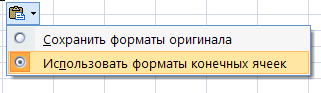
**Лабораторная работа 1**

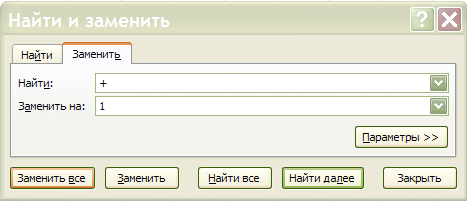
При создании использовалась монография «**Тестирование учебных достижений»** **В.С. КИМ** <http://uss.dvfu.ru/static/kim_testing_monograph/index.html>

1. Создаем нашу матрицу – копируем одну из предложенных в Excel.

(вставляем из Word). Мы взяли матрицу номер 2.



Удаляем столбцы Баллы и Оценка – они нам не нужны, хотя Баллы это и есть Xi. Но лучше пересчитать. Если ваша матрица такова, что там – и +, то сделаем замену на 0 и 1 соответственно. Для этого также скопируем матрицу и заменим: **Ctrl+F –** появится меню, выберем вкладку **Заменить.** Впишем нужные значения и нажмем **Заменить все**. Для минуса аналогично. (Найти- – Заменить на - 0)



1. Рассчитать ,

Xi – это индивидуальный бал испытуемого

рассчитывается как сумма по соответствующей строчке.

Можно воспользоваться **авто-суммой**  (в правом верхнем углу)

(главное не захватить в диапазон номер – с этим внимательнее, иначе у нас будут неверные результаты)



Затем формулу – метим ячейку с посчитанной суммой . Правый нижний угол – появится черный крестик – зажимаем левую кнопку и протягиваем вниз до последней строчки – Xi посчитаны.

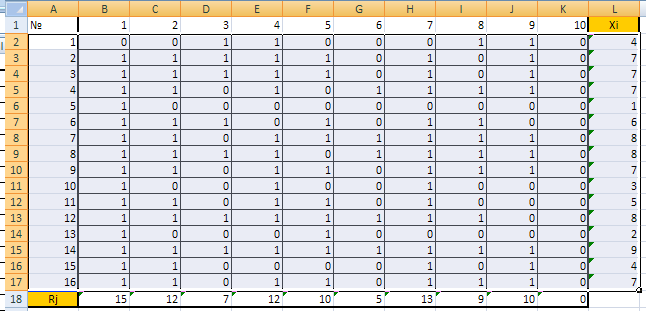
Rj – количество верных ответов на j-тое задание.

Считается так же – сумма, только по столбикам. Опять если пользуетесь авто-суммой не забывайте исключить номер из диапазона.

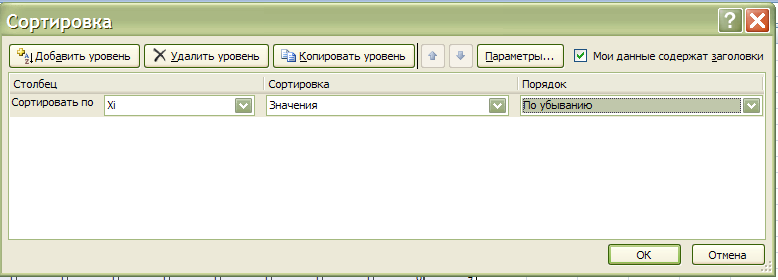
Когда значения Rj и Xi посчитаны, можно проверить себя, сумма всех Xi и Rj должна быть равной, в нашем случае 93.

1. Упорядочить бинарную матрицу

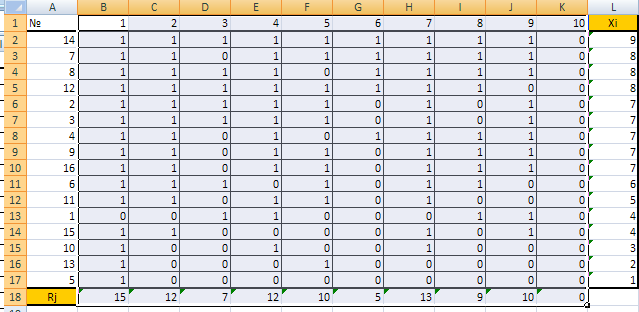
Сначала упорядочим по Xi. Метим матрицу таким образом: верхнюю строчку не захватываем и нижнюю с Rj тоже.



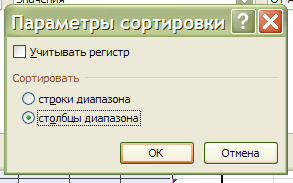
Затем открываем вкладку **Данные** (Панель инструментов) – **Сортировка** и выбираем в появившемся окне следующие варианты: Xi, Значения, По убыванию. Жмем **Ок**.



Теперь отсортируем по Rj. Метим следующим образом:

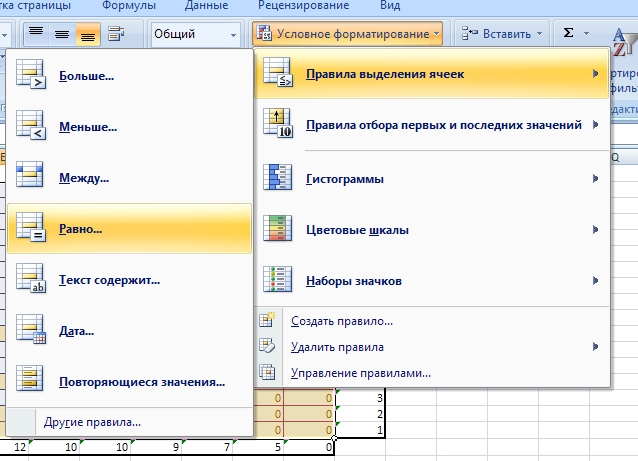


Опять вкладка **Данные** – **Сортировка**, но теперь внимание – нам надо сортировать по столбцам, а не по строкам. В открывшемся меню Сортировка открываем вкладку **Параметры** и там выбираем **Столбцы Диапазона** (до этого по умолчанию стоит Строки диапазона, поэтому мы не открывали эту вкладку когда сортировали по Xi.



Затем выбираем такие варианты: Строка 18 (или та строка, где у вас располагаются значения Rj), Значения, По убыванию.

После первой и второй сортировок наши номера будут идти вовсе не по порядку, так и должно быть. Чтобы себя проверить – выделим в матрице нули, они должны располагаться преимущественно в правом нижнем углу матрицы. Чтобы выделить нули выделим нашу матрицу и вкладка **Главная** – затем **Условное Форматирование – Правила выделения ячеек – Равно.**

****

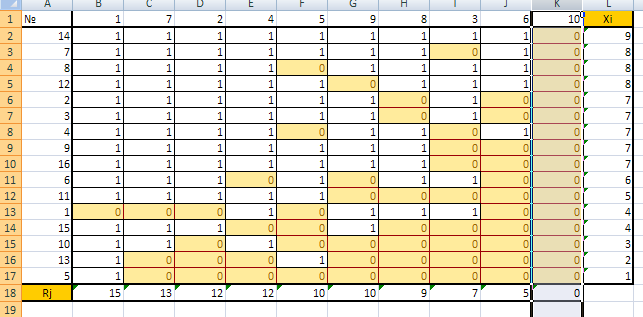
В появившемся окне выбираем «0» и способ заливки.

Бинарная матрица имеет характерную особенность - почти все нули и единицы распределились относительно диагонали, идущей из левого нижнего угла в правый

верхний. Это стало видно, после заливки нулей. Это значит, что мы все правильно отсортировали.

1. При необходимости редуцировать матрицу

Т.е если в матрице есть столбец или строка где все значения либо нули либо единицы – то нам их нужно удалить. Только внимание такие столбцы или строки или и строки мы можем удалять, если они есть изначально. Если же допустим, как в нашем случае, мы удалили нулевой столбец под номер 10, а затем у нас получились все единицы в строке номер 1, то удалять мы уже её не можем.

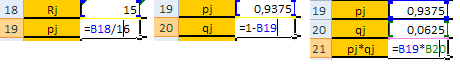


Удаляем столбец номер 10. Щелкаем по названию столбца по букве К правой кнопкой и выбираем удалить. Чтобы не палить всю матрицу и не высматривать нулевые столбцы – можно просто посмотреть есть ли нулевые значения по Rj или Xi. Для единиц соответственно не нулевые, а максимальные значения. После таких операций наша матрица называется редуцированной.

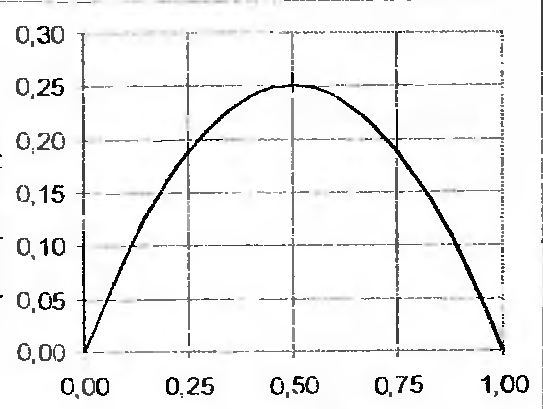
1. Рассчитать , ,

– доля правильных ответов, рассчитывается как отношение верных ответов к максимальному количеству заданий. То есть формула для будет такая:

**= /N**. Где N – общее число испытуемых после редуцирования (т.е сколько всего у нас осталось строчек после редуцирования). По сути это вероятность, но уже статистическая. Опять протягиваем и получаем строчку со значениями .

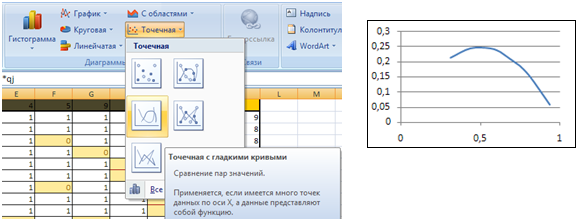
  
 считается как доля неправильных ответов, то есть из теории вероятности 1-. Протягиваем и получаем строчку с соответствующими значениями. Затем рассчитываем произведение \*.

1. Построить распределение Гутмана

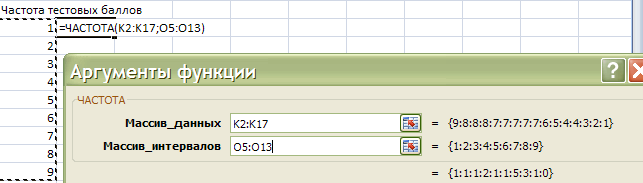
Это график, который должен выглядеть подобным образом. Строится на значениях и \*.

Выделяем строку - щелкаем левее самой ячейке , т.е. по номеру строки (19 в нашем случае) – тогда выделится вся строка. Затем удерживая клавишу **Ctrl**, таким же образом выделяем строку \*. Затем **Вставка – Точечная –** на отделе **Диаграммы.**

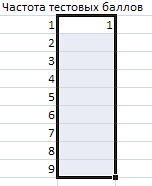
Наш график получился каким-то недорисованным, но это не беда, просто у нас минимальные значения не достигают околонулевых значений.

****

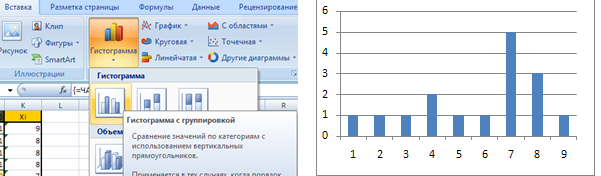
1. Представить графически тестовые баллы

Снова будем строить графики. Для этого нам необходимо рассчитать частоту тестовых баллов. Необходимо построить табличку, где будут значения баллов и напротив количество повторений или вхождений в Xi. Можно это сделать в принципе в ручную, или воспользоваться **функцией Частота**, которая находится в разделе **Статистические**. Выписываем в таблицу значения баллов, т.е., от 1 до 9 – в нашем случае. Лучше именно в порядке возрастания. Затем в соседней ячейке пишем формулу **Частота**, выбираем массив данных – это наш столбик Xi, естественно без ячейки Xi, а массив интервалов – только что созданный столбик из значений баллов. Жмем Ок. 

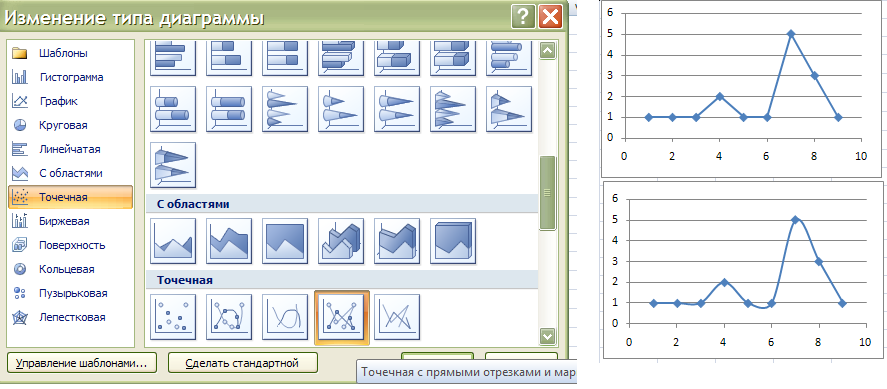
Пока у нас только одно значение. Делаем следующее. Выделяем вот такой диапазон

Жмем F2 и видим формулу, а затем клавиши CTRL+SHIFT+ВВОД. И наши частоты для баллов посчитаны. Можно проверить вручную. Или посчитать вручную вообще не пользуясь формулой.

Строим графики. Сначала построим гистограмму. Выделяем в нашей таблице столбец с частотой. Как показано на рисунке (только там уже будут значения), затем **Вставка** – **Гистограмма**. Выберем первый вариант.



Важно: если расположить баллы в порядке убывания, то график может быть неправильным. А так же если у вас возможные балы не от 1 до 9, а предположим, 3,4,5,6 проделываем тоже, просто потом изменяем подпись оси икс. Когда вы построите гистограмму, у вас будет нумерация с 1. Для этого щелкаем по графику правой кнопкой и выбираем в меню: **Выбрать данные – в категории Подписи горизонтальной оси – Изменить -** и выбираем диапазон уже с нашими балами (который расположен в нашей табличке слева от соответствующих частот)  
Затем можно построить полигон частот. Просто дублируем нашу гистограмму, **Копировать – Вставить,** а затем правой кнопкой мыши по графику – **Изменить тип.** Выбираем нужный нам график, построим полигон и эмпирическую кривую, на рисунке показан выбор полигона, эмпирическая кривая находится левее в том же разделе Точечная.



1. Вычислить меры центральной тенденции

Их три – мода, медиана, и среднее арифметическое. Считаем для массива Xi.

Среднеарифметическое значение считается по одноименной формуле.

Мода – часто встречающееся значение в выборке. Считается по одноименно формуле, в разделе статистические.

Соглашения об использовании моды:

1) Если все значения в группе встречаются одинаково часто, то мода отсутствует.   
Например, в группе (1, 1, 2, 2, 13, 13) моды нет.

2) Когда два соседних значения имеют одинаковые частоты и они больше частоты любого другого значения, мода есть среднее этих двух значений.   
Например, в группе (1, 2, 2, 5, 5, 5, 6, 6, 6, 9, 9, 10) мода равна 5,5.

3) Если два несмежных значения в группе  имеют равные частоты и они больше частот любого другого значения, то существуют две моды. В этом случае говорят, что группа оценок является *бимодальной*.   
Например, в группе (1,4,4,4,7,7,9,9,9,10) модами являются 4 и 9. На рис.3.3.3 показано бимодальное распределение с модами 1,5  и 6.

Наибольшей модой в группе называется единственное значение, удовлетворяющее определению моды. Однако во всей группе может быть несколько меньших мод.  Эти моды представляют собой локальные максимумы распределения частот.

Медиана – это значение, которое делит упорядоченное множество данных пополам. Так же по формуле. Выбираем нужную формулу и захватываем диапазон со значениями Xi.

1. Проверить распределение на нормальность

Если среднее арифметическое примерно равно утроенному стандартному отклонению,

то можно считать дисперсию оптимальной, а распределение тестовых баллов близким к нормальному. Дисперсия считается по формуле «ДИСП» выделяем диапазон со значениями Xi, а сигма, т.е. стандартное отклонение – корень из дисперсии.

Наше среднее арифметическое = 5,8 отличается от утроенного сигма = 7,1. Значит в нашем случае распределение по такому критерию ненормальное. Отметим, что это утверждение справедливо не для всех случаев. Так же можно посмотреть на значения моды, медина, и ср. ариф. Если они равные, то распределение близко к нормальному.

В качестве грубой оценки нормальности распределения можно рекомендовать проверку следующего соотношения: ~~X~~ - 3σ ≤ X ≤ ~~X~~ + 3σ, где ~~X~~ – ср. арифметическое. Если почти все значения тестовых баллов *X* укладываются в этот интервал, то в первом приближении можно считать эмпирическое распределение нормальным. Для нашего случая по этому критерию распределение оказывается нормальным. (-1,3≤ X ≤12,9). То есть у нас получается неоднозначная картина. Тогда скажем, что наше распределение не дотягивает до нормального.

1. Построить эмпирическую кривую

Мы проделали это задание, когда строили графики.

1. Проверить гипотезу о нормальном распределении по критерию Пирсона

Этот пункт мы пропускаем.