



心理データ解析演習（2011.4.20）

# 尺度開発の方法 - データ解析の前に -

PD 井関龍太

# 心理尺度を作るという問題

## □国内での尺度を用いた研究で出会う問題

- 因子構造が再現されない
- 他の測度との間に理論的に想定される関連が見いだせない  
尺度開発の方法論上の問題：測定器具が妥当でなければ妥当な研究はできない  
(開発の中でもデータ解析が生かされる)

## □Clark & Watson (1995) を紹介

- Loevinger (1957) の理論ベースの尺度開発アプローチを解説した, 実践的ガイドライン

# 今回の流れ

## □ 構想段階

- 概念化
- 文献レビュー
- 項目プールの生成

- ✓ 内的一貫性の検討
- ✓ 希薄化のパラドクス
- ✓ 因子分析の利用
- ✓ 下位尺度の作成

## □ 評価段階

- テスト構成の方針
- 初回データ収集
- 心理測定学的評価
  - ✓ 項目分布の分析

# 構想段階

- 測定したい**構成概念**（**ターゲット構成概念**）を明確にし，初回項目プールを作成する
  - 実際にデータを収集する前の段階
- ここでは紹介されないアプローチ（Streiner & Norman, 2008）
  - インタビュー，自由回答調査をもとに項目を作成する
  - 臨床的観察，専門家の意見に基づいて項目を作成する

# 概念化

- **理論的背景**の必要性：ごく狭い範囲の構成概念を測る場合でも，理論的枠組みに収めた方がよい
  - 同じ現象の測度でも，開発者の**理論的観点**によって違ったものができる
- **最初の重要なステップ**：ターゲット構成概念とその理論的文脈の**正確かつ詳細な概念化**
  - 構成概念の**短い形式的な記述**を書き出す  
自分のモデルを明確化することに有効

# 文献レビュー（１）

- 同じ問題に対する他の研究者のアプローチを調べる
    - 同じ構成概念と関連する構成概念の両方についての研究
      - ✓ “絶望”の新しい尺度を作るなら，絶望だけでなく，抑うつ，楽観主義，悲観主義も参照
    - 直接的には関連しない構成概念も調べて，ターゲット構成概念の境界を明確にする
      - ✓ “絶望”と他のネガティブ感情（抑うつ，不安，敵愾心，罪悪感，恥，不平）の関係を明確にする
- 優れた収束的・弁別的相関パターン

# 文献レビュー（２）

## □ 包括的な文献レビューの重要性

- ターゲット構成概念の内容の性質と範囲を明確にするのに役立つ
- 既存の尺度の問題点を同定するのに役立つ
  - ✓ 不明確な教示，問題のある回答形式など
- 作ろうとしている尺度が本当に必要かどうかを明らかにする
  - ✓ 扱いたい問題をすでにカバーしている尺度があるなら，それを使って研究すればよい

# 項目プールの生成

- ターゲット構成概念に**潜在的に関連するすべての内容**を系統的に抽出する
  - 初回プールは，自分の理論的見解よりも**広く，包括的に**する
  - 初回プールには，**後で無関連であることがわかるような内容**も含めた方がよい
- 下位概念を構成する**各内容領域**に適切な項目サンプルが入るようにする
  - 各内容領域に費やす項目の割合は，**構成概念における重要度**に比例させるとよい
    - ✓ 特定の内容領域をカバーする項目が1つか2つしかないとその内容が反映されにくくなる



# 項目執筆の基本原則

## □よい項目の条件

- 表現が**単純で直截的**，ターゲット母集団の**読解レベル**に適切なもの
- いまふうの表現，口語表現，年齢やジェンダーによって熟知度が異なるような表現は避ける
- 誰にでも当てはまる，誰も当てはまらない項目は避ける（妥当でない反応を評価する意図がある場合は除く）
- 複雑・両義的な項目**を避ける（2つ以上の特徴を評価してしまう）

# 尺度形式の選択（１）

- **チェックリスト**：当てはまる項目のみにチェックする
    - 全項目に回答を求める形式よりも**反応バイアス**が大きい（Bentler, 1969; Green et al., 1993）
  - **強制選択**：いろいろな構成概念を表す項目から当てはまるものを選ぶ
    - 評価する構成概念の**相対的な個人内強度**しか反映しない；標準的な**個人間情報**をもたらさない
- この２つはあまり勧められない形式

## 尺度形式の選択（２）

- **数直線**：定義した連続体に沿って好きな位置にチェックする
  - **得点化**に労力を要する（PC実施が楽）
- **二値反応**：２つの選択肢から選ぶ
  - **極端な回答分布**の場合，相関を歪める
    - ✓ 極端な回答率の項目を除くとよい
  - 同じ時間内に**多くの項目**に回答できる
- **Likert尺度**：頻度，程度，類似性，賛成など様々に使われる
  - **選択肢の性質**が項目内容を制約する
  - **選択肢の数**を増やすことは，必ずしも信頼性・妥当性を高めない

# 評価段階

□ 実際に尺度を実施して、データに基づいて構成を整える

- 尺度を使った最初の調査
- 理想的には、複数回の項目執筆と分析を繰り返す反復的過程
- 構成概念妥当性は、1回の観察からは推論できない
  - ✓ 尺度の因子構造，他の測度との相関，選抜群間の違い，時間の経過や実験操作への反応についての仮説に沿った変化

# テスト構成の方針

## □ 尺度のモデルを確定する

- 量的（次元）モデル：個人をターゲット構成概念の度合いやレベルについて区別
- クラスモデル：個人を質的に異なるグループに分類
- ダイナミックモデル：さらに複雑

## □ 構造的妥当性を確保する

- 尺度の内的構造（項目間相関）がターゲット特性の外的構造（特性間の相互関係）と平行になるようにする

## □ 3つの主な項目選択方略

- 組み合わせて用いる（ひとつに頼らない）

# 3つの項目選択方略（1）

□ **基準ベースの方法**：尺度と理論的に関連する基準との相関に基づく  
（**妥当性**の決定的な証拠）

● 初回項目プールを**広く不均質なサンプル**に実施

✓ e.g., 健常群と臨床群の両方を含む

● サンプルの**サブグループ**を適切に**識別**する項目を選択

✓ e.g., 健常vs.臨床，臨床範囲内で行動パターンや診断の異なる群

● **欠点**：交差妥当化，新しい環境や他の母集団への一般化が困難；理論を進展させない

## 3つの項目選択方略（2）

□ **内的的一貫性**による方法：因子分析，項目 - 全体相関に基づく

● 使用上の注意：

✓ 盲目的に少数の**単純な規則**に従わないこと

✓ **優れた理論**と**慎重な考察**に代わるものはない  
（冗長な項目を含めると，内的的一貫性の推定値は増すが，スパンの狭い尺度になる）

● **構成概念の中核**を反映するはずの項目が強い相関を示さなかった場合，なぜそうなったのかをよく考えること

✓ 単純に項目を消すことは適切でない

## 3つの項目選択方略（3）

□ **項目反応理論（IRT）**：各テスト項目について，反応と特性の関係を単調増加関数で記述できる（項目特性得点）

● **利点：**

- ✓ 個々の回答者の特性レベルの**推定の正確さ**を示す統計量が得られる
- ✓ 特性レベルの推定は，実施した**特定の項目セットとは独立**に行える  
（知能検査等で，非常に易しい項目・難しい項目を実施していない場合に特性レベルを推定するなど）



# 初回データ収集

- 比較のための **アンカー尺度** を含める
  - **他の構成概念** を評価するマーカーとなる尺度を含める方がよい（ターゲット構成概念の境界を調べる）
- **サンプル** の考慮
  - 適度なサイズのサンプルで予備テスト（e.g., 100 ~ 200名の大学生）
  - その後、より大きな、適度に不均質なサンプルでテスト（最低でも300名）

# 項目分布の分析

- 個々の項目の**反応分布**を調べる
  - **歪度**が高く分布が**不均衡**な項目を除外
- 不均衡な項目が望ましくない理由
  - ほとんどの回答者が同じ回答をするなら、その項目はほとんど情報を持たない
  - ばらつきが限られているため、他の項目と**弱くしか相関しない**
  - **不安定な相関**を生じることがある
    - ✓ 2つの項目に対して1名が両方ともノー、199名がイエスと回答したら相関は1.0になる
- **分布の範囲が広い項目**は残した方がよい

# 内的一貫性の検討

- 理論ベースの測度を作るときに重要
- 係数やKR20などの指標が高ければよいのではない
  - これらの指標は、項目数と項目間の平均相互相関の関数（項目数を増やす，相互相関の高い項目を増やすと高くなる）
    - 平均項目間相関の方が有用：.15～.50の範囲を推奨（Briggs & Cheek, 1986）
      - ✓広い高次の構成概念（外向性など）：.15～.20
      - ✓狭い構成概念（話し好きなど）：.40～.50
  - ただし，平均相互相関だけで十分ではない

# 希薄化のパラドクス

- 一定以上に**内的一貫性を増やす**ことは、**構成概念妥当性を高め**ない；むしろ、**妥当性を低く**することもある
  - 尺度は、**中程度**に相互に相関する、**いろいろな項目**を含む場合により多くの情報を与える（より妥当な測度となる）
  - 内的一貫性を高めすぎると、**内容の狭い**尺度になる（冗長な項目を含めない）

# 因子分析の利用（１）

## □ 因子構造の分析

- 主因子法か主成分分析にかけ，最初の 2，3 因子を調べる
- 項目の第一（非回転）因子・成分への負荷を調べる
  - ✓ 第一因子への負荷が弱く（主因子.35以下，主成分.40以下）他の項目との相関が弱い
  - ✓ 第二以降の因子に第一因子より高い負荷  
除外の候補
  - ✓ 第一因子への負荷が強く，以降の因子への負荷が弱い  
残す候補

# 因子分析の利用（２）

- 弁別的妥当性を高めるための利用
  - よく似た構成概念の尺度項目（ターゲットとだいたい同数）をいっしょに因子分析にかける
    - 理想的には，直行回転で２因子に分かれるはず（因子負荷のパターンから，除外・残す候補をしぼれる）
  - 確認的因子分析も，尺度項目の相互相関と弁別的妥当性の評価に利用できる（Floyd & Widaman, 1995）

# 下位尺度の作成

- 妥当な下位尺度の作成の難しさ
  - 無関連な項目や下位尺度をひとつの全体得点に組み合わせることは、心理測定学的に無意味
  - 相当な相互相関のある項目プールを取り上げて、任意に独立の下位尺度に分割することも、心理測定学的に無意味
- 下位尺度間の項目の相関は、高すぎても低すぎても問題になる
  - $0 < r < .20$ くらいに収める？

# まとめ

## □理論ベースの重要性

- 理論的な位置づけを明確にすることで、より**構成概念妥当性**の高い尺度に
- 類似する構成概念**との違いを意識した方がよい（尺度開発の意義にもつながる）

## □複数回のデータ収集の重要性

- 信頼性だけでなく、**妥当性**（特に、基準ベース）を意識した尺度開発
- 他の母集団，状況への**一般化可能性**を確保する



# 文献

- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319.
- DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd edition). Sage Publications.
  - 今年7月出版予定
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, 51, 493-504.
- 村上宣寛 (2006). *心理尺度のつくり方* 北大路書房
- Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4th edition). Oxford University Press.