

古典的テスト理論

前川眞一

19951020,21,19971121,20020517,20070519,22,23,29,30,31,20080906, 20130524, 20140826,1210,20,
20150104,08,20160119,20170106,08,09,12dnc,13,14,15,19,0304,06,0612dnc,0908,
2018180511,20210126,20230507

目次

1	古典的テスト理論	2
1.1	基本的モデル	2
2	テストの信頼性係数	5
2.1	信頼性係数	5
2.2	複数のテスト	6
2.3	合計点とその信頼性	7
2.4	テストモデルの分類と合計点の信頼性係数	9
2.5	差得点の信頼性	14
2.6	信頼性係数の求め方	14
2.7	真値の区間推定	18
A	付録	21
A.1	期待値と分散・共分散に関して	21
A.2	確率変数間の線形回帰	27
A.3	信頼区間について	29
A.4	付録：不等式	30
A.4.1	池田央 (1973) の p83 の (4.101)	30
A.4.2	池田央 (1973) の p83 の (4.102)	31
A.4.3	コーシー・シュヴァルツの不等式	31
A.4.4	相加平均と相乗平均	31

1 古典的テスト理論

1.1 基本的モデル

古典的テスト理論では、ある個人のテスト得点 X をその真値 (true score) T と誤差得点 E に分け以下のことを仮定する。

$$X = T + E \quad (1)$$

$$\mathcal{E}(E|T) = 0 \quad (2)$$

$$\mathcal{E}(X|T) = T \quad (3)$$

すなわち、真値が T である個人が繰り返し同じテストを受けた場合、その得点に加わる誤差得点の平均は 0、したがって、得点の平均は真値 T となる。すなわち、上記の仮定は、繰り返し測定の期待値としてテスト得点の真値を定義していることになる。なお、(1) 式と (2) 式を仮定すれば (3) 式を、また、(1) 式と (3) 式を仮定すれば (2) 式を導くことができる。また、誤差得点の標準偏差 $\sqrt{\text{var}(E|T)}$ を測定の標準誤差 (standard error of measurement) と呼ぶが、この値は、一般的に真値 T に依存する値である。

次に、ある母集団における真値の (周辺) 分布を仮定し、その平均と分散を

$$\mathcal{E}(T) = \mu \quad \text{and} \quad \text{var}(T) = \sigma_T^2 \quad (4)$$

と置く。すると、その集団におけるテスト得点の (周辺) 分布の平均と分散は

$$\mathcal{E}(X) = \mu \quad (5)$$

$$\text{var}(X) \equiv \sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (6)$$

となる。なぜなら、付録 A.1.0.5 の 2 に示す条件付き期待値や分散の公式を用いて

$$\mathcal{E}(X) = \mathcal{E}(\mathcal{E}(X|T)) = \mathcal{E}(T) = \mu \quad (7)$$

$$\text{var}(X) = \mathcal{E}(\text{var}(X|T)) + \text{var}(\mathcal{E}(X|T)) = \mathcal{E}(\text{var}(X|T)) + \text{var}(T) \quad (8)$$

となるからである。ただし、誤差分散 (error variance) σ_E^2 を

$$\sigma_E^2 \equiv \mathcal{E}(\text{var}(E|T)) = \mathcal{E}(\text{var}(T + E|T)) = \mathcal{E}(\text{var}(X|T)) \quad (9)$$

と、個人ごとの誤差分散 (測定の標準誤差の 2 乗) の平均として定義した。したがって、これの平方根を平均的測定の標準誤差 (SEM) (average standard error of measurement) と呼ぶことがあるが、一般的に、測定の標準誤差 (SEM) という場合には、(9) 式の平方根を指すことが多い。(Harvill, 1991 を参照。)

この、観測される得点、真値、ならびに誤差得点の関係を 図 1 に示した。図中、横軸は T 、縦軸は X であるが、傾き 45 度の右上がりの線が、 X の T への回帰直線、すなわち T を与えたときの X の期待値 (すなわち T) を示している。また、誤差得点の分布は、その平均が 0 であるため、分布の中心が回帰直線上に乗っているが、その散らばりの大きさが $\text{var}(X|T) = \text{var}(E|T)$ であり、その平方根が測定の標準誤差 (SEM) である。

なお、誤差得点の (周辺) 分布に関しては、その平均が 0 に、また、その分散が、上記に定義した σ_E^2 になることが以下のようにして示される。

$$\mathcal{E}(E) = \mathcal{E}(\mathcal{E}(E|T)) = \mathcal{E}(0) = 0 \quad (10)$$

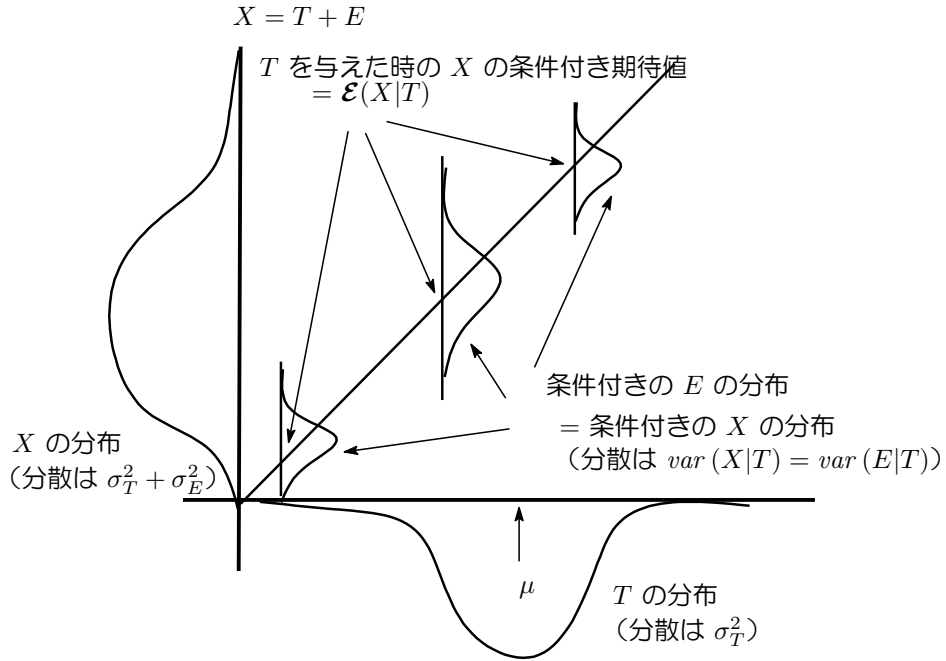


図 1: Classical Test Theory

$$\text{var}(E) = \mathcal{E}(\text{var}(E|T)) + \text{var}(\mathcal{E}(E|T)) = \mathcal{E}(\text{var}(E|T)) + 0 = \sigma_E^2 \quad (11)$$

したがって、確率変数の和の分散の公式を用いると

$$\text{var}(X) = \text{var}(T + E) = \text{var}(T) + \text{var}(E) + 2\text{cov}(T, E) \quad (12)$$

より

$$\text{cov}(T, E) = 0 \quad (13)$$

を得る。

なお、確率変数 X, T, E に個人を表す添え字 i や繰り返し測定を表す添え字 k を付けても同様の結果を導くことができるが、その際は、たとえば

$$\mathcal{E}(X_i) = \mu, \quad \text{and} \quad \text{var}(X_i) = \sigma_T^2 + \sigma_E^2, \quad i = 1, 2, \dots$$

の様な式を、 X_i がすべて独立で同一の分布をする確率変数であると考え、(所謂 iid) その共通の平均と分散を示していると考えればよい。厳密に考えると、真値で条件付けた場合の得点分布 (同じ真値を持つ個人を集めて得られる得点分布) と、個人内での繰り返しによる得点分布は異なるとも考えられるが、テストの場合、繰り返し測定には無理があるため、個人の違いは真値の違いのみで表現されるという潜在変数モデルに共通の立場に立ち、この二つの分布を区別していない。

上記のモデルの記述、すなわち (1) 式と (2) 式もしくは (1) 式と (3) 式のみを仮定し、(4) 式の定義を用いて、(5) 式, (6) 式, (10) 式, (13) 式を導くやり方は、ほぼ、Novick et al. (1971) や Lord (1980) の第 2 章によるものであるが、逆に、(1) 式と (10) 式ならびに (13) 式を基本仮定とし、(4) 式と (11) 式の定義を用いて、(5) 式や (6) 式を導くやり方もある。

すなわち

$$X = T + E \quad (1) \text{ 式}$$

$$\mathcal{E}(E) = 0 \quad (10) \text{ 式}$$

$$\text{cov}(T, E) = 0 \quad (13) \text{ 式}$$

ならびに

$$\sigma_T^2 \equiv \text{var}(T), \quad \sigma_E^2 \equiv \text{var}(E), \quad \sigma_X^2 \equiv \text{var}(X), \quad \mu \equiv \mathcal{E}(T) \quad (14)$$

より

$$\mathcal{E}(X) = \mathcal{E}(T + E) = \mathcal{E}(T) + \mathcal{E}(E) = \mathcal{E}(T) = \mu \quad (15)$$

$$\text{var}(X) = \text{var}(T + E) = \text{var}(T) + \text{var}(E) + 2\text{cov}(T, E) = \sigma_T^2 + \sigma_E^2 \quad (16)$$

が導かれる。また、(1) 式を X の T への線形の回帰式と解釈するか、(13) 式が T と E の独立を意味するものと解釈するか、(10) 式がすべての下位集団において誤差得点の平均が 0 を意味すると解釈すれば、

$$\mathcal{E}(E|T) = 0 \quad (17)$$

となるため

$$\mathcal{E}(X|T) = \mathcal{E}(T + E|T) = \mathcal{E}(T|T) + \mathcal{E}(E|T) = T \quad (18)$$

である。また、(11) 式より

$$\mathcal{E}(\text{var}(E|T)) = \text{var}(E) - \text{var}(\mathcal{E}(E|T)) = \text{var}(E) - 0 = \sigma_E^2 \quad (19)$$

を得る。

なお、古典的テスト理論における真値や誤差得点は、潜在変数として定義されており、決して直接的に観測されることはない。観測されるのはテスト得点 X のみであることに注意されたい。

また、テスト X の信頼性係数 (reliability coefficient) を

$$\rho(X) = \frac{\sigma_T^2}{\sigma_X^2} \quad (20)$$

と定義する。

通常、平均的測定の標準誤差 (SEM) は未知であるがテストの信頼性係数が既知であれば、上式から

$$\sigma_E = \sqrt{1 - \rho(X)} \times \sigma_X \quad (21)$$

で与えられる。

また、Lord(1959) によると、経験的に、テストが中程度の困難度 (通過率) を持つものであれば、 n 個の項目からなるテストの正答数得点 (n 点満点) の測定の標準誤差 (SEM) は

$$\sigma_E \approx 0.45\sqrt{n} \quad (22)$$

であり、それより易しい場合には上記の値より小さいことが知られている。(Harvill, 1991)

2 テストの信頼性係数

2.1 信頼性係数

テスト X の信頼性係数 (reliability coefficient) $\rho(X)$ は

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} = \frac{\sigma_T^2}{\sigma_X^2} \quad (23)$$

と定義される。またこれは、(6) 式を用いると

$$\rho(X) = 1 - \frac{\text{var}(E)}{\text{var}(X)} = 1 - \frac{\sigma_E^2}{\sigma_X^2} \quad (24)$$

と表現される。いずれにしても信頼性係数は

$$0 \leq \rho(X) \leq 1 \quad (25)$$

を満たす。また、古典的テストのモデルから

$$\text{cov}(X, T) = \text{cov}(T + E, T) = \text{cov}(T, T) + \text{cov}(E, T) = \text{var}(T) + 0 = \sigma_T^2 \quad (26)$$

したがって、

$$\text{corr}(X, T) = \frac{\text{cov}(X, T)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(T)}} = \frac{\sigma_T^2}{\sqrt{\sigma_X^2}\sqrt{\sigma_T^2}} = \frac{\sqrt{\sigma_T^2}}{\sqrt{\sigma_X^2}} = \sqrt{\frac{\sigma_T^2}{\sigma_X^2}} \quad (27)$$

であるため、

$$\rho(X) = \text{corr}(X, T)^2 \quad (28)$$

とも書ける。

すなわち、信頼性係数とは、テスト得点の分散のうち真値の分散が占める割合であり、また、テスト得点とその真値との相関係数の 2 乗である。この相関係数の事を信頼性指標 (reliability index) と呼ぶことがある。

古典的テスト理論で用いられる公式の多くは、上記の信頼性係数が既知の場合として導かれている。たとえば、信頼性係数が解っている場合には、

$$\text{var}(T) = \rho(X)\text{var}(X) \quad \text{and} \quad \text{var}(E) = (1 - \rho(X))\text{var}(X) \quad (29)$$

であるから、データを基にして、真値の分散と誤差の分散を推定することが出来る。以下の記述では信頼性係数が既知であることを仮定する。

2.1.0.1 回帰分析的見方

(1) 式並びに (3) 式は X を T に線形に回帰したときの回帰式であると考えられる。したがって、その傾きは 1、切片は 0 である。なお、これは、付録に示した回帰の公式を、 $X \rightarrow V$, $T \rightarrow U$ として用いても

$$b = \frac{\text{cov}(T, X)}{\text{var}(T)} = \frac{\text{var}(T)}{\text{var}(T)} = 1 \quad (30)$$

$$a = \mathcal{E}(X) - b\mathcal{E}(T) = \mathcal{E}(X) - 1 \times \mathcal{E}(T) = \mathcal{E}(T) - \mathcal{E}(T) = 0 \quad (31)$$

導くことが出来る。そして、その分散説明率は

$$R^2 = \text{corr}(X, T)^2 = \rho(X) \quad (32)$$

を得る。すなわち、テストの信頼性係数とは、テスト得点をその真値へ線形に回帰したときの分散説明率と解釈することが出来る。

2.2 複数のテスト

m 個のテストを添字付き変数 X_1, X_2, \dots, X_m で表し、それぞれに上記のモデルが当てはまっているとする。(本節以降、添字はテストを表すものであり、個人を表すものではないことに注意。) すなわち、

$$X_j = T_j + E_j \quad (33)$$

$$\mathcal{E}(E_j) = 0 \quad \text{and} \quad \text{var}(E_j) = \sigma_{E_j}^2 \quad (34)$$

$$\mathcal{E}(X_j) = \mathcal{E}(T_j) = \mu_j \quad (35)$$

$$\text{var}(X_j) = \text{var}(T_j) + \text{var}(E_j) \quad (36)$$

$$\rho(X_j) = \frac{\text{var}(T_j)}{\text{var}(X_j)} = 1 - \frac{\text{var}(E_j)}{\text{var}(X_j)} \quad (37)$$

また、

$$\text{cov}(E_j, E_k) = 0, \quad j \neq k, \quad \text{and} \quad \text{cov}(T_j, E_k) = 0, \quad j, k = 1, 2, \dots, m \quad (38)$$

であるとする。

この場合、 $j \neq k$ に関して、

$$\text{cov}(X_j, X_k) = \text{cov}(T_j + E_j, T_k + E_k) \quad (39)$$

$$= \text{cov}(T_j, T_k) + \text{cov}(T_j, E_k) + \text{cov}(E_j, T_k) + \text{cov}(E_j, E_k) \quad (40)$$

$$= \text{cov}(T_j, T_k) \quad (41)$$

である。なお、 $j = k$ の場合を含めると

$$\text{cov}(X_j, X_k) = \begin{cases} \text{var}(X_j) = \text{var}(T_j) + \text{var}(E_j) & \text{if } j = k \\ \text{cov}(T_j, T_k) & \text{if } j \neq k \end{cases} \quad (42)$$

と書ける。同様に

$$\text{cov}(X_j, T_k) = \text{cov}(T_j, X_k) = \text{cov}(T_j, T_k) \quad (43)$$

である。

また、これらの関係を利用すると

$$\text{corr}(X_j, T_k) = \frac{\text{cov}(X_j, T_k)}{\sqrt{\text{var}(X_j)\text{var}(T_k)}} \quad (44)$$

$$= \frac{\text{cov}(X_j, X_k)}{\sqrt{\text{var}(X_j)\rho(X_k)\text{var}(X_k)}} \quad (45)$$

$$= \frac{\text{corr}(X_j, X_k)}{\sqrt{\rho(X_k)}} \geq \text{corr}(X_j, X_k) \quad (46)$$

$$\text{corr}(T_j, T_k) = \frac{\text{cov}(T_j, T_k)}{\sqrt{\text{var}(T_j)\text{var}(T_k)}} \quad (47)$$

$$= \frac{\text{std}(X_j)\text{std}(X_k)\text{corr}(X_j, X_k)}{\sqrt{\rho(X_j)\text{std}(X_j)}\sqrt{\rho(X_k)\text{std}(X_k)}} \quad (48)$$

$$= \frac{\text{corr}(X_j, X_k)}{\sqrt{\rho(X_j)\rho(X_k)}} \geq \text{corr}(X_j, X_k) \quad (49)$$

となるが、これらを**稀薄化の修正公式 (correction formula for attenuation)**と呼ぶ。要するに、観測されるテスト得点は誤差を含むので、見かけの相関が低くなるが、テストの信頼

性が分かっているならば、上記の修正公式を用いることにより、真値の間の相関を求めることが出来る、ということである。また、この修正に関しては、誤差が取り除かれるにしたがって高くなっていく

$$\text{corr}(X_j, X_k) \leq \frac{\text{corr}(X_j, T_k)}{\text{corr}(T_j, X_k)} \leq \text{corr}(T_j, T_k) \quad (50)$$

ということである。また、(49) 式より

$$\text{corr}(X_j, X_k) = \text{corr}(T_j, T_k) \sqrt{\rho(X_j)\rho(X_k)} \quad (51)$$

であるが、相関係数ならびに信頼性係数は 1 以下の値を取るため

$$\text{corr}(X_j, X_k) \leq \sqrt{\rho(X_j)\rho(X_k)} = \sqrt{\rho(X_j)}\sqrt{\rho(X_k)} \leq \sqrt{\rho(X_j)} \quad (52)$$

となるが、このことは、あるテスト j とそれ以外のテストとの相関係数は、自分自身の信頼性の平方根を超えないことを示している。すなわち、信頼性が低いテストは、どのようなテストとも相関が低い。したがって、外的基準との相関を妥当性の指標とする場合、テストの妥当性（相関係数）の値は、信頼性の値を超えることがない。このことを、**テストの信頼性は妥当性の必要条件である**と表現することがある。すなわち、信頼性が高くない限りそのテストの妥当性は高くなることはない。

2.3 合計点とその信頼性

各テストの得点を合計をした合計点を考えると、それはやはり真値と誤差に分かれ、

$$X = \sum_{j=1}^m X_j \quad (53)$$

$$= \sum_{j=1}^m (T_j + E_j) \quad (54)$$

$$= \sum_{j=1}^m T_j + \sum_{j=1}^m E_j \quad (55)$$

$$= T + E \quad (56)$$

その平均と分散は以下のように様々な形で書くことが出来る。

$$\mathcal{E}(X) = \sum_{j=1}^m \mu_j \quad \text{and} \quad \mathcal{E}(E) = 0 \quad (57)$$

$$\text{var}(T) = \sum_{j=1}^m \sum_{k=1}^m \text{cov}(T_j, T_k) \quad (58)$$

$$= \sum_{j=1}^m \text{var}(T_j) + \sum_{j \neq k} \text{cov}(T_j, T_k) \quad (59)$$

$$= \sum_{j=1}^m \text{var}(X_j)\rho(X_j) + \sum_{j \neq k} \text{cov}(X_j, X_k) \quad (60)$$

$$\text{var}(E) = \sum_{j=1}^m \text{var}(E_j) \quad (61)$$

$$= \sum_{j=1}^m \text{var}(X_j)(1 - \rho(X_j)) \quad (62)$$

$$\text{var}(X) = \sum_{j=1}^m \sum_{k=1}^m \text{cov}(X_j, X_k) \quad (63)$$

$$= \sum_{j=1}^m \text{var}(X_j) + \sum_{j \neq k} \text{cov}(X_j, X_k) \quad (64)$$

$$= \sum_{j=1}^m \text{var}(X_j) + \sum_{j \neq k} \text{cov}(T_j, T_k) \quad (65)$$

なお、(41)式を用いた。

いずれにしても

$$\text{var}(X) = \text{var}(T) + \text{var}(E) \quad (66)$$

となる。また、合計点の信頼性係数は

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} \quad (67)$$

$$= \frac{\sum_{j=1}^m \text{var}(T_j) + \sum_{j \neq k} \text{cov}(T_j, T_k)}{\text{var}(X)} \quad (68)$$

$$= \frac{\sum_{j=1}^m \text{var}(X_j)\rho(X_j) + \sum_{j \neq k} \text{cov}(X_j, X_k)}{\text{var}(X)} \quad (69)$$

$$= 1 - \frac{\text{var}(E)}{\text{var}(X)} \quad (70)$$

$$= 1 - \frac{\sum_{j=1}^m \text{var}(X_j)(1 - \rho(X_j))}{\text{var}(X)} \quad (71)$$

と書くことが出来るが、(71)式の形は、合計点の信頼性係数が、個々のテストの分散および共分散と信頼性係数とで書き表された形となっている。

この形から、各テスト間の相関構造が同じであれば、各テストの信頼性が高いほど、合計点の信頼性が高いことが解る。また、各テストの信頼性係数と分散が一定の場合、合計点の信頼性を高めるためには(71)式の名分母に現れる $\text{var}(X)$ を大きくすれば良く、そのためには、(63)式が示すように、各テスト間の共分散、すなわち相関を高めればよいことが解る。すなわち、**内的整合性が高いテストの合計点(合計点)の信頼性係数は高くなる。**

また、合計点と各テストの共分散は

$$\text{cov}(X, X_j) = \text{cov}(\sum_{k=1}^m X_k, X_j) \quad (72)$$

$$= \sum_{k=1}^m \text{cov}(X_k, X_j) = \sum_{k=1}^m \text{cov}(X_j, X_k) \quad (73)$$

であるから、合計点の分散は(63)式より

$$\text{var}(X) = \sum_{j=1}^m \sum_{k=1}^m \text{cov}(X_j, X_k) = \sum_{j=1}^m \text{cov}(X, X_j) = \sum_{j=1}^m \text{corr}(X, X_j) \sqrt{\text{var}(X)\text{var}(X_j)} \quad (74)$$

と書ける。そしてこの両辺を $\sqrt{\text{var}(X)}$ で除ると

$$\sqrt{\text{var}(X)} = \sum_{j=1}^m \text{corr}(X, X_j) \sqrt{\text{var}(X_j)} \quad (75)$$

したがって、各テスト得点とその合計点の共分散または相関が高い場合にも合計点の信頼性係数が高くなる。この関係を根拠に、合計点との相関が低い項目を取り除くことがある。

2.4 テストモデルの分類と合計点の信頼性係数

m 個のテスト $X_j = T_j + E_j$ が存在するとし、真値と誤差、ならびに誤差同士は無相関であるとする。

2.4.0.1 平行テスト

いま、これらの m 個のテストの真値が全て等しく、誤差分散も全て等しい

$$T_j = T_* \quad \text{and} \quad \text{var}(E_j) = \sigma_E^2, \quad j = 1, 2, \dots, m \quad (76)$$

と仮定できる場合、それらの m 個のテストを平行テスト（強平行テスト parallel test）と呼ぶ。個人を表す添え字を付けて表現すると、

$$X_{ij} = T_{i*} + E_{ij}, \quad i = 1, 2, \dots, N \quad (77)$$

となり、誤差得点はその平均と分散が等しいことのみを仮定している。この場合、共通の真値 T_* の分散を $\sigma_{T_*}^2$ とすると、各テストの平均と分散は以下のようになり

$$\mathcal{E}(X_j) = \mathcal{E}(T_* + E_j) = \mathcal{E}(T_*) = \mu \quad (78)$$

$$\text{var}(X_j) = \text{var}(T_* + E_j) = \text{var}(T_*) + \text{var}(E_j) = \sigma_{T_*}^2 + \sigma_E^2 \equiv \sigma_{X_*}^2 \quad (79)$$

全て等しくなる。（なお、この結果を $\sigma_{X_*}^2$ と呼ぶ。）したがって、信頼性係数も等しくなり、

$$\text{cov}(X_j, X_k) = \text{cov}(T_j, T_k) = \text{cov}(T_*, T_*) = \text{var}(T_*) = \sigma_{T_*}^2 \quad (80)$$

$$\text{corr}(X_j, X_k) = \frac{\text{cov}(X_j, X_k)}{\sqrt{\text{var}(X_j)\text{var}(X_k)}} = \frac{\sigma_{T_*}^2}{\sigma_{X_*}^2} = \rho^* \quad (81)$$

として与えられる。すなわち、平行なテスト間の相関係数は、それらのテストの信頼性係数を表している。

また、この場合の単純合計点 X の信頼性係数は

$$\rho(X) = \frac{m\rho^*}{1 + (m-1)\rho^*} \quad (82)$$

となる。何故なら、平行テストの合計点は

$$X = \sum_{i=1}^m X_i = \sum_{i=1}^m (T_* + E_i) = mT_* + \sum_{i=1}^m E_i = T + E \quad (83)$$

であり、したがって、

$$\text{var}(T) = \text{var}(mT_*) = m^2\sigma_{T_*}^2 \quad (84)$$

$$\text{var}(X) = \text{var}(mT_* + \sum_{i=1}^m E_i) = \text{var}(mT_*) + \text{var}(\sum_{i=1}^m E_i) \quad (85)$$

$$= \text{var}(mT_*) + \sum_{i=1}^m \text{var}(E_i) = m^2\sigma_{T_*}^2 + m\sigma_E^2 \quad (86)$$

となるが、これを信頼性係数の定義式に代入し、分子分母を $\sigma_{X^*}^2$ で除すと

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} = \frac{m^2 \sigma_{T^*}^2}{m^2 \sigma_{T^*}^2 + m \sigma_E^2} \quad (87)$$

$$= \frac{m^2 \sigma_T^2 / \sigma_{X^*}^2}{m^2 \sigma_{T^*}^2 / \sigma_{X^*}^2 + m \sigma_E^2 / \sigma_{X^*}^2} = \frac{m^2 \rho^*}{m^2 \rho^* + m(1 - \rho^*)} \quad (88)$$

$$= \frac{m \rho^*}{m \rho^* + (1 - \rho^*)} = \frac{m \rho^*}{1 + (m - 1) \rho^*} \quad (89)$$

となるからである。これを **Spearman-Brown の公式**と呼ぶ。

Spearman-Brown の公式は、それを項目単位で用いた場合に、テストの長さ、すなわち、テストに含まれる項目数が、そのテストの信頼性に与える影響を表す公式として利用される。

また、 $m = 2$ の場合には、テストを二つの平行な部分に分割して信頼性を求める述べる折半法の公式となる。(後述)

2.4.0.2 τ 等価テスト

次に、誤差分散が等しいという仮定を落とし、真値が等しいこと

$$T_j = T_*, j = 1, 2, \dots, m \quad (90)$$

のみを仮定したモデルを **τ 等価テスト (tau-equivalent test)**、また、真値が定数だけずれているということ

$$T_j = b_j + T_*, j = 1, 2, \dots, m \quad (91)$$

のみを仮定したモデルを**本質的 τ 等価テスト (essentially tau-equivalent test)**と呼ぶ。この両者は**弱平行テスト**とも呼ばれることがある。本質的 τ 等価テストなテストの場合、

$$\mathcal{E}(X_j) = \mathcal{E}(b_j + T_* + E_j) = \mathcal{E}(b_j + T_*) = b_j + \mu \quad (92)$$

$$\text{var}(X_j) = \text{var}(b_j + T_* + E_j) = \text{var}(T_*) + \text{var}(E_j) = \sigma_{T^*}^2 + \text{var}(E_j) \quad (93)$$

$$\text{cov}(X_j, X_k) = \text{var}(T_*) = \sigma_{T^*}^2, j \neq k \quad (94)$$

$$\text{corr}(X_j, X_k) = \frac{\sigma_{T^*}^2}{\sqrt{\text{var}(X_j)}\sqrt{\text{var}(X_k)}} = \sqrt{\frac{\sigma_{T^*}^2}{\text{var}(X_j)}}\sqrt{\frac{\sigma_{T^*}^2}{\text{var}(X_k)}} \quad (95)$$

$$= \sqrt{\rho(X_j)}\sqrt{\rho(X_k)} = \text{corr}(X_j, T_j)\text{corr}(X_k, T_k) \quad (96)$$

となる。(28)式を利用。)

また、単純合計点に関しては

$$X = \sum_{i=1}^m X_i = \sum_{i=1}^m (T_* + E_i) = mT_* + \sum_{i=1}^m E_i = T + E \quad (97)$$

より

$$\text{var}(T) = \text{var}(mT_*) = m^2 \sigma_{T^*}^2 \quad (98)$$

$$\text{var}(X_i) = \text{var}(T_*) + \text{var}(E_i) = \sigma_{T^*}^2 + \text{var}(E_i) \quad (99)$$

$$\text{var}(X) = \text{var}(mT_* + \sum_{i=1}^m E_i) = \text{var}(mT_*) + \text{var}(\sum_{i=1}^m E_i) \quad (100)$$

$$= m^2 \sigma_{T^*}^2 + \sum_{i=1}^m \text{var}(E_i) \quad (101)$$

であるが、(99) 式の両辺の和を取ることで

$$\sum_{i=1}^m \text{var}(X_i) = m\sigma_{T^*}^2 + \sum_{i=1}^m \text{var}(E_i) \quad (102)$$

したがって

$$\sum_{i=1}^m \text{var}(E_i) = \sum_{i=1}^m \text{var}(X_i) - m\sigma_{T^*}^2 \quad (103)$$

であるため、これを (101) 式に代入すると

$$\text{var}(X) = m^2\sigma_{T^*}^2 + \sum_{i=1}^m \text{var}(E_i) = m^2\sigma_{T^*}^2 + \sum_{i=1}^m \text{var}(X_i) - m\sigma_{T^*}^2 \quad (104)$$

$$= m(m-1)\sigma_{T^*}^2 + \sum_{i=1}^m \text{var}(X_i) \quad (105)$$

を得る。また、この表現から、真値の分散が

$$\sigma_{T^*}^2 = \frac{\text{var}(X) - \sum_{i=1}^m \text{var}(X_i)}{m(m-1)} \quad (106)$$

という形で観測される得点の分散として表されるため、信頼性係数は

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} = \frac{m^2\sigma_{T^*}^2}{\text{var}(X)} \quad (107)$$

$$= \frac{m}{m-1} \frac{\text{var}(X) - \sum_{i=1}^m \text{var}(X_i)}{\text{var}(X)} \quad (108)$$

$$= \frac{m}{m-1} \left(1 - \frac{\sum_{i=1}^m \text{var}(X_i)}{\text{var}(X)} \right) \equiv \alpha(X) \quad (109)$$

と書くことが出来る。この形の信頼性係数の計算方法は、**Cronbach の α 係数**と呼ばれており、 $\alpha(X)$ と記される。当然であるが、テストが平行である場合、Cronbach の α 係数は Spearman-Brown の公式に一致する。

また、各テストが 0-1 で採点される項目得点の場合、項目 j の通過率 p_j を用いて

$$\text{var}(X_j) = p_j(1-p_j) \quad (110)$$

と書けるため α 係数は

$$\alpha(X) = \frac{m}{m-1} \left(1 - \frac{\sum_{i=1}^m p_j(1-p_j)}{\text{var}(X)} \right) \quad (111)$$

と書けるが、この表現を **KR20 (Kuder-Richardson の公式 20)** と呼ぶ。

最後に、テストが本質的 τ 等価でない場合、すなわち、(94) 式が満たされない場合、 α 係数は正しい信頼性係数 (71) 式より小さいことが知られている。なぜなら、(106) 式に示される (94) 式を仮定して推定される真値の分散の分子は、(63) 式や (41) 式より

$$\text{var}(X) - \sum_{i=1}^m \text{var}(X_i) = \sum_{i \neq j} \text{cov}(X_i, X_j) = \sum_{i \neq j} \text{cov}(T_i, T_j) \quad (112)$$

であるため、(106) 式は以下のように $m(m-1)$ 個の共分散の平均として真値の分散を推定している

$$\sigma_{T^*}^2 = \frac{1}{m(m-1)} \sum_{i \neq j} \text{cov}(T_i, T_j) \quad (113)$$

と解釈できる。したがって、

$$m^2 \sigma_{T^*}^2 = m^2 \frac{1}{m(m-1)} (\text{var}(X) - \sum_{i=1}^m \text{var}(X_i)) = \frac{m}{m-1} \sum_{i \neq j} \text{cov}(T_i, T_j) \quad (114)$$

であるが、付録に示す (303) 式を用いることにより

$$\text{var}(T) = \text{var}(\sum_{i=1}^m T_i) \geq \frac{m}{m-1} \sum_{i \neq j} \text{cov}(T_i, T_j) = m^2 \sigma_{T^*}^2 \quad (115)$$

である。したがって、 α 係数の定義式のうち (107) 式の部分は、テストが本質的 τ 等価でない場合、すなわち、(94) 式が満たされない場合には

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} \geq \frac{m^2 \sigma_{T^*}^2}{\text{var}(X)} = \alpha(X) \quad (116)$$

となる。このことを称して、 α 係数は真の信頼性係数の下限となっていると言う。

2.4.0.3 同族テスト

最後に、単に、各テストの真値間の相関が 1 の場合、すなわち、

$$T_j = b_j + \lambda_j T_*, \quad j = 1, 2, \dots, m \quad (117)$$

の場合を同族テスト (congeneric test) と呼ぶ。この場合は、

$$X_j = b_j + \lambda_j T_* + E_j \quad (118)$$

となり、いわゆる一次元の因子分析モデルに一致する。観測値の平均や分散は、 μ と $\sigma_{T^*}^2$ を T_* の平均ならびに分散として

$$\mathcal{E}(X_j) = b_j + \lambda_j \mu \quad \text{and} \quad \text{var}(X_j) = \lambda_j^2 \sigma_{T^*}^2 + \text{var}(E_j) \quad (119)$$

$$\text{cov}(X_j, X_k) = \lambda_j \lambda_k \sigma_{T^*}^2 \quad (120)$$

であるが、 b_j と μ の間の加算的不定性と λ_j と T_* の間の乗法的不定性を避けるため

$$\mu = 0, \quad \text{and} \quad \sigma_{T^*}^2 = 1 \quad (121)$$

と置くことが多いが、 $\lambda_j > 0$ を仮定して

$$\mu = 0, \quad \text{and} \quad \sum_{j=1}^m \lambda_j = 1 \quad (122)$$

を用いても良い。

各テストの信頼性係数は

$$\rho(X_j) = \frac{\text{var}(T_j)}{\text{var}(X_j)} = \frac{\lambda_j^2 \sigma_{T^*}^2}{\lambda_j^2 \sigma_{T^*}^2 + \text{var}(E_j)} \quad (123)$$

となるが、(27) 式より、これは、各テスト得点と真値 T_j もしくは T_* の間の相関係数

$$\text{corr}(X_j, T_*) = \frac{\text{cov}(X_j, T_*)}{\sqrt{\text{var}(X_j) \sigma_{T^*}^2}} = \frac{\lambda_j \sigma_{T^*}^2}{\sqrt{\lambda_j^2 \sigma_{T^*}^2 + \text{var}(E_j)} \sqrt{\sigma_{T^*}^2}} = \sqrt{\rho(X_j)} \quad (124)$$

を 2 乗したものに等しい。いずれにしてもこれらの値は λ_j の単調増加関数となっている。

また、単純合計点に関しては

$$X = \sum_{j=1}^m (b_j + \lambda_j T_* + E_j) = \sum_{j=1}^m b_j + \left(\sum_{j=1}^m \lambda_j \right) T_* + \sum_{j=1}^m E_j \quad (125)$$

$$= T + E \quad (126)$$

であるから

$$\text{var}(T) = \left(\sum_{j=1}^m \lambda_j \right)^2 \sigma_{T_*}^2 \quad (127)$$

$$\text{var}(X) = \left(\sum_{j=1}^m \lambda_j \right)^2 \sigma_{T_*}^2 + \sum_{j=1}^m \text{var}(E_j) \quad (128)$$

$$(129)$$

となり、その信頼性係数は、

$$\rho(X) = \frac{\left(\sum_{j=1}^m \lambda_j \right)^2 \sigma_{T_*}^2}{\left(\sum_{j=1}^m \lambda_j \right)^2 \sigma_{T_*}^2 + \sum_{j=1}^m \text{var}(E_j)} \quad (130)$$

となる。(121) 式を用いると、 λ_j の大きなテストの合計点の信頼性が高いことが分かる。

各テストと合計点との共分散は

$$\text{cov}(X_j, X) = \text{cov}\left(b_j + \lambda_j T_* + E_j, \sum_{j=1}^m b_j + \left(\sum_{j=1}^m \lambda_j \right) T_* + \sum_{j=1}^m E_j\right) \quad (131)$$

$$= \text{cov}\left(\lambda_j T_* + E_j, \left(\sum_{k=1}^m \lambda_k \right) T_* + E_j\right) = \lambda_j \left(\sum_{k=1}^m \lambda_k \right) \sigma_{T_*}^2 + \text{var}(E_j) \quad (132)$$

である。したがって、各テストと合計点との相関係数は

$$\text{corr}(X_j, X) = \frac{\lambda_j \left(\sum_{k=1}^m \lambda_k \right) \sigma_{T_*}^2 + \text{var}(E_j)}{\sqrt{\lambda_j^2 \sigma_{T_*}^2 + \text{var}(E_j)} \sqrt{\left(\sum_{j=1}^m \lambda_j \right)^2 \sigma_{T_*}^2 + \sum_{j=1}^m \text{var}(E_j)}} \quad (133)$$

であるが、(122) 式を使うと

$$\text{corr}(X_j, X) = \frac{\lambda_j \sigma_{T_*}^2 + \text{var}(E_j)}{\sqrt{\lambda_j^2 \sigma_{T_*}^2 + \text{var}(E_j)} \sqrt{\sigma_{T_*}^2 + \sum_{j=1}^m \text{var}(E_j)}} \quad (134)$$

となる。そして、この λ_j に関する一次微分

$$\frac{d \text{corr}(X_j, X)}{d \lambda_j} = \frac{(1 - \lambda_j) \text{var}(E_j) \sigma_{T_*}^2}{\sqrt{\sigma_{T_*}^2 + \sum_{k=1}^m \text{var}(E_k)} (\lambda_j^2 \sigma_{T_*}^2 + \text{var}(E_j))^{3/2}} \quad (135)$$

は非負であるため、各テストと合計点との相関係数は λ_j の単調増加関数であることが分かる。

各テスト X_j と自分自身を除く合計点 X_{-j} との共分散は、(122) 式を使うと

$$\text{cov}(X_j, X_{-j}) = \text{cov}\left(b_j + \lambda_j T_* + E_j, \sum_{k \neq j}^m b_k + \left(\sum_{k \neq j}^m \lambda_k \right) T_* + \sum_{k \neq j}^m E_k\right) \quad (136)$$

$$= \text{cov}\left(\lambda_j T_* + E_j, \left(\sum_{k \neq j}^m \lambda_k \right) T_*\right) = \lambda_j \left(\sum_{k \neq j}^m \lambda_k \right) \sigma_{T_*}^2 = \lambda_j (1 - \lambda_j) \sigma_{T_*}^2 \quad (137)$$

である。したがって、各テストと合計点との相関係数は

$$\text{corr}(X_j, X) = \frac{\lambda_j(1 - \lambda_j)\sigma_{T^*}^2}{\sqrt{\lambda_j^2\sigma_{T^*}^2 + \text{var}(E_j)}\sqrt{(1 - \lambda_j)^2\sigma_{T^*}^2 + \sum_{k \neq j}^m \text{var}(E_k)}} \quad (138)$$

である。

2.5 差得点の信頼性

一人の人がふたつのテスト X_1 と X_2 を受けている時、その差を

$$X_D = X_2 - X_1 = T_2 + E_2 - (T_1 + E_1) = T_2 - T_1 + E_2 - E_1 = T_D + E_D \quad (139)$$

ただし

$$T_D = T_2 - T_1 \quad \text{and} \quad E_D = E_2 - E_1 \quad (140)$$

とする。このとき、以下のことが言える。ただし、ふたつのテストの信頼性係数を $\rho(X_1)$, $\rho(X_2)$ とする。

$$\text{var}(X_D) = \text{var}X_1 + \text{var}(X_2) - 2\text{cov}(X_1, X_2) \quad (141)$$

$$= \text{var}(X_1) + \text{var}(X_2) - 2\sqrt{\text{var}(X_1)\text{var}(X_2)}\text{corr}(X_1, X_2) \quad (142)$$

$$\text{var}(T_D) = \text{var}(T_1) + \text{var}(T_2) - 2\text{cov}(T_1, T_2) \quad (143)$$

$$= \text{var}(X_1)\rho(X_1) + \text{var}(X_2)\rho(X_2) - 2\sqrt{\text{var}(X_1)\text{var}(X_2)}\text{corr}(X_1, X_2) \quad (144)$$

したがって差得点の信頼性係数は

$$\rho(X_D) = \frac{\text{var}(T_D)}{\text{var}(X_D)} = \frac{\text{var}(X_1)\rho(X_1) + \text{var}(X_2)\rho(X_2) - 2\sqrt{\text{var}(X_1)\text{var}(X_2)}\text{corr}(X_1, X_2)}{\text{var}(X_1) + \text{var}(X_2) - 2\sqrt{\text{var}(X_1)\text{var}(X_2)}\text{corr}(X_1, X_2)} \quad (145)$$

となる。

ここで、ふたつのテストの分散が等しいと仮定すると、上式は、

$$\text{corr}(X_1, X_2) \leq \frac{1}{2}(\rho(X_1) + \rho(X_2)) \leq 1 \quad (146)$$

の関係から、

$$\rho(X_D) = \frac{\frac{1}{2}(\rho(X_1) + \rho(X_2)) - \text{corr}(X_1, X_2)}{1 - \text{corr}(X_1, X_2)} \leq \frac{1}{2}(\rho(X_1) + \rho(X_2)) \quad (147)$$

となる。すなわち、差得点の信頼性係数は、元のテストの信頼性係数の平均を超えることはなく、ふたつのテストの相関が高いほど、その値は低くなる。

2.6 信頼性係数の求め方

2.6.0.1 集団への依存性

信頼性係数はテストの精度を表す統計量であるが、その定義からわかる様に、その大きさは、それを計算する集団の特性にも左右される。

たとえば、小学生の間に習うべき漢字の習熟度を測るテストがあり、そのテストの信頼性を、集団1として全学年のデータを用いて計算した場合と、集団2として4-5年生のデータのみを用いて計算した場合を考えてみる。この場合、集団 k におけるテスト得点、真値、並びに誤差を $X_{(k)}, T_{(k)}, E_{(k)}$ とすると、

$$\text{var}(X_{(1)}) > \text{var}(X_{(2)}) \quad (148)$$

であると考えられるが、その理由は、テストの特性である誤差分散（測定の標準誤差の2乗の平均）が変わるといよりは、真値の分散が減少していると考えた方が納得がいく。（必ずしもこうなるとは限らないが。）すなわち、

$$\text{var}(T_{(1)}) > \text{var}(T_{(2)}) \quad \text{and} \quad \text{var}(E_{(1)}) \approx \text{var}(E_{(2)}) \equiv \sigma_E^2 \quad (149)$$

であると考えられる。したがって、各集団ごとに計算された信頼性係数を

$$\rho(X_{(k)}) = \frac{\text{var}(T_{(k)})}{\text{var}(X_{(k)})} \approx 1 - \frac{\sigma_E^2}{\text{var}(X_{(k)})} \quad (150)$$

とすると、誤差分散は等しいという上記の仮定は

$$\text{var}(X_{(1)})(1 - \rho(X_{(1)})) \approx \text{var}(X_{(2)})(1 - \rho(X_{(2)})) \quad (151)$$

となり、

$$\text{var}(X_{(1)}) > \text{var}(X_{(2)}) \quad \rightarrow \quad \rho(X_{(1)}) > \rho(X_{(2)}) \quad (152)$$

を得る。すなわち **得点分布の分散が小さな集団から計算された信頼性係数の値は低くなる**。また、本来の集団における得点分布の分散の値がわかっている場合には、以下の式を使って、およその補正をすることが出来る。

$$\rho(X_{(1)}) \approx 1 - (1 - \rho(X_{(2)})) \frac{\text{var}(X_{(2)})}{\text{var}(X_{(1)})} \quad (153)$$

2.6.0.2 再テストもしくは平行テスト法

テスト X の信頼性係数を求めるための方法として、受験者に X に平行なテスト Y を受験させたデータを利用する方法がある。当然であるが、 X 自身は X に平行である。この場合、(81) 式より、平行なテスト得点間の相関係数を計算すればそれが信頼性係数の推定値となる。

ただし、多くのテストの場合、再テストには記憶の影響が出ることは不可避であり、また、平行テスト法は、被験者に二つのテストを受けさせるという負担を求めなければならない。

2.6.0.3 折半法

テスト X が $2n$ 個の項目で構成されている場合、それをふたつに分けてテスト X_1 と X_2 を作る。

もしもふたつのテスト X_1, X_2 が平行に作られているならば、その間の相関係数は、半分の長さのテストの信頼性係数の推定値となる。

$$\text{corr}(X_1, X_2) = \rho(X_1) = \rho(X_2) = \rho^* \quad (154)$$

したがって、元のテストは $X_1 + X_2$ であるから、Spearman-Braun の公式 (89) 式に於いて $m = 2$ とすれば得られる。すなわち

$$\rho(X) = \frac{2\rho^*}{1 + \rho^*} = \frac{2\text{corr}(X_1, X_2)}{1 + \text{corr}(X_1, X_2)} \quad (155)$$

である。

ふたつのテストが完全に平行ではなく、 τ 等価（弱平行）である場合には、 $m = 2$ の場合の α 係数を求めれば良いが

$$\rho(X) = \alpha(X) = 2 \left(1 - \frac{\text{var}(X_1) + \text{var}(X_2)}{\text{var}(X_1 + X_2)} \right) \quad (156)$$

この式は Guttman の L_4 と呼ばれる。

なお、 $m=2$ の場合に α 係数は以下の様な形に変形することが出来るが

$$\alpha(X) = 1 - \frac{\text{var}(X_1 - X_2)}{\text{var}(X_1 + X_2)} \quad (157)$$

$$= 4 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1 + X_2)} \quad (158)$$

これらの形を、上から Rulon の公式、そして Flanagan の公式と呼ぶ。なお、Rulon の公式は、 τ 等価（弱平行）な二つのテスト得点の差が

$$X_1 - X_2 = b_1 + T_* + E_1 - (b_2 + T_* + E_2) = b_1 - b_2 + E_1 - E_2 \quad (159)$$

となるため、その分散が

$$\text{var}(X_1 - X_2) = \text{var}(b_1 - b_2 + E_1 - E_2) = \text{var}(E_1) + \text{var}(E_2) = \text{var}(E_1 + E_2) \quad (160)$$

となることを利用している。また、Flanagan の公式は、

$$X_1 + X_2 = b_1 + T_* + E_1 + b_2 + T_* + E_2 = b_1 + b_2 + 2T_* + E_1 + E_2 \quad (161)$$

の真値 $b_1 + b_2 + 2T_*$ の分散が $4\text{var}(T_*)$ として計算されることを利用している。

なお、上記の α 係数は (155) 式に示す Spearman-Braun の公式を用いて同じテストから計算された値よりも小さくなることが知られている。なぜなら、Flanagan の公式から

$$\alpha(X) = 4 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1 + X_2)} \quad (162)$$

$$= 4 \frac{\sqrt{\text{var}(X_1)}\sqrt{\text{var}(X_2)}\text{corr}(X_1, X_2)}{\text{var}(X_1) + \text{var}(X_2) + 2\sqrt{\text{var}(X_1)}\sqrt{\text{var}(X_2)}\text{corr}(X_1, X_2)} \quad (163)$$

$$= \frac{2\text{corr}(X_1, X_2)}{\frac{1}{2} \left[\frac{\text{var}(X_1) + \text{var}(X_2)}{\sqrt{\text{var}(X_1)}\sqrt{\text{var}(X_2)}} \right] + \text{corr}(X_1, X_2)} \quad (164)$$

を得るが、上式の大括弧の中は $\frac{\sqrt{\text{var}(X_1)}}{\sqrt{\text{var}(X_2)}}$ と $\frac{\sqrt{\text{var}(X_2)}}{\sqrt{\text{var}(X_1)}}$ の相加平均であるため相乗平均を下回ることがない、すなわち

$$\frac{1}{2} \left[\frac{\sqrt{\text{var}(X_1)}}{\sqrt{\text{var}(X_2)}} + \frac{\sqrt{\text{var}(X_2)}}{\sqrt{\text{var}(X_1)}} \right] \geq \sqrt{\frac{\sqrt{\text{var}(X_1)}}{\sqrt{\text{var}(X_2)}} \times \frac{\sqrt{\text{var}(X_2)}}{\sqrt{\text{var}(X_1)}}} = 1 \quad (165)$$

であるため、(164) 式 の分母の値は (155) 式 の分母の値を下回ることがないからである。なお、等式が成立するのは、 $\text{var}(X_1) = \text{var}(X_2)$ の場合、すなわち二つのテストが平行である場合である。

2.6.0.4 α 係数の利用

テスト X が m 個の項目から作られている場合、それぞれの項目を 0-1 の値を取るテストと考える。もしもそれらの m 個のテストが τ 等価（弱平行）であれば、 α 係数を計算することにより合計点、すなわちテスト X の信頼性係数の推定値を求めることが出来るが、この場合、(111) 式に示す KR-20 の形で計算する事も可能である。

α 係数を用いる根拠として以下のような考え方も存在する。

信頼性係数は二つの平行なテスト、 X_1 と X_2 の間の相関係数 $corr(X_1, X_2)$ として定義されているが、いま、各テストが m 個の項目、 $X_{ik}, i = 1, 2, \dots, m, k = 1, 2$ を含む場合を考える。

すると二つのテストの共分散は

$$cov(X_1, X_2) = cov\left(\sum_{i=1}^m X_{i1}, \sum_{i=1}^m X_{i2}\right) \quad (166)$$

$$= \sum_{i=1}^m \sum_{j=1}^m cov(X_{i1}, X_{j2}) = m^2 \bar{c} \quad (167)$$

と、全ての項目間の共分散の和として書くことが出来る。ただし、 \bar{c} は全ての共分散の平均である。他方、各テストの分散は

$$var(X_k) = var\left(\sum_{i=1}^m X_{ik}\right) \quad (168)$$

$$= \sum_{i=1}^m var(X_{ik}) + \sum_{i \neq j} cov(X_{ik}, X_{jk}) \quad (169)$$

と書き表せるため、テスト k から計算される項目間共分散の平均値 \bar{c}_k は

$$\bar{c}_k = \frac{1}{m(m-1)} \left(var(X_k) - \sum_{i=1}^m var(X_{ik}) \right) \quad (170)$$

となる。二つのテストは平行であるから、この \bar{c}_k は $k=1, 2$ にかかわらず似た値になると考えられる。そして、この値を (167) 式に現れる \bar{c} の推定値として用いれば、

$$cov(X_1, X_2) = m^2 \times \frac{1}{m(m-1)} \left(var(X_k) - \sum_{i=1}^m var(X_{ik}) \right) \quad (171)$$

$$= \frac{m}{(m-1)} \left(var(X_k) - \sum_{i=1}^m var(X_{ik}) \right) \quad (172)$$

を得る。従って、信頼性係数、すなわち相関係数は

$$\rho(X_k) = corr(X_1, X_2) = \frac{cov(X_1, X_2)}{\sqrt{var(X_1)}\sqrt{var(X_2)}} \quad (173)$$

$$= \frac{m}{m-1} \frac{var(X_k) - \sum_{i=1}^m var(X_{ik})}{var(X_k)}, \quad k = 1, 2 \quad (174)$$

と、一つのテスト X_1 または X_2 から得られる統計量のみで計算することが出来る。

前節では、テストを構成する m 個の項目得点が本質的 τ 等価であるという強い仮定の元に、信頼性係数が (109) 式の形となることを示したが、ここでは、(同じ長さの) 平行テストを構成する項目得点間の共分散の平均は等しいという仮定の下に、 α 係数と同じ形の信頼性係数の推定値を導いた。なお、詳細は南風原・杉沢・平井 (2009) を参照のこと。

2.7 真値の区間推定

2.7.0.1 真値の信頼区間

古典的テスト理論においては、母数 T が所与の場合の統計量 X の分布の条件付き平均は $\mathcal{E}(X|T) = T$ であり、 X を観測した場合、

$$\hat{T} = X \quad (175)$$

を用いて真値 T を推定するとすることは自然である。このとき、 \hat{T} 、したがって、 X の真値 T を与えたときの条件付き分布が正規分布で近似できる場合には

$$\hat{T}|T \sim N\left(T, \text{var}\left(\hat{T}|T\right)\right) \quad (176)$$

となるため、 $z_{1-\alpha/2}$ を標準正規分布の $100 * (1 - \alpha/2)\%$ の分位点であるとすると

$$\Pr\left(T - z_{1-\alpha/2}\sqrt{\text{var}\left(\hat{T}|T\right)} < \hat{T} < T + z_{1-\alpha/2}\sqrt{\text{var}\left(\hat{T}|T\right)}\right) \approx 1 - \alpha \quad (177)$$

であるが、この不等式の同値変形より

$$\Pr\left(\hat{T} - z_{1-\alpha/2}\sqrt{\text{var}\left(\hat{T}|T\right)} < T < \hat{T} + z_{1-\alpha/2}\sqrt{\text{var}\left(\hat{T}|T\right)}\right) \approx 1 - \alpha \quad (178)$$

を得る。したがって、 $\hat{T} = X = x$ を観測した時の真値の $100(1 - \alpha)\%$ 信頼区間は、 $\text{var}\left(\hat{T}|T\right)$ を $\text{var}\left(\hat{T}|T = x\right)$ で推定するとすれば、

$$x \pm z_{(1-\alpha/2)} \sqrt{\text{var}\left(X|T = x\right)} \quad (179)$$

で与えられる。そして、この、上限と下限を I_R, I_L とする。これを用いると、「貴方の得点は x 点ですが、貴方の真の得点は、ほぼ、 I_R 点と I_L 点の間に入っている事は確実です。」といった言い方が可能となる。「ほぼ」は α の大きさに依存するが、例えば $z_{(1-\alpha/2)} = 1$ とすると $\alpha \approx 0.3$ であるので「ほぼ」は「3回のうち2回は」という解釈となる。(実際は、信頼区間なので、区間が変動していることになる。)

なお、後に述べる項目反応理論を背後に仮定した場合には、測定の標準誤差 (SEM) は個人の実験値の値に依存する値となるが、古典的テスト理論の枠組み内でも、Feldt, Steffan, & Gupta(1985) 等が示すように、測定の標準誤差 (SEM) は個人に依存することが知られている。しかし、

$$\text{var}\left(X|T\right) \approx \mathcal{E}\left(\text{var}\left(X|T\right)\right) \equiv \sigma_E^2 \quad (180)$$

と見なされる場合には

$$x \pm z_{(1-\alpha/2)} \sigma_E \quad (181)$$

である。したがって、

$$100(1 - \alpha)\% \text{信頼区間の幅} = I_R - I_L = 2 z_{(1-\alpha/2)} \sigma_E \quad (182)$$

であるが、 $z_{(1-\alpha/2)} = 1$ の場合は $w = 2\sigma_E$ となる。

たとえば、 $\text{var}\left(X\right) = 12^2, \rho\left(X\right) = 0.75$ の試験の場合、 $\sigma_E^2 = 12^2 \times 0.25 = (12 \times 0.5)^2 = 6^2$ となるため、70 点を取った人の真値の 68% 信頼区間は (64, 76) となる。

2.7.0.2 真値の信用区間

真値と誤差の分布に独立な正規分布を仮定した場合、すなわち、

$$X|T \sim N(T, \sigma_E^2) \quad \text{and} \quad T \sim N(\mu, \sigma_T^2) \quad (183)$$

とした場合、 X を観測した時の T の条件付き分布 (事後分布) は、ベイズ統計学における正規分布の母平均に関する議論 (分散が既知の場合で自然共役事前分布を用いた場合) から、

$$T|X \sim N\left(\mu_{T|X}, \sigma_{T|X}^2\right) \quad (184)$$

ただし、

$$\sigma_{T|X}^2 = \frac{1}{\frac{1}{\sigma_E^2} + \frac{1}{\sigma_T^2}} = \frac{\sigma_E^2 \sigma_T^2}{\sigma_T^2 + \sigma_E^2} \quad (185)$$

$$\mu_{T|X} = \sigma_{T|X}^2 \left(\frac{X}{\sigma_E^2} + \frac{\mu}{\sigma_T^2} \right) = \sigma_{T|X}^2 \frac{\sigma_T^2 X + \sigma_E^2 \mu}{\sigma_T^2 \sigma_E^2} = \frac{\sigma_T^2 X + \sigma_E^2 \mu}{\sigma_T^2 + \sigma_E^2} \quad (186)$$

となる。したがって、 $X = x$ を観測した時の T の信用区間は

$$x \pm z_{(1-\alpha/2)} \sigma_{T|X} \quad (187)$$

そして、その幅は

$$100(1 - \alpha)\% \text{信用区間の幅} = 2 z_{(1-\alpha/2)} \sigma_{T|X} \quad (188)$$

となるが、これは確率変数 T の分布を表しているため、信頼区間よりも直感的に分かりやすい確率表現をすることが出来る。すなわち、この場合には、「貴方の真の得点 m_T は、ほぼ、 $t - 1$ 点と $t + 1$ 点の間に入っています。」といった言い方が可能となる。

なお、 T の分布に局所的一様分布、すなわち、 $1/\sigma_T^2 \rightarrow 0$ を仮定した場合、上式の第1項から

$$\sigma_{T|X}^2 \rightarrow \frac{1}{\frac{1}{\sigma_E^2} + 0} = \sigma_E^2 \quad (189)$$

$$\mu_{T|X} \rightarrow \sigma_{T|X}^2 \left(\frac{X}{\sigma_E^2} + 0 \right) = X \quad (190)$$

となるため、前節に述べた信頼区間は T の分布に局所的一様分布を用いた場合の信用区間と形式的に一致する。

また、上式は、古典的テストの基本的公式と信頼性係数を用いて

$$\mu_{T|X} = \rho(X) X + (1 - \rho(X)) \mu \quad (191)$$

$$\sigma_{T|X}^2 = \sigma_X^2 (1 - \rho(X)) \rho(X) = \sigma_E^2 \rho(X) \quad (192)$$

と書くことが出来るが、この式は、以下の様に導出することも出来る。

テスト理論では、真値を与えたときのテスト得点の回帰 (条件付き期待値) が真値に関して線形であることを仮定しているが、それに加えて、真値のテスト得点への回帰がテスト得点に関して線形であると仮定すると、付録 A.2 に示した回帰の公式を、 $T \rightarrow V$, $X \rightarrow U$ として用いて

$$b = \frac{\text{cov}(X, T)}{\text{var}(X)} = \frac{\text{var}(T)}{\text{var}(X)} = \rho(X) \quad (193)$$

$$a = \mathcal{E}(T) - b \mathcal{E}(X) = (1 - \rho(X)) \times \mu \quad (194)$$

を得る。したがって、この線形回帰を用いた真値の予測式は

$$t = a + b x = (1 - \rho(X))\mu + \rho(X) x \quad (195)$$

となり、先ほどの $\mu_{T|X}$ の形と一致する。

これを Kelly の公式と呼ぶ。(Kelly, 1947) (池田(1973) の公式 5.4、もしくは Harvill(1991) を参照のこと。)

A 付録

A.1 期待値と分散・共分散に関して

A.1.0.1 密度関数を用いた期待値と分散・共分散の定義

X, Y, X_1, X_2 等を確率変数とし、その密度関数を密度関数

$$f(x), f(x, y) \text{ 等、ただし } \int_{\mathbf{X}} f(x) dx = 1, \int_{\mathbf{Y}} \int_{\mathbf{X}} f(x, y) dx dy = 1 \quad (196)$$

と表す。確率変数が多変量（ベクトル）の場合は同時密度関数を以下のように書く。

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_m), \text{ ただし} \quad (197)$$

$$\int_{\mathbf{X}} f(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{X}_1} \int_{\mathbf{X}_2} \cdots \int_{\mathbf{X}_m} f(x_1, x_2, \dots, x_m) dx_m dx_{m-1} \cdots dx_1 = 1 \quad (198)$$

また、各確率変数の周辺密度関数を

$$f_j(x_j) = \int_{\mathbf{X}_1} \int_{\mathbf{X}_2} \cdots \int_{\mathbf{X}_{j-1}} \int_{\mathbf{X}_{j+1}} \cdots \int_{\mathbf{X}_m} f(x_1, x_2, \dots, x_m) dx_m dx_{m-1} \cdots dx_{j+1} dx_{j-1} \cdots dx_1 \quad (199)$$

で定義する。当然であるが、

$$\int_{\mathbf{X}_j} f_j(x_j) dx_j = 1 \quad (200)$$

である。

- 期待値

$$\mathcal{E}(X) = \int_{\mathbf{X}} x f(x) dx \quad (201)$$

X が m 変量の確率変数ベクトルの場合は X の各要素の期待値を X と同じ形に並べたもの

$$\mathcal{E}(\mathbf{X}) = \begin{pmatrix} \mathcal{E}(X_1) \\ \mathcal{E}(X_2) \\ \vdots \\ \mathcal{E}(X_m) \end{pmatrix} \quad (202)$$

ただし、

$$\mathcal{E}(X_j) = \int_{\mathbf{X}_j} x_j f_j(x_j) dx_j \quad (203)$$

である。したがって、

$$\mathcal{E}(\mathbf{X}') = \mathcal{E}(\mathbf{X})' \quad (204)$$

である。

- 分散

$$\text{var}(X) = \int_{\mathbf{X}} (x - \mathcal{E}(X))^2 f(x) dx = \mathcal{E}(X^2) - \mathcal{E}(X)^2 \quad (205)$$

- 共分散

$$\text{cov}(X, Y) = \int_{\mathbf{Y}} \int_{\mathbf{X}} (x - \mathcal{E}(X))(y - \mathcal{E}(Y))f(x, y)dx dy = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) \quad (206)$$

したがって、

$$\text{cov}(X, X) = \text{var}(X) \quad (207)$$

である。

- 相関係数

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} \quad (208)$$

- m 変量の確率変数の分散共分散行列

$$\mathcal{D}(\mathbf{X}) = \{\text{cov}(X_i, X_j)\} \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, m \quad (209)$$

$$= \mathcal{E}(\mathbf{X}\mathbf{X}') - \mathcal{E}(\mathbf{X})\mathcal{E}(\mathbf{X})' \quad (210)$$

分散共分散行列は対称で対角要素には分散 $\text{var}(X_j)$ が入り非対角部分には共分散 $\text{cov}(X_i, X_j)$ が入る。同様に $\text{corr}(X_i, X_j)$ を並べた行列を相関行列と呼び、その対角要素は全て 1 となる。

m 個の変数を含む変数群 \mathbf{X} と n 個の変数を含む変数群 \mathbf{Y} の共分散からなる $m \times n$ の行列を

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = \{\text{cov}(X_i, Y_j)\} = \mathcal{E}(\mathbf{X}\mathbf{Y}') - \mathcal{E}(\mathbf{X})\mathcal{E}(\mathbf{Y})' \quad (211)$$

と書く。

A.1.0.2 標本統計量の総和記号を用いた定義

$x_i, i = 1, 2, \dots, N$ を N 人分のテスト X の得点とし、それを縦に並べた $N \times 1$ のベクトルを \mathbf{x} と記す。また、テストが複数ある場合テスト $X_j, j = 1, 2, \dots, m$ の個人 i の得点を x_{ij} と記し、それを $N \times m$ の行列の形に並べたものを \mathbf{X} と記す。また、 \mathbf{X} の第 j 列、すなわち、 N 人分のテスト j の得点を縦に並べたものを $\mathbf{x}_{(j)}$ と記す。

- 期待値

$$\mathcal{E}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N x_i \quad (212)$$

多変量数の場合は

$$\mathcal{E}(\mathbf{X}) = \begin{pmatrix} \mathcal{E}(\mathbf{x}_{(1)}) \\ \mathcal{E}(\mathbf{x}_{(2)}) \\ \vdots \\ \mathcal{E}(\mathbf{x}_{(m)}) \end{pmatrix} \quad (213)$$

ただし、

$$\mathcal{E}(\mathbf{x}_{(j)}) = \frac{1}{N} \sum_{i=1}^N x_{ij} \quad (214)$$

- 分散

$$\text{var}(\mathbf{x}_{(j)}) = \frac{1}{N} \sum_{i=1}^N (x_{ij} - \mathcal{E}(\mathbf{x}_{(j)}))^2 = \frac{1}{N} \sum_{i=1}^N x_{ij}^2 - (\mathcal{E}(\mathbf{x}_{(j)}))^2 \quad (215)$$

- 共分散

$$\text{cov}(\mathbf{x}_{(j)}, \mathbf{x}_{(k)}) = \frac{1}{N} \sum_{i=1}^N (x_{ij} - \mathcal{E}(\mathbf{x}_{(j)})) (x_{ik} - \mathcal{E}(\mathbf{x}_{(k)})) \quad (216)$$

$$= \frac{1}{N} \sum_{i=1}^N x_{ij} x_{ik} - \mathcal{E}(\mathbf{x}_{(j)}) \mathcal{E}(\mathbf{x}_{(k)}) \quad (217)$$

したがって、

$$\text{cov}(\mathbf{x}, \mathbf{x}) = \text{var}(\mathbf{x}) \quad (218)$$

とする。

- 相関係数

$$\text{corr}(\mathbf{x}, \mathbf{y}) = \frac{\text{cov}(\mathbf{x}, \mathbf{y})}{\sqrt{\text{var}(\mathbf{x})} \sqrt{\text{var}(\mathbf{y})}} \quad (219)$$

- m 変量の確率変数の分散共分散行列

$$\mathcal{D}(\mathbf{X}) = \{ \text{cov}(\mathbf{x}_{(j)}, \mathbf{x}_{(k)}) \} \quad j = 1, 2, \dots, m, \quad k = 1, 2, \dots, m \quad (220)$$

A.1.0.3 期待値、分散等の演算子の性質

\mathbf{x} および \mathbf{y} を確率変数、 \mathbf{A} , \mathbf{c} および \mathbf{d} を定数とし、以下の演算子を定める。定義は、密度関数を用いるものでも総和記号を用いるものでもどちらでも良い。

$$\mathcal{E}(\mathbf{x}) = \text{平均ベクトル} \quad \text{and} \quad \mathcal{D}(\mathbf{x}) = \text{分散共分散行列} \quad \text{and} \quad \text{Cov}(\mathbf{x}, \mathbf{y}) = \text{共分散行列} \quad (221)$$

そのとき、

$$\mathcal{E}(\mathbf{c}) = \mathbf{c}, \quad \mathcal{D}(\mathbf{c}) = \mathbf{0}, \quad \text{Cov}(\mathbf{c}, \mathbf{d}) = \mathbf{0}, \quad \text{Cov}(\mathbf{x}, \mathbf{c}) = \mathbf{0} \quad (222)$$

$$\mathcal{E}(\mathbf{x} + \mathbf{y}) = \mathcal{E}(\mathbf{x}) + \mathcal{E}(\mathbf{y}) \quad (223)$$

$$\mathcal{E}(\mathbf{A}\mathbf{x} + \mathbf{c}) = \mathbf{A} \mathcal{E}(\mathbf{x}) + \mathcal{E}(\mathbf{c}) = \mathbf{A} \mathcal{E}(\mathbf{x}) + \mathbf{c} \quad (224)$$

$$\mathcal{D}(\mathbf{x} + \mathbf{y}) = \mathcal{D}(\mathbf{x}) + 2\text{Cov}(\mathbf{x}, \mathbf{y}) + \mathcal{D}(\mathbf{y}) \quad (225)$$

$$\mathcal{D}(\mathbf{A}\mathbf{x} + \mathbf{c}) = \mathcal{D}(\mathbf{A}\mathbf{x}) + 2\text{Cov}(\mathbf{A}\mathbf{x}, \mathbf{c}) + \mathcal{D}(\mathbf{c}) \quad (226)$$

$$= \mathbf{A} \mathcal{D}(\mathbf{x}) \mathbf{A}' \quad (227)$$

$$\mathcal{D}(\mathbf{x}) = \mathcal{E}(\mathbf{x}\mathbf{x}') - \mathcal{E}(\mathbf{x})\mathcal{E}(\mathbf{x})' \quad (228)$$

$$\text{Cov}(\mathbf{x} + \mathbf{y}, \mathbf{u} + \mathbf{v}) = \text{Cov}(\mathbf{x}, \mathbf{u}) + \text{Cov}(\mathbf{x}, \mathbf{v}) + \text{Cov}(\mathbf{y}, \mathbf{u}) + \text{Cov}(\mathbf{y}, \mathbf{v}) \quad (229)$$

$$\begin{aligned} \text{Cov}(\mathbf{A}\mathbf{x} + \mathbf{c}, \mathbf{B}\mathbf{y} + \mathbf{d}) &= \text{Cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) + \text{Cov}(\mathbf{A}\mathbf{x}, \mathbf{d}) + \text{Cov}(\mathbf{c}, \mathbf{B}\mathbf{y}) + \text{Cov}(\mathbf{c}, \mathbf{d}) \\ &= \mathbf{A} \text{Cov}(\mathbf{x}, \mathbf{y}) \mathbf{B}' \end{aligned} \quad (230)$$

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = \mathcal{E}(\mathbf{x}\mathbf{y}') - \mathcal{E}(\mathbf{x})\mathcal{E}(\mathbf{y})' = \text{Cov}(\mathbf{y}, \mathbf{x})' \quad (231)$$

$$\text{Cov}(\mathbf{x}, \mathbf{x}) = \mathcal{D}(\mathbf{x}) \quad (232)$$

$$\text{Corr}(\mathbf{x}, \mathbf{y}) = (\text{diag}(\mathcal{D}(\mathbf{x})))^{-\frac{1}{2}} \text{Cov}(\mathbf{x}, \mathbf{y}) (\text{diag}(\mathcal{D}(\mathbf{y})))^{-\frac{1}{2}} \quad (233)$$

である。ただし、各変数がスカラーの場合は

$$\mathcal{D}(X) = \text{var}(X), \quad \mathcal{Cov}(X, Y) = \text{cov}(x, y), \quad \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} \quad (234)$$

と読み替える。

A.1.0.4 確率変数の和

いくつかの確率変数の和 $\sum_{j=1}^m X_j$ の平均と分散は、

$$\mathbf{A} = \{1, 1, \dots, 1\} \quad (235)$$

を m 個の 1 を横に並べたものとすれば、

$$\sum_{j=1}^m X_j = \mathbf{A}\mathbf{x} \quad (236)$$

であるため、(224) 式と (227) 式より

$$\mathcal{E}\left(\sum_{j=1}^m X_j\right) = \mathcal{E}(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathcal{E}(\mathbf{x}) = \sum_{j=1}^m [\mathcal{E}(\mathbf{x})]_j \quad (237)$$

$$\text{var}\left(\sum_{j=1}^m X_j\right) = \mathcal{D}(\mathbf{A}\mathbf{x}) = \mathbf{A}\mathcal{D}(\mathbf{x})\mathbf{A}' = \sum_{i=1}^m \sum_{j=1}^m [\mathcal{D}(\mathbf{x})]_{ij} \quad (238)$$

したがって、

$$\mathcal{E}\left(\sum_{j=1}^m X_j\right) = \sum_{j=1}^m \mathcal{E}(X_j) \quad (239)$$

$$\text{var}\left(\sum_{j=1}^m X_j\right) = \sum_{j=1}^m \sum_{k=1}^m \text{cov}(X_j, X_k) = \sum_{j=1}^m \text{var}(X_j) + \sum_{j \neq k} \text{cov}(X_j, X_k) \quad (240)$$

$$(241)$$

となる。重み付き和は

$$\mathbf{A} = \{w_1, w_2, \dots, w_m\} \quad (242)$$

とすればよく

$$\mathcal{E}\left(\sum_{j=1}^m w_j X_j\right) = \sum_{j=1}^m w_j \mathcal{E}(X_j) \quad (243)$$

$$\text{var}\left(\sum_{j=1}^m w_j X_j\right) = \sum_{j=1}^m \sum_{k=1}^m w_j w_k \text{cov}(X_j, X_k) = \sum_{j=1}^m w_j^2 \text{var}(X_j) + \sum_{j \neq k} w_j w_k \text{cov}(X_j, X_k) \quad (244)$$

となる。

また、確率変数ベクトル \mathbf{x} とそれに対応する平均や分散共分散行列が以下のように

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad (245)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \end{bmatrix} \quad (246)$$

に分割されている場合、 \mathbf{Ax}

$$\mathbf{Ax} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 \\ \mathbf{A}_{21}\mathbf{x}_1 + \mathbf{A}_{22}\mathbf{x}_2 \end{pmatrix} \quad (247)$$

の分布は、(227) 式 により

$$\mathcal{E}(\mathbf{Ax}) = \mathbf{A}\boldsymbol{\mu} \quad \text{and} \quad \mathcal{D}(\mathbf{Ax}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' \quad (248)$$

となるが、 $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ 行列を $\mathbf{S} = (\mathbf{S}_{11}, \mathbf{S}_{12})$ と置くと、

$$\mathbf{S}_{11} = \mathbf{A}_{11}\boldsymbol{\Sigma}_{11}\mathbf{A}'_{11} + \mathbf{A}_{11}\boldsymbol{\Sigma}_{12}\mathbf{A}'_{12} + \mathbf{A}_{12}\boldsymbol{\Sigma}_{21}\mathbf{A}'_{11} + \mathbf{A}_{12}\boldsymbol{\Sigma}_{22}\mathbf{A}'_{12} \quad (249)$$

$$\mathbf{S}_{12} = \mathbf{A}_{11}\boldsymbol{\Sigma}_{11}\mathbf{A}'_{21} + \mathbf{A}_{11}\boldsymbol{\Sigma}_{12}\mathbf{A}'_{22} + \mathbf{A}_{12}\boldsymbol{\Sigma}_{21}\mathbf{A}'_{21} + \mathbf{A}_{12}\boldsymbol{\Sigma}_{22}\mathbf{A}'_{22} \quad (250)$$

である。

A.1.0.5 分散の計算

分散・共分散の計算 1

$$\text{var}(X) = \mathcal{E}(X^2) - \mathcal{E}(X)^2 \quad (251)$$

$$\text{cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) \quad (252)$$

証明 $\mathcal{E}(X)$ と $\mathcal{E}(Y)$ には確率変数 X と Y は含まれないことに注意すると

$$\text{cov}(X, Y) = \int_X \int_Y (x - \mathcal{E}(X))(y - \mathcal{E}(Y)) f(x, y) dy dx \quad (253)$$

$$= \int_X \int_Y (xy - \mathcal{E}(X)y - x\mathcal{E}(Y) + \mathcal{E}(X)\mathcal{E}(Y)) f(x, y) dy dx \quad (254)$$

$$= \int_X \int_Y xy f(x, y) dy dx - \mathcal{E}(X) \int_X \int_Y y f(x, y) dy dx \\ - \mathcal{E}(Y) \int_X \int_Y x f(x, y) dy dx + \mathcal{E}(X)\mathcal{E}(Y) \int_X \int_Y f(x, y) dy dx \quad (255)$$

$$= \int_X \int_Y xy f(x, y) dy dx - \mathcal{E}(X) \int_Y y \int_X f(x, y) dx dy - \mathcal{E}(Y) \int_X x \int_Y f(x, y) dy dx \\ + \mathcal{E}(X)\mathcal{E}(Y) \int_X \int_Y f(x, y) dy dz = \mathcal{E}(XY) - \mathcal{E}(X)\mathcal{E}(Y) \quad (256)$$

証明終わり

条件付き期待値と分散・共分散の計算 2

$$\mathcal{E}(X) = \mathcal{E}(\mathcal{E}(X|Y)) \quad (257)$$

$$\text{var}(X) = \mathcal{E}(\text{var}(X|Y)) + \text{var}(\mathcal{E}(X|Y)) \quad (258)$$

証明

$$\mathcal{E}(X) = \int_Y \int_X x f(x, y) dx dy \quad (259)$$

$$= \int_Y \int_X x f(x|y) f(y) dx dy = \int_Y \int_X x f(x|y) dx f(y) dy \quad (260)$$

$$= \int_Y \mathcal{E}(X|y) f(y) dy = \mathcal{E}(\mathcal{E}(X|Y)) \quad (261)$$

$$\text{var}(X) = \mathcal{E}(X^2) - \mathcal{E}(X)^2 = \mathcal{E}(\mathcal{E}(X^2|Y)) - \mathcal{E}(\mathcal{E}(X|Y))^2 \quad (262)$$

$$= \mathcal{E}(\text{var}(X|Y) + \mathcal{E}(X|Y)^2) - \mathcal{E}(\mathcal{E}(X|Y))^2 \quad (263)$$

$$= \mathcal{E}(\text{var}(X|Y)) + \mathcal{E}(\mathcal{E}(X|Y)^2) - \mathcal{E}(\mathcal{E}(X|Y))^2 \quad (264)$$

$$= \mathcal{E}(\text{var}(X|Y)) + \text{var}(\mathcal{E}(X|Y)) \quad (265)$$

証明終わり

したがって以下の関係を得る。

$$\mathcal{E}(\text{var}(X|Y)) = \text{var}(X) - \text{var}(\mathcal{E}(X|Y)) \leq \text{var}(X) \quad (266)$$

多変量の場合は、

$$\mathcal{E}(X) = \mathcal{E}(\mathcal{E}(X|Y)) \quad (267)$$

$$\mathcal{D}(X) = \mathcal{E}(\mathcal{D}(X|Y)) + \mathcal{D}(\mathcal{E}(X|Y)) \quad (268)$$

証明

$$\mathcal{D}(X) = \mathcal{E}(XX') - \mathcal{E}(X)\mathcal{E}(X)' \quad (269)$$

$$= \mathcal{E}(\mathcal{E}(XX'|Y)) - \mathcal{E}(\mathcal{E}(X|Y))\mathcal{E}(\mathcal{E}(X|Y))' \quad (270)$$

$$= \mathcal{E}(\mathcal{D}(X|Y) + \mathcal{E}(X|Y)\mathcal{E}(X|Y)') - \mathcal{E}(\mathcal{E}(X|Y))\mathcal{E}(\mathcal{E}(X|Y))' \quad (271)$$

$$= \mathcal{E}(\mathcal{D}(X|Y)) + \mathcal{E}(\mathcal{E}(X|Y)\mathcal{E}(X|Y)') - \mathcal{E}(\mathcal{E}(X|Y))\mathcal{E}(\mathcal{E}(X|Y))' \quad (272)$$

$$= \mathcal{E}(\mathcal{D}(X|Y)) + \mathcal{D}(\mathcal{E}(X|Y)) \quad (273)$$

証明終わり

A.2 確率変数間の線形回帰

確率変数間の線形回帰

二つの確率変数 U と V の間に、

$$\mathcal{E}(V|U) = \hat{V} = a + bU \quad (274)$$

という線形の回帰式を仮定した場合の回帰係数 (a, b) の最小2乗推定値は

$$a = \mathcal{E}(V) - b\mathcal{E}(U) \quad \text{and} \quad b = \frac{\text{cov}(U, V)}{\text{var}(U)} \quad (275)$$

で与えられる。また、 V の分散のうち、 \hat{V} で説明される割合（分散説明率、決定係数、重相関係数の2乗）は

$$R^2 = \frac{\text{var}(\hat{V})}{\text{var}(V)} = \text{corr}(U, V)^2 \quad (276)$$

である。

証明

最小2乗基準

$$Q = \mathcal{E}\left(\left(V - \hat{V}\right)^2\right) \quad (277)$$

$$= \mathcal{E}\left(V^2 - 2V\hat{V} + \hat{V}^2\right) \quad (278)$$

$$= \mathcal{E}\left(V^2 - 2aV - 2bUV + a^2 + 2abU + b^2U^2\right) \quad (279)$$

$$= \mathcal{E}\left(V^2\right) - 2a\mathcal{E}(V) - 2b\mathcal{E}(UV) + a^2 + 2ab\mathcal{E}(U) + b^2\mathcal{E}(U^2) \quad (280)$$

のパラメタ a による偏微分

$$\frac{\partial Q}{\partial a} = 2(-\mathcal{E}(V) + a + b\mathcal{E}(U)) \quad (281)$$

を 0 と置くことにより、

$$a = \mathcal{E}(V) - b\mathcal{E}(U) \quad (282)$$

を得る。この関係を最小2乗基準のパラメタ b による偏微分に代入したもの

$$\frac{\partial Q}{\partial b} = 2(-\mathcal{E}(UV) + a \mathcal{E}(U) + b \mathcal{E}(U^2)) \quad (283)$$

$$= 2(-\mathcal{E}(UV) + (\mathcal{E}(V) - b \mathcal{E}(U))\mathcal{E}(U) + b \mathcal{E}(U^2)) \quad (284)$$

$$= 2(-\mathcal{E}(UV) + \mathcal{E}(U)\mathcal{E}(V) + b(-\mathcal{E}(U)^2 + \mathcal{E}(U^2))) \quad (285)$$

$$= 2(-cov(U, V) + b(var(U))) \quad (286)$$

を0と置くことにより

$$b = \frac{cov(U, V)}{var(U)} \quad (287)$$

を得る。

また、 V の分散のうち、 \hat{V} で説明される割合は

$$var(\hat{V}) = var(a + bU) = b^2 var(U) \quad (288)$$

$$= \left(\frac{cov(U, V)}{var(U)}\right)^2 var(U) = corr(U, V)^2 var(V) \quad (289)$$

となるため、その値は

$$R^2 = \frac{var(\hat{V})}{var(V)} = corr(U, V)^2 \quad (290)$$

で与えられる、

証明終わり

A.3 信頼区間について

A.3.0.1 正規近似を用いる方法

母数 θ を持つ分布からのデータ $Y_i, i = 1, 2, \dots, N$ から θ の点推定値となる統計量 S を計算し、それを用いて母数 θ を区間推定することを考える。いま、 S の分布が近似的に

$$S \sim N(\theta, \sigma(\theta)^2) \quad (291)$$

と見なせるような点推定量 S があれば、

$$\Pr(\theta - z_{1-\alpha/2} \sigma(\theta) < S < \theta + z_{1-\alpha/2} \sigma(\theta)) \approx 1 - \alpha \quad (292)$$

であるが、この不等式を同値変形することにより

$$\Pr(S - z_{1-\alpha/2} \sigma(\theta) < \theta < S + z_{1-\alpha/2} \sigma(\theta)) \approx 1 - \alpha \quad (293)$$

を得る。したがって、 $\sigma(\theta)$ を $\sigma(S)$ で推定すると $(S - z_{1-\alpha/2} \sigma(S), S + z_{1-\alpha/2} \sigma(S))$ がおよそその信頼水準 (confidence level) または信頼係数 (confidence coefficient) が $100(1 - \alpha)\%$ の θ の信頼区間と見なせる。

A.3.0.2 Neyman 流の正確な信頼区間

いま、母数 θ を与えたときの統計量 S の条件付き分布に関して、その上位 $100 \alpha/2 \%$ 分位点を θ の関数と見たもの、ならびに下位 $100 \alpha/2 \%$ 分位点を θ の関数と見たものをそれぞれ、 $t_1(\theta)$ 、 $t_2(\theta)$ とする。この関係を表したものが図 A である。(条件付き期待値を $\mathcal{E}(S|\theta)$ で示している。) 定義により、この図において、母数が θ_0 の時に、統計量 S が区間 (S_L, S_U) に入る確率は $1 - \alpha$ である。いま、統計量 S が与えられたときに、 t_1 と t_2 の逆関数 t_1^{-1} および t_2^{-1} を用いて計算される θ の値を $\theta_L(S) = t_1^{-1}(S)$ 、 $\theta_U(S) = t_2^{-1}(S)$ とする。 S が区間 (S_L, S_U) を下から上に移動するとき、それに従って付随する θ の区間 (点線) は (A, B) から (C, D) を経て (E, F) まで移動するが、その際、区間 (A, B) から (E, F) までの間にのみ θ_0 が含まれることになる。したがって、確率変数 S が区間 (S_L, S_U) に含まれるという事象と確率的な区間 $(\theta_L(S), \theta_U(S))$ が θ_0 を含むという事象は同値であり、その確率は $1 - \alpha$ である。この意味で、区間 $(\theta_L(S), \theta_U(S))$ を信頼水準が $100(1 - \alpha)\%$ の Neyman(1937) の信頼区間と呼ぶ。

なお、 S は母数 θ の点推定量である必要は無く、より一般的には、母数を $\xi = h(\theta)$ と変換したものの信頼区間も、 ξ を与えたときの S の分位点を表す関数、 $t_1(\xi)$ 、 $t_2(\xi)$ やその逆関数を用いて求めることが出来る。さらに、 S は純粋な統計量 (データのみの関数) である必要は無く、母数 θ に依存する量 $S(\theta)$ であっても、 $t_1(\theta)$ と $t_2(\theta)$ を $S(\theta)$ の分布の分位点であるとすれば、同様の議論を用いて、 $\theta_L(S(\theta)) = t_1^{-1}(S(\theta))$ 、 $\theta_U(S(\theta)) = t_2^{-1}(S(\theta))$ として信頼水準 $100(1 - \alpha)\%$ 信頼区間を求めることが出来る。

項目反応理論において特性値 θ の信頼区間の導出にこの方法を利用する際には、テスト得点 S が離散量であることに注意が必要であるが、適当に補間することにより、近似的な信頼区間を数値的に求めることが可能である。なお、図 1 の横軸を θ に付け替えたもので示される得点の条件付き分布の分位点を繋いだものが図 A となっている。

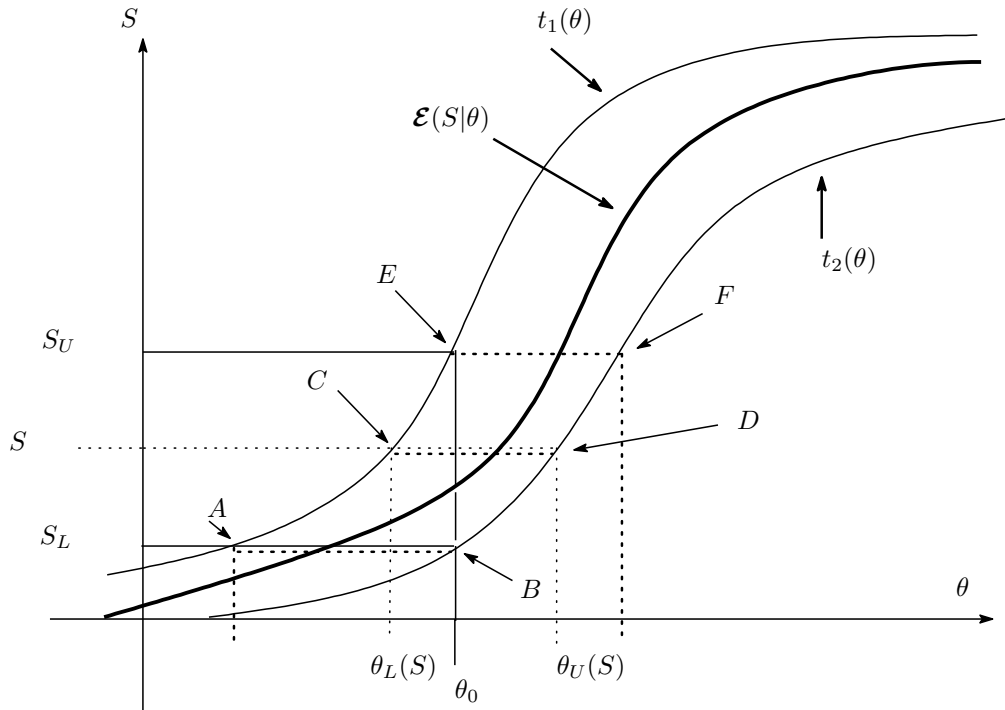


図 A: Conditional Distribution of S given θ and Confidence Interval

A.4 付録：不等式

A.4.1 池田央 (1973) の p83 の (4.101)

$$\text{var}(T_i - T_j) = \text{var}(T_i) + \text{var}(T_j) - 2\text{cov}(T_i, T_j) \geq 0 \quad (294)$$

より

$$\text{var}(T_i) + \text{var}(T_j) \geq 2\text{cov}(T_i, T_j) \quad (295)$$

両辺の $i \neq j$ に関する和をとると

$$\sum_{i=1}^m \sum_{j=1, j \neq i}^m (\text{var}(T_i) + \text{var}(T_j)) \geq \sum_{i=1}^m \sum_{j=1, j \neq i}^m 2\text{cov}(T_i, T_j) \quad (296)$$

左辺は

$$\sum_{i=1}^m \sum_{j=1, j \neq i}^m (\text{var}(T_i) + \text{var}(T_j)) = \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{var}(T_i) + \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{var}(T_j) \quad (297)$$

$$= 2(m-1) \sum_{i=1}^m \text{var}(T_i) \quad (298)$$

したがって

$$(m-1) \sum_{i=1}^m \text{var}(T_i) \geq \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \quad (299)$$

または

$$\sum_{i=1}^m \text{var}(T_i) \geq \frac{1}{m-1} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \quad (300)$$

または

$$\frac{1}{m} \sum_{i=1}^m \text{var}(T_i) \geq \frac{1}{m(m-1)} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \quad (301)$$

A.4.2 池田央 (1973) の p83 の (4.102)

(300) 式の両辺に $\sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j)$ を加えると

$$\begin{aligned} \sum_{i=1}^m \text{var}(T_i) + \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) &\geq \frac{1}{m-1} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) + \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \\ &= \frac{m}{m-1} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \end{aligned} \quad (302)$$

となるが、左辺は $\text{var}(\sum_{i=1}^m T_i)$ であるため

$$\text{var}(\sum_{i=1}^m T_i) \geq \frac{m}{m-1} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \quad (303)$$

または

$$\text{var}\left(\frac{1}{m} \sum_{i=1}^m T_i\right) \geq \frac{1}{m(m-1)} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \text{cov}(T_i, T_j) \quad (304)$$

を得る。

A.4.3 コーシー・シュヴァルツの不等式

$$\left(\sum_{k=1}^n X_k^2\right) \left(\sum_{k=1}^n Y_k^2\right) \geq \left(\sum_{k=1}^n X_k Y_k\right)^2 \quad (305)$$

$$\text{var}(X)\text{var}(Y) \geq \text{cov}(X, Y)^2 \quad (306)$$

$$\left(\sqrt{\text{var}(X)} + \sqrt{\text{var}(Y)}\right)^2 \geq \text{var}(X+Y) \quad (307)$$

A.4.4 相加平均と相乗平均

$X_k \geq 0, k = 1, 2$ の場合

$$\frac{1}{2}(X_1 + X_2) \geq \sqrt{X_1 X_2} \quad (308)$$

$X_k \geq 0, k = 1, 2, \dots, n$ の場合

$$\frac{1}{n} \sum_{k=1}^n X_k \geq \left(\prod_{k=1}^n X_k\right)^{1/n} \quad (309)$$

参考文献

- 池田央 (1973) テスト II. 東京大学出版会
- 南風原朝和 (1991) 第 1 章 項目反応理論反応理論の概要 芝祐順編 (1991) 項目反応理論
基礎と応用 東京大学出版会
- 南風原朝和 (2014) 報告 2 入試選抜試験の測定問題 独立行政法人大学入試センター 入学者
選抜試験研究に関する調査室報告書 2 「大学入試の日本的風土は代えられるか」 p61-74.
- Bartlett, M. S. (1947) The Use of Transformations. *Biometrics*, Vol. 3, No.1, pp. 39-52
- Birnbaum, A.(1968) Some Latent Trait Models. In F. M. Lord and M. R. Novick, *Statistical
Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley.
- Feldt, L. S., Steffan, M., & Gupta, N. C. (1985) A comparison of five methods for esti-
mating the standard error of measurement at specific score levels. *Applied Psychological
Measurement*, 9, p351-361.
- Harvill (1991) An NCME Instructional Module on Standard Error of Measurement. *Educa-
tional Measurement: Issues and Practice*, v10 n2 p33-41
- Kelly, T. L. (1947) *Fundamentals of Statistics*. Cambridge: Harvard University Press.
- Lord, F. M. (1959). Tests of the same length do have the same standard error of measure-
ment. *Educational and Psychological Measurement*, 19, 233-239.
- Lord, F. M. (1980). *Applications of Item Response Theory To Practical Testing Problems*.
Lawrence Erlbaum Associates. Mahwah, New Jersey
- Lord, F. M. and Novick, M. R. (1968) *Statistical Theories of Mental Test Scores*. Reading,
Mass.: Addison-Wesley.
- Mayekawa, S. (2017) *lazy.irt: An R package*.
URL: <http://mayekawa.in.coocan.jp/Rpackages.html>
- Mayekawa, S., & Arai, S. (2008) Distribution of the Sum of Scored Multinomial Random
Variables and Its Application to the Item Response Theory. In K. Shigemasu, A. Okada,
T. Imaizumi, & T. Hoshino (Eds.) *New Trends in Psychometrics*. Tokyo: Universal
Academic Press.
- Neyman, J. (1937) Outline of a theory of statistical estimation based on the classical theory
of probability. *Philos. Trans. Roy. Soc. London. Ser. A. , 236 pp. 333-380*
- Novick, M. R. , Jackson, P. H., & Thayer, D. T. (1971) Bayesian Inference and the Classical
Test Theory Model: Reliability and True Scores. *Psychometrika*. vol. 36, No.3.
- Ver Hoef, J.M. (2012) Who Invented the Delta Method? *The American Statistician*. vol.
66, no. 2, pp 124-127