Лабораторная работа 1

Получена бинарная матрица результатов тестирования.

МАТРИЦА ТЕСТОВЫХ РЕЗУЛЬТАТОВ a_{ij} – это матрица размерности NxM, содержащая числовые обозначения градации индикатора, связанного с изучаемой латентной переменной, где M-число индикаторов, N-число испытуемых.

1. Рассчитать X_i, R_j

X_i - индивидуального балла i-го испытуемого

$$X_i = \sum_{j=1}^{M} a_{ij}$$

количество верных ответов R_j, на j-е задание

$$R_j = \sum_{i=1}^{N} a_{ij}$$

2. Упорядочить бинарную матрицу

Для удобства визуального анализа в ячейках с нулевыми значениями необходимо сделать заливку цветом. (Главная-Условное форматирование-Правила выделения ячеек-Меньше).

Далее, необходимо упорядочить бинарную матрицу. Сначала выполним упорядочение по величине X_i. Для этого выделяем все данные, кроме первой и последней строки. Открываем вкладку Данные – Сортировка. Выбираем сортировать по X_i по убыванию. Аналогично выполняется упорядочение по значению R_j, но теперь сортируем по столбцам, в параметрах сортировки необходимо изменить «строки диапазона» на «столбцы диапазона». Выбираем сортировать по строке (в которой содержится R_j) по убыванию.

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

3. При необходимости редуцировать матрицу

Если в матрице есть задания, которые успешно выполнили все испытуемые или задания, с которыми не справился ни один испытуемый, то эти столбцы необходимо удалить, т.к. они не позволяют дифференцировать испытуемых.

Кроме того, если есть испытуемые, которые успешно выполнили все задания, то эти строки тоже необходимо удалить, т.к. они не дают никакой информации, кроме того, что для испытуемого все задания слишком легкие.

4. Рассчитать $w_j, p_j, q_j, \ p_j \cdot q_j$.

количество неверных ответов W_j на j-е задание

$$W_j = N - R_j$$

доля верных ответов p_j и доля неверных ответов q_j

$$p_j = \frac{R_j}{N}$$

5. Построить распределение Гутмана.

Согласно Гуттману это разграничение должно быть идеальным. Если испытуемый, верно ответил на трудное задание, то он тем более, должен справиться с более легкими заданиями. Это должно приводить к строгому разграничению единиц и нулей диагональю матрицы.

В действительности же это не совсем так. Например, если испытуемый справился с самым трудным заданием, но не справился с более легкими заданиями. Профиль испытуемого искажен. Если бы испытуемый, верно ответил на трудные вопросы, но не справился с легкими, то говорят, у него инвертированный профиль.

Инвертированный профиль свидетельствует либо о неверной структуре знаний испытуемого, либо о нарушении процедуры тестирования (списывание, угадывание и т.д.), либо о недостатках тестовых заданий (по форме и (или) по содержанию).

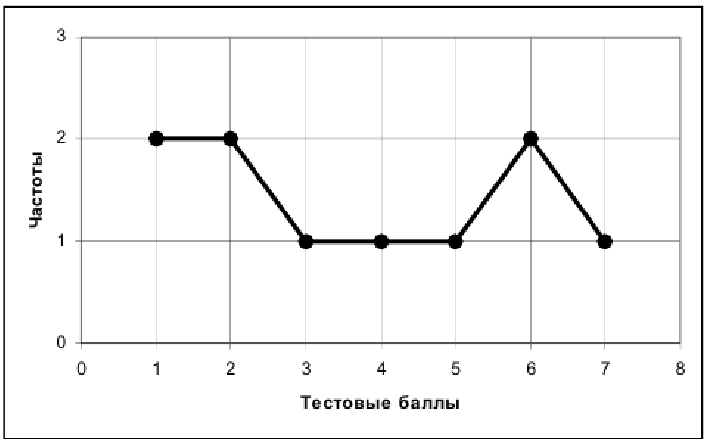
Распределение строится по строкам p_j и $p_j * q_j$ выбираем точечную диаграмму.

6. Представить графически тестовые баллы

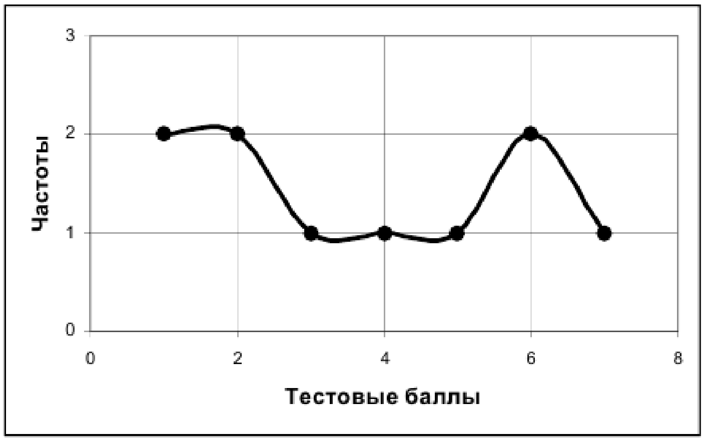
Полученные данные удобно анализировать, используя их графическое представление. Для этого надо найти частоты тестовых баллов.

ЧАСТОТА ТЕСТОВОГО БАЛЛА – это количество испытуемых, имеющих данный тестовый балл. Для расчета добавляем таблицу, состоящую из двух столбцов. В первом указываем все возможные значения индивидуальных баллов испытуемых. Во втором рассчитываем частоты индивидуальных баллов испытуемых (можно воспользоваться функцией excel ЧАСТОТА).

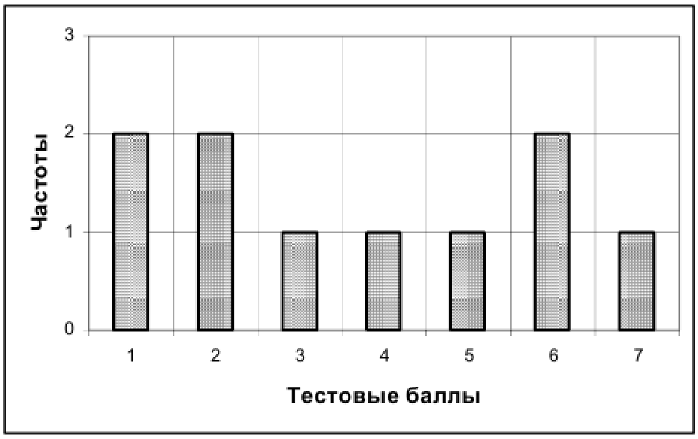
На основании таблицы частот можно построить полигон частот. По оси абсцисс отложены тестовые баллы, а по оси ординат – соответствующие частоты.



Вместо полигона частот иногда используют сглаженную кривую, проходящую максимально близко к экспертным точкам.



Для графического представления данных можно также использовать гистограммы.



Желательно, чтобы распределение частот было близко к нормальному (Гауссовому).

7. Вычислить меры центральной тенденции.

Тестовые баллы испытуемых обычно группируются вблизи некоторых, наиболее вероятных значений, которые можно охарактеризовать тремя мерами центральной тенденции – модой, медианой и средним.

МОДА- это такое значение в множестве наблюдений, которое встречается наиболее часто.

МЕДИАНА - это значение, которое делит упорядоченное множество данных пополам, так что одна половина значений оказывается больше медианы, а другая - меньше.

СРЕДНЕЕ АРИФМЕТИЧЕСКОЕ определяется по формуле

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

где N - количество элементов в группе, Xi - величина i-го элемента группы.

Какую из мер центральной тенденции выбрать - решать исследователю.
В педагогике очень часто в качестве меры центральной тенденции выбирается среднее арифметическое.

8. **Проверить распределение на нормальность (три способа)**

Рассчитываем дисперсию, используя статистическую функцию ДИСП, для которой надо указать диапазон ячеек со значениями индивидуальных баллов испытуемых.

Величина дисперсии тестовых баллов позволяет судить о качестве теста, о его дифференцирующей способности. Малая величина дисперсии говорит о том, что тест плохо различает испытуемых по уровню знаний, не позволяет с приемлемой точностью ранжировать их. Слишком большая дисперсия указывает на сильную неоднородность группы испытуемых, на возможные нарушения процедуры тестирования, на недостаточно ясные формулировки заданий и т.п. В случае оптимальной величины дисперсии, распределение тестовых баллов близко к нормальному.

С дисперсией связан еще один важный параметр - стандартное отклонение

$$s_x = \sqrt{s_x^{\,2}}$$

М.Б.Челышкова считает, что если среднее арифметическое примерно равно утроенному стандартному отклонению,

$$\overline{X} \approx 3s_x$$

то можно считать дисперсию оптимальной, а распределение тестовых баллов близким к нормальному.

Отметим, что это утверждение справедливо не для всех случаев. Возможны ситуации, когда среднее арифметическое гораздо больше утроенного стандартного отклонения, но распределение тестовых баллов, тем не менее, достаточно близко к нормальному.

В качестве грубой оценки нормальности распределения можно рекомендовать проверку следующего соотношения:

$$\overline{X} - 3s_X \le X \le \overline{X} + 3s_X$$

- если почти все значения тестовых баллов X укладываются в этот интервал, то в первом приближении можно считать эмпирическое распределение нормальным.

Для корректного решения вопроса о степени близости эмпирических данных нормальному распределению необходимо использовать более строгие доказательства, например, проверить гипотезу о нормальном распределении генеральной совокупности по критерию Пирсона.