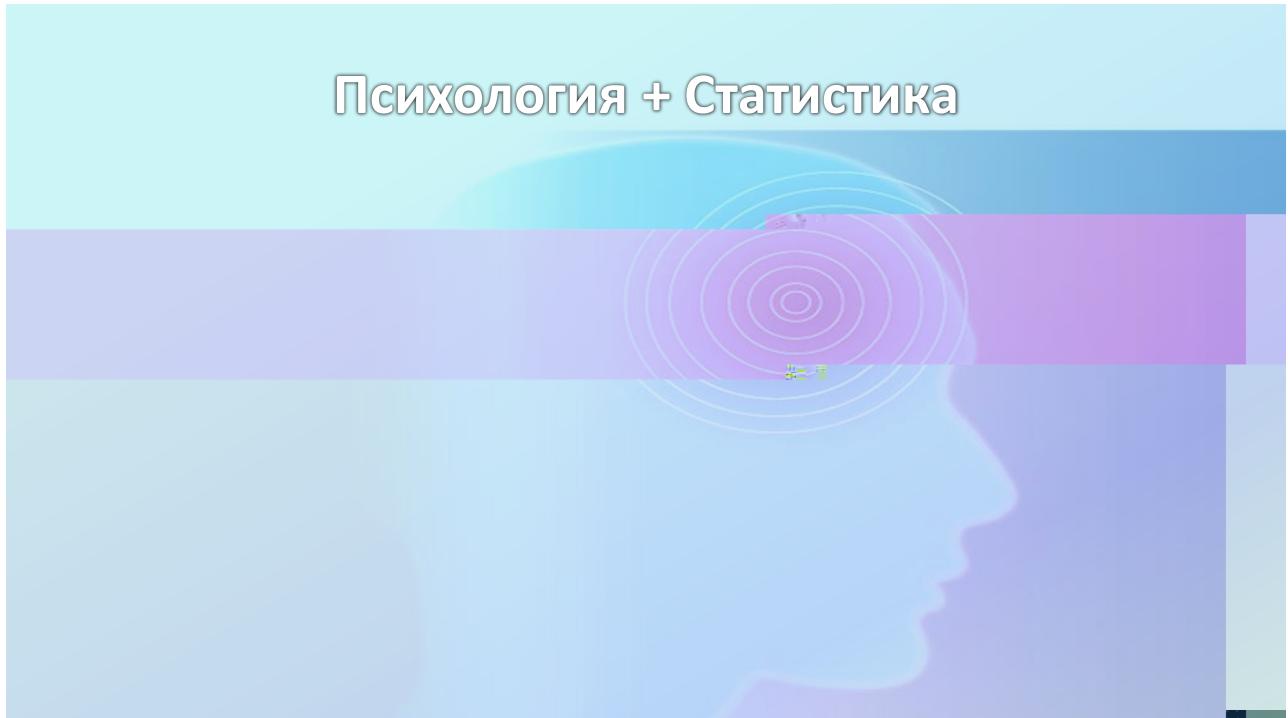




Психология + Статистика



Цель психометрики:

Достоверность
Надежность
Честность
Безопасность



Item Response theory (IRT) - Теория тестовых заданий

Современная теория тестирования

Набор методов, позволяющий оценить вероятность правильного ответа испытуемых на задания различной трудности

Используется для того, чтобы избавиться от плохих (неинформационных) вопросов в опросе (тестировании)

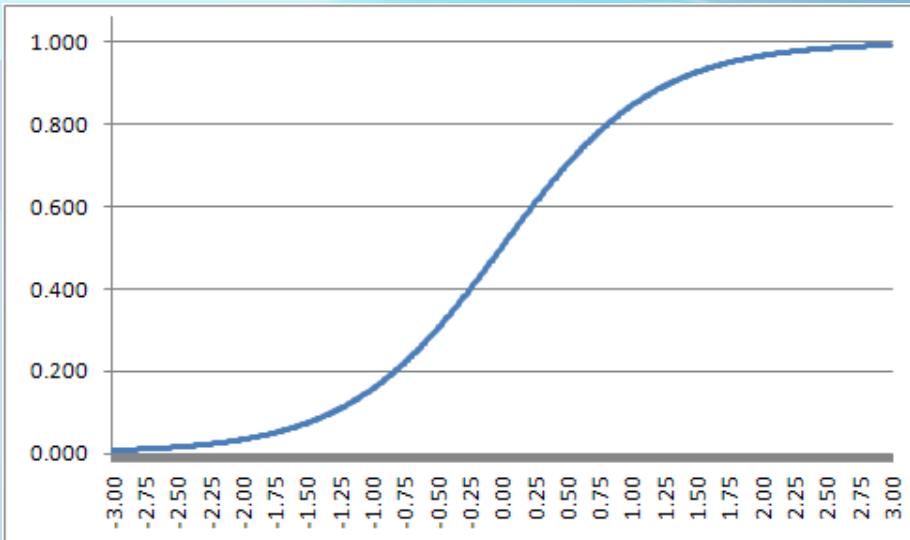
Основана на взаимосвязи между качеством ответов на вопросы и общим качеством знания, которое проверяется в teste

Не предполагает, что каждый вопрос одинаково трудный

Item Response theory (IRT)



Модель Раша



Анализ заданий и интерпретация результатов

Анализ трудности задания

Анализ дискриминационной (различительной) эффективности задания

Анализ вариантов ответа на вопрос задания

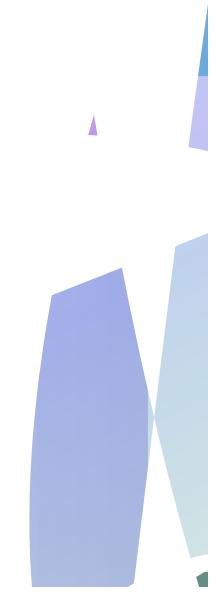
Сравнительный анализ групп экзаменуемых

Анализ трудности задания



Анализ для генерализации (вспомогательной)

Формулы для задачи



Анализ вариантов ответа на вопрос задания

Необходимо оценивать эффективность вариантов ответов

Наличие вариантов ответа, которые никто не выбрал:

Неправдоподобен

Имеет структурный недостаток

Частота ответа на вариант ответ больше, чем у правильного ответа

Возможное наличие двух правильных ответов

Возможно правильный ответ назначен ошибочно

Группа	A	B	C+	D	E
1 группа	6	0	89	1	4
2 группа	15	0	61	1	5
Всего	16	0	75	1	8
Значение Р: 75			Индекс дискриминации: +0,33		

Группа	A	B	C	D	E+
1 группа	3	1	43	2	51
2 группа	17	21	22	20	20
Всего	8	14	31	12	35
Значение Р: 35			Индекс дискриминации: +0,30		

Группа	A+	B	C	D	E
1 группа	50	6	8	16	20
2 группа	22	14	19	21	24
Всего	35	11	14	18	22

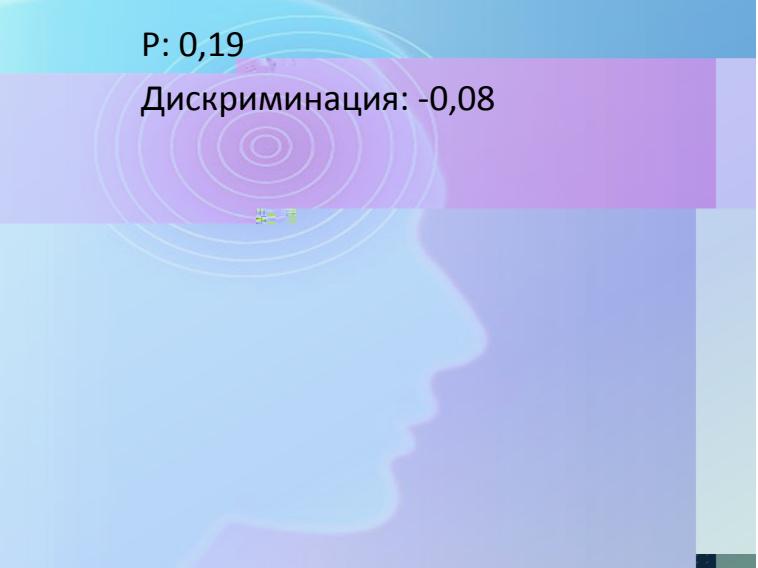
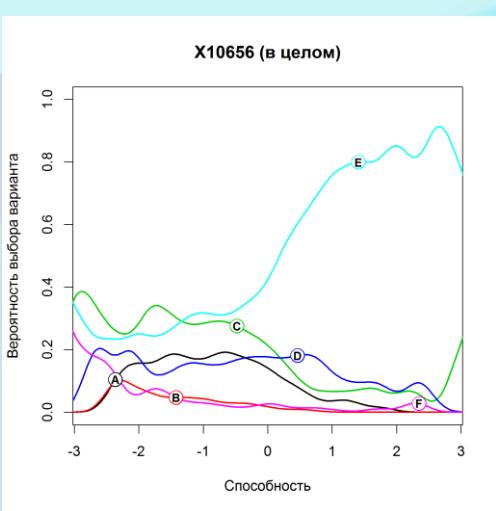
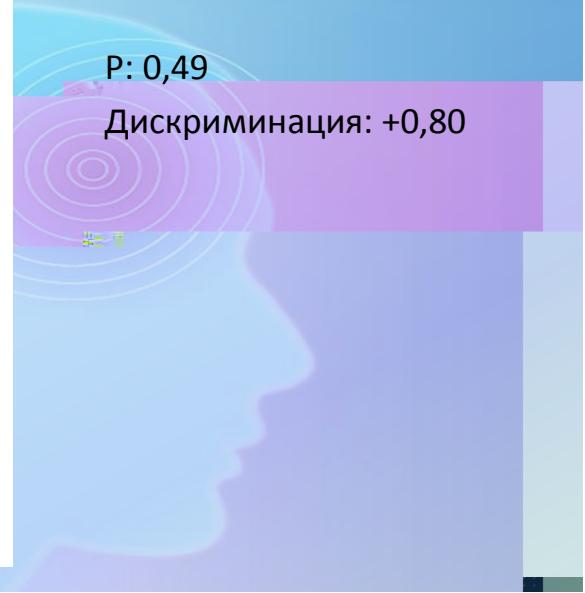
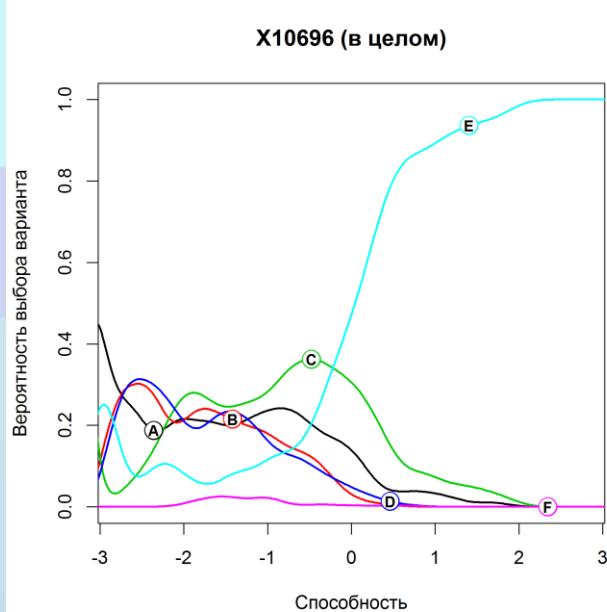
Значение Р: 35

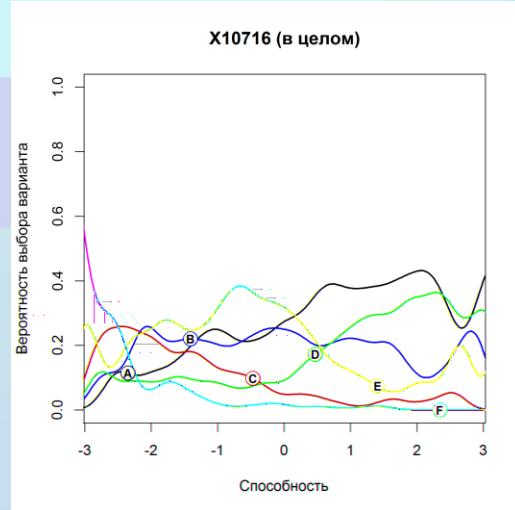
Индекс дискриминации: +0,30

Группа	A	B+	C	D	E
1 группа	10	39	5	44	2
2 группа	23	25	12	37	3
Всего	17	30	7	44	2

Значение Р: 30

Индекс дискриминации: -0,09





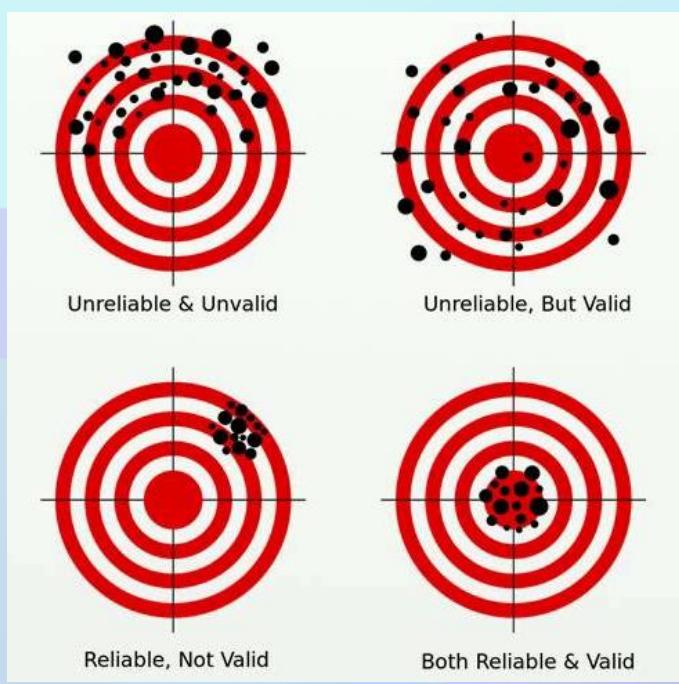
P: 0,21

Дискриминация: 0,01

Психометрический анализ

Альфа Кронбаха/ Кудер Ричардсон – оценка надежности теста





Differential Item Functioning (DIF)

Сравнение успешного ответа на вопрос между двумя определёнными группами (например между мужчинами и женщинами)

Сложность задания должна быть только по отношению оцениваемым компетенциям претендента

Сложность задания не должна быть связана с факторами, которые не имеют отношения к оцениваемым компетенциям претендента

Сложность задания должна быть единая для всех

Претест

15 – 25 претестовых заданий в первых 60 вопросах настоящего экзамена

Претестовые задания разбросаны среди 60 вопросов экзамена (максимум 2 претестовых вопроса подряд)

Не влияют на результат претендента

Должны быть не отличимыми от настоящих вопросов для претендента

Апробация

Проверка экзамена в реальных условиях

Условия должны полностью повторять настоящий экзамен:

Время

Аудитория

Длительность экзамена

Формат экзамена

База тестовых заданий

Психометрический анализ

Установка порогового балла (Standard setting)



Набор панели экспертов

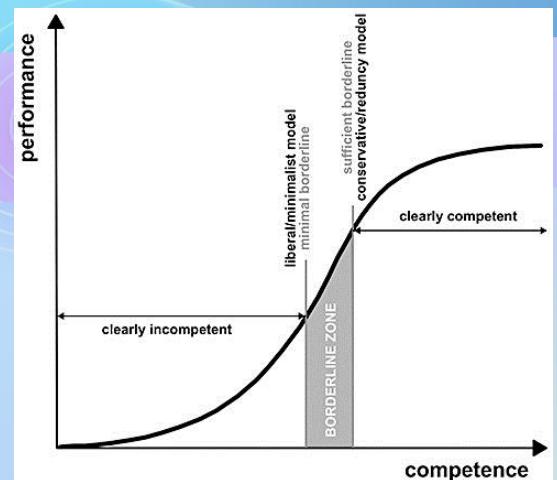
Кто должен быть в панели:

1. Медицинская сестра со стажем работы больше 3х лет
2. Главная (старшая) медсестра
3. Представитель обучающей организации, имеющий непосредственный контакт с медицинскими сестрами (КГМИПиПК)
4. Представитель ассоциации медицинских сестер
5. Представитель СД МЗ КР

Определение кандидата с минимумом обязательных компетенций

При установки порогового балла упор идет на кандидатах, которые имеют «достаточный» уровень знаний и умений для работы в практике эффективно и безопасно для общества.

Кандидаты, которые не являются «суперзвездами» и не самыми худшими



Обучение панели экспертов

Метод Angoff:

Эксперты должны определить, каков шанс того, что кандидат с минимумом обязательных компетенций (МОК) ответит на задание правильно

Другими словами, из 100 кандидатов с МОК ответят правильно на данное задание

Эксперты дают рейтинг от 0 до 100

Меньше рейтинг = тяжелее вопрос

Метод Angoff

№ Задания	Оценка Экспертов (%)						Ср. оценка
	№1	№2	№3	№4	№5	№6	
1	65	70	65	65	70	65	67
2	85	70	60	80	70	80	74
3	70	65	85 -> 75	70	45 -> 50	70	67
4	75	60	65	70	75	70	69
Общая средняя оценка							69