001. – 170 с.
3. Жеребин В.М., Романов А.Н. Уровень жизни населения. – М.:ЮНИТИ-ДАНА, 2002. – 592 с.
4. Інформаційне забезпечення державного та регіонального соціального управління: Монографія / О.Г.Осауленко, О.Ф.Новікова, Н.С.Власенко, І.В.Калачова та ін. / НАН України. Інститут економіки промисловості; Держкомстат України. – Київ, Донецьк, 2004. – 656 с.
5. Смит А. Исследование о природе и причинах богатства народов // Антология экономической классики: В 2 т. – М., 1991. – Т.1. – 375 с.
6. Федоренко Н.П., Римашевская Н.М Система экономико-математических моделей для анализа и прогноза уровня жизни. – М.: Наука, 1986. – 264 с.
7. Черенько Л.М. Бідність пенсіонерів в Україні //Україна: аспекти праці. – 2003. – №8. – С. 3-8.
8. Четыркин Е.М. Статистические методы прогнозирования / Изд. 2-е, перераб. и доп. – М.: Статистика, 1977. – 200 с.

1. Василенко О.А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. /О.А.Василенко, І.А.Сенча. – Одеса:ОНАЗ ім.О.С.Попова, 2011.– 166с.
2. Головач А.В., Ерина А.М., Трофимов В.П Критерии математической статистики в экономических исследованиях. М: Статистика, 1973. – 136 с.


### Допоміжна


1. Трофимов В.П. Измерение взаимосвязей социально-экономических явлений. М: Статистика, 1975. – 149 с.

2. О.Г.Ханін Прийняття статистичних рішень в економіці та фінансах  із застосуванням Excel Режим доступу: https://fisfm.education/content/files/um/kl/umklmn6jr4t3vflhifd6jhtmehagqjv7.pdf