

1. データ尺度の種類
2. 度数分布表とヒストグラム
3. 代表値と散布度
4. 平均と標準偏差

【データ尺度の種類】

● 名義尺度 (nominal scale)

(質的変数とも)

● 順序尺度 (ordinal —)

● 間隔尺度 (interval —)

● 比率尺度 (ratio —)

(教科書 p. 41–48)

★ 上位の尺度のほうが使える演算が豊富

★ 上位の尺度は下位の尺度の特徴を兼ね備えている

➡ 分析手法の選択幅が広い

私たちが測定するものは、
たいてい順序尺度以下である。

★上位の尺度への変換には
一定の理論的根拠が必要

【度数分布表】

Frequencies コマンドを使う

- ★ 度数
- ★ 相対度数（%）
- ★ 累積度数・累積相対度数
- ★ 欠損値のあつかい

【棒グラフとヒストグラム】

- 棒グラフ……棒同士の間空白をあける。高さ(長さ)をよむ。
- histogram (柱グラフ)……柱の間隔をあけない。面積をよむ。

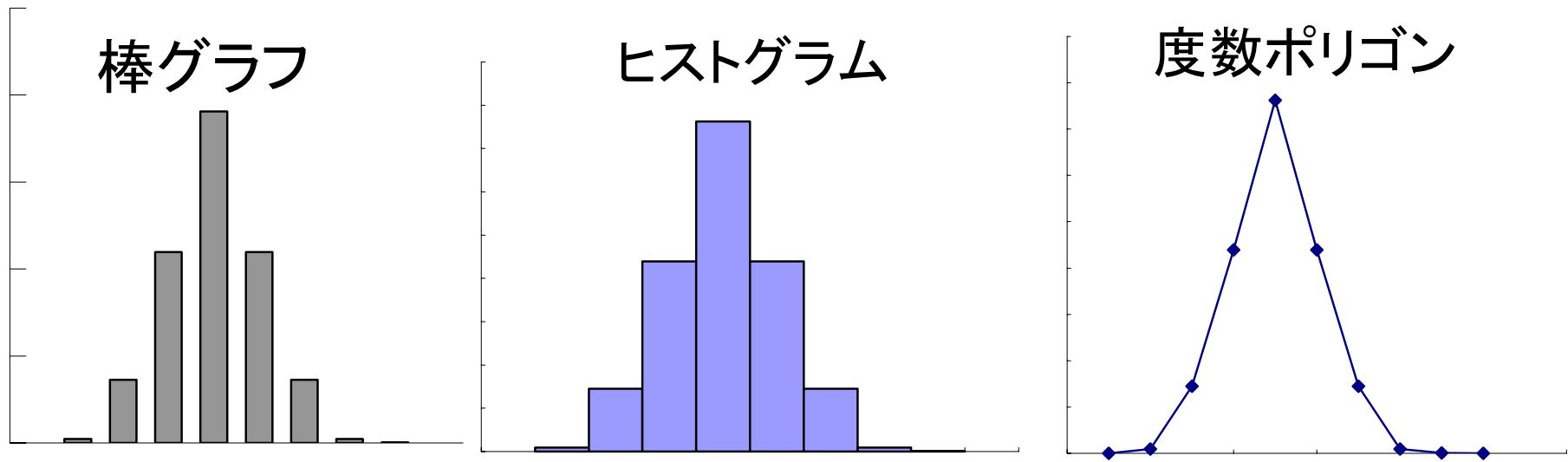
※縦軸は度数または%

★ 連続量を階級分けした場合

→ ヒストグラム

★ それ以外の場合 (質的変数 / 離散量) → 棒グラフ

※度数多角形 (polygon) は複数の変数の分布を比較するときに便利。



SPSS では histogram は書きにくい

★ recode で整形した上で度数分布表のメニューで「図表…」指定。棒グラフを書く

★ グラフ→インタラクティブ→ヒストグラム
では等間隔の区間に分割してくれる

【代表値と散布度】

★ 平均値 (mean) — 標準偏差 (SD)
(間隔尺度以上)

★ 中央値 (median) — 四分位偏差 (Q)
(順序尺度以上)

【平均値】

総和をデータ数で割ったもの

【標準偏差】

平均値からの偏差の2乗値の平均が「分散」
分散の平方根が「標準偏差」

★ 平均値と標準偏差はセットで使う

【SPSS のコマンド】

「記述統計」 → 「記述統計」

→ 変数とオプションを指定

【平均値を使うときの注意事項】

- ★平均値ははずれ値の影響を受けやすい。
あまりにかけはなれたケースがあるときは
 - 上下数%を取りのぞいたデータセットで計算する (調整平均)
 - 順位に変換したり中央値を使って分析

★平均値・標準偏差は**間隔尺度**以上のデータ
に対してしか意味をもたない。

順序尺度の平均値をとっていいのは

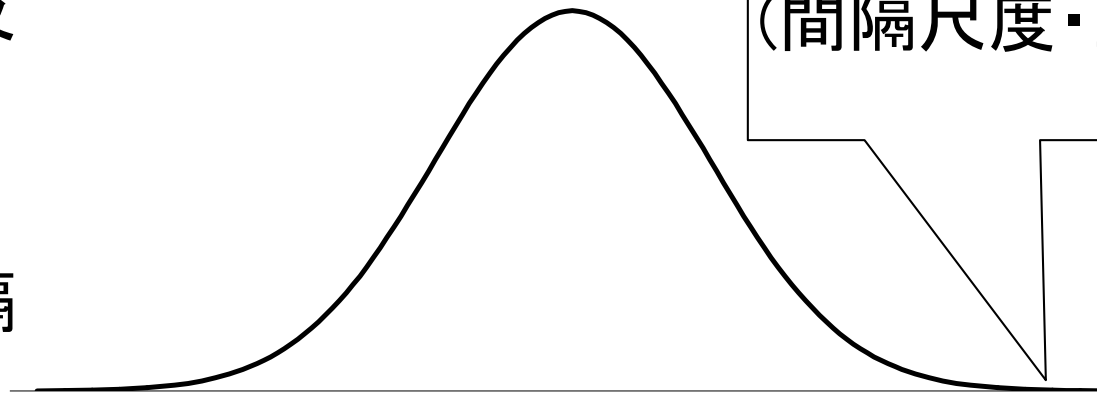
- ・潜在的には間隔尺度のはず

- ・測定のポイントが一定間隔

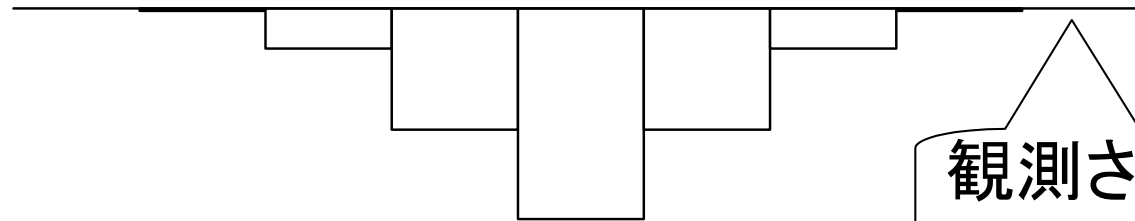
という2条件をとともに満たす場合

観測変数が潜在変数の尺度を反映していると推測できる場合のみ、順序尺度の観測変数を間隔尺度とみなしてよい

潜在変数
(間隔尺度・正規分布)



1 2 3 4 5 6 7 8 9



観測された変数

➔ 具体的には

● 4 点以上での測定している

● 正規分布に近似している (教科書 p.15) :

- 単峰性
- 左右対称性 (歪度)
- 中央への集中度 (尖度)

ヒストグラムを描いて検討するとよい。

正規分布との乖離度を統計的に検討する手法もある

- ➡ これらの条件を満たさない場合は
- 非線形変換 (対数・平方根など)
- 順位に変換したり中央値を使って分析

※ 間隔尺度のデータでも、
左右対称でないものについては
平均値よりも中央値のほうが
適当であることが多い

典型例：収入・人口など

【課題】

- (1) 本人年齢の度数分布表を出力し、中央値と上側 25% 点に印をつける
- (2) 適当な変数について棒グラフまたはヒストグラムを作成し、横軸上に平均値と標準偏差を書き入れる