The therapist’s dilemma

The problem of using nomothetic data for the individual case

Bob Radstaak & Nora Loretan

February 6 , 2020

Allport (1961) recognized eloquently that "the dilemma of uniqueness haunts the house of clinical psychology" (p. 21)

Yet, the therapist's dilemma applies to most situations regarding individual assessment, e.g.

- Education and school
  - Testing school admission
  - Achievement testing (e.g. CITO)
    - Annual spending roughly $ 1,7 billion in de US (Chingos, 2012)
- Clinical and counseling
  - Diagnostic assessment
  - Therapeutic change
  - Forensic
  - Neuropsychological
  - Child custody evaluation
- Industrial and organizational psychology, e.g.
  - Job performance
  - Personnel selection
  - Personality measurements

Nomothetic:
The word has its roots in *nomos,* meaning "law," and -*thetēs,* meaning "one who establishes." Merriam-Webster. (n.d.).

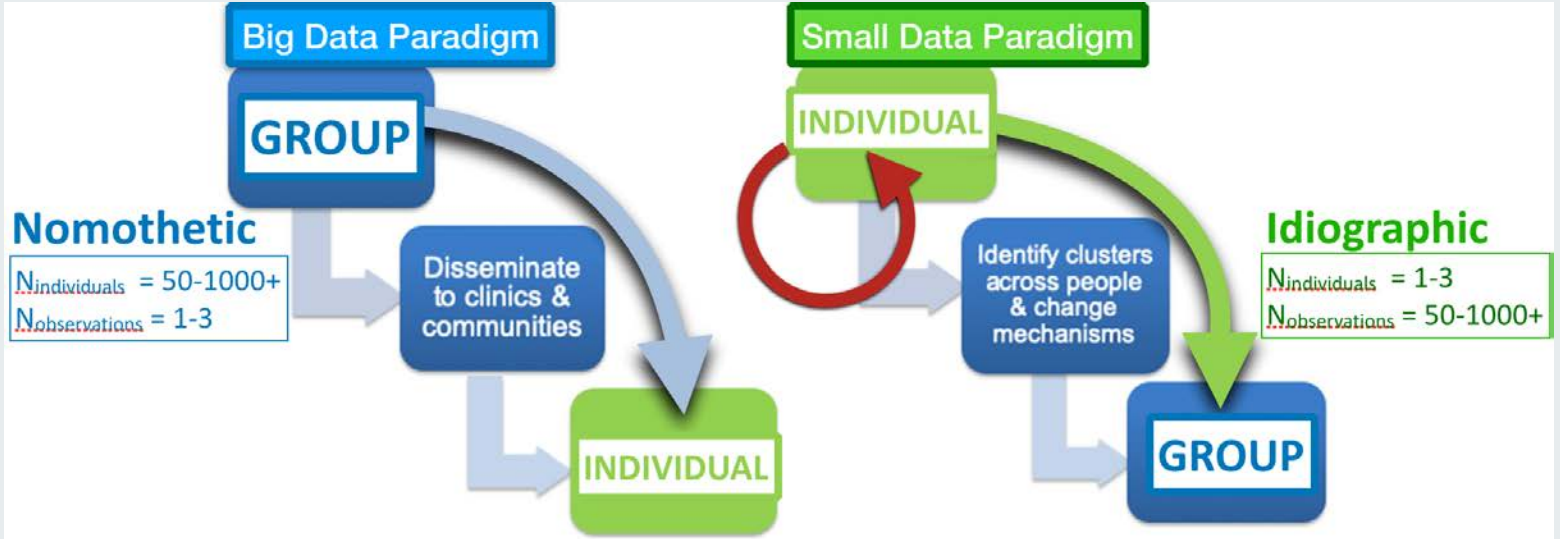


Fig. 1 Small versus big data paradigm pathways to help individuals and transportable knowledge

Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., & Sim, I. (2019). Why we need a small data paradigm. *BMC medicine*, *17*(1), 1-9.

Idiographic:
"The root "idio" refers to "personal" or "distinct." Haynes et al., 2000, p. 110)

"**Nomothetic** research strategies focus on identifying covariations between variables using data from groups of persons. They are used to develop models of behavior problems that are generalizable to the "*average*" person."
(Haynes et al. 2000, p. 110)

"Inferences made in social and medical research typically result from statistical tests conducted on aggregated data. The implicit assumption is that group-derived estimates can be applied to understanding individual phenomenology, physiology, and behavior."
(Fisher et al. 2018, p. E6106)

"**Idiographic** research strategies emphasize the measurement and analysis of variables for a single person. Covariation estimates are used to develop models that may be valid only for that person. Inferences from idiographic research may be, but are not necessarily, generalizable across persons." (Haynes et al. 2000, p. 110)

Difference
"Idiographic assessment is the measurement of variables and functional relations that have been individually selected, or derived from assessment stimuli or contexts that have been individually tailored, to maximize their relevance for a particular client. Idiographic assessment contrasts with nomothetic assessment in which judgments about a client are based on a comparison of the client with other persons using data from the same assessment instrument administered in a standardized manner." (Haynes et al., 2018, p. 191)

How is the reliability of a test score determined?

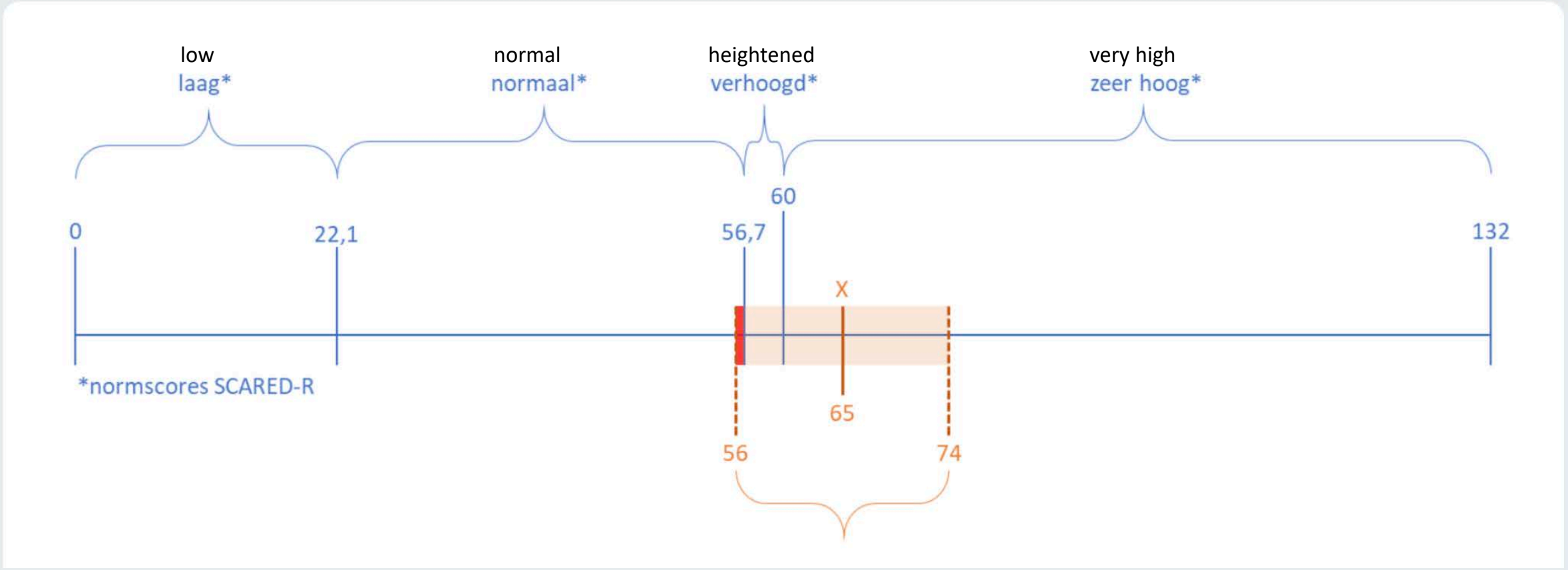
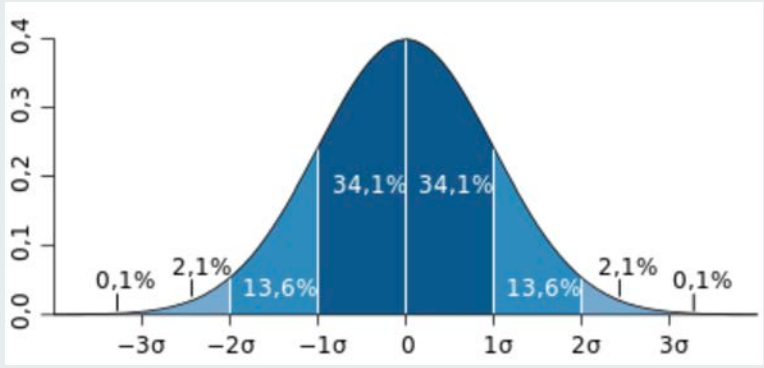
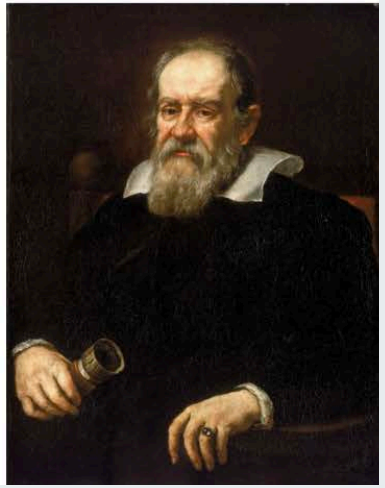


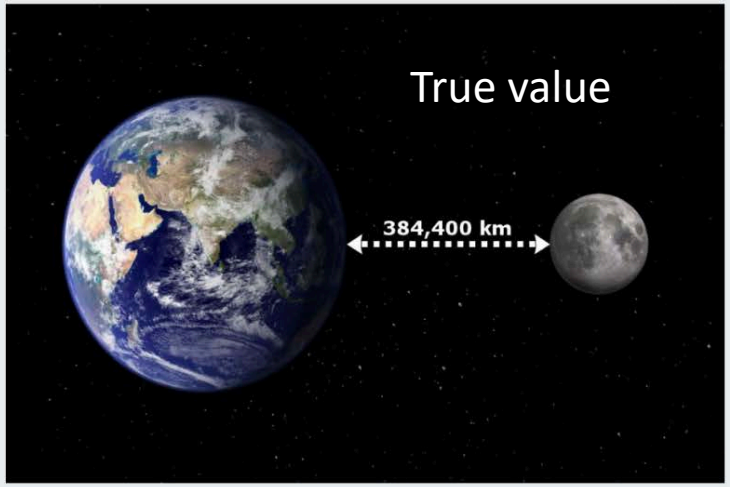
Figure 1: Standardized confidence interval

This confidence interval is based on the well-known

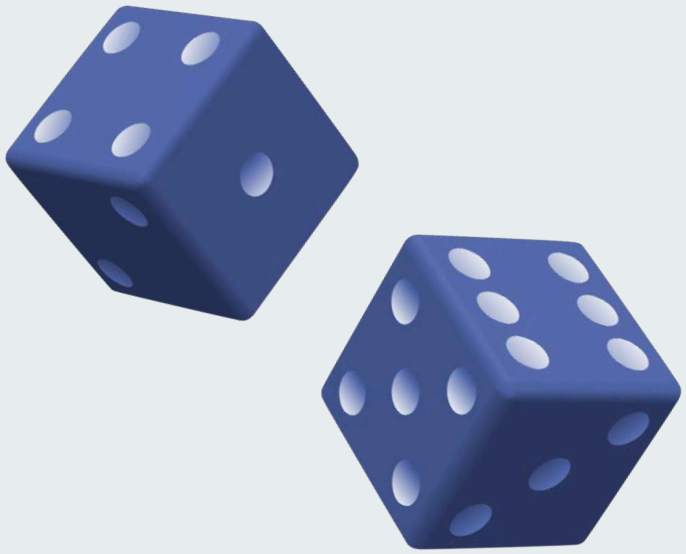




Galileo (1564 - 1642) was the first to formulate an rudimentary theory of errors



1. There is only one number which gives the distance of the star from the center of the earth, the *true distance*.
2. All observations are *encumbered with errors*, due to the observer, the instruments, and the other observational conditions.
3. The observations are *distributed symmetrically* about the *true value*; that is, the errors are distributed symmetrically about zero.
4. *Small* errors occur *more frequently* than *large* errors.



(based on e.g. Gauss, Simpson, Lambert, Laplace, Pascal, Bernoulli, Huygens and Leibniz)

Fare game objects

Probability of sum = 2
<
Probability of sum = 7

A priori = beforehand

Calculation of the probability of a certain outcome (based upon the physical characteristics of the game object)
= **deduction**

The frequentist interpretation of probability
(± 19e eeuw)

The frequentist interpretation of probability is based on the law of large numbers

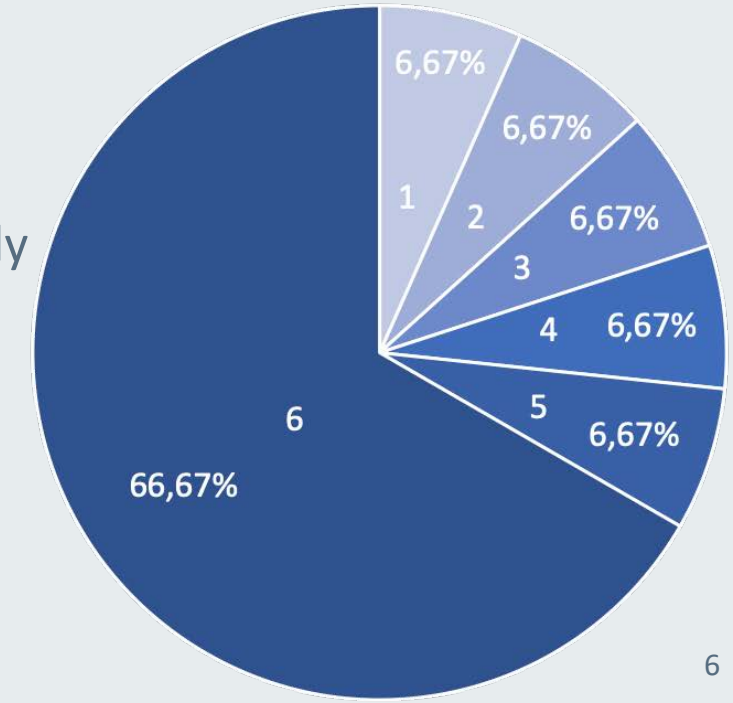
- **infinitely** many repetitions
- **identical** conditions
- repetitions are **mutually independent**
- relative frequencies and the weighted average have a **limit**

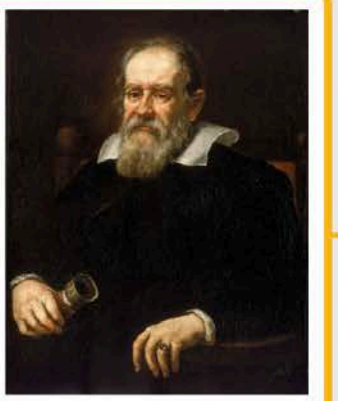
Unfair die with **unknown** physical properties

15 millionfold throw under **equal conditions**

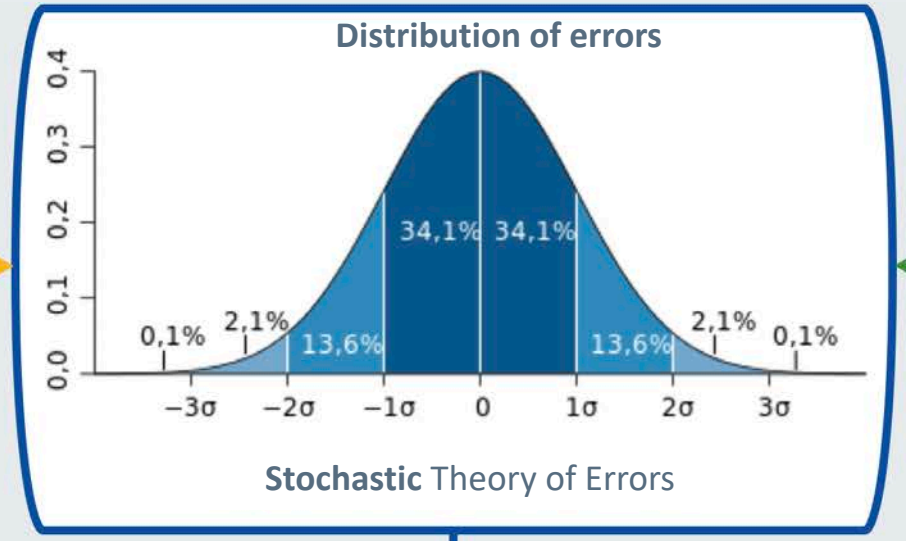
The outcome of 6 pips appears to occur 10 times more frequently than 1 - 5 pips.

A posteriori = afterwards
Calculation of the probability of a certain outcome **= induction**





Theory of errors



Stochastic Theory of Errors

Frequentist interpretation of probability

infinitely repeated measurements yield a normal distribution

measurement errors cancel each other out

expected value = weighted average

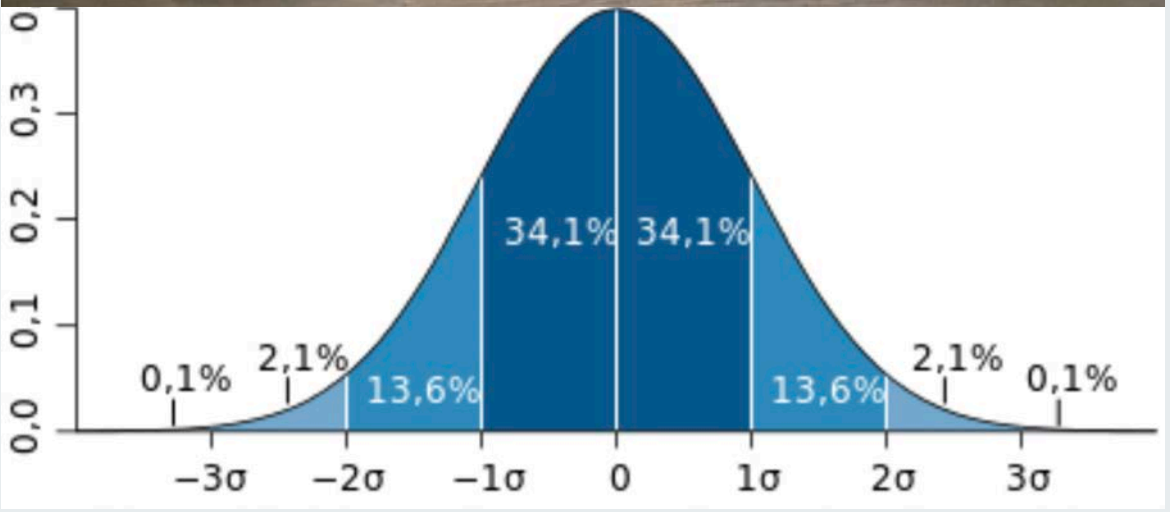
Theory of errors provided insight into the distribution of measurement errors

Probability theory added theoretical insights regarding chance

The **standard deviation** (SD) made it possible to quantify the **uncertainty/error** of a **measurement/observation**.
The larger the SD, the more **unreliable** the measurement procedure is (see Garfield's scale on the left)

Henceforth, one was able to determine the **reliability** of a **observation/measurement(procedure)**.

The **stochastic theory of errors** makes it possible to link observations to a probability distribution



Distribution of measurement errors

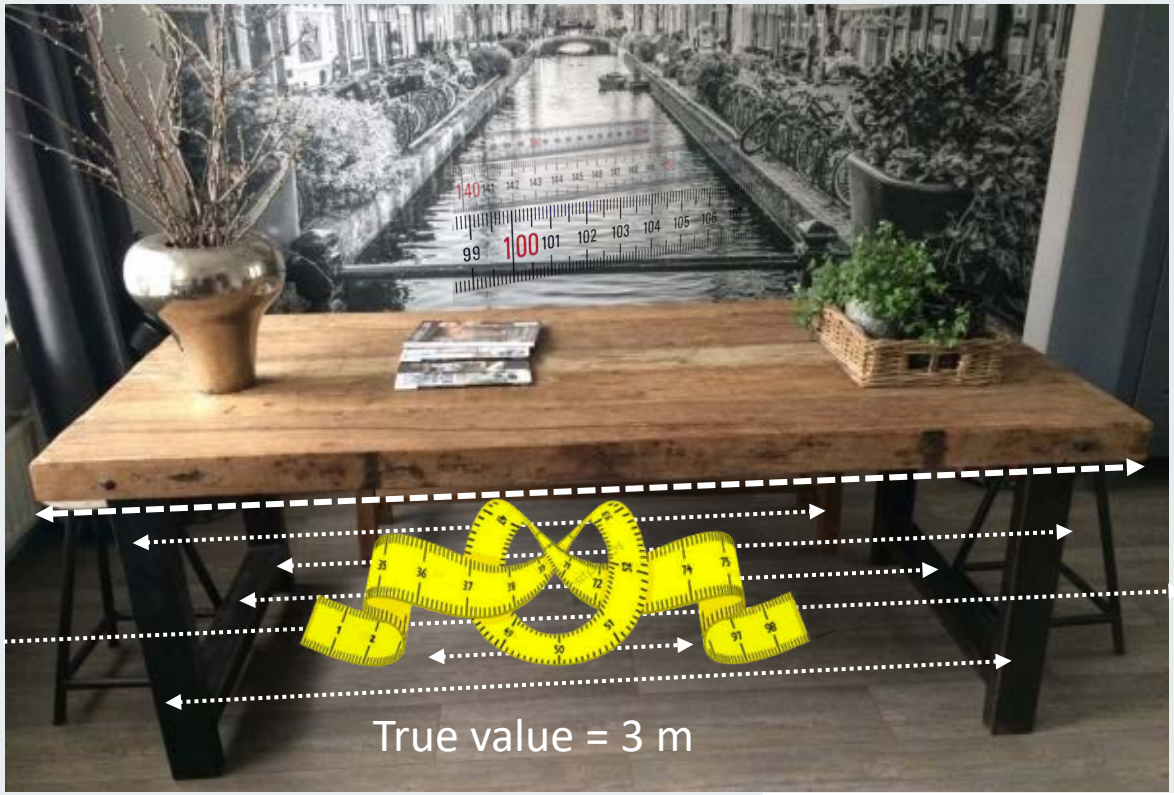
Assumptions:

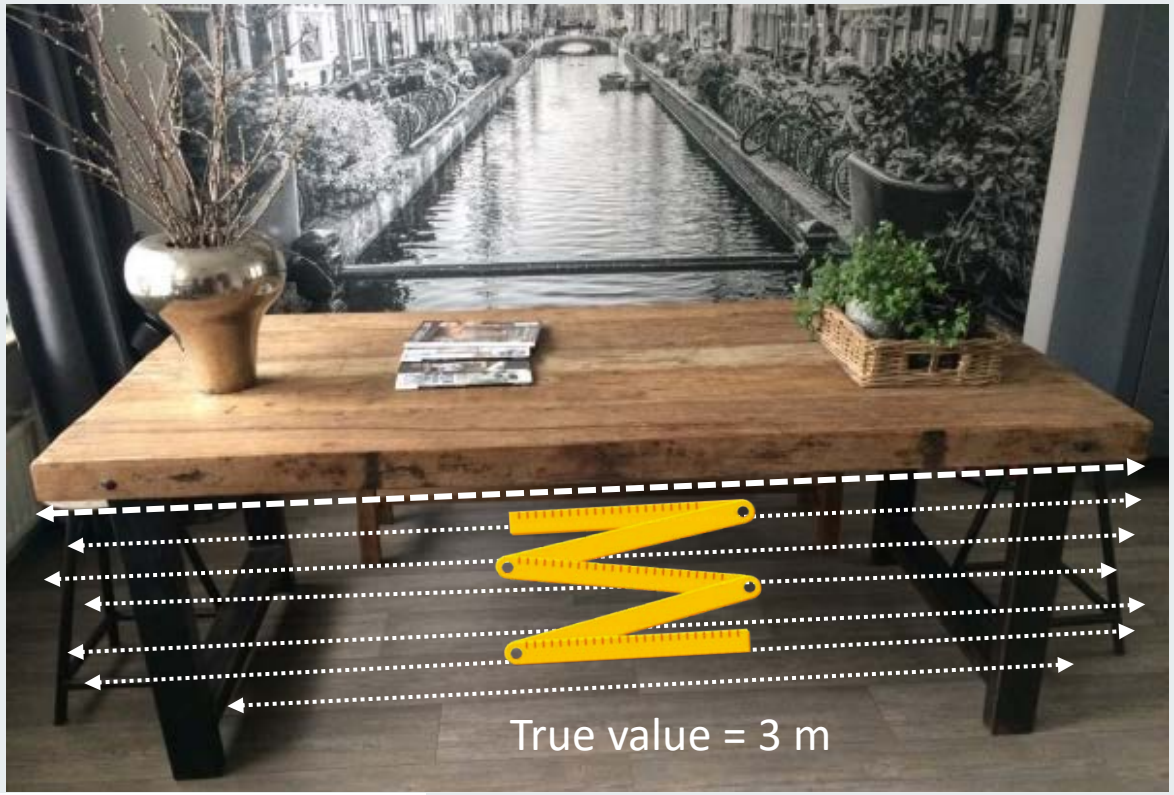
- The observations need to be mutually independent and repeated ad infinitum
- Under identical conditions
- The measurable phenomenon has to be static, in the sense that it has to possess a limit (expected value). Hence, the table is characterized by a single true length that never changes.

If these assumptions are met, then…

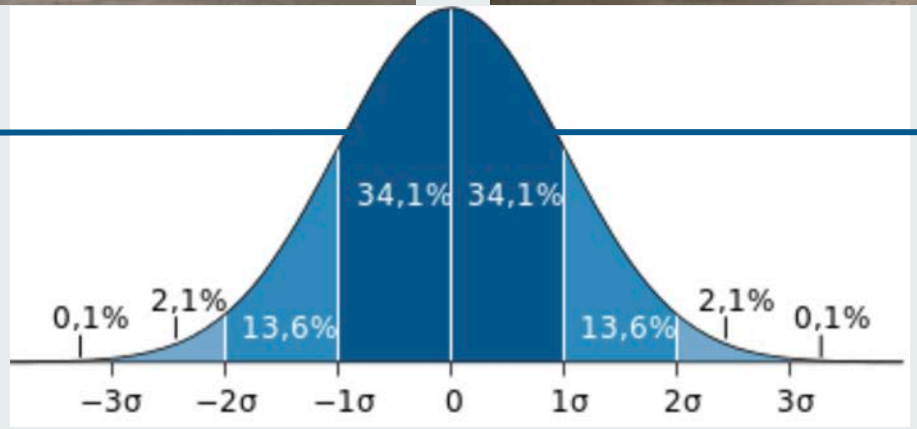
- …the true value = the average of infinitely many repeated observations/measurements
- …the standard deviation = the standard measurement of error (SE)

The SE allows one to calculate the probability of a certain observation given a certain average.
See the next slide…





SD = 60 mm



SD = 6 mm

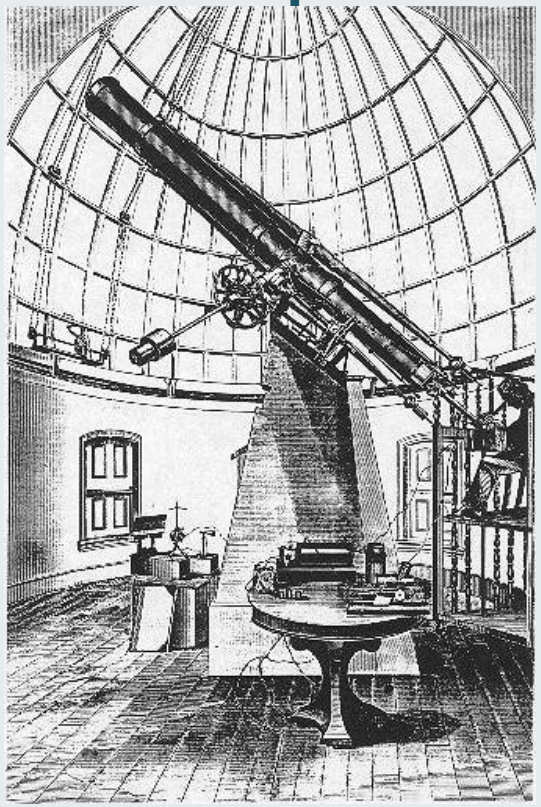
Hence: this procedure is more reliable

Given the SD's, the probability of measuring 5 meters is higher when using

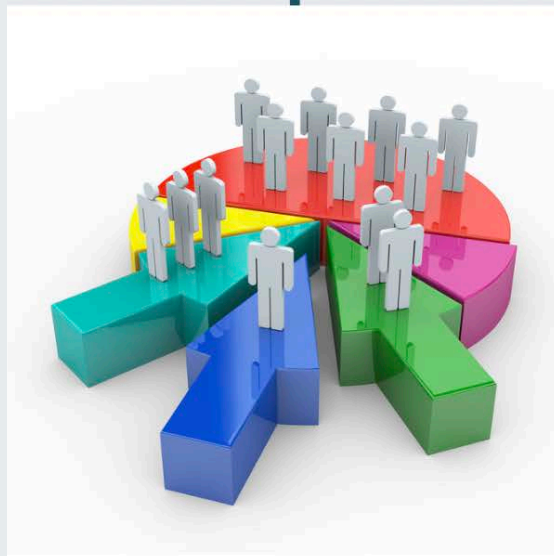
than when using

Adolphe Quetelet

1796 - 1874



Royal Observatory Belgium
1826



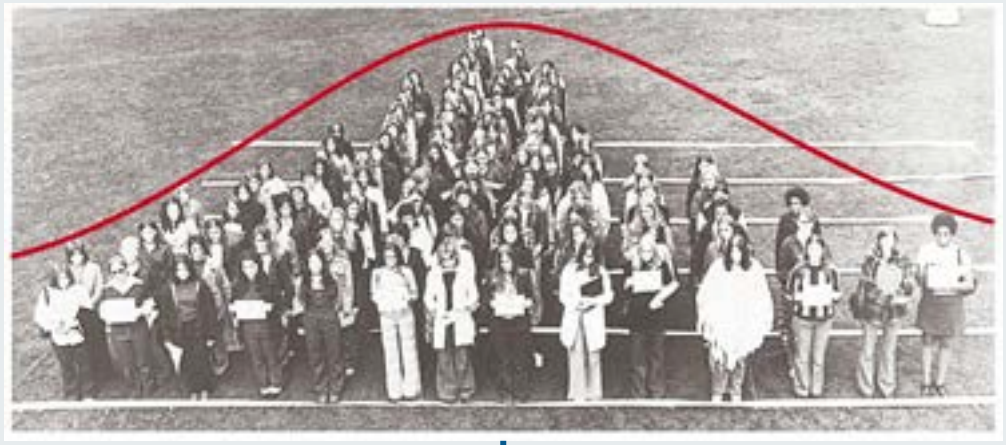
Social Physics &
public policy

A posteriori determination of probability

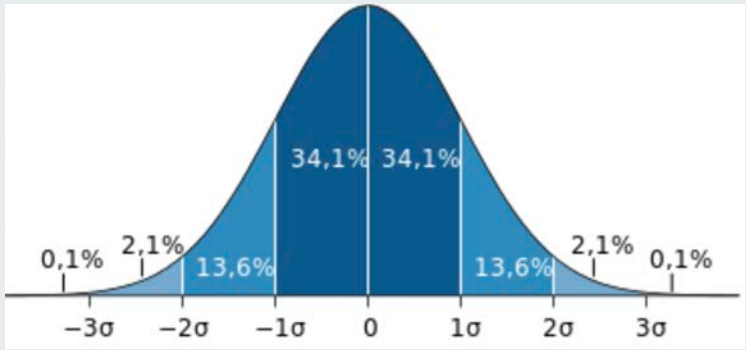
e.g. insurances, policy

Distribution of
measurement errors

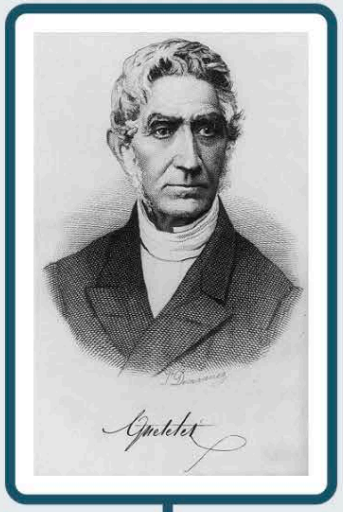
Distribution of
deviating individuals



average
=
true value
(**real** entity)



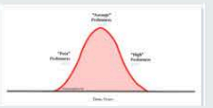
average
=
norm
(**fictitious** ideal human)



5738 x 1 soldier

1 x 5738 soldiers

symmetrical distribution



symmetrical distribution

expected value =
weighted average =
true value

independent

measurement errors are normally distributed

inter- and intra-individual data are interchangeable

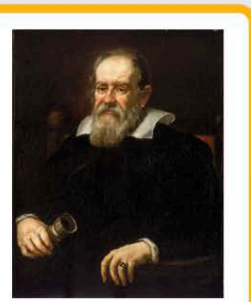
variation in a population is normally distributed

expected value =
weighted average =
the average man

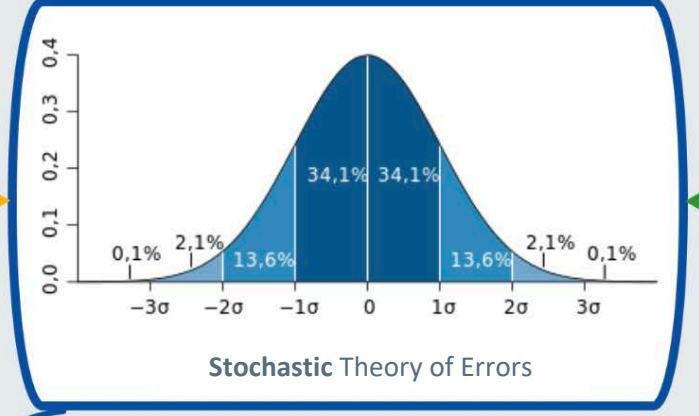
independent

Summary

- In order to reliably determine a test score, one utilizes the stochastic theory of errors
- Which is based on astronomy and physics
- Adolphe Quetelet was the first mathematician who applied the stochastic theory of errors on social data and thereby introduced the ‘average man’.
- He postulated that inter- and intra-individual data are interchangeable
- This principle is also discernable in the Classical Test Theory (CTT) by Lord & Novick (1968), which is the most commonly used test theory in psychology and education
- CTT is based on the frequentist interpretation of probability

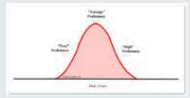


Theory of errors



Stochastic Theory of Errors

Frequentist interpretation of probability



symmetrical distribution

expected value = weighted average = **true value**

independent

measurement errors are normally distributed

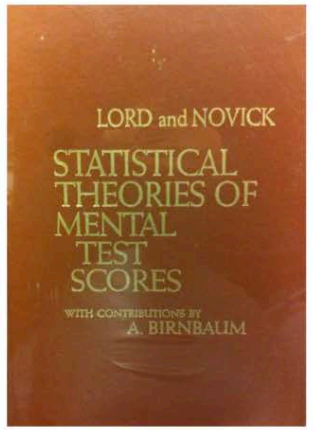
inter- and intra-individual data are interchangeable

variation in a population is normally distributed

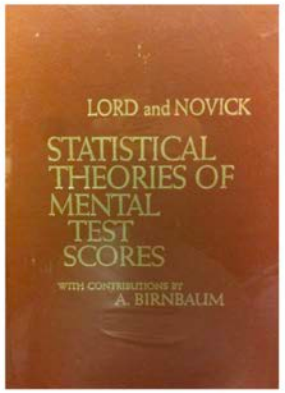
symmetrical distribution

expected value = weighted average = **the average man**

independent



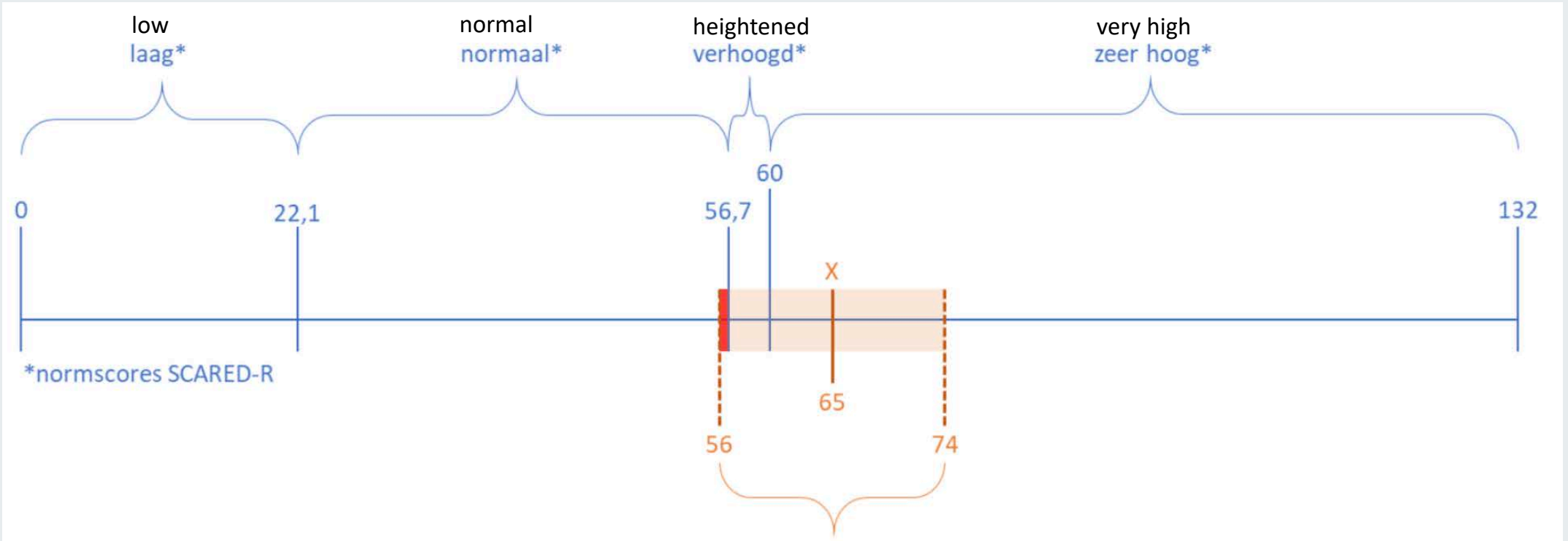
KTT: X = T + E



KTT: X = T + E

Dolora

How is the reliability of a test score determined?



How is a confidence interval determined?

X = 65

Confidence interval = 18 (± 9)
(according to the norms in the manual)

The CTT demands an intra-individual measurement of Dolora, which entails that we……

…**infinitely** often…

….administer the **same test**….

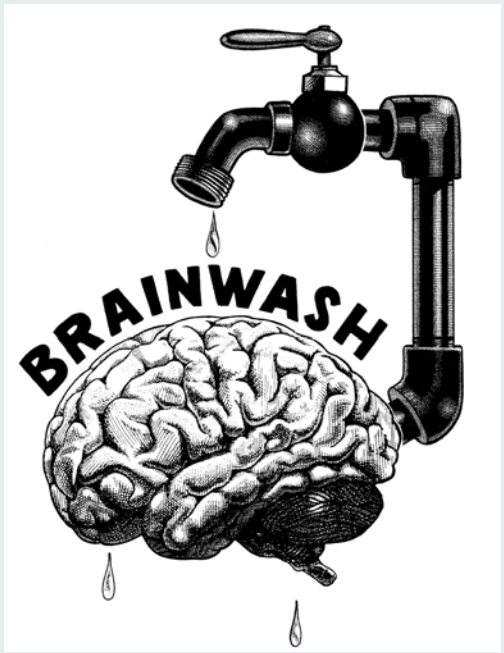
….under **equal conditions**…

……while maintaining mutually **independence** of all test administrations

Stochastic theory of errors

Thought experiment in Lord and Novick (1968)

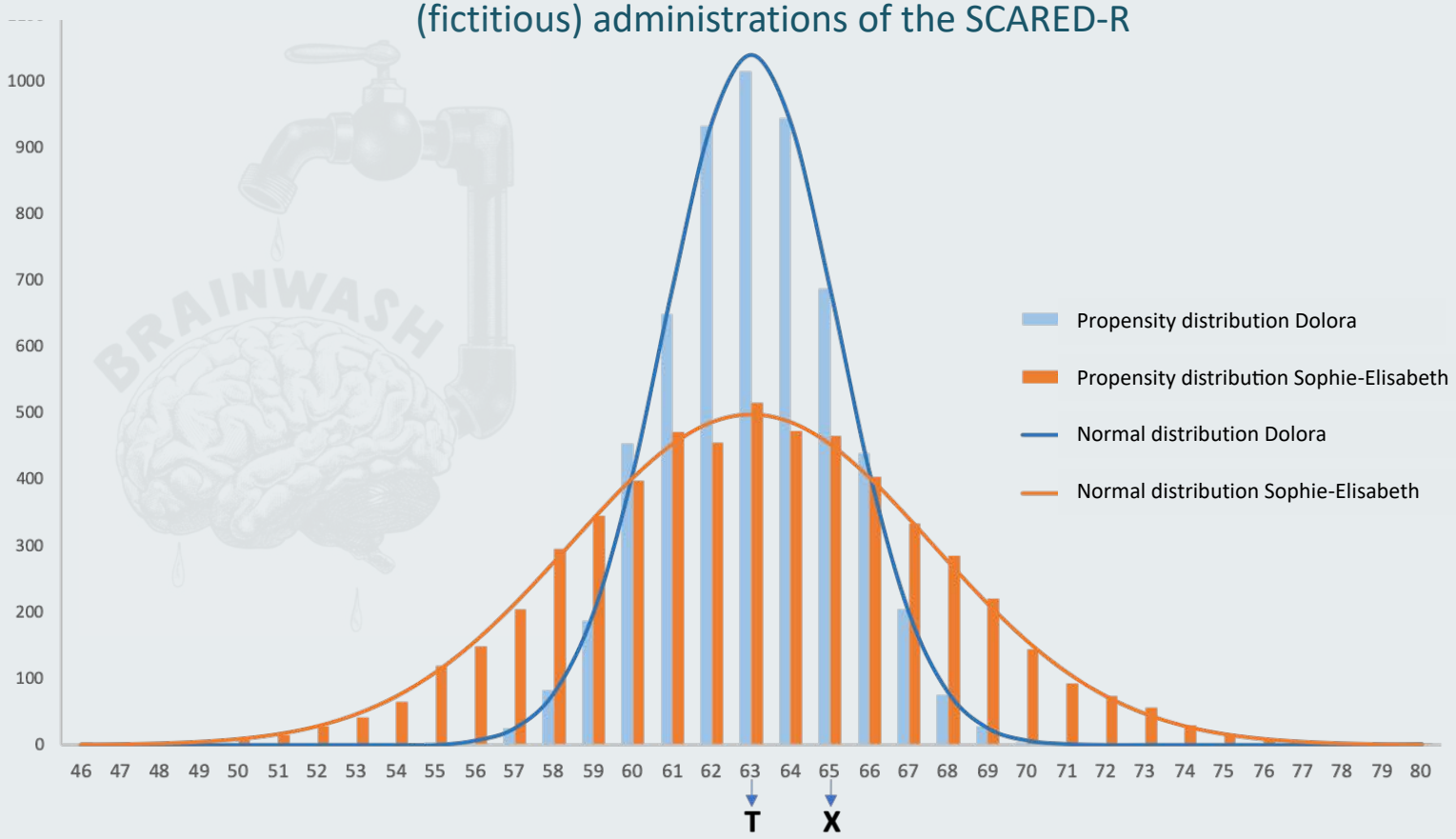
Suppose we ask an individual, Mr Brown, repeatedly whether he is in favour of the United Nations; suppose further that after each question we 'wash his brains' and ask him the same question again. Because Mr Brown is not certain as to how he feels about the United Nations, he will sometimes give a favorable and sometimes an unfavorable answer. Having gone through this procedure many times, we then compute the proportion of times Mr Brown was in favor of the United Nations. (Lazarsfeld, 1959; quoted in Lord and Novick, 1968, pp. 29–30)



Dolora may not learn

The expected value (limit) has to exist, hence Dolora may not change

Two simulated propensity distributions based on 5.738 hypothetical (fictitious) administrations of the SCARED-R



UNKNONW, because this procedure is only possible in a thought experiment

T_D

Dolora's true score

SE_D

Dolora's standard error of measurement

As one **lacks** Dolora's **propensity distribution**, one uses the **data of a sample** comprised of different individuals to **estimate** the unobtainable information regarding Dolora

For pragmatic reasons, the CTT uses a measurement of a sample of different individuals (**interindividual**)

$\bar{x}$

average score in the **sample**

S

standard deviation (=average absolute deviation of the sample mean) in the **sample**

How is the reliability of a test score determined?

A common method is to equal the estimated true (value) score to the observed score of an individual

$$\hat{T}_D$$

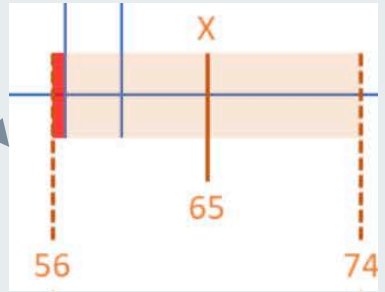
Estimated true value

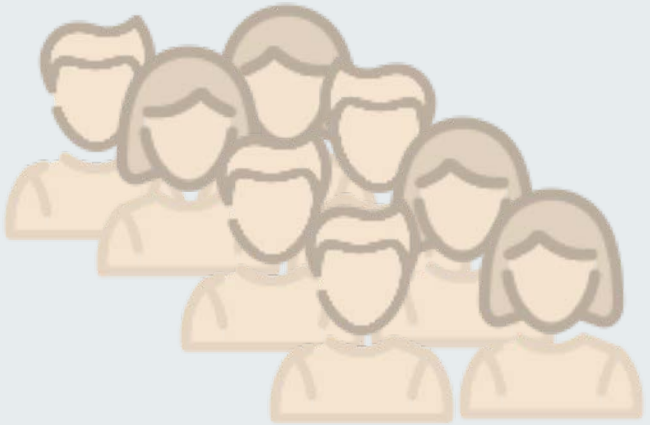
The estimation of the standard error of measurement is most commonly based on Cronbach's alpha and the sample variance

$$\widehat{SE}_D$$

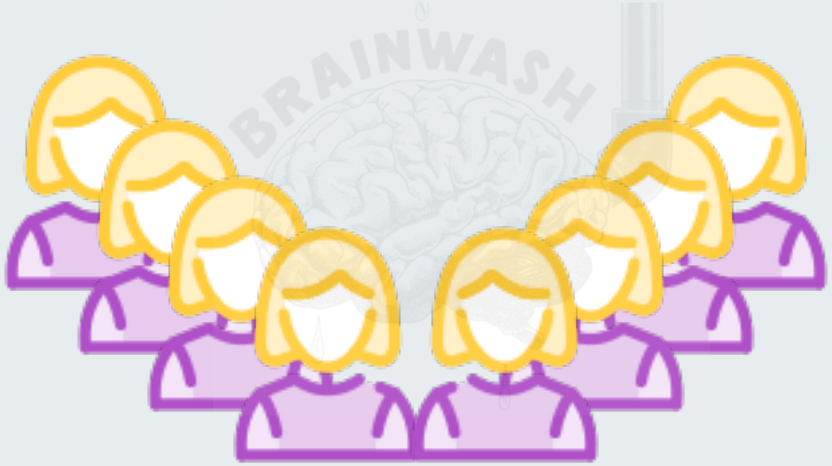
Estimated standard error of measurement

times (±1,96)

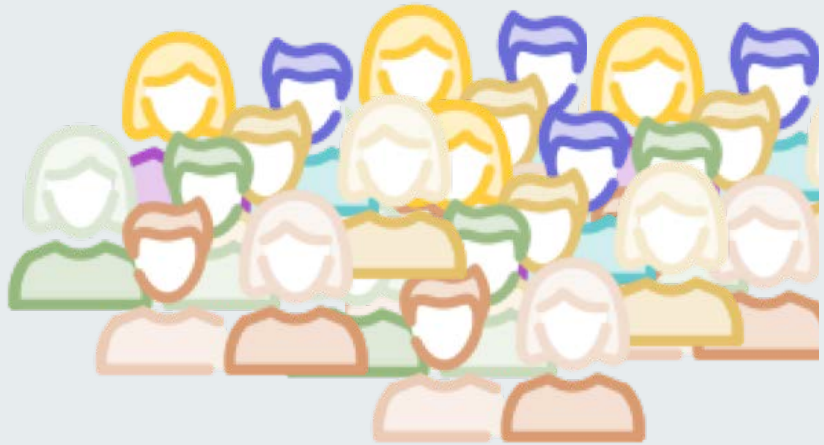




Each and every individual that takes the test is allocated the **same estimated standard error** of measurement and consequently the **same confidence interval**



Test practice:
Intra- and interindividual data are interchanged



CTT requires infinitely many test administrations for the same individual, Dolora (intra-individual measurements)

Is one allowed to do this freely?
NO, because …

For pragmatic reasons, the CTT uses a single test administration for a sample of different individuals (inter-individual)

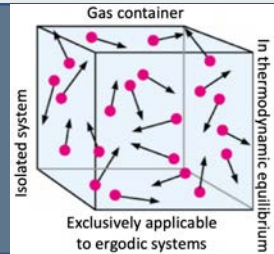
First, the characteristics of Dolora need to remain constant through time (stationary).

This does not apply to human beings

Second, the sample needs to exist of identical (homogeneous) copies of Dolora.

This does not apply to human beings

Only if these two assumptions are met, one is allowed to interchange intra- and inter-individual data. Such strict conditions are exclusively applicable to ergodic systems, and therefore not to human beings.



Humans are not ergodic

If one nevertheless replaces intra-individual data by inter-individual data…

…then it is conceivable that one reaches **incorrect** conclusions

Yet, one will **never** be able to verify exactly how incorrect ones conclusions really are

Hence: it is **impossible** to draw an informed conclusion, either correct or incorrect

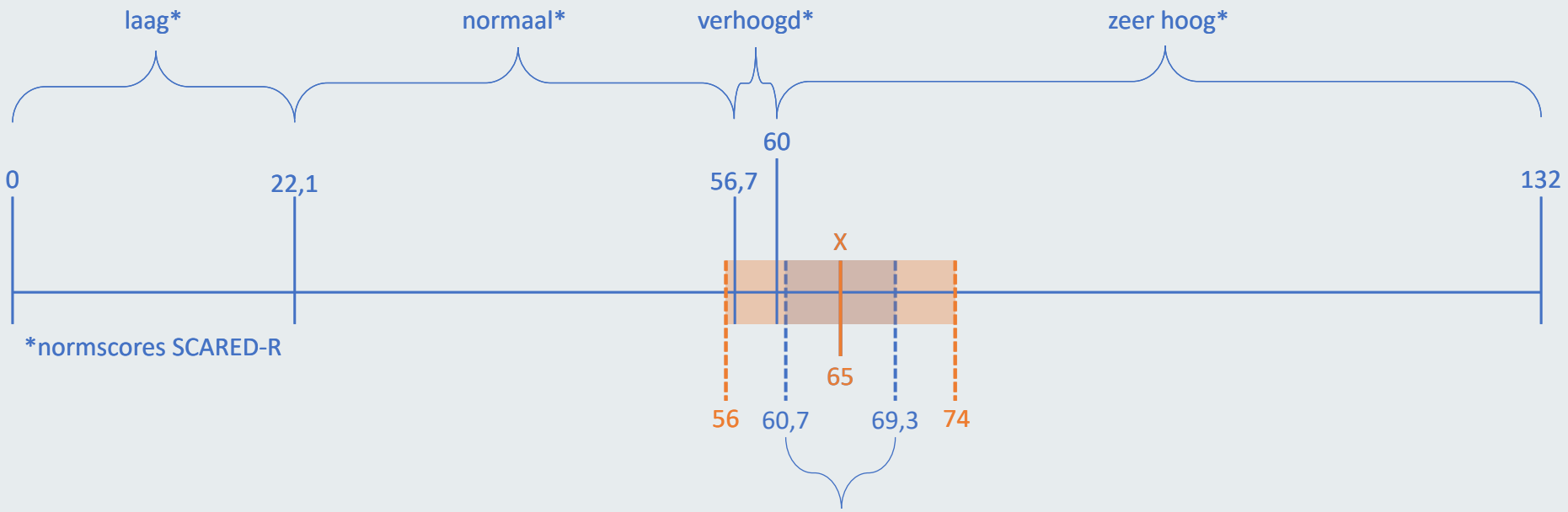


Figure 4: Dolora's confidence interval with her fictitious own SE and the SE of the test

Conclusion: As a result, the use of tests based on nomothetic data are discouraged for the assessment of an individual