In[33]:=
```
ClearAll["Global`*"]
```

Экспериментальные основания теории вероятностей

Математическая модель

1. Рассматривается эксперимент со случайными исходами - определение числа гербов при бросании трех монет

1.1. Сделать 10 бросков трех монет и зафиксировать результаты эксперимента, в каждом эксперименте фиксировать число полученных гербов, соответствующее значение 0 или 1 или 2 или 3. Полученные значения оформить в виде списка

1.2. Провести имитацию этого эксперимента с помощью датчика случайных чисел, функция RandomChoice[] с аргументом - список {“Г”,”Р”}. Символ “Г” - обозначает выпадение герба, а символ “Р” - появление решки. Вызов этой функции с дополнительным аргументом 3 имитирует результат броска трех монет. Построить с помощью функции Table[] последовательность таких троек. Общее количество проведенных экспериментов (объем выборки) задать самостоятельно (значение от 300 до 500)

1.2.1. Для подключения датчика случайных чисел используется функция SeedRandom, ее использование смотри в тексте работы

1.3. Провести обработку полученных результатов, вычислив количества появления герба в каждом эксперименте. Объединить вместе результаты, полученные в первом и втором пункте, в виде списка randcol. Получить объем (количество проведенных экспериментов) новой объединенной выборки

1.4. Провести визуализацию полученных данных с помощью функции Histogram[randcol], эта функция без дополнительных атрибутов дает значения в абсолютных единицах, с дополнительными атрибутами Automatic и "Probability" дает значения в относительных единицах

1.5. С помощью функции Count[] определить зависимость текущего количества появления гербов при троекратном бросании монеты от объема выборки, то есть текущую абсолютную частоту появления количества гербов от объема выборки. Функция Count[randcol , k] - число гербов равное k, то есть Count[randcol , 0] опрределяет количество экспериментов, в которых зафиксировано значение 0, Count[randcol , 1] - число экспериментов, в которых зафиксировано значение 1 и так далее.

1.6. Построить зависимость текущей относительной частоты выпадений количества гербов от объема выборки (то есть текущее количество появления гербов деленное на объем выборки)

1.7. Построить графики зависимости относительной частоты от объема выборки, обратить внимание на тенденции поведения построенных зависимостей, определить значения к которым стремятся каждая из относительных частот

1.8. Добавить к графику линии с теоретическими значениями вероятностей соответствующих событий, для значений 0 и 3 теоретическое значение равно p03teor = $\frac{1}{8}$, для значений 1 и 2 теоретическое знаечние равно p12teor= $\frac{3}{8}$

1.9. Определить визуально скорость сходимости абсолютной величины разности частоты и теоретической вероятности от объема выборки. Для этого сравнить абсолютную величину разности с величиной $\frac{\text{sigma}}{n^{\alpha}}$, где sigma=`sig03teor` = $\frac{\sqrt{7}}{8}$ для значений 0 и 3 и sigma=`sig12teor` = $\frac{\sqrt{15}}{8}$ для значений 1 и 2. Скорость сходимости (показатель степени α) менять в пределах от 0 до 1. Сделать вывод о значении скорости сходимости для каждой из кривых. Теория предсказывает значение $\alpha = \frac{1}{2}$

2. Рассматривается эксперимент со случайными исходами - определение среднего числа гербов при бросании трех монет (вычисляется общая сумма очков при трех бросаниях и она делится на 3)

2.1. Провести имитацию этого эксперимента с помощью датчика случайных чисел. Построить с помощью функции Table[] последовательность сгенерированных троек . Общее количество проведенных экспериментов, объем выборки, задать самостоятельно (значение от 300 до 500)

2.2. Построить таблицу значений для текущего среднего числа выпавших гербов и эту таблицу вывести ее в виде графика, на этом графике также указать теоретическое значение sred = 1.5

2.3. Определить визуально скорость сходимости (значение показателя α) абсолютной величины разности текущего среднего и теоретического значения от объема выборки. Для этого сравнить указанное значение с величиной $\frac{\text{sigmasr}}{n^{\alpha}}$, где sigmasr = $\frac{3}{4}$. Показатель степени α (скорость сходимости) менять в пределах от 0 до 1. Сделать вывод о значении скорости сходимости. Теория предсказывает значение $\alpha = \frac{1}{2}$

3. Сделать выводы из наблюдаемых результатов экспериментов (можно повторить все эксперименты, увеличив объемы выборки до 1000)

Использование средств пакета Mathematica

- Определение функции и переменных в виде отложенного выражения
- Запрет вывода результата на экран, символ ; в конце выражения
- Функция вычисления псевдослучайного числа RandomInteger[]
- Функция задания условия для работы генератора псевдослучайных чисел SeedRandom[], вызов без аргумента - каждый раз новая серия чисел, с любым параметром - серия чисел воспроизводима
- Функция построения таблицы данных Table[]
- Функция объединения списков Join[]
- Функция вычисления суммы списка элементов Total[]
- Функция построения гистограммы для выборки Histogram[]
- Функция определения числа элементов в списке Dimensions[]
- Функция определения количества элементов в списке, удовлетворяющих заданному условию Count[]
- Операция ;; указания диапазона индексов в списке. Пример, 1;;10 (диапазон от 1 до 10)
- Функция вывода на экран элементов последовательности ListPlot[], атрибут Joined → True точки соединить непрерывной линией
- Функция построения прямой линии или нескольких линий Line[]
- Функция интерактивного вывода на экран Manipulate[]

Варианты заданий

1. Провести натурный эксперимент по одновременному бросанию 3-х монет. Зафиксировать результаты 10-ти проведенных экспериментов
2. Провести имитацию этого эксперимента, используя датчик псевдослучайных чисел, объем выборки от 300 до 500
3. Выполнить все задания, указанные в описании

Отчет по лабораторной работе

1. Общие требования к отчету для любой лабораторной работы

Отчет должен содержать - название лабораторной работы, цель лабораторной работы, основные формулы и зависимости, основные графики, вывод о проделанных вычислениях.
Отчет должен быть представлен в рукописном или печатном виде.
Срок сдачи отчета - одна неделя после проведения лабораторной работы.
К защите лабораторной работы должны быть представлены - исходный файл в формате блокнота (.nb) и проверенный отчет.

2. Дополнительные требования к этой лабораторной работе

2.1. Результаты физического эксперимента

2.2. Гистограмма объединенной выборки

2.3. График зависимости относительной частоты от количества экспериментов с указанием теоретических предельных значений

2.4. Результаты определения показателей скорости сходимости частоты к предельному значению

2.5. График выборочного среднего в зависимости от количества проведенных экспериментов

2.6. Результаты определения показателя скорости сходимости среднего к предельному значению

Пример решения задачи

■ Эксперимент с бросанием трех монет (определение частот)

Результаты проведенного физического (ручного) эксперимента

In[35]:=
```
experim = {1, 2, 0, 0, 3, 0, 1, 2, 2, 2}
```

Out[35]=
```
{1, 2, 0, 0, 3, 0, 1, 2, 2, 2}
```

Эксперимент - имитация бросания трех монет, фиксация результата
“Г” - появление герба, “Р” - появление решетки.
Построение выборки объема Nvyb (задаем, например, значение 300)

In[36]:=
```
Nvyb = 300
```

Out[36]=
```
300
```

С помощью функции SeedRandom обеспечиваем повторяемость результатов, если в качестве аргумента функции задать какое-либо конкретное число, то результаты при всех вычислениях будут повторяться, чтобы сделать результаты уникальными - вызвать функцию SeedRandom без аргумента.
Аргумент SeedRandom задать самостоятельно (9 - 11 знаков в числе, число нечетное)

In[37]:=
```
SeedRandom[1234567];
imit = RandomChoice[{"Г", "Р"}, {Nvyb, 3}];
```

Значение imit[[1]] - результат первого имитационного эксперимента (Г означает, что выпал герб, Р означает, что выпала решетка)

In[39]:=
```
imit[[1]]
```

Out[39]=
```
{Р, Г, Г}
```

В нашем случае получилось - Решетка, Герб, Герб

Несколько (10 штук) первых результатов

In[40]:=
```
imit[[1 ;; 10]]
```

Out[40]=
```
{{Р, Г, Г}, {Г, Г, Г}, {Г, Р, Г}, {Р, Г, Р},
 {Р, Р, Г}, {Г, Р, Р}, {Р, Р, Г}, {Р, Р, Р}, {Р, Г, Р}, {Г, Г, Р}}
```

Простейшая обработка - расчет количества появления герба

In[41]:=
```
rand = Table[Count[imit[[i]], "Г"], {i, 1, Nvyb}];
(*Count[#,"Г"]&/@imit;*)
```

Первые 10 значений

In[42]:=
```
rand[[1 ;; 10]]
```

Out[42]=
```
{2, 3, 2, 1, 1, 1, 1, 0, 1, 2}
```

Объединение результатов эксперимента, определение нового объема объединенной выборки

In[43]:=
```
randcol = Join[experim, rand];
```

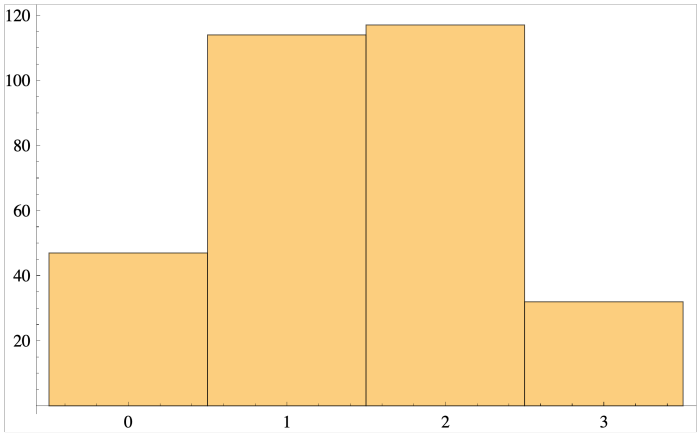
In[44]:=
```
nrandcol = Length[randcol]
```

Out[44]=
```
310
```

Визуализация результата - построение гистограммы объединенной выборки

In[45]:=
```
Histogram[randcol]
```

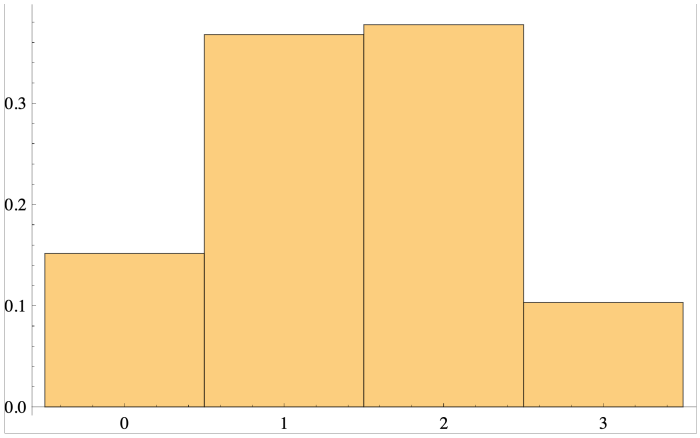
Out[45]=



В относительных единицах

In[46]:=
```
Histogram[randcol, Automatic, "Probability"]
```

Out[46]=



Вычисление абсолютной частоты - сумма гербов равная 0 появилась 47 раз , сумма гербов равная 1 появилась 114 раза , сумма гербов равная 2 появилась 117 раз и сумма гербов равная 3 появиалсь 32 раз.

In[47]:=
```
Table[Count[randcol, i], {i, 0, 3}]
(*Count[randcol,#]&/@{0,1,2,3}*)
```

Out[47]=
```
{47, 114, 117, 32}
```

Или использование стандартной функции

In[48]:=
```
BinCounts[randcol, {0, 4, 1}]
```

Out[48]=
```
{47, 114, 117, 32}
```

Ваши результаты могут отличаться от приведенных выше!!!

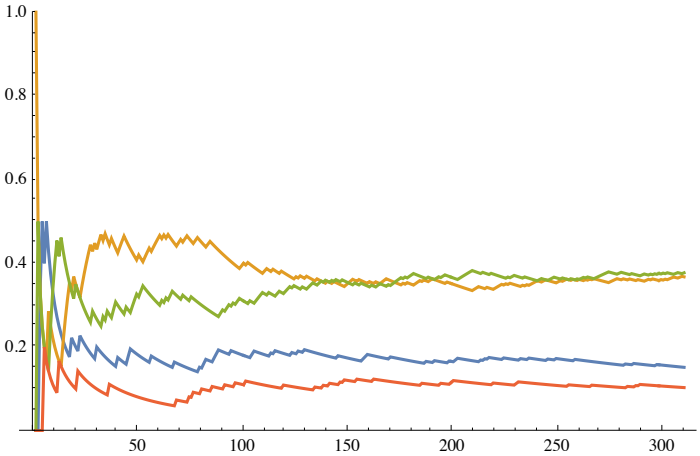
Построение зависимости текущего относительной частоты числа выпадений гербов от объема выборки (текущий объем выборки k меняется от 1 до nrandcol
val0 - для общего числа гербов 0,
val1 - для общего числа гербов 1,
val2 - для общего числа гербов 2,
val3 - для общего числа гербов3.

In[49]:=
```
val0 = Table[Count[randcol[[1 ;; k]], 0] / k, {k, 1, nrandcol}] // N;
val1 = Table[Count[randcol[[1 ;; k]], 1] / k, {k, 1, nrandcol}] // N;
val2 = Table[Count[randcol[[1 ;; k]], 2] / k, {k, 1, nrandcol}] // N;
val3 = Table[Count[randcol[[1 ;; k]], 3] / k, {k, 1, nrandcol}] // N;
```

График зависимости относительной частоты от объема выборки - наблюдается стабилизация частоты с ростом объема выборки

In[53]:=
```
ListPlot[{val0, val1, val2, val3}, PlotRange → {0, 1}, Joined → True]
```

Out[53]=



Сравним экспериментальные данные с теоретическими для идеальных монет

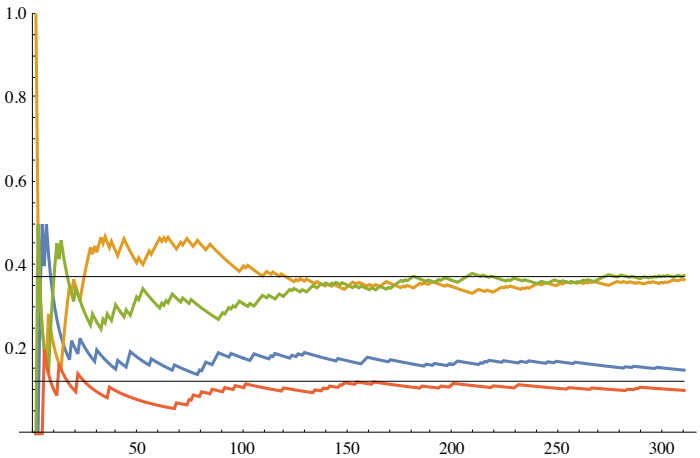
In[54]:=
```
{p03teor, p12teor} = {1 / 8, 3 / 8}
```

Out[54]= $\left\{\frac{1}{8}, \frac{3}{8}\right\}$

Графики вместе с теоретическими значениями

In[55]:=
```
ListPlot[{val0, val1, val2, val3}, PlotRange → {0, 1}, Joined → True,
 Epilog →
  Line[{{{0, p03teor}, {nrandcol, p03teor}}, {{0, p12teor}, {nrandcol, p12teor}}}]]
```

Out[55]=



Теоретическая оценка скорости сходимости частоты к соответствующей вероятности, теоретические значения среднеквадратичныъ отклонений

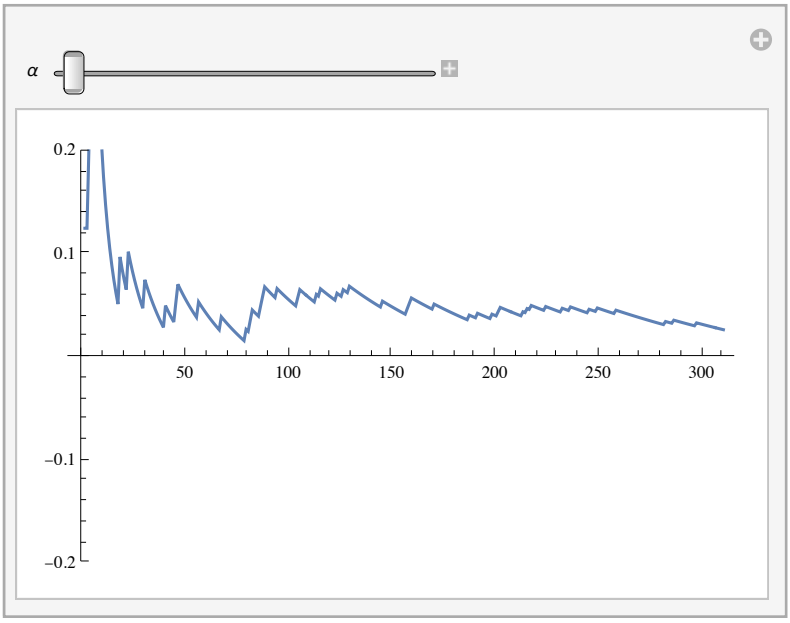
In[56]:=
```
{sig03teor, sig12teor} = {Sqrt[7] / 8, Sqrt[15] / 8}
```

Out[56]= $\left\{\frac{\sqrt{7}}{8}, \frac{\sqrt{15}}{8}\right\}$

Определение скорости сходимости модуля разности val0 и p03teor (теоретические значения для параметра $\alpha = 0.5$)

In[57]:=
```
Manipulate[ListPlot[{Table[Abs[(val0[[k]] - p03teor)], {k, 1, nrandcol}],
  Table[sig03teor / k^α, {k, 1, nrandcol}]},
 Joined → True, PlotRange → {-0.2, 0.2}], {α, 0, 1}]
```

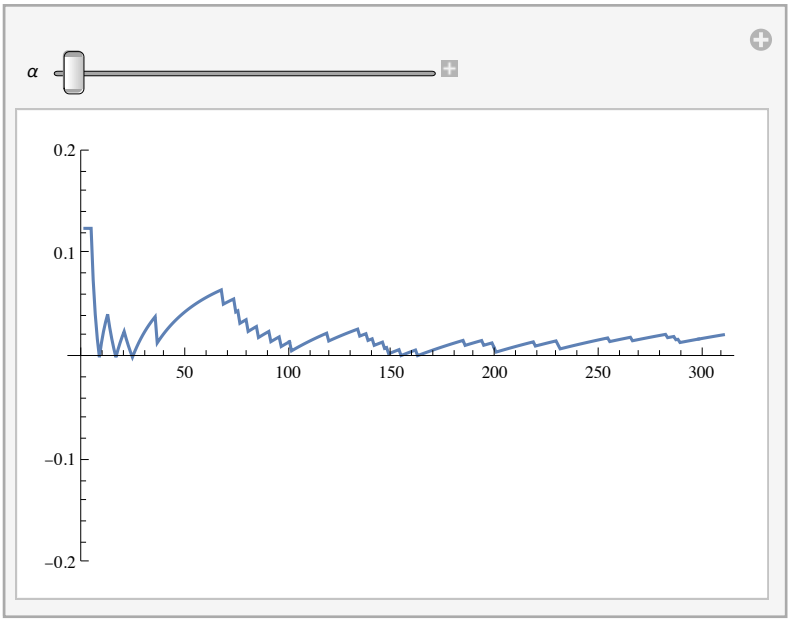
Out[57]=



Ответ - значение α приблизительно равно 0.433 (для данной выборки)
Ваши результаты могут отличаться от полученных!!!

In[58]:=
```
Manipulate[ListPlot[{Table[Abs[(val3[[k]] - p03teor)], {k, 1, nrandcol}],
  Table[sig03teor / k^α, {k, 1, nrandcol}]},
 Joined → True, PlotRange → {-0.2, 0.2}], {α, 0, 1}]
```

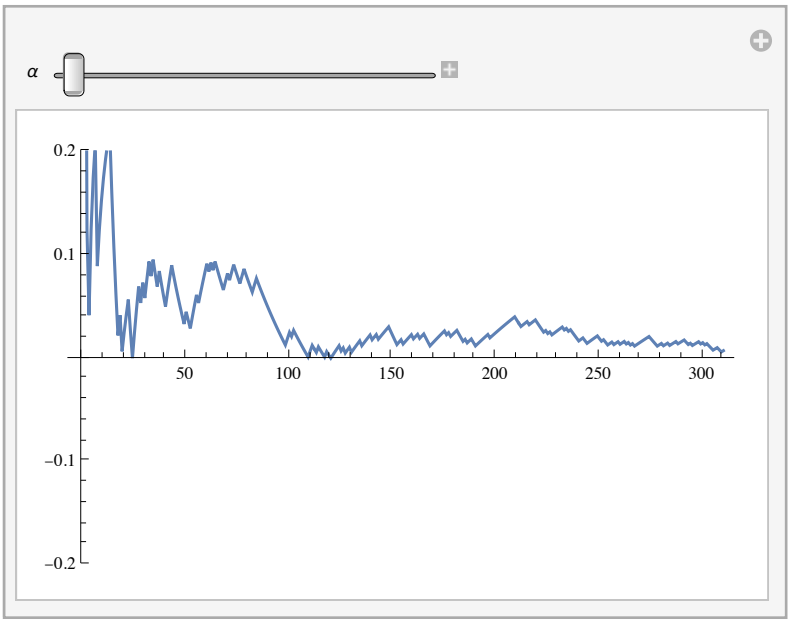
Out[58]=



Ответ - подходящее значение $\alpha = 0.52$ (для данной выборки)

In[59]:=
```
Manipulate[ListPlot[{Table[Abs[(val1[[k]] - p12teor)], {k, 1, nrandcol}],
   Table[sig12teor/k^α, {k, 1, nrandcol}]},
  Joined → True, PlotRange → {-0.2, 0.2}], {α, 0, 1}]
```

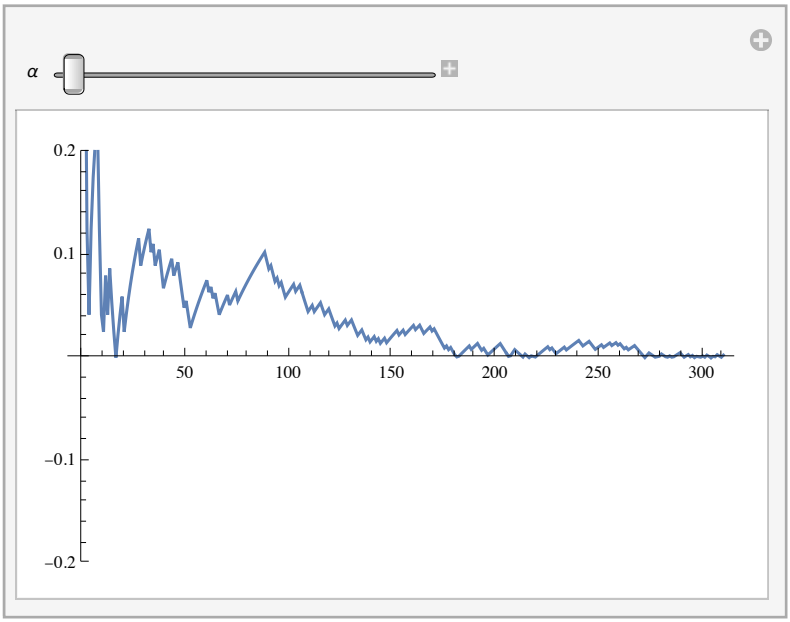
Out[59]=



Ответ - подходящее значение $\alpha = 0.618$ (для данной выборки)

In[60]:=
```
Manipulate[ListPlot[{Table[Abs[(val2[[k]] - p12teor)], {k, 1, nrandcol}],
   Table[sig12teor/k^α, {k, 1, nrandcol}]},
  Joined → True, PlotRange → {-0.2, 0.2}], {α, 0, 1}]
```

Out[60]=



Ответ - подходящее значение $\alpha = 0.759$ (для данной выборки)

■ Эксперимент с вычислением среднего значения

Для проведения этого эксперимента можно в функции SeedRandom взять то же самое затравочное число, что и в первой задаче.

Объем выборки можно взять больше

In[61]:=
```
Nvyb1 = 500
SeedRandom[1 234 567];
imit1 = Table[RandomChoice[{"Г", "Р"}, 3], {i, 1, Nvyb1}];
```

Out[61]= 500

In[63]:=
```
randtab = Table[Count[imit1[[i]], "Г"], {i, 1, Nvyb1}];
```

Теоретические значения параметров

In[64]:=
```
sred = 1.5
sigmasr = 0.75
```

Out[64]= 1.5

Out[65]= 0.75

Если выбрано то же самое затравочное число, то данные должны совпадать с результатом первой задачи

In[66]:=
```
randtab[[1 ;; 10]]
```

Out[66]= {2, 3, 2, 1, 1, 1, 1, 0, 1, 2}

Вычисление текущего среднего значения количества выпавших гербов

In[67]:=
```
sredtek = Table[1/k Total[randtab[[1 ;; k]]], {k, 1, Nvyb1}];
```

Для примера первые 10 значений

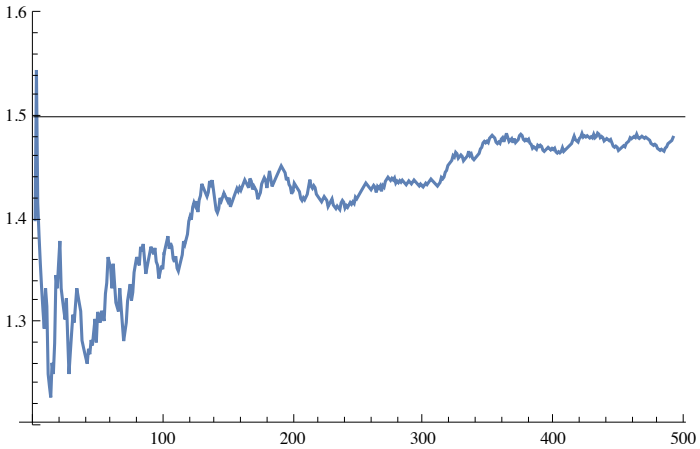
In[68]:=
```
sredtek[[1 ;; 10]] // N
```

Out[68]= {2., 2.5, 2.33333, 2., 1.8, 1.66667, 1.57143, 1.375, 1.33333, 1.4}

Из теории вытекает, что среднее значение должно стабилизироваться к величине sred

In[69]:=
```
ListPlot[sredtek[[10 ;;]], Joined → True, PlotRange → {1.2, 1.6},
 Epilog → Line[{{0, sred}, {Nvyb1, sred}}]]
```

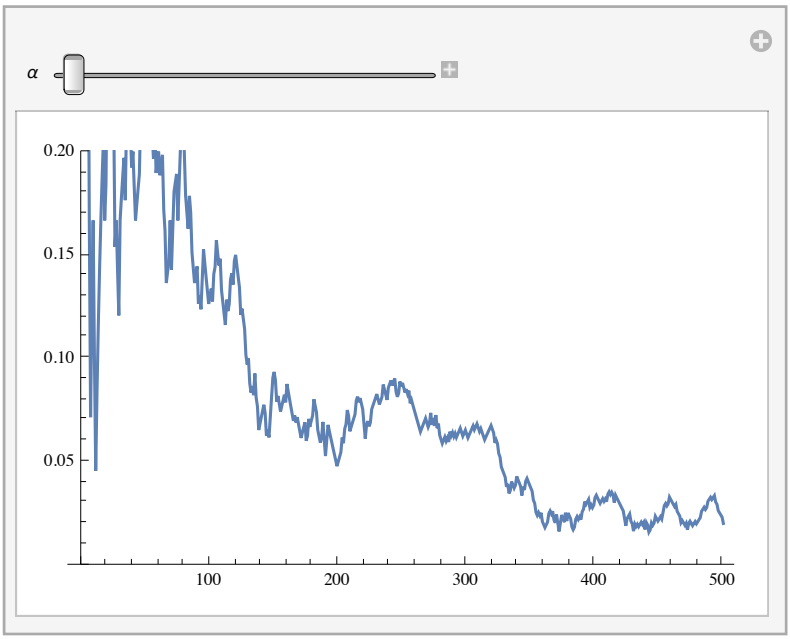
Out[69]=



Из теории вытекает, что скорость стремления отклонения к нулю определяется показателем степени 0.5

In[70]:=
```
Manipulate[ListPlot[{Table[Abs[(sredtek[[k]] - sred)], {k, 1, Nvyb1}],
  Table[sigmasr/k^α, {k, 1, Nvyb1}]}, Joined → True, PlotRange → {0, 0.2}], {α, 0, 1}]
```

Out[70]=



Возможный вариант значения $\alpha = 0.556$

■ Выводы из проделанных экспериментов

1. Эксперименты с частотой получения количества гербов.

При увеличении количества экспериментов происходит стабилизация частот к некоторым предельным значениям - теоретическим вероятностям.

Скорость стабилизации описывается зависимостью $\frac{\text{Const}}{n^{\alpha}}$, где n - объем выборки, а величина α порядка 0.5

2. Экспермент со средним количеством гербов

При увеличении количества экспериментов происходит стабилизация среднего значения к некоторому предельному значению - теоретическому среднему значению.

Скорость стабилизации описывается зависимостью $\frac{\text{Const}}{n^{\alpha}}$, где n - объем выборки, а величина α порядка 0.5

In[256]:= `ClearAll["Global`*"]`

Точечные и интервальные оценки параметров генеральной совокупности

Математическая модель

1. Получить выборку из нормального закона с параметрами (m,σ) заданного объема n

2. Получить точечную оценку для математического ожидания

3. Получить точечную несмещенную оценку для дисперсии

4. Построить доверительный интервал с надежностью γ для оценки математического ожидания

4.1. В предположении известной дисперсии σ

4.2. В предположении неизвестной дисперсии σ

4.3. Построить таблицу зависимости величины доверительного интервала от величины надежности γ

5. Построить доверительный интервал с надежностью γ для оценки дисперсии

5.1. Построить таблицу зависимости величины доверительного интервала от величины надежности γ

6. Оценить необходимый объем выборки для получения точечной оценки параметра закона распределения с точностью 1% с заданной надежностью γ

6.1. Оценить необходимый объем выборки для получения оценки математического ожидания с точностью 1% с заданной надежностью γ. Сравнить значения с полученными экспериментально

Использование средств пакета Mathematica

- Определение функции и переменных в виде отложенного выражения
- Функция генерации псевдослучайных чисел RandomReal[]
- Функция определения нормального закона распределения NormalDistribution[]
- Функция определения t-распределения Стьюдента StudentTDistribution[]
- Функция определения χ^2-распределения ChiSquareDistribution
- Функция для обратной плотности вероятности закона распределения InverseCDF[]
- Функция определения квантиля заданного закона распределения Quantile[]
- Функция определения среднего значения списка Mean[]
- Функция определения дисперсии списка Variance[]
- Функция определения суммы элементов списка Total[]
- Функция построения таблицы значений Table[]
- Функция объединения списков Join[]
- Функция определения числа элементов в списке Lenght[]
- Функция построения точечного графика ListPlot[]
-Функция определения корня уравнения FindRoot[]

Варианты заданий

1. Получить выборку из нормального закона распределения с параметрами (m,σ) заданного объема n.

Параметры выбрать самостоятельно, 5 <= m <= 10, 1 <= σ <=3, n >= 200
2. Решить все задачи, указанные в пунктах 2 - 6

Пример решения задачи

Генерация выборки и списка доверительных вероятностей

Пусть m = 2, σ = 1, объем выборки n = 200

In[257]:=
```
m = 2
σ = 1
n = 200
```

Out[257]= 2

Out[258]= 1

Out[259]= 200

Генерация выборки, распределенной по нормальному закону

In[260]:=
```
vyb = RandomReal[NormalDistribution[m, σ], n];
```

Получение точечных оценок

In[261]:=
```
mean = Mean[vyb]
```

Out[261]= 2.00795

In[262]:=
```
m1 = Total[vyb] / n
```

Out[262]= 2.00795

In[263]:=
```
disp1 = Total[(vyb - m)^2] / n
sig1 = √disp1
```

Out[263]= 0.77244

Out[264]= 0.878885

In[265]:=
```
disp2 = Total[(vyb - mean)^2] / n
sig2 = √disp2
```

Out[265]= 0.772376

Out[266]= 0.878849

In[267]:=
```
disp3 = Total[(vyb - mean)^2] / (n - 1)
sig3 = √disp3
```

Out[267]= 0.776258

Out[268]= 0.881055

In[269]:=
```
disp = Variance[vyb] (* совпадает с disp3  *)
sig = √disp
```

Out[269]= 0.776258

Out[270]= 0.881055

■ Построение доверительного интервала для оценки математического ожидания при известной дисперсии

Построение таблицы задаваемых доверительных вероятностей

In[271]:=
```
γ = Table[0.9 + 0.01 i, {i, 0, 9}]
```

Out[271]=
```
{0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99}
```

In[272]:=
```
γ1 = Join[γ, {0.995, 0.999}]
```

Out[272]=
```
{0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 0.999}
```

При доверительной вероятности γ используется квантиль $\frac{1+\gamma}{2}$ для стандартного нормального закона

Определение доверительного интервала для заданных доверительных вероятностей

In[273]:=
```
t1 = {γ1, mean - σ/√n Quantile[NormalDistribution[], (1 + γ1)/2],
  mean + σ/√n Quantile[NormalDistribution[], (1 + γ1)/2]}
```

Out[273]=
```
{{0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 0.999},
 {1.89164, 1.88807, 1.88416, 1.87983, 1.87496, 1.86936, 1.86273, 1.8545,
  1.84345, 1.82581, 1.80946, 1.77528}, {2.12426, 2.12783, 2.13174, 2.13607,
  2.14094, 2.14654, 2.15317, 2.1614, 2.17245, 2.19009, 2.20644, 2.24063}}
```

In[274]:=
```
qq = Quantile[NormalDistribution[], (1 + γ1)/2]
```

Out[274]=
```
{1.64485, 1.6954, 1.75069, 1.81191, 1.88079,
 1.95996, 2.05375, 2.17009, 2.32635, 2.57583, 2.80703, 3.29053}
```

Вывод в виде таблицы

In[275]:=
```
Join[{{"Дов. вер.", "Нижняя гр.", "Верхняя гр."}}, Transpose[t1]] // MatrixForm
```

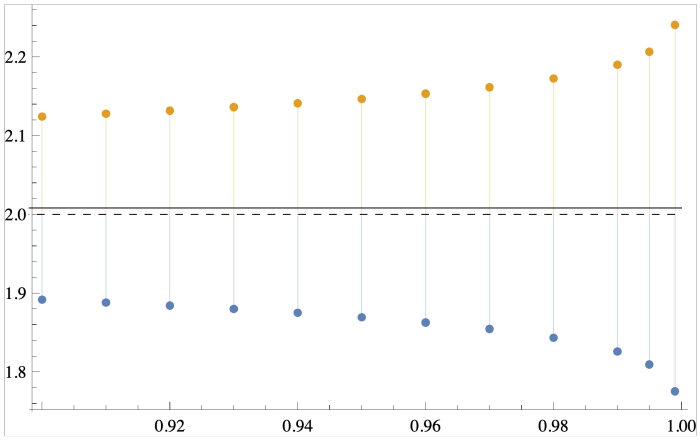
Out[275]//MatrixForm=

Дов. вер.	Нижняя гр.	Верхняя гр.
0.9	1.89164	2.12426
0.91	1.88807	2.12783
0.92	1.88416	2.13174
0.93	1.87983	2.13607
0.94	1.87496	2.14094
0.95	1.86936	2.14654
0.96	1.86273	2.15317
0.97	1.8545	2.1614
0.98	1.84345	2.17245
0.99	1.82581	2.19009
0.995	1.80946	2.20644
0.999	1.77528	2.24063

Графическое изображение доверительных интервалов, чем больше доверительная вероятность, тем больше доверительный интервал и, тем самым, хуже точность определения параметра. Приведены также оценка математического ожидания mean, и значение, использованное в симуляции для построения выборки. Определить, попадает ли значение, использованное для симуляции в построенный доверительный интервал

In[276]:=
```
ListPlot[{Transpose[{t1[[1]], t1[[2]]}], Transpose[{t1[[1]], t1[[3]]}]}, Filling → 2,
 Epilog → {Line[{{0, mean}, {1, mean}}], Dashed, Line[{{0, m}, {1, m}}]}]
```

Out[276]=



■ Построение доверительного интервала для оценки математического ожидания при неизвестной дисперсии

При доверительной вероятности γ используется квантиль $\frac{1+\gamma}{2}$ для t-распределния Стьдента с числом степеней свободы (n-1)

In[277]:=
```
t2 = {γ1, mean - sig/Sqrt[n - 1] Quantile[StudentTDistribution[n - 1], (1 + γ1)/2],
  mean + sig/Sqrt[n - 1] Quantile[StudentTDistribution[n - 1], (1 + γ1)/2]}
```

Out[277]=
```
{{0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 0.999},
 {1.90474, 1.90155, 1.89805, 1.89417, 1.88981, 1.88479, 1.87884, 1.87144,
  1.86148, 1.84552, 1.83066, 1.79934}, {2.11116, 2.11436, 2.11785, 2.12173,
  2.12609, 2.13111, 2.13707, 2.14447, 2.15443, 2.17039, 2.18525, 2.21656}}
```

Вывод в виде таблицы

In[278]:=
```
Join[{{"Дов. вер.", "Нижняя гр.", "Верхняя гр."}}, Transpose[t2]] // MatrixForm
```

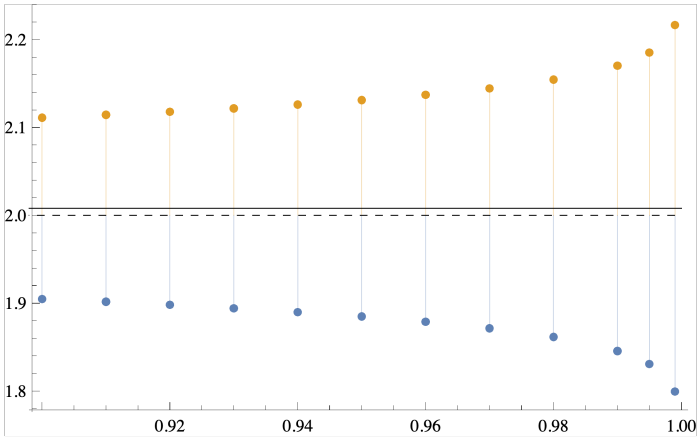
Out[278]//MatrixForm=

Дов. вер.	Нижняя гр.	Верхняя гр.
0.9	1.90474	2.11116
0.91	1.90155	2.11436
0.92	1.89805	2.11785
0.93	1.89417	2.12173
0.94	1.88981	2.12609
0.95	1.88479	2.13111
0.96	1.87884	2.13707
0.97	1.87144	2.14447
0.98	1.86148	2.15443
0.99	1.84552	2.17039
0.995	1.83066	2.18525
0.999	1.79934	2.21656

Графическое изображение доверительных интервалов, чем больше доверительная вероятность, тем больше доверительный интервал и, тем самым, хуже точность определения параметра. Приведены также оценка математического ожидания mean, и значение, использованное в симуляции для построения выборки. Определить, попадает ли значение, использованное для симуляции в построенный доверительный интервал

In[279]:=
```
ListPlot[{Transpose[{t2[[1]], t2[[2]]}], Transpose[{t2[[1]], t2[[3]]}]}, Filling → 2,
 Epilog → {Line[{{0, mean}, {1, mean}}], Dashed, Line[{{0, m}, {1, m}}]}]
```

Out[279]=



■ Построение доверительного интервала для дисперсии

При доверительной вероятности γ используется квантили уровня $\frac{1+\gamma}{2}$ и $\frac{1-\gamma}{2}$ для распределения χ^2 с числом степеней свободы (n-1)

In[280]:=
```
t3 = {γ1, (n sig^2) / Quantile[ChiSquareDistribution[n - 1], (1 + γ1)/2],
   (n sig^2) / Quantile[ChiSquareDistribution[n - 1], (1 - γ1)/2]}
```

Out[280]=
```
{{0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99, 0.995, 0.999},
 {0.666568, 0.663374, 0.659905, 0.656091, 0.651835, 0.64699, 0.641313, 0.634362,
  0.625186, 0.610901, 0.598053, 0.572339}, {0.927645, 0.932639, 0.938144, 0.944289,
  0.951267, 0.959372, 0.969091, 0.98133, 0.998091, 1.02564, 1.05208, 1.11033}}
```

Вывод в виде таблицы

In[281]:=
```
Join[{{"Дов. вер.", "Нижняя гр.", "Верхняя гр."}}, Transpose[t3]] // MatrixForm
```

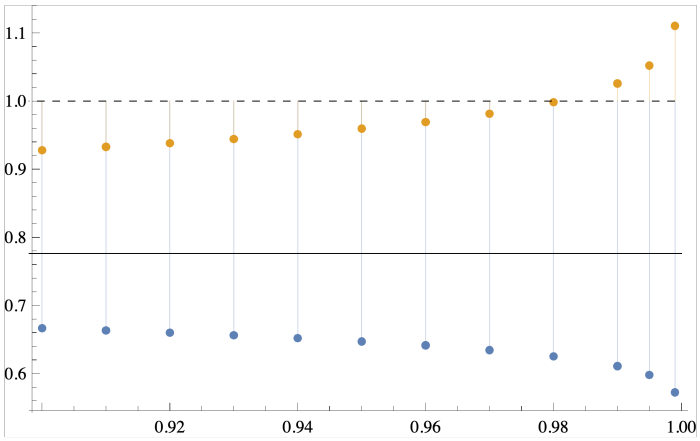
Out[281]//MatrixForm=

Дов. вер.	Нижняя гр.	Верхняя гр.
0.9	0.666568	0.927645
0.91	0.663374	0.932639
0.92	0.659905	0.938144
0.93	0.656091	0.944289
0.94	0.651835	0.951267
0.95	0.64699	0.959372
0.96	0.641313	0.969091
0.97	0.634362	0.98133
0.98	0.625186	0.998091
0.99	0.610901	1.02564
0.995	0.598053	1.05208
0.999	0.572339	1.11033

Графическое изображение доверительных интервалов, чем больше доверительная вероятность, тем больше доверительный интервал и, тем самым, хуже точность определения параметра. Приведены также оценка дисперсии sig^2, и значение, использованное в симуляции для построения выборки. Определить, попадает ли значение, использованное для симуляции в построенный доверительный интервал

In[282]:=
```
ListPlot[{Transpose[{t3[[1]], t3[[2]]}], Transpose[{t3[[1]], t3[[3]]}]}, Filling → 1,
 Epilog → {Line[{{0, sig^2}, {1, sig^2}}], Dashed, Line[{{0, σ^2}, {1, σ^2}}]}]
```

Out[282]=



■ Оценка необходимого объема выборки

Заданная точность p=5% для оценки математического ожидания
Исходное уравнение для оценки n

$\frac{\text{sig}}{\sqrt{n-1}} t_{\frac{1+\gamma}{2},n-1} = \text{mean} * p$, где $t_{\alpha,n}$ - квантиль t-распределения Стьюдента

In[283]:=
```
p = 0.05
```

Out[283]=
```
0.05
```

Решим задачу для надежности 0.9

In[284]:=
```
γ1[[1]]
```

Out[284]=
```
0.9
```

In[287]:=
```
FindRoot[
 sig/Sqrt[nn - 1] Quantile[StudentTDistribution[nn - 1], (1 + γ1[[1]])/2] == mean * p, {nn, 200}]
```

FindRoot: Failed to converge to the requested accuracy or precision within 100 iterations.

Out[287]=
```
{nn → 211.365}
```

Результат - для достижения 5% точности с заданной надежностью 0.9 (доверительной вероятностью) необходимо получить выборку объема не менее 212.

Проведем вычисления для всех заданных надежностей (доверительных вероятностей)

(Вычисления могут занять значительное время)

In[288]:=
```
Table[{γ1[[i]],
  FindRoot[sig/Sqrt[nn - 1] Quantile[StudentTDistribution[nn - 1], (1 + γ1[[i]])/2] == mean * p,
   {nn, 200}]}, {i, 1, Length[γ1]}] // TableForm
```

FindRoot: Failed to converge to the requested accuracy or precision within 100 iterations.

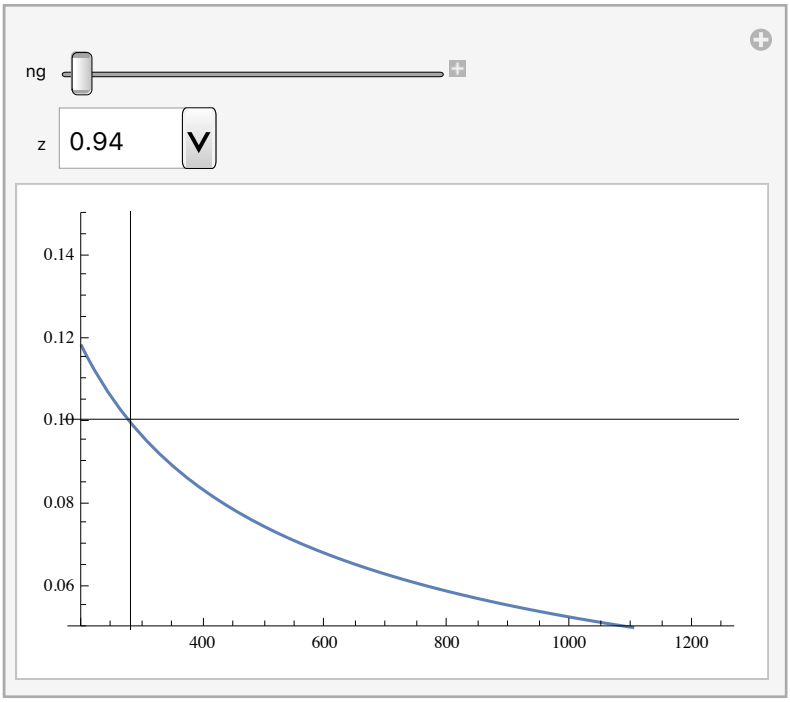
FindRoot: Failed to converge to the requested accuracy or precision within 100 iterations.

Out[288]//TableForm=

0.9	nn → 211.365
0.91	nn → 224.295
0.92	nn → 239.063
0.93	nn → 255.969
0.94	nn → 275.685
0.95	nn → 299.255
0.96	nn → 328.432
0.97	nn → 366.521
0.98	nn → 420.981
0.99	nn → 515.777
0.995	nn → 612.243
0.999	nn → 840.755

In[289]:=
```
Manipulate[Plot[sig/Sqrt[nn - 1] Quantile[StudentTDistribution[nn - 1], (1 + z)/2],
  {nn, 200, 1250}, PlotRange → {0.05, 0.15},
  Epilog → {Line[{{1, mean * p}, {1500, mean * p}}], Line[{{ng, -1}, {ng, 1}}]}],
 {ng, 280, 1200}, {z, γ1}]
```

Out[289]=



При выборках большого объема уравнение можно заменить на более простое

$\frac{\text{sig}}{\sqrt{n}} t_{\frac{1+\gamma}{2}} = \text{mean} * p$, где t_α- квантиль стандартного нормального закона

```
Table[{γ1[[i]], FindRoot[sig/Sqrt[nn] Quantile[NormalDistribution[], (1 + γ1[[i]])/2] == mean p,
    {nn, 500}]}, {i, 1, Length[γ1]}] // TableForm
```

```
0.9      nn → 289.968
0.91     nn → 308.063
0.92     nn → 328.483
0.93     nn → 351.86
0.94     nn → 379.122
0.95     nn → 411.711
0.96     nn → 452.055
0.97     nn → 504.722
0.98     nn → 580.024
0.99     nn → 711.1
0.995    nn → 844.484
0.999    nn → 1160.45
```

Как видно, отличия этих двух таблиц незначительно.

In[230]:=
```
ClearAll["Global`*"]
```

Проверка гипотезы о равенстве средних двух генеральных совокупностей

Математическая модель

1. Даны две выборки - данные по расходу сырья на единицу продукции в зависимости от использования новой и старой технологии. Необходимо определить, является ли новая технология более эффективнее или нет.

Если снять комментарий со следующих команд, то можно автоматически создать необходимый файл. После однократного выполнения этих команд необходимо восстановить комментарий, поставив символы (* *)

In[231]:=
```
(* data={{303,307,307,307,307,308,308,308,308},
  {303,303,304,304,304,304,304,304,306,306,306,306,308}};
Export[FileNameJoin[{NotebookDirectory[],"data.txt"}],data,"CSV"];  *)
```

1.1. Загрузить данные из файла data.txt, расположенного в той же папке, что и рабочая книга

1.2. Разбить загруженные данные на старые - первая строка загруженной таблицы, и новые - вторая строка загруженной таблицы

1.3. Определить выборочные средние и выборочные дисперсии для полученных выборок

1.4. Проверить нулевую гипотезу H_0 о равенстве генеральных средних. В качестве альтернативы берется гипотеза о преимуществе новой технологии над старой. Для этого составить статистику

$$\frac{\bar{X}-\bar{Y}}{\sqrt{\left(\frac{1}{n_x}+\frac{1}{n_y}\right)\frac{S_x^2 n_x+S_y^2 n_y}{n_x+n_y-2}}},$$

которая должна иметь t-распределения Стьюдента с $\left(n_x + n_y - 2\right)$ степенями свободы

1.5. Задать ошибку первого рода α и вычислить квантиль уровня (1-α) для t-распределения с $\left(n_x + n_y - 2\right)$ степенями свободы

1.6. Сравнить полученные значения и сделать вывод о принятии или отклонении гипотезы H_0

2. Во время эпидемии гриппа изучалась эффективность прививок против этого заболевания. Получены следующие данные

После прививки		**Без прививки**	
Заболели	**Не заболели**	**Заболели**	**Не заболели**
4 чел .	**192 чел .**	**34 чел .**	**111 чел.**

2.1. Указывают ли эти данные на эффективность прививок? Уровень значимости принять равным 0.01

Использование средств пакета Mathematica

- Определение функции и переменных в виде отложенного выражения
- Функция генерации псевдослучайных чисел RandomReal[]
- Функция определения нормального закона распределения NormalDistribution[]
- Функция определения t-распределения Стьюдента StudentTDistribution[]

- Функция определения χ^2-распределения ChiSquareDistribution
- Функция для обратной плотности вероятности закона распределения InverseCDF[]
- Функция определения среднего значения списка Mean[]
- Функция определения дисперсии списка Variance[]
- Функция определения суммы элементов списка Total[]
- Функция построения таблицы значений Table[]
- Функция объединения списков Join[]
- Функция определения числа элементов в списке Lenght[]
- Функция построения точечного графика ListPlot[]
-Функция определения корня уравнения FindRoot[]

Пример решения задачи

Загрузка двух выборок, данные по старой технологии и данные по новой технологии

Данные хранятся в текстовом файле data.txt в виде двух строк, данные в каждой строке отделены друг от друга запятой

In[232]:=
```
data = Import[FileNameJoin[{NotebookDirectory[], "data.txt"}], "CSV"]
```

Out[232]=
```
{{303, 307, 307, 307, 307, 308, 308, 308, 308},
 {303, 303, 304, 304, 304, 304, 304, 304, 306, 306, 306, 306, 308}}
```

In[233]:=
```
dataOld = data[[1]] (* Старая технология*)
dataNew = data[[2]](* Новая технология*)
```

Out[233]=
```
{303, 307, 307, 307, 307, 308, 308, 308, 308}
```

Out[234]=
```
{303, 303, 304, 304, 304, 304, 304, 304, 306, 306, 306, 306, 308}
```

Вычисление выборочных характеристик

Объем каждой из выборок

In[235]:=
```
nOld = Length[dataOld]
nNew = Length[dataNew]
```

Out[235]=
```
9
```

Out[236]=
```
13
```

Выборочные средние и дисперсия

In[237]:=
```
meanOld = Mean[dataOld] // N
meanNew = Mean[dataNew] // N
```

Out[237]=
```
307.
```

Out[238]=
```
304.769
```

In[239]:=
```
dispOld = Variance[dataOld] // N
dispNew = Variance[dataNew] // N
```

Out[239]=
```
2.5
```

Out[240]=
```
2.19231
```

Вычисление статистики и сравнение ее значения с критическим

Проверка нулевой гипотезы H_0 - средние значения используемого сырья одинаковы

Статистика

In[241]:=
$$\text{stat} = \frac{\text{meanOld} - \text{meanNew}}{\sqrt{\left(\frac{1}{\text{nOld}} + \frac{1}{\text{nNew}}\right) \frac{\text{dispOld nOld} + \text{dispNew nNew}}{\text{nOld} + \text{nNew} - 2}}}$$

Out[241]= 3.22156

Уровень значимости

In[242]:=
```
α = 0.05
```

Out[242]= 0.05

In[243]:=
```
krit = Quantile[StudentTDistribution[nOld + nNew - 2], 1 - α]
```

Out[243]= 1.72472

In[244]:=
```
If[stat < krit, "Гипотеза принимается", "Гипотеза отклоняется"]
```

Out[244]= Гипотеза отклоняется

■ Эффективность прививок

Гипотеза H_0 – эффекта о прививок нет, то есть вероятности заболеваний одинаковы

In[245]:=
```
vyb1 = {4, 192}
vyb2 = {34, 111}
```

Out[245]= {4, 192}

Out[246]= {34, 111}

In[247]:=
```
n1 = Total[vyb1]
n2 = Total[vyb2]
```

Out[247]= 196

Out[248]= 145

In[249]:=
$$\text{w1} = \frac{\text{vyb1[[1]]}}{\text{n1}} \text{ // N}$$

$$\text{w2} = \frac{\text{vyb2[[1]]}}{\text{n2}} \text{ // N}$$

Out[249]= 0.0204082

Out[250]= 0.234483

In[251]:=
$$\text{pocen} = \frac{\text{vyb1[[1]]} + \text{vyb2[[1]]}}{\text{n1} + \text{n2}} \text{ // N}$$

Out[251]= 0.111437

In[252]:=
$$\text{statis} = \text{Abs}\left[\frac{\text{w1} - \text{w2}}{\sqrt{\left(\frac{1}{\text{n1}} + \frac{1}{\text{n2}}\right) \text{pocen (1 - pocen)}}}\right]$$

Out[252]= 6.21071

In[253]:=
```
α0 = 0.01
```

$$\text{krits} = \text{Quantile}\left[\text{NormalDistribution[]}, 1 - \frac{\alpha 0}{2}\right]$$

Out[253]= 0.01

Out[254]= 2.57583

In[255]:=
```
If[statis < krits, "Гипотеза принимается", "Гипотеза отклоняется"]
```

Out[255]= Гипотеза отклоняется

In[305]:=
```
ClearAll["Global`*"]
```

Зависимость между стоимостью жилья Y(долларах) и размером жилой площади X(квадратные метры)

Для выполнения задания необходим дополнительно файл с именем flat.xls, должен находитmся в той же папке.

- **Как задавать исходные данные**

Данные можно импортировать либо из текстового файла с расширением .CSV или .TXT (необходимо указывать атрибут “Data” при импорте). (Данные такого вида можно подготовить в любом текстовом редакторе, включая Word, Excel или OpenOffice),
либо прямо из электронной таблицы - файл с расширением .XLS (Excel или OpenOffice)
Если таких файлов нет, то можно непосредственно вводить данные в пакете *Mathematica* в виде матрицы.

Загрузка электронной таблицы - получается трехмерный массив (массив размерности 3), предварительно необходимо из результата извлечь первый элемент, далее результат такой же как и в первом случае

In[306]:=
```
datatxt0 = Import[FileNameJoin[{NotebookDirectory[], "flat.xls"}], "Data" ];
datatxt = datatxt0[[1]]
```

Out[307]=
```
{{Жилая площадь (кв. метры), Стоимость ( долл)}, {30.2, 5000.},
 {32., 5200.}, {32., 5350.}, {37., 5880.}, {30., 5430.}, {30., 5430.},
 {30., 5430.}, {29., 5350.}, {33., 5740.}, {31., 5570.}, {30., 5530.},
 {34., 6020.}, {38., 7010.}, {31., 6420.}, {39., 7150.}, {39.5, 7190.}}
```

In[308]:=
```
nd = Length[datatxt]
```

Out[308]=
```
17
```

Получение исходных данных из введенного datatxt

In[309]:=
```
data = datatxt[[2 ;; nd]]
```

Out[309]=
```
{{30.2, 5000.}, {32., 5200.}, {32., 5350.}, {37., 5880.}, {30., 5430.},
 {30., 5430.}, {30., 5430.}, {29., 5350.}, {33., 5740.}, {31., 5570.}, {30., 5530.},
 {34., 6020.}, {38., 7010.}, {31., 6420.}, {39., 7150.}, {39.5, 7190.}}
```

Определение границ переменных на графике

In[310]:=
```
Min[data[[All, 1]]]
Max[data[[All, 1]]]
```

Out[310]=
```
29.
```

Out[311]=
```
39.5
```

In[312]:=
```
Min[data[[All, 2]]]
Max[data[[All, 2]]]
```

Out[312]=
```
5000.
```

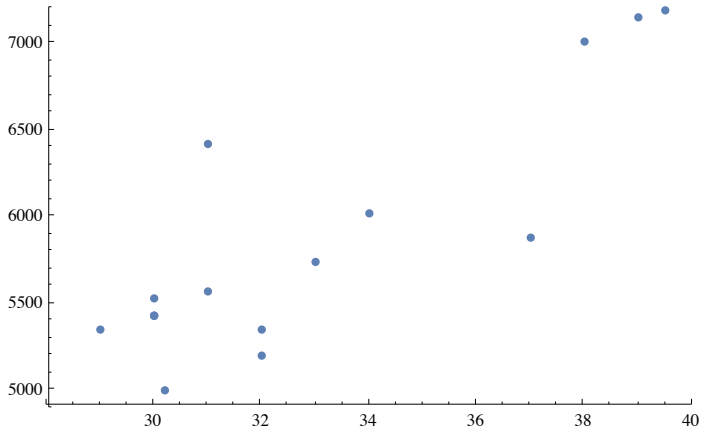
Out[313]=
```
7190.
```

- **Облако наблюдения - визуализация исходных данных**

Облако наблюдения или Поле корреляции или Диаграмма рассеяния

In[314]:=
```
plot1 = ListPlot[data, PlotRange → {{28, 40}, {4900, 7200}}]
```

Out[314]=



- **Прямое построение нормальных уравнений и их решения**

Построение массива X и Y

In[315]:=
```
{xdata, ydata} = {data[[All, 1]], data[[All, 2]]}
ndata = Length[xdata]
```

Out[315]=
```
{{30.2, 32., 32., 37., 30., 30., 30., 29., 33., 31., 30., 34., 38., 31., 39., 39.5},
 {5000., 5200., 5350., 5880., 5430., 5430., 5430., 5350.,
  5740., 5570., 5530., 6020., 7010., 6420., 7150., 7190.}}
```

Out[316]=
```
16
```

Построение системы нормальных уравнений

In[317]:=
```
ma = ( 1                 Mean[xdata]
       Mean[xdata]   xdata.xdata/ndata )
mb = {Mean[ydata], xdata.ydata/ndata}
```

Out[317]=
```
{{1, 32.8563}, {32.8563, 1091.39}}
```

Out[318]=
```
{5856.25, 194 433.}
```

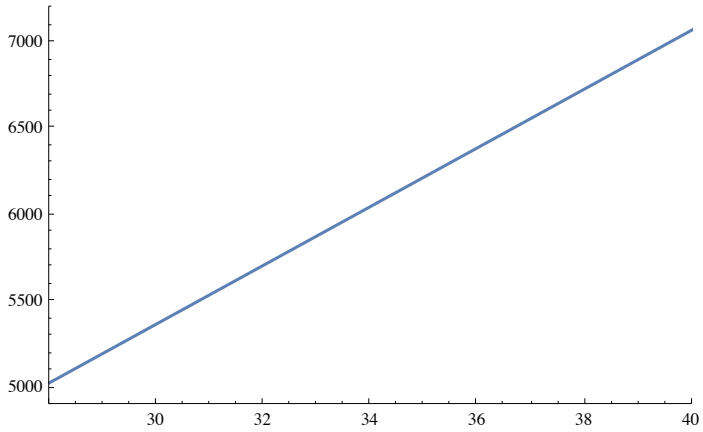
In[319]:=
```
{alfaReg, betaReg} = LinearSolve[ma, mb] // N
```

Out[319]=
```
{262.847, 170.239}
```

Прямая регрессии

In[320]:=
```
plot2 = Plot[alfaReg + betaReg x, {x, 28, 40}, PlotRange → {{28, 40}, {4900, 7200}}]
```

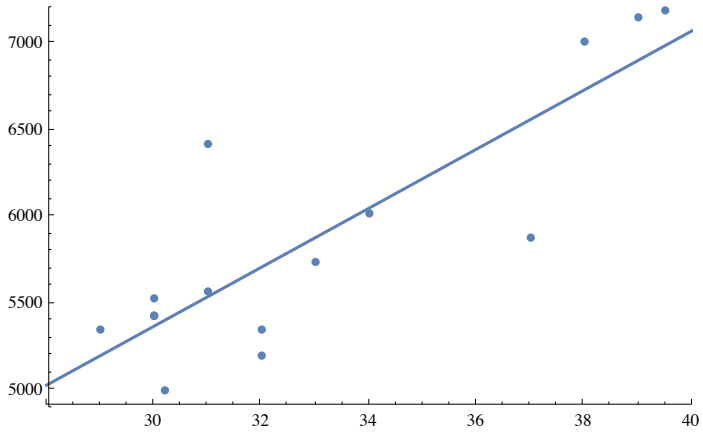
Out[320]=



Исходные данные и прямая регрессии

In[321]:=
```
Show[plot1, plot2]
```

Out[321]=



- **Интерпретация коэффициента регрессии betaReg.**

  При изменении жилой площади на 1 метр стоимость жилья увеличивается (так как betaReg > 0) на betaReg = 170.239 долларов.

In[322]:=
```
betaReg
```

Out[322]=
```
170.239
```

- **Решение системы нормальных уравнений матричным методом**

  Матрица X

In[323]:=
```
matX = Transpose[{Array[1.0 &, ndata], xdata}]
```

Out[323]=
```
{{1., 30.2}, {1., 32.}, {1., 32.}, {1., 37.}, {1., 30.}, {1., 30.}, {1., 30.}, {1., 29.},
 {1., 33.}, {1., 31.}, {1., 30.}, {1., 34.}, {1., 38.}, {1., 31.}, {1., 39.}, {1., 39.5}}
```

  Матрица $X^T X$

In[324]:=
```
xtx = Transpose[matX].matX
```

Out[324]=
```
{{16., 525.7}, {525.7, 17462.3}}
```

  Определение вектора Y

In[325]:=
```
vy = ydata
```

Out[325]=
```
{5000., 5200., 5350., 5880., 5430., 5430., 5430.,
 5350., 5740., 5570., 5530., 6020., 7010., 6420., 7150., 7190.}
```

  Вектор $X^T Y$

In[326]:=
```
xy = Transpose[matX].vy
```

Out[326]=
```
{93700., 3.11094 × 10^6}
```

  Определение матрицы $\left(X^T X\right)^{-1}$

In[327]:=
```
invx = Inverse[Transpose[matX].matX] // N
```

Out[327]=
```
{{5.75146, -0.173147}, {-0.173147, 0.00526983}}
```

  Определение коэффициентов регрессии $\left(X^T X\right)^{-1} X^T Y$

In[328]:=
```
{b0, b1} = invx.xy
```

Out[328]=
```
{262.847, 170.239}
```

  После получения матричного уравнения для его решения можно использовать и стандартные методы (метод Гаусса) решения системы линейных уравнений

In[329]:=
```
LinearSolve[Transpose[matX].matX, xy] // N
```

Out[329]=
```
{262.847, 170.239}
```

- **Вычисление коэффициента детерминации - определение качества построенного уравнения регрессии**

Вычисление среднего значения Y

In[330]:=
```
my = Mean[ydata]
```

Out[330]=
```
5856.25
```

```
5856.25
```

TSS - общая дисперсия

In[331]:=
```
tss = (ydata - my).(ydata - my)
```

Out[331]=
```
7.55238 × 10^6
```

```
7.55238 × 10^6
```

ESS - остаточная дисперсия

In[332]:=
```
yregr = b0 + b1 xdata
```

Out[332]=
```
{5404.05, 5710.48, 5710.48, 6561.68, 5370.01, 5370.01, 5370.01, 5199.77,
 5880.72, 5540.24, 5370.01, 6050.96, 6731.91, 5540.24, 6902.15, 6987.27}
```

```
{5404.05, 5710.48, 5710.48, 6561.68, 5370.01, 5370.01, 5370.01, 5199.77,
 5880.72, 5540.24, 5370.01, 6050.96, 6731.91, 5540.24, 6902.15, 6987.27}
```

In[333]:=
```
ess = (ydata - yregr).(ydata - yregr)
```

Out[333]=
```
2.05292 × 10^6
```

```
2.05292 × 10^6
```

RSS - объясненная часть дисперсии

In[334]:=
```
rss = (yregr - my).(yregr - my)
```

Out[334]=
```
5.49945 × 10^6
```

```
5.49945 × 10^6
```

Коэффициент детерминации $-R^2$

In[335]:=
$$\text{rsquare} = \frac{\text{rss}}{\text{tss}}$$

Out[335]=
```
0.728175
```

```
0.728175
```

In[336]:=
$$1 - \frac{\text{ess}}{\text{tss}}$$

Out[336]=
```
0.728175
```

```
0.728175
```

Вывод - значение коэффициента детерминации достаточно велико, следовательно уравнение регрессии удовлетворительно объясняет исходные экспериментальные данные.

Для парной регрессии коэффициент детерминации связан с коэффициентом корреляции и коэффициентом регрессии

Коэффициент корреляции r - положителен и достаточно велик

In[337]:=
```
sigmax = Sqrt[Mean[(xdata - Mean[xdata])^2]]
sigmay = Sqrt[Mean[(ydata - Mean[ydata])^2]]
r = (Mean[xdata * ydata] - Mean[xdata] Mean[ydata]) / (sigmax sigmay)
```

Out[337]= 3.44383

Out[338]= 687.04

Out[339]= 0.853332

Его квадрат совпадает с коэффициентом детерминации R^2

In[340]:=
```
r^2
```

Out[340]= 0.728175

Коэффициент корреляции выражается через коэффициент регрессии (важно, что их знаки совпадают)

In[341]:=
```
betaReg sigmax/sigmay
```

Out[341]= 0.853332

In[342]:=
```
ClearAll["Global`*"]
```

Имеются данные о годовых ставках месячных доходов по трем акциям за шестимесячный период (файл Акции.xls).
Есть основания предполагать, что доходы Y по акциям типа C зависят от доходов X1 и X2 по акциям типа A и B.
Необходимо:
а) Построить линейную регрессионную модель с помощью LinearModelFit
б) Получить уравнение линейной регрессии и значения параметров регрессионной модели
в) Проверить исходные данные, матрицу независимых переменных и зависимую переменную
г) Вычислить предсказанные значения зависимой переменной, построить график исходного и предсказанного значения
д) Построить график остатков модели в зависимости от номера наблюдения
е) Определить статистические характеристики остатков
ж) Определить значения F-статистик и определить значимость регрессионных переменных
з) Определить значение F-статистики для регрессии в целом и определить ее значимость
и) Определить доверительные интервалы для доверительной вероятности 0.9 и 0.95
к) Определить значения F-статистик для коэффициентов и определить значимость коэффициентов регрессии
л) Определить ковариационную матрицу коэффициентов модели
м) Определить ковариационную и корреляционную матрицу факторов и проверить их мультиколлинеарность

■ Ввод данных

In[343]:=
```
datatxt0 = Import[FileNameJoin[{NotebookDirectory[], "Акции.xls"}]];
datatxt = datatxt0[[1]]
nn = Length[datatxt]
```

Out[344]=
```
{{Доход по месяцам, , }, {Акции типа A, Акции типа Б , Акции типа C}, {5.4, 6.3, 9.2},
 {5.3, 6.2, 9.2}, {4.9, 6.1, 9.1}, {4.9, 5.8, 9.}, {5.4, 5.7, 8.7}, {6., 5.7, 8.6}}
```

Out[345]= 8

In[346]:=
```
data = datatxt[[3 ;; nn]]
```

Out[346]=
```
{{5.4, 6.3, 9.2}, {5.3, 6.2, 9.2}, {4.9, 6.1, 9.1},
 {4.9, 5.8, 9.}, {5.4, 5.7, 8.7}, {6., 5.7, 8.6}}
```

In[347]:=
```
ndata = Length[data]
```

Out[347]= 6

а) Построить линейную регрессионную модель с помощью LinearModelFit

In[348]:=
```
lmf = LinearModelFit[data, {1, x1, x2}, {x1, x2}]
```

Out[348]= FittedModel[«1»]

б) Получить уравнение линейной регрессии и значения параметров регрессионной модели

In[349]:=
```
Normal[lmf]
```

Out[349]= 5.6154 − 0.238581 x1 + 0.774254 x2

In[350]:=
```
fregression[x1_, x2_] = lmf["BestFit"]
```

Out[350]= 5.6154 − 0.238581 x1 + 0.774254 x2

In[351]:=
```
lmf["BestFit"]
```

Out[351]= 5.6154 − 0.238581 x1 + 0.774254 x2

In[352]:=
```
lmf["BestFitParameters"]
```

Out[352]= {5.6154, −0.238581, 0.774254}

в) Проверить исходные данные, матрицу независимых переменных и зависимую переменную

In[353]:=
```
lmf["Data"] // MatrixForm
```

Out[353]//MatrixForm=

$$\begin{pmatrix} 5.4 & 6.3 & 9.2 \\ 5.3 & 6.2 & 9.2 \\ 4.9 & 6.1 & 9.1 \\ 4.9 & 5.8 & 9. \\ 5.4 & 5.7 & 8.7 \\ 6. & 5.7 & 8.6 \end{pmatrix}$$

In[354]:=
```
lmf["DesignMatrix"] // MatrixForm
```

Out[354]//MatrixForm=

$$\begin{pmatrix} 1. & 5.4 & 6.3 \\ 1. & 5.3 & 6.2 \\ 1. & 4.9 & 6.1 \\ 1. & 4.9 & 5.8 \\ 1. & 5.4 & 5.7 \\ 1. & 6. & 5.7 \end{pmatrix}$$

In[355]:=
```
lmf["Response"] // MatrixForm
```

Out[355]//MatrixForm=

$$\begin{pmatrix} 9.2 \\ 9.2 \\ 9.1 \\ 9. \\ 8.7 \\ 8.6 \end{pmatrix}$$

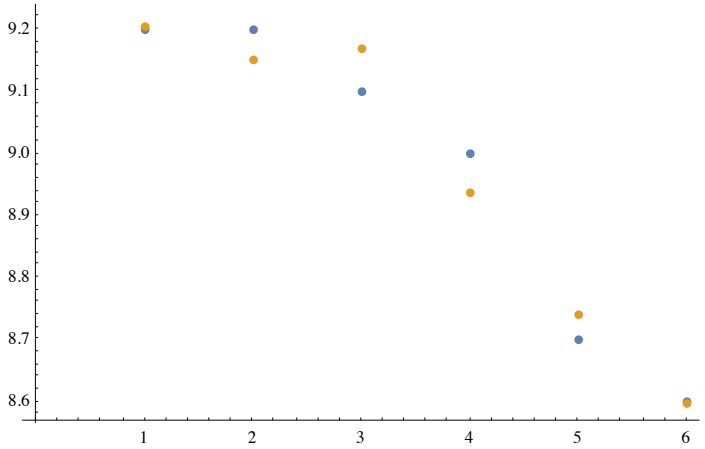
г) Вычислить предсказанные значения зависимой переменной, построить график исходного и предсказанного значения

In[356]:=
```
fregression[lmf["Data"][[All, 1]], lmf["Data"][[All, 2]]]
```

Out[356]=
```
{9.20487, 9.1513, 9.16931, 8.93703, 8.74032, 8.59717}
```

In[357]:=
```
ListPlot[{lmf["Response"], fregression[lmf["Data"][[All, 1]], lmf["Data"][[All, 2]]]}]
```

Out[357]=



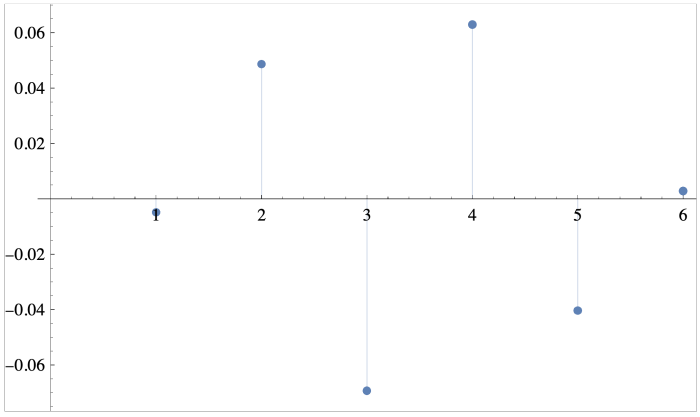
д) Построить график остатков модели в зависимости от номера наблюдения

In[358]:=
```
lmf["FitResiduals"]
```

Out[358]=
```
{-0.00486976, 0.0486976, -0.0693092, 0.0629672, -0.0403171, 0.00283126}
```

In[359]:=
```
ListPlot[lmf["FitResiduals"], Filling → Axis]
```

Out[359]=



е) Определить статистические характеристики остатков

In[360]:=
```
lmf["ANOVATable"]
```

Out[360]=

	DF	SS	MS	F-Statistic	P-Value
x1	1	0.128826	0.128826	30.2001	0.0118568
x2	1	0.19171	0.19171	44.9415	0.00677267
Error	3	0.0127973	0.00426576		
Total	5	0.333333			

ж) Определить значения F - статистик и определить значимость регрессионных переменных

In[361]:=
```
lmf["ANOVATableFStatistics"]
```

Out[361]=
```
{30.2001, 44.9415}
```

In[362]:=
```
lmf["ANOVATablePValues"]
```

Out[362]=
```
{0.0118568, 0.00677267}
```

При уровне значимости 5% значения P-values меньше 0.05. Следовательно, обе переменные являются значимыми

з) Определить значение F-статистики для регрессии в целом и определить ее значимость

In[363]:=
```
lmf["RSquared"]
```

Out[363]=
```
0.961608
```

In[364]:=
```
fstat1 = lmf["RSquared"] / (1 - lmf["RSquared"]) (ndata - 2 - 1) / 2
```

Out[364]=
```
37.5708
```

In[365]:=
```
alfa = 0.05
```

Out[365]=
```
0.05
```

In[366]:=
```
fkrit1 = InverseCDF[FRatioDistribution[2, ndata - 2 - 1], 1 - alfa]
```

Out[366]=
```
9.55209
```

In[367]:=
```
fstat1 > fkrit1
```

Out[367]=
```
True
```

Значение fstat1 больше, чем критические значения, поэтому регрессия значима на уровне значимости 5%

и) Определить доверительные интервалы для доверительной вероятности 0.9 и 0.95

In[368]:=
```
p1 = 0.9
p2 = 0.95
```

Out[368]= 0.9

Out[369]= 0.95

In[370]:=
```
lmf["ParameterConfidenceIntervals", ConfidenceLevel → p1]
```

Out[370]= {{3.50169, 7.72912}, {-0.416097, -0.0610642}, {0.502455, 1.04605}}

При доверительной вероятности p2=0.95 - границы доверительного интервала для первого коэффициента включают значение ноль, что еще раз свидетельствует о не значимости первого коэффициента

In[371]:=
```
lmf["ParameterConfidenceIntervals", ConfidenceLevel → p2]
```

Out[371]= {{2.75703, 8.47377}, {-0.478636, 0.00147435}, {0.406701, 1.14181}}

к) Определить значения Т-статистик для коэффициентов и определить значимость коэффициентов регрессии

In[372]:=
```
lmf["ParameterTable"]
```

Out[372]=

	Estimate	Standard Error	t-Statistic	P-Value
1	5.6154	0.898167	6.25207	0.00825621
x1	−0.238581	0.0754309	−3.1629	0.0507575
x2	0.774254	0.115494	6.70384	0.00677267

На уровне значимости 5% коэффициент b2 - значим, а коэффициент b1 - не значим

л) Определить ковариационную матрицу коэффициентов модели

In[373]:=
```
lmf["CovarianceMatrix"] // MatrixForm
```

Out[373]//MatrixForm=

$$\begin{pmatrix} 0.806705 & -0.0462647 & -0.093858 \\ -0.0462647 & 0.00568983 & 0.00268388 \\ -0.093858 & 0.00268388 & 0.0133389 \end{pmatrix}$$

In[374]:=
```
lmf["CorrelationMatrix"] // MatrixForm
```

Out[374]//MatrixForm=

$$\begin{pmatrix} 1. & -0.682878 & -0.904803 \\ -0.682878 & 1. & 0.308073 \\ -0.904803 & 0.308073 & 1. \end{pmatrix}$$

м) Определить ковариационную и корреляционную матрицу факторов и проверить их мультиколлинеарность

In[375]:=
```
Covariance[data[[All, {1, 2}]]]
```

Out[375]= {{0.165667, -0.0333333}, {-0.0333333, 0.0706667}}

In[376]:=
```
Correlation[data[[All, {1, 2}]]]
```

Out[376]= {{1., -0.308073}, {-0.308073, 1.}}

Коэффициент корреляции факторов X1 и X2 равен -0.308073, что меньше по модулю 0.8, следовательно, можно считать, что мультиколлинеарность слабая.

In[377]:=
```
ClearAll["Global`*"]
```

Имеются данные о весе, возрасте 13 индюшек, выращенных в областях A, B, C. Есть основания предполагать, что на вес индюшек оказывает влияние не только их возраст, но и область происхождения. Необходимо

а) найти уравнение парной регрессии Y на X и оценить его значимость,

б) введя соотвествующие фиктивные переменные, найти общее уравнение множественной регрессии Y по всем объясняющим переменным, включая фиктивные,

в) оценить значимость общего уравнения множественной регрессии по F-критерию и значимость его коэффициентов по T-критерию на уровне 5%,

г) проследить за изменением скорректированного коэффициента детерминации при переходе от парной к множественной регрессии,

д) оценить совместную объясняющую способность фиктивных переменных,

е) оценить на уровне 5% значимость различия между свободными членами уравнений , получаемых из общего уравнения множественной регрессии Y для каждой области.

- **Ввод данных**

In[378]:=
```
datatxt0 = Import[FileNameJoin[{NotebookDirectory[], "Индейки.xls"}]];
datatxt = datatxt0[[1]] ;
nn = Length[datatxt]
```

Out[380]=
```
15
```

In[381]:=
```
datatxt
```

Out[381]=
```
{{Имеются данные о весе, возрасте 13 индюшек, выращенных в областях
    A, B, C. Есть основания предполагать, что на вес индюшек оказывает
    влияние не только их возраст, но и область происхождения., , },
 {Возраст в неделях, Вес в фунтах, Область происхождения},
 {28., 13.3, A},
 {20., 8.9, A},
 {32., 15.1, A},
 {22., 10.4, A},
 {29., 13.1, B},
 {27., 12.4, B}, {28., 13.2, B},
 {26., 11.8, B}, {21., 11.5, C},
 {27., 14.2, C}, {29., 15.4, C},
 {23., 13.1, C}, {25., 13.8, C}}
```

In[382]:=
```
dataf = datatxt[[3 ;; nn]]
```

Out[382]=
```
{{28., 13.3, A}, {20., 8.9, A}, {32., 15.1, A}, {22., 10.4, A},
 {29., 13.1, B}, {27., 12.4, B}, {28., 13.2, B}, {26., 11.8, B},
 {21., 11.5, C}, {27., 14.2, C}, {29., 15.4, C}, {23., 13.1, C}, {25., 13.8, C}}
```

In[383]:=
```
ndata = Length[dataf]
```

Out[383]=
```
13
```

- **Введение фиктивных переменных**

Так вид области имеет три градации, необходимо ввести две фиктивные переменные Z1 и Z2,
Z1 = 1 для области A и 0 для остальных, Z2 = 1 для области B и 0 для остальных. то есть для области C переменные Z1 и Z2 обе равны 0.

Исходная переменная X1

In[384]:=
```
xx1 = dataf[[All, 1]]
```

Out[384]=
```
{28., 20., 32., 22., 29., 27., 28., 26., 21., 27., 29., 23., 25.}
```

Фиктивная переменная Z1

In[385]:=
```
zz1 = dataf[[All, 3]] /. {"A" → 1, "B" → 0, "C" → 0}
```

Out[385]=
```
{1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0}
```

Фиктивная переменная Z2

In[386]:=
```
zz2 = dataf[[All, 3]] /. {"A" → 0, "B" → 1, "C" → 0}
```

Out[386]=
```
{0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0}
```

Объясняемая переменная

In[387]:=
```
y = dataf[[All, 2]]
```

Out[387]=
```
{13.3, 8.9, 15.1, 10.4, 13.1, 12.4, 13.2, 11.8, 11.5, 14.2, 15.4, 13.1, 13.8}
```

Исходные данные

In[388]:=
```
data = Transpose[{xx1, zz1, zz2, y}]
```

Out[388]=
```
{{28., 1, 0, 13.3}, {20., 1, 0, 8.9}, {32., 1, 0, 15.1},
 {22., 1, 0, 10.4}, {29., 0, 1, 13.1}, {27., 0, 1, 12.4},
 {28., 0, 1, 13.2}, {26., 0, 1, 11.8}, {21., 0, 0, 11.5},
 {27., 0, 0, 14.2}, {29., 0, 0, 15.4}, {23., 0, 0, 13.1}, {25., 0, 0, 13.8}}
```

In[389]:=
```
data // MatrixForm
```

Out[389]//MatrixForm=

28.	1	0	13.3
20.	1	0	8.9
32.	1	0	15.1
22.	1	0	10.4
29.	0	1	13.1
27.	0	1	12.4
28.	0	1	13.2
26.	0	1	11.8
21.	0	0	11.5
27.	0	0	14.2
29.	0	0	15.4
23.	0	0	13.1
25.	0	0	13.8

■ Построение моделей

■ а) найти уравнение парной регрессии Y на X и оценить его значимость,

In[390]:=
```
data1 = data[[All, {1, 4}]]
model1 = LinearModelFit[data1, {1, X1}, X1]
```

Out[390]=
```
{{28., 13.3}, {20., 8.9}, {32., 15.1}, {22., 10.4}, {29., 13.1}, {27., 12.4}, {28., 13.2},
 {26., 11.8}, {21., 11.5}, {27., 14.2}, {29., 15.4}, {23., 13.1}, {25., 13.8}}
```

Out[391]=
```
FittedModel[ 1.98333 «1» 0.416667 «1» «2» ]
```

In[392]:=
```
Normal[model1]
```

Out[392]=
```
1.98333 + 0.416667 X1
```

In[393]:=
```
model1["ANOVATable"]
model1["ParameterTable"]
model1["AdjustedRSquared"]
```

Out[393]=

	DF	SS	MS	F-Statistic	P-Value
X1	1	26.2019	26.2019	21.8102	0.000682393
Error	11	13.215	1.20136		
Total	12	39.4169			

Out[394]=

	Estimate	Standard Error	t-Statistic	P-Value
1	1.98333	2.33273	0.850218	0.413327
X1	0.416667	0.0892194	4.67013	0.000682393

Out[395]= 0.63426

In[396]:=
```
alfa = 0.05
fstat1 = (model1["RSquared"] / (1 - model1["RSquared"])) / (ndata - 1 - 1) / 1
fkrit1 = InverseCDF[FRatioDistribution[1, ndata - 1 - 1], 1 - alfa]
fstat1 > fkrit1
```

Out[396]= 0.05

Out[397]= 21.8102

Out[398]= 4.84434

Out[399]= True

- **б) введя соотвествующие фиктивные переменные, найти общее уравнение множественной регрессии Y по всем объясняющим переменным, включая фиктивные,**

In[400]:=
```
data2 = data
model2 = LinearModelFit[data2, {1, X1, Z1, Z2}, {X1, Z1, Z2}]
```

Out[400]=
```
{{28., 1, 0, 13.3}, {20., 1, 0, 8.9}, {32., 1, 0, 15.1},
 {22., 1, 0, 10.4}, {29., 0, 1, 13.1}, {27., 0, 1, 12.4},
 {28., 0, 1, 13.2}, {26., 0, 1, 11.8}, {21., 0, 0, 11.5},
 {27., 0, 0, 14.2}, {29., 0, 0, 15.4}, {23., 0, 0, 13.1}, {25., 0, 0, 13.8}}
```

Out[401]= FittedModel[«1»]

In[402]:=
```
Normal[model2]
```

Out[402]= 1.43088 + 0.486765 X1 − 1.91838 Z1 − 2.19191 Z2

In[403]:=
```
model2["ANOVATable"]
model2["ParameterTable"]
model2["AdjustedRSquared"]
```

Out[403]=

	DF	SS	MS	F-Statistic	P-Value
X1	1	26.2019	26.2019	290.71	3.69079×10^{-8}
Z1	1	2.71652	2.71652	30.1397	0.000385157
Z2	1	9.68731	9.68731	107.481	2.6481×10^{-6}
Error	9	0.811176	0.0901307		
Total	12	39.4169			

Out[404]=

	Estimate	Standard Error	t-Statistic	P-Value
1	1.43088	0.657442	2.17644	0.0575071
X1	0.486765	0.0257435	18.9083	1.48876×10^{-8}
Z1	−1.91838	0.201803	−9.50621	5.44523×10^{-6}
Z2	−2.19191	0.211426	−10.3673	2.6481×10^{-6}

Out[405]= 0.972561

- **в) оценить значимость общего уравнения множественной регрессии по F-критерию и значимость его коэффициентов по T-критерию на уровне 5%,**

In[406]:=
```
alfa = 0.05
fstat2 = (model2["RSquared"] / (1 - model2["RSquared"])) * ((ndata - 3 - 1) / 3)
fkrit2 = Quantile[FRatioDistribution[3, ndata - 3 - 1], 1 - alfa]
fstat2 > fkrit2
```

Out[406]= 0.05

Out[407]= 142.777

Out[408]= 3.86255

Out[409]= True

T-статистики и их значимость для коэффициентов уравнения регрессии

In[410]:=
```
model2["ParameterTStatistics"]
```

Out[410]= {2.17644, 18.9083, -9.50621, -10.3673}

In[411]:=
```
model2["ParameterPValues"]
```

Out[411]= $\{0.0575071, 1.48876 \times 10^{-8}, 5.44523 \times 10^{-6}, 2.6481 \times 10^{-6}\}$

- **г) проследить за изменением скорректированного коэффициента детерминации при переходе от парной к множественной регрессии,**

Скорректированный коэффициент детерминации для исходной модели

In[412]:=
```
model1["AdjustedRSquared"]
```

Out[412]= 0.63426

Скорректированный коэффициент детерминации для модели с фиктивными переменными

In[413]:=
```
model2["AdjustedRSquared"]
```

Out[413]= 0.972561

Наблюдается значительное увеличение скорректированного коэффициента детерминации

- **д) оценить совместную объясняющую способность фиктивных переменных,**

Для проверки гипотезы о том, что все коэффициенты при фиктивных переменных равны 0, необходимо составить F-статистику, включающую сумму квадратов остатков в модели с фиктивными переменными и в модели без фиктивных переменных.

In[414]:=
```
model1["ANOVATableSumsOfSquares"]
```

Out[414]= {26.2019, 13.215, 39.4169}

In[415]:=
```
ostosn = model1["ANOVATableSumsOfSquares"][[2]]
```

Out[415]= 13.215

In[416]:=
```
model2["ANOVATableSumsOfSquares"]
```

Out[416]= {26.2019, 2.71652, 9.68731, 0.811176, 39.4169}

In[417]:=
```
ostfict = model2["ANOVATableSumsOfSquares"][[4]]
```

Out[417]= 0.811176

In[418]:=
```
fstatfict = ((ostosn-ostfict)/(3-1)) / (ostfict/(ndata-3-1))
```

Out[418]= 68.8102

In[419]:=
```
alfa = 0.05
fkritfict = Quantile[FRatioDistribution[3 - 1, ndata - 3 - 1], 1 - alfa]
fstatfict > fkritfict
```

Out[419]= 0.05

Out[420]= 4.25649

Out[421]= True

In[422]:=
```
pvalfict = 1 - CDF[FRatioDistribution[3 - 1, ndata - 3 - 1], fstatfict]
```

Out[422]= $3.517359818372 \times 10^{-6}$

In[423]:=
```
pvalfict < alfa
```

Out[423]= True

- **e) графики, показывающие все регрессии**

In[424]:=
```
datafA = Select[dataf, #[[3]] == "A" &]
```

Out[424]= {{28., 13.3, A}, {20., 8.9, A}, {32., 15.1, A}, {22., 10.4, A}}

In[425]:=
```
datafA = Select[dataf, #[[3]] == "A" &][[All, 1 ;; 2]]
```

Out[425]= {{28., 13.3}, {20., 8.9}, {32., 15.1}, {22., 10.4}}

In[426]:=
```
datafB = Select[dataf, #[[3]] == "B" &][[All, 1 ;; 2]]
```

Out[426]= {{29., 13.1}, {27., 12.4}, {28., 13.2}, {26., 11.8}}

In[427]:=
```
datafC = Select[dataf, #[[3]] == "C" &][[All, 1 ;; 2]]
```

Out[427]= {{21., 11.5}, {27., 14.2}, {29., 15.4}, {23., 13.1}, {25., 13.8}}

In[428]:=
```
reg1[X1_] = Normal[model1]
```

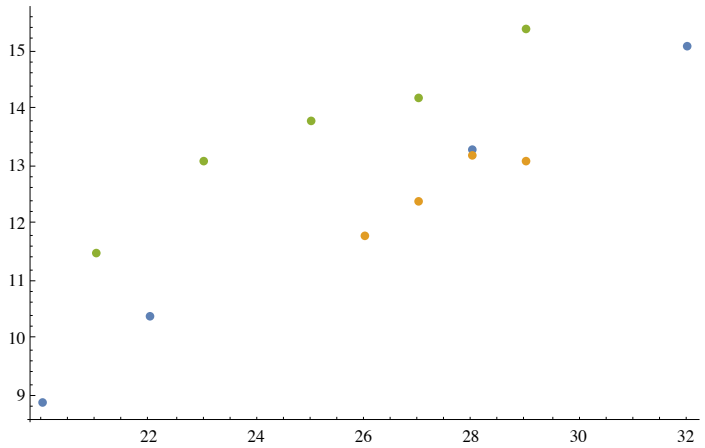
Out[428]= 1.98333 + 0.416667 X1

In[429]:=
```
reg2[X1_, Z1_, Z2_] = Normal[model2]
```

Out[429]= 1.43088 + 0.486765 X1 − 1.91838 Z1 − 2.19191 Z2

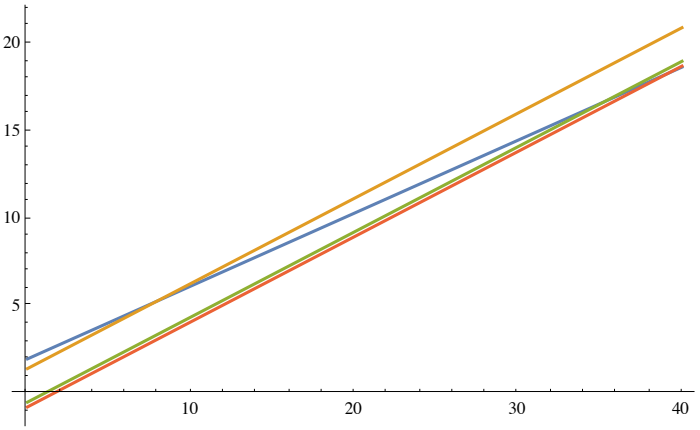
In[430]:=
```
g1 = ListPlot[{datafA, datafB, datafC}]
```

Out[430]=



In[431]:=
```
g2 = Plot[{reg1[x], reg2[x, 0, 0], reg2[x, 1, 0], reg2[x, 0, 1]}, {x, 0, 40}]
```

Out[431]=



In[432]:=
```
Show[g1, g2]
```

Out[432]=