CBI Yuki Ito

# Statistical Thinking for the 21st Century

## Chapter1,2

# Chapter 1 Introduction

1. What is statistical thinking?

2. What can statistics do for us?

3. Fundamental concepts of statistics.

4. Causality concepts of statistics.

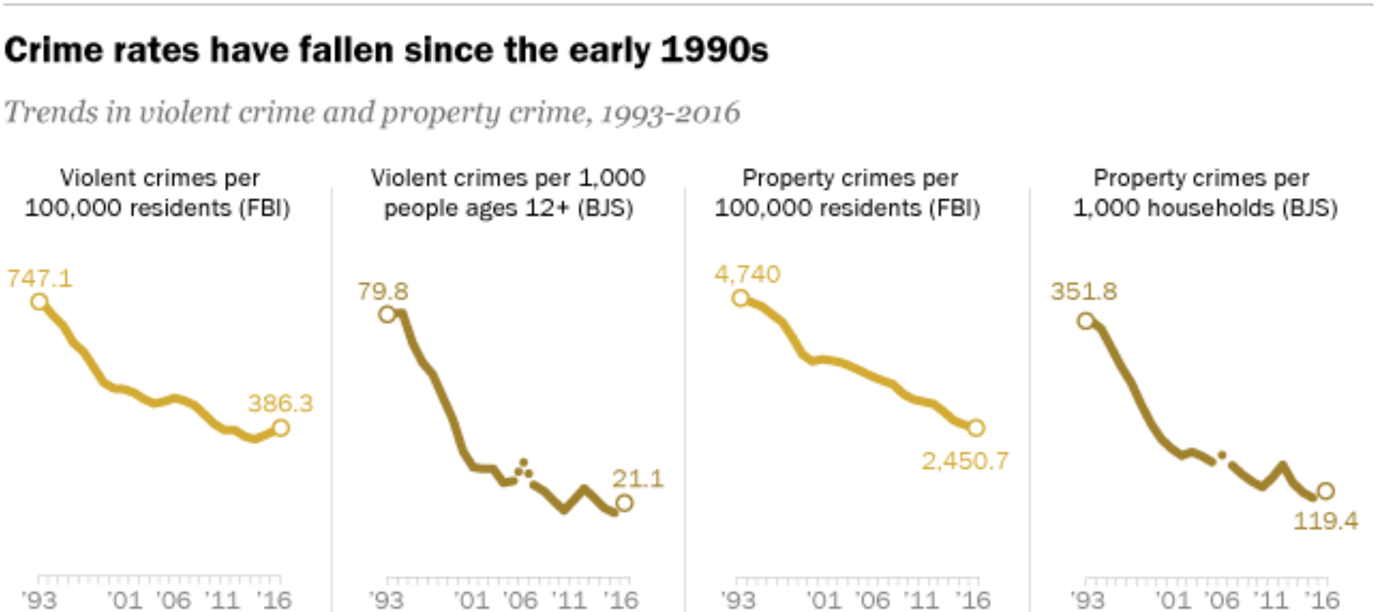
5. Suggested readings.

# 1.1 What is statistical thinking?

● Statistical thinking captures essential aspects of its structure, and also provides us with some idea of how uncertain we are about our knowledge.

Human intuition often tries to answer the questions that we can answer using statistical thinking, but often gets the answer wrong.

For example…



# 1.2 What can statistics do for us?

There are three major things that we can do with statistics:

- *Describe*:
  The world is complex and we often need to describe it in a simplified way that we can understand.

- *Decide*:
  We often need to make decisions based on data, usually in the face of uncertainty.

- *Predict*:
  We often wish to make predictions about new situations based on our knowledge of previous situations.

# 1.2 What can statistics do for us?

<u>Is saturated fat in our diet a bad thing?</u>

*Describe*:
If this number is greater than 1, people in the group are *more* likely to die than are people in the lowest quintile. If it's less than 1, people in the group are *less* likely to die.

*Decide*:
People who ate more saturated fat were *less* likely to die, and the more they ate, the bigger this effect was.

*Predict*:
Based on the data we would like to make predictions about future outcomes.
An important aspect of prediction is that it requires us to generalize from the data.

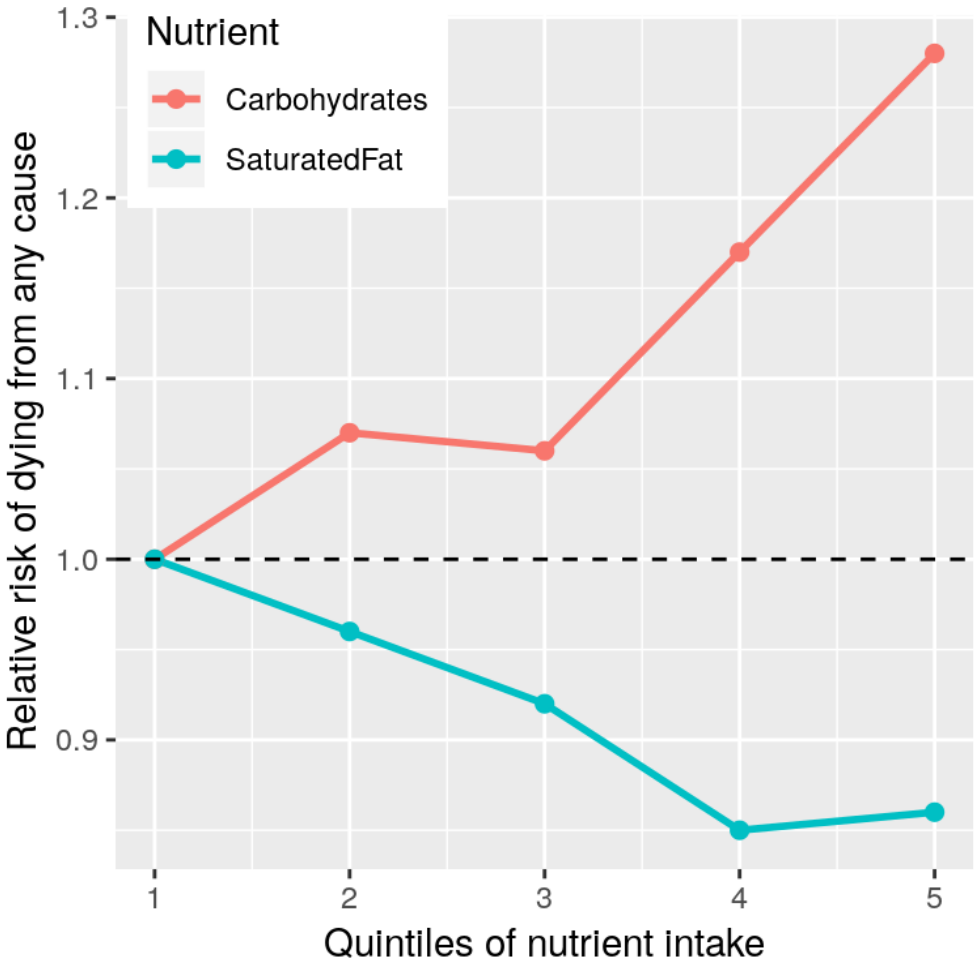



Figure 1.1:
This plot shows the relationship between death from any cause and the relative intake of saturated fats and carbohydrates.

# 1.3 Fundamental concepts of statistics

## 1.3.1 Learning from data

- One way to think of statistics is as a set of tools that enable us to learn from data.
- Statistics provides a way to describe how new data can be best used to update our beliefs.

## 1.3.2 Aggregation

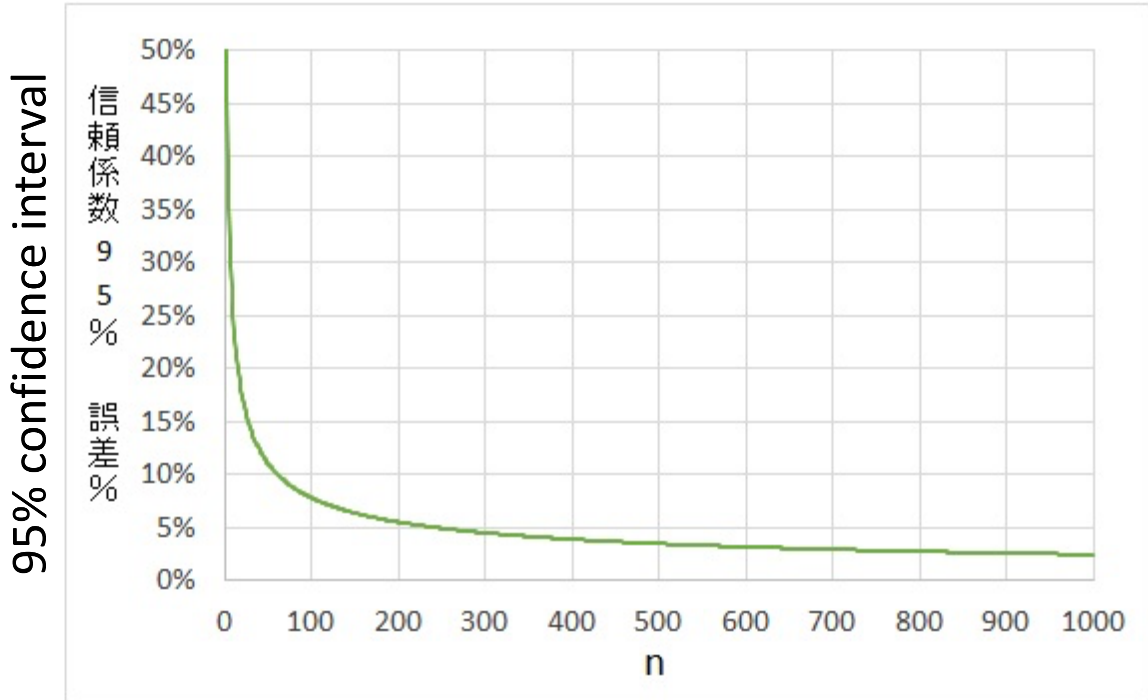
- Another way to think of statistics is "the science of throwing away data".

## 1.3.3 Uncertainty

- Statistics provides the tools to characterize uncertainty, to make decisions under uncertainty, and to make predictions whose uncertainty we can quantify.
- Statistical analysis can never "prove" a hypothesis, unlike mathematical proof. <u>Statistics can provide us with evidence, but it's always tentative and always subject to the uncertainty.</u>

# 1.3.4 Sampling

- How much data we need? Samples obtained in the right way can summarize the entire population based on a small sample of the population.




- 95% confidence interval:

$1.96 \times \sqrt{(p \times (1-p)/n)}$

- Sample proportion: 0.2

- $1.96 \times \sqrt{(0.2 \times (1-0.2)/600)}$ $=0.032$

20%±3.2%

# 1.4 Causality and statistics

- **If we eat more saturated fat, will that cause us to live longer?**

➢ This is because we don't know whether there is a direct causal relationship between eating saturated fat and living longer.

Correlation does not imply causation.

Causation can be demonstrated using studies that experimentally control and manipulate a specific factor.

➢ Such a study is referred to as a ***randomized controlled trial*****(RCT)**.

## 1.4 Causality and statistics

- In the RCT, a sampling group is assigned them to either a treatment group or a control group.
  - ➢ There are no other differences between the groups that would confound the treatment effect.

- We still can't be certain because sometimes randomization yields treatment vs control groups that do vary in some important way.
  - ➢ Removing the influence of a confound from the data can be very difficult.

# Chapter 2 Working with data

1. What are data?
   i. Qualitative data
   ii. Quantitative data
2. Scale of measurement
   i. Why do scales of measurement matter?
3. What makes a good measurement?
   i. Reliability
   ii. Validity

# 2.1 What are data?

## 2.1.1 Qualitative data

- Data are composed of variables, where a variable reflects a unique measurement or quantity. Some variables are qualitative, meaning that they describe a quality rather than a numeric quantity.
  - Gender, Blood Type, Weather etc.
- We will often code qualitative data using numbers in order to make them easier to work.

## 2.1.2 Quantitative data

- More commonly in statistics we will work with *quantitative* data, meaning data that are numerical.
  - Length, weight, volume, temperature, time etc.

| Why are you taking this class? | Number of students |
| --- | --- |
| It fulfills a degree plan requirement | 105 |
| It fulfills a General Education Breadth Requirement | 32 |
| It is not required but I am interested in the topic | 11 |
| Other | 4 |

Table 2.1: Counts of the prevalence of different responses to the question “Why are you taking this class?”

※Note that the students’ answers were qualitative, but we generated a quantitative summary of them by counting how many students gave each response.

# 2.1.2.1 Types of numbers

- **Binary numbers**:

  The simplest are binary numbers – that is, 0 or 1.

  We will often use binary numbers to represent whether something is true or false, or present or absent.

- **Integers**:

  Integers are whole numbers with no fractional or decimal part.

- **Real numbers**:

  Most commonly in statistics work with real numbers.

## 2.2 Scales of measurement

**There are four ways in which the different values of a variable can differ.**

- **Identity**:

  Each value of the variable has a unique meaning.

- **Magnitude**:

  The values of the variable reflect different magnitudes and have an ordered relationship to one another.

- **Equal intervals**:

  Units along the scale of measurement are equal to one another.

- **Zero is the minimum value**:

  The scale has a true zero point, below which no values exist.

# 2.2 Scales of measurement

There are four different scales of measurement that go along with these different ways that values of a variable can differ.

- **Nominal scale**:

  A nominal variable satisfies the criterion of **identity**, such that each value of the variable represents something different, but the numbers simply serve as qualitative labels.

  - For example, Male=1, Female=0.

- **Ordinal scale**:

  An ordinal variable satisfies the criteria of **identity** and **magnitude**, such that the values can be ordered in terms of their magnitude.

  - For example, **ranking of sales**.

| Rank | Day | CD Title | Artist | Sales (Ten thousand) |
|---|---|---|---|---|
| 1 | 1999/3/10 | First Love | 宇多田ヒカル | 765.0 |
| 2 | 1998/5/20 | B'z The Best "Pleasure" | B'z | 513.6 |
| 3 | 1997/10/1 | REVIEW～BEST OF GLAY | GLAY | 487.6 |
| 4 | 2001/3/28 | Distance | 宇多田ヒカル | 447.2 |

# 2.2 Scales of measurement

● Interval scale:

An interval scale has **all of the features of an ordinal scale**, but in addition the **intervals** between units on the measurement scale can be treated as equal.

- Physical temperature like Celsius.

● Ratio scale:

A ratio scale variable has all features:
identity, magnitude, equal intervals, and a minimum value of zero.
**The difference between a ratio scale variable and an interval scale variable is that the ratio scale variable has a true zero point.**

- Physical temperature like Kelvin.

## 2.2.1 Why do scales of measurement matter?

| | Equal/not equal | Greater than/less than | Add/subtract | Multiply/divide |
|---|---|---|---|---|
| Nominal | OK | | | |
| Ordinal | OK | OK | | |
| Interval | OK | OK | OK | |
| Ratio | OK | OK | OK | OK |

Table 2.2: Different scales of measurement admit different types of numeric operations.

- Statistics that simply involve counting of different values : **Mode** — Any of the variable types.
- Statistics are based on ordering or ranking of values : **Median** — Ordinal, Interval, Ratio scale.
- Statistics that involve adding up values : ***Average, Mean*** — Interval, Ratio scale.

# 2.3 What makes a good measurement?

- In many fields such as psychology, the thing that we are measuring is not a physical feature, but instead is an unobservable theoretical concept.
- It is usually impossible to measure a construct without some amount of error like miss-read.
  - However, we generally want our measurement error to be as low as possible.

We usually distinguish two different aspects of a good measurement.

➡ Reliability, Validity

## 2.3.1 Reliability

Reliability quantifies the consistency of our measurements.

- **Test-retest reliability**:
  It measures how well the measurements agree if the same measurement is performed twice.

- **Inter-rater reliability**:
  In the case that the data includes subjective judgments, it measures and assesses to executing multiple ratings, comparing their ratings, and make sure that they agree well with each other.

Reliability is important if we want to compare one measurement to another.

## 2.3.2 *Validity*

There are three types of validity that are commonly discussed:

- **Face validity**:
  Does the measurement make sense on its face?
- **Construct validity**:
  Is the measurement related to other measurements in an appropriate way?
  - **Convergent validity** :
    This means that the measurement should be closely related to other measures that are thought to reflect the same construct.
  - **Divergent validity /Discriminant validity**:
    Measurements thought to reflect different constructs should be unrelated.
- **Predictive validity**:
  If our measurements are valid, then they should also predict other outcomes.

# 2.3.1 Reliability/2.3.2 Validity

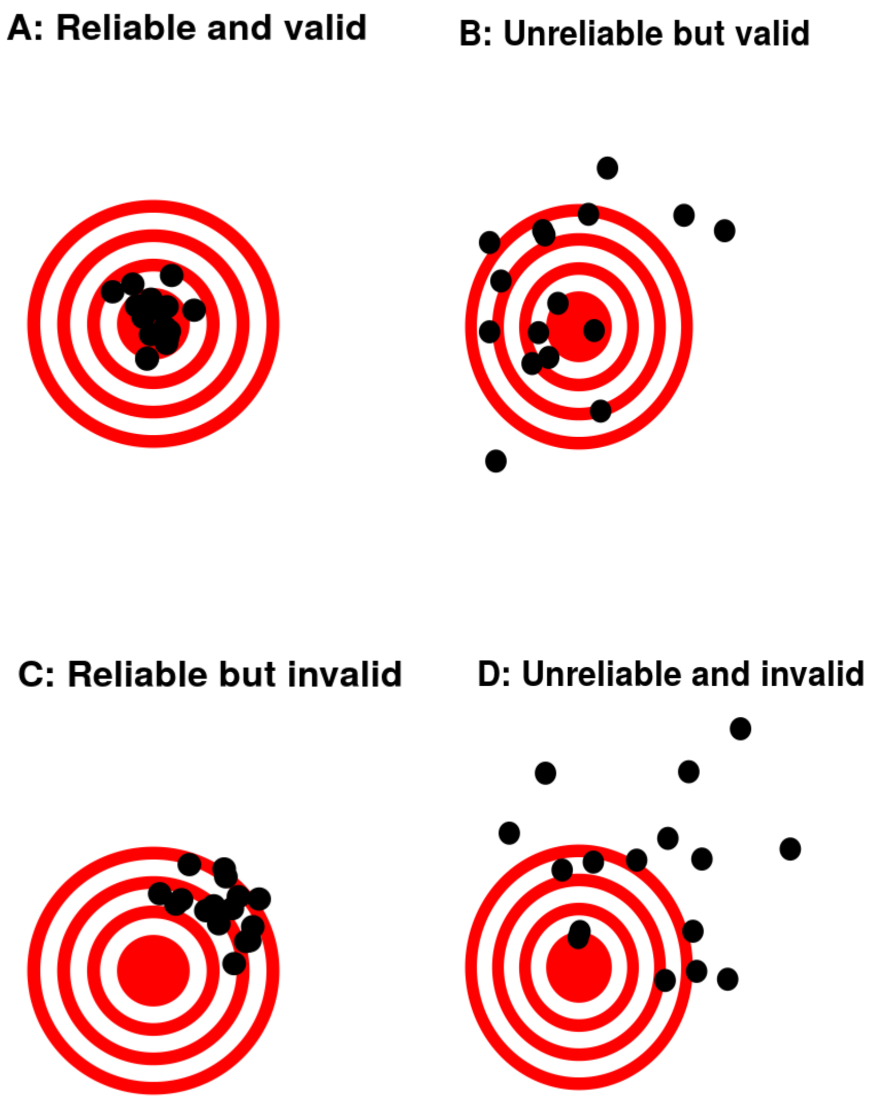



Figure 2.1:

- A figure demonstrating the distinction between reliability and validity.
- Reliability refers to the consistency of location of shots.
- Validity refers to the accuracy of the shots with respect to the center of the bullseye.

If you are more interested in this topic, please check the below:

- https://healthpolicyhealthecon.com/2014/12/15/validity-and-reliability/
- http://kichikawaginjiro.blogspot.com/2013/06/blog-post_27.html

# 補助スライド

| 信頼性のタイプ(推定方法) | コンセプト | 評価方法(定量化・数値化する方法) |
|---|---|---|
| 評価者間信頼性(Inter-rater reliability) | 複数の評価者がいる場合、評価者が変わっても測定結果が安定していて影響を受けないかどうかの指標 | クラス内相関(Intra-class correlation) |
| 試験・再試験信頼性(Test-retest reliability) | テストを複数回やってみたときに、同じ結果が得られるかどうかの指標 | Test-retest method |
| 方法間信頼性(Inter-method reliability) | 異なった測定方法で同じターゲットを測定したときに、測定結果が測定方法に左右されるかどうかの指標 | (Spearman-Brown) split half |
| 内的整合性(Internal consistency reliability) | テストに含まれる項目全体が、同一の特性に対する測定を実現しているかどうかの指標 | Cronbachの$\alpha$ |

https://healthpolicyhealthecon.com/2014/12/15/validity-and-reliability/

## 信頼性係数
**reliability coefficient**

心理測定やテストの信頼性を表す指標で、どの程度再現性を持っているか評価する際に用いる。クロンバックの$\alpha$係数など複数の信頼性係数がある。

$X$が心理尺度による測定値、$T$がその真値、$E$は誤差であるとすると、これらの関係は次のように表される。

$$X = T + E$$

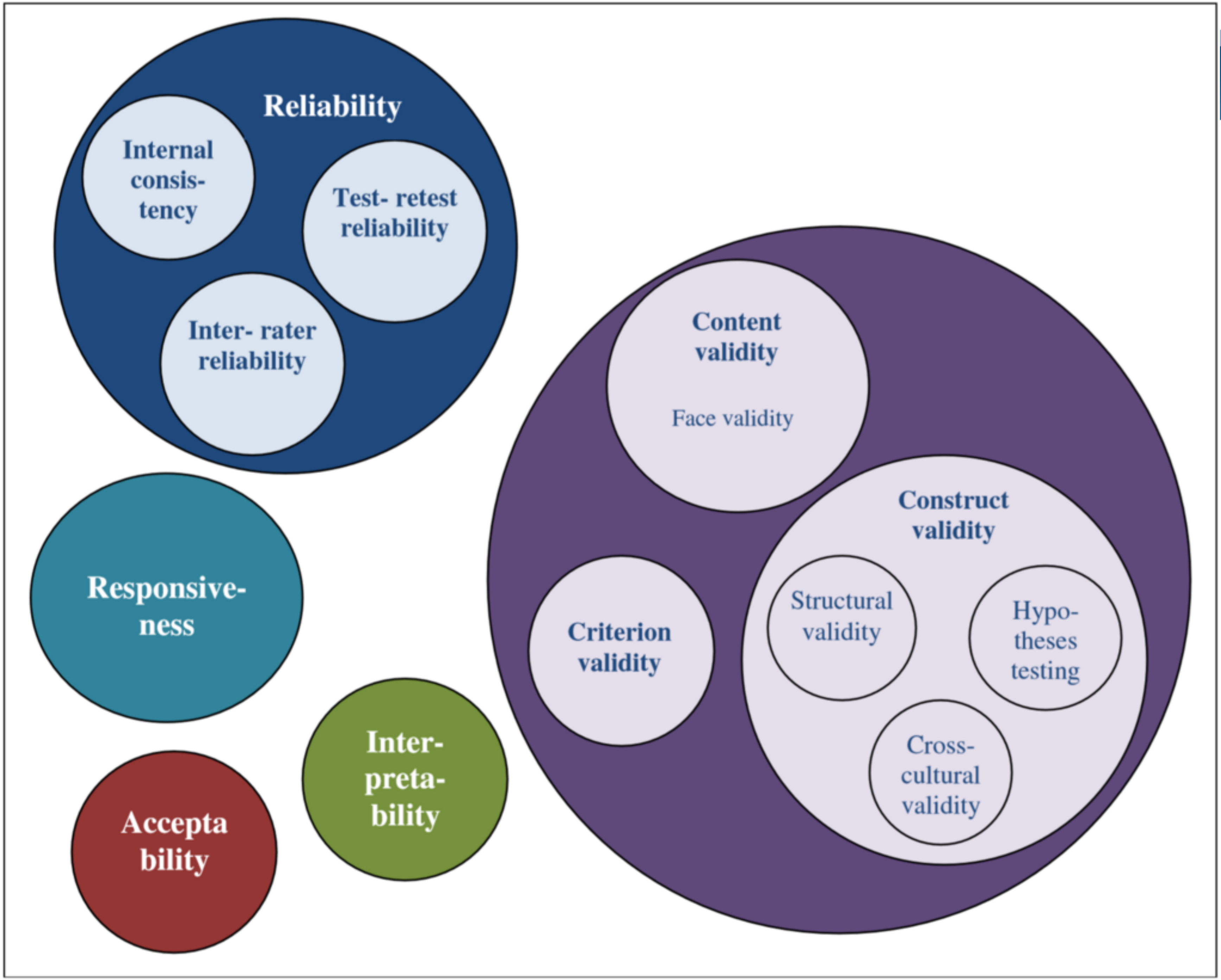
このとき、それらの測定値の分散は次のように表される

$$\sigma_x^2 = \sigma_T^2 + \sigma_E^2$$

信頼性係数$\rho$は以下の式により定義される。信頼性係数は全体の得点の分散に占める真値の分散の割合と言える。

$$\rho = \frac{\sigma_T^2}{\sigma_E^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

**Figure 1. Overview of measurement properties**




http://arichan.hatenablog.jp/entry/2017/12/20/162404

# 様々な妥当性概念

- （１）表面的妥当性（face validity）：直感的、印象的に主張される妥当性
  - ［例］特定の職務の適性検査が、当の職務内容と関連することを調べていなければ、その適性検査の表面的妥当性は低い
- （２）内容的妥当性（content validity）：テストの内容が測定すべき領域の適切な標本（sample）になっていること
  - ［例］世界史の学力試験：試験問題がいろんな時代や地域を適切にカヴァーしていれば、試験問題の内容的妥当性は高い

- （３）基準測度（criterion measure）に基づく妥当性
  - テストの得点と基準測度との比較によって判断される妥当性
- （３－１）予測的妥当性（predictive validity）
  - 将来の行動の測定値が基準測度となる
  - ［例］入学試験　＝　「入学後の学科履修の適性を測る」
    - 基準測度　＝　入学後の成績
    - 入学試験の成績と入学後の成績の相関が高い
      - →　その入学試験は予測的妥当性が高い
- （３－２）併存的妥当性（concurrent validity）
  - ２つのテストを併用する場合
  - ［例］精神科医の診断（Ａ）とパーソナリティ・テスト（Ｂ）を併用
    - 両者の相関が高ければ、Ａ［Ｂ］はＢ［Ａ］と高い併存的妥当性を持つと判断できる

- （４）構成概念妥当性（construct validity）
  - テストが予め計画された構成体（construct）を測定していること
  - 構成体（仮設構成体）　－　直接は観測できず、研究者が「ある」と仮定したもの
    - 態度、パーソナリティ、知能、など
- （４－１）収束的妥当性（convergent validity）
  - 同一対象に対する測定値間の相関が高くなること
- （４－２）弁別的妥当性（discriminant validity）
  - 異なった対象に対する測定値間の相関が低くなること
  - 相関が高い場合、２つの測定は同じ対象を測っている可能性がある