МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ № 1

В контрольной работе № 1 студенты решают шесть задач. Вариант задания выбирается в зависимости от начальной буквы, с которой начинается фамилия студента.

При выполнении работы следует придерживаться следующих требований:

– перед решением задачи необходимо привести ее условие;

– решение задач следует объяснять: сопровождать необходимыми формулами, расчетами, краткими пояснениями. Задачи, в которых приведены только ответы без расчетов, считаются невыполненными;

– все расчеты относительных показателей производятся с принятой в статистике точностью до 0,001, процентов – до 0,1;

– в конце работы указывается перечень использованной литературы, ставится подпись и дата.

1. Методические указания к решению задачи № 1 по темам "Средние показатели" и "Показатели вариации"

Средняя величина является обобщенной количественной характеристикой признака в статистической совокупности. Вычисление средней величины должно осуществляться с учетом экономического содержания изучаемого показателя.

Средняя арифметическая является одной из наиболее распространенных форм средней величины:

средняя арифметическая простая

$$\bar{x} = \frac{\sum x}{n} \quad ;$$

средняя арифметическая взвешенная

$$\bar{x} = \frac{\sum x\, f}{\sum f} \quad ,$$

где $\boldsymbol{x}$ – значения (варианты) признака;

$\boldsymbol{n}$ – число вариантов (число наблюдений), из которых рассчитывается средняя;

$\boldsymbol{f}$ – статистический вес (число повторений значения признака).

Средняя гармоническая вычисляется из обратных значений признака:

средняя гармоническая простая

$$\bar{x} = \frac{n}{\sum \frac{1}{x}};$$

средняя гармоническая взвешенная

$$\bar{x} = \frac{\sum w}{\sum \frac{1}{x} w},$$

где $w = xf$.

Средняя геометрическая невзвешенная рассчитывается следующим образом:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2, \ldots, x_n} = \sqrt[n]{\prod x}.$$

Средняя геометрическая применяется, например, в статистических расчетах при вычислении средних темпов роста.

При определении средних величин в интервальном вариационном ряду в случае открытых крайних интервалов необходимо определить нижнюю границу первого и верхнюю границу последнего интервалов. Для этого используются величины других, закрытых интервалов: величина интервала первой группы условно принимается равной величине интервала последующей, а величина интервала последней группы – величине интервала предыдущей. В интервальном ряду распределений необходимо варианты признака выразить одним числом (дискретным). За такое дискретное число принимается середина интервала.

Для измерения степени варьирования (колеблемости) признака служит вариация, показателями которой являются: размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, средний квадрат отклонений (дисперсия), коэффициент вариации, коэффициент детерминации, эмпирическое корреляционное отношение.

Размах вариации (*R*) характеризует пределы вариации (изменения) индивидуальных значений (или вариантов) признака ($\boldsymbol{x}$) в статистической совокупности

$$R = x_{\max} - x_{\min},$$

где $x_{\max}$, $x_{\min}$ – наибольшее и наименьшее значение признака.

Среднее линейное отклонение вычисляется по формулам

средней арифметической взвешенной

$$\bar{d} = \frac{\sum | x_i - \bar{x} | \, f_i}{\sum f_i}$$

или средней арифметической простой (невзвешенной)

$$\bar{d} = \frac{\sum \left| x_i - \bar{x} \right|}{n},$$

где x_i – варианты признака.

Среднее квадратическое отклонение рассчитывается по формуле взвешенной

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$

или невзвешенной

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}},$$

где x_i – i –е значение признака x;

$\bar{x}$ – средняя величина признака x;

f_i – статистический вес i –го значения признака;

n – число членов совокупности.

При анализе связей признаков в статистической совокупности, разбитой на группы, рассчитываются следующие дисперсии: групповая, межгрупповая, внутригрупповая и общая.

Групповая дисперсия (частная) характеризует вариацию признака в группе, обусловленную действием на него всех прочих факторов, кроме признака, положенного в основание группировки (группировочного признака):

$$\sigma_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 f_{ij}}{\sum_{i=1}^{n_j} f_{ij}},$$

где x_{ij} – значение признака $\boldsymbol{i}$ –й единицы $\boldsymbol{j}$ –й группы;

$\bar{x}_j$ – частная (групповая) средняя величина признака в j –й группе;

f_{ij} – вес значения признака i –й единицы в j –й группе;

n_j – численность единиц j –й группы.

Межгрупповая дисперсия измеряет степень колеблемости (вариацию) признака во всей статистической совокупности за счет фактора, положенного в основание группировки (группировочного признака):

$$\delta^2 = \frac{\sum_{j=1}^{J} \left(\bar{x}_j - \bar{x}\right)^2 F_j}{\sum_{j=1}^{J} F_j},$$

где $\bar{x}_j$ – средняя j –й группы;

$\bar{x}$ – общая средняя;

F_j – вес j –й группы, представляющий собой численность единиц в j –й группе;

J – количество групп в статистической совокупности.

Внутригрупповая дисперсия измеряет степень колеблемости признака во всей совокупности в целом за счет действия на него всех прочих факторов (признаков), кроме группировочного признака:

$$\overline{\sigma^2} = \frac{\sum \sigma_j^2 F_j}{\sum F_j},$$

где σ_j^2 – групповая дисперсия j –й группы.

Общая дисперсия измеряет степень колеблемости признака, порождаемую всей совокупностью действующих на него факторов:

$$\sigma^2 = \overline{\sigma^2} + \delta^2 \; .$$

Коэффициент вариации вычисляется по формуле

$$v = \frac{\sigma}{\bar{x}},$$

где σ – среднее квадратическое отклонение;

$\bar{x}$ – средняя величина признака.

Коэффициент вариации выражается обычно в процентах и дает представление о степени однородности статистической совокупности.

Коэффициент детерминации показывает, какую часть общей вариации изучаемого признака составляет межгрупповая вариация, т.е. обусловленная группировочным признаком. Определяется следующим образом:

$$\eta^2 = \frac{\delta^2}{\sigma^2} \quad .$$

Эмпирическое корреляционное отношение характеризует степень тесноты связи между факторным (группировочным) и результативным признаками

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

где σ^2 – общая дисперсия результативного признака.

2. Методические указания к решению задачи № 2 по теме "Индексы и индексные системы "

Основной формой сводного индекса является *агрегатный индекс*, который вычисляется по следующим формулам:

агрегатный индекс с весами текущего периода используется для изучения динамики цен, себестоимости, производительности труда

$$I_x = \frac{\sum x_1 f_1}{\sum x_0 f_1},$$

например, агрегатный индекс цены

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1};$$

агрегатный индекс с весами базисного периода используется для изучения динамики физического объема продукции

$$I_x = \frac{\sum x_1 f_0}{\sum x_0 f_0},$$

например, индекс физического объема продукции (объема производства в неизменных ценах)

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0};$$

агрегатный индекс совместного изменения обеих величин используется для изучения динамики стоимости продукции, затрат на производство продукции

$$I_{xf} = \frac{\sum x_1 f_1}{\sum x_0 f_0},$$

например, индекс стоимости продукции (объема производства в фактических ценах)

$$I_Q = \frac{\sum p_1 q_1}{\sum p_0 q_0},$$

где x_0, x_1 – индексируемая величина в базисном и текущем периодах;
f_0, f_1 – веса индексов в базисном и текущем периодах;
p_0, p_1 – цена изделия в базисном и отчетном периодах;
q_0, q_1 – физический объем продукции в базисном и отчетном периодах.

Индексы с постоянными весами, как правило, вычисляются при анализе объемных (суммарных) показателей, например, физического объема продукции.

Индексы с постоянными весами рассчитываются следующим образом:

– базисные индексы с постоянными весами

$$I_1^b = \frac{\sum x_1 f_0}{\sum x_0 f_0}; I_2^b = \frac{\sum x_2 f_0}{\sum x_0 f_0};\ldots\ ; I_i^b = \frac{\sum x_i f_0}{\sum x_0 f_0};\ldots; I_n^b = \frac{\sum x_n f_0}{\sum x_0 f_0};$$

– цепные индексы с постоянными весами

$$I_1^c = \frac{\sum x_1 f_0}{\sum x_0 f_0}; I_2^c = \frac{\sum x_2 f_0}{\sum x_1 f_0};\ldots\ ; I_i^c = \frac{\sum x_i f_0}{\sum x_{i-1} f_0};\ldots; I_n^c = \frac{\sum x_n f_0}{\sum x_{n-1} f_0},$$

где I_i^b – базисный индекс $\boldsymbol{i}$ –го периода;

I_i^c – цепной индекс i –го периода;

x_0 – значения показателя в базисном периоде;

x_i – значения показателя в i –м периоде;

f_0 – веса показателя x в базисном периоде.

Например, базисные I_q^b и цепные I_q^c индексы физического объема продукции с постоянными весами – ценами базисного периода p_0 :

$$I_{q1}^b = \frac{\sum q_1 p_0}{\sum q_0 p_0}; I_{q2}^b = \frac{\sum q_2 p_0}{\sum q_0 p_0}; \ldots; I_{qn}^b = \frac{\sum q_n p_0}{\sum q_0 p_0} \quad ;$$

$$I_{q1}^c = \frac{\sum q_1 p_0}{\sum q_0 p_0}; I_{q2}^c = \frac{\sum q_2 p_0}{\sum q_1 p_0}; \ldots; I_{qn}^c = \frac{\sum q_n p_0}{\sum q_{n-1} p_0} \quad .$$

Индексы с переменными весами строятся для таких показателей, как цена, себестоимость единицы продукции, производительность труда.

Индексы с переменными весами рассчитываются следующим образом:

– базисные индексы с переменными весами

$$I_1^b = \frac{\sum x_1 f_1}{\sum x_0 f_1}; I_2^b = \frac{\sum x_2 f_2}{\sum x_0 f_2}; \ldots \quad ; I_i^b = \frac{\sum x_i f_i}{\sum x_0 f_i}; \ldots; I_n^b = \frac{\sum x_n f_n}{\sum x_0 f_n};$$

– цепные индексы с переменными весами

$$I_1^c = \frac{\sum x_1 f_1}{\sum x_0 f_1}; I_2^c = \frac{\sum x_2 f_2}{\sum x_1 f_2}; \ldots \quad ; I_i^c = \frac{\sum x_i f_i}{\sum x_{i-1} f_i}; \ldots; I_n^c = \frac{\sum x_n f_n}{\sum x_{n-1} f_n},$$

где f_i – веса показателя x в i –м периоде.

Например, базисные I_{pi}^b и цепные I_{pi}^c индексы цен с переменными весами – физическими объемами продукции q_i :

$$I^{b}_{p1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; I^{b}_{p2} = \frac{\sum p_2 q_2}{\sum p_0 q_2}; \ldots; I^{b}_{qn} = \frac{\sum p_n q_n}{\sum p_0 q_n},$$

$$I^{c}_{p1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}; I^{c}_{p2} = \frac{\sum p_2 q_2}{\sum p_1 q_2}; \ldots; I^{c}_{pn} = \frac{\sum p_n q_n}{\sum p_{n-1} q_n}.$$

Индекс постоянного (*фиксированного*) *состава* вычисляется с весами, фиксируемыми на уровне какого-либо периода:

$$I_x = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1} = \frac{\sum x_1 f_1}{\sum x_0 f_1}.$$

Например, индекс себестоимости постоянного состава

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1},$$

где z_0, z_1 – себестоимость единицы продукции в базисном и текущем периоде;

q_1 – количество продукции в текущем периоде.

Индекс себестоимости постоянного состава показывает изменение средней себестоимости изделия под влиянием изменения себестоимостей отдельных видов продукции.

Индекс переменного состава показывает соотношение средних уровней изучаемого явления, относящихся к разным периодам времени или разным территориям. Индекс переменного состава вычисляется следующим образом:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \frac{\sum x_1 f_1}{\sum x_0 f_0} : \frac{\sum f_1}{\sum f_0},$$

где $\bar{x}_0, \bar{x}_1$ – среднее значение показателя в базисном и текущем периодах;

f_0, f_1 – веса показателя в базисном и отчетном периодах.

Например, индекс переменного состава себестоимости продукции

$$I_{\bar{z}} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0},$$

где z_0, z_1 – средняя себестоимость единицы продукции в базисном и текущем периодах;

q_0, q_1 – количество единиц продукции, изготовленной в базисном и текущем периодах.

Индекс себестоимости переменного состава показывает изменение средней себестоимости изделия под влиянием изменения себестоимости отдельных видов продукции и структуры выпускаемой продукции.

Индекс структурных сдвигов (*индекс влияния изменения структуры*) характеризует влияние структурных сдвигов (изменения структуры изучаемого явления) на динамику среднего уровня этого явления. В общем виде этот индекс записывается следующим образом:

$$I_{\text{стр}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \frac{\sum x_0 f_1}{\sum x_0 f_0} : \frac{\sum f_1}{\sum f_0}.$$

Если в качестве весов используется относительная величина структуры f', то

$$I_{\text{стр}} = \frac{\sum x_0 f_1'}{\sum x_0 f_0'}.$$

Например, индекс структурных сдвигов при изучении динамики себестоимости

$$I_{\text{стр}} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0}$$

показывает изменение средней себестоимости изделия под влиянием изменения структуры выпускаемой продукции.

Часто индекс структурных сдвигов вычисляется как частное от деления индекса переменного состава на индекс постоянного (фиксированного) состава.

Индекс себестоимости переменного состава показывает изменение средней себестоимости изделия под влиянием изменения себестоимости отдельных видов продукции и структуры выпускаемой продукции.

Индекс структурных сдвигов (*индекс влияния изменения структуры*) характеризует влияние структурных сдвигов (изменения структуры изучаемого явления) на динамику среднего уровня этого явления. В общем виде этот индекс записывается следующим образом:

$$I_{\text{стр}} = \frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0} = \frac{\sum x_0 f_1}{\sum x_0 f_0} : \frac{\sum f_1}{\sum f_0}.$$

Если в качестве весов используется относительная величина структуры f', то

$$I_{\text{стр}} = \frac{\sum x_0 f_1'}{\sum x_0 f_0'}.$$

Например, индекс структурных сдвигов при изучении динамики себестоимости

$$I_{\text{стр}} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0}$$

показывает изменение средней себестоимости изделия под влиянием изменения структуры выпускаемой продукции.

Часто индекс структурных сдвигов вычисляется как частное от деления индекса переменного состава на индекс постоянного (фиксированного) состава.

3. Методические указания к решению задачи № 3 по теме "Общие принципы сводки и группировки статистических данных "

Для изучения взаимосвязей между явлениями, их признаками, структуры совокупности используется статистическая группировка.

Группировка заключается в образовании групп единиц совокупности, однородных в каком-либо существенном отношении, а также имеющих одинаковые или близкие значения группировочного признака. В зависимости от задач исследования используются различные виды группировки статистической информации (табл.1).

Первичная группировка производится непосредственно по первичным данным статистического наблюдения.

Вторичная группировка используется для образования новых групп на основе ранее произведенной первичной группировки. Необходимость в перегруппировке данных возникает в тех случаях, когда первичная группировка содержит больше (или меньше) групп, чем это необходимо для характеристики типичных отношений и связей, и когда для целей сравнения необходимо получить сопоставимые данные по нескольким группировкам.

Таблица 1

Виды группировок

Признак	Вид группировки
Содержание статистической информации	Первичная Вторичная
Число группировочных признаков	Простая Комбинированная Многомерная
Задачи, решаемые с помощью группировки	Типологическая Структурная Аналитическая Территориальная

При простой группировке объединение единиц совокупности в группы производится по одному какому-либо признаку.

При комбинированной группировке производится разбиение статистической совокупности на группы по двум и более признакам, взятым в сочетании (комбинации). Сначала образуются группы по одному признаку, затем выделенные группы подразделяются на подгруппы по другому признаку, в свою очередь выделенные подгруппы разделяются на подгруппы по следующему признаку и т.д.

Многомерная группировка производится по величине средней многомерной.

Аналитическая группировка служит для выявления взаимосвязей между изучаемыми явлениями и их признаками. Взаимосвязанные признаки делятся на факторные признаки и результативные признаки. При этом группы образуются по факторному признаку, а для каждой выделенной группы рассчитывается либо среднее значение результативного признака, если он количественный, либо относительные величины, если он – качественный.

С использованием типологической группировки в изучаемой совокупности явлений выделяются однокачественные в существенном отношении группы, прежде всего классы и социально-экономические типы. Для выделения типов часто берутся не отдельные изолированные признаки, а совокупность признаков.

Структурная группировка выявляет состав (строение) однородной в качественном отношении статистической совокупности.

При территориальной группировке осуществляется распределение сводных статистических данных по экономико-географическому и административно-территориальному признакам.

В процессе решения задачи по данной теме выполняются две таблицы: разработочная (рабочая) и результативная (итоговая). В разработочной таблице должны быть представлены упорядоченные (в соответствии с установленными интервалами) данные отдельно по каждой единице совокупности и итоговые данные по группам. В результативной таблице приводятся суммарные и средние показатели по группам и по всей совокупности в целом.

Интервалы могут быть равными и неравными, закрытыми и открытыми. Закрытыми называются интервалы группировок, у которых обозначены обе границы интервалов, открытыми – интервалы, у которых указана только одна граница: верхняя – у первого, нижняя – у последнего интервала группировок. Величина равных интервалов определяется как разность между максимальным и минимальным значениями признака в совокупности, деленная на заранее заданное число образуемых групп. При обработке статистических данных открытые интервалы необходимо закрывать. Величина первого интервала принимается равной величине второго, а последнего – величине предыдущего. Для определения величины интервала группировок может быть использована формула Стерджесса (обычно при незначительной вариации признаков)

$$\Delta = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \cdot \lg N},$$

где $x_{\max}, x_{\min}$ – максимальное и минимальное значение признака ряда распределения;

N – число единиц совокупности.

При решении задачи необходимо обратить внимание на правильность оформления результативной таблицы:

в заголовке отражаются основное содержание таблицы, место и время изучаемого явления;

общие для ряда показателей названия выносятся в заголовок;

используются только общепринятые сокращения;

однотипные показатели приводятся с одинаковой степенью точности;

таблица должна содержать итоговые или средние показатели по всей статистической совокупности.

По результатам решения задачи делаются выводы об особенностях единиц совокупности по выделенным группам и о связи между признаками.

4. Методические указания к решению задачи № 4 по теме "Модели рядов динамики"

В зависимости от характера изучаемого явления, а также вида статистических данных ряды динамики подразделяются на ряды абсолютных, относительных и средних величин. Исходными являются ряды динамики абсолютных величин, которые могут быть моментными или интервальными рядами динамики. Для изучения изменения явлений во времени в рядах динамики вычисляются средняя хронологическая величина; абсолютные приросты; темпы роста; темпы прироста; абсолютное значение одного процента прироста.

Для интервального ряда средняя хронологическая:
невзвешенная

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

взвешенная

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i},$$

где y_i – i –й уровень ряда;

n – число уровней ряда;

t_i – период времени, отделяющий один уровень ряда от другого.

Абсолютный прирост представляет собой разность двух уровней ряда динамики. Абсолютный прирост характеризует скорость (в абсолютном выражении) изменения уровней ряда динамики в единицу времени.

Если производится сравнение каждого данного уровня y_i с непосредственно ему предшествующим y_{i-1}, то получаются цепные абсолютные приросты:

$$\Delta_i^c = y_i - y_{i-1}.$$

Если каждый уровень y_i сравнивается с начальным y_1 или каким-либо другим, принятым за постоянную базу сравнения y_0, то получаются базисные абсолютные приросты

$$\Delta_i^b = y_i - y_0.$$

Сумма цепных приростов за определенный период времени равна базисному абсолютному приросту за весь этот период или разности между конечным и начальным уровнями ряда динамики:

$$\sum_{i=1}^{n-1} \Delta_i^c = \sum_{i=1}^{n-1} (y_i - y_{i-1}) = \Delta_n^b = y_n - y_1.$$

Средний абсолютный прирост

$$\overline{\Delta} = \frac{\sum_{i=1}^{n-1} \Delta_i^c}{n-1} = \frac{y_n - y_1}{n-1}.$$

Темп роста представляет собой отношение двух уровней ряда динамики. Темпы роста могут быть вычислены с постоянной и переменной базой сравнения. Первые называются цепными $T_{\text{р}}^c$, вторые – базисными $T_{\text{р}}^b$:

$$T_{\text{р}i}^c = \frac{y_i}{y_{i-1}} \cdot 100,\%; \qquad T_{\text{р}i}^b = \frac{y_i}{y_0} \cdot 100,\%.$$

Обобщающим показателем темпа роста уровней ряда динамики служит их средняя величина, называемая средним темпом роста:

$$\overline{T}_{\text{р}} = \sqrt[n-1]{\prod_{i=1}^{n-1} \frac{y_{i+1}}{y_i}} = \sqrt[n-1]{\frac{y_n}{y_0}}.$$

Темп прироста представляет собой отношение абсолютного прироста к уровню динамики, по сравнению с которым он рассчитан. Для ряда динамики темпы прироста могут быть вычислены с переменной базой сравнения, они называются цепными $T_{\text{пр}}^{c}$, и с постоянной – называются базисными $T_{\text{пр}}^{b}$. Они рассчитываются следующим образом:

$$T_{\text{пр}i}^{c} = \frac{y_i - y_{i-1}}{y_{i-1}}; \qquad T_{\text{пр}i}^{b} = \frac{y_i - y_0}{y_0}.$$

Темпы прироста могут быть определены и как разность между темпом роста, выраженном в процентах, и 100%, т.е.

$$T_{\text{пр}i} = T_{\text{р}i} - 100.$$

Обобщающим показателем темпов прироста уровней ряда динамики служит их средняя величина, называемая средним темпом прироста. Он обычно вычисляется как разность между средним темпом роста и 100%.

Абсолютное значение одного процента прироста представляет собой отношение абсолютного прироста к темпу прироста, выраженному в процентах:

$$|A_i| = \frac{\Delta_i^c}{T_{\text{пр}i},\%} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}100} = 0{,}01 \cdot y_{i-1}.$$

Абсолютное значение одного процента прироста представляет собой отношение абсолютного прироста к темпу прироста, выраженному в процентах:

$$|A_i| = \frac{\Delta_i^c}{T_{\text{пр}i},\%} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}100} = 0{,}01 \cdot y_{i-1}.$$

5. Методические указания к решению задачи № 5 по теме "Корреляционно-регрессионный анализ связей социально-экономических явлений "

Одним из методов изучения корреляционных связей является корреляционно-регрессионный анализ, позволяющий определить форму связи и тесноту связи между признаками.

Если с увеличением факторного признака результативный признак равномерно возрастает или убывает, то такая зависимость является линейной и выражается уравнением прямой

$$y_x = a_0 + a_1 x = f(x, a_0, a_1),$$

где y_x – теоретическая зависимость результативного признака;

x – факторный признак;

a_0, a_1 – параметры уравнения прямой (уравнения регрессии).

Параметры уравнения прямой a_0 и a_1 определяются путем решения системы уравнений, полученных по методу наименьших квадратов:

$$\begin{cases} \sum_{i=1}^{n} y_i = n a_0 + a_1 \sum_{i=1}^{n} x_i; \\ \sum_{i=1}^{n} x_i y_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2, \end{cases}$$

где x_i, y_i – индивидуальные значения факторного и результативного признаков;

n – число индивидуальных значений признака.

Коэффициент регрессии a_1 показывает, к какому изменению средней величины результативного признака приводит изменение факторного признака на одну единицу.

Если связь между признаками нелинейная и с возрастанием факторного признака происходит ускоренное возрастание или убывание результативного признака, то корреляционная зависимость может быть выражена параболой

$$y_x = a_0 + a_1 x + a_2 x^2.$$

Значения параметров параболы a_0, a_1, a_2 определяются из решения системы нормальных уравнений

$$\begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2; \\ \sum_{i=1}^{n} x_i y_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3; \\ \sum_{i=1}^{n} y_i x_i^2 = a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4. \end{cases}$$

Если результативный признак с увеличением факторного признака возрастает (или убывает) не бесконечно, а стремится к конечному пределу, то для анализа такого признака применяется уравнение гиперболы

$$y_x = a_0 + a_1 1/x.$$

Для определения параметров этого уравнения используется система нормальных уравнений

$$\begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1 \sum_{i=1}^{n} \frac{1}{x_i}; \\ \sum_{i=1}^{n} \frac{1}{x_i} y_i = a_0 \sum_{i=1}^{n} \frac{1}{x_i} + a_1 \sum_{i=1}^{n} (\frac{1}{x_i})^2. \end{cases}$$

Для оценки адекватности уравнения регрессии используется показатель средней ошибки аппроксимации

$$\bar{\varepsilon} = \frac{1}{n} \sum \frac{\left| y_i - y_{x_i} \right|}{y_i} \cdot 100, \ \%,$$

где y_i, y_{x_i} – эмпирические (фактические) и теоретические значения результативного признака.

При линейной зависимости для определения тесноты связи между результативным и факторным признаками и оценки степени влияния факторного признака на результативный используются коэффициенты корреляции и детерминации.

Коэффициент корреляции

$$r_{xy} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sigma_x \sigma_y} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}\ \sqrt{n\sum_{i=1}^{n} y^2{}_i - (\sum_{i=1}^{n} y_i)^2}}.$$

Величина коэффициента корреляции находится в пределах от –1 до +1. Чем ближе по абсолютной величине коэффициент корреляции к 1, тем теснее связь.

Коэффициент детерминации

$$D_{xy} = r_{xy}^2$$

характеризует долю влияния факторного признака на вариацию результативного.

При нелинейных зависимостях для определения тесноты связи между результативным и факторным признаками и оценки степени влияния факторного признака на результативный используются индексы корреляции и детерминации.

Индекс корреляции

$$R_{xy} = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}},$$

где $\sigma_{y_x}^2$ – факторная дисперсия результативного признака y ;

σ_y^2 – общая дисперсия результативного признака.

Величина индекса корреляции находится в пределах от –1 до +1. Чем ближе по абсолютной величине коэффициент корреляции к 1, тем теснее связь.

Факторная дисперсия результативного признака

$$\sigma_{y_x}^2 = \frac{\sum_{i=1}^{n} (y_{x_i} - \bar{y})^2}{n},$$

где y_{x_i} – теоретические значения результативного признака (значения линии регрессии) при значениях факторного признака x_i;

$\bar{y}$ – среднее значение результативного признака.

Общая дисперсия результативного признака

$$\sigma_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n},$$

где y_i – эмпирические значения результативного признака.

Индекс детерминации

$$B_y = R_{xy}^2 = \frac{\sigma_{y_x}^2}{\sigma_y^2}$$

выражает долю факторной дисперсии в общей дисперсии, т.е. характеризует, какая часть общей вариации результативного признака y объясняется изучаемым фактором x.

6. Методические указания к решению задачи № 6 по теме "Выборочный метод в статистических исследованиях"

При статистическом исследовании экономических явлений часто производится наблюдение не всех единиц генеральной совокупности, а лишь их части, и по этой части (выборке) судят о всей совокупности в целом.

Отбор единиц из генеральной совокупности производится таким образом, чтобы выборочная совокупность была представительна (репрезентативна) и наиболее полно и адекватно представляла свойства генеральной совокупности.

При отборе объектов из генеральной совокупности используются следующие методы: индивидуальный, групповой (серийный), комбинированный, повторный (возвратный), бесповторный (безвозвратный), одноступенчатый, многоступенчатый, собственно-случайный, механический и типический отбор.

В основе теории выборочного наблюдения лежат теоремы законов больших чисел, которые позволяют решить два взаимосвязанных вопроса выборки: рассчитать ее объем при заданной точности исследования и определить ошибку при данном объеме выборки.

Расхождение между значениями изучаемого признака выборочной и генеральных совокупностей является ошибкой репрезентативности (представительности). Она может быть систематической и случайной.

Систематическая ошибка репрезентативности возникает вследствие нарушения правил отбора единиц генеральной совокупности, в частности принципа беспристрастного, непреднамеренного отбора. Случайная ошибка возникает в силу того, что выборочное статистическое наблюдение является несплошным наблюдением. Определение случайной ошибки производится по формулам ошибки выборочной доли и ошибки выборочной средней.

Выборочная доля представляет собой отношение числа единиц, обладающих данным признаком или данным его значением m к общему числу единиц выборочной совокупности n

$$w = \frac{m}{n}.$$

Ошибка выборочной доли представляет собой расхождение (разность) между долей в выборочной совокупности w и долей в генеральной совокупности p, возникающее вследствие несплошного характера наблюдения. Величина ошибки выборочной доли определяется как предел отклонения w от p, гарантируемый с заданной вероятностью:

$$|w - p| < \tau\mu_w,$$

где τ – гарантийный коэффициент, зависящий от вероятности P_τ, с которой гарантируется невыход разности $w - p$ за пределы $\tau\mu_w$; μ_w – средняя ошибка выборочной доли.

Значения гарантийного коэффициента τ и соответствующие им вероятности P_τ приведены в табл. 2. Обычно вероятность принимается равной 0,9545 или 0,9973, а τ при этом равно 2 и 3.

Средняя ошибка выборочной доли определяется по формуле

$$\mu_w = \sqrt{\frac{\sigma_w^2}{n}},$$

где σ_w^2 – дисперсия выборочной доли.

Таблица 2

Значения гарантийного коэффициента τ

τ	P_τ	τ	P_τ	τ	P_τ
1,00	0,6827	1,70	0,9109	2,40	0,9836
1,10	0,7287	1,80	0,9281	2,50	0,9876
1,20	0,7699	1,90	0,9426	2,60	0,9907
1,30	0,8064	2,00	0,9545	2,70	0,9931
1,40	0,8385	2,10	0,9643	2,80	0,9949
1,50	0,8664	2,20	0,9722	2,90	0,9963
1,60	0,8904	2,30	0,9786	3,00	0,9973

Для выборочной доли дисперсия определяется по формуле

$$\sigma_w^2 = w(1-w).$$

Приведенная формула средней ошибки выборочной доли применяется при повторном отборе.

При бесповторном отборе численность генеральной совокупности сокращается, поэтому дисперсия умножается на коэффициент $1-\frac{n}{N}$. Формулы расчета средних ошибок выборочной доли для различных способов отбора единиц из генеральной совокупности приведены в табл. 3.

Таблица 3

Формулы расчета средних ошибок выборочной доли и выборочной средней

Метод отбора выборки	**Средняя ошибка**	
	выборочной доли	**выборочной средней**
Механический и собственно–случайный повторный отбор	$\mu_w = \sqrt{\frac{\sigma_w^2}{n}}$	$\mu_x = \sqrt{\frac{\sigma_x^2}{n}}$

Продолжение табл. 3

Метод отбора выборки	**Средняя ошибка**	
	выборочной доли	**выборочной средней**
Механический и собственно–случайный бесповторный отбор	$\mu_w = \sqrt{\frac{\sigma_w^2}{n}(1-\frac{n}{N})}$	$\mu_x = \sqrt{\frac{\sigma_x^2}{n}(1-\frac{n}{N})}$
Серийный отбор при бесповторном отборе серий	$\mu_w = \sqrt{\frac{\delta_w^2}{r}(1-\frac{r}{R})}$	$\mu_x = \sqrt{\frac{\delta_x^2}{r}(1-\frac{r}{R})}$
Типический отбор при повторном случайном отборе внутри групп	$\mu_w = \sqrt{\frac{\overline{\sigma_w^2}}{n}}$	$\mu_x = \sqrt{\frac{\overline{\sigma_x^2}}{n}}$
Типический отбор при бесповторном случайном отборе внутри групп	$\mu_w = \sqrt{\frac{\overline{\sigma_w^2}}{n}(1-\frac{n}{N})}$	$\mu_x = \sqrt{\frac{\overline{\sigma_x^2}}{n}(1-\frac{n}{N})}$

В табл. 3 приняты следующие обозначения:

N – численность генеральной совокупности;

δ_w^2 – межсерийная дисперсия выборочной доли;

r – число отобранных серий;

R – число серий в генеральной совокупности;

$\overline{\sigma_w^2}$ – средняя из групповых дисперсий;

σ_x^2 – дисперсия признака x ;

δ_x^2 – межсерийная дисперсия выборочных средних;

$\overline{\sigma_x^2}$ – средняя из выборочных дисперсий типических групп.

Дисперсии в формулах расчета средних ошибок выборочной доли в табл. 3 рассчитываются следующим образом:

– межсерийная дисперсия выборочной доли

$$\delta_w^2 = \frac{\sum_{j=1}^{r}(w_j - \overline{w})^2}{r},$$

где w_j – выборочная доля в j –й серии;

$\overline{w}$ – средняя величина доли во всех сериях;

– средняя из групповых дисперсий

$$\overline{\sigma_w^2} = \overline{w(1-w)} = \frac{\sum_{j=1}^{k} w_j(1-w_j)n_j}{\sum_{j=1}^{k} n_j},$$

где w_j – выборочная доля в j –й типической группе;

n_j – число единиц в j –й типической группе;

k – число типических групп.

Предельное значение ошибки выборочной доли определяется по следующей формуле:

$$\Delta_w = \mu_w \tau.$$

Доля альтернативного признака в генеральной совокупности равна

$$p = w \pm \Delta_w.$$

Ошибка выборочной средней представляет собой расхождение (разность) между выборочной средней $\widetilde{x}$ и генеральной средней $\overline{x}$, возникающее вследствие несплошного выборочного характера наблюдения. Величина ошибки выборочной средней определяется как предел отклонения $\widetilde{x}$ от $\overline{x}$, гарантируемый с заданной вероятностью

$$|\tilde{x} - \bar{x}| < \tau\mu_x,$$

где μ_x – средняя ошибка выборочной средней.

При повторном отборе средняя ошибка определяется следующим образом:

$$\mu_x = \sqrt{\frac{\sigma_x^2}{n}},$$

где σ_x^2 – средняя величина дисперсии количественного признака x, которая рассчитывается по формуле средней арифметической невзвешенной

$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})^2}{n}$$

или средней арифметической взвешенной

$$\sigma_x^2 = \frac{\sum_{i=1}^{n}(x_i - \tilde{x})^2 f_i}{\sum_{i=1}^{n} f_i},$$

где f_i – статистический вес.

Формулы расчета средней ошибки выборочной средней для различных способов отбора выборочной совокупности приведены в табл. 3.

Межсерийная дисперсия выборочных средних δ_x^2 и средняя из выборочных дисперсий типических групп $\overline{\sigma_x^2}$ вычисляются следующим образом:

$$\delta_x^2 = \frac{\sum_{j=1}^{r}(\tilde{x}_j - \tilde{x})^2}{r};$$

$$\sigma_x^2 = \frac{\sum_{j=1}^{k} \sigma_j^2 n_i}{\sum_{j=1}^{k} n_i},$$

где $\widetilde{x}_j$ – среднее значение показателя в j – й серии;

σ_j^2 – дисперсия признака x в j – й типической группе;

n_j – число единиц в j –й типической группе.

Предельная ошибка выражается следующим образом:

$$\Delta_x = \mu_x \tau \, .$$

Средняя величина количественного признака в генеральной совокупности определяется с учетом предельной ошибки выборочной средней

$$\overline{x} = \widetilde{x} \pm \Delta_x .$$

Определение необходимого объема выборки n основывается на формулах предельных ошибок выборочной доли и выборочной средней. Например, для повторного отбора предельные ошибки равны

$$\Delta_w = \tau \sqrt{\frac{w(1-w)}{n}};$$

$$\Delta_x = \tau \sqrt{\frac{\sigma_x^2}{n}}.$$

Отсюда объемы выборок для расчета выборочной доли n_w и выборочной средней n_x следующие:

$$n_w = \frac{\tau^2 w(1-w)}{\Delta_w^2};$$

$$n_x = \frac{\tau^2 \sigma_x^2}{\Delta_x^2}.$$

Аналогичным образом определяются объемы выборок при различных способах отбора выборочной совокупности. Для серийного отбора определяется число отобранных серий. Формулы расчета приведены в табл. 4.

Таблица 4

Формулы расчета объема выборки

Метод отбора выборки	**Объем выборки или число серий для определения**	
	выборочной доли	**выборочной средней**
Механический и собственно-случайный повторный отбор	$n_w = \frac{\tau^2 w(1-w)}{\Delta_w^2}$	$n_x = \frac{\tau^2 \sigma_x^2}{\Delta_x^2}$
Механический и собственно-случайный бесповторный отбор	$n_w = \frac{\tau^2 w(1-w)N}{\Delta_w^2 N + \tau^2 w(1-w)}$	$n_x = \frac{\tau^2 \sigma_x^2 N}{\Delta_x^2 N + \tau^2 \sigma_x^2}$
Серийный отбор при бесповторном отборе серий	$r_w = \frac{\tau^2 \delta_w^2 R}{\Delta_w^2 (R-1) + \tau^2 \delta_w^2}$	$r_x = \frac{\tau^2 \delta_x^2 R}{\Delta_x^2 (R-1) + \tau^2 \delta_x^2}$
Типический отбор при повторном случайном отборе внутри групп	$n_w = \frac{\tau^2 \overline{\sigma_w^2}}{\Delta_w^2}$	$n_x = \frac{\tau^2 \overline{\sigma_x^2}}{\Delta_x^2}$

Типический отбор при бесповторном случайном отборе внутри групп	$n_w = \dfrac{\tau^2 \overline{\sigma_w^2} N}{\Delta_w^2 N + \tau^2 \overline{\sigma_w^2}}$	$n_x = \dfrac{\tau^2 \overline{\sigma_x^2} N}{\Delta_x^2 N + \tau^2 \overline{\sigma_x^2}}$

В табл. 4 приняты следующие обозначения:

n_w, n_x – объемы выборок для определения ошибок выборочной доли и выборочной средней;

r_w, r_x – число отобранных серий для определения ошибок выборочной доли и выборочной средней;

Δ_w, Δ_x – предельные ошибки выборочной доли и выборочной средней.