

1. Статистическое исследование производства AAA на Территории (в РФ по Федеральным округам и областям) за Период (с хх по ххх).

- 1.1. **Цель исследования:**
Выявить закономерности и особенности производства AAA на Территории за Период.
- 1.2. **Задачи исследования:**
Изучить структуру производства AAA на Территории и ее динамику за Период.
Изучить особенности распределения производства AAA на Территории за Период, оценить значения показателей распределения.
Выявить наличие и оценить проявление взаимосвязей значений отдельных показателей производства AAA на Территории за Период.
Изучить динамику производства AAA на Территории за Период.
- 1.3. **Предмет исследования:**
Структура производства AAA на Территории, распределение значений показателей, их динамика за Период.
- 1.4. **Период наблюдения:** Период с хх по ххх производства AAA на Территории.
- 1.5. **Объект наблюдения:** Производство AAA на Территории за Период.
- 1.6. **Единицы наблюдения:** Единицы Территории, производящие AAA, в отдельные временные интервалы / моменты за Период с хх по ххх.
- 1.7. **Статистическая единица:** Госкомстат РФ.
2. **Статистические признаки единиц наблюдения, подлежащих исследованию.**
 - 2.1. **Статистическая конкретизация признаков:**
Полные наименования соответствующих абсолютных, относительных и средних показателей, характеризующих объект наблюдения и его отдельные элементы.
Характеристика особенностей их регистрации.
 - 2.2. **Результатирующие признаки единиц наблюдения:**
Абсолютные показатели объема производства AAA на отдельной единице Территории в отдельные временные интервалы / моменты за Период, измеренные в (ед. изм.).
Представленные значения - результат сводки и группировки. Они характеризуют суммарный (полученный сложением) результат производства AAA всеми производителями на отдельной единице Территории в отдельные временные интервалы / моменты за Период с хх по ххх.
Обладает свойством аддитивности, суммируемости:
при сложении значений признака у двух единиц наблюдения, получается валовой, итоговый показатель объема производства для этих двух наблюдений.
 - 2.3. **Факторные признаки единиц наблюдения:**
 - 2.3.1. **Названия единиц Территории,** производящих AAA (атрибутивные признаки).
Отдельные единицы Территории могут обладать свойством вложенности (области входят в Федеральные округа), отражающим территориальную структуру объекта наблюдения.
 - 2.3.2. **Значения отдельных временных интервалов / моментов за Период,** для которых собираются сведения о производстве AAA на Территории.
Они задают структуру наблюдения динамики производства на Периоде.

3. Статистическая структура исходных данных.

- 3.1. **Описание структуры данных.**
Представленные данные – результат сводки и группировки данных от производителей по территориальным признакам и динамическим.
- 3.2. **Логический и содержательный контроль данных с точки зрения статистических свойств рассматриваемой совокупности:** (опечатки, пропуски значений, большие отклонения – выбросы значений);
если исходные представляют сгруппированные данные, (переизбыточность наблюдений доминантной группы, наличие нескольких малозначимых групп и др.).
Комментарии и рекомендации по реагированию на обнаруженные особенности.
4. **Ряды распределения по каждому из наблюдаемых признаков.**
Ряды распределения представляются в табличной или графической форме.
Они отражают, какое значения признака соответствует данному наблюдению.

5. **Сводка и первичная группировка данных по значениям вариантов результирующего показателя.**
Это структурная группировка - разделение однородной совокупности на группы по варьирующему признаку.
Возможны другие группировки.
Типологическая группировка – разделение качественно разнородной совокупности на качественно однородные группы, например, территориальная.
Аналитическая группировка – разделение совокупности по факторному признаку, например, динамическая.
Обоснование количества групп, их размеров и выбора границ для каждой группы.
6. **Для каждой группы можно рассчитать показатели:**
 - ✓ Объемные показатели численности/ размера совокупностей.
 - ✓ Объемные валовые, граничные показатели количественных признаков объектов наблюдения.
 - ✓ Показатели вариации, изменчивости значений признака;
 - ✓ Средние значения количественных признаков объектов наблюдения для единицы наблюдаемой совокупности;
 - ✓ Порядковые и ранговые показатели значений признаков, измеренных в шкалах, устанавливающих порядок следования, и некоторые другие показатели.

и оценить уровень ее однородности относительно значений варьирующего признака.
- 6.1. **а) Объемный показатель численности.** Абсолютный показатель, как правило, натуральный, не именованный.
Количество единиц наблюдения, относящихся к рассматриваемой группе: f .
б) При выборочном наблюдении совокупности интервально или точно на некоторой Генеральной совокупности, единицы измерения соответствуют размерности замеров для Генеральной совокупности.
в) При наблюдении во времени, Генеральная совокупность – Период. Единица замера наблюдения либо момент времени, либо временной интервал.
Объемный показатель численности для динамических наблюдений – Период наблюдения.
- 6.2.1. **Объемный, валовой показатель вариант (x) признака в группе численности f:** X. Абсолютный показатель, единицы измерения, как у варианты. $X = \sum x = x_1 + x_2 + \dots + x_f$.
- 6.2.2. **Доля варианты x в объемном показателе X:** d. Относительный показатель. $d = \frac{x}{X}$.
- 6.2.2. **Граничные показатели вариант (x) признака.** Абсолютный показатель, единицы измерения, как у варианты.
 - а) **Граничные показатели вариант (x) признака по наблюдениям:** x_{\min} и x_{\max} – минимальное и максимальное значение вариант.
Все значения вариант в группе из интервала $[x_{\min}, x_{\max}]$.
 - б) **Граничные показатели вариант (x) признака аналитически,** на основе теоретических представлений о граничных значениях вариант в группе k: x_{k-1} и x_k – границы интервала значений вариант. Все значения вариант в группе из интервала (x_{k-1}, x_k) .
- 6.2.3. **Расстояние варианты x от фиксированного числа a:** δ_a . Абсолютный показатель, неотрицательный, единицы измерения, как у варианты. $\delta_a = |x - a|$.
- 6.2.4. **Квадратичное расстояние варианты x от фиксированного числа a:** δ_a^2 . Абсолютный показатель, неотрицательный, единицы измерения – квадрат единицы измерения варианты. $\delta_a^2 = (x - a)^2$.
- 6.2.5. **Отклонение отдельной варианты от фиксированного числа a:** Абсолютный показатель, единицы измерения, как у варианты. $s_a = (x - a)$.
- 6.3. **Показатели вариации, изменчивости значений признака.**
 - 6.3.1. **а) Размах: R, показатель максимальной вариации** значений признака. Абсолютный показатель, неотрицательный, единицы измерения, как у варианты.
Разность верхней и нижней границ признака.
$$R = x_{\max} - x_{\min}$$

б) **Длина интервала значений вариант в группе** Δ_x . Абсолютный показатель, неотрицательный, единицы измерения, как у варианты. $\Delta_x = x_k - x_{k-1}$.

6.3.2. Дисперсия или квадратичная вариация признака: σ^2 . Абсолютный показатель, неотрицательный, единицы измерения - квадрат единицы измерения варианты. $\sigma^2 = \frac{\sum (x - \bar{x})^2}{f} = \frac{\sum \delta_x^2}{f}$,

где f - количество наблюдений в исследуемой группе; \bar{x} - среднее значение варианты, δ_x^2 - квадратичное расстояние варианты x от \bar{x} .

6.3.3. Простое среднеквадратическое отклонение варианты от своего среднего или стандартное отклонение: σ . Абсолютный показатель, неотрицательный, единицы измерения, как у варианты.

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{f}}, \text{ или } \sigma = \sqrt{\sigma^2} - \text{корень квадратный из дисперсии признака.}$$

6.3.4. Линейное отклонение варианты от своего среднего: σ_x . Абсолютный показатель, неотрицательный, единицы измерения, как у варианты.

$$\sigma_x = \frac{\sum |x - \bar{x}|}{f} = \frac{\delta_x^2}{f}.$$

6.3.5. Оценка средней вариации признака: σ_{cp} . Абсолютный показатель, неотрицательный, единицы измерения, как у варианты.

$$\sigma_{cp} = \frac{R}{6} = \frac{x_{\max} - x_{\min}}{6}.$$

6.3.6. Коэффициент вариации: K_{var} . Относительный показатель, неотрицательный, измерение в %.

$$K_{var} = \frac{\sigma}{\bar{x}} \times 100\%.$$

6.4. Средние значения варьирующего показателя:

6.4.1. Среднее арифметическое простое: \bar{x} . Среднее значение, единицы измерения, как у варианты.

$$\text{Усреднение валового показателя } X \text{ по численности наблюдений } f: \bar{x} = \frac{X}{f} = \frac{\sum x}{f} = \frac{x_1 + x_2 + \dots + x_f}{f}.$$

Свойства среднего арифметического:

а) Основное свойство \bar{x} : Если в совокупности наблюдений все значения вариант x признака заменить на значение \bar{x} , то валовой показатель признака X останется неизменным:

$$X = \sum x = x_1 + x_2 + \dots + x_f = \sum \bar{x} = f \times \bar{x} = f \bar{x} \text{ или } X = f \bar{x}.$$

б) Этот показатель чувствителен к большим отклонениям.

При замене в однородной совокупности в отдельном наблюдении значения признака на существенно большее значение, величина \bar{x} отреагирует увеличением своего значения.

в) Если в тексте приводится термин среднее значение, то имеется в виду среднее арифметическое.

6.4.2. Середина интервала значений варианты x_c . Среднее значение, единицы измерения, как у варианты.

$$x_c = (x_{\min} + x_{\max}) / 2.$$

6.4.3. Медиана варьирующего признака: (M_e). Абсолютный показатель, единицы измерения, как у варианты.

Свойства медианы.

а) Медиана - это срединное значение упорядоченной последовательности вариант.

б) Этот показатель устойчив к большим отклонениям.

При замене в однородной совокупности в отдельном наблюдении значения признака на существенно большее значение, величина медианы не изменится, либо изменится незначительно.

в) Медиана совпадает с порядковым показателем варианты уровня 0,5. Т.е. $M_e = L_{0,5}$.

6.5. Порядковые показатели уровня α или квантили уровня α : (L_α). Абсолютный показатель, единицы измерения, как у варианты.

L_α - это максимальная величина варьирующего показателя y ($\alpha \times 100\%$) самых "мелких" элементов совокупности, если размер элемента определяется значением варьирующего признака.

6.5.1. Медиана: $M_e = L_{0,5}$. Абсолютный показатель, единицы измерения, как у варианты.

6.5.2. Первый дециль: $L_{0,1}$. Абсолютный показатель, единицы измерения, как у варианты.

6.5.3. Десятый дециль: $L_{0,9}$. Абсолютный показатель, единицы измерения, как у варианты.

6.5.4. Квантили: $L_{0,25}$ и $L_{0,75}$. Абсолютные показатели, единицы измерения как у варианты.

6.7. Степень однородности

Качественная оценка однородности группы осуществляется при сопоставлении различных показателей вариации признака для группы и других показателей, а так же анализом графического представления значений варьирующего признака для изучаемой группы наблюдений.

7. Вариационные ряды.

7.1. Табличное представление вариационного ряда.

7.1.1. Дискретный вариационный ряд.

7.1.2. Интервальный вариационный ряд.

7.2. Показатели строк (групп) вариационного ряда.

Частота - численность наблюдений в группе k : f_k . **Накопленная частота:** $F_k = F_{k-1} + f_k$.

Частость, доля численности в группе относительно численности совокупности: p_k . **Накопленная частость:** $P_k = P_{k-1} + p_k$.

Плотность численности наблюдений f на единицу длины интервала значений вариант Δ : y .

$$y = \frac{f}{\Delta}.$$

7.3. Графики вариационного ряда:

7.1.1. Гистограмма плотностей (варианта x) / (плотность численности $y = \frac{f}{\Delta}$), где Δ - длина интервала, а f - частота.

7.1.2. Гистограмма частот (или частостей) (варианта x / частота f (или частость $p = \frac{f}{F}$ численности), где F - накопленная частота, совпадающая с численностью наблюдений.

7.1.3. Кумулята (варианта x / накопленная частость P численности)

7.1.4. Линия Лоренца (накопленная доля (частость) P численности / накопленная доля объема D варианты)

7.2. Показатели вариационного ряда:

7.2.1. Средний размер варьирующего признака

$$(\text{среднее арифметическое взвешенное варианты } x), \quad \bar{x} = \frac{\sum (x \cdot f)}{F} = \sum (x \cdot p).$$

7.2.2. Максимальное и минимальное значение варьирующего признака

7.2.3. Размах вариации (максимальное значение - минимальное значение варианты)

7.2.4. Типичное значение / Мода (M_o) варьирующего признака аналитически (на модальном интервале, на котором плотность принимает наибольшее значение). M_o = начало интервала x_{Mod-1} +

длина интервала Δ_{Mod} , помноженная на долю $\times \frac{a}{a+b}$, где $a = y_{Mod} - y_{Mod-1}$, величина превышения

плотности модального интервала над плотностью предшествующего интервала, и $b = y_{Mod} - y_{Mod+1}$, величина превышения плотности модального интервала над плотностью последующего интервала), т.е.

$$M_0 = x_{Mod-1} + \Delta_{Mod} \times \frac{a}{a+b} = x_{Mod-1} + \Delta_{Mod} \times \frac{(y_{Mod} - y_{Mod-1})}{(y_{Mod} - y_{Mod-1}) + (y_{Mod} - y_{Mod+1})}$$

7.2.5. **Типичное значение / Мода (M_0)** варьирующего признака **графически** (по гистограмме плотностей путем дополнительных построений на модальном интервале).

7.2.6. **Медиану (M_d)** варьирующего признака **аналитически** (на медианном интервале, на котором накопленная частота P_{Med} впервые превышает значение 0,5).

M_e = начало интервала x_{Med-1} + длина интервала Δ_{Med} , помноженная на долю

$$\frac{0,5 - P_{Med-1}}{P_{Med} - P_{Med-1}} = \frac{0,5 - P_{Med-1}}{P_{Med}}, \text{ т.е.}$$

$$M_e = x_{Med-1} + \Delta_{Med} \times \frac{0,5 - P_{Med-1}}{P_{Med} - P_{Med-1}} = x_{Med-1} + \Delta_{Med} \times \frac{0,5 - P_{Med-1}}{P_{Med}}$$

7.2.7. **Медиану (M_d)** варьирующего признака **графически** (по кумуляте определяется минимальное значение варианты, при котором накопленная частота P принимает значение 0,5).

7.2.8. **Уровень дифференциации** элементов совокупности по варьирующему признаку **графически** и **аналитически** (на основе коэффициентов дифференциации):

$$\text{Фондового. Относительный показатель, единицы измерения (\%)} \quad K_{\text{фонд}} = \frac{d_{zL_0}}{d_{zL_1}} \times 100\%,$$

$$\text{Квартильного. Относительный показатель, единицы измерения (\%)} \quad K_{\text{кварт}} = \frac{L_{0,75}}{L_{0,25}} \times 100\%,$$

$$\text{Децильного. Относительный показатель, единицы измерения (\%)} \quad K_{\text{дец}} = \frac{L_{0,9}}{L_{0,1}} \times 100\%,$$

где L_q - минимальное значение варианты, при котором накопленная частота P принимает значение q , на графике кумуляты, или другими словами L_q - минимальное значение варианты, при котором доля наблюдений не превосходящих L_q не превышает q ;

d_{zL_1} (соответственно d_{zL_0}) - доля валового объема варианты для наблюдений с значениями варианты не превышающими L_1 (соответственно не меньшими чем L_1).

В некоторых случаях более содержательным представляется показатель, когда значение d с соответствующим индексом, рассматривается не как доля валового объема варианты, а как среднее значение варианты для наблюдений с соответствующими значениями признака.

Степень концентрации варьирующего признака в обследуемой совокупности (на основе коэффициента Герфиндаля);

$$H = \sum (d_k)^2$$

где d_k - доля валового объема варианты группы k в объеме для всей совокупности, $d_k = \frac{x_k}{\sum x}$.

1) аналитически и графически L_9 - минимальную величину варьирующего показателя у 10% самых крупных элементов совокупности, L_1 - максимальную величину варьирующего показателя у 10% самых мелких элементов совокупности оценивают методами аналогичными для определения M_e ;

2) аналитически и графически абсолютную и относительную численность элементов совокупности, имеющих значение варьирующего показателя не ниже модального уровня форму распределения на основе коэффициента асимметрии по Пирсону. Относительный показатель, единицы измерения (%).

$$K_{ac} = \frac{\bar{x} - M_0}{\sigma} \times 100\%.$$

8. Перейдя от абсолютной шкалы к порядковой, определите для каждого периода времени доминантную (60 - 80)%, приоритетную и малозначимую группу единиц, проанализируйте структурные сдвиги за весь период наблюдения на основе коэффициента неравномерности распределения.

Коэффициент неравномерности распределения:

$$k_{\text{неравн}} = \frac{K * L}{K - L} * \sum (w_i - p)^2, \quad 0 < k_{\text{неравн}} < 1$$

p - доля групп при абсолютно равномерном распределении, $p = \frac{1}{K}$;

K - число групп в сравниваемых структурах,

L - число доминантных групп, принимается равным L_{max} по всем сравниваемым структурам для обеспечения сопоставимости полученных по каждой структуре коэффициентов неравномерности

$0 < k_{\text{неравн}} < 1$,

9. Проведите логический и содержательный контроль результатов группировки для каждого года наблюдения.

10. Проанализируйте аналитически и графически динамику обобщающих показателей распределения элементов совокупности по объёму варьирующего признака X . Сравните по годам наблюдения среднесезонные значения темпов прироста варьирующего показателя, абсолютных приростов и абсолютного содержания 1% прироста. $T_{\text{прироста}} = T_{\text{роста}} - 100$. $T_{\text{роста}} = K_{\text{роста}} * 100\%$.

11. Проведите выравнивание динамических рядов типичного и медианного значения варьирующего показателя методом скользящей средней. Обоснуйте выбор числа интервалов осреднения. Изобразите результаты графически и сделайте выводы относительно динамики обобщающих показателей в течение рассматриваемого периода.

12. Проведите выравнивание динамического ряда, построенного на основе децильного коэффициента дифференциации методом скользящей средней, изобразите результаты графически и сделайте выводы относительно изменения уровня дифференциации элементов совокупности по величине варьирующего показателя в течение рассматриваемого периода.

13. Проведите динамическую группировку данных для начального и конечного уровней ряда наблюдения за динамикой показателя. Постройте шахматную таблицу, характеризующую зависимость величины варьирующего показателя от фактора времени и сделайте выводы относительно тесноты этой зависимости на основе критерия χ^2 .

Критерий χ^2 :

$$\chi^2_{\text{факт}} = \sum \frac{\tilde{y}_i - \tilde{f}_i}{\tilde{f}_i}$$

$\chi^2_{\text{табл}}$ при $\alpha=0,05$ и $df=(m-2)-1$, где m - число групп

14. Для начального и конечного уровней ряда наблюдения за динамикой показателя сравните структуру совокупности по величине варьирующего показателя графически и аналитически (на основе среднего квадратического коэффициента Казинца и интегральных коэффициентов структурных различий Салаи и Гатева). Сопоставьте полученные на основе различных коэффициентов результаты. Оцените изменение приоритетов.

Линейный коэффициент структурных сдвигов Л. Казинца [%]

$$\Delta_{w_1-w_0} = \frac{\sum |w_1 - w_0|}{k}, \quad k - \text{число групп}$$

Квадратический коэффициент структурных сдвигов Л. Казинца [%]

$$\sigma_{w_1-w_0} = \sqrt{\frac{\sum (w_1 - w_0)^2}{k}}, \quad k - \text{число групп}$$

Показывает на сколько удельный вес групп (по группировочному признаку) различается в сравниваемых структурах.

3. **Интегральный коэффициент структурных сдвигов К. Гитева:**

$$k = \sqrt{\frac{\sum (w_1 - w_0)^2}{\sum (w_1^2 + w_0^2)}}, \quad 0 < k < 1$$

учитывает интенсивность изменений по отдельным группам и удельный вес групп в сравниваемых структурах.

4. **Интегральный коэффициент структурных различий Салаи:**

$$I = \sqrt{\frac{\sum ((w_1 - w_0)/(w_1 + w_0))^2}{k}}, \quad 0 < I < 1, \quad k - \text{число групп}$$

15. Сравните уровень концентрации и однородности структур по состоянию на начало и конец периода наблюдения и оцените существенность различий дисперсий в сравниваемых структурах, используя критерий Бартлетта.

$$1) \text{ Коэффициент Лоренца } L = \sum |d - p|/2,$$

где d – доля i-й группы в общем объеме признака, в кратном отношении,

p – частость i-й группы в общем объеме признака, в кратном отношении

$0 < L < 1$; $L \rightarrow 0 \Rightarrow$ равномерное распределение

Для проверки нулевой гипотезы H_0 о равенстве дисперсий в нескольких структурах самым мощным критерием считается **критерий Бартлетта** (несколько дисперсий одновременно, не ограничен попарными сравнениями)

$$M = |\ln(\bar{\sigma}_{арифм}) * \sum n - (\bar{\sigma}_{geom})|$$

Порядок расчета критерия при $n > 50$:

1. Вычисляется средняя арифметическая из дисперсий сравниваемых групп

$$\bar{\sigma}_{арифм} = \frac{\sum x_i * n_i}{\sum n_i}$$

2. Находим натуральный логарифм средней арифметической

$$\ln(\bar{\sigma}_{арифм})$$

3. Определяется $(\sum n) * \ln(\bar{\sigma}_{арифм})$

4. Вычисляется средняя геометрическая из дисперсий сравниваемых групп

$$(\bar{\sigma}_{geom}) = \sum (n_i * \ln x_i)$$

5. Определяется $M = |\ln(\bar{\sigma}_{арифм}) * \sum n - (\bar{\sigma}_{geom})|$

6. Отношение M/C распределено как χ^2 с числом степеней свободы $df=m-1$, где m – число групп.

16 Проведите логический и содержательный контроль результатов анализа по п. п. 5 – 15. Проведите нормативный анализ и интерпретацию полученных показателей и сформулируйте выводы.

17. Продумайте возможные перспективы развития исследования.

Готовая оформленная исследовательская работа студента должна представлять собой завершенную, самостоятельную научную работу, включающую описание всех этапов

проведённого учащимся полноценного статистического исследования: от организации наблюдения до интерпретации данных и формулировки выводов.

В основной текст готовой работы должны быть включены следующие элементы, соответствующие последовательным этапам статистического исследования:

1. Цель исследования и ее актуальность, предмет исследования, период исследования, основные задачи исследования.
2. Описание используемой учащимся базы данных: предмета и объекта исследования, единицы наблюдения и статистической единицы.
3. Описание используемой системы показателей с пояснением подхода к конкретизации признаков для статистического применения, а также методов экономико-статистического анализа с обоснованием их применимости к исследуемой базе данных.
4. Обоснование и описание построенных группировок.
5. Основные результаты первичной и вторичной обработки данных в форме обобщённых таблиц, пригодных для визуального анализа, а также описание результатов проведённого анализа представленных в основном тексте таблиц. **Внимание!** Каждая таблица должна иметь номер, название и подробные заголовки в подлежащем и сказуемом, позволяющие однозначно определить наименование используемых в таблице показателей.
6. Необходимый графический материал в виде рисунков (графиков различного вида) и описание визуального анализа. **Внимание!** Каждый рисунок должен иметь номер, название и подробные заголовки осей, позволяющие однозначно определить наименование используемых при построении графика показателей.
7. Интерпретация полученных результатов на основе нормативных теоретических знаний, полученных студентом в результате всего предшествующего обучения.
8. Обобщающее заключение по исследованию в целом с выделением основных полученных выводов.
9. Возможные перспективы дальнейшего исследования.
10. Список использованной литературы.

Использованные формулы, проведённые расчёты и промежуточные таблицы должны быть вынесены за пределы основного текста, в приложение, чтобы обеспечить преподавателю возможность проверки, как самих расчётов, так и интерпретации показателей.

Объём основного текста работы должен составлять не менее 12 – 15 страниц (шрифт 12 pt, одинарный интервал).