

量的研究デザインの方法
本田勝久(k-honda@faculty.chiba-u.jp)
千葉大学(教育学部)

0. 要旨

本発表では、具体的な研究テーマを例として、量的手法で研究を行う場合の研究デザインについて提示する。具体的には、(1)リサーチ・クエスションの設定、(2)データ収集の方法、(3)データ分析の方法、(4)研究の評価に関して、それぞれ提示していく。

昨年度と同様に、統計的検定における検出力 (power) に基づくサンプルサイズの設計方法について触れる。検出力やサンプルサイズの設計は、統計的検定を行うにあたって基本的な概念であり、重要である。英語教育における量的手法では、自由度 ($n-1$) の t 分布の両側 5%点より大きければ「有意差あり」と判定し、「帰無仮説を棄却して対立仮説が成り立っている」と判断することがある。しかし、検出力とサンプルサイズの関係を検討すれば、実質的な意味のある差について考察することができる。

- ・対立仮説が成り立っている→サンプルサイズ(大)、検出力(高)
- ・帰無仮説が成り立っている→サンプルサイズ(小)、検出力(低)

さらに本年度は、APA Publication Manual 第 5 版 (2001) 以降に報告が必要とされている効果量 (effect size) について解説する。いくつかある効果量の指標の中で、どの分析でどの指標を選ぶのかは諸説あるが、効果量の基本的な考え方を理解し、適切な指標を選択することは、今後の量的研究において不可欠である。

本発表では、正規分布、 t 分布、 χ^2 分布、 F 分布の分布関数を計算できる (例えば、表計算ソフト Excel などを使って) ことを想定している。統計的検定におけるサンプルサイズの設計では、検出力の考え方と計算方法が密接に関連している。検定力、サンプルサイズ、効果量などの用語を整理し、それぞれの関係を検討することによって、量的研究デザインの方法について議論したい。

1. 理論的視点

研究デザインを行う際に、研究者が立つ理論的視点・枠組み(paradigm)は大きく3つに分類される。

① 実証主義的視点 (positivist/scientific paradigm)

人々の行為には普遍的な法則・規則がある。客観性、予測、反復可能性を重視し、因果関係を説明するための調査を目的とする。科学的で実験主義的な調査を通して得られた知識は客観的で定量的であり、現実世界(reality)は静的で観察・測定可能なものである。

② 解釈主義的視点 (interpretive paradigm)

人々の経験は文脈に縛られており、時、場所、人間という行為者の心と切り離すことはできない。多元的な現実世界(reality)は人々により社会的に構築されている。参加者の行為や経験、教育のプロセスの意味を探究し理解することを目的とする。

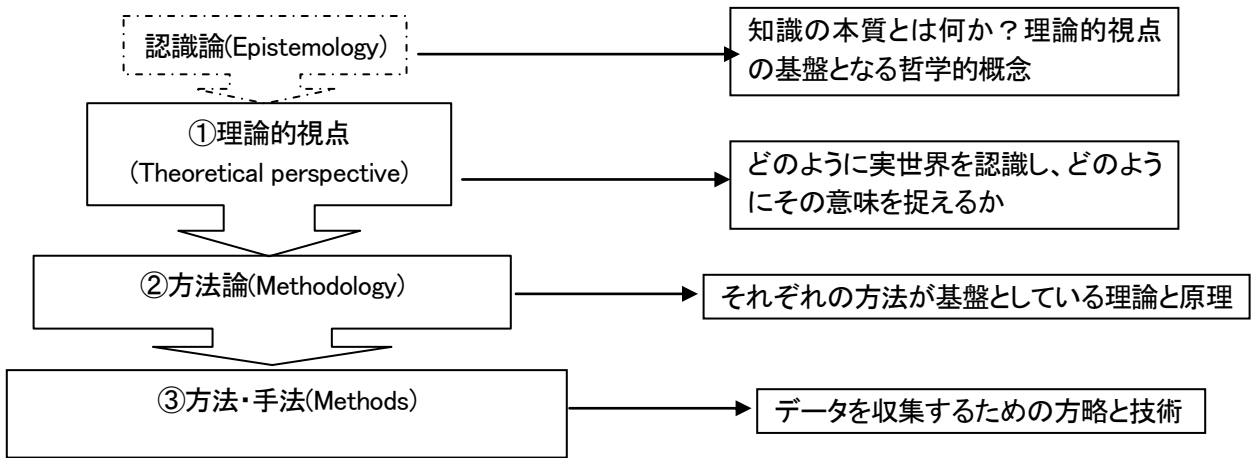
③ 批判的視点 (critical paradigm)

客観的や中立的な知識は存在せず、知識は常に社会的利害に影響されている。人間の行動を解釈し、理解するだけでなく、社会的批判をすることで、社会的あるいは組織的な変化を起こすことを目的とする。

2. 研究計画

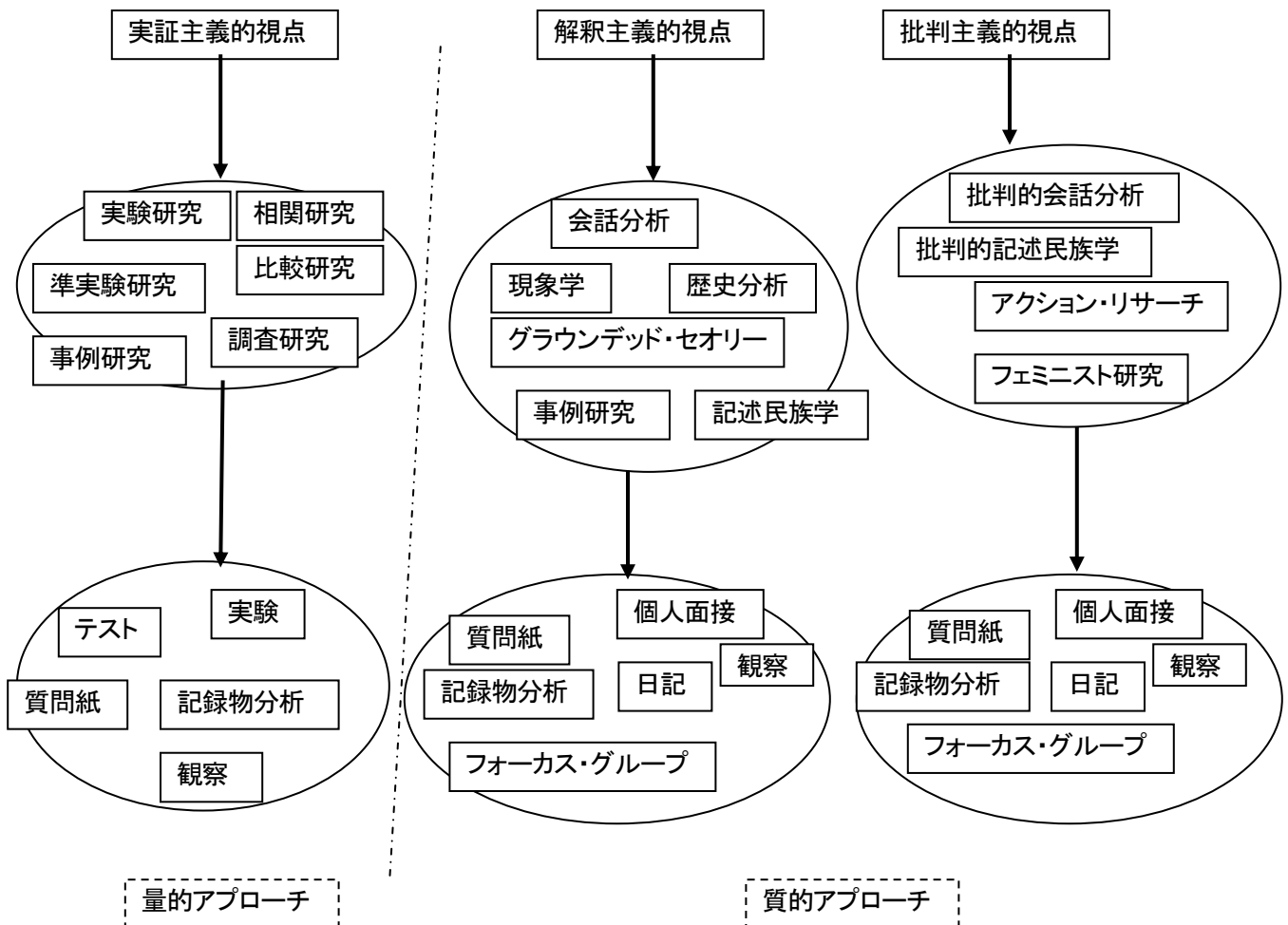
2.1 研究計画に関わる4要素

研究計画を立てるときに以下の4つの要素について認識しておく必要がある。



Crotty (1998, p.4)より

2.2. 量的アプローチと質的アプローチ



3. 研究課題の設定

3.1 一般から特定へ

教師としての経験や関心(あるいは勘)⇒日々の教育実践の中で抱く疑問や課題から



測定される特定の変数の用語によって具体的に言い表されたものへ

- ・研究論文は、研究課題が興味を引くに足ることを立証しなければならない。
- ・研究課題は、その研究が適切かつ重要であり、新たな寄与をおこなうものであることを示さなければならない。

「目的」はその研究の目標についての一般的な表現

「仮説」はその実験で起こると思われることについてのより特定化した予想(量的)

(一般)学生が好んで選択する心理学のコースは、教える講師のタイプと関係があるか

(特定)個人的・社会的なスキルを強調する講師は、臨床のコースでより好まれるか

分析的・客観的なスキルを強調する講師は、実験的なコースでより好まれるか

(Findlay 1993, p. 17改訂)

3.2. 先行研究から

・関連文献を探す(雑誌論文, 電子ジャーナル, 人と人のコンタクト etc.)

学習者要因に関する多くの研究論文は、*The Modern Language Journal*や*Language Learning*に代表されるSLAの主なジャーナルにも掲載されている。

・関連文献をまとめる(文献目録, コンピュータのデータベース etc.)

「対象者の属性」「何がなされたか」「何がわかったか」「どのような結論になったか」などの関連文献ごとのリストを作成する。

・用語を整理する(検索エンジン, 用語辞典 etc.)

questionnaires, inventories, forms, tests, batteries, checklists, scales, surveys, schedules, studies, profiles, index/indicators, sheets, etc.

先行研究の追試や先行研究への疑問から研究課題が見つかることも...

3.3 自分なりの研究課題の設定へ

(1)位置づけができること→先行研究から自分の研究はどのような位置づけになるのか?

(2)オリジナリティがあること→先行研究との違いは何か?単なる追試になっていないか?

(3)検証可能であること→無理な課題設定はしていないか?変数の統制や条件は整備できるか?(量的)

先行研究では

(1)何が主張されていて..

⇒

(2)何がわかっている..

研究課題では

(1)何を知らなければならないのか..

⇒

(2)何をすることができるのか..

研究方法の決定へ

では、

⇒ どのようにしたら、

知ることができるか?

4. 量的アプローチによるリサーチ・デザイン: 調査研究

4.1 調査研究とは？

実験研究 (experimental study)

特定の事象・現象を分析する目的のために、人為的に整えられた条件(例: 実験(experimental group)・統制群 (control group))のもとで、特定の道具や装置を用いた実験を行い、そこで得られる実験結果を分析するもの。

調査研究 (survey)

観察、検査、アンケート、面接などを行いそれによって得られた資料を分析するもので、人為的な条件は設定しない。

白畑知彦・富田祐一・村野井仁・若林茂則 (1999, p.261)

①観察によるもの⇒「見ること」により学習者を理解しようとするもの

・観察法: 学習者の行動を観察・記録・分析し、行動の質的・量的特徴や行動の法則性を解明すること

②言語を媒介とするもの⇒「聞くこと」により学習者を理解しようとするもの

・質問紙法&面接法: 行動そのものよりも学習者の感情や価値観、動機など、心の内面を理解すること

4.2 研究手法の決定

リサーチ・デザインを決定する際、選択する研究手法によりプロセスが異なることに留意する。

例えば...質問紙法では、

①測定対象を明確にする

↓

②項目の候補を収集する

- * 自分で考える
- * 人に尋ねる
- * 関連文献にあたる

↓* 項目の分類, 絞込み

③予備データを収集する

↓

④項目を決定する

- * 反応分布の検討
- * G-P分析, I-T相関分析

↓* 因子分析

⑤本調査を行う

↓

⑥信頼性の検討

- * 再検査法
- * 折半法

↓* α 係数

⑦妥当性の検討

- * 基準関連妥当性
- * 構成概念妥当性

例えば...観察法では、

①対象行動の適切性の検討

- * 予備観察

↓* 継続時間, 観察単位

②観察対象者の決定

↓

③行動の低位カテゴリーとその操作的定義の決定

↓

④適切な時間間隔と観察回数の決定

↓

⑤記録用紙の設計

- * 自由記述法
- * サンプリング法 (1/0, ポイント)
- ↓* サインシステム, カテゴリーシステム

⑥観察の実行

↓

⑦一致度の算出

- * 信頼性の検討
- ↓* カッパ係数

⑧データ処理

これらのプロセスを視野に入れて
リサーチ・デザインを決定するのが望ましい...

(1)調査対象を定義する→調査目的や方法に応じて調査対象の範囲を想定し、直接の調査対象者(サンプリング)と母集団との関係を決定すること

(2)調査内容を吟味する→調査内容の候補を収集・精選し、予備調査や分析(信頼性&妥当性)などを通して実際の調査項目を作成すること

(3)調査方法の種類を知る→いろいろな調査方法の短所と長所を知り、研究課題や時間・労力・資金などに合わせて調査方法を選択すること(面接調査, 集団調査, 委託調査, 郵送調査 etc.)

(4)分析手法やデータ処理を予測する→調査対象者の数やその構成(性別・年齢)、調査項目への回答の分布、欠損値の有無などのデータ内容の記述統計的な概略から、データの集計と統計的分析手法を予測すること(質的・量的データ, ノンパラメトリック検定 etc.)

(5)倫理の問題を考慮する→調査対象者とのあいだに信頼関係を築き、人間尊重の精神で調査を実施すること(説明と同意, プライバシーの保護, 結果報告, データの保存または破棄 etc.)

「はじめに研究手法ありき」「ただやっているだけ」ではなく、まずは自分の研究課題をじっくり明らかにすることから、「よい良い研究手法」を決定して欲しい...

4.3 データ処理

4.3.1 統計を用いたデータ処理の5つの条件(なぜデータ処理をするのか?)

Systematic Research: A statistical study has a clear structure with definite procedural rules that must be followed.

Logical Research: The rules and procedures underlying these studies form a straightforward, logical pattern—a step-by-step progression of building blocks, each of which is necessary for the logic to succeed.

Tangible Research: Statistical research is tangible in that it is based on the collection and manipulation of data from the real world.

Replicable Research: Statistical research should be also replicable. The researcher's proper presentation and explanation of the system, logical data collection, and data manipulation in a study should make it possible for the reader to replicate the study (do it again under the same conditions).

Reductive Research: Statistical research can reduce the confusion of facts that language and language teaching frequently present, sometimes on a daily basis.

(Brown 1988, pp. 4-5 参照)

4.3.2 記述統計と推測統計(統計における基礎的な述語とは?)

変数, 構成概念, 独立変数, 従属変数, 母集団, 標本, 検定, 統計的有意, 危険率, 水準 etc.

→記述統計(descriptive statistics): 標本のcentral tendencyとdispersionを調べること
平均, 分散, 標準偏差, 尖度, 歪度, z得点/偏差値, 中央値, 最頻値, パーセンタイル etc.

→推測統計(inferential statistics): 標本のnormal distributionから母集団を推測すること
正規分布, 等分散検定, 5%水準, 片側&両側検定, 対応ありと対応なし, 繰り返し要因 etc.

- (1)パラメトリック検定:母集団の数理的特性を推定し、母集団の分布型についての仮定を含める検定(母集団が正規であることを前提とする)→t検定, 分散分析, ピアソンの相関係数 etc.
 (2)ノンパラメトリック検定:標本の統計的特性から、母集団の分布が特定の形であることを前提としない検定(分布にかかわらず統計法)→ χ^2 検定, マン-ホイットニーの検定 etc.

4.3.3 相関研究とカテゴリカルデータ分析

カテゴリカルデータとは、相互に排他的なカテゴリーに分類されたデータで、数量としての意味をもたない数字や、記号によって示される名義尺度上のデータをいう。

①数量として平均値や標準偏差を求めない ②母集団について特別な仮定をもたない
 ↓でも...
 そのカテゴリーが量的に変化に沿って設定されていて、量的な意味を完全に失っていない場合には、量的なデータとしてそれに対応する分析を行うこともある。
 ↓重要なのは...
尺度 (scale) が決まれば、検定 (test) が決まる
 4つの尺度の使い分け、差の有無、関係の有無

例えば...関係の検討(分析タイプII)では、
 名義尺度→ χ^2 検定, 順序尺度→順位相関係数, 間隔&比尺度→ピアソンの積率相関係数 etc.

		大小	加算
名義尺度 (nominal scale)	性別, 職種, 所属クラブ, 好きな食べ物 etc.	×	×
順序尺度 (ordinal scale)	満足度, 階級, コンテスト順位(1<2<3) etc.	○	△
間隔尺度 (interval scale)	身長, 体重, 年齢, 金額, 試験成績, 等間隔に点数	○	○
比尺度 (ratio scale)	化された回答 etc.	(順序)	(平均)

4.3.4 データに対応する分析方法

χ^2 検定, t検定, 分散分析, マン・ホイットニーの検定, クラスカル・ウォリスの検定, ピアソンの積率相関係数, 順位相関係数, 重回帰分析などたくさんあるが...

分析によっては特別な知識や訓練が必要になる。「とりあえずやる」ではなく、あくまでもデータに合った処理方法を選択する必要がある。

◎分析タイプ I

・項目のデータの特徴を記述する。

- (1)度数分布を書く。
- (2)代表値(平均値、中央値、最頻値)を求める
- (3)散布度(標準偏差、四分位偏差)を求める。

◎分析タイプ II ~ VI

(1)関係 or 差の有無 (2)尺度の違い ⇒ 分析方法の決定

- ・分析タイプII...関係・差→例えば、対応ありのt検定、ピアソンの積率相関係数
- ・分析タイプIII...差(2つの場合)→例えば、対応なしのt検定、マン・ホイットニーの検定
- ・分析タイプIV...差(3つ以上)→例えば、1元配置の分散分析、クラスカル・ウォリスの検定
- ・分析タイプV...差(2要因: 交互作用)→例えば、2元配置の分散分析、逆正弦変換法
- ・分析タイプVI...関係・差(多変量)→例えば、重回帰分析、因子分析、フリードマンの検定

4.4 例題

これまでのセミナーでは、 χ^2 検定、ウイルクソンの順位和検定、重回帰分析、スピアマンの順位相関係数、対応ありのt検定、1元配置の分散分析と異なった検定を取り上げてきた。本セミナーでは、昨年度の分析手法を取り上げ、分析方法の違いについて解説する。これまでの分析結果および本年度の発表資料などは、浦野 研(北海学園大学)先生のHPからダウンロードできるので、ご参照頂きたい。

<http://www.urano-ken.com/research/seminar/>

4.4.1 対応ありのt検定 (dependent t-test)

mouth movement の提示は学習者の音素の識別に有効であると考えている N 先生は、「音声だけを用いた学習よりも音声と映像を併用した学習の方が音素の識別には有効である。」ということを実証するために、以下のような実験を試みた。まず、被験者である大学生 40 名を TOEIC (listening section) のスコアによって均等になるような2つのグループ(音声だけを用いた学習をするグループ: G1)と、音声と映像を併用した学習をするグループ: G2)に分類した。さらに、弁別が困難であるといわれる /æ/, /ʌ/, /ɑ:/, /ɔ:/ の4母音を取り上げ、それぞれの音素を識別する pre-test を実施した。pre-test は音声テープにより 1 問 30 秒(15 秒×2 回)で計 40 問であった。両グループの違いが「音声」と「音声+映像(ビデオテープ)」になるように、それ以外の変数はできるだけ統制するような授業を4週間にわたって行った。最後に、post-test として pre-test と同じテストを実施し、以下のようなデータが得られた。

データ表

(G1)				(G2)			
No	Listening	Pre-test	Post-test	No	Listening	Pre-test	Post-test
1	155	25	22	1	80	16	19
2	185	28	38	2	155	14	17
3	125	18	10	3	240	45	43
4	150	28	29	4	125	18	25
5	180	37	33	5	145	29	36
6	115	9	19	6	140	19	16
7	95	23	20	7	195	29	33
8	165	37	44	8	165	25	31
9	120	16	18	9	25	12	15
10	185	41	40	10	125	27	29
11	95	7	10	11	205	21	33
12	145	10	8	12	150	16	13
13	175	12	15	13	55	10	15
14	85	17	13	14	145	15	26
15	180	33	35	15	145	21	39
16	95	18	19	16	250	27	44
17	155	41	38	17	120	33	45
18	115	10	12	18	120	18	24
19	135	16	14	19	145	27	37
20	205	29	21	20	110	19	30

どのような統計処理をしたらよいのでしょうか？また、N先生の考えは支持されたのでしょうか？

対応なし／対応ありのt検定 (independent/dependent t-test)

(1) t検定とはどのような時に用いるのだろうか

2つの標本があって、それからそのそれぞれの母集団を推測して、2つの平均値の間に統計的に有意な差があるかどうかを確かめる時に用いる。

(2) t検定を行うにあたって必要な条件とは何だろうか。

- ・2つの標本が正規分布(normal distribution)であること。
- ・2つの標本がほぼ等分散であること。

(3) 片側検定(one-tailed)&両側検定(two-tailed)とは何だろう。

片側検定→どちらかの標本(母集団)が理論的に明らかに平均値が高いと分かっている時に用いる。例えば、中学生と大学生の語彙力測定テストの点数を比べる時など。危険率が両側の半分になる。
両側検定→どちらかの標本(母集団)が平均値が高いかは、理論的には分からない時に用いる。

4.4.2 1元配置の分散分析 (one-way ANOVA)

「リーディングテキストにおける未知語の推測は、学習者の語彙のリテンションに効果的である」と考えているH先生は、4週間にわたって次のような調査を行った。まず、Nation (1990) のVocabulary Levels Test (2,000-3,000 word level) を30人の大学生に実施し、それぞれ10人からなる語彙力において等質な3グループを作成した。それら3つのグループに異なるインストラクションの授業 (Group1: vocabulary guessing, Group2: vocabulary glossing, Group3: accessing dictionaries) を行い、vocabulary test を4週間後に実施した。このvocabulary testは、Flesch grade level 14のリーディングテキストから、パイロットスタディとして学習者が未知語として選出した29語の中から16語をターゲットとして作成されたものである (correct=1, incorrect=-0.33, I don't know=0)。それぞれのグループにおけるvocabulary testの結果は以下の表のとおりであった。はたして、H先生の考えは支持されたのであろうか。

Group 1		Group 2		Group 3	
Nation's Vocabulary Test	Vocabulary Test (4週間後)	Nation's Vocabulary Test	Vocabulary Test (4週間後)	Nation's Vocabulary Test	Vocabulary Test (4週間後)
25	43.6	24	27.4	24	18.3
28	56.8	23	38.9	36	21.7
31	27.3	31	59.4	30	29.5
19	35.0	42	43.2	26	15.6
40	48.4	18	15.9	18	9.7
32	42.4	15	22.2	14	16.0
23	25.3	24	52.4	20	7.5
19	51.7	27	28.9	42	31.7
20	31.4	21	30.6	37	13.4
22	29.1	28	25.7	31	22.9

※上記の研究・デザインやデータはあくまで架空のものです。

1元配置の分散分析 (one-way ANOVA)

(1) 1元配置の分散分析はどのような時に用いるのだろうか。

3つ以上の標本があって、それからそのそれぞれの母集団を推測して、3つの平均値の間に統計的に有意な差があるかどうかを確かめる時に用いる。

(2) 多重比較とは何か。

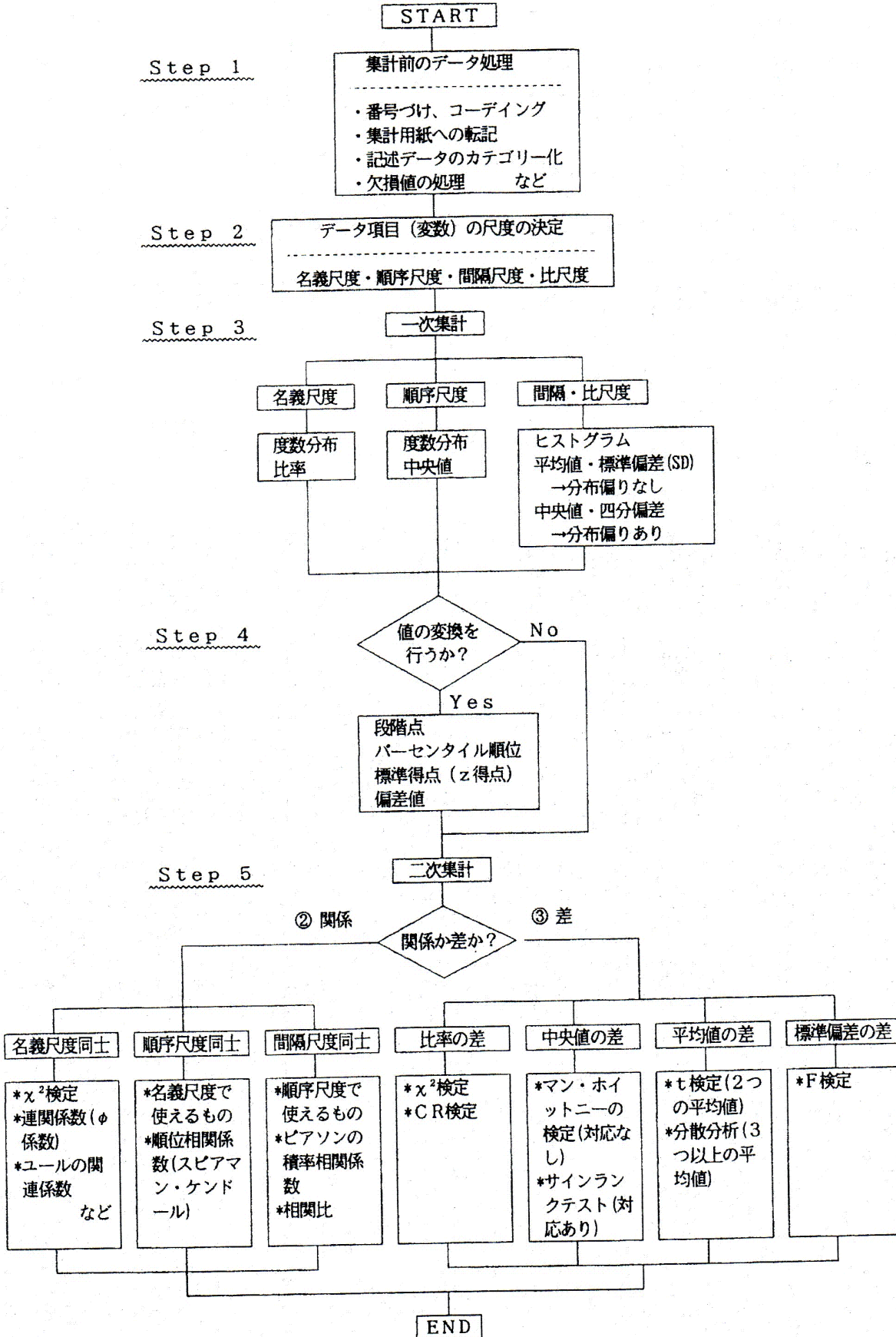
1元配置の分散分析の結果「標本間に差はない」が棄却された時に行う手法(必ずしもはじめに分散分析をする必要はない)。具体的に、どことどの標本間で差があるのかを調べてくれる。

(3) 3つの標本を比較する場合、t検定を繰り返せば(すなわち、AとB・BとC・CとAの平均を比較する)いいはずである。どうしてダメなのか？

複数回の比較をセットとして見立てた場合 (familywise)、1回またはそれ以上第1種の誤りを犯す確率は回数が多いほど高くなる。

$$\alpha \text{ FW} = 1 - (1 - \alpha)^m \text{ (3回} \rightarrow (1 - 0.05)(1 - 0.05)(1 - 0.05) = 0.86)$$

4.5 データ処理の手順



5. 研究の評価

研究の評価を行う際には、量的アプローチに適した判断基準で、研究の評価を行う必要がある。

研究の判断基準

- ・ 内的妥当性(internal validity)

測定したい対象が性格に測定できているか。データから結果が導けるか。

→独立変数(原因となる変数)と従属変数(結果となる変数)の間の因果関係について、その因果関係が存在するという記述(因果検証)

- ・ 外的妥当性(external validity)

結果がどれだけ一般化できるか。条件を変えて研究をしても同じ結果が得られるか。

→研究結果の一般化可能性(研究の結果がどの程度一般的なものであるか)を確認。異なる被験者や違う状況設定での追試実験(一度きりの結果を一般化していないか)

- ・ 信頼性(reliability)

結果に一貫性があるか。同じ調査方法で、同じような結果が得られるか。

→同じ測定を同じ被験者や同じ状況で実施した時、何度やっても同じ結果になるかという検査法の一貫性。信頼性の測定法として再検査法(ある検査を日時を置いて2度施行した場合の相関)。信頼性の指標として信頼性係数(coefficient of reliability: 例 Cronbach α)。内的整合性の検証としての α 係数。

- ・ 客観性(objectivity)

中立的で偏りがないか。

→あらゆる被験者や場面に適応する測定。採点者の個人的判断が採点に影響していないこと。客観的検査として、真偽法、多肢選択法、組み合わせ法、完成法。

6. サンプルサイズ的设计

6.1 サンプルサイズの重要性

検定では、限定的な形で帰無仮説 H_0 を設定し、それが棄却されたときの受け皿として対立仮説 H_1 を設定する。 H_0 を棄却できるとき「有意である(significant)」と述べることができる。

・ 本当は帰無仮説 H_0 が成り立っているのに、 H_0 を棄却する誤り→第1種の誤り(確率 α)

・ 本当は帰無仮説 H_0 が成り立っていないのに、 H_0 を棄却しない誤り→第2種の誤り(確率 β)

検定における2種類の誤り

		本当に成り立っているのは...	
		帰無仮説 H_0	対立仮説 H_1
検定結果	H_0	正しい (その確率: $1-\alpha$)	第2種の誤り (その確率: β)
	H_1	第1種の誤り (その確率: α)	正しい (その確率: $1-\beta$ =検出力)

永田 靖『サンプルサイズの決め方』朝倉書店 (2003)

検定における2種類の誤りに関する基本的事項

- (1) 帰無仮説は限定的なので、第1種の誤りの確率(有意水準) α は1つに定まる。
- (2) 対立仮説は複合的なので、第2種の誤りの確率 β はパラメータの値が異なると変化する。
- (3) α は小さく設定できたとしても、 β は非常に大きな値になりうる。
- (4) α を大きくすると β は小さくなる。
- (5) サンプルサイズが大きくなると β は小さくなる。

では、どのようにデータをとればよいのか、どのくらい(サンプルサイズ)データをとればよいのか.....

6.2. 検定力とサンプルサイズ的设计

本当は対立仮説 H_1 が成り立っているときに、 H_0 を正しく棄却する確率を検出力または検定力と呼ぶ(英語では power) ⇒ 第 2 種の誤りを犯さない確率 ($1-\beta$)

※計算上は、第 1 種の誤りの確率 (α) も検出力 ($1-\beta$) も、どちらも H_0 を棄却する確率になる。

・検定では第 1 種の誤りの確率である有意水準 α を小さく設定できても、第 2 種の誤りの確率 β が大きくなるかもしれない。

⇒ H_0 が棄却できるときには H_1 の内容を結論として断定できるが、 H_0 を棄却できないときには H_0 の内容を結論として断定はできない。

・では、 H_0 の内容を断定的に述べたい場合はどうするか…

⇒ β の値が小さいことを示す。つまり、検出力 $1-\beta$ が大きいことを示す必要がある。

H_0 とほとんど変わらない H_1 の状態なら (p が 0.5 に非常に近い場合)、 β の値を小さくするためには非常に大きなサンプルサイズが必要になる。しかし、コストなどで通常はサンプルサイズに制限がある。⇒ サンプルサイズが大きくないときに帰無仮説が棄却されるのなら、意味のある差(有意差)があると判定する。

6.3 検出力の計算方法

サンプルサイズを決定する方法 ⇒ 検定力分析 (power analysis)

(1) 実験を実施した後に、サンプルサイズ・効果量・有意水準 (α) から、検定力 ($1-\beta$) を確認する。

(2) 実験を実施する前に、これまでの先行研究からわかっている(推測される)効果量、有意水準 (α)、目指している検定力 ($1-\beta$) からサンプルサイズを決定する。

※検定力分析は、フリーソフト G*Power を使って簡単に確認することができる。

<http://www.nuis.ac.jp/pda-j/doc/00008/gpowermanual.html>

統計 web/統計 tips を活用 ⇒ 必要なサンプルサイズの計算

母比率の区間推定(無限母集団)の場合

http://software.ssri.co.jp/statweb2/cgi-bin/tips_1.cgi

無限母集団から抽出したサンプル中である条件を満たす対象の比率(サンプル比率)が母集団のそれと近い値であって欲しい場合、サンプル比率がある程度の誤差範囲内で収まるために必要なサンプルサイズを、誤差、信頼度、母比率を指定することで求めることができる。

母平均の区間推定(母分散未知)の場合

http://software.ssri.co.jp/statweb2/cgi-bin/tips_2.cgi

これから集めるサンプルにおける何らかの測定値の平均値が、母集団のそれとあまり変わらないようにしたい場合、サンプルの平均値がある程度の誤差範囲内で収まるために必要なサンプルサイズを誤差、信頼度、標準偏差を指定することで求めることができる。

7. 効果量(effect size)

APA Publication Manual 第 5 版 では「効果量(effect size)」の報告が必要であると記載されている。

For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section. (p. 25)

これまでの統計的検定で見ているものは、「効果があるかどうか」であり、「効果の大きさ」は見えていない。サンプルサイズに影響されない効果の大きさを示す必要がある。⇒効果量

水本・竹内（2008）には、効果量の指標と大きさの目安がまとめられている。また、効果量を計算する Excel による計算シートも掲載されている。

<http://www.izumoto.com/stats/effectsize.xls>

表 1

検定(分析)の種類ごとに見る代表的な効果量の指標と大きさの目安

使用される検定(分析)	対象と注意	効果量の指標	効果量の目安			
			小 (Small)	中 (Medium)	大 (Large)	
(1) 相関分析		r	.10	.30	.50	
(2) 重回帰分析		R^2	.02	.13	.26	
		f^2	.02	.15	.35	
(3) t 検定 (t -test)	対応あり・なしとも同じ	r	.10	.30	.50	
		d	.20	.50	.80	
		η^2	.01	.06	.14	
(4) 一元配置分散分析 (One-way ANOVA)	全体の差の検定	partial η^2	-	-	-	
		ω^2	.01	.09	.25	
		f	.10	.25	.40	
	多重比較	r	.10	.30	.50	
		η^2	.01	.06	.14	
(5) 二元配置分散分析 (Two-way ANOVA)	主効果	partial η^2	-	-	-	
		ω^2	.01	.09	.25	
多元配置分散分析* (Multi-way ANOVA) *三元配置以上の分散分析	交互作用	η^2	.01	.06	.14	
		partial η^2	-	-	-	
		ω^2	.01	.09	.25	
	多重比較	r	.10	.30	.50	
(6) 共分散分析 (ANCOVA)	共変量の影響を取り除いて分析し、主効果、交互作用、多重比較の効果量は (4) や (5) と同じ					
(7) 多変量分散分析 (MANOVA)	多変量検定	multivariate η^2 (multivariate R^2)	-	-	-	
		multivariate partial η^2	-	-	-	
多変量共分散分析 (MANCOVA)	従属変数ごとの分散分析	主効果、交互作用、多重比較の効果量は (4) や (5) と同じ				
(8) カイ2乗検定 (χ^2 test)	2×2の分割表	$\phi (= W)$.10	.30	.50	
	2×2以外	Cramer's V	.10	.30	.50	
(9)	マン・ホイットニーの U 検定 ウィルコクソンの符号順位和検定 クラスカル・ウォリスの順位和検定 フリードマン検定	検定統計量を Z に変換して r を求める	r	.10	.30	.50

Note. Cohen (1998; 1992), Field (2005), Tabachnick and Fidell (2006) などを基に作成。効果量の大きさはあくまで目安であるので研究分野によって変わる。(3) d , (4) f , (8) W についての詳細は, Cohen (1988) を参照のこと。 η^2 の大きさの目安は文献によっては, r を2乗した r^2 に合わせて, $\eta^2 = .01$ (効果量小), $\eta^2 = .09$ (効果量中), $\eta^2 = .25$ (効果量大)としているものもある。また, partial η^2 の効果の大きさの基準は明確なものがない。multivariate η^2 と multivariate partial η^2 の値は従属変数(dependent variable)の数によって変わるため, 効果量の目安は Cohen (1998) を参照。

6. 推薦文献

英語教師のための研究法の必読書

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston: Heinle & Heinle.

Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.

_____ (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. London: Lawrence Erlbaum Associates.

Press. Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.

英語教師のための研究法の必読書として上記の5冊を推薦したい。リサーチ・クエスチョンの設定から変数の扱い、名義尺度やノンパラメトリック検定を含む統計手法の意義などが解説されている。言語テストや動機づけ研究の第一人者からのメッセージは、傾聴に値する。

最近出版された研究法書籍

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University

Mackey, A., & Gass, S. (2005). *Second Language Research: Methodology and Design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.

Wall, D. (2006). *Studies in Language Testing 22: The impact of high-stakes testing on classroom*. Cambridge: Cambridge University Press.

比較的新しい書籍としては上記の4冊を推薦したい。Dörnyeiは応用言語学における研究手法を包括的に扱っている一冊。効果量についても記載あり。Mackey & Gassは、リサーチ・クエスチョンの設定からデータ収集、さらにはコーディングからレポート作成におけるチェックリストまで、質的・量的手法ともかなりを網羅した内容になっている。Ellis & Barkhuizenは、SLAのアプローチにもとづく分析を紹介している(Computer-based analysesはMichael Barlow)。Language TestingシリーズのAppendixの豊富さには驚かされる(今回はスリランカのプロジェクト、他)。

リサーチデザインの設定

三浦省吾(監修) 前田啓朗・山森光陽(編) 磯田貴道・廣森友人(2004)『英語教師のための教育データ分析入門』東京:大修館書店

清川英男・濱岡美郎・鈴木純子(2003)『英語教師のためのExcel活用法』東京:大修館書店

リサーチ・デザインの設定に役立つものとして上記の2冊を推薦したい。英語教育の事例が紹介され、実践に即したデータを使って解説されている。Q&A集や使用するソフトの情報源が掲載され、実証研究に取り組む人にとってはイメージがつかみやすい構成になっている。

心理学研究法

中澤 潤・大野木裕明・南 博文(編)(1997)『心理学マニュアル 観察法』京都:北大路書房

鎌倉雅彦・宮下一博・大野木裕明・中澤 潤(編)(1998)『心理学マニュアル 質問紙法』京都:北大路書房

大野木裕明・中澤 潤・保坂 亨(編)(2000)『心理学マニュアル 面接法』京都:北大路書房

C.ウイリッグ(著) 上淵 寿・大塚まゆみ・小松孝至(訳)(2003)『心理学のための質的研究法入門[創造的な探求に向けて]』東京:培風館

心理学研究法の入門書としては上記の4冊を推薦したい。それぞれの手法における基本的な問題だけでなく、集計や統計的解析、尺度構成や信頼性の検討などの問題も含んでいる。『心理学マニュアル』シリーズとして『要因計画法』や『事例研究法』に関するものも刊行されている。

心理統計法の解説書

山内光哉 (1998) 『心理・教育のための統計法<第2版>』東京:サイエンス社

森 敏昭・吉田寿夫 (1990) 『心理学のためのデータ解析テクニカルブック』京都:北大路書房

山田剛史・村井潤一郎 (2004) 『よくわかる心理統計』京都:ミネルヴァ書房

心理統計法の解説書としては上記の3冊を推薦したい。分布や平均の意味からノンパラメトリック検定まで詳細に解説されている。多変量解析は少ないが、特に分散分析については大変詳しい。さらなる分析手法を望む場合には、分野別の解説書が必要となるので注意して欲しい。

サンプルサイズと効果量の解説書

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

豊田秀樹(編著)(2009). 『検定力分析入門—R で学ぶ最新データ解析—』東京:東京図書

永田 靖 (2003) 『サンプルサイズの決め方』東京: 朝倉書店

吉田寿夫(編) (2007) 『心理学研究法の新しいかたち』東京:誠信書房

サンプルサイズと効果量については上記の4冊を推薦したい。Cohenには効果量とその大きさの目安が書かれている。メタ分析を勉強したい人、サンプルサイズの設計が必要な人、効果量と検定力分析について総合的に学びたい人には、上記の書籍がお薦め。サンプルサイズ・有意水準・検定力・効果量の関係について詳しく書かれている。

American Psychological Association. (2001). *Publication manual of the American Psychological Association (5th ed.)*. Washington, DC: American Psychological Association.

水本 篤・竹内 理 (2008) 「研究論文における効果量の報告のために—基礎的概念と注意点—」『英語教育研究』31, 57-66.

APA第5班ならびに、水本先生・竹内先生の論文は、英語教育関係者は必読。

統計手法の解説書

Field, A. (2005). *Discovering statistics using SPSS (2nd ed.)*. London: Sage Publications.

市川伸一・岸本淳司・大橋靖雄・浜田知久馬 (1993) 『SASによるデータ解析入門<第2版>』東京:東京大学出版会

石村貞夫 (2001) 『SPSSによるカテゴリカルデータ分析の手順』東京:東京書籍

柳井久江 (2004) 『4 Stepsエクセル統計<第2版>』東京:オーエムエス出版

統計パッケージ関係としては上記の4冊を推薦したい。使用するソフトによって使い分けをして欲しいが、いずれもデータの読み方を知る上では有用な本である。統計の知識がないと理解しにくい部分もありが、マニュアルやプログラムの掲載本として活用して欲しい。

英語教育・統計用語辞典

石村貞夫・デズモンドアレン (1997) 『すぐわかる統計用語』東京:東京書籍

白畑知彦・富田祐一・村野井仁・若林茂則 (1999) 『英語教育用語辞典』東京:大修館書店

Longman (2005). *Longman dictionary of English language and culture: Gets to the heart of the language (3rd ed.)*. Edinburgh: Pearson Education Limited.

用語辞典としては上記の3冊を推薦したい。『すぐわかる統計用語』には英語で論文を書く時の基礎知識が付記してある。Longman Dictionary (3rd ed.) は40,000語の語義と15,000の文化的レファレンスボックスを最新化し、アメリカ文化の項目を増やしている。