

# 大規模集団における抑うつ評価尺度の項目反応と 総スコアの分布の数学的研究

富高 辰一郎<sup>1), 2)</sup> 川崎 洋平<sup>3)</sup> 井出 和希<sup>2), 4)</sup> 芥川 麻衣子<sup>4)</sup> 大野 裕<sup>5)</sup> 古川 壽亮<sup>2)</sup>

1) パナソニック健康保険組合 2) 京都大学 3) 千葉大学 4) 静岡県立大学 5) 認知行動療法研修開発センター

## <要 旨>

抑うつ症状には、抑うつ気分、意欲の低下、睡眠障害、自殺念慮といった症状が含まれる。我々は大規模データを用いれば、一般人口における抑うつ評価尺度の項目反応や総スコアの分布に何らかの特徴的な数理パターンが見出せるのではないかと仮説を立て、研究を開始した。日本の保健福祉動向調査や米国政府が行った健康調査（NHANES, BRFSS, NHIS）の大規模データを用いて、CES-D, K6, PHQ-8, PHQ-9 といった抑うつ尺度の項目反応や総スコアの分布を解析したところ、いずれのデータにおいても、項目反応や総スコアの分布に指数分布に関連した数理分布を示すことを見出した。抑うつ尺度のスコアがこのような特徴的な数理分布に従うメカニズムとしては、抑うつ症状が同じ潜在特性を持ち、かつその潜在特性が指数分布に従うことが示唆された。抑うつ評価尺度の総スコアや項目反応の数理分布（非正規分布）が確立すれば、正規分布を想定した統計手法を見直す必要があると思われた。

## <キーワード> 抑うつ症状、項目反応、分布、指数分布、数理モデル、抑うつ評価尺度

### 【はじめに】

自然科学の研究と心理学研究は方法論において大きく異なる部分がある。

物理化学は、自然現象を観察し、そのデータの中に再現性のある数理パターンを見出し、最終的にはその数理パターンから数理モデルを作成することによって発展してきた。こういった再現性の高いデータのパターンから帰納的に作成されたモデルを帰納モデル（inductive model）と呼ぶ。

一方心理学研究では、様々な統計モデル（例えば因子分析は、一般的に正規分布モデル、線形モデルを使う）を使って観察データを分析し、統計量や有意差の有無を決定するというアプローチが一般的である。一般的に、統計モデルは帰納モ

デルではなく、演繹モデル（deductive model）である。演繹モデルとは一種のデータの見方（perspective）と言ってもよい。一方、帰納モデルは事実に裏づけされた再現性のある法則である。

言うまでもないが、帰納的モデルは、再現性、予測性において非常に優れている。近代になって物理化学が急激に発展したのは、自然現象において帰納モデル（ボイル・シャルルの法則、オームの法則等）が次々と見つかったことが大きい。では心理現象においても物理現象と同じように観察データから数理パターンを見出し、帰納モデルを作成することはできないのだろうか。

我々はサンプルサイズが非常に大きくなれば、心理現象においても再現性の高い数理パターンが出現するのではないかと仮説を立てた。実は物理現象も個々の分子や電子の動きや方向は予測が難しい。しかし非常に多数の分子や電子が存在することによって、肉眼的に物理現象は安定した数理パターンを形成しているように見える。例えば大気中に存在する分子のスピードや方向は様々だが、非常に大量の分子が存在することによって、大気圧は高度や温度に対して再現性の高い数理パターンを示す。

我々は、大規模データを用いれば、一般人口における抑うつ評価尺度の項目反応や総スコアの分布に何らかの特徴的な数理パターンが見出せるのではないかと仮説を立て（大規模の経験分布は理論分布に近づく）、研究を開始した。

## 【方法】

本研究では、様々な国の一般人口における大規模な抑うつ評価尺度のデータを手に入れ、解析を行った。解析対象は原則的にパブリックデータ（研究者に公開されている匿名化された生データ）とした。本研究のような場合、国や公的機関の大規模調査で得られたオープンデータを使用した方が独自に疫学調査を行うよりも利点が多い。パブリックデータは、国や大規模な研究組織が多額の予算と人的予算を使って収集するため、サンプル数、データの質の高さ、透明性、いずれの点でも優れている。実際、こういった透明性の高い大規模なオープンデータを利用すると、結果の再現性が容易に検証できるので学術的にも信頼されやすい。さらに大規模研究の匿名化されたオープンデータを二次利用すれば研究のためのコストや時間が大幅に削減される。なお当該施設

（パナソニック健康保険組合）の倫理委員会は、公開された匿名化されたデータを解析する研究を、倫理委員会の審査を必要と見なさない。

データ解析には、日本で行われた保健福祉動向調査（2000）の Center for Epidemiologic Studies Depression Scale（CES-D）のデータ<sup>1,2</sup>、米国で行われた National Health and Nutrition Examination Survey（NHANES）の Patient Health Questionnaire-9（PHQ-9）データ<sup>3</sup>、National Survey of Midlife Development in the United States（MIDUS）の Kessler Screening Scale for Psychological Distress（K6）データ<sup>4</sup>、Behavioral Risk Factor Surveillance Survey の Patient Health Questionnaire-8（PHQ-8）データ<sup>5</sup>、アイルランドの The Irish Longitudinal Study on Ageing（TILDA）の CES-D データを使用した。これらのデータを用いて、抑うつ評価尺度の項目反応や総スコアの分布のパターン解析を行った。

## 【結果】

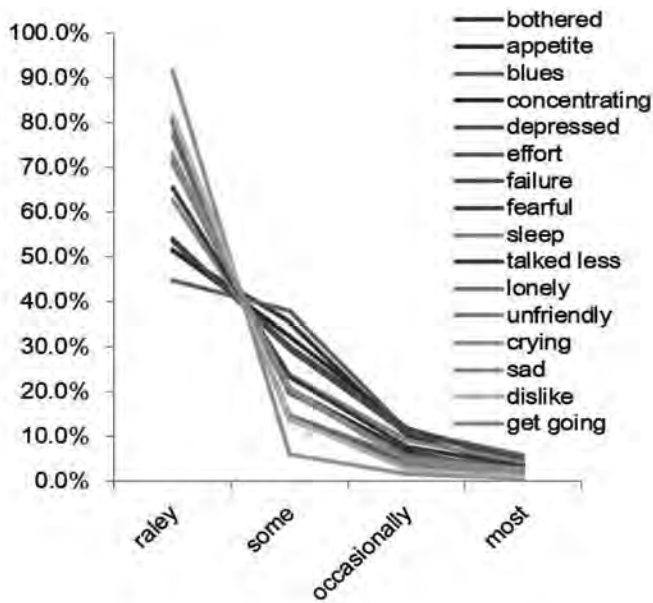
### 1. 保健福祉動向調査の CES-D の項目反応

2000 年に行われた保健福祉動向調査（一般人口約 3 万人）の CES-D 評価尺度の項目反応と総スコアの分布のパターンについて示す<sup>1,2</sup>。

図 1 は CES-D20 項目の中のポジティブ感情を除いた 16 項目の項目反応の結果である。CES-D はそれぞれの抑うつ症状が過去一週間にどの程度存在するか、4 つの応答分類、rarely（一日以下）、some（1-2 日）、occasionally（3-5 日）、most（6 日以上）から選択する形式となっている。

図 1 が示すように 16 項目の項目反応は共通する数理パターンを示している。

図1 保健福祉動向のCES-Dの項目反応



rarely と some の間において一点でクロスし、some から most まではクロスすることもなく右肩下がりに低下した (図1)。Some と most の間の分布を片対数グラフにプロットすると、ほぼ同じ傾きで線形を示した (図2)。

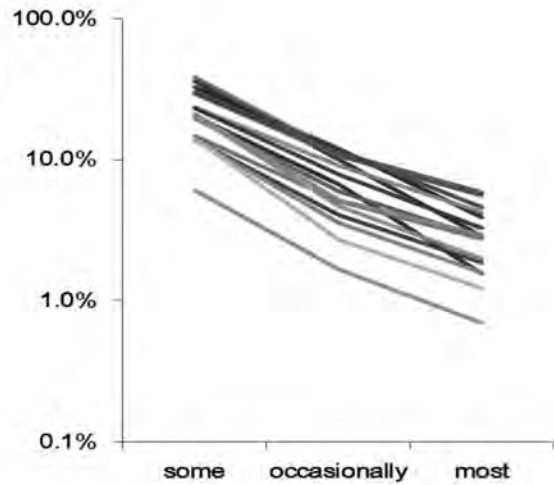
これらのデータから示唆されることは、抑うつ各項目の項目反応にはある種の規則性が存在ということである。

これまでこのような所見の報告がなかったため、慎重に再現性を検討した。7つの大規模疫学調査のデータを解析したが、全てのデータで再現性が確認できた。紙面の都合で全ての解析結果を示すことは難しいので、PHQ-8を用いた米国政府が毎年行うBRFSS(2015)の項目反応のデータ(約2万人)を呈示する。

## 2. 米国 BRFSS データの PHQ-8 の項目反応

PHQ-8は大うつ病のDSMの8項目に基づいて作られた評価尺度であり、通常は過去2週間の状態を5つの応答分類(not at all, several days, more than half the days, nearly every day)か

図2 図1の some から most の対数グラフ



ら選択する形式となっている。しかしBRFSSバージョンでは応答分類を具体的な日数(0日から14日)で答えるようになっている。図3(次のページ)はPHQ-8の8項目の項目反応の結果を示している<sup>5</sup>。

図3が示すように8項目の症状に対する項目反応はいずれも0日と答える比率が非常に高く、逆に6日、9日、13日に対する比率は非常に少ない。5日、7日、10日が隣の日数より比率が高いのはおそらく、キリがよい数字(end-digit preference bias)だからであろう。0日と他の日数の比率の差が大きすぎて、8項目に共通するパターンがあるか評価しにくいので、0日と1日、1日と14日にグラフを分けて、拡大してみる。

図4は図3の0日と1日の部分を拡大した図である。0日と1日の間において矢印が示すように8項目全てが一点でクロスしていることがわかる。つまりCES-Dの項目反応(図1)が、rarelyとsomeの間において一点でクロスしているのと同じパターンである。

図5は図3の1日と14日の部分を拡大した図

図3 PHQ-8の項目反応(過去14日間の症状日数)

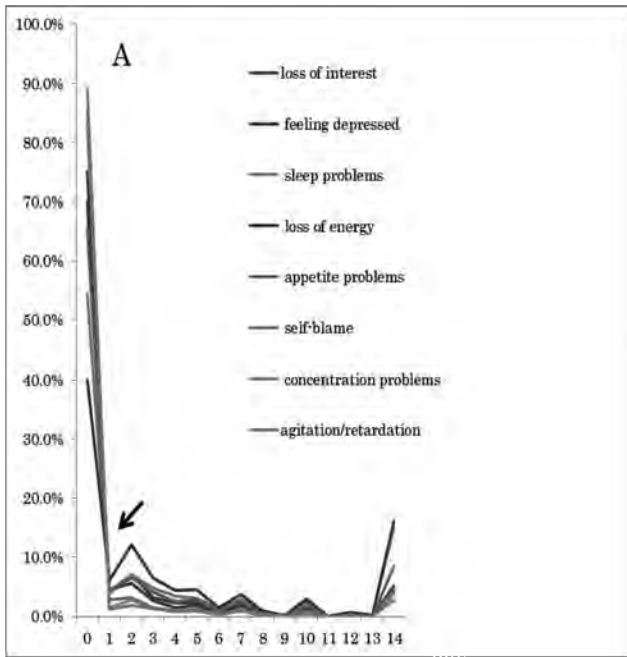


図5 図3の1日と14日間の拡大図

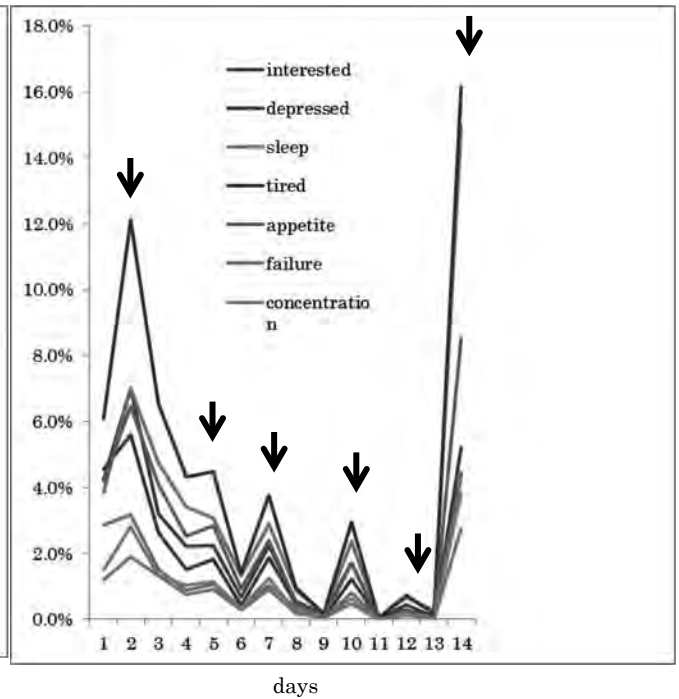


図4 図3の0日と1日間の拡大図

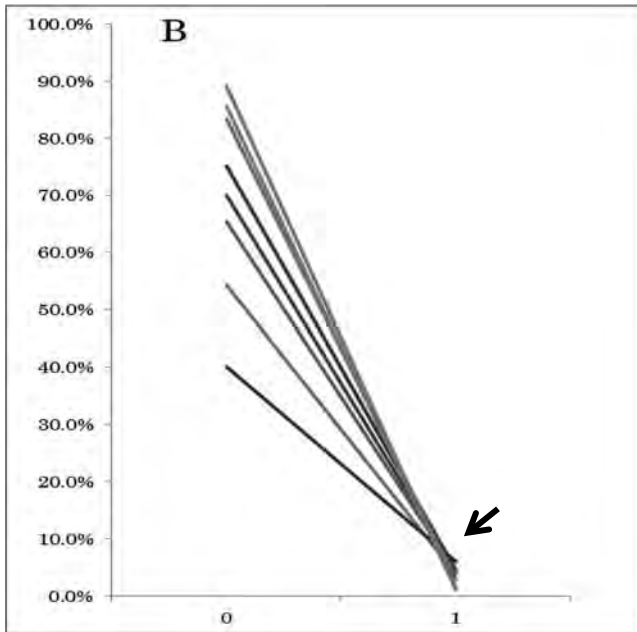
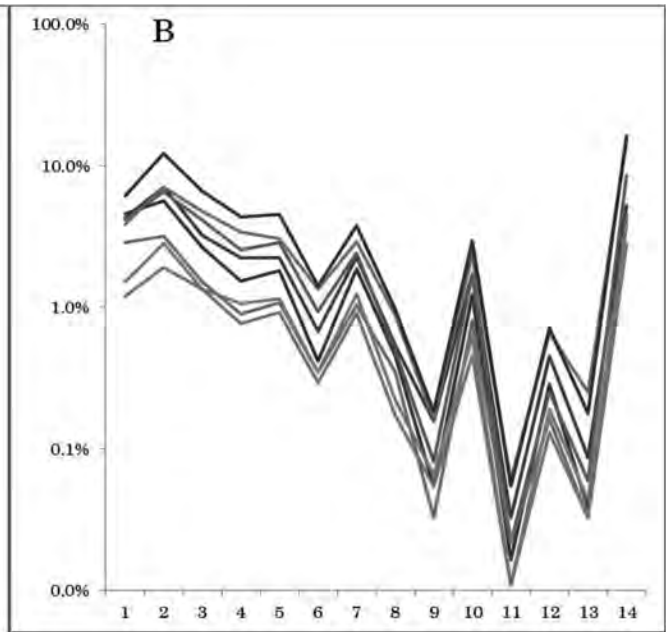


図6 図5の1日と14日間の対数グラフ



である。1日から14日まではクロスすることもなく山と谷を形成している。山と谷になる部分は8つの項目で共通している。2日、5日、7日、10日、14日が山になっているのは、キリがよい数字 (end-digit preference bias) だからであろう。

図5の分布を片対数グラフにプロットすると、ほぼ同じ傾きでジグザグの線形を示した(図6) 8項目の互いのグラフは平行に近い。つまりCES-Dのsomeとmostの間の対数グラフ(図2)において、16項目のラインが平行になっているのと同じパターンである。

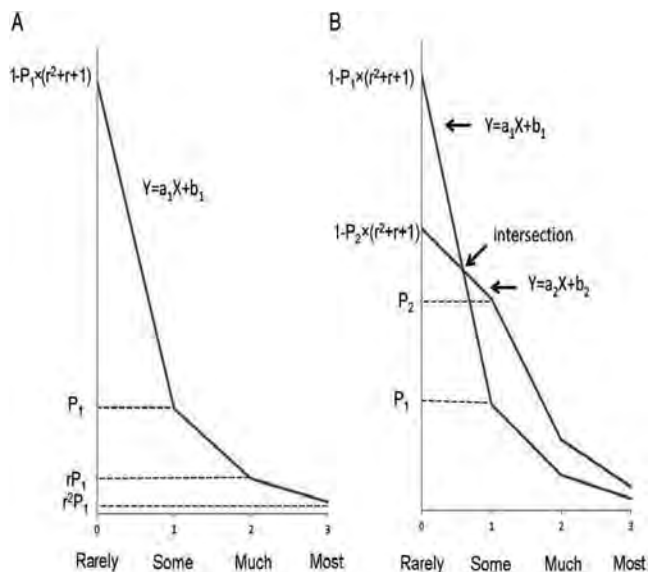
米国で行われた NHANES の PHQ-9 データ<sup>3</sup> や MIDUS や NHIS の K6 データ<sup>4</sup>でもこのような抑うつ症状に共通する項目反応のパターンが起きることが確認できた。

### 3. 項目反応の帰納モデル

複数の大規模データの項目反応を解析することによって、抑うつ尺度に関係なく共通するパターンが存在することが確認できた。CES-D の項目反応 (図 1, 2) でその共通するパターンを、説明すると、全ての項目が、some と occasionally の比率、および occasionally と most の比率が近似していた。項目反応の最小スコアの選択枝 (rarely) はそれ以外の選択枝とこのような関係はなかった。

PHQ-8 の場合でも同じで、0 日を除く 1 日から 14 日までの隣り合う選択枝の確率の比率が近似していた。抑うつ尺度に関係なく共通するパターンが見出されたので、そのパターンを用いて項目反応の分布の帰納モデルを作成した (図 7)<sup>1</sup>。図 1 の帰納モデルが図 7A である。

図 7 CES-D の項目反応の数理モデル



この数理モデルでは、CES-D の 16 項目の抑うつ症状が rarely と occasionally (some) の間において全ての項目反応の線グラフが一点で交わることが明らかになった。また対数グラフでは、occasionally (some) から most の間が全ての項目反応の線グラフが平行になることも明らかになった。

### 4. CES-D の総スコアの分布のパターン

項目反応に何らかのパターンがあるなら、項目スコアの和の分布にも特徴的なパターンが

存在するかもしれない。その仮説に従って前述の保健福祉動向調査の CES-D のスコアの和の分布の解析を行ったところ、16 項目の和の分布が最小スコアを除いて指数分布に禁じることがわかった。なお抑うつ症状のスコアの和が指数分布に近似することは、任意の数の任意の項目を選択しても成立した<sup>6</sup>。

図 8 CES-D の 4 項目の和の分布

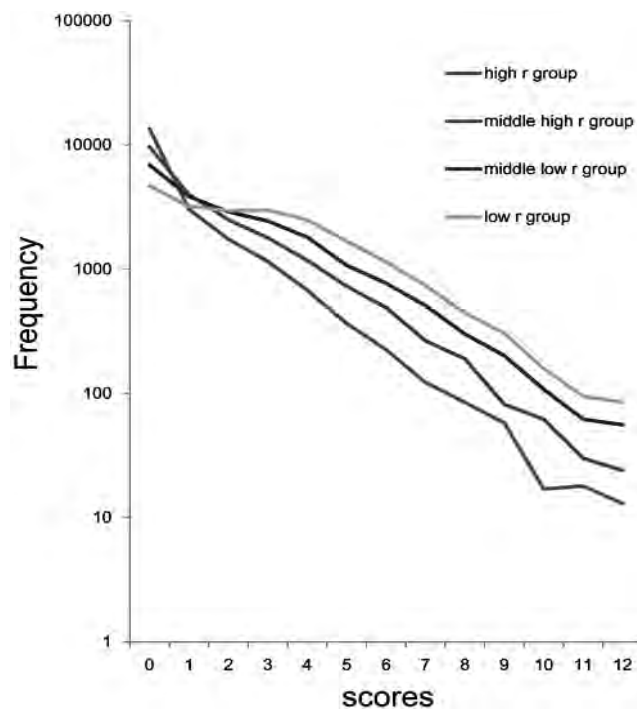


図 8 は前述した保健福祉動向調査(2000 年)の



CES-D データの 16 項目を（症状なし）の確率の大きさから 4 つのグループに分けて、それぞれの総スコアの分布を片対数グラフにプロットしたものである。最少スコア部分を除いていずれのグループも指数パターン（片対数グラフでは線形を示す）に従っていることがわかる。

総スコアが指数パターンに従うという所見も、前述の欧米の大規模疫学調査のデータを使って追試したところ、再現性が確認された。

### 【考察】

一般的に研究は、1) 新たな現象を発見する、2) 既知の現象のメカニズムを明らかにする、の二つの方向に大別できる。本研究の意義は、明らかに前者であり、新しい現象を見出したことにある。抑うつ評価尺度の項目反応や総スコアの分布に共通するパターンが存在するという事は、集団として見ると抑うつ症状にはマクロなルールが存在するという事である。

新たな現象の研究の場合、研究の具体的意義について考察するのは難しい部分もあるが、次のようなことが挙げられる。

意義① 集団における抑うつ評価尺度の項目反応や総スコアの分布の確率モデルが明らかになれば、当然その知見はうつ病のスクリーニングや診断、職場でのストレスチェック等において分布を検証することができるので、検査が適正に行われたか検証することが可能性になる。

意義② 抑うつ評価尺度の項目数、項目の内容、選択肢数（4 件法等）に関係なく、項目反応や総スコアが同じ数理パターンを示すのなら、抑うつ評価尺度に統一する理論を作成することができる可能性がある。また統一された理論ができれば、異なった評価尺度同士の比較が可能となる。

意義③ これまで抑うつ評価尺度の総スコアや項目反応が正規分布に従うことを想定して行われた各種の統計手法は見直しを行う必要がある可能性がある。

意義④ 抑うつ評価尺度は順序尺度の一種であるが、抑うつ評価尺度で測定した項目反応や総スコアが特定のパターンを示した事実から、抑うつ尺度だけではなく、様々な心理尺度のメカニズムを明らかにできる可能性がある。

意義⑤ マクロの視点で見ると抑うつ症状にルールが存在するという事は、抑うつという現象は心理社会的に調節されている可能性がある。Bio—psycho—social という言葉が精神医学ではよく使われるが、本研究の結果は psycho—social 研究のための重要な材料を提供する可能性があると思われる。

今後はこの現象の再現性を様々なデータを使って確認していくと共に、なぜこのような数理パターンが起きるのか<sup>7,8</sup>、その解明を進めていきたい。

### 【文献】

1 Tomitaka S, Kawasaki Y, Furukawa T. A distribution model of the responses to each depressive symptom item in a general population: a cross-sectional study. *BMJ open* 5 (9), e008599 2015

2 Tomitaka S, Kawasaki Y, Furukawa T. Right Tail of the Distribution of Depressive Symptoms Is Stable and Follows an Exponential Curve during Middle Adulthood. *PloS one* 10 (1), e0114624-e0114624

3 Tomitaka S, Kawasaki Y, Ide K, et al. Distributional patterns of item responses and total scores on the PHQ-9 in the general population: data from the National Health and Nutrition Examination Survey. *BMC psychiatry* 18 (1), 108 2018

4 Tomitaka S, Kawasaki Y, Ide K, et al. Pattern analysis of total item score and item response of the Kessler Screening Scale for Psychological Distress (K6) in a nationally representative sample of US adults. *PeerJ* 5, e2987 2017

5 Tomitaka S, Kawasaki Y, Ide K, et al. Item response Patterns on the Patient health Questionnaire-8 in a nationally representative sample of Us adults. *Frontiers in psychiatry* 8 2017

6 Tomitaka S, Kawasaki Y, Ide K, et al. Relationship between item responses of negative affect items and the distribution of the sum of the item scores in the general population. *PloS one* 11 (11), e0165928 2016

7 Tomitaka S, Kawasaki Y, Ide K, et al. Exponential distribution of total depressive symptom scores in relation to exponential latent trait and item threshold distributions: a simulation study. *BMC research notes* 10 (1), 614 2017

8 Tomitaka S, Kawasaki Y, Ide K, et al. Boundary

curves of individual items in the distribution of total depressive symptom scores approximate an exponential pattern in a general population *PeerJ* 4, e2566 2016