



Национальный исследовательский университет
«Высшая школа экономики»

Факультет экономических наук

Образовательная программа: Экономика

Название дисциплины: Основы статистических наблюдений

Отчёт о выполнении домашнего задания №1

«Статистический анализ мирового ВВП
и динамика экономики Китая (1960–2022)»

Выполнил
Студент группы БЭК 251
Ярёменко Степан Алексеевич

Москва 2025



1.

Тема моего исследования – применение методов описательной статистики для анализа мирового и национального ВВП.

2.

Первая задача моего исследования – найти и проанализировать пространственные данные о ВВП стран мира за 2022 год.

Современная экономика глобализирована, и распределение ВВП по странам отражает расстановку мировых экономических сил. Анализ пространственных данных позволяет увидеть, где сосредоточены ключевые центры производства и потребления.

2022 год — особенный период в мировой экономике:

- последствия пандемии COVID-19,
- геополитическая напряжённость и изменение цепочек поставок,
- структурные сдвиги в мировой торговле и энергетике.

Поэтому исследование пространственного распределения ВВП именно за этот год позволяет выявить новые тенденции в мировой экономической географии.

Пространственный анализ ВВП помогает не просто сравнивать страны по уровню производства, но и оценивать степень неравномерности развития, концентрацию богатства и экономическую мощь отдельных регионов (США, ЕС, Китай, Юго-Восточная Азия и др.).

Такие данные являются основой для прогнозов и стратегий развития: они используются в международной политике, инвестиционном анализе, при формировании макроэкономической и внешнеэкономической политики государств.

Вторая задача – проанализировать динамику и статистические характеристики ВВП Китая за период 1960–2022 гг. с использованием методов описательной статистики.

Китай — одно из крупнейших и наиболее быстро развивающихся государств мира. С середины XX века его экономика прошла путь от аграрной к индустриальной и затем — к технологически развитой, что делает анализ её динамики уникально показательным.

Длительный период наблюдения (1960–2022 гг.) охватывает ключевые этапы экономического развития Китая:



- начало реформ и «открытия» в конце 1970-х годов,
- ускоренный рост в 1990–2010-х,
- замедление темпов и переход к инновационной модели в 2020-х.

Это позволяет выявить долгосрочные тенденции и структурные переломы в экономике.

Использование методов описательной статистики (средние, дисперсия, асимметрия, эксцесс и др.) даёт возможность количественно охарактеризовать процесс экономического роста — оценить устойчивость, волатильность и степень концентрации значений.

Полученные результаты служат основой для экономической интерпретации: позволяют определить, насколько устойчив был рост ВВП Китая, какие периоды характеризовались резкими изменениями, и как формировалась современная экономическая мощь страны.

3.

Для анализа я использовал показатель: GDP (current US\$).

Это номинальный ВВП, то есть стоимость всех конечных товаров и услуг, произведённых в стране за год, в текущих рыночных ценах.

Единицы измерения – доллары США (USD) в миллиардах долларов (billions of current US\$). Данные не скорректированы на инфляцию (в отличие от real GDP или GDP (constant 2015 US\$)).

4.

Источники: World Bank national accounts data и Kaggle Datasets.

5.

Валовой внутренний продукт (ВВП, Gross Domestic Product, GDP) — это ключевой макроэкономический показатель, который характеризует итоговую рыночную стоимость всех конечных товаров и услуг, произведённых в стране за определённый период времени (обычно за год).

Показатель ВВП является основным индикатором экономической активности и масштаба национальной экономики. Он используется для сравнения экономического потенциала стран, анализа темпов роста, а также для оценки уровня благосостояния населения.



6.

Все расчёты по исследованию проводились на платформе colab.research.google.com, на языке python, с использованием как популярных модулей **numpy**, **pandas**, так и специализированного статистического пакета **scipy.stats**. Графики построены с использованием пакетов **seaborn** и **matplotlib.pyplot**.

Ссылка на блокнот:

https://colab.research.google.com/drive/1MqBmsz06NwSE0wFITWD3Sxd84dYoYmyM?usp=drive_link

Датасеты доступны для загрузки в репозитории проекта по адресу:

<https://github.com/sayaryomenko/statistic/>

Дальнейшее описание исследования делится на две части, соответствующие пространственным и временным данным.

Пространственные данные

Датасет GDP-by-Country-1999-2022 имеет следующий вид:

Country	1999	2000	2001	2002	2003	2004	2005	2006	2007	20
Afghanistan	0	0	0	4.084	4.585	5.971	7.309	8.399	9.892	11.
Albania	3.444	3.695	4.096	4.456	5.6	7.452	8.376	9.133	10.163	11.
Algeria	48.845	54.749	55.181	57.053	68.013	85.016	102.38	114.322	116.158	12.
Angola	6.153	9.135	8.936	11.386	13.956	19.8	30.632	43.759	55.37	67.
Antigua and Barbuda	0.652	0.678	0.71	0.718	0.754	0.818	0.875	0.962	1.026	1.0
Argentina	283.523	284.204	268.697	97.732	127.643	151.958	181.549	212.702	247.096	27
Armenia	1.845	1.912	2.118	2.376	2.807	3.573	4.903	6.41	7.666	8.5
Australia	401.782	390.017	368.191	412.901	527.588	639.356	712.436	754.816	822.089	85.
Austria	213.39	194.407	193.345	208.566	255.842	293.194	305.338	321.934	348.659	36.
Azerbaijan	4.581	5.273	5.708	6.236	7.276	8.682	12.561	19.817	28.699	40.
Bahamas	4.704	5.004	5.131	5.389	5.503	5.661	5.87	6.223	6.624	7.0
Bahrain	5.517	7.066	7.037	8.446	8.734	11.18	12.378	15.065	17.505	18.

Из этих данных мы программно получаем нужный нам столбец из 180 строк:

```
2022
0    0.000
1   18.012
2  168.195
3   74.953
4    1.534
...
```



После обработки этих данных статистическими функциями и методами мы получили следующие результаты:

Минимальное значение (x_{\min}) = 0.0
Максимальное значение (x_{\max}) = 24796.08
Размах (R) = 24796.08
Среднее значение (mean) = 556.7953555555555
Мода (mode) = 0.0
Медиана (median) = 40.186
Дисперсия (variance) = 5702244.923001037
Стандартное отклонение (СКО, std) = 2387.937378366744
Первый квартиль (Q1) = 8.885000000000002
Третий квартиль (Q3) = 305.18025
Интерквартильный размах (IQR) = 296.29525
Коэффициент вариации = 4.288716409970991
Коэффициент асимметрии (skewness) = 8.31903412812399
Коэффициент эксцесса (kurtosis) = 74.5777857967683

Даже без графиков и дополнительных вычислений, этот набор статистических метрик уже позволяет сделать некоторые выводы.

ВВП варьируется от микроскопических экономик (например, малые островные государства) до крупнейших экономик мира (США, Китай).

Даже если отбросить нулевые (в нашем случае 0 соответствует пропущенным значениям), видно, что размах (R) почти равен максимальному значению, что подтверждает сильную асимметрию вправо – множество малых значений и несколько огромных «гигантов» тянут распределение вверх.

Среднее значение в 14 раз выше медианы.

Это явный признак правосторонней асимметрии – большая часть стран имеет относительно низкий ВВП, но несколько экономических лидеров (США, Китай, Япония и др.) резко увеличивают среднее значение.

Вывод: Среднее не отражает «типичный» уровень ВВП страны, а медиана более показательная для описания большинства стран.

Следующим шагом мной были подсчитаны выбросы в исходном ряду.:

Q1 = 8.885000000000002
Q3 = 305.18025
IQR = 296.29525
 $U_c (-)$ для $1.5 \cdot IQR$ = -435.557875
 $U_c (+)$ для $1.5 \cdot IQR$ = 749.6231250000001
 $U_c (-)$ для $3 \cdot IQR$ = -880.00075
 $U_c (+)$ для $3 \cdot IQR$ = 1194.066
Выбросы ($1.5 \cdot IQR$), штук = 21
Выбросы ($3 \cdot IQR$), штук = 16



Для лучшего понимания я дополнительно вывел список «стран-выбросов». Страны-выбросов по критерию $3 \cdot IQR$:

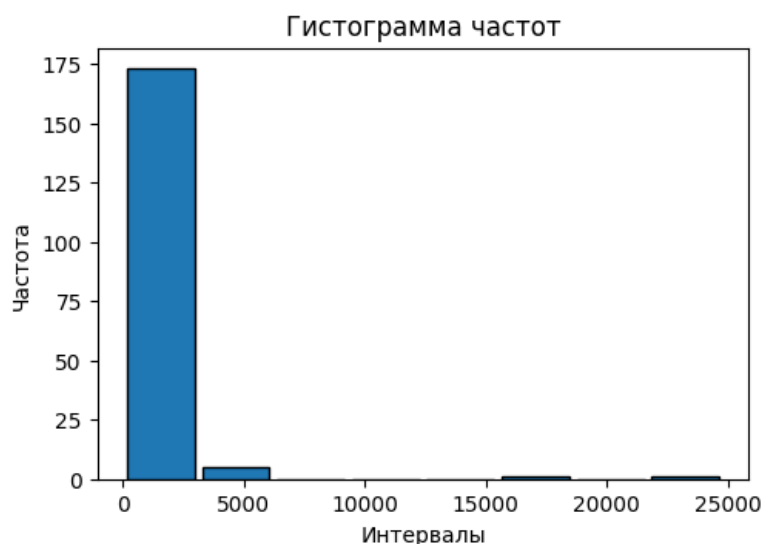
	Country	2022
7	Australia	1677.45
22	Brazil	1810.61
29	Canada	2189.79
34	China	18463.13
57	France	3140.03
61	Germany	4557.35
74	India	3250.08
75	Indonesia	1247.35
79	Italy	2272.27
81	Japan	5383.68
86	Korea	1907.66
106	Mexico	1371.64
131	Russia	1703.53
145	Spain	1570.91
170	UK	3442.21
171	US	24796.08

Россия тоже в числе стран-выбросов!

Следующим этапом исследования была подготовка к построению гистограммы. Для этого нам первым делом необходим интервальный ряд. Для нахождения числа интервалов я воспользовался [критерием \(формулой\) Стёрджера](#):

$$k = 1 + 3.322 \cdot \log_{10}(n), \quad \text{где } n \text{ — количество измерений.}$$

Не стану приводить полученные значения в табличном виде (можно посмотреть в блокноте), а сразу перейду к получившейся гистограмме:





Мы видим, что нашей гистограммы большие проблемы:

Диапазон ВВП слишком большой — значения стран варьируются от миллионов до десятков триллионов долларов и из-за этого распределение получается очень скошенным.

Используется линейная шкала, из-за чего выбросы (США, Китай и т.д.) «вытягивают» ось. Без логарифма почти все страны сгруппировались в первых интервалах, а несколько богатейших «вытянули» шкалу, делая остальные различия невидимыми.

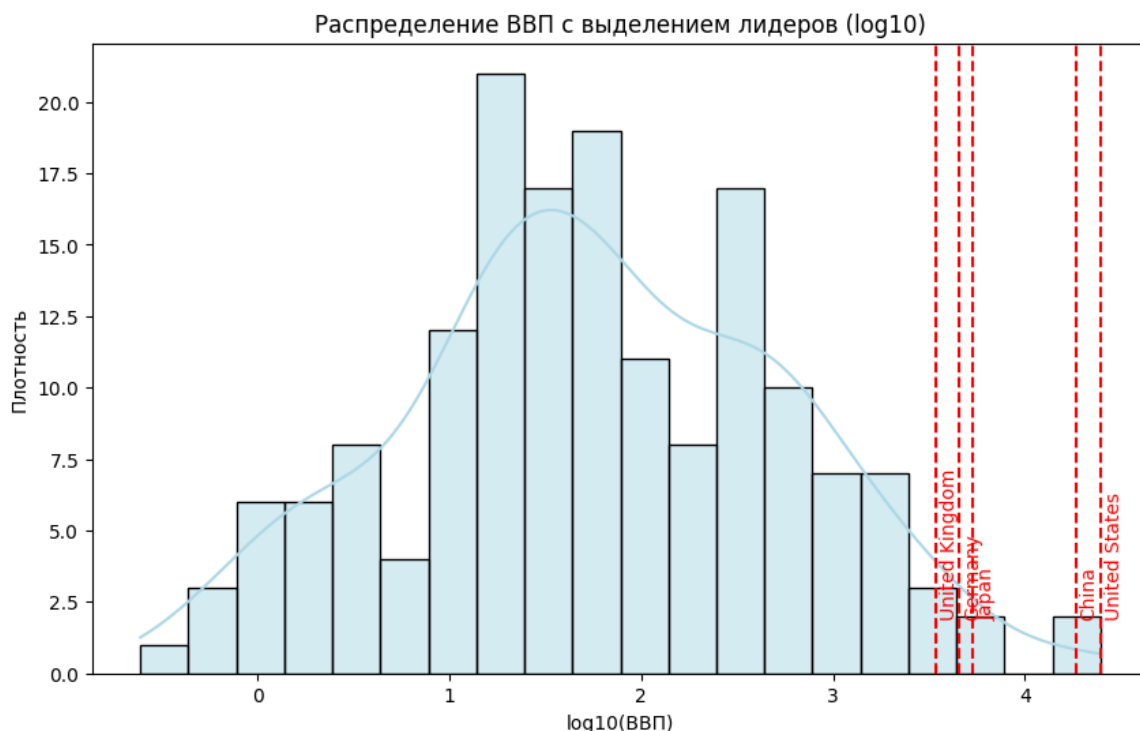
Поэтому попробуем использовать логарифмическую шкалу. ВВП различается в сотни раз между странами значит, применим логарифм с основанием 10.

Данные получились гораздо лучше, покажу их таблично:

$h = 0.626$

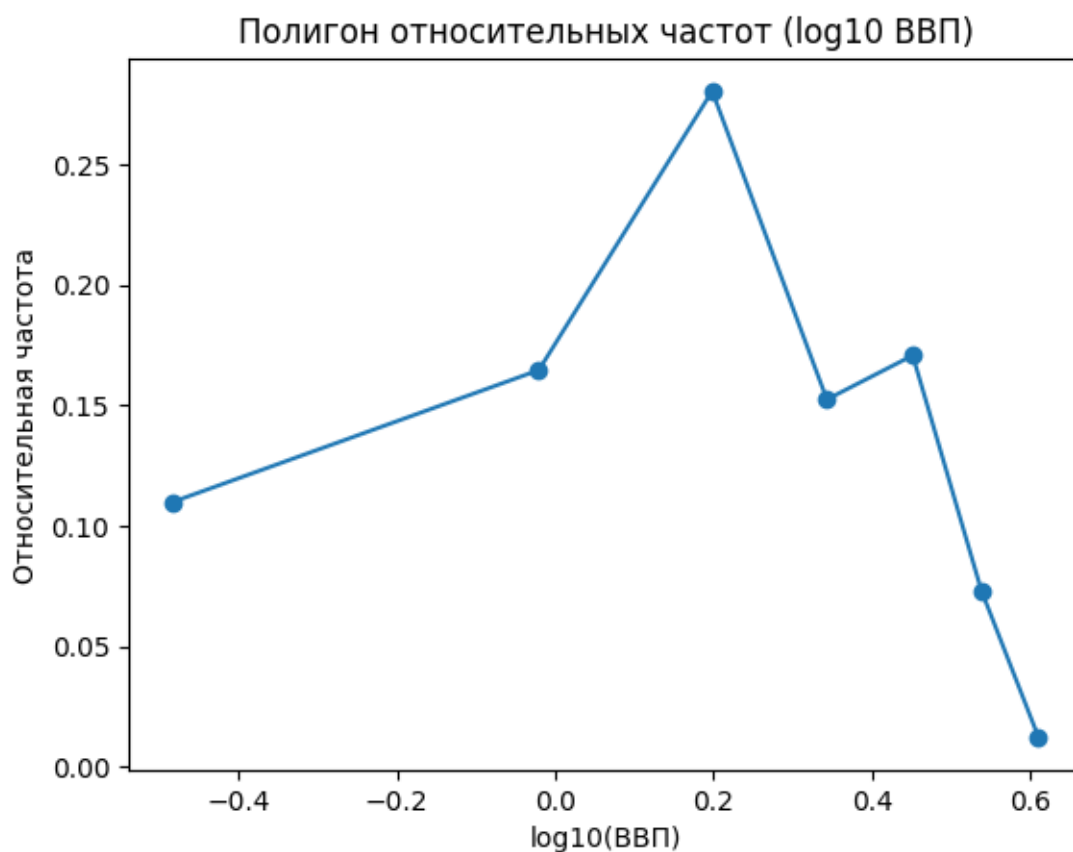
Интервалы (\log_{10})	Частота	Середины интервалов (\log_{10})	Накопленная частота	$(x - \text{среднее})^2 * f$	$(x - \text{среднее})^3 * f$	$(x - \text{среднее})^4 * f$
0	6	-0.298008	6	25.260831	-51.831743	106.351594
1	18	0.327644	24	36.613353	-52.218333	74.474312
2	27	0.953296	51	17.304112	-13.852944	11.090085
3	46	1.578948	97	1.407234	-0.246133	0.043050
4	25	2.204601	122	5.079305	2.289478	1.031974
5	28	2.830253	150	32.441739	34.920236	37.588086
6	12	3.455905	162	34.763712	59.169594	100.709637
7	2	4.081557	164	10.836399	25.223915	58.713773

Теперь гистограмма станет показывать реальные кластеры стран (бедные, средние, богатые), а не просто одну «гору слева»:

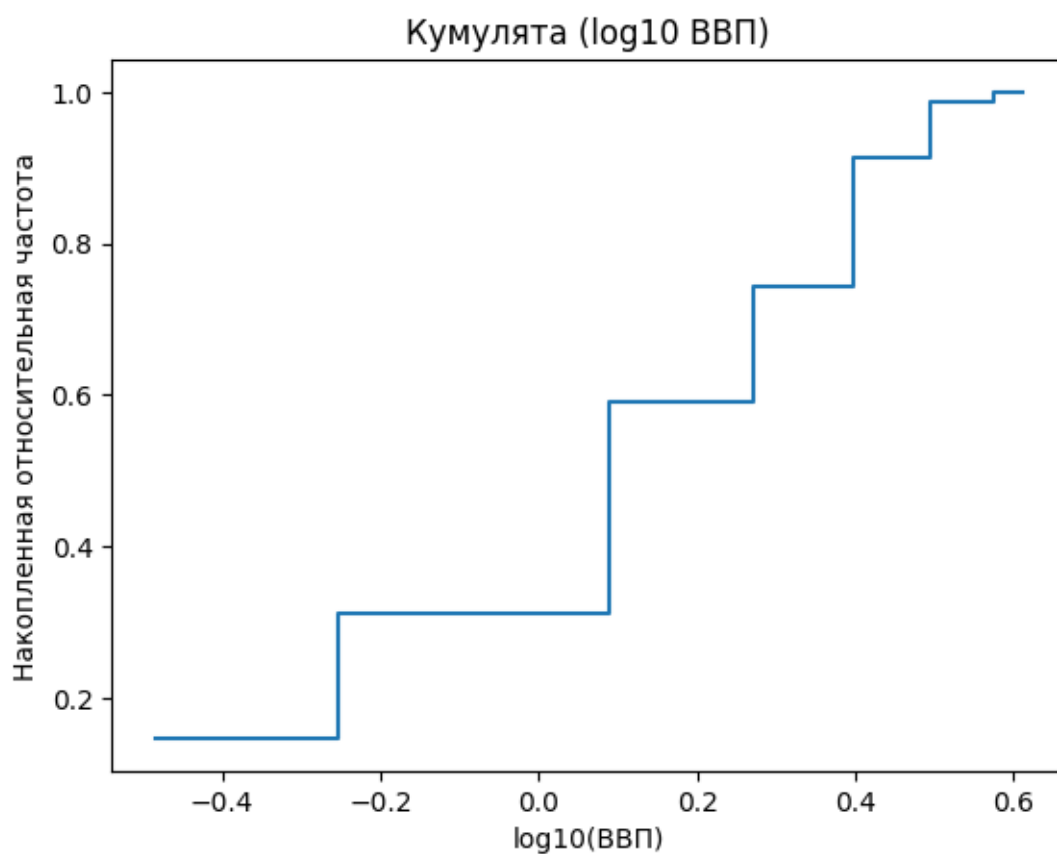




Лучше понять распределение помогли следующие графики:



Линия полигона помогает легче показывает нам асимметрию, пики и хвосты распределения, чем обычная гистограмма.

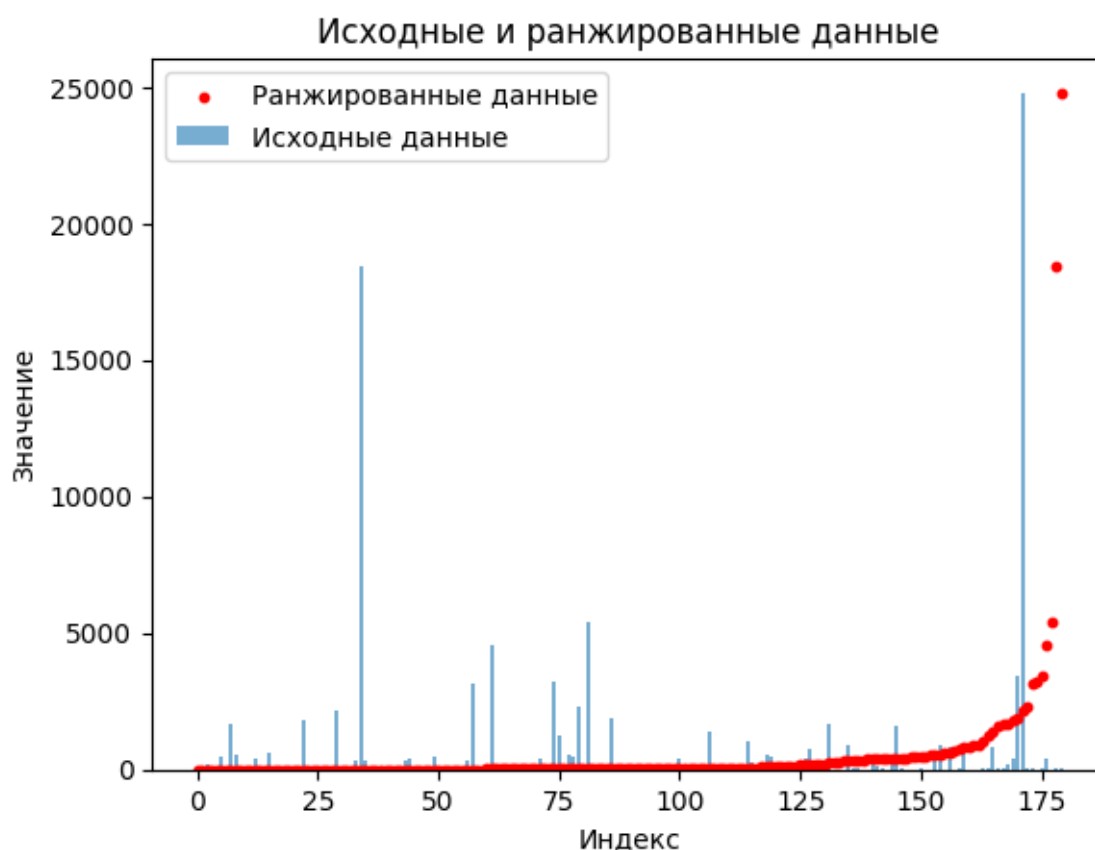




Здесь участок с плавным ростом – это множество стран с низким и средним ВВП.

Резкий подъём в конце – это несколько стран с очень высоким ВВП (США, Китай, Япония и др.).

Высокие «ступеньки» – отдельные страны сильно выделяются по масштабу ВВП.



Ранжированные данные отобразим без логарифмирования. Здесь как раз нужно видеть реальную диспропорцию. В логарифме теряется ощущение того, насколько огромен разрыв между странами.

Общие выводы:

Распределение ВВП крайне неравномерное, даже после логарифмирования: большинство стран сосредоточено на низких значениях, а несколько лидеров «поднимают хвост» графика.

Мы увидели концентрацию экономической мощи в руках нескольких стр



Временные данные

Для этих целей предыдущий датасет не годится (требование 30 уровней и более).

Поэтому я взял данные с меньшим выбором по странам, но с большим временным интервалом: Countries GDP 1960-2020:

https://www.kaggle.com/datasets/rinichristy/countries-gdp-19602020?utm_source=chatgpt.com&select=Countries+GDP+1960-2020.csv

Данные можно посмотреть [репозитории](#) проекта.

Из этих данных я выбрал для исследования Китай, как самую большую и динамичную экономику мира.

Метрики

Я извлёк данные из общей таблицы и сохранил их в обычный список. Метрики распределения получились следующие:

Минимальное значение (x_{\min}) = 47209359006.0

Максимальное значение (x_{\max}) = 14700000000000.0

Размах (R) = 14652790640994.0

Среднее значение (mean) = 2576957252390.3276

Мода (mode) = 47209359006.0

Медиана (median) = 361000000000.0

Дисперсия (variance) = 1.7804006467061449e+25

Стандартное отклонение (СКО, std) = 4219479407114.277

Первый квартиль (Q1) = 150000000000.0

Третий квартиль (Q3) = 2290000000000.0

Интерквартильный размах (IQR) = 2140000000000.0

Коэффициент вариации = 1.6373882039371754

Коэффициент асимметрии (skewness) = 1.7358498430672726

Коэффициент эксцесса (kurtosis) = 1.6115109512293113

Выводы из этого:

Минимум (≈ 47 млрд) и максимум ($\approx 14,7$ трлн) демонстрируют колоссальный рост ВВП за рассматриваемый период. Данные охватывают практически весь путь Китайской экономики от начала реформ до современной сверхдержавы.

Размах ($\approx 1,46 \times 10^{13}$) показывает очень большую вариативность — значения ВВП изменяются на несколько порядков.



Среднее значение ($\approx 2,58$ трлн) значительно больше медианы (≈ 361 млрд), что указывает на асимметрию вправо — т.е. последние, высокие значения сильно вытягивают среднее.

Мода = минимум означает, что самые маленькие значения встречались чаще — вероятно, первые годы наблюдений.

Коэффициент вариации $> 1,6$ указывает на крайнюю нестабильность и высокую изменчивость данных. Это типично для периода быстрого экономического роста.

Коэффициент асимметрии $\approx 1,74$ подтверждает сильное смещение вправо: большинство значений ниже среднего, но отдельные большие значения (поздние годы) поднимают распределение.

Коэффициент эксцесса $\approx 1,61$ (> 0). Значения сильно сосредоточены вокруг низов, с тяжёлым "длинным хвостом" больших наблюдений.

Интерквартильный размах ($\approx 2,14$ трлн) и СКО ($\approx 4,22$ трлн) также свидетельствуют об очень широкой дисперсии и неравномерности роста.

Выбросы

Наши данные это ВВП Китая за 1960–2020, то есть временной ряд с ярко выраженным ростом.

ВВП Китая рос экспоненциально, особенно после 1980-х.

Из-за этого формула с IQR почти наверняка посчитает «выбросами» последние годы с высоким ВВП, хотя это реальный тренд, а не ошибка данных.

Вывод:

Искать выбросы не будем.

Интервалы и гистограмма

ВВП Китая растёт экспоненциально (особенно после 1980-х).

Если построить гистограмму по исходным значениям, то большая часть интервалов будет пустая или почти пустая для первых 40 лет. А последние 5–10 лет будут сильно выделяться.

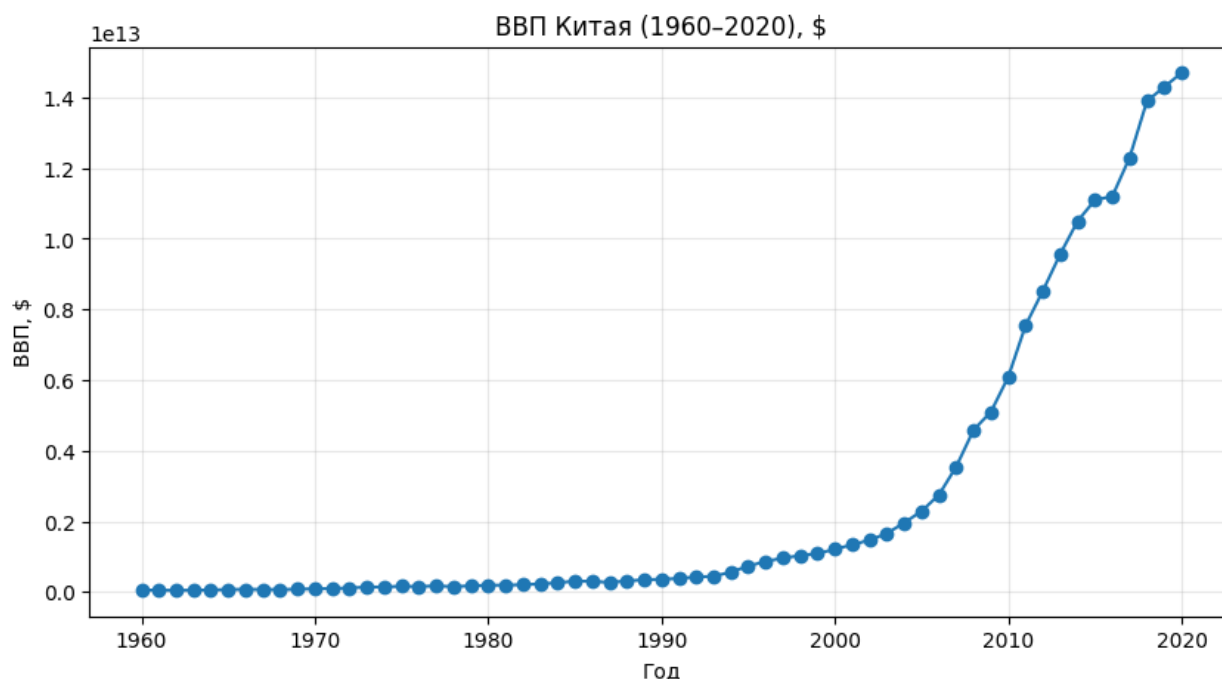
Гистограмма будет неинформативной, почти все столбцы будут «пустыми», кроме верхних.

Вывод:

Строить таблицу интервалов и гистограмму не будем.

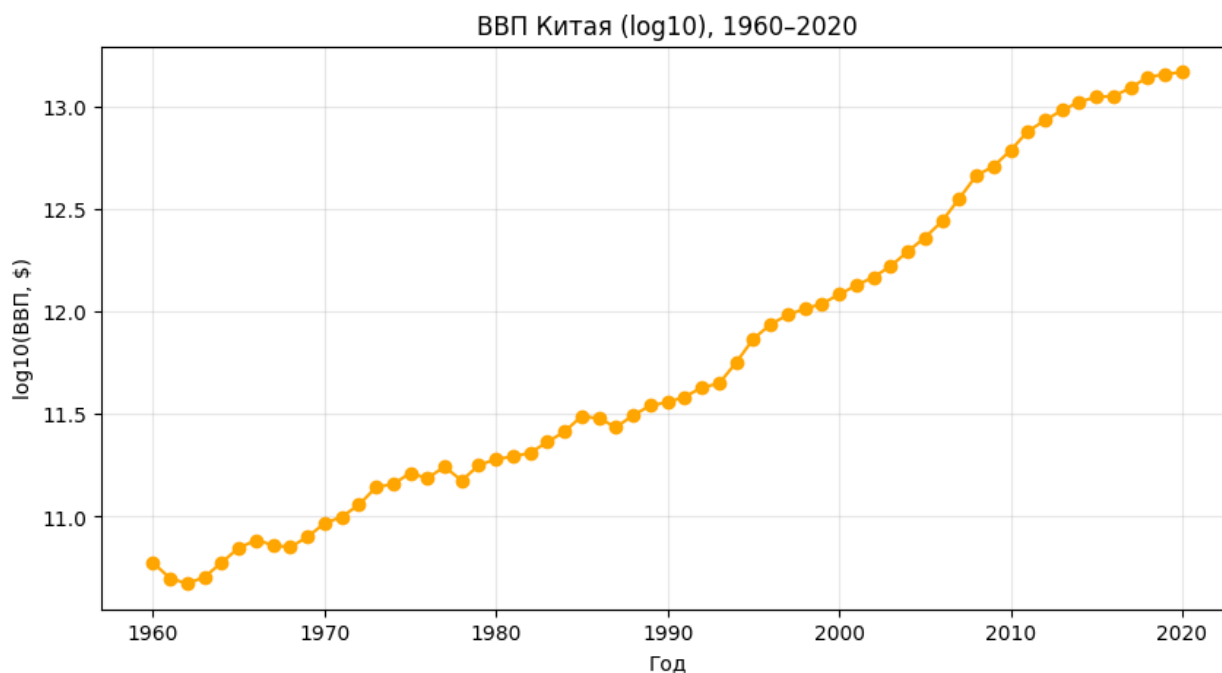


Линейный график ВВП



Линейный график показывает абсолютный рост ВВП по годам. Сразу видно, что в правой части график сильно наклонён. Попробуем прологарифмировать.

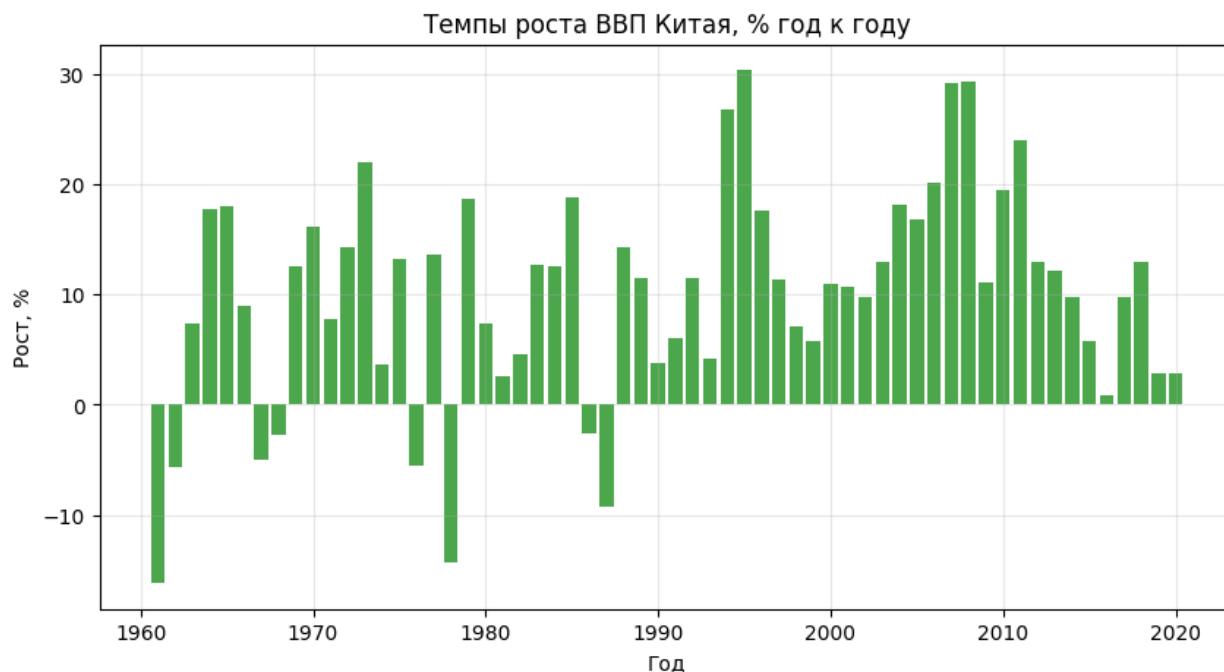
Логарифмический график ВВП



Теперь мы видим рост в относительных масштабах (линейный тренд, если рост экспоненциальный). На таком графике хорошо видны ускорения и замедления роста.

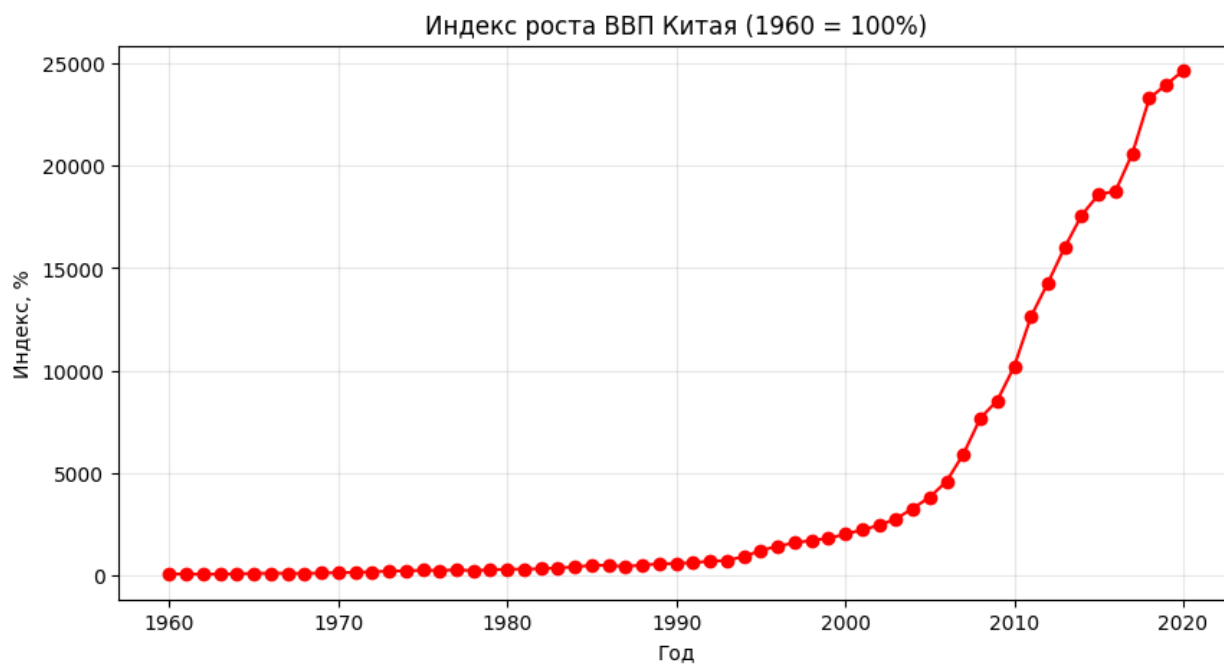


Темпы роста (год к году)



Это процентное изменение ВВП относительно предыдущего года. Здесь хорошо видны экстремальные изменения (спады, кризисы).

Кумулятивный рост (индекс роста)



Здесь видим, во сколько раз ВВП увеличился относительно базового года.



Общие выводы:

ВВП современного Китая демонстрирует экспоненциальную динамику роста: на ранних этапах развития значения малы и плотны, затем резкий подъём поднимает верхнюю границу в десятки раз.

Распределение неравномерное, с правой асимметрией характерно для стран с долгосрочным быстрым ростом экономики.