

Die gebruik van effekgrootte-indekse by die bepaling van praktiese betekenisvolheid

HS STEYN JR. & SM ELLIS

Statistiese Konsultasiediens, Noordwes-Universiteit, Privaat sak X6001, Potchefstroom, 2520
skdhss@puk.ac.za

UITTREKSEL

Die bepaling van betekenisvolheid van verskille in gemiddeldes of dié van verbande tussen veranderlikes is in baie empiriese studies belangrik. Gewoonlik word net die statistiese betekenisvolheid gerapporteer, wat egter nie noodwendig op belangrike (prakties betekenisvolle) verskille of verbande dui nie. By studies gebaseer op waarskynlikheidsteekproewe behoort effekgrootte-indekse ook gerapporteer te word om aanvullend tot die statistiese betekenisvolheid uitspraak oor die praktiese betekenisvolheid te kan gee. Waar met volledige populasies of gerieflikheidsteekproewe gewerk word, is bepaling van statistiese betekenisvolheid streng gesproke nie van toepassing nie en lewer effekgrootte-indekse 'n basis om betekenisvolheid te kan beoordeel. In hierdie artikel word die aandag gevestig op die gebruik van effekgrootte-indekse om praktiese betekenisvolheid te bepaal. Verder word gewys op die gebruik daarvan in 'n paar toepassingsvelde van statistiek asook hoe dit aandag kry in die statistiese literatuur en rekenaarpakkette. Die gebruik van effekgroottes word toegelig deur enkele voorbeelde uit die navorsingsliteratuur.

ABSTRACT

The use of effect size indices to determine practical significance

The determination of significance of differences in means and of relationships between variables is of importance in many empirical studies. Usually only statistical significance is reported, which does not necessarily indicate an important (practically significant) difference or relationship. With studies based on probability samples, effect size indices should be reported in addition to statistical significance tests in order to comment on practical significance. Where complete populations or convenience samples are worked with, the determination of statistical significance is strictly speaking no longer relevant, while the effect size indices can be used as a basis to judge significance. In this article attention is paid to the use of effect size indices in order to establish practical significance. It is also shown how these indices are utilized in a few fields of statistical application and how it receives attention in statistical literature and computer packages. The use of effect sizes is illustrated by a few examples from the research literature.

1. INLEIDING EN AGTERGROND

By empiriese navorsing is vergelyking van groepe baie keer van belang, of andersins die bepaling van verbande tussen veranderlikes wat gemeet word. Daar word dan gevra na betekenisvolheid van die verskil tussen gemiddeldes of betekenisvolheid van 'n verband. Onder “betekenisvolheid” word gewoonlik verstaan dat 'n sogenaamde “nulhipotese” van geen verskil in gemiddeldes (of geen verband), op 'n voorafbepaalde betekenispeil (gewoonlik 5%) verwerp word, of anders gestel, dat die sogenaamde “p-waarde” kleiner is as 0,05. Hierdie “betekenisvolheid”, ook genoem “statistiese betekenisvolheid”, beteken eintlik maar net dat die waarskynlikheid klein (byvoorbeeld $\leq 0,05$) is dat die nulhipotese verkeerdlik verwerp kan word. Dus dui dit daarop dat die verbande of verskille wat die waarskynlikheidsteekproef of -steekproewe oplewer, nie maar 'n toevalligheid is nie, want die kans vir so 'n toevalligheid is klein (sê 5%). Wat dit egter nie sê nie, is hoe belangrik die verskille of verband wel is. Om uitspraak oor die belangrikheid van verskille of verbande te gee, word van effekgrootte-indekse gebruik gemaak.

2. PRAKTIESE BETEKENISVOLHEID

Soos met statistiese betekenisvolheid, waarby die nulhipotese verwerp moet word, is die vraag: wanneer is 'n verskil of verband groot genoeg om belangrik te wees? Hiervoor kan effekgrootte-indekse gebruik word in die sin dat sulke indekse se grootte direk eweredig is aan die belangrikheid van 'n verskil in gemiddeldes of 'n verband tussen veranderlikes. As 'n indeks groot genoeg is, kan ons die resultaat as prakties betekenisvol bestempel. Dit is 'n algemene term wat in verskillende kontekste gebruik kan word. By kliniese proewe staan dit bekend as byvoorbeeld “klinies betekenisvol” en as die opvoedkunde dit gebruik, kan daarna verwys word as byvoorbeeld “opvoedkundig betekenisvol”.

Pogings om riglynwaardes te koppel aan effekgrootte-indekse, om as sogenaamde afsnypunte te dien vir “klein”, “medium” en “groot” effekte, is deur skrywers soos Cohen¹ aangewend. Weens die feit dat sulke waardes arbitrêr kan wees, is daar ook baie kritiek daarteen.

3. WANNEER IS EFFEKGROOTTE-INDEKSE NODIG?

Gestel dat die gemiddelde diastoliese bloeddruk van 25 hipertensiewe pasiënte met byvoorbeeld 10 mmHg verlaag na 'n sekere behandeling. Gestel hierdie verlaging was statisties betekenisvol op 'n 1%-peil. Die navorser kan op grond hiervan sê dit was ook 'n prakties betekenisvolle verlaging, want hy/sy beoordeel dit op die bekende mmHg-skaal waarby dit bekend is dat 10 eenhede groot genoeg is om belangrik te wees. Hier is 'n effekgrootte-indeks nie nodig nie. 'n Ander voorbeeld is die verkryging van 'n hoogs betekenisvolle korrelasie $p < 0,0001$ van 0,8 tussen 'n nuwe psigometriese toets en 'n standaardtoets om byvoorbeeld depressie op 'n proefgroep van 200 persone te meet, wat daarop dui dat die nuwe toets geldig is vir die proefgroep, want 'n 0,8-korrelasie is uit die sielkundige se ervaring groot genoeg om op geldigheid te dui.

In beide die voorbeelde was kennis van die skaal of vorige ervaring genoegsaam om sonder meer uitspraak te kon lewer oor praktiese betekenisvolheid. Die bepaling van 'n effekgrootte-indeks en evaluering daarvan ten opsigte riglynwaardes was dus nie nodig nie.

Daar is egter gevalle waar effekgrootte-indekse wel nodig is om te bepaal, ten einde uitspraak oor praktiese betekenisvolheid al dan nie, te kan gee. Steyn² gee 'n lys van sulke gevalle met bespreking. Die belangrikste gevalle is:

- Die skaal waarop verskille in gemiddeldes gemeet word, is onbekend, byvoorbeeld in 'n ongestandaardiseerde vraelys waar vrae op 'n 4-punt-Likertskaal gevra word.
- Waarskynlikheidssteekproewe (soos ewekansige of gestratifiseerde steekproewe) word uit populasies getrek en 'n statisties betekenisvolle resultaat word verkry, maar onsekerheid oor die belangrikheid van verskille of verbande bestaan. Hier word effekgrootte-indekse bereken as 'n tweede stap nadat vir statistiese betekenisvolheid getoets is. Indien met groot steekproewe gewerk word, is die resultaat baie keer statisties betekenisvol, sodat die effekgrootte dan daarop gaan dui of dit ook prakties belangrik is.
- By volledige opnames (sensusse) waar volledige populasies betrek word, is bepaling van effekgroottes soms al manier om praktiese betekenisvolheid te beoordeel. In die praktyk kom volledige opnames baie voor. Vir voorbeelde, kyk byvoorbeeld Steyn.^{2,3,4,5}

4. GEBRUIK VAN EFFEKGROOTTE IN TOEPASSINGSVELDE VAN STATISTIEK

Weens die feit dat daar oor jare 'n debat in psigologietydskrifte was oor die gebruik van statistiese betekenistoetse, het die American Psychological Association (APA)⁶ die Task Force on Statistical Inference (TFSI) in die lewe geroep. Hulle verslag (Wilkinson & TFSI)⁷ lewer aanbevelings betreffende data-ontledings, en een van hulle hoofaanbevelings is volgens Kline⁸ (p.13) dat

effekgroottes vir primêre uitkomst gerapporteer moet word, en in elk geval waar p-waardes gegee word.

Steyn² gee voorbeelde van verdere verwysings uit redaksionele artikels in opvoedkundige psigologie, sport- en oefenwetenskap en ook verbruikerswetenskap waarin outeurs gewys word op die belangrikheid van die gebruik van effekgrootte-indekse.

5. EFFEKGROOTTES IN STATISTIESE LITERATUUR EN REKENAARPAKKETTE

Min standaardhandboeke oor statistiese metodes bevat enige materiaal betreffende effekgroottes. 'n Uitsondering is Sheskin⁹ wat 'n verskeidenheid effekgrootte-indekse bespreek. Verder gee Tabachnick en Fidell¹⁰ aandag aan meerveranderlike effekgrootte-indekse, en so ook Huberty.¹¹ Omdat so min statistiese handboeke effekgroottes behandel, het dit tot nou toe nie inslag gevind by bekende statistiese rekenaarpakette soos SAS (SAS Institute Inc.)¹² nie, terwyl STATISTICA (StatSoft Inc.)¹³ en SPSS (SPSS Inc.)¹⁴ se uitvoer beperk is tot die rapportering van eta-kwadrat by ANOVA en MANOVA.

Met die doel om 'n onderskeidingsvermoë en steekproefgrootte te kan bepaal by aanwending van 'n verskeidenheid statistiese betekenistoetse, bespreek Cohen¹ effekgroottes breedvoerig. Omdat sy doel nie die gebruik van effekgrootte is ten einde praktiese betekenisvolheid te bepaal nie, gee hy geen aandag aan die beraming van effekgroottes nie. In die tydskrif *Statistics in Medicine* gee D'Agostino¹⁵ in 'n redaksionele artikel wel riglyne betreffende praktiese betekenisvolheid (hy noem dit “kwantitatiewe of kliniese betekenisvolheid”). Dit is gedoen na aanleiding van 'n artikel van Feinstein,¹⁶ waarin baie breedvoerig na praktiese betekenisvolheid by die vergelyking van twee groepe gekyk is.

Uit bostaande bespreking is dit duidelik dat selfs in die statistiese literatuur effekgroottes as 'n “vreemde” konsep beskou word en dit dus geen wonder is dat dit tot nou toe min of geen inslag gevind het in Toegepaste Statistiek nie. Om daardie rede het 'n handleiding vir die bepaling van effekgrootte-indekse en praktiese betekenisvolheid (Steyn)² onlangs die lig gesien.

6. ENKELE VOORBEELDE VAN ARTIKELS WAAR DIE GEBRUIK VAN EFFEKGROOTTE-INDEKSE AANBEVEEL SOU KON WORD

Ons beperk ons tot twee artikels wat die afgelope tyd in *Die Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* verskyn het:

6.1 Betekenisvolle verskil tussen twee groepgemiddeldes

In 'n studie rakende die toelating van studente tot studierigtings het Maree¹⁷ die hele populasie van eerstejaars binne die Skool vir Onderwysopleiding by die Universiteit van Pretoria betrek en daarby genoem dat die resultate beperkte veralgemeningswaarde besit. In sy tabel 5 word dan onder andere 'n t-toets (waar gelyke variansies aanvaar word) gedoen om die hipotese te toets dat die gemiddelde M-tellings van studente wat gedruip of studies gestaak het, verskil van dié wat geslaag het ($t = -2,345$, $p = 0,021$). Na ons mening sou die effekgrootte-indeks d (Cohen)¹ se rapportering hier meer sinvol gewees het. Dit lewer in hierdie geval die waarde $-0,55$ op (kyk Steyn,² vir die metode van berekening daarvan uit t en die groepgroottes). Volgens Cohen se riglyne is dit 'n medium effekgrootte ($0,8$ sou groot gewees het), wat dus op 'n “matige” verskil in gemiddeldes dui. Die gevolgtrekking van Maree dat daar 'n betekenisvolle verskil op grond van $p < 0,05$ bestaan, is om twee redes ons insiens nie geregverdig nie: (1) statistiese betekenisvolheid is nie relevant nie, vanweë die feit dat hierdie nie 'n waarskynlikheidssteekproef uit 'n groter populasie is nie, maar reeds die populasie self is, en (2) die resultaat nie noodwendig op 'n prakties betekenisvolle verskil dui nie.

6.2 Betekenisvolle verband tussen twee kategoriese veranderlikes

In 'n ondersoek na laerisikogroepe vir bilharziabesmetting in 'n endemiese gebied rapporteer Wolmarans et al.¹⁸ statisties betekenisvolle verbande met behulp van die Fisher-eksakte toets tussen ouderdomsgroepe (op twee vlakke) en besmetting ($p < 0,05$). Die phi-koëffisiënt, wat as 'n effekgrootte-indeks dien (Steyn),^{2,5} kon ook hier bereken word en varieer in hierdie gevalle tussen 0,03 en 0,33, wat dui op klein tot medium grootte verbande. 'n Prakties betekenisvolle verband tussen ooreenstem met 'n groot effek (phi-koëffisiënt van 0,5 of groter, kyk Cohen).¹ In hierdie geval is p-waardes streng gesproke nie relevant nie omdat met 'n gerieflikheidsteekproef gewerk word.

BIBLIOGRAFIE

1. Cohen, J. (1988). *Statistical Power Analysis*. Second Edition. Academic Press, Inc., New York.
2. Steyn, H.S. jr. (2005). Handleiding vir bepaling van effekgrootte-indeks en praktiese betekenisvolheid. Beskikbaar op internet by: www.puk.ac.za/fakulteite/natuur/skd/index.html
3. Steyn, H.S. jr. (1992). Die gebruik van effekgroottes in hipotesetoetsing by volledige populasies. *S.A. Tydskrif vir Sielkunde*, 22(2):97.
4. Steyn, H.S. jr. (2000). Practical significance of the difference in means. *S.A. Tydskrif vir Bedryfsielkunde*, 26(3):1-3.
5. Steyn, H.S. jr. (2002). Practically significant relationship between two variables. *S.A. Tydskrif vir Bedryfsielkunde*, 28(3):10-15.
6. APA (2001). *Publication manual of the American Psychological Association*, Fifth edition. Washington, DC.
7. Wilkinson, L. & TFSI (1999). Statistical methods in psychological journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
8. Kline, R.B. (2004). *Beyond significance testing: reforming data analysis methods in behavioural research*. American Psychological Association. Washington, DC.
9. Sheskin, D.J. (2000). *Handbook of parametric and nonparametric statistical procedures*. Second edition. London: Chapman & Hall.
10. Tabachnick, B.G. & Fidell, L.S. (2000). *Using multivariate statistics*. Fourth Edition. Boston: Allyn & Bacon.
11. Huberty, C.J. (1994). *Applied discriminant analysis*. New York: John Wiley & Sons.
12. SAS Institute Inc. (2002-2003). SAS 9.1 TS Level 1 MO XP_PRO platform. SAS Institute Inc., Cary, NC, USA. www.sas.com
13. StatSoft, Inc. (2005). STATISTICA (data analysis software system), version 7.1 www.statsoft.com
14. SPSS Inc. (1989-2005). SPSS 14.0 for Windows. www.spss.com.
15. D'Agostino, R.B. snr. (1999). Editorial: quantifying the comparison of two groups. *Statistics in Medicine*, 18, 2551-2555.
16. Feinstein, A.R. (1999). Indexes of contrast and quantitative significance for comparisons of two groups. *Statistics in Medicine*, 18, 2557-2581.
17. Maree, J.G. (2003). There is a tide in affairs of men... Die uitdaging om werkbare en billike(-r) keuringsmeganismes te ontwerp. *SA Tydskrif vir Natuurwetenskap en Tegnologie*, 22(1):19-26.
18. Wolmarans, C.T., De Kock, K.N. & Le Roux, J. (2004). Die gebruik van vraelyste en parasitologiese analises om laerisikogroepe vir bilharziabesmetting, asook die faktore wat daartoe mag bydra, in 'n gemeenskap in 'n endemiese gebied in Suid-Afrika te identifiseer. *SA Tydskrif vir Natuurwetenskap en Tegnologie*, 23(4):99-103.

Faans Steyn is hoof van die Statistiese Konsultasiediens by die Noordwes-Universiteit se Potchefstroomkampus sedert 1981. Hy is vanaf 1969 verbonde as dosent aan die Departement Statistiek van die toenmalige PU vir CHO waar hy ook in 1973 sy PhD in Statistiek behaal. Hy is geregistreer as professionele natuurwetenskaplike by SACNASP en is lid van die Suid-Afrikaanse Statistiese Vereniging. As statistiese konsultant help hy navorsers oor 'n wye veld met statistiese beplanning, ontledings en interpretasie in hulle navorsing wat neerslag gevind het in nagenoeg 70 artikels. Ook het hy 'n 12-tal artikels gepubliseer oor basiese en toegepaste navorsing in Statistiek.



Suria Ellis is senior vakkundige by die Statistiese Konsultasiediens van die Noordwes-Universiteit se Potchefstroomkampus. Sy het die volgende nagraadse kwalifikasies aan die destydse PU vir CHO behaal: MSc (Fisika), MSc (Statistiek) en PhD (Statistiek). Sy is geregistreer as professionele natuurwetenskaplike by SACNASP, is lid van die Suid-Afrikaanse Statistiese Vereniging en was ook betrokke by die organisering van die *International Conference for Mathematics in Finance 2005*. Sy het deelgeneem aan nege internasionale kongresse en is outeur of mede-outeur van net soveel nasionale en internasionale publikasies.

