

研究者に知っておいてほしい 統計分析の基礎知識

2024年10月

石井 秀宗

名古屋大学大学院教育発達科学研究科

- 統計分析法の選択
- 統計分析の役割
- 尺度水準と統計図表
- パラメトリック・ノンパラメトリック
- 統計分析におけるいくつかの用語
- 標準偏差・標準誤差
- 統計的検定
- 標本サイズ
- 効果量
- その他（時間があれば）

統計分析法の選択

- 標本の記述か母集団に関する推測か
- 分析に用いる変数はどれか
- 各変数の尺度水準は何か
- 母集団の「パラメタ」が関心対象か
- 関心は変数間の相関関係か説明関係か
- 独立変数および従属変数は何か
- 条件や集団(群)は何か, いくつあるか
- 対応のあるデータか対応のないデータか

統計的検定法の分類

表 説明関係を検討する統計的検定法の分類

	従属変数		独立変数（名義尺度，順序尺度）		
	尺度水準	特性	群数	対応の有無	
				対応あり	対応なし
パラメトリック法	比 間隔	平均値	2群	対応のある t 検定	対応のない t 検定
			多群	被験者内要因分散分析	被験者間要因分散分析
ノンパラメトリック法	順序	分布の位置	2群	ウィルコクソンの符号つき順位検定	ウィルコクソンの順位和検定 = マン・ホイットニの検定
			多群	フリードマンの検定	クラスカル・ウォリスの検定
	名義	度数比率	2群	マクネマーの検定	リスク差・リスク比・オッズ比の検定
			多群	コ克兰のQ検定	カイ2乗検定

表 分析目的による統計分析法の分類

変数間の相関関係を表す指標

量的変数間：相関係数

質的変数間：連関係数，ファイ係数，カッパ係数

従属変数に対する量的な独立変数の影響を検討する分析法

従属変数が量的変数：回帰分析

従属変数が質的変数：ロジスティック回帰分析，判別分析

変数間の関連を示すモデルを作る分析法

量的変数間：パス解析，構造方程式モデリング

質的変数間：対数線形モデル

変数や個体のまとまりを作る分析法

相関関係に基づく方法：主成分分析，因子分析

類似度に基づく方法：多次元尺度構成法，クラスター分析

統計分析の役割

- 標本の記述か母集団に関する推測か
- 分析に用いる変数はどれか
- 各変数の尺度水準は何か
- 母集団の「パラメタ」が関心対象か
- 関心は変数間の相関関係か説明関係か
- 独立変数および従属変数は何か
- 条件や集団(群)は何か, いくつあるか
- 対応のあるデータか対応のないデータか

- 統計的記述（記述統計）

データ（標本）の持つ情報を分かりやすくまとめる
統計図表，記述統計量

- 統計的推測（推測統計）

母集団における状態を推測する
推定値，信頼区間，統計的仮説検定

- （構成概念の）測定

性格，能力，感情など構成概念の程度を数量化する
因子分析（確認的，探索的），信頼性，妥当性

- 統計学の教科書には必ず「母集団と標本」のような項目がある
- データを使って何かを言うとき,
そのデータ(標本)についての説明なのか,
そのデータの元となる母集団についての推測なのか
を明確に意識し区別する
- 「有意ではなかったが平均値に差があった」のような
どっちつかずの主張はしないこと

尺度水準と統計図表

- 標本の記述か母集団に関する推測か
- 分析に用いる変数はどれか
- 各変数の尺度水準は何か
- 母集団の「パラメタ」が関心対象か
- 関心は変数間の相関関係か説明関係か
- 独立変数および従属変数は何か
- 条件や集団(群)は何か, いくつあるか
- 対応のあるデータか対応のないデータか

- 尺度水準

数値データを数量的に扱える程度の基準

- 名義尺度：単なる分類記号 背番号, 国番号, マイナンバー
- 順序尺度：順序性のみ表す 順位, 病状ステージ
- 間隔尺度：数値の差が特性の程度の差を表す 心理尺度
- 比(例)尺度：数値の比が特性の程度の比を表す 物理量

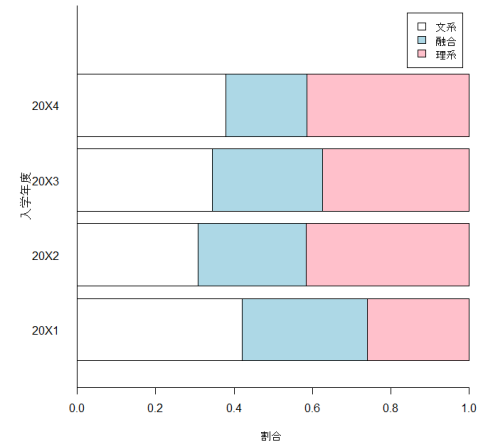
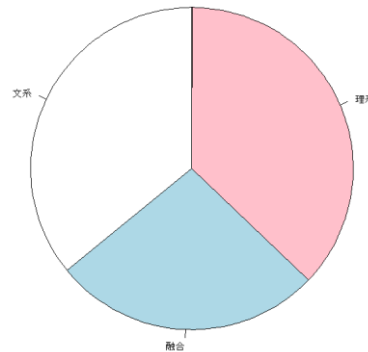
- 名義尺度・順序尺度のデータ (質的データ・質的変数)
 - 度数, 相対度数(割合), 最頻値, 連関係数, 一致係数
- 間隔尺度・比(例)尺度のデータ (量的データ・量的変)
 - 度数, 相対度数(割合), 累積度数, 累積相対度数, 最頻値, 平均値, 中央値, 最小値, 最大値, 四分位数, 標準偏差, 分散, 範囲, 四分位範囲, 相関係数
- 間隔尺度・比(例)尺度のデータのほうが適用できる統計手法が多い

- あてはまる程度を「1.まったくあてはまらない」～「5.とてもあてはまる」の5段階で評定したデータなど
- 本来的には**順序尺度**
1つの**従属変数**として分析する場合は、分布がよく分からないので、順序尺度として扱うほうがよい → ノンパラメトリック法
- **間隔尺度**とすることも多い
性格検査など**複数の項目を合計**（データ同士の足し算）して分析する場合は、間隔尺度として扱っている
→ パラメトリック法

- 各値の度数や割合の比較

円グラフ, 帯グラフ

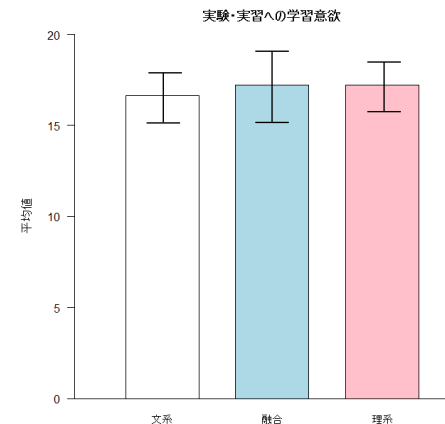
文系, 理系, 融合領域
の学生人数の割合
(全体及び年度別)



- 値別に群分けしたときの何らかの統計量の比較

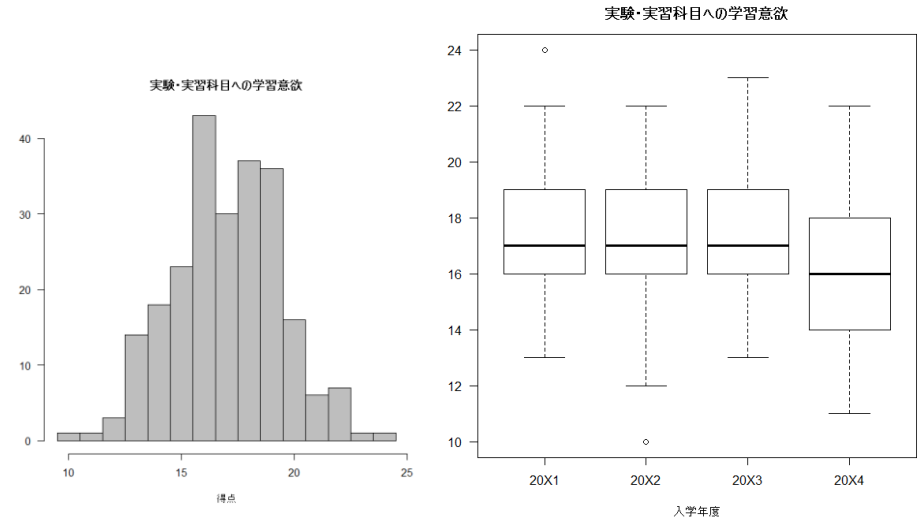
棒グラフ

文系, 理系, 融合領域の学生の
学習意欲得点の平均値



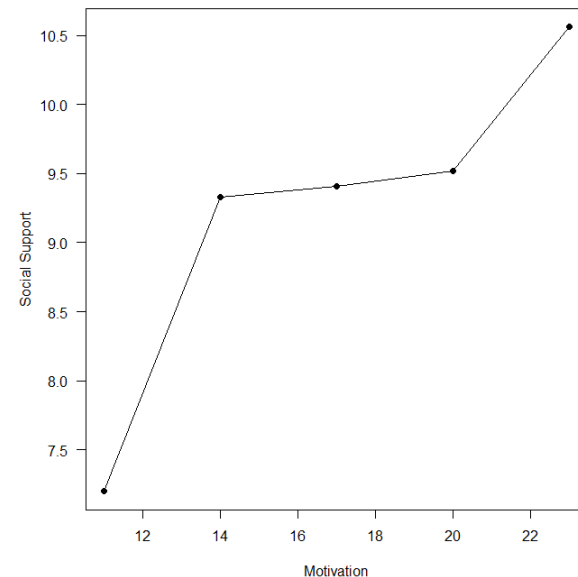
- 各値(階級)の度数分布
ヒストグラム, 箱ひげ図

実験・実習科目の学習意欲得点
(全体及び年度別)



- 値で群分けしたときの
何らかの統計量の比較
折れ線グラフ

学習意欲得点で5群に分けたときの
ソーシャルサポート得点の平均値



- 基準値がある場合は, それを使うかまず検討
BDI (ベック抑うつ質問票)
GHQ (精神健康調査票) など
- 臨床群を構成するのであれば基準値を用いる
- 相対的な程度の違いで群分けするので良ければ, 基準値を使わないこともある
- とくに決まりはない

- 基準値がない, もしくは基準値があっても使わない場合は, 領域の慣習も考慮して設定

M群... $100/M\%$ 各群の人数が均等になるように群分け
得点範囲/M 得点範囲を均等に分割して群分け

3群... 平均 $\pm 1SD$ 高群, 低群の人数が少なくなりやすい
平均 $\pm 0.5SD$ 5段階評価の1と2, 3, 4と5 に相当

2群... 平均値 分布に偏りがあると人数が均等にならない
中央値 人数が均等になる

パラメトリック ノンパラメトリック

- 標本の記述か母集団に関する推測か
- 分析に用いる変数はどれか
- 各変数の尺度水準は何か
- 母集団の「パラメタ」が関心対象か
- 関心は変数間の相関関係か説明関係か
- 独立変数および従属変数は何か
- 条件や集団(群)は何か, いくつあるか
- 対応のあるデータか対応のないデータか

- パラメトリックな分析手法
- 母集団におけるデータの分布に確率分布を仮定するなどとして
- 母平均 μ や母相関係数 ρ など確率分布の「パラメタ」と言われるものに関する推測を行う分析手法
- 間隔尺度や比(例)尺度のデータを分析することが多い

- ノンパラメトリックな分析手法
- 母集団におけるデータの分布に確率分布を仮定せず
- 母集団における, データ分布の位置, 度数, 割合, 比率などに関する推測を行う分析手法
- 名義尺度や順序尺度のデータを分析することが多い

- パラメトリック法, ノンパラメトリック法の選択は
- 母集団におけるデータの分布に確率分布を仮定するか否か, 母集団パラメタに関する推測か否かの問題
- 標本サイズ(データ数)の大小の問題ではない
- 「データが少ないからノンパラメトリック」とするのは誤り

- IQなど母集団分布がある程度分かっている変数
- 段階評定項目の合計得点
- 母集団分布に確率分布を仮定することにさほど問題はない
- 合計得点を算出している時点で、段階評定を間隔尺度として扱っている
- 人数の多寡に関わらず、パラメトリックな方法で分析することに一定の合理性がある

- 1つの段階評定項目（順序尺度変数）の度数分布
- 名義尺度・順序尺度の変数の度数分布やクロス集計表
- 母集団分布に確率分布を仮定しづらい
- 関心対象が母平均や母相関係数などパラメタではない
- 人数の多寡に関わらず, ノンパラメトリックな方法で分析するのが適切

- 入院患者とその家族における入院生活に対する満足度（5段階評価）の比較
- ノンパラメトリック法：ウィルコクソンの符号順位検定
- パラメトリック法：対応のある t 検定
- いずれの方法を用いるにしろ統計的に有意だが、この場合はノンパラメトリック法のほうが適切

表 入院満足度データの度数分布表と分析結果

	カテゴリ					Total	M	SD
	1	2	3	4	5			
患者	7 14%	16 32%	12 24%	13 26%	2 4%	50 100%	2.74	1.12
家族	1 2%	13 26%	18 36%	12 24%	6 12%	50 100%	3.18	1.02

V	p
220.5	.025
t	p
-2.45	.009

統計分析における いくつかの用語

- 標本の記述か母集団に関する推測か
- 分析に用いる変数はどれか
- 各変数の尺度水準は何か
- 母集団の「パラメタ」が関心対象か
- 関心は変数間の相関関係か説明関係か
- 独立変数および従属変数は何か
- 条件や集団(群)は何か, いくつあるか
- 対応のあるデータか対応のないデータか

• 相関関係

変数間の対称な関連性

ある変数の値が〇〇であると，他の変数の値は××であるという傾向

身長が高い児童は体重が重い
＝ 体重が重い児童は身長が高い } 意味することは同じ

• 説明関係・予測関係

変数間の非対称な（方向性のある）関連性

一方の変数の値の違いが，他方の変数の値にどのように影響するか
（説明するか）という関係

親が読書好きだと子どもの国語の成績は良くなる
≠ 子どもの国語の成績が良いと親は読書好きになる } 意味することが異なる

- 説明関係・予測関係を記述する変数

- 独立変数

説明する側の変数，条件の違いを表す変数

説明変数，予測変数，要因，因子

ex. 親が読書好きな程度

- 従属変数

説明される側の変数，条件の違いが影響するかの関心対象となる変数

基準変数，反応変数，結果変数，目的変数

ex. 子どもの国語の成績

- 要因 (Factor, 因子)

条件や集団の違いを表す独立変数：{親が読書好きな程度}

- 水準 (Level)

要因における各条件，各集団：{高，中，低}

- 群 (Group)

各水準に含まれる研究参加者またはその従属変数データの集合

親が読書好き高群の子どもの国語成績：{88, 92, 74, ...}

親が読書好き中群の子どもの国語成績：{73, 66, 78, ...}

親が読書好き低群の子どもの国語成績：{58, 62, 75, ...}

- 研究目的を達成するためには、何を独立変数, 何を従属変数に設定すべきかを, 明確にする
- 独立変数と従属変数を逆にしないこと
- 自尊感情の高さが共感性に影響するという研究なら, 独立変数は自尊感情(の高低), 従属変数は共感性
- 共感性の高群と低群の, 自尊感情の平均値を比較しないこと

標準偏差・標準誤差

- 標準偏差がデータの散らばりを表す論理
- 標準誤差の考え方
- 母集団と標本の意識化
- 交互作用の理解
- 統制変数, 調整変数, 媒介変数の理解

など

- Standard Deviation: SD

量的データの散らばりの大きさを表す指標の1つ

- 0以上の値を取る

- $SD=0$ のとき, データは散らばらず, 全て同じ値

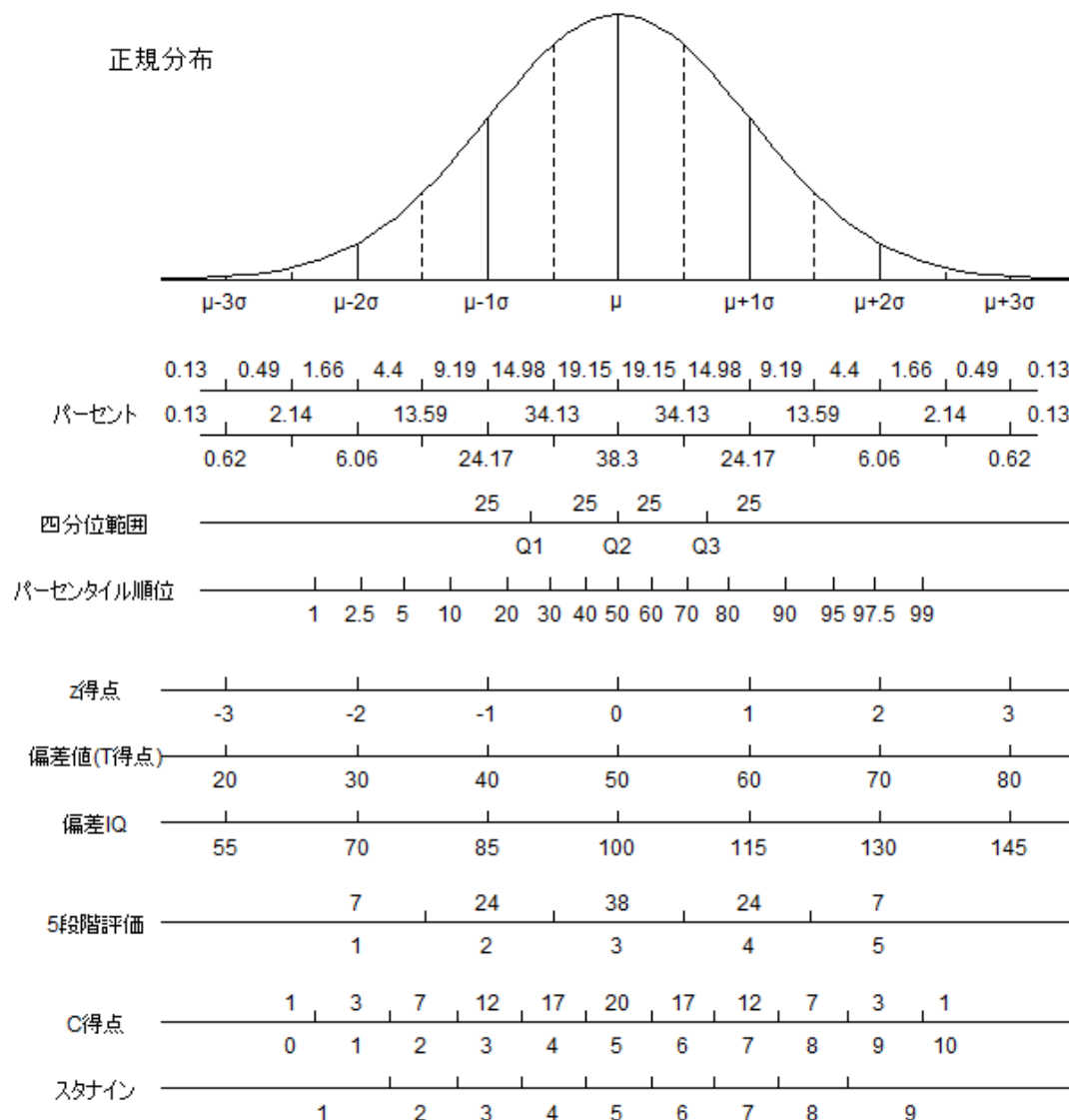
- 標準偏差の値は, データの測定単位に依存するので, 値がいくつ以上なら散らばりが大きいという基準はない

- データ分布が正規分布に近い場合,

平均 ± 1 SDの範囲に約7割

平均 ± 0.5 SDの範囲に約4割

のデータが入る



- 各データの平均からの偏差 ($X_i - \bar{X}$) は, 正に大きい, 負に大きい, 0に近いなど, いろいろある
- 偏差の平均を求めたら, データの散らばりの度合いが分かりそう
- しかし, 偏差の平均 (合計) は, 正負が相殺され, どんなデータでも必ず 0 になってしまう $\sum(X_i - \bar{X}) = 0$
- 偏差を2乗すればすべて0以上の値になるので, 正負が相殺されなくなる

- 偏差の2乗の平均を求めれば、データの散らばりの度合いが分かりそう

- それを求めたものが分散 $S^2 = \frac{\sum (X_i - \bar{X})^2}{N}$

- 2乗したのを元の単位に戻したものが標準偏差

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

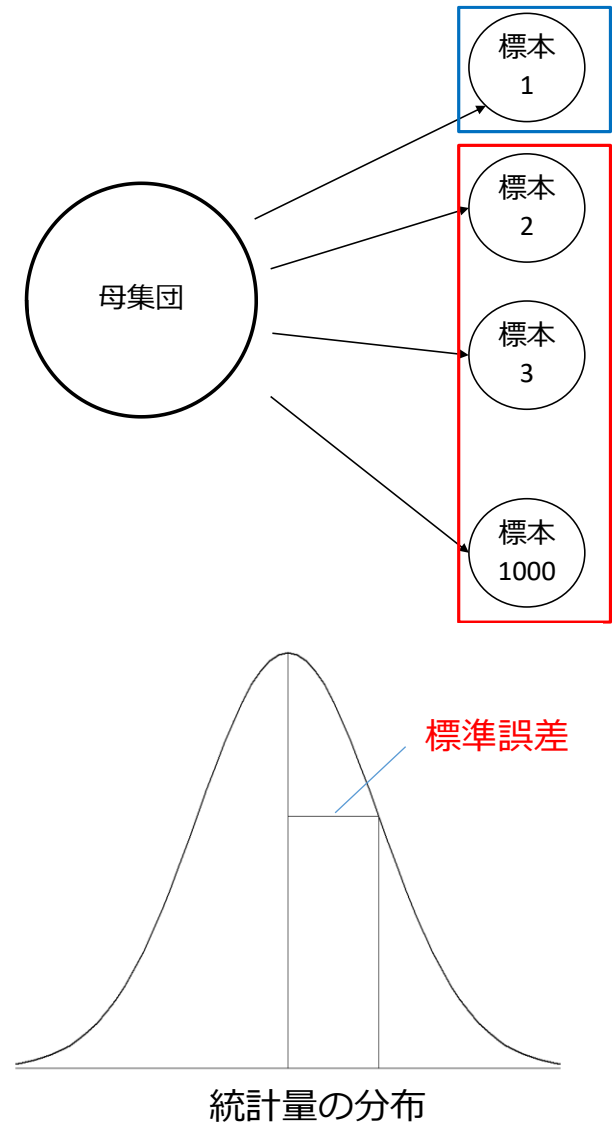
- 標準偏差は、標準的な偏差の大きさ

標準：2乗値の平均の平方根を取り、元の単位に戻したもの

- Standard Error: SE

- 標本平均や標本相関係数など, 何らかの統計量の, 母集団値からの標準的な変動の大きさを表す指標
- 統計量
標本から求められるある値, または, それを求める関数
- 標本が異なれば統計量の値は変わってくる
- 統計量と母集団値のズレ(標本誤差)も, 標本と共に変動する

- データを収集して統計量の値を求めることを何度も繰り返したと考える
(実際にはしない。思考実験)
- 統計量の変動を表す架空の分布をイメージすることができる(標本分布)
- 統計量の変動の標準的な大きさ
= 統計量の標本分布の標準偏差
= 標準誤差
- 標準誤差
何らかの統計量の標準的な変動(誤差)の大きさ



- 標準偏差は, データ分布におけるデータの散らばり
- 標準誤差は, 標本分布における統計量の散らばり

標本1		標本2		標本M		統計量の分布の	
ID	データ	ID	データ	ID	データ	平均	標準誤差
1	112	1	125	1	105	← 標本平均の分布	105.42 9.77
2	101	2	98	2	94		
...		...		...			
N	106	N	116	N	113		
平均	108.72	平均	120.51	平均	100.20		
SD	15.07	SD	15.84	SD	14.29	← 標本SDの分布	14.23 0.88

各標本の平均, 標準偏差

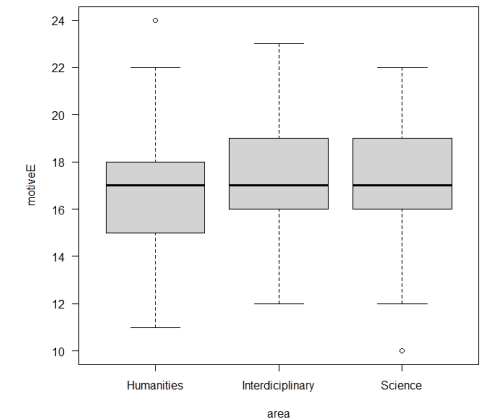
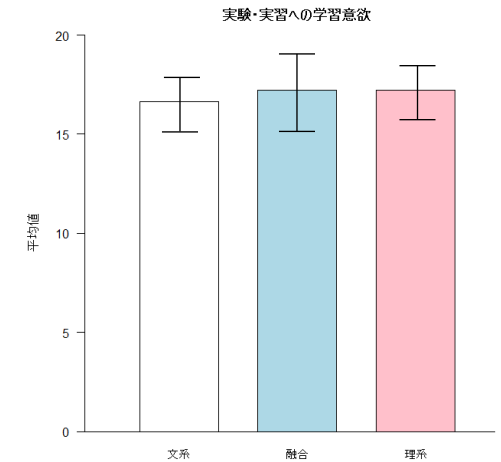
標本平均の平均: $\bar{\bar{X}}$

標本平均の標準誤差: $\frac{s}{\sqrt{N}}$

標本SDの平均: $\bar{s}$

標本SDの標準誤差: 省略 (難しい式)

- 平均値を比較する棒グラフを描くとき、エラーバーは標準偏差(SD)と標準誤差(SE)のどちらが良いか？
- SDはデータの散らばり
- SEは平均値の散らばり
- 平均値がどの程度の変動性を有しているかを表すSEのほうが良い
- データの散らばりを示すなら、ヒストグラムや箱ひげ図のほうが良い



統計的検定

- 度数, 割合, 比率, 平均値, 相関係数などが, ある値である, もしくは, 群間で差があると主張できるかどうかに関心がある
- 「ある値である」「差がある」などの判断を, 一定の合理性に基づいて行うために, 統計的検定を行う
- 実はいろいろ問題がある

対応のある2群の平均値の比較

- 日本の大学生における, 教養科目に対する学習意欲と専門科目に対する学習意欲を比較したい
- 237名の大学生からデータを収集
- 標本平均は, 教養16.72, 専門17.01で, 差がある
- 母集団 (日本の大学生) において差があるといえるか?

	N	M	SD	r
教養科目	237	16.72	2.12	.43
専門科目		17.01	2.40	

- 統計的検定を行う

- 帰無仮説 $H_0: \mu_1 = \mu_2$

母集団において, 教養科目と専門科目の学習意欲の平均値は等しい

- 対立仮説 $H_1: \mu_1 \neq \mu_2$

母集団において, 教養科目と専門科目の学習意欲の平均値は等しくない

- 差得点を考える

差得点 $d = \text{教養科目} - \text{専門科目}$

- 帰無仮説 $H_0: \mu_d = 0$

- 対立仮説 $H_1: \mu_d \neq 0$

- 検定統計量, p値の計算

$$SE = \frac{s_d}{\sqrt{N}} \rightarrow \frac{2.42}{\sqrt{237}} = 0.157$$

$$t = \frac{\bar{x}_d - \mu_d}{SE} \rightarrow \frac{-0.29 - 0}{0.157} = -1.85$$

$$df = N - 1 \rightarrow 236$$

	N	M	SD	r
教養科目	237	16.72	2.12	.43
専門科目		17.01	2.40	

	N	M	SD
差得点	237	-0.29	2.42

- $df = 236$ の t 分布において, $t = -1.85$ に対応する p 値は 0.065

$$p = 0.065 \geq 0.05 (\alpha)$$

- 帰無仮説を採択(保持)し, 対立仮説を棄却
- 母集団において, 教養科目に対する学習意欲と, 専門科目に対する学習意欲の平均値に, 差があるとは言えない
- 「母平均は等しい」と言うのは誤り

- 確率分布のかたちを決めるパラメタの1つ
df : degrees of freedom
- 統計分析では, 標本サイズ, 群数, カテゴリ数, 分析デザインなどによって, 自由度の値が決まる
- データの中身 (値) には影響されない
- 自由度は, 標本サイズや研究デザインと, 対応する確率分布をつなぐもの

t 値の意味すること

- $t = \frac{\overline{x_d} - \mu_d}{SE}$
- μ_d は帰無仮説で設定した母集団値
- SE (標準誤差) は, 統計量の, 母集団値からの標準的な変動の大きさ
- t 値は, 標準的な変動の大きさに対して, 手元のデータの変動(ズレ)はどの程度の規模かを表す

- t 値が過度に大きければ, そのような値は滅多に出ないので, そもそも帰無仮説が間違っていたと判断する
- t 値が基準 (限界値) を超えれば, 帰無仮説を棄却し, 対立仮説を採択する
- t 値が基準 (限界値) を超えなければ, 帰無仮説を採択 (保持) し, 対立仮説を棄却する
- 検定統計量 (t 値) が限界値を超えることを「統計的に有意」と言う

- 有意なのは「検定統計量」

平均値差に意味があるとは言っていない！

確かに、平均値差が大きければ t 値は大きくなるが、
標本サイズ(N)が大きくても t 値は大きくなる

- 有意なのは「統計的に」

実質的に意味があるとは言っていない！

t 値が限界値より大きくなるという数量的な関係を「統計的に有意」と言っているだけ

- 統計的有意性から「有意な平均値差」や「有意な相関」と言うことができるなら、まったく同じ論理で「有意な標本サイズ」と言うこともできるはず
- 「有意な標本サイズ」と言っている論文をみたことはない
- 研究者は「有意な平均値差」「有意な相関」など自分に都合のよい言い方だけを採用している

- 標本平均の標準誤差 $SE = \frac{\sigma}{\sqrt{N}}$

- $t = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}} = \frac{\bar{x} - \mu}{\sigma} \times \sqrt{N}$

- 検定統計量 = 効果量 × 標本サイズ

- 検定統計量は効果量, 標本サイズの増加関数
データをたくさん集めれば統計的に有意な結果,
データを集めなければ統計的に有意でない結果,
にすることができる

- 検定結果を「有意」にしたければ, 頑張ってデータを集めれば良い?
- 反対に, 平均値は等しいなど帰無仮説を採択したければ, データを集めなければ良い?
- 統計的検定は, 結果が標本サイズに依存するという, 重大な欠陥のある分析法
- 標本サイズの設計, 効果量の計算, 信頼区間の推定などが必要

標本サイズ

- 統計分析に耐える最小の標本サイズ：30程度
- 相関分析：150程度 ($r \geq$ 約0.3 で統計的に有意)
- 多変量データ解析, 構造方程式モデリング：300程度
- 探索的因子分析：項目数の10倍程度
- 尺度作成：300以上は欲しい

- 標本サイズについて経験的に言われている値は、分析結果が比較的安定するという観点から、経験的に導かれた値
- しかし、統計的検定の結果が標本サイズに依存することもあり、標本サイズは研究計画の段階で決めておくことが必要
- 標本サイズの設計には、検定力に基づく方法と、信頼区間に基づく方法がある

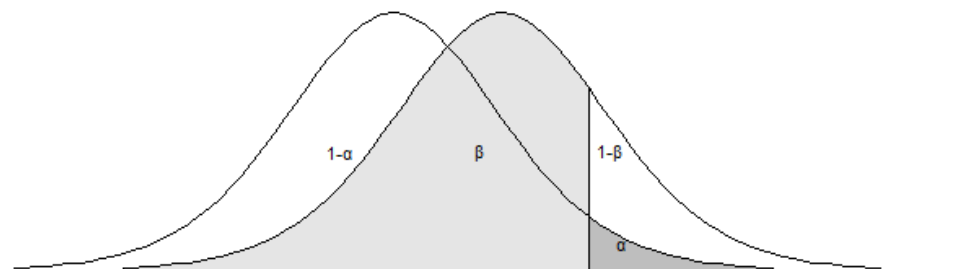
- 検定力・検出力

対立仮説が正しいときに、
正しくそれを採択する確率
($1-\beta$)

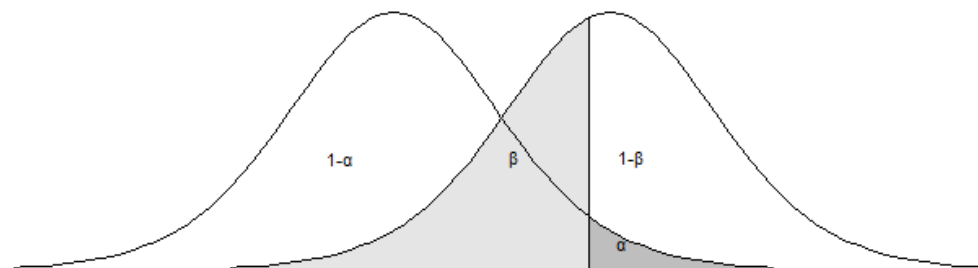
- 検定力は、効果量, 標本
サイズ, 有意水準の増加
関数

- どの程度の母効果量を,
どれくらいの検定力で検
出するかを考え, 必要な
標本サイズを推定

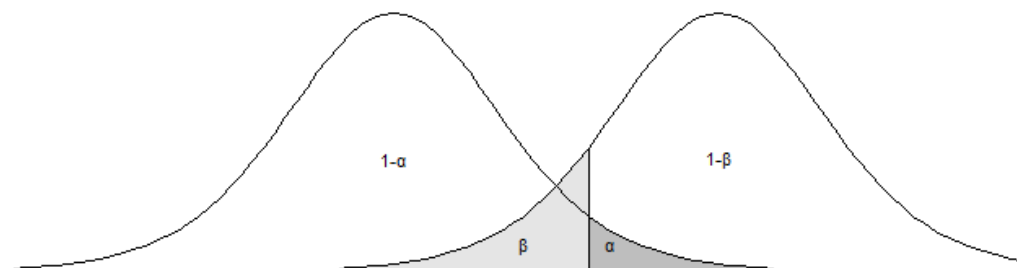
$\delta = \text{小}$, または $n = \text{小}$, $\alpha = 0.05$, $1-\beta = 0.3$)



$\delta = \text{中}$, または $n = \text{中}$, $\alpha = 0.05$, $1-\beta = 0.6$)



$\delta = \text{大}$, または $n = \text{大}$, $\alpha = 0.05$, $1-\beta = 0.8$)

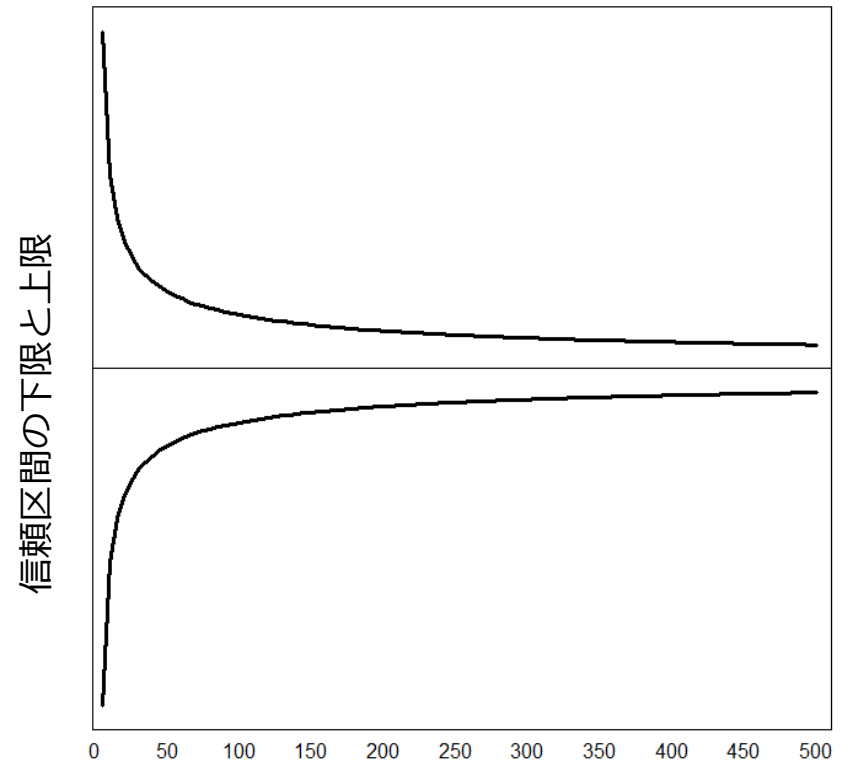


- 設定する条件： 効果量， 検定力， 有意水準
- よく用いられる設定
 - 有意水準 α : 0.05
 - 検定力 $1-\beta$: 0.8
 - 効果量 : 当該領域の先行研究等を参考に設定
- 統計解析ソフト (GPower, Rのpwrパッケージなど) を使って, 標本サイズを計算

- 有意水準 $\alpha = 0.05$, 検定力 $1 - \beta = 0.8$

分析	効果量		N (切り上げ)
対応のある2群の平均値	δ	0.2	199
	δ	0.5	34
対応のない2群の平均値	δ	0.2	394×2 群
	δ	0.5	64×2 群
相関係数	ρ	0.1	782
	ρ	0.3	85
3水準1B要因分散分析	η^2	0.01	319×3 群
	η^2	0.06	52×3 群

- 信頼区間の幅は、標準誤差、信頼係数などで決まる
- 平均値差の信頼区間
$$\left[\bar{x}_d - t_0 SE, \bar{x}_d + t_0 SE \right]$$
- 標本サイズが大きくなると、標準誤差が小さくなり、信頼区間の幅は狭くなる
- 信頼区間の幅は、標本サイズの減少関数



標本サイズ

- 信頼区間の幅をどの程度にしたいかを考え, 必要な標本サイズを推定
- 2群の平均値の場合, 信頼区間の半幅を標準偏差の何倍にするかと考えると, その倍率は効果量に等しい
- 半幅: 平均値に関する信頼区間 $[\bar{x} - t_0 SE, \bar{x} + t_0 SE]$ の場合 $t_0 SE$
- 設定する条件: 信頼区間の幅, 信頼係数

- 信頼係数=0.95 (95%信頼区間)

分析	半幅	N (切り上げ)	備考
対応のある2群の平均値	0.2 σ	195	対応のある変数間の 相関係数 $r=0.0$
	0.5 σ	34	
対応のない2群の平均値	0.2 σ	194×2	検定力 $1 - \beta \doteq 0.5$ に相当
	0.5 σ	32×2	
相関係数	0.1	320	$r=0.3$
	0.1	219	$r=0.5$

- 対応のない2群の平均値の標本サイズが, 検定力に基づいた場合と大きく異なっている

- 検定力に基づく方法 ($\delta=0.5$, $\alpha=0.05$, $1-\beta=0.8$)
→ $N=64 \times 2 \text{群} = 128 \text{名}$
- この標本サイズの場合, $d=0.35$ 以上で統計的に有意
- 検定力は 0.8 ではなく 0.5
- 検定力に基づく方法は, 想定した効果量より小さい効果量でも統計的に有意となり, そのときの検定力は0.8より小さい

- 信頼区間に基づく方法 (半幅= 0.5σ , 信頼係数=0.95)
→ $N=32 \times 2 \text{群} = 64 \text{名}$
- この標本サイズの場合, $d=0.5$ 以上で統計的に有意
- 検定力は 0.5
- 信頼区間に基づく方法は, 想定した効果量以上の効果が得られてはじめて統計的に有意となる

- 検定力に基づく方法は、標本サイズを多めに推定することがある
- 統計的に有意になったときの検定力が0.8あるとは限らない
- 標本サイズの設計の参考資料
村井・橋本 (2017). 「心理学のためのサンプルサイズ設計入門」 講談社

石井 「統計解析ソフトRのスクリプト集」(現在, Ver.15)

<https://www.educa.nagoya-u.ac.jp/~ishii-h/archives.html>

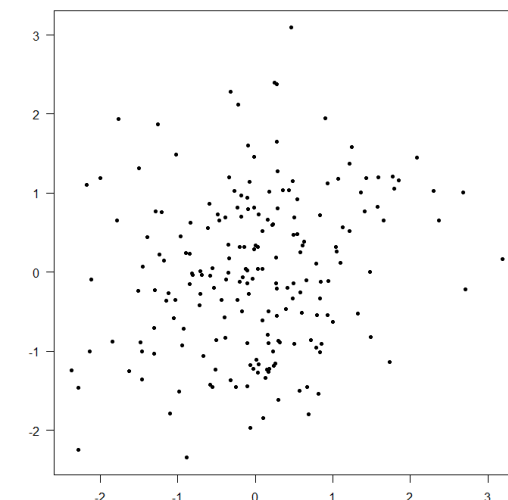
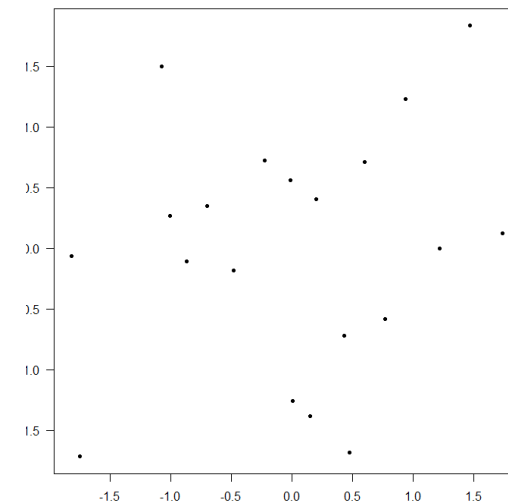
- 標本サイズが小さければ, 統計的に有意になりにくい
- 標本サイズが小さいときに統計的に有意になったら, 大きな効果があると主張できるか?
- 標本サイズが小さいときは, 標本変動の影響が大きい
- 結果の安定性が乏しいので, 強い主張はできない
- 直感的に考えて, 少ない情報(データ)から強い主張をすることはできない

- 予備調査と位置づけて、探索的な研究にする
有意水準を 0.05 ではなく 0.1 にする
有意水準を緩めて「統計的に有意」になる要因を拾い、
そこから焦点を絞って本調査を行う
- 事例研究にする
母集団に関する推測（検定）ではなく、標本（データ）の
記述に努める
- 質的研究法を用いて分析する

効果量

- 統計的検定の結果は標本サイズに依存する
- 標本サイズが大きくなれば p 値は小さくなるので、 p 値の小ささ(*の数)で評価するのは不適切
- 昨今の学術誌では、検定統計量の値や p 値だけでなく、信頼区間、や効果量の報告を求められることが多い
- 報告を求められなかったとしても、**標本サイズ, 平均値, 平均値差, 相関係数, それらの信頼区間や効果量等を確認し, 実際どの程度の効果があるかを検討すべき**

- 統計的に有意でも, 効果量が小さければ, 実質的な意味はない
- 相関係数はどちらの散布図も 0.2
人数: 上図 20, 下図 200
- 帰無仮説: 母相関係数=0
- p 値: 上図 .398, 下図 .005
(下図はいわゆる「有意な相関」)
- 「有意」だからと言って, $r = 0.2$ 程度で相関ありと主張するのは適切でない



相関係数の一般的解釈

絶対値	解釈
0.0～0.2	関連なし
0.2～0.3	ほぼ関連なし
0.3～0.4	弱い関連あり
0.4～0.6	中程度の関連あり
0.6～0.8	強い関連あり
0.8～1.0	非常に強い関連あり

- 経験的に言われているものであり, 客観的な基準ではない
- 領域によって基準が異なる場合もある

相関係数	例
0.3～0.4	性格検査と態度尺度得点の関連
0.5～0.6	国語と英語のテスト得点間の相関
0.5～0.7	地域別の平均世帯年収と児童の平均学力
0.8程度	PM2.5 (微小粒子状物質) 濃度とSPM (浮遊粒子状物質) 濃度の相関
0.9以上	イオンクロマトグラフィにおける試料中のイオン濃度と面積値

効果量の大きさの目安

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

分析	効果量の指標	小	中	大
t 検定	d, g, Δ	0.2	0.5	0.8
分散分析	η^2	0.01	0.06	0.14
多重比較	$r = \sqrt{\frac{t^2}{t^2 + df}}$	0.1	0.3	0.5
相関分析	r	0.1	0.3	0.5
クロス集計	V	0.1	0.3	0.5
回帰分析	R^2	0.02	0.13	0.26

その他

- 対応のあるデータ

同一個体に対して, ある変数を異なる複数の条件下で測定したデータ(の部分)

- 対応のないデータ

異なる複数の集団において, ある変数を測定したデータ(の部分)

- 可能なら対応のあるデータを収集したほうが良い

研究参加者の違いによる誤差を小さくできる

研究参加者が, 条件の違いを相対比較できる

順序効果等が生じないように注意

- 主効果または交互作用の有意水準を保ちながら, 水準間の平均値の比較を行う方法
- 1要因3水準のデータがあるとき, 水準間の平均値の比較は**総当たり**で3通りある 1と2 1と3 2と3
(2要因の場合も, 総当たりの組合わせを考える)
- それぞれの比較を $\alpha=0.05$ で普通に検定すると, 3つ全体での有意水準は0.143と大きくなってしまう
- これを0.05に抑えるようにして, 水準間の平均値の比較を行う

- 群間の平均値の比較の場合, Tukey法を用いるのが一般的
- Scheffe法, Bonferroni法, Sidak法も保守的なのでお勧め
- 各群の標準偏差が大きく異なるときは, Games-Howell法が頑健と言われている (Kirk, R.E. (2012) Experimental Design, SAGE)
- LSD, ライアン(REGW), SNKなどは, 有意水準の調整に問題があるので, お勧めしない
- 統計ソフトで多重比較を行う場合は, 調整後の p 値が表示されるので, 各比較の p 値が0.05を下回るか否かで判断すれば良い

- 欠測が多い研究参加者は, 有効回答者から除外する
- 欠測が少ない研究参加者は, 当該変数を使う分析のみ除外する
- 尺度得点の算出
 - 欠測が多い研究参加者の尺度得点は, 欠測とする
 - 欠測が少ない研究参加者の尺度得点は, 当該項目を除いた平均得点を使う
- 完全情報最尤推定法
 - 分布を仮定し, 観測されたデータをフルに利用して分析する方法
- 補完法
 - 平均値代入, 回帰予測, マッチング代入, 傾向スコア法 など