

研究法

古典的テスト理論

1.1 モデル

個人のテスト得点を真値 (true score) と誤差得点に分け、誤差に関して以下のことを仮定する。

$$X_i = T_i + E_i \tag{1}$$

$$\mathcal{E}(E_i) = 0 \text{ 誤差のバイアス無し。} \tag{2}$$

$$\text{var}(E_i) = \sigma_E^2 \text{ 誤差分散は個人に依存しない。} \tag{3}$$

この仮定は、テスト得点の条件付き分布に以下のことを仮定すること同値である。

$$\mathcal{E}(X_i | T_i) = T_i \quad \text{and} \quad \text{var}(X_i | T_i) = \sigma_E^2 \tag{4}$$

この σ_E のことを測定の標準誤差 (standard error of measurement) と呼ぶ。

古典的テスト理論

古典的テスト理論

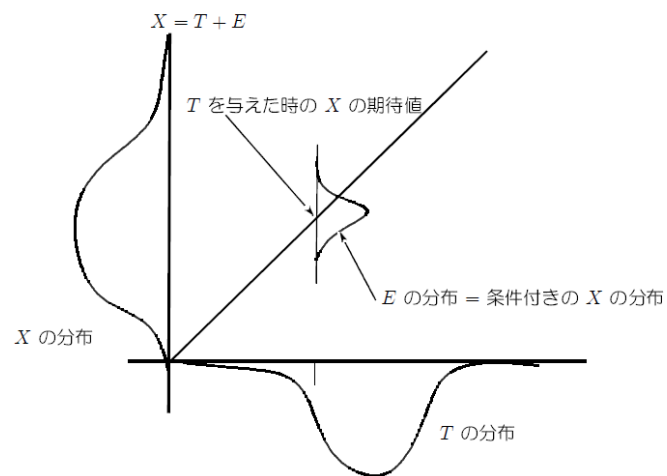


図 1: Classical Test Theory

ある集団における真値の分布として

$$\mathcal{E}(T_i) = \mu \tag{5}$$

$$\text{var}(T_i) = \sigma_T^2 \tag{6}$$

$$\text{cov}(E_i, T_i) = 0 \text{ (真値と誤差は無相関)} \tag{7}$$

を仮定する。すると、その集団におけるテスト得点の (周辺) 分布は

$$\mathcal{E}(X_i) = \mu \tag{8}$$

$$\text{var}(X_i) = \sigma_X^2 = \sigma_T^2 + \sigma_E^2 \tag{9}$$

となるが、これは、 $X_i, i = 1, 2, \dots, N$ が iid (所謂ランダムサンプル) でその平均は μ 、分散は σ_X^2 であることを示している。従って、以下では個人を表す添え字 i を省略して (X, T, E) を用いる。すなわち、

$$X = T + E \tag{10}$$

$$\mathcal{E}(E) = 0 \quad \text{and} \quad \text{var}(E) = \sigma_E^2 \quad \text{and} \quad \text{cov}(T, E) = 0 \tag{11}$$

$$\mathcal{E}(X|T) = T \quad \text{and} \quad \text{var}(X|T) = \sigma_E^2 \tag{12}$$

$$\mathcal{E}(X) = \mathcal{E}(T) = \mu \tag{13}$$

$$\text{var}(X) = \text{var}(T) + \text{var}(E) \quad \text{or} \quad \sigma_X^2 = \sigma_T^2 + \sigma_E^2 \text{ (ANOVA decomposition)} \tag{14}$$

である。

テストの信頼性

1.2 テストの信頼性係数

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

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} = \frac{\sigma_T^2}{\sigma_X^2} \quad (15)$$

と定義される。またこれは、(14) 式を用いると

$$\rho(X) = 1 - \frac{\text{var}(E)}{\text{var}(X)} = 1 - \frac{\sigma_E^2}{\sigma_X^2} \quad (16)$$

$$(17)$$

と表現される。いずれにしても信頼性係数は

$$0 \leq \rho(X) \leq 1 \quad (18)$$

5

テストの信頼性

また、古典的テストのモデルから

$$\text{cov}(X, T) = \text{cov}(T + E, T) = \text{cov}(T, T) + \text{cov}(E, T) = \text{var}(T) + 0 = \sigma_T^2 \quad (19)$$

したがって、

$$\text{corr}(X, T) = \frac{\text{cov}(X, T)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(T)}} = \frac{\sigma_T^2}{\sqrt{\sigma_X^2}\sqrt{\sigma_T^2}} = \frac{\sqrt{\sigma_T^2}}{\sqrt{\sigma_X^2}} \quad (20)$$

であるため、

$$\rho(X) = \frac{\text{var}(T)}{\text{var}(X)} = \frac{\sigma_T^2}{\sigma_X^2} = (\text{corr}(X, T))^2 \quad (21)$$

とも書ける。

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、

$$\text{var}(T) = \rho(X)\text{var}(X) \quad \text{and} \quad \text{var}(E) = (1 - \rho(X))\text{var}(X) \quad (22)$$

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。

6

テストの信頼性

1.3 複数のテストと真値の間の相関

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。 (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、

$$X_j = T_j + E_j \quad (23)$$

$$\mathcal{E}(E_j) = 0 \quad \text{and} \quad \text{var}(E_j) = \sigma_{E_j}^2 \quad (24)$$

$$\mathcal{E}(X_j) = \mathcal{E}(T_j) = \mu_j \quad (25)$$

$$\text{var}(X_j) = \text{var}(T_j) + \text{var}(E_j) \quad (26)$$

$$\rho(X_j) = \frac{\text{var}(T_j)}{\text{var}(X_j)} = 1 - \frac{\text{var}(E_j)}{\text{var}(X_j)} \quad (27)$$

また、

$$\text{cov}(E_j, E_k) = 0, \quad j \neq k, \quad \text{and} \quad \text{cov}(T_j, E_k) = 0, \quad j, k = 1, 2, \dots, m \quad (28)$$

であるとする。

この場合、 $j \neq k$ に関して、

$$\text{cov}(X_j, X_k) = \text{cov}(T_j + E_j, T_k + E_k) \quad (29)$$

$$= \text{cov}(T_j, T_k) + \text{cov}(T_j, E_k) + \text{cov}(E_j, T_k) + \text{cov}(E_j, E_k) \quad (30)$$

$$= \text{cov}(T_j, T_k) \quad (31)$$

である。なお、 $j = k$ の場合を含めると

$$\text{cov}(X_j, X_k) = \begin{cases} \text{var}(X_j) = \text{var}(T_j) + \text{var}(E_j) & \text{if } j = k \\ \text{cov}(T_j, T_k) & \text{if } j \neq k \end{cases} \quad (32)$$

7

テストの信頼性

希薄化の修正公式 correction formula for attenuation

$$\text{corr}(T_j, T_k) = \frac{\text{cov}(T_k, T_j)}{\sqrt{\text{var}(T_j)}\sqrt{\text{var}(T_k)}} \quad (37)$$

$$= \frac{\text{std}(X_j)\text{std}(X_k)\text{corr}(X_j, X_k)}{\sqrt{\rho(X_j)}\text{std}(X_j)\sqrt{\rho(X_k)}\text{std}(X_k)} \quad (38)$$

$$= \frac{\text{corr}(X_j, X_k)}{\sqrt{\rho(X_j)}\rho(X_k)} \geq \text{corr}(X_j, X_k) \quad (39)$$

$$\text{corr}(X_j, X_k) \leq \frac{\text{corr}(X_j, T_k)}{\text{corr}(T_j, X_k)} \leq \text{corr}(T_j, T_k) \quad (40)$$

$$\text{corr}(X_j, X_k) \leq \sqrt{\rho(X_j)}\sqrt{\rho(X_k)} \leq \sqrt{\rho(X_j)} \quad (42)$$

8

平行テスト

1.5.1 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

$$T_j = T_* \quad \text{and} \quad \text{var}(E_j) = \sigma_E^2, \quad j = 1, 2, \dots, m \quad (67)$$

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。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

$$\mathcal{E}(X_j) = \mathcal{E}(T_* + E_j) = \mathcal{E}(T_*) = \mu \quad (68)$$

$$\text{var}(X_j) = \text{var}(T_* + E_j) = \text{var}(T_*) + \text{var}(E_j) = \sigma_{T_*}^2 + \sigma_E^2, \quad (69)$$

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、

$$\text{cov}(X_j, X_k) = \text{cov}(T_j, T_k) = \text{cov}(T_*, T_*) = \text{var}(T_*) = \sigma_{T_*}^2 \quad (70)$$

$$\text{corr}(X_j, X_k) = \frac{\text{cov}(X_j, X_k)}{\sqrt{\text{var}(X_j)\text{var}(X_k)}} = \frac{\sigma_{T_*}^2}{\sigma_{X_*}^2} = \rho^* \quad (71)$$

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。

テストの合計点の信頼性

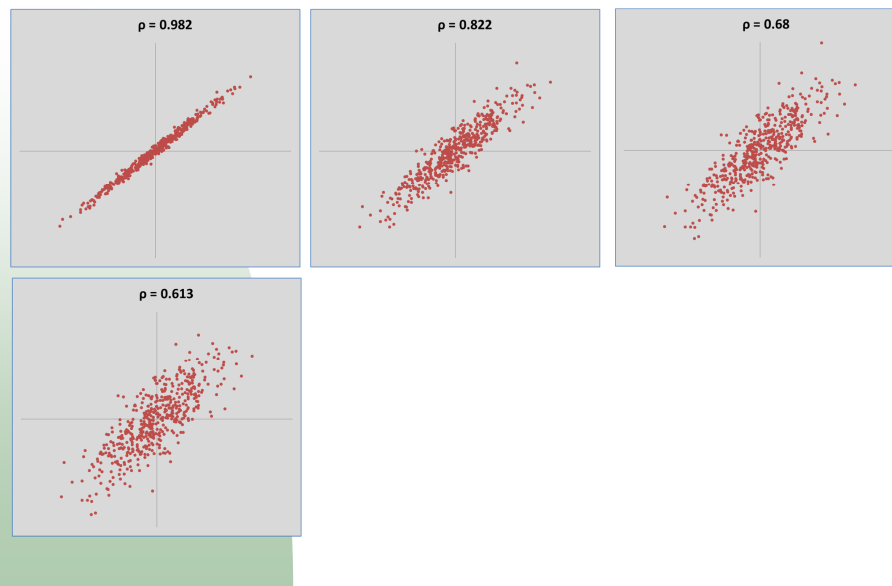
$$\rho(X) = \frac{m\rho^*}{1 + (m-1)\rho^*} \quad \text{Spearman-Brown の公式} \quad (72)$$

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

$$= \frac{m}{m-1} \left(1 - \frac{\sum_{i=1}^m p_j(1-p_j)}{\text{var}(X)} \right) \quad \text{KR20 (Kuder-Richardson の公式 20)}$$

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

信頼性係数



真値の信頼区間

信頼性係数 ρ が既知の場合を考える。

X を観測した時の真値の $100(1-\alpha)\%$ 信頼区間は

$$X \pm z_{(1-\alpha/2)} \sigma_E \quad (9)$$

で与えられる。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かです。」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。(信頼区間なので、区間が変動していることをどう表現するかは難しいけれど。)

また、この信頼区間の幅は

$$w = I_R - I_L = 2 z_{(1-\alpha/2)} \sigma_E \quad (10)$$

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

たとえば、 $\text{var}(X) = 12^2, \rho = 0.75$ 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。

真値の信頼区間

段階評価では、これを1点幅に縮めれば良いので、

$$Y = \frac{(X - \mu_X)}{z_{(1-\alpha/2)}\sigma_E} + \mu_Y \quad (11)$$

という変換を施せば良い。ただし、 μ_Y は任意に選んだ新尺度の平均点である。

この新しい尺度において Y 点を得た時の真値の信頼区間は $(Y - 1, Y + 1)$ となる。したがって、「貴方の得点は Y 点ですが、貴方の真の得点は、ほぼ、 $Y - 1$ 点と $Y + 1$ 点の間に入っている事は確かです。」という言い方が可能となる。

なお、(11) 式の変換により、 Y に含まれる誤差部分の分散が $1/z_{(1-\alpha/2)}^2$ となるが、 Y の分散そのものは

$$\text{var}(Y) = \frac{\sigma_X^2}{(z_{(1-\alpha/2)}\sigma_E)^2} = \frac{1}{z_{(1-\alpha/2)}^2 (1 - \rho)} \quad (12)$$

となる。(当然であるが、信頼性係数は不変である。)

段階数の目安

$$4 \times \sqrt{\text{var}(Y)} = \frac{4}{z_{(1-\alpha/2)} \sqrt{(1 - \rho)}}$$