

Rex. B. Kline (2004)

Beyond significance testing: Reforming data analysis methods in behavioral research,

APA

Chapter5

---

---

## Nonparametric Effect Size Indexes

---

---

Rep.脇本 竜太郎<sup>1</sup>

### INTRODUCTION

- 結果変数は連続変数でなく、離散変数であることがある。離散的結果変数の水準は相互排他的(=ケースはいずれか1つの水準にのみ分類される)
- 離散的結果変数のノンパラメトリックな効果量指標を本章では紹介。これは、医学、疫学、遺伝学でよく使用されるもので、メタ分析でも頻繁に検討されている。
- 離散的結果変数は、対数線形モデルやロジスティック回帰分析等の手法でも分析可能  
→ノンパラメトリックな効果量指標を背後にある統計モデルに基づかせるという意味合い
- しかし、後述する方法で算出したノンパラメトリック効果量指標は記述統計として看做すべきである。

### CATEGORICAL OUTCOMES

- 最も単純な離散変数は二値変数(ex 再発あり・再発なし)
  - ・2群が2値変数について比較されるとき、データは $2 \times 2$ の表で表される
- 離散変数は2つ以上のカテゴリを持ちうる(ex 賛成、反対、わからない)
  - ・このような変数について2群が比較される時、表は $2 \times 2$ より大きくなる。  
→ $2 \times 2$ の時の効果量指標の一部しか適用できない。適用できる方法については、3群以上を有する離散変数について比較する場合にも適用可能。
- 離散変数の水準には、非順序的(unordered: 人種、婚姻状態等)なものと同順序的(ordered; 少し賛成—賛成—強く賛成などのリッカート尺度)なものがある
  - ・順序的な離散変数に特化した方法もあるが、それは非順序的変数についてのものよりも発展しておらず、また知られていない。  
→順序的な変数を間隔尺度と看做してパラメトリックな手法を適用する、水準を崩して臨床的に意味がある相互排他的な2値変数に再編し、 $2 \times 2$ の方法を適用するといった方法が存在。
- 他の枠組み: 感度(sensitivity), 特異度(specificity), 適中率(predictive value)モデル
  - ・医学領域で、病気のスクリーニングテストの正確さを評価するための方法としてよく知られ

---

<sup>1</sup> 日本学術振興会・東京大学

ているもの。…検査結果（陽性・陰性）×実際の状態(病気あり・病気なし)  
 →うつや学習障害などの心理検査にも適用可能であり，その応用は益多きこと。

### EFFECT SIZE INDEXES FOR 2×2 TABLES

Parameters：2つの異なる母集団での相対的リスク<sup>2</sup>の程度を反映する4つのパラメタを紹介。

■処置群と統制群で，再発ありとなしを比較する研究を想定。各パラメタを以下のように定義  
 処置群で再発； $\pi_T$ ，処置群で非再発； $1-\pi_T$ ，統制群で再発； $\pi_C$  統制群で非再発  $1-\pi_C$

■母集団リスク比(risk ratio, 或いは rate ratio)； $\pi_C/\pi_T$

→2つの母集団での，再発リスクの比率の差を示している

Ex.  $\pi_C/\pi_T=1.3$ …統制群の再発リスクは処置群の1.3倍

■母集団オッズ比( $\omega$ )； $\Omega_C/\Omega_T$

・統制群での再発オッズ； $\Omega_C=\pi_C/1-\pi_C$

Ex.  $\pi_C=0.60$ ， $\Omega_C=1.50$ なら，統制群で再発する可能性は再発しない可能性の1.5倍

・処置群での再発オッズ； $\Omega_T=\pi_T/1-\pi_T$

Ex.  $\pi_T=0.40$ ， $\Omega_T=0.67$ なら，処置群で再発する可能性は再発しない可能性の0.67倍

・このとき母集団オッズ比 $\omega=1.50/0.67=2.25$

→再発のオッズ比は統制群で処置群の2.25倍

■母集団相関

$$\phi = \frac{\pi_{CR}\pi_{TNR} - \pi_{CNR}\pi_{TR}}{\sqrt{\pi_C \cdot \pi_T \cdot \pi_{\cdot R} \cdot \pi_{\cdot NR}}}$$

・分子は2×2で有り得る組み合わせを示しており，その合計は1.

・ $\pi_C$ と $\pi_T$ ， $\pi_{\cdot R}$ ， $\pi_{\cdot NR}$ はそれぞれ周辺比率。

### Statistics and Evaluation

■Table5.1は処置あり・なしと再発有・無しの2×2の表で，各セルの観測度数がアルファベットで表されている。一方，Table5.2では，前節で紹介したパラメタの標本推定量を示している。

TABLE 5.1  
 A Fourfold Table for an Observed Group Contrast on a Dichotomy

Group	Relapsed	Not relapsed
Control	A	B
Treatment	C	D

Note. The letters A–D represent observed cell frequencies.

<sup>2</sup> 再発等のように，1つの水準が望ましくない結果である場合の呼称。望ましさに違いがない水準を比べる場合，単純に比率の差。

TABLE 5.2  
Definitions of Effect Size Statistics for 2 × 2 Contingency Tables

Parameter	Statistic	Equation
Proportions of undesirable outcome		
$\pi_C$	$p_C$	$A/(A+B)$
$\pi_T$	$p_T$	$C/(C+D)$
Comparative risk		
$\pi_C - \pi_T$	RD	$p_C - p_T$
$\pi_C/\pi_T$	RR	$p_C/p_T = \frac{A(A+B)}{C(C+D)}$
$\omega = \Omega_C/\Omega_T$	OR	$\frac{p_C}{p_T} = \frac{p_C(1-p_C)}{p_T(1-p_T)} = \frac{A/B}{C/D} = AD/BC$
Measure of association		
$\phi$	$\phi$	$\frac{AD-BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} = \sqrt{\chi^2(1)/N}$

Notes. RD = risk difference; RR = risk ratio; OR = odds ratio. The letters A-D represent observed cell frequencies in Table 5.1. If A, B, C, or D = 0 in the computation of OR, add .5 to all cells.

■ 標本比率差(RD)は母集団比率差の推定量。

- ・ 解釈はしやすいが、重大な限界あり。それは、RD の分散<sup>3</sup>が  $\pi_C$  と  $\pi_T$  の値に依存すること。  
→ それぞれの比率が .5 に近くなるほど分散は大きくなる。  
→  $\pi_C$  と  $\pi_T$  の値のが大きく違うような異なる研究間では、RD の値は比較可能ではない。

■ 標本リスク比(RR)は、再発の観測比率の処置あり群と統制群の差を示す。

- ・ これもまた解釈が容易な指標であるが、やはり欠点がある、
- ・ 分子にある群のリスクが相対的に小さいことを示すのは値が 0-1.0 の限られた範囲だが、分母にある群のリスクが相対的に小さいを示す範囲は 1.0 から無限大  
→ RR の可能な値の範囲は分母に依存  
Ex.  $p_T=0.4$  のサンプルでは、RR の範囲は 0-2.50、一方  $p_T=0.6$  のサンプルでは、RR の範囲は 0-1.67  
…異なるサンプルの結果を比較する標準的指標としての価値は制限される。

■ 標本オッズ比(OR)は群内オッズの比。

- ・ 全ての周辺度数が等しい 2×2 の表で、 $OR=RR^2$
- ・ RR に関する範囲の問題は、OR でも同様。この問題は OR を対数変換し、その結果を真数変換することで対処できる(RR の場合も同様)

■ OR が便利なのは、2×2 表の標準化平均値差の指標として知られるロジット d に変換できること。

- ・ ロジット =  $\ln(OR)$  \*OR の自然対数
- ・ ロジスティック分布は近似的に正規分布であり、標準偏差  $pi/3^{1/2} \doteq 1.8138$
- ・ ロジット  $d = \frac{\ln(OR)}{pi/\sqrt{3}} = \frac{\ln(o_C) - \ln(o_T)}{pi/\sqrt{3}}$  ; これは同じ群の連続的結果変数の比較の、標準化された平均値差と直接的に比較可能。

<sup>3</sup> 原文では range だけど、範囲ではなく分散のことだと思う。

Ex.  $p_c=.60$ ,  $p_r=.20$  であるとき,  $o_c=1.50$ ,  $o_r=.25$ , そして  $OR=6.00$ 。このとき, グループの対比を示す  $d$  の値は,  $\frac{\ln(6)}{1.8138} = \frac{\ln(1.50) - \ln(.25)}{1.8131} = .9878$

\*Shadish et al (1999)の ES program は自動的に 2 値変数のロジット  $d$  を計算。また, 逆正弦変換やプロビット変換など 2 値変数の調節の他の方法については, Lipsey & Wilson(2000, pp.52-58)を参照。

■ 標本オッズ比は今まで見た指標の中で, 直観的には最もわかりにくいもの, しかし, それは統計的に最も良い性質を持っている(特に病気の危険因子の疫学研究にとって)。

→OR は予測研究(暴露・非暴露母集団からランダムサンプリングを行う)にも回顧研究(病気の有無で被験者を分類するような研究)にも用いることができる

→RR は回顧的な研究には使用できないし,  $\hat{\phi}$  はランダムサンプリングを行っていない研究では有効でない。

■  $\phi$  の標本推定量 ;  $\hat{\phi}$

・ 2 値変数が 0 と 1 でコーディングされている場合, 通常の  $r$  の式で計算可能であるが, 各セルの度数から直接求めることも可能であり, そのほうが便利かもしれない。

・ 値は-1 から 1 だが, 符号はセルをどう配置するかで決まるので, 絶対値を報告する研究者は多い。但し,  $2 \times 2$  表の効果は方向があるものを頭に入れておく必要がある(ex 再発・非再発等)。

・ また,  $\hat{\phi}$  の絶対値は自由度 1 の  $\chi^2$  乗値をサンプルサイズで割ったものの平方根と同じ値。

・  $\hat{\phi}^2$  は群要因で説明される分散の比率。

・ 行・列の周辺度数が等しい場合,  $RD = \hat{\phi}$

■  $\hat{\phi}$  の問題点

・ margin-bound であること

… 行と列の周辺度数が等しい時にのみ, 最大の絶対値 1 をとる

… バランスドデザインの下で, 理論的最大値 1 は周辺度数が 0.5 の時のみ可能。0.5 から離れるほど最大値は 0 に近づく, … セル度数が異なれば最大値が異なる(2 値以上のものから 2 カテゴリーを作る時, どこで区分するかで最大値が異なる)

→標準効果量指標としては不適

・ 行と列を対象に扱っている(独立変数と結果変数の区別がない)→行列を入れ替えても値は同じ。予測変数と結果変数を区別し, 非対称に扱う指標も存在→Darlington(1996)を参照

Interval Estimation

■ Table5-3.に、大標本の下での各統計量の推定された標準誤差(漸近法による)を表示。

**TABLE 5.3**  
**Asymptotic Standard Errors for Sample Proportions**

Statistic	Standard error
<b>Proportions of undesirable outcome</b>	
$p_c$	$\sqrt{\frac{p_c(1-p_c)}{n_c}}$
$p_r$	$\sqrt{\frac{p_r(1-p_r)}{n_r}}$
<b>Comparative risk</b>	
RD	$\sqrt{\frac{p_c(1-p_c)}{n_c} + \frac{p_r(1-p_r)}{n_r}}$
ln (RR)	$\sqrt{\frac{1-p_c}{n_c p_c} + \frac{1-p_r}{n_r p_r}}$
ln (OR)	$\sqrt{\frac{1}{n_c p_c (1-p_c)} + \frac{1}{n_r p_r (1-p_r)}}$

Note: RD = risk difference, RR = risk ratio, and OR = odds ratio, and ln = natural log.

- $\hat{\phi}$  の標準誤差は極めて複雑なので、関心のある人は Fleiss(1994, p.249)を参照
- 統計量の  $1-\alpha\%$ 信頼区間は、各標準偏差と z 分布の両側確率  $\alpha$  の臨界値との積。

Ex.

- 2 群のサンプル数が 100 ずつ,  $p_c=.60$ ,  $p_r=.40$  RD=.20 の時,  
RD の SD=.0693, RD の 95%信頼区間  $.20-.693(1.96) \leq RD \leq .20+.693(1.96)$   
.06 $\leq$ RD $\leq$ .34 $\rightarrow$ RD=.20 は母集団リスク差.06 $\sim$ .34 と信頼水準 95% で整合的
- RR=.60/.40=1.50, OR=(.60/.40)/(.40/.60)=2.25 RR と OR そのものは一般的に正規分布に従わないが、対数変換したものは近似的に正規分布に従う。それゆえ、対数に基づくものに信頼区間変換(第 2 章参照)をほどこして信頼区間を求めることができる。  
ln(2.25)=.8109, ln (OR) の SD=.2887 95%信頼限界 .8109 $\pm$ .2887(1.96)  
95%信頼区間 .2450 $\leq$ RR $\leq$ 1.3768  
これをもとの尺度に戻すと  
ln<sup>-1</sup>(.2450)=e<sup>.2450</sup>=1.2776 ln<sup>-1</sup>(1.3768)=e<sup>1.3768</sup>=3.9622  
 $\rightarrow$ OR=2.25 は信頼水準 95%で $\omega=1.28\sim 3.96$  と整合的

**EFFECT SIZE ESTIMATION FOR LARGER TWO-WAY TABLES**

- 水準あるいは群が 3 以上ある時、クロス集計表は 2 $\times$ 2 より大きくなる、
- RD,RR, OR は、どこかを除外するか結合するかして 2 $\times$ 2 に直した時にのみ使用可能
- 2 $\times$ 2 より大きい集計表の、おそらく最も有名な指標は Cramer の V

$$V = \sqrt{\frac{\chi^2(r-1, c-1)}{\min(r-1, c-1) \times N}}$$

r=行数, c=列数, min(r-1, c-1)は r-1, c-1 の小さいほうを選択する意。

$\hat{\rho}$  は  $2 \times 2$  の時の  $V$ 。

- ・但し、 $V$  は 0-1 の値をとるものの相関ではないので、 $V$  の平方を分散説明率と看做すことはできない。また、 $V$  は  $\hat{\rho}$  の一般形なので、同様に margin-bound であり、問題点も共有している。

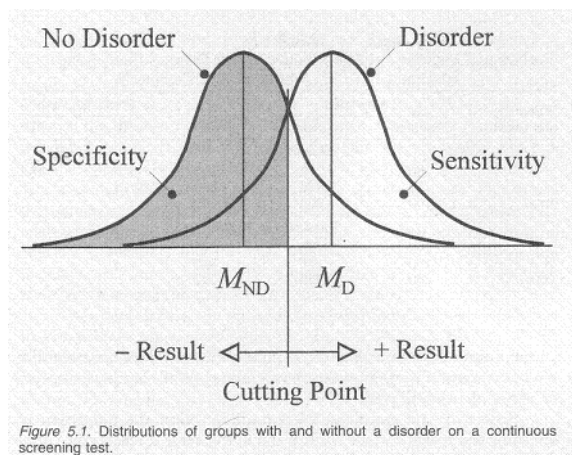
### SENSITIVITY, SPECIFICITY, AND PREDICTIVE VALUE

■ 診断のための確立された診断法(gold standard)と、スクリーニングテスト双方が存在するような病気を考える。

- ・ 確立された診断法は個別化されており相対的に高額(一連の検査など)
- ・ スクリーニングテストは確立された診断法ほど正確ではないが、費用は安くすむ。また、個人ではなく集団に実施することも可能。

■ スクリーニングテストはしばしば連続的な指標(ex. 血中濃度, 質問項目の“あてはまる”の数)

- ・ そのような指標では、陽性と陰性を区分するカットングポイントが存在
- ・ 実際に陰性の群と陽性の群にスクリーニングテストを行った場合、両群の結果の分布は重なる傾向がある→Figure 5.1; 偽陰性(false negative), 偽陽性(false positive)が存在



■ Table 5.4 の上部は、実際の病気の有無とスクリーニングテストの結果の関係を示した 4 分表。各セルのアルファベットは観測頻度。A, D は、有病、無病を正しく診断できているケース、B, C はそれぞれ偽陰性、偽陽性。

■ 同表の下部には、感度 (sensitivity), 特異度 (specificity), 陽性・陰性適中率 (positive/negative predictive value) が定義されている。

- ・ 感度：実際に病気を有しているケースのうち、スクリーニングテスト結果が陽性となる割合  
 $= A / (A + C)$
- ・ 特異度：実際に病気ではないケースのうち、スクリーニングテスト結果が陰性となる割合  
 $= B / (B + D)$

- ・感度と特異度はカッティングポイントによって決まる。ゆえに、カッティングポイントの設定に際しては偽陽性と偽陰性の相対的なコストが考慮される。

**TABLE 5.4**  
Definitions of Sensitivity, Specificity, Predictive Value, and Base Rate

Screening test result	Prediction	True status	
		Disorder	No disorder
+	Disorder	A	B
-	No disorder	C	D
Statistic		Definition	
Sensitivity		$A/(A + C)$	
Specificity		$D/(B + D)$	
Predictive value			
Positive (+PV)		$A/(A + B)$	
Negative (-PV)		$D/(C + D)$	
Base rate		$(A + C)/(A + B + C + D)$	

*Note.* The letters A-D represent observed cell frequencies.

- 感度と特異度は適中率に影響。一般的に、それぞれが高くなれば適中率も高くなる。
  - ・陽性適中率：スクリーニングで陽性とされたケースのうち、実際に有病であるケースの割合
  - ・陰性適中率：スクリーニングで陰性とされたケースのうち、実際に無病であるケースの割合
- さらに適中率はベースレートにも影響を受ける
  - ・Table 5.5 上段はベースレートが 10% の場合の、下段は 75% の場合の陽性・陰性適中率を示している。双方とも感度.80, 特異度.70。
  - ・ベースレート 10% の場合、陰性適中率は.97 と高いが、陽性適中率は.23  
→病気の可能性を排除するには有効だが、病気の存在を確認するには有効ではない。
  - ・ベースレート 75% の場合(寄生虫によって生じる病気だとこれくらい高いことがある)、陽性適中率は.89 で陰性適中率は.54  
→10% の場合とは対照的に、病気の可能性を排除するより病気の存在を確認することに有効
- \*一般的に、ベースレートが 0 に近づくとも陽性適中率は低くなり、陰性適中率は高くなる。逆にベースレートが高くなるほど陽性適中率は高くなり、陰性適中率は低くなる。

**TABLE 5.5**  
Positive and Negative Predictive Values at Two Different Base Rates for a Screening Test 80% Sensitive and 70% Specific

Screening test result	True status			Predictive value	
	Disorder	No disorder	Total	+PV	-PV
Base rate = .10					
+	80	270	350	.23	.97
-	20	630	650		
Total	100	900	1,000		
Base rate = .75					
+	600	75	675	.89	.54
-	160	175	325		
Total	750	250	1,000		

- ベースレートの陽性・陰性適中率に対する影響は大きいのだが、しばしば看過される。それは経験を積んだ診断者にも当てはまること(Medin & Edelson, 1988)
- ・よくある間違いは、ベースレートに影響されない感度と特異度を、影響される適中率と混同

してしまうこと。

→診断者が、ベースレートの変化に応じてスクリーニングテストの正確さの推測を修正できていないことを示唆

## RESEARCH EXAMPLES

### *Smoking and the Risk of Coronary Heart Disease (Glass and Hopkins, 1996)*

■ある企業で、120名の冠動脈疾患を持つ従業員と、120名の疾患のない従業員をマッチング。その後、喫煙量（非喫煙、1日1パック未満、1日1パック未満、1日1パック以上）でグループ分け

→回顧的な研究なので、オッズ比がもっとも良い効果量の指標。しかし、教育的目的のためその他の指標についても算出する。

■Table 5.6に各セルの観測度数を表示。Cramerの $V=0.20$ 。さらに、喫煙群を1つに統合し4分表にしたものおよびそれに基づく効果の指標をTable 5.7に表示。

・RD=0.16；喫煙群は非喫煙群よりも心臓病のリスクが16%高い。信頼区間より、これは母集団値0.03~0.29と信頼水準95%で整合的

・RR=1.39；喫煙者の有病率は非喫煙者の約1.4倍。この値は母集団値1.06~1.84と整合的。

・観測オッズ比=1.92，母集団オッズ比1.14~3.22と整合的

→全体として信頼区間が広く、精度が低いが、これはサンプルサイズの小ささによるもの。

・ $\phi=0.16$ ，分散説明率は2.5%。これはこの領域では典型的な値であり、「小さい」として軽視されるべきものではない。

**TABLE 5.6**  
Contingency Table for Level of Smoking and Coronary Heart Disease

Group	n	Coronary heart disease	
		Yes	No
Smoking (packs/day)			
< 1	42	19	23
1	64	39	25
> 1	31	20	11
Nonsmoking	103	42	61

Notes. From *Statistical Methods in Education and Psychology* (3rd ed., p. 335), by G. Glass and K. Hopkins, 1986, Boston: Allyn and Bacon. Copyright 1996 by Pearson Education. Adapted with permission.

**TABLE 5.7**  
Fourfold Table for Smokers Versus Nonsmokers for Presence Versus Absence of Coronary Heart Disease for the Data in Table 5.6

Group	n	Coronary heart disease		Effect size					
		Yes	No	$\rho_b$	$\rho_k$	RD	RR	OR	$\phi$
Smoking	137	78	59	.57	.41	.16 <sup>a</sup>	1.39 <sup>b</sup>	1.92 <sup>c</sup>	.16
Nonsmoking	103	42	61						

Notes. S = smoking, N = nonsmoking, RD = risk difference, RR = risk ratio, OR = odds ratio,  $\chi^2(1) = 6.14$ ,  $p = .013$ .

<sup>a</sup>Approximate 95% confidence interval for  $\rho_b - \rho_k$  is .03-.29.

<sup>b</sup>Approximate 95% confidence interval for  $\rho_b/\rho_k$  is 1.06-1.84.

<sup>c</sup>Approximate 95% confidence interval for OR is 1.14-3.22.



Driving Outcomes of Youth With Attentional Difficulties

■ Woodward, Fergusson, & Horwood(2000)

- ・ 13 歳時点の注意欠陥とその後の運転記録の関連についての検討。
- ・ 941 名の少年から構成されるコーホートが、親もしくは教師からの報告に基づいて注意欠陥の観点から 5 群に分類。
- ・ 運転記録は 18-21 歳時点で調査。

■ 下の研究の 11 個の指標のうち、3 つについてのみ今回は報告。全体として、13 歳時点の注意欠陥の程度は運転の問題と関連(Cramer の  $V=.10\sim.23$ )

■ Table5.9 は評定 2~5 の者を注意問題群として結合した 4 分表。

- ・ RD と  $\hat{\phi}$  の値は負。これは 13 歳時点で注意問題の程度が弱かった青年で、運転上の問題を経験する割合が少ないことを示している。
- ・ RR, OR の値は .05 の周辺・・・問題群では問題なし群に比べて、危険運転行動をする割合が約 2 倍

TABLE 5.9  
Effect Size Statistics for Driving Outcomes by Lower Versus Higher Levels of Attentional Difficulties for the Data in Table 5.8

Outcome at age 21 years	Attentional difficulties at age 13 years		$\chi^2(1)$	$\hat{\phi}$	Effect size		
	Lower	Higher			RD	RR	OR
<i>n</i>	476	466					
Injury-related accident (%)	3.6	6.4	4.05 <sup>a</sup>	-.07	-.03	.56	.58
Arrested for drinking and driving (%)	2.9	7.7	10.41 <sup>b</sup>	-.11	-.05	.38	.40
Driving without license (%)	8.2	19.5	25.30 <sup>b</sup>	-.16	-.11	.42	.48

Notes. RD = risk difference; RR = risk ratio; and OR = odds ratio.  
<sup>a</sup> $p < .044$ .    <sup>b</sup> $p < .001$ .

Predictive Values of a Cognitive Test Profile in Detecting Learning or Attentional Problems

■ 教師によってスクールカウンセラーに紹介される児童の多くが、WISC-IIIを受けさせられる。

- ・ WISC-IIIでは 13 の課題から構成され、4 つの群指標(言語理解、注意記憶<sup>4</sup>、処理速度、知覚統合)で個人のプロフィールを作成
- ・ 言語理解(VC)は言語能力と事実についての知識を問うものであり、学習障害の児童では成績が低くなる。
- ・ 注意記憶(FD)と処理速度(PS)を構成する課題は頭文字をとって SCAD<sup>5</sup>と呼ばれる。SCAD は短期記憶を要するものなので、注意スパンが短いと成績が低下する。
- ・ 知覚統合(PO)に含まれる 4 つの課題は、視空間推理に関するものなので、学習障害や注意障害により成績が影響されにくい
- ・ 学習障害や注意欠陥の児童は PO-SCAD の得点が高いことが知られている。次から、この差

<sup>4</sup> Freedom from Distractibility

<sup>5</sup> Symbol search, Coding, Arithmetic, Digit span

得点の適中率について検討していく。

■Table5.10 には、PO-SCAD の得点範囲(1-18)と、そこをカッティングポイントにした場合の感度、特異度を表示。

- ・ここで、感度は学習障害もしくは注意欠陥と診断された子供のうち、その得点以上をとった者の割合。
- ・一方、特異度は WISC-III の規準標本で、その得点以下をとる者の割合。
- ・差得点 9 以上をとる子供の割合は、障害群(54%)と非障害群(26%)で 5%水準の差。この有意水準が 1%となるのは 12 点以上の場合。臨床家は有意でない限り差を解釈しないであろうから、以降は 9 と 12 を取り上げて検討。

TABLE 5.10  
Estimated Sensitivity and Specificity Values for Differences Between  
Scores on Visual-Spatial and Short-Term Memory Tasks for Normal  
and Attentional-Learning Problem Children

Value of (PO-SCAD)	Sensitivity <sup>a</sup>	Specificity <sup>b</sup>
18	.17	.98
17	.21	.97
16	.23	.96
15	.26	.96
14	.29	.94
13	.33	.93
12 <sup>c</sup>	.37	.91
11	.43	.89
10	.49	.86
9 <sup>d</sup>	.54	.84
8	.58	.81
7	.63	.77
6	.65	.74
5	.68	.70
4	.74	.66
3	.80	.61
2	.82	.58
1	.84	.52

Notes. PO = sum of the tasks for the Perceptual Organization factor and SCAD = sum of the tasks for Freedom from Distractibility factor and Processing Speed factor of the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1981).

<sup>a</sup>Sensitivity values were computed based on data from 164 children classified as either learning-disabled or as having attention deficit hyperactivity disorder and reported in "Base Rates of WISC-III Diagnostic Subtest Patterns Among Normal, Learning-Disabled, and ADHD Samples," by A. Prifitera and J. Denck, 1993. In B. A. Bracken and R. S. McCullum (Eds.), *Journal of Psychoeducational Assessment Monograph Series, Advances in Psychological Assessment: Wechsler Intelligence Scale for Children—Third Edition*, p. 39. Copyright 1993 by the Psychoeducational Corporation. Adapted with permission.

<sup>b</sup>Specificity values were computed based on data from 2,158 children in the WISC-III normative sample and reported in *Intelligent Testing With the WISC-III*, by A. S. Kaufman, 1984, New York: Wiley, p. 220. Copyright 1984 by John Wiley and Sons. Adapted with permission.

<sup>c</sup>p < .05 for an individual profile.  
<sup>d</sup>p < .01 for an individual profile.

■Table5.11 で、PO-SCAD>9 の適中率を、1000 人の仮想サンプルで、ベースレートを 10%、25% として検討(10%のみ現実的だが、比較してみるのには教育的に有意義)

- ・ベースレート 10%の時、感度.54、特異度.84 と想定すると、陽性適中率は.27、陰性適中率は.94
- ・ベースレート 25%の時、陽性適中率は.53、陰性適中率は.84
- ・PO-SCAD>12 としても、ベースレート 10%だと陽性適中率は.31、陰性適中率は.93  
→現実的なベースレートの範囲では、カッティングポイントを PO-SCAD の差得点が大きき意方向に変化させても、陽性適中率は上がらない。

TABLE 5.11  
Estimated Predictive Values for a Difference Between Scores on  
Visual-Spatial and Short-Term Memory Tasks of Nine Points at  
Two Different Base Rates

Result (PO-SCAD)	Prediction	True status		Total	Predictive value	
		Learning-attentional problem	No problem		+PV	-PV
Base rate = .10						
≥ 9	Learning-attention problem	54	144	198	.27	.94
< 9	No problem	46	756	802		
	Total	100	900	1,000		
Base rate = .25						
≥ 9	Learning-attention problem	135	120	255	.53	.84
< 9	No problem	115	630	745		
	Total	250	750	1,000		

Note. Sensitivity = .54 and specificity = .84 for (PO-SCAD) ≥ 9; see Table 5.10. PO = sum of the tasks for the perceptual organization factor and SCAD = sum of the tasks for freedom from distractibility factor and processing speed factor of the Wechsler Intelligence Scale for Children—Third Edition (Wechsler, 1991).

■ 勿論学習障害・注意欠陥の有無は差得点のみで判断されるものではないが、認知テスト研究の文脈では、認知的・学習的・精神的問題の発見に役立つとされる PO-SCAD のような差得点が多く挙げられている。しかしながら、それらが感度、特異度、適中率から検討されるのは極めて稀。その枠組みからすれば、診断者はベースレートが適中率に及ぼす影響を考えなければいけないのに、これは残念なこと。

#### Conclusion

- 本章では離散的な結果変数について群を比較する時の効果量指標を紹介。
- 最もよい統計的性質を持つのはオッズ比。
- 4分表のφは周辺度数の比率に影響されるという意味で効果量指標としては問題。
- さらに本章では感度、特異度、適中率の枠組みを紹介。この枠組みでは、スクリーニングテストの結果による病気の有無の判断の正確さに及ぼす母集団ベースレートの影響を明示的に説明。
- 一般的に、スクリーニングテストは稀なケースの特定については正確ではないが、その可能性の排除についてはかなり正確。