


Работа с собственными экспериментальными данными





Тот, кто серьезно думает о перспективах своей деятельности, обязательно будет накапливать информацию об окружающем мире, пытаясь выделить закономерности из случайностей, чтобы опереться на них в своих действиях, и он будет постоянно искать стабильные и обоснованные критерии выбора, позволяющие стандартизировать процесс принятия решений.

Почему не стоит бояться математики?

Математики смотрят на математику как на область исследования, где они могут делать все, что хотят и как они хотят. Остальные из нас используют математику как точный язык для выражения отношений между величинами реального мира и в качестве инструмента для получения количественных выводов из этих отношений.





Самое главное - правильно выбрать метод анализа и четко понимать следующее:

- Как формулируется тот вопрос, на который Вы хотите ответить с помощью статистического анализа?

2) Какие статистические процедуры наиболее адекватны для поиска ответа на данный вопрос?

3) Как следует интерпретировать результаты?

**Исследование должно
правильно планироваться,
а результаты - правильно
интерпретироваться.**

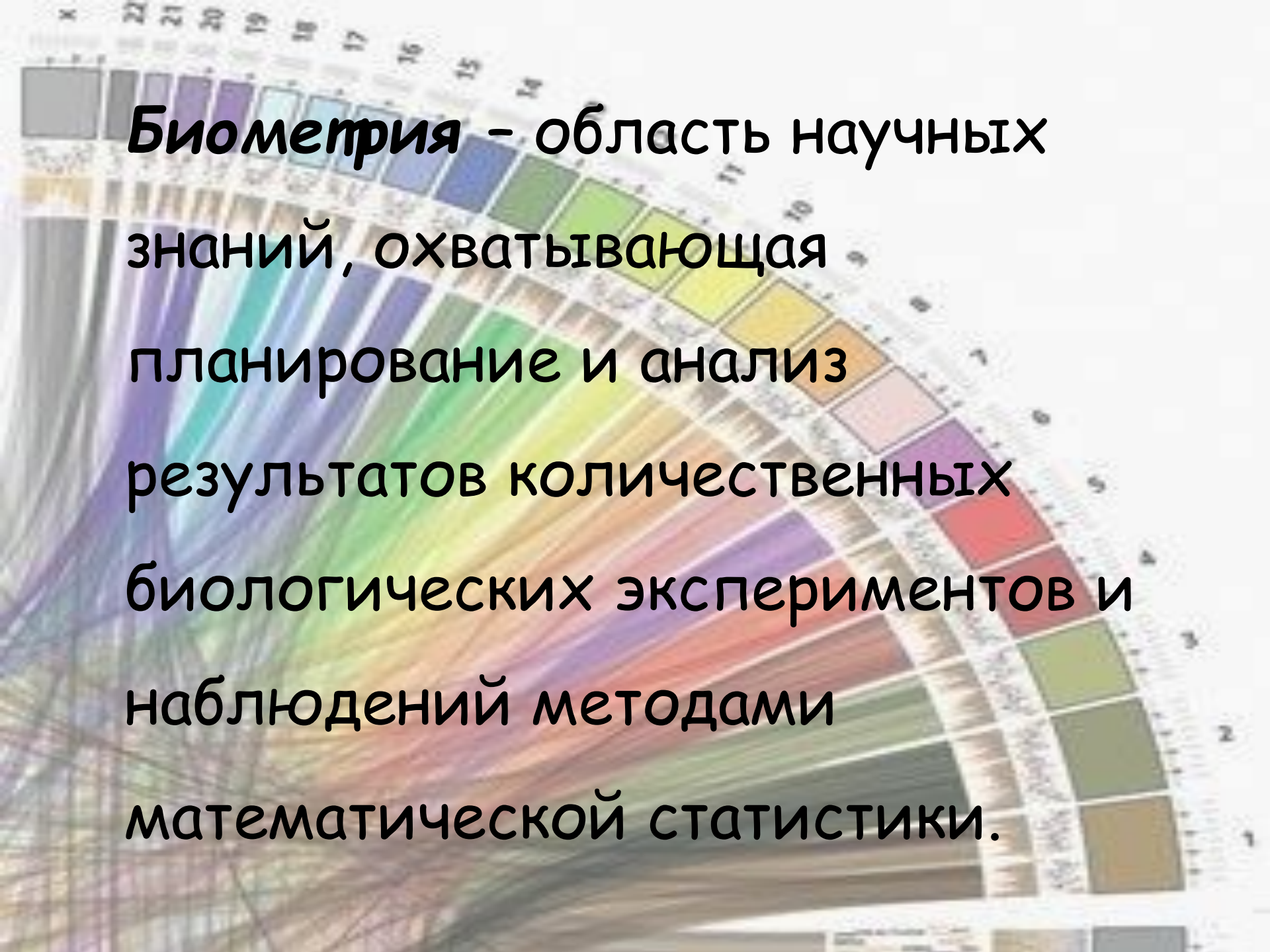
До недавнего времени вся медицинская практика представляла то, что Lewis Thomas описывал как фривольное и безответственное экспериментирование на человеке, основанное ни на чем другом, кроме проб и ошибок, которые происходят именно в такой последовательности.

Математическая статистика –
раздел математики, посвященный
математическим методам
систематизации, обработки и
использования статистических
данных для научных и
практических выводов.

Задачи математической статистики

1) кратко описать большой массив данных;

2) проверить некую статистическую гипотезу.



Биометрия - область научных знаний, охватывающая планирование и анализ результатов количественных биологических экспериментов и наблюдений методами математической статистики.

Разделы математической статистики

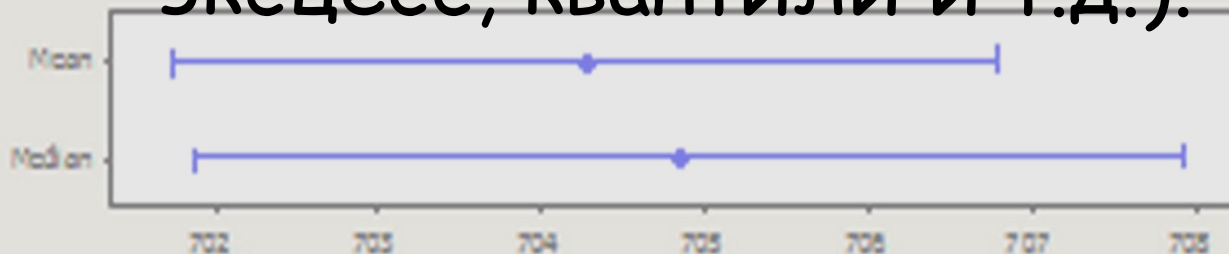
Описательная статистика

позволяет рассчитать основные параметры, характеризующие собранную информацию в целом:

среднее значение, разброс значений, или стандартное отклонение,

характеристики формы распределения данных (мода, медиана, асимметрия, эксцесс, квантили и т.д.).

A-Squared	0,30
P-Value	0,562
Mean	704,27
StDev	8,93
Variance	79,66
Skewness	-0,454499
Kurtosis	0,062888
N	50
Minimum	681,05
1st Quartile	689,34
Median	704,86
3rd Quartile	710,97
Maximum	722,56
95% Confidence Interval for Mean	701,73 706,80
95% Confidence Interval for Median	701,88 707,92
95% Confidence Interval for StDev	7,46 11,12





1

2

3

**Критерии различий
оценивают различия между двумя
(или более) совокупностями
данных. Результат применения**



4

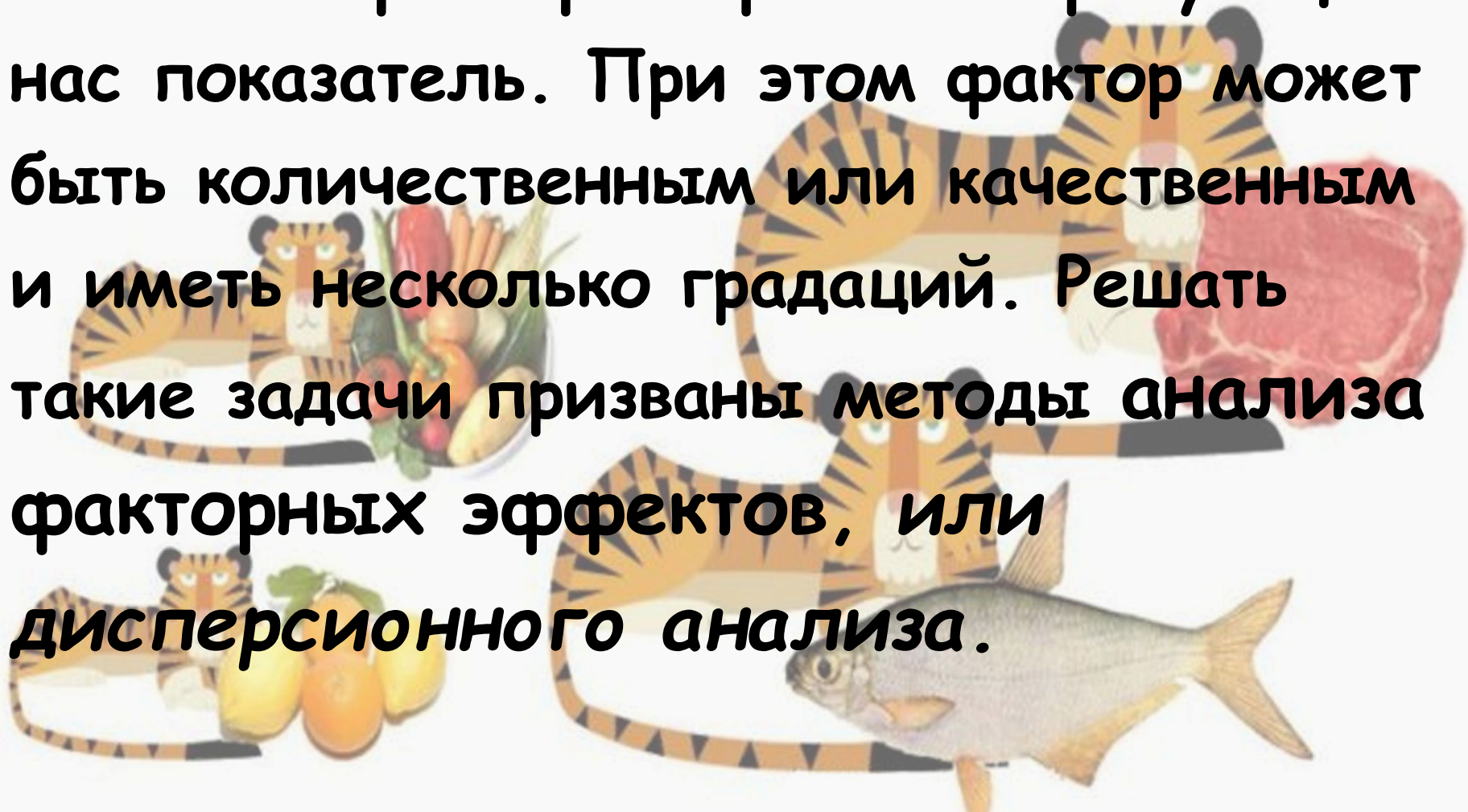
5

6

**критериев различий покажет,
случайны или неслучайны
различия двух числовых выборок.**

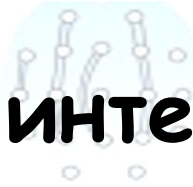
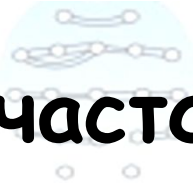
Дисперсионный анализ

Часто необходимо выяснить, влияет или нет некоторый фактор на интересующий нас показатель. При этом фактор может быть количественным или качественным и иметь несколько градаций. Решать такие задачи призваны методы анализа факторных эффектов, или дисперсионного анализа.

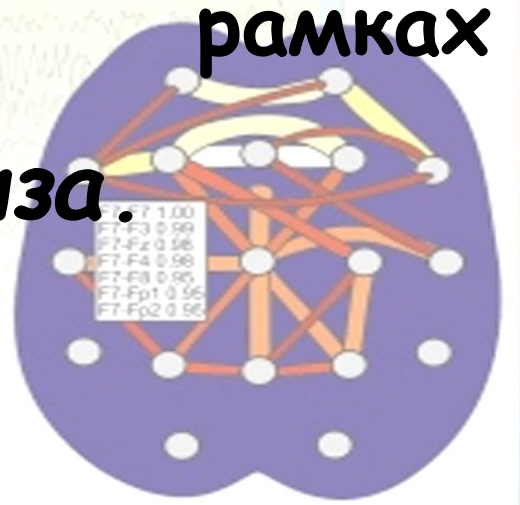


Корреляционный анализ

Нас часто интересует характеристика тесноты (силы) связи между двумя переменными, выраженная одним числом. Расчет этого параметра осуществляется в рамках корреляционного анализа.



01-A1
P4-A2
P3-A1
C3-A1
F4-A2
F3-A1
Fp2-A2
Fp1-A1
T6-A2
T5-A1
T4-A2
T3-A1
F8-A2
F7-A1
0z-A2
Pz-A1
Cz-A2
Fz-A1



Регрессионный анализ. Вопросы моделирования статистических зависимостей между двумя или несколькими переменными решаются в разделе регрессионного анализа. С помощью таких моделей можно не только построить прогноз, но и определить оптимальные области для последующего управления и контроля.

$$\begin{aligned}
 y(t) &= \frac{1}{5} \left(-\frac{dx}{dt} + 2x + 3 \right) = \\
 &= \frac{1}{5} \left(-e^{-2t} \left((-2C_1 + 3C_2) \cos 3t + (-3C_1 - 2C_2) \sin 3t \right) + 2(e^{-2t} (C_1 \cos 3t + C_2 \sin 3t) + 1) + 3 \right) = \\
 &= \frac{1}{5} \left(e^{-2t} \left((2C_1 - 3C_2) \cos 3t + (3C_1 + 2C_2) \sin 3t \right) + 2C_1 \cos 3t + 2C_2 \sin 3t + 5 \right) = \\
 &= \frac{1}{5} \left(e^{-2t} \left((2C_1 - 3C_2 + 2C_1) \cos 3t + (3C_1 + 2C_2 + 2C_2) \sin 3t \right) + 5 \right) = \\
 &= e^{-2t} \left(\left(\frac{4C_1 - 3C_2}{5} \right) \cos 3t + \left(\frac{3C_1 + 4C_2}{5} \right) \sin 3t \right) + 1
 \end{aligned}$$

Методы контроля качества

Вильфредо
Парето

20% Ваших усилий
дают
80% результатов

Многих может интересовать, когда необходимо немедленно вмешиваться в технологический или управленческий процесс. Не упустить этот важный момент помогут методы контроля качества.

Многомерные методы

Факторный анализ вычислит новую, сокращенную систему координат, на которую изучаемые объекты будут проецироваться более экономным и информативным образом.

Кластерный анализ построит дерево классификации объектов (дендрограмму), на котором ветви отходят от ствола соответственно взаимной удаленности (несхожести) объектов.

Дискриминантный анализ подберет уравнение, разбивающее объекты на заданное число классов, используя которое можно успешно классифицировать новые объекты.

Этапы анализа данных

1. Сбор данных для анализа.

Сбор данных для анализа обычно осуществляется одним из трех способов:

- в ходе **активного эксперимента** с технической регистрацией значений соответствующих переменных;
- в результате **наблюдения или опроса** с ручной регистрацией соответствующих показателей;
- данные извлекаются из **литературных источников**.

Процедура сбора данных определяется задачей исследования.

2. Ввод данных

Процесс систематизации результатов массовых наблюдений, объединения их в относительно однородные группы по некоторому признаку называется **группировкой**. Нельзя группировать в одну и ту же совокупность неоднородные по составу данные, необдуманно выбирать способ группировки. Группировка должна отвечать требованию поставленной задачи и соответствовать содержанию изучаемого явления.

Способы группировки данных:

1. **Таблицы.** Наиболее распространенной формой группировки являются *статистические таблицы*.

Они бывают простыми (табл. 1) и сложными.

2. **Статистические ряды.** *Статистическим* называют ряд числовых значений признака, расположенных в определенном порядке.

8	95.76	97.61
9	159.6	100
10		

3. Визуальный анализ

Регрессия: $x_2=f(x_1)$

*E1

После ввода на данные в электронной таблице следует просто посмотреть, чтобы составить общее представление о характере их изменения, специфических особенностях и закономерностях, что очень важно для выбора стратегии и тактики дальнейшего исследования. Для этого можно использовать как исходное числовое представление, так и различные формы графического изображения.

Как я выгляжу, зебра?

4. Преобразование данных

1. Удаление из введенных данных **выбросов** (резко отличающихся от остальных данных по величине, которые могут быть, например, результатом некорректных измерений) и посторонних данных, а также замена (или удаление) пропущенных (неизмеренных) значений.
2. **Корректировка исходных данных** методами ручного редактирования или полуавтоматического преобразования к виду, адекватному методу выбранного анализа.

Нормально



5. Статистический анализ

Всех X=2, x3=0

Таблица данных

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
1	0	0								
2	0	0								
3	7.98	7.86								
4	15.86	22.42								
5	23.54	52.15								
6	31.92	81.9								
7	61.82	9								
8	95.76	97.61								
9	159.6	100								
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										

Выбор метода анализа, адекватного поставленной задаче, и его последовательное исполнение.

Статистические методы

Параметрические тесты

- 1=Описательная статистика
- 2=Полное ранжирование
- 3=Корреляция
- 4=Стьюдента и Фишера

Непараметрические тесты

- 5=Хи-квадрат
- 6=Сдвига/положения
- 7=Масштаба/рассеяния
- 8=Произвольных альтернатив
- 9=Корреляция/независимость
- А=Кросстабуляция

Дисперсионный анализ

- В=Однофакторный
- С=Двухфакторный
- с=Групповой
- Д=Многофакторный
- Е=Ковариационный

Анализ временных рядов

- Г=Корреляционный
- Ж=Спектральный
- И=Сглаживание/фильтрация
- Л=ARIMA-модели
- М=Фурье-модели

Регрессионный анализ

- К=Сравнение 2-х регрессий
- Л=Простая регрессия/тренд
- М=Множественная линейная
- Н=Пошаговая регрессия
- О=Общая/нелинейная модель

Многомерные методы

- Р=Факторный
- Q=Кластерный
- В=Дискриминантный
- С=Шкалирование

Распределения и частоты

- Т=Вычисление вероятностей
- У=Согласие распределений
- В=Согласие частот
- W=Последовательный анализ
- w=Анализ выживаемости

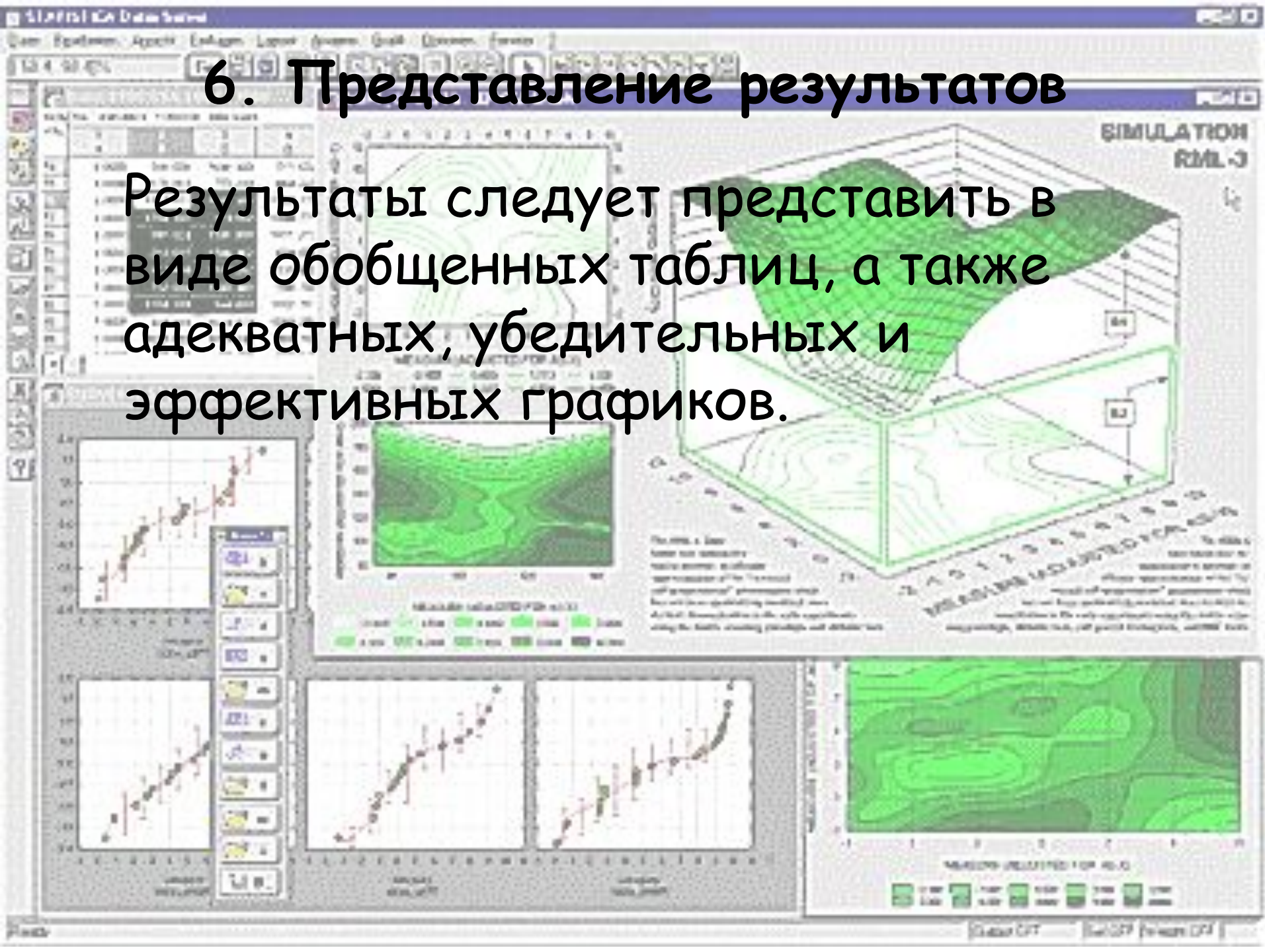
Контроль качества

- Х=Гистограмма
- У=Диаграмма Парето
- Z=Контрольные карты

X <Esc>=Отменить

6. Представление результатов

Результаты следует представить в виде обобщенных таблиц, а также адекватных, убедительных и эффективных графиков.



7. Выводы



ТЫ МОЖЕШЬ
В ЭТО ПОВЕРИТЬ?!
МЫ ПРОДЕЛАЛИ ВЕСЬ ЭТОТ
ДОЛГИЙ ПУТЬ И ВЕРНУЛИСЬ
К ТОМУ С ЧЕГО
НАЧИНАЛИ!

Если выводы полностью исчерпывают поставленную задачу, то исследователь может перейти к постановке и решению новой задачи. В противном случае может понадобиться выбор и исполнение другого метода анализа, выполнение новых преобразований, их повторное визуальное изучение, ввод или извлечение из архива дополнительной информации.