Which IRT model?

Start from what one item score looks like, answer a question or two, and end at a model and the line of R code that fits it. When in doubt, let `select_model()` compare the candidates for you.

Every path ends in the same call

```
library(mirtverse)
mod <- mirt(data, 1, itemtype = "2PL")
interpret(mod) explains the result
```

What does one item's score look like?

Right or wrong (0/1)

Ability tests, quizzes, licensure exams

Could low scorers guess the right answer?
Multiple choice, few options

- Yes → **3PL**
  Adds a floor for guessing
  `itemtype = "3PL"`
  Needs about 1,000+ people
- No → **Should every item count the same in the score?**
  The sum score tells the story
  - Yes → **Rasch (1PL)**
    One slope shared by all items
    `itemtype = "Rasch"`
    Works from about 100 people
  - No → **2PL**
    Each item gets its own slope
    `itemtype = "2PL"`
    Aim for about 500+ people

Ordered categories (0, 1, 2, ...)

Rubric points, Likert ratings, partial credit

Are points earned step by step on a rubric?
Constructed responses, tasks

- Yes → **Partial credit (GPCM)**
  Steps, each item its own slope
  `itemtype = "gpcm"`
  Rasch version (PCM): "Rasch"
- No → **Same rating scale, and a Rasch model wanted?**
  All items share the labels
  - Yes → **Rating scale model**
    Thresholds shared by all items
    `itemtype = "rsm"`
    The most compact choice
  - No → **Graded response**
    The usual pick for Likert scales
    `itemtype = "graded"`
    Aim for about 500+ people

Unordered options

Which option was chosen

Nominal response
Models every option, so you see which distractors pull whom
`itemtype = "nominal"`
Large samples help

Agree only if close

Ideal-point attitude items

Unfolding (GGUM)
People on both extremes disagree with a middle statement
`itemtype = "ggum"`
Large samples help

Not sure? Let the data compare.

```
sel <- select_model(data)
```

Fits Rasch, 2PL and 3PL (or PCM, GPCM and graded for ordered items), ranks them by BIC, runs likelihood-ratio tests and M2 fit, and explains the pick in plain words.

`sel$best` *the recommended model, ready for interpret()*

Sample sizes on this page are rough rules of thumb, not guarantees.

Then think about the structure

Each question adds one argument or one function

More than one trait?
Check local dependence first, then fit two or more factors.
`gg_q3(mod)`
`mirt(data, 2)`
`mirt.model("F1 = 1-5 ...")`

Items in clusters?
Testlets, shared passages or scenarios: a general factor plus specific ones.
`bfactor(data, specific)`

Groups to compare?
Check items for DIF, then link the groups on one scale.
`dif(mod, group)`
`multipleGroup(data, 1, group)`

Item features matter?
Explain difficulty with item properties (LLTM, explanatory IRT).
`predict_difficulty()`
`eirm::eirm()`
`mixedmirt()`

Response times too?
Model speed with ability and flag unusual timing.
`fit_rt(mod, rt)`

Classes, not a continuum?
Mastery groups or latent profiles instead of a trait.
`mdirt(data, 2)`

Before you trust the model

`glance(mod)` overall fit (M2, RMSEA, CFI)
`gg_itemfit(mod)` items that misfit
`gg_trace(mod)` curves that look strange
`interpret(mod)` a plain-language summary

A simpler model that fits is easier to explain. Move to a richer one only when the fit checks or the purpose of the test call for it.

Online version with examples
okanbulut.com/mirtverse/articles/which-model.html