PRIRODOSLOVNO-MATEMATIČKI FAKULTET
SVEUČILIŠTE U SPLITU

Ivana Ćulav

Statističke metode za procjenjivanje veličine učinka

DIPLOMSKI RAD

SPLIT, rujan, 2008.

Studijska grupa: Matematika i informatika
Predmet: Primjena računala u nastavi

Statističke metode za procjenjivanje veličine učinka

DIPLOMSKI RAD

Student: Ivana Ćulav

Mentor: prof.dr.sc. Slavomir Stankov
Neposredni voditelj: mr.sc. Ani Grubišić

SPLIT, rujan, 2008.

Sadržaj

1. Uvod

Prvu meta-analizu je proveo Karl Pearson, 1904. godine. Želio je savladati problem u istraživanjima s malom veličinom uzorka, gdje je umanjena "moć" statistike. Analizirajući rezultate istraživanja dolazio je do mnogo preciznijih analiziranih podataka. Njegovim razmišljanjima su potaknute mnoge 'klasične' statističke metode koje su i danas u upotrebi. Neki od njegovih doprinosa su pojam *korelacije*, *korelacijskog koeficijenta*, kao i pojam *veličine učinka*, itd [WIKI2008].

Meta-analizu možemo definirati kao statističku tehniku za povezivanje različitih rezultata istraživanja, sažimanje rezultata u manje skupove ne gubeći bitne informacije i ocjenjivanje kvalitete kvantitativnih rezultata. Jednostavnije rečeno, meta-analiza nam može dati odgovor na možda najvažnije pitanje, a to je "Hoće li ono što možemo izmjeriti ili proračunati biti uistinu novi rezultat?".

Najzanimljivije u meta-analizi je efekt kombiniranja rezultata više istraživačkih timova u nekoj znanstvenoj problematici prema jednom velikom istraživanju. Do kombiniranja rezultata dolazi se korištenjem pravila i raznih metoda meta-analize, a pomoću njih se lako mogu uspješno istražiti rijetki događaji, ali i kvalitetno otkloniti prijašnji pogrešni rezultati. Za svaki rezultat današnjih istraživanja, u svim granama znanosti, podrazumijeva se da je dobiven korištenjem naprednih alata i mogućnosti meta-analize.

Potrebno je naglasiti da se meta-analiza primjenjuje samo na istraživanjima sa empirijskim rezultatima. Odnosno, primjenjuje se samo na istraživanjima koja daju kvantitativne rezultate, tj. ona koja koriste kvantitativno mjerenje varijabli [LIPS2001]. Kada imamo istraživanja koja udovoljavaju tom uvjetu, možemo krenuti sa meta-analizom, ali postepeno i po koracima.

Najvažniji korak je usporedba odabranih rezultata istraživanja nekom vrstom numeričke skale vrijednosti, tako da se krajnje dobiveni rezultati mogu međusobno smisleno analizirati. Rezultati pojedinog istraživanja se prikazuju u obliku *veličine učinka*, a to je pojam koji otkriva kvantitativnu informaciju iz svakog rezultata istraživanja. On otkriva kolika je učinkovitost eksperimentalnog faktora, tj. je li učinkovit, te kolika je točno učinkovitost u jedinicama standardne devijacije. S ciljem da usporedimo rezultate više različitih istraživanja, moramo najprije izračunati veličinu učinka posebno za svako istraživanje. Nakon toga se dobiveni rezultati mogu uspoređivati, ali i obrađivati, kao npr. traženje aritmetičke sredine, varijance, itd. Sljedeći korak, ali i najbitniji korak u meta-analizi, je definiranje metode izračuna veličine učinka. Koja će metoda izračuna veličine učinka biti najprikladnija ovisi o prirodi rezultata, statističkom obliku u kojem su izraženi, ali i o pitanjima na koja odgovara meta-analiza.

E-učenje danas sve češće susrećemo u praktičnoj primjeni, ono omogućava korisnicima kvalitetno sudjelovanje u nastavi i kada to pitanje udaljenosti, rasporeda i sličnih okolnosti praktički čine nemogućim. Oblici poučavanja, kao što su *web-temeljeno poučavanje* (engl. web-based learning), *računalno-temeljeno poučavanje* (engl. computer based learning), *digitalne* *medije* (engl. digital colaboration), čine e-učenje [WIKI2008]. Implementacija e-učenja ostvaruje se u sustavima e-učenja. Oni su zamišljeni kao „osobni učitelji“ za svakog učenika. S ciljem da se postigne veća učinkovitost u procesu učenja i poučavanja, nastoji se sustave e-učenja što više prilagoditi svakom pojedinom korisniku.

Jedan od istraživača, kojemu je cilj unapređenje procesa učenja i poučavanja, je i Benjamin S. Bloom. On je pokazao da je veličina učinka individualnog poučavanja 2 standardne devijacije u odnosu na tradicionalno, tj. da je student u tutorskoj grupi za oko dvije standardne devijacije bolji od studenta u tradicionalnoj grupi, dok je prosječan student u grupi učenja s provjeravanjem za oko jednu standardnu devijaciju bolji od prosječnog studenta u tradicionalnoj grupi. Bloom je smatrao da je ova razlika od dvije standardne devijacije ideal prema kojem proces učenja i poučavanja treba stremiti, te je zato ovaj problem nazvao 2-sigma problem [BLOO1984].

U drugom poglavlju su opisane parametrijske i neparametrijske procjene veličine učinka iz jednog istraživanja. Pokazano je kako se interpretira veličina učinka, te su dane četiri metode procjene veličine učinka, kao i njihovo uspoređivanje međusobno. Nadalje je prikazano izračunavanje intervala pouzdanosti, te distribucije veličine učinka.

U trećem poglavlju su opisane parametrijske procjene veličine učinka iz niza eksperimenata, kao što su težinska linearna procjena, procjena maksimalnih vjerojatnosti i procjene temeljene na transformiranim procjenama. Prikazano je ispitivanje homogenosti veličine učinka, ali i procjena veličine učinka za male uzorke. Te na kraju su dani utjecaji pogrešaka u mjerenju i nevaljanost.

U četvrtom poglavlju su obrađene i prikazane procjene veličine učinka kada nisu svi rezultati promatrani, kao i pristranost vezana za uzorak. Opisane su procjene veličine učinka koje se temelje samo na značajnim rezultatima, u slučaju jednog istraživanja, te u slučaju kada se procjenjuje veličina učinka iz niza nezavisnih istraživanja.

U petom poglavlju provedena je meta-analiza rezultata dobivenih na temelju istraživanja o učinkovitosti sustava xTEx-Sys na postignuća učenika i studenata. Istraživanja su provedena u akademskoj godini 2005/2006 i 2006/2007 na 499 ispitanika. Prikazano je izračunavanje raznih metoda procjene veličine učinka, kao i izračunavanje testa homogenosti, intervala pouzdanosti, te na kraju prosječne veličine učinka.

2. Parametrijske i neparametrijske procjene veličine učinka iz jednog eksperimenta

Ovo poglavlje je posvećeno analizi parametrijskih i neparametrijskih metoda za procjenu veličine učinka iz pojedinačnog eksperimenta, tj. istraživanja. Zamislimo da je provedeno nekoliko istraživanja. Ocjenjivač bi prvo htio procijeniti veličinu učinka za svako istraživanje. Ako postoji razlog za vjerovati da je veličina učinka jednaka u svakom istraživanju, tada mi procjenjujemo zajedničku veličinu učinka koristeći informacije iz svih istraživanja. Važno je prepoznati da je procjena i interpretacija zajedničke veličine učinka u biti jednaka kroz studije. U protivnom, procjena srednje vrijednosti (eng. mean) učinka može važne razlike među studijama napraviti nevidljivima.

2.1. Procjena veličine učinka iz jednog eksperimenta

U ovom dijelu raspravljamo o nekoliko točaka procjene veličine učinka δ iz pojedinačnog dvo-grupnog eksperimenta. Ove procjene su temeljene na primjeru standardne razlike aritmetičkih sredina (eng. standardized mean difference), koje ovise o uključenim veličinama uzorka (eng. sample sizes).

Statistička svojstva procjene veličine učinka ovise o modelu promatranja u eksperimentu. Prikladan i često realističan model je da pretpostavimo da su opažanja neovisno normalno distribuirana među grupama u eksperimentu. Odnosno, pretpostavljamo da su eksperimentalna opažanja $Y_1^E,...Y_{n^E}^E$ normalno distribuirana sa srednjom vrijednošću μ^E i varijancom σ^2. Slično, kontrolna opažanja $Y_1^C,...Y_{n^C}^C$ su distribuirana sa srednjom vrijednošću μ^C i varijancom σ^2. Odnosno,

$$\begin{aligned} Y_j^E &\approx N\left(\mu^E,\sigma^2\right), \quad j=1,...n^E \quad \text{i} \\ Y_j^C &\approx N\left(\mu^C,\sigma^2\right), \quad j=1,...n^C \end{aligned} \tag{1}$$

Veličina učinka δ je definirana kao

$$\delta = (\mu^E - \mu^C)/\sigma \tag{2}$$

2.1.1. Interpretiranje veličine učinka

Veličine učinka se koriste u sintezi eksperimentalnih rezultata. Najizravnija interpretacija veličine učinka populacije $\delta = (\mu^E - \mu^C)/\sigma$ je ta da je δ srednja vrijednost koja će se postići ako se zavisna varijabla izmjeri tako da ima elementarnu razliku među grupama u eksperimentu. Dakle, veličina učinka je samo srednja vrijednost ponovno izražena u skaliranim jedinicama.

Kada su opažanja u eksperimentalnoj i kontrolnoj grupi normalno distribuirana, δ može biti korištena za kvantitativno određivanje stupnja preklapanja između distribucija opažanja u eksperimentalnoj i kontrolnoj grupi. Kako je $\delta = (\mu^E - \mu^C)/\sigma$ standardizirani rezultat (z-vrijednost) srednje vrijednosti eksperimentalne grupe u distribuciji kontrolne grupe, $\Phi(\delta)$ predstavlja udio rezultata kontrolne grupe koja je manja nego prosječni rezultat eksperimentalne grupe. Prema tome, veličina učinka od δ = 0.5 implicira da rezultat prosječne individue u eksperimentalnoj grupi prelazi 69% individua u kontrolnoj grupi. Slično, veličina učinka od -0.5 implicira da rezultat prosječne individue u eksperimentalnoj grupi prelazi samo 31% individua u kontrolnoj grupi.

Druga interpretacija veličine učinka je dobivena pretvaranjem δ u procjenu koeficijenta korelacije. U ovoj situaciji je

$$\rho^2 \cong \frac{\delta^2}{\delta^2 + (n^E + n^C - 2)/\tilde{n}} \tag{3}$$

gdje je $\tilde{n} = n^E n^C / (n^E + n^C)$. Ako je $n^E = n^C = n$ onda se formula svodi na

$$\rho^2 \cong \frac{\delta^2}{\delta^2 + 4(n-1)/n} \cong \frac{\delta^2}{\delta^2 + 4} \tag{4}$$

Autori [ROSE1979] predlažu upotrebu binomnog prikaza veličine učinka (eng. binomial effect size display) kao jednostavnog načina za ilustriranje jačine učinka (eng. magnitude of effects) izražene kao koeficijent korelacije. Oni su raščlanili nezavisnu varijablu na eksperimentalnu i kontrolnu grupu, a zavisnu varijablu na uspjeh i neuspjeh. Navode da postoji jednostavna povezanost između koeficijenta korelacije i udjela uspjeha i neuspjeha u eksperimentalnoj i kontrolnoj grupi. Pretpostavimo da je općenito udio uspješnih 0.50. Tada je učestalost uspjeha u eksperimentalnoj grupi $0.50 + \rho/2$, a učestalost uspjeha u kontrolnoj grupi je, $0.50 - \rho/2$ kao što je prikazano u Tablici 1.

Tablica 1. Povezanost između koeficijenta korelacije i udjela uspjeha u binomnom prikazu veličine učinka

Grupa	Ishod	
	Neuspjeh	Uspjeh
Kontrolna	$0.50 + \rho/2$	$0.50 - \rho/2$
Eksperimentalna	$0.50 - \rho/2$	$0.50 + \rho/2$

Razlika između učestalosti uspjeha u eksperimentalnoj i kontrolnoj grupi je uvijek $\rho/2+\rho/2=\rho$ dakle, čak i prilično mali udio obračunate varijance može odgovarati značajnoj promjeni u učestalosti uspjeha između eksperimentalne i kontrolne grupe. Npr., vrijednost od ρ^2=0.10, koja se može činiti prilično malom, odgovara (kako je ρ=0.32) priličnom povećanju stope uspjeha od 34% do 66 %.

2.1.2. Procjena veličine učinka temeljena na standardnoj razlici aritmetičkih sredina

Glass je predložio metodu procjene δ temeljene na primjeru vrijednosti standardne razlike aritmetičkih sredina [GLAS1976]. Ako su $\overline{Y}^E$ i $\overline{Y}^C$ odgovarajuće srednje vrijednosti eksperimentalne i kontrolne grupe, tada je standardna razlika aritmetičkih sredina:

$$\left(\overline{Y}^E-\overline{Y}^C\right)/s^* \qquad (5)$$

gdje je s^* standardna devijacija. Različiti odabiri s^* rezultiraju različitim metodama procjene. Npr., može se s^* definirati kao s^C (standardna devijacija kontrolne grupe) ili s^E (standardna devijacija eksperimentalne grupe). Umjesto toga možemo koristiti zajedničku standardnu devijaciju (eng. pooled standard deviation) koja kombinira s^E i s^C.

Nadalje, predlaže metodu procjene koja koristi s^C za standardnu razliku aritmetičkih sredina:

$$g'=\left(\overline{Y}^E-\overline{Y}^C\right)/s^C \qquad (6)$$

Njegov argument je da zajednički parovi varijanci (eng. pooling pairs of variances) mogu dovesti do različitih standardnih vrijednosti identičnih razlika srednjih vrijednosti u eksperimentu gdje je nekoliko postupaka uspoređeno sa kontrolnom grupom [GLAS1976].

U mnogim slučajevima, pretpostavka jednakih varijanci populacijske (eng. equal population variance) je razumna, pa se predlaže najpreciznija metoda procjene populacijske varijance dobivena sa udruživanjem (eng. pooling)

Kako smo uzeli u obzir samo dvije grupe po eksperimentu, s jednakim populacijskim varijancama, Glassov argument možemo primijeniti. Stoga, je izmijenjena metoda procjene dobivena korištenjem procjene zajedničke standardne devijacije.

Jedna varijacija od g' je

$$g=\left(\overline{Y}^E-\overline{Y}^C\right)/s \qquad (7)$$

gdje je s primjer zajedničke standardne devijacije

$$s=\sqrt{\frac{\left(n^E-1\right)\left(s^E\right)^2+\left(n^C-1\right)\left(s^C\right)^2}{n^E+n^C-2}} \qquad (8)$$

n^E i n^C su primjeri veličine uzorka (eng. sample size) eksperimentalne i kontrolne grupe, za svaku posebno.

Zbog toga što su g i g' statistike uzorka (eng. sample statistic), svaki sadrži i distribuciju uzorka (eng. sampling distribution). Distribucija uzorka od obje statistike je usko vezan za necentalnu t-distribuciju (eng. noncentral t-distribution) osobito ako je

$$\tilde{n} = \frac{n^E n^C}{n^E + n^C} \tag{9}$$

tada svaki od $\sqrt{\tilde{n}}g$ i $\sqrt{\tilde{n}}g'$ ima necentralnu t-distribuciju sa necentralnim parametrom $\sqrt{\tilde{n}}\delta$ i odgovarajućim stupnjem slobode $n^E + n^C - 2$ i. $n^C - 1$

Kao što će biti prikazano, pristranost (eng. bias) i varijanca od g su obje manje od pristranosti i varijance od odgovarajućeg g'. Slijedom navedenog, g je bolja procjena nego g' bez obzira na vrijednost od δ, i g' se više neće razmatrati.

Očekivana srednja vrijednost od g definirana sa (7) je otprilike

$$E(g) \cong \delta + \frac{3\delta}{4N - 9} \tag{10}$$

gdje je $N = n^E + n^C$.

Točna srednja vrijednost je

$$E(g) = \delta / J(N - 2) \tag{11}$$

gdje je $J(m)$ konstanta koja je prikazana u Tablici 2. za vrijednost m od 2 do 50 (ova konstanta $J(m)$ se pojavljuje kao $c(m)$ u mnogim objavljenim literaturama, [HEDG1981], mi smo promijenili notaciju da ne dođe do sukoba sa ostalim notacijama koje će biti korištene kasnije). Konstanta $J(m)$ približno iznosi

$$J(m) = 1 - \frac{3}{4m - 1} \tag{12}$$

Varijanca od g je približno

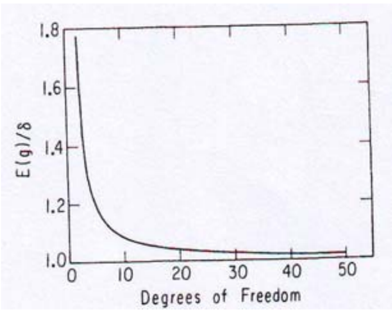
$$Var(g) \cong \frac{1}{\tilde{n}} + \delta^2 \frac{1}{2(N - 3.94)} \tag{13}$$

Iz (10) možemo vidjeti da pristranost u procjeni δ sa g je približno $Bias(g) \cong 3\delta / (4N - 9)$ pa za N dovoljno malen kao npr. 12, pristranost je samo 0.08δ, što je maleno, osim ako je δ prilično velika.

Grafički prikaz pristranosti od g kao funkcije stupnjeva slobode N-2 je dan na Slici 1.

Tablica 2. Točne vrijednosti od faktora za korekciju pristranosti

m	$J(m)$	m	$J(m)$	m	$J(m)$	m	$J(m)$
2	0.5642	15	0.9490	27	0.9719	39	0.9806
3	0.7236	16	0.9523	28	0.9729	40	0.9811
4	0.7979	17	0.9551	29	0.9739	41	0.9816
5	0.8408	18	0.9577	30	0.9748	42	0.9820
6	0.8686	19	0.9599	31	0.9756	43	0.9824
7	0.8882	20	0.9619	32	0.9764	44	0.9828
8	0.9027	21	0.9638	33	0.9771	45	0.9832
9	0.9139	22	0.9655	34	0.9778	46	0.9836
10	0.9228	23	0.9670	35	0.9784	47	0.9839
11	0.9300	24	0.9684	36	0.9790	48	0.9843
12	0.9359	25	0.9699	37	0.9796	49	0.9846
13	0.9410	26	0.9708	38	0.9801	50	0.9849
14	0.9453						



Slika 1. Grafički prikaz pristranosti od *g*, [HEDG1985]

2.1.3. Nepristrana procjena veličine učinka

Razmatranje Slike 1. otkriva da g ima malu pristranost uzorka. Jednostavno je ukloniti tu pristranost definirajući novu metodu procjene

$$d = J(N-2) \qquad g = J(N-2)\frac{\overline{Y}^E - \overline{Y}^C}{s} \tag{14}$$

ili približno

$$d \cong \left(1 - \frac{3}{4N-9}\right) g \tag{15}$$

Zato toga su što su i pristranost i varijanca od d manji od pristranosti i varijance od g, d ima manju srednju kvadratnu pogrešku (eng. mean squared error) nego g i dominira g kao metoda procjene od δ. Kada je $n^E = n^C$, d nije samo nepristrana metoda procjene, nego također i jedinstvena nepristrana procjena najmanje varijance od δ [HEDG1981].

Ako usporedimo d i g, dobivamo

$$d / g = J(N-2) \tag{16}$$

što ovisi o ukupnoj veličini uzorka. Kako se N povećava, $J(N-2)$ se približava jedinstvenom, tako da distribucije i d i g imaju tendenciju prema normalnoj distribuciji sa identičnim srednjim vrijednostima i varijancama u velikim uzorcima. Kako d ima tendenciju prema g, pri vjerojatnosti da se N povećava, d i g su u biti ista metoda procjene u većim uzorcima.

2.1.4. Procjenjena maksimalna vjerojatnost veličine učinka

Metode procjene maksimalne vjerojatnosti je konzistentna i djelotvorna. Kako god, procjena maksimalne vjerojatnosti je također široko korištena, čak i kad potreba za velikim uzorkom nije ispunjena.

Metode procjene maksimalne vjerojatnosti $\hat{\delta}$ od δ ja dana sa

$$\hat{\delta} = \sqrt{\frac{N}{N-2}} \frac{\overline{Y}^E - \overline{Y}^C}{s} = \sqrt{\frac{N}{N-2}} g \tag{17}$$

$\hat{\delta}$ je pristrana na malim uzorcima i ima veću pristranost nego Glassova metoda procjene g'.

Također, metoda procjene maksimalne vjerojatnosti $\hat{\delta}$ ima veću varijancu nego d. Sve tri metode procjene $\hat{\delta}$, g i d veličine učinka iz pojedinog eksperimenta imaju jednaku asimptotsku distribuciju i jednaki su kod većine uzoraka.

2.1.5. Procjenjivanje veličine učinka metodom smanjivanja

Poznato je da nepristrana procjena najmanje varijance ne treba biti procjena najmanje srednje kvadratne pogreške. Npr., srednja kvadratna pogreška od procijenjene maksimalne vjerojatnosti varijance normalne distribucije, može biti reducirana metodom smanjivanja prema nuli (Stein, 1964.). Bez obzira na vrijednost parametra δ,

$$\tilde{g} = \frac{N-4}{N-2} \frac{g}{J(N-2)} = \frac{N-4}{N-2} \frac{d}{[J(N-2)]^2} \tag{18}$$

ima manju glavnu kvadratnu pogrešku nego d.

2.1.6. Uspoređivanje parametrijskih metoda procjena veličine učinka

Predloženi su sljedeći procjenitelji veličine učinka:

$$g = \frac{\overline{Y}^E - \overline{Y}^C}{s}, \quad d = J(N-2)g, \tag{19}$$

$$\hat{\delta} = \sqrt{\frac{N}{N-2}} g, \quad \tilde{g} = \frac{N-4}{N-2} \frac{g}{J(N-2)}$$

Zbog konstanti u definicijama od d, $\hat{\delta}$ i g se približavaju jedinstvenoj vrijednosti kako se veličina uzorka povećava, a sve četiri metode procjene imaju jednake distribucije velikih uzoraka, i stoga i jednaka svojstva velikih uzoraka.

PRIMJER

Pretpostavimo da istraživanje sa $n^E = n^C = 10$ ima veličinu učinka od g =0.60. Koristeći (19) i Tablicu 2., nepristrana metoda procjene d za ovo istraživanje je $d = J(18)0.60 = (0.9577)0.60 = 0.57$.

Metoda procjene maksimalne vjerojatnosti za veličinu učinka za ovo istraživanje je $\hat{\delta} = \sqrt{20/18}0.60 = 0.63$.

Procjena smanjivanjem veličine učinka $\hat{\delta}$ za ovo istraživanje je $\hat{\delta} = [16/(18)(0.9577)]0.60 = 0.51$.

Primijetimo, da čak i u ovom primjeru, gdje je $n^E = n^C = 10$, vrijednosti 0.60, 0.57, 0.63, i 0.51 za četiri metode procjene se ne razlikuju mnogo. Četiri metode procjene δ su pouzdane, tako da je $\hat{\delta}^2 \geq g^2 \geq d^2 \geq \tilde{g}^2$.

Iz definicije (19) slijedi da varijance su poredane na isti način: $Var(\hat{\delta}) \geq Var(g) \geq Var(d) \geq Var(\tilde{g})$.

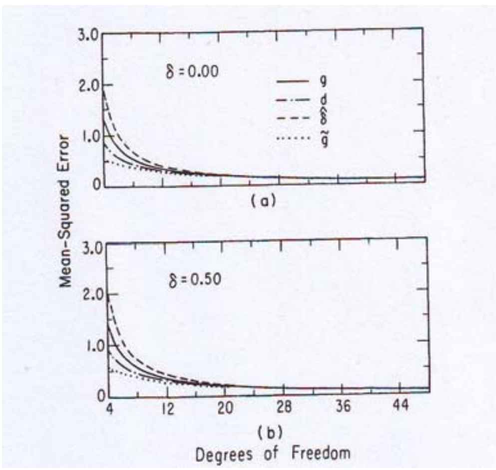
Ipak, srednje kvadratne pogreške ne moraju biti poredane na isti način.

Tablica 3. prikazuje varijancu, pristranosti i srednje kvadratne pogreške za četiri metode procjene za odabranu veličinu uzorka i veličinu učinka δ. $Bias(\hat{\delta})$ i $bias(g)$ su pozitivni kad je δ pozitivna, a negativni kad je δ negativna. Najbolja metoda procjene po kriteriju srednje kvadratne pogreške je $\tilde{g}$, i razlike među metodama procjena su veće kada je veličina uzorka manja.

Slika 2. predstavlja grafički prikaz srednje kvadratne pogreške za svaku od četiri metode procjene kao funkciju stupnjeva slobode. Slika prikazuje da iako se metode procjene razlikuju za konačno velike uzorke, razlike u tim metodama su male za 16 ili više stupnjeva slobode. Uistinu, razlike su primjetne samo za vrlo malene vrijednosti stupnjeva slobode, što je nerealno u većini primjena. Mnoge primjene obuhvaćaju veličine uzorka od najmanje 10 subjekata po grupi, i u ovim slučajevima razlike između ove četiri metode procjene su zanemarive. Kao osnovnu metodu procjene veličine učinka za pojedinačno istraživanje koristimo *d*, jer ima dobra svojstva za male uzorke.

Tablica 3. Varijanca, pristranost i srednja kvadratna pogreška za četiri metode procjene veličine učinka za veličinu učinka δ =0.0, 0.2, 0.5, i 1.0, i za različite stupnjeve slobode, [HEDG1985]

Degrees of freedom	Variance				Bias				Mean-squared error			
	δ	g	d	$\tilde{g}$	δ	g	d	$\tilde{g}$	δ	g	d	$\tilde{g}$
						$\delta = 0.0$						
4	2.000	1.333	0.849	0.524	0.000	0.000	0.000	0.000	2.000	1.333	0.849	0.524
6	1.000	0.750	0.566	0.442	0.000	0.000	0.000	0.000	1.000	0.750	0.566	0.442
8	0.667	0.533	0.435	0.368	0.000	0.000	0.000	0.000	0.667	0.533	0.435	0.368
10	0.500	0.417	0.355	0.313	0.000	0.000	0.000	0.000	0.500	0.417	0.355	0.313
15	0.308	0.271	0.244	0.226	0.000	0.000	0.000	0.000	0.308	0.271	0.244	0.226
20	0.222	0.202	0.187	0.177	0.000	0.000	0.000	0.000	0.222	0.202	0.187	0.177
25	0.174	0.161	0.151	0.145	0.000	0.000	0.000	0.000	0.174	0.161	0.151	0.145
50	0.083	0.080	0.078	0.076	0.000	0.000	0.000	0.000	0.083	0.080	0.078	0.076
100	0.041	0.040	0.039	0.039	0.000	0.000	0.000	0.000	0.041	0.040	0.039	0.039
						$\delta = 0.2$						
4	2.026	1.351	0.860	0.530	0.107	0.051	0.000	−0.043	2.037	1.353	0.860	0.532
6	1.009	0.757	0.571	0.446	0.066	0.030	0.000	−0.023	1.014	0.758	0.571	0.446
8	0.672	0.538	0.438	0.371	0.048	0.022	0.000	−0.016	0.674	0.538	0.438	0.371
10	0.504	0.420	0.357	0.315	0.037	0.017	0.000	−0.012	0.505	0.420	0.357	0.316
15	0.310	0.273	0.246	0.228	0.024	0.011	0.000	−0.008	0.310	0.273	0.246	0.228
20	0.224	0.203	0.188	0.178	0.018	0.008	0.000	−0.005	0.224	0.203	0.188	0.178
25	0.175	0.162	0.152	0.146	0.014	0.006	0.000	−0.004	0.175	0.162	0.152	0.146
50	0.084	0.081	0.078	0.076	0.007	0.003	0.000	−0.002	0.084	0.081	0.078	0.076
100	0.041	0.040	0.039	0.039	0.004	0.002	0.000	−0.001	0.041	0.040	0.039	0.039
						$\delta = 0.5$						
4	2.161	1.441	0.917	0.566	0.267	0.127	0.000	−0.107	2.233	1.457	0.917	0.577
6	1.058	0.794	0.599	0.468	0.165	0.076	0.000	−0.058	1.085	0.799	0.599	0.471
8	0.670	0.560	0.456	0.386	0.119	0.054	0.000	−0.040	0.714	0.563	0.456	0.388
10	0.523	0.436	0.371	0.327	0.094	0.042	0.000	−0.030	0.531	0.437	0.371	0.328
15	0.320	0.282	0.254	0.235	0.061	0.027	0.000	−0.019	0.324	0.283	0.254	0.236
20	0.231	0.210	0.194	0.184	0.045	0.020	0.000	−0.014	0.233	0.210	0.194	0.184
25	0.180	0.167	0.157	0.150	0.036	0.016	0.000	−0.011	0.182	0.167	0.157	0.150
50	0.087	0.083	0.080	0.079	0.018	0.008	0.000	−0.005	0.086	0.083	0.083	0.079
100	0.042	0.041	0.041	0.040	0.009	0.004	0.000	−0.002	0.042	0.041	0.041	0.040
						$\delta = 1.0$						
4	2.644	1.763	1.122	0.692	0.535	0.253	0.000	−0.215	2.930	1.827	0.122	0.738
6	1.233	0.925	0.698	0.545	0.329	0.151	0.000	−0.116	1.341	1.240	0.698	0.558
8	0.799	0.639	0.521	0.441	0.239	0.108	0.000	−0.080	0.856	0.770	0.521	0.448
10	0.591	0.492	0.419	0.370	0.187	0.084	0.000	−0.060	0.626	0.565	0.419	0.374
15	0.357	0.315	0.284	0.263	0.122	0.054	0.000	−0.038	0.372	0.448	0.284	0.264
20	0.256	0.232	0.215	0.203	0.090	0.040	0.000	−0.027	0.264	0.318	0.215	0.204
25	0.199	0.184	0.173	0.166	0.072	0.031	0.000	−0.022	0.204	0.185	0.173	0.166
50	0.094	0.090	0.088	0.086	0.054	0.015	0.000	−0.010	0.096	0.091	0.088	0.086
100	0.046	0.045	0.044	0.044	0.018	0.008	0.000	−0.005	0.046	0.045	0.044	0.044



Slika 2. Grafički prikaz srednje kvadratne pogreške za svaku od četiri metode procjene kao funkciju stupnjeva slobode, [HEDG1985]

2.2. Distribucija i intervali pouzdanosti za veličine učinka

Kad god procjenjujemo parametar, trebamo znati distribuciju metode procjene, kao i intervale pouzdanosti. Često točna distribucija parametrijskih metoda procjena veličine učinka ima relativno komplicirану analitičku formu. Postoji nekoliko izvrsnih aproksimacija za distribuciju procjene veličine učinka koje se mogu koristiti za izračun intervala pouzdanosti ako je kombinirana veličina uzorka $n^E + n^C$ umjerena do velika.

2.2.1. Asimptotska distribucija procjena veličine učinka

Četiri prethodno objašnjene parametrijske metode procjene veličine učinka temeljene na pojedinačnom istraživanju imaju istu distribuciju na velikim uzorcima. Distribucija od $d = J(N-2)\left(\overline{Y}^E - \overline{Y}^C\right)/s$ na velikom uzorku teži prema normalnoj distribuciji. Konkretno, ako se n^E i n^C se jednako povećavaju, tj. ako n^E/N i n^C/N ostaju nepromijenjeni, tada je asimptotska distribucija od d normalna sa srednjom vrijednošću δ i asimptotskom varijancom

$$\sigma_\infty^2(d) = \frac{n^E + n^C}{n^E n^C} + \frac{\delta^2}{2\left(n^E + n^C\right)} \tag{20}$$

Ova formalna asimptotska distribucija se može koristiti za postizanje odlične aproksimacije distribuciju od d za velike uzorke. U praksi, aproksimacija je korištena kao zamjena d sa δ u izrazu (20) za varijancu. Procijenjena varijanca $\hat{\sigma}^2(d)$ je dana sa:

$$\hat{\sigma}^2(d) = \frac{n^E + n^C}{n^E n^C} + \frac{d^2}{2\left(n^E + n^C\right)} \tag{21}$$

$100(1-\alpha)$-postotak intervala pouzdanosti (δ_L, δ_U) za δ je dan sa

$$\delta_L = d - C_{\alpha/2}\hat{\sigma}(d)\,, \quad \delta_U = d + C_{\alpha/2}\hat{\sigma}(d) \tag{22}$$

gdje je $C_{\alpha/2}$ dvosmjerna kritična vrijednost (eng. two-tailed critical value) standardne normalne distribucije.

PRIMJER

Pretpostavimo da proučavamo veličinu učinka g = 0.60 u studiji sa $n^E = n^C = 10$, te da želimo postići 95-postotni interval pouzdanosti za δ.

Podsjećamo se da je $d = J(18)g = 0.57$.

Koristeći (21) za izračunavanje varijance $\hat{\sigma}^2(d)$ od d, dolazimo do $\hat{\sigma}^2(d) = \frac{10+10}{(10)(10)} + \frac{(0.57)^2}{2(10+10)} = 0.208$. Koristeći (22) 95-postotni interval pouzdanosti za δ je

$0.57 \pm 1.96\sqrt{0.208}$ ili $-0.32 \le \delta \le 1.46$. Ova aproksimacija je detaljno obrađena od strane N. L. Johnson i B.L. Welch koji su zaključili da je aproksimacija prilično dobra kada je δ malen i N umjereno velik [JOHN1939].

Empirijski razmjeri intervala pouzdanosti koji sadrže δ za različite značajne razine su dani u Tablici 4. Očekivani intervali pouzdanosti koji sadrže δ su jako blizu uočenim vjerojatnosnim sadržajem intervala pouzdanosti. Aproksimacija distribucije metode procjene veličine učinka za veliki uzorak je ispravna za veličine uzorka koji premašuju 10 ispitanika po grupi, kada je veličina učinka između 0.25 i 1.50.

Tablica 4. Mali uzorak preciznosti intervala pouzdanosti za δ temeljen na normalnoj aproksimaciji distribucije od *d*, [HEDG1985]

Sample size $n^E = n^C$	δ	Mean of d	Variance of d	Proportion of confidence intervals containing δ with nominal significance level					
				0.60	0.70	0.80	0.90	0.95	0.99
10	0.25	0.252	0.206	0.621	0.714	0.813	0.910	0.955	0.991
20	0.25	0.255	0.103	0.604	0.704	0.806	0.903	0.951	0.990
30	0.25	0.246	0.067	0.608	0.708	0.809	0.908	0.954	0.989
40	0.25	0.248	0.051	0.601	0.703	0.800	0.900	0.950	0.991
50	0.25	0.250	0.039	0.612	0.709	0.809	0.904	0.952	0.991
100	0.25	0.251	0.020	0.600	0.705	0.808	0.903	0.949	0.990
10	0.50	0.504	0.214	0.620	0.714	0.807	0.906	0.954	0.990
20	0.50	0.497	0.105	0.609	0.709	0.807	0.904	0.955	0.990
30	0.50	0.500	0.070	0.603	0.699	0.800	0.903	0.952	0.990
40	0.50	0.499	0.052	0.599	0.700	0.803	0.903	0.952	0.991
50	0.50	0.496	0.042	0.597	0.696	0.799	0.904	0.953	0.990
100	0.50	0.498	0.021	0.606	0.698	0.799	0.906	0.954	0.989
10	1.00	0.993	0.241	0.602	0.703	0.801	0.901	0.952	0.992
20	1.00	0.993	0.112	0.609	0.707	0.808	0.907	0.955	0.992
30	1.00	0.993	0.078	0.594	0.697	0.799	0.897	0.952	0.991
40	1.00	0.995	0.057	0.607	0.709	0.805	0.901	0.953	0.992
50	1.00	0.996	0.046	0.603	0.706	0.806	0.900	0.951	0.989
100	1.00	1.000	0.023	0.607	0.702	0.805	0.905	0.953	0.990
10	1.50	1.498	0.282	0.599	0.696	0.797	0.901	0.953	0.990
20	1.50	1.501	0.137	0.601	0.697	0.797	0.898	0.950	0.989
30	1.50	1.505	0.091	0.600	0.699	0.797	0.899	0.950	0.991
40	1.50	1.502	0.069	0.594	0.693	0.796	0.899	0.949	0.990
50	1.50	1.500	0.056	0.593	0.691	0.794	0.897	0.949	0.991
100	1.50	1.500	0.026	0.603	0.705	0.810	0.906	0.955	0.990

2.2.2. Intervali pouzdanosti za veličine učinka temeljene na transformacijama

U nekim slučajevima metoda procjene *d* može biti transformirana u novu slučajnu varijablu koja ima jednostavniju distribuciju. Nakon dobivanja intervala pouzdanosti za transformirani parametar, možemo ga transformirati nazad da dobijemo interval pouzdanosti za δ. U sadašnjem slučaju, zato što varijanca od *d* ovisi o nepoznatom parametru δ, koristimo transformaciju stabiliziranja varijance (eng. variance-stabilizing transformation)

$$h(d) = \sqrt{2}\sinh^{-1}\frac{d}{a} = \sqrt{2}\log\left(\frac{d}{a} + \sqrt{\frac{d^2}{a^2}+1}\right) \tag{23}$$

gdje je $\sinh^{-1}$ inverzna hiperbolička sinus funkcija, i

$$a = \sqrt{4 + 2\left(n^E / n^C\right) + 2\left(n^C / n^E\right)} \tag{24}$$

Točna forma transformacije ovisi o ravnoteži n^E / n^C u eksperimentu. Zbog pojednostavljenja, označimo transformiranu vrijednost procjene sa h i transformiranu vrijednost parametra sa η: $h = h(d)$ i $\eta = h(\delta)$. Tada transformirana procjena h, kada je na odgovarajući način normalizirana, ima aproksimiranu standardnu distribuciju. Preciznije,

$$\sqrt{N}(h-\eta) \approx N(0,1) \tag{25}$$

Dakle, $100(1-\alpha)$-postotak intervala pouzdanosti (η_L, η_U) za η je dan sa

$$\eta_L = h - C_{\alpha/2} / \sqrt{N}, \quad \eta_U = h + C_{\alpha/2} / \sqrt{N}, \tag{26}$$

gdje je $C_{\alpha/2}$ dvosmjerna kritična vrijednsot (eng. two-tailed critical value) standardne normalne distribucije. Granice pouzdanosti (eng. confidence limits) η_L i η_U su invertirane da stvaraju interval pouzdanosti (δ_L, δ_U) za δ tražeći vrijednosti δ_L i δ_U koje odgovaraju vrijednostima η_L i η_U:

$$\delta_L = h^{-1}(\eta_L) \qquad \text{i} \qquad \delta_U = h^{-1}(\eta_U) \tag{27}$$

gdje je $h^{-1}(x) = a\sinh\left(x/\sqrt{2}\right)$.

PRIMJER

Vratimo se na naš primjer eksperimenta sa $n^E = n^C = 10$ koji dopušta uzorak veličine učinka od g = 0.60 i nepristranu procjenu veličine učinka od d = 0.57. Izračunali smo 95-postotni interval pouzdanosti za δ koristeći (23) sa $a = \sqrt{4 + 2(10/10) + 2(10/10)} = \sqrt{8}$, $h = h(d) = \sqrt{2}\sinh^{-1}\left(0.57/\sqrt{8}\right) = 0.283$.

Izraz (26) daje granice pouzdanosti η_L i η_U za η: $\eta_L = 0.283 - 1.96\sqrt{1/(10+10)} = -0.155$, $\eta_U = 0.283 + 1.96\sqrt{1/(10+10)} = 0.721$.

Koristeći $h^{-1}(x) = \sqrt{8}\sinh\left(x/\sqrt{2}\right)$, dobivamo granice pouzdanosti $\delta_L = \sqrt{8}\sinh\left(-0.155/\sqrt{2}\right) = -0.31$, $\delta_U = \sqrt{8}\sinh\left(0.721/\sqrt{2}\right) = 1.50$ i interval pouzdanosti $-0.31 \le \delta \le 1.50$.

Drugačija transformacija je predložena od strane Kaemera. Ova transformacija, [KRAE1983], iskorištava činjenicu da su distribucija procjene veličine učinka (necentralna t-distribucija) i distribucija „product-moment“ koeficijenta korelacije povezane [KRAE1979]. Kraemer, [KRAE1983], pokazuje da ako je

$$r = d/\left(d^2 + v\right)^{1/2} \qquad \text{i} \qquad \rho = \delta/\left(\delta^2 + v\right)^{1/2} \tag{28}$$

gdje je $v = N(N-2)/n^E n^C$, tada transformirana slučajna varijabla

$$u = u(r, \rho) = (\rho - r)(1 - r\rho) \tag{29}$$

ima otprilike nul distribuciju od "product-moment"" koeficijenat korelacije distribucije. Pokazuje se da (slučajna) varijabla

$$\sqrt{N-2}\,u/\left(1-u^2\right)^{1/2} \tag{30}$$

ima aproksimativnu Studentovu t-distribuciju sa N-2 stupnja slobode, što može biti korišteno za dobivanje granica pouzdanosti iz percentilnih točaka (eng. percentage points) od (centralne) t-distribucije [KRAE1979]..

Nakon dobivanja (bilo kojom metodom) granica pouzdanosti u_L i u_U, dobivamo granice pouzdanosti ρ_L i ρ_U tražeći vrijednosti od ρ koje odgovaraju vrijednostima u_L i u_U :

$$\rho_L = (u_L - r)/(u_L r - 1), \qquad \rho_U = (u_U - r)/(u_U r - 1) \tag{31}$$

Napokon, dobivamo interval pouzdanosti od δ transformirajući ρ_L i ρ_U u δ_L i δ_U :

$$\delta_L = \rho_L \sqrt{v}/\left(1-\rho_L^2\right)^{1/2}, \qquad \delta_U = \rho_U \sqrt{v}/\left(1-\rho_U^2\right)^{1/2} \tag{32}$$

PRIMJER

Ako se vratimo na primjer istraživanja sa $n^E = n^C = 10$ i dobivene veličine učinka g = 0.60, te nepristrane procjene veličine učinka d = 0.57. za računanje 95-postotne pouzdanosti za δ, prvo trebamo izračunati $v = 18/20\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 3.6$, te $r = 0.57/\left(0.57^2 + 3.6\right)^{1/2} = 0.288$.

Koristeći činjenicu da $u = (r-\rho)(1-r\rho)$ ima aproksimativno nul distribuciju od koeficijenta korelacije za n = 20 promatranja, možemo dobiti 2.5 i 97.5 percentilne točke iz tablice kritičnih vrijednosti korelacije. Ove kritične vrijednosti su $u_L = -0.444$ i $u_U = 0.444$. Uvrštavajući ove vrijednosti u (31), dobivamo

$\rho_L = (0.444 - 0.288)/[(0.444)(0.288) - 1] = -0.179$
$\rho_U = (-0.444 - 0.288)/[-(0.444)(0.288) - 1] = 0.649$

Uvrštavajući vrijednosti od ρ_L i ρ_U u (32) rezultira s $\delta_L = -0.179\sqrt{3.6}/\left[1-(0.179)^2\right]^{1/2} = -0.34$, $\delta_U = 0.649\sqrt{3.6}/\left[1-(0.649)^2\right]^{1/2} = 1.62$. Stoga je 95-postotni interval pouzdanosti za δ koristeći trenutnu metodu je $-0.34 \le \delta \le 1.62$

Relativno malo je poznato o točnosti ove metode transformacije. Moguće su veće značajne razlike za veće veličine učinka ili za jako male veličine uzorka. Nijedna metoda za stvaranje intervala pouzdanosti temeljena na transformacijama nije bolja od metode temeljene na asimptotskoj distribuciji od g.

2.2.3. Točni intervali pouzdanosti za veličine učinka

Iako neke metode rezultiraju točnim intervalima pouzdanosti za δ, mi raspravljamo o aproksimacijama, jer je točna metoda prilično komplicirana. Preporučamo upotrebu aproksimativnog intervala pouzdanosti danog u (22) kada je manji od n^E i n^C umjeren ili velik (veći od 10). Za manje veličine uzorka, ovi asimptotski intervali pouzdanosti, mogu biti nedovoljno precizni. Točni intervali pouzdanosti za δ su dobiveni koristeći točnu distribuciju veličine učinka metode procjene *g*.

Statistika $g\sqrt{n^E n^C / N}$ ima necentralnu t-distribuciju sa N-2 stupnjeva slobode i necentralni parametar $\delta\sqrt{n^E n^C / N}$. Označimo funkciju kumulativne distribucije metode procjene g sa $F(g; N-2, \delta)$. Tada granice pouzdanosti δ_L i δ_U su definirane kao rješenja jednadžbi:

$$F(g; N-2, \delta_L) = \alpha/2, \qquad F(g; N-2, \delta_U) = 1-\alpha/2 \tag{33}$$

Ovdje nailazimo na različite integrale niza prikaza funkcija distribucije, koji se ne mogu lako ručno izračunati.

2.3. Robusna i neparametrijska procjena veličine učinka

Do sada smo koristili parametrijski model (1) za izvođenje metoda procjene veličine učinka. Uobičajeni model je često prikladan, ali ne uvijek i realističan. Kada se ne susreću pretpostavke za normalnost podataka, veličina učinka izračunavana iz podataka prije i poslije transformacije može se značajno ne slagati. Jedan pristup metode procjene δ je taj da koristimo veličinu učinka izračunatu pomoću transformiranih podataka, koji su vjerojatno, razumno normalni. Alternativno je da razvijemo metodu procjene veličine učinka koja je nepromijenjena transformacijama opažanja. To je bila motivacija za neparametrijske metode procjene veličine učinka koje je predložio Kraemer i Andrews, [KRAE1982] i kao proširenje metoda procjene koje su dali Hedges i Olkin [HEDG1984].

Standardna razlika aritmetičkih sredina omogućava preklapanje distribucija rezultata u eksperimentalnoj i kontrolnoj grupi. Ali kada podaci nisu normalno distribuirani, interpretacija veličine učinka u terminima percentila standardne normalne distribucije je neupotrebljiva. Ako podaci mogu biti transformirani tako da postanu više normalno distribuirani, tada je važno imati metodu procjene koja ostaje nepromijenjena u procesu transformacije.

Započinjemo raspravljajući o metodama procjene veličine učinka koristeći robusne metode procjene za μ^E, μ^C i σ, i proučavajući neparametrijske metode procjene veličine učinka, [KRAE1982].

2.3.1. Robusna procjena veličine učinka naspram vanjskih vrijednosti

Veličina učinka δ je omjer razlike aritmetičkih sredina prema standardnoj devijaciji. Da bi dobili metodu procjene veličine učinka koja je otporna na vanjske vrijednosti, procijenimo i razliku aritmetičkih sredina i standardnu devijaciju pomoću robusnih metoda procjene. Npr, srednja vrijednost populacije svake grupe može biti procijenjena medijanom, a standardna devijacija svake grupe može biti procijenjena rangom. Možemo procijeniti

$$\tilde{\delta} = \left[Med\left(Y^E\right) - Med\left(Y^C\right)\right] / \tilde{\sigma} \tag{34}$$

gdje su $Med\left(Y^E\right)$ i $Med\left(Y^C\right)$ medijani zapažanja u eksperimentalnoj i kontrolnoj grupi, i

$$\tilde{\sigma} = a_2 y_{(2)}^C + ... + a_{n-1} y_{(n-1)}^C \tag{35}$$

gdje su $y_{(1)}^C \le y_{(2)}^C \le ... \le y_{(n-1)}^C \le y_{(n)}^C$ uređene vrijednosti u kontrolnoj grupi. Optimalni koeficijenti $a_2,...,a_{n-1}$ su odabrani tako da minimaliziraju varijancu od $\tilde{\sigma}$, [SARH1962]. U svrhu umanjivanja osjetljivosti prema ekstremnim vrijednostima, najmanja i najveća vrijednost su izostavljene u metodi procjene od σ. Kako procjene $\tilde{\delta}$ od δ dane u (34) ne uključuju dva ekstremna opažanja, one su otporne na pomake najekstremnijih opažanja.

2.3.2. Neparametrijske metode procjene veličine učinka

Neparametrijski pristup u procjeni veličine učinka je predložen od strane Kraemera i Andrewsa (KA). Njihova tehnika zahtjeva da rezultati inicijalnih (x) i završnih (y) mjerenja budu dostupni za svakog pojedinca u eksperimentalnoj i kontrolnoj grupi (Tablica 5.). KA metoda procjene služi za razjašnjavanje kako se dodatne metode procjene mogu konstruirati, [KRAE1982] Za svaku grupu određujemo proporciju uzorka (eng. sample proportion) $\hat{p}$ od x rezultata koji leže ispod medijana y rezultata. Ova proporcija odgovara standardnoj normalnoj devijaciji $\tilde{\delta}$:

$$\Phi\left(\hat{\delta}\right) = \hat{p} \qquad \text{ili} \qquad \hat{\delta} = \Phi^{-1}\left(\hat{p}\right) \tag{36}$$

gdje je $\Phi(x)$ standardna normalna kumulativna distribucija funkcije. Upotreba ove procedure za eksperimentalnu i kontrolnu grupu procjenjuje veličine učinka $\hat{\delta}^E$ i $\hat{\delta}^C$ za eksperimentalnu i kontrolnu grupu redom. Ukupna veličina učinka $\hat{\delta}$ je dobivena kao

$$\hat{\delta} = \hat{\delta}^E - \hat{\delta}^C \tag{37}$$

Tablica 5. Podaci inicijalnih i završnih mjerenja sistoličkog krvnog tlaka dobiveni na 40 ispitanika

	Eksperimentalna grupa		Kontrolna grupa	
	Inicijalna mjerenja	Završna mjerenja	Inicijalna mjerenja	Završna mjerenja
	134	130	139	130
	135	131	140	131
	135	135	141	144
	136	136	143	146
	145	136	151	128
	147	138	152	156
	148	124	152	161
	150	126	153	162
	151	104	153	160
	153	142	154	131
	153	114	154	158
	155	166	159	166
	156	153	160	150
	158	169	160	186
	162	127	162	188
	165	130	163	153
	167	120	165	144
	168	121	169	147
	179	149	175	169
	180	150	176	170
Medijan	153	133	154	154.5

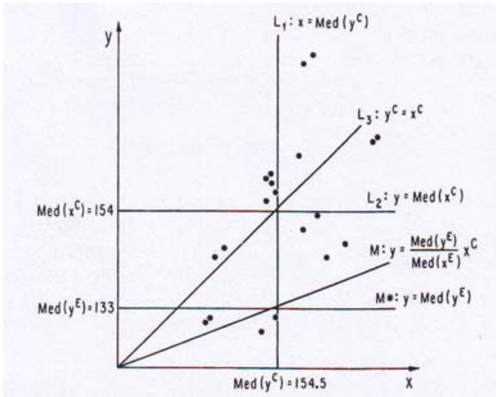
PRIMJER

Podaci od *sistoličkom krvnom tlaku* su dani u [KRAE1982], i reproducirani su u Tablici 5. rezultatima inicijalnih (x) i završnih (y) mjerenja tretmana ispitanika eksperimentalne i kontrolne grupe te su prikazani u uzlaznom nizu rezultata inicijalnih mjerenja. Medijan y rezultata u kontrolnoj grupi je 154.5. Broj x rezultata manjih od 154.5 je 11, pa slijedi $\hat{\delta}^C = \Phi^{-1}\left(\frac{11}{20}\right) = 0.126$.

Na sličan način dolazimo do $\hat{\delta}^E = \Phi^{-1}(0)$. U ovakvim slučajevima Kraemer i Andrews preporučuju korištenje $p = 1/(n+1) = \frac{1}{21} = 0.048$ umjesto nule, tako da je $\hat{\delta}^E = \Phi^{-1}(0.048) = -1.665$. Kombinirajući $\hat{\delta}^E$ i $\hat{\delta}^C$, dolazimo do procjene veličine učinka $\hat{\delta} = \hat{\delta}^E - \hat{\delta}^C = -1.665 - 0.126 = -1.79$.

2.3.3. Procjene temeljene na razlikama kontrolnih i eksperimentalnih proporcija

Logika neparametrijskih metoda procjene veličine učinka je najbolje ilustrirana grafičkim prikazom rezultata inicijalnih (x) mjerenja napsram rezultata završnih (y) mjerenja (Slika 3.).



Slika 3. Rezultati inicijalnih (x) mjerenja naspram rezultata završnih (y) mjerenja, [HEDG1985]

KA procedura promatra proporciju *p* točke (x,y), tako da je x manji od medijana od y, tj. izračunava udio točaka na lijevoj strani od pravca L_1 za kontrolnu grupu. Nakon toga koristimo inverznu normalnu kumulativnu distribuciju funkcije Φ^{-1} da bi dobili procjenu veličine učinka mjerenja iz *p*. KA metoda procjene je

$$\hat{\delta}_1 = \Phi^{-1}\left(\hat{p}_1^E\right) - \Phi^{-1}\left(\hat{p}_1^C\right) \tag{38}$$

Kada su promatranja normalno distribuirana, $\hat{\delta}_1$ je procjena od

$$\frac{\mu_y^E - \mu_x^E}{\sigma_x^E} - \frac{\mu_y^C - \mu_x^C}{\sigma_x^C} \tag{39}$$

gdje su μ_y^E, μ_x^E i σ_x^E završna srednja vrijednost, inicijalna srednja vrijednost i inicijalna standardna devijacija eksperimentalne grupe, redom, i μ_y^C, μ_x^C i σ_x^C su analogni parametri za kontrolnu grupu.

KA proceduru karakterizira to da što je veći učinak tretmana, manja je proporcija rezultata inicijalnih mjerenja eksperimentalne grupe koja premašuje medijan završnih mjerenja.

Povezana metoda procjene $\hat{\delta}_2$ je temeljena na proporciji $\hat{p}_2^C$ rezultata završnih mjerenja u kontrolnoj grupi koja premašuje medijan inicijalnih mjerenja u kontrolnoj grupi, i proporcija $\hat{p}_2^E$ završnih rezultata mjerenja u eksperimentalnoj grupi koja premašuje medijan inicijalnih mjerenja u eksperimentalnoj grupi. Za kontrolnu grupu relevantna proporcija odgovara proporciji promatranja iznad pravca L_2. Gore navedeni opis je preveden u formulu za $\hat{\delta}_2$:

$$\hat{\delta}_2 = \Phi^{-1}\left(\hat{p}_2^E\right) - \Phi^{-1}\left(\hat{p}_2^C\right) \tag{40}$$

Za Kraemer-Andrews podatke je $\hat{p}_2^E = \frac{2}{20} = 0.10$ i $\hat{p}_2^C = \frac{10}{20} = 0.50$. Dakle, $\hat{\delta}_2^E = \Phi^{-1}(0.10) = -1.282$, $\hat{\delta}_2^C = \Phi^{-1}(0.50) = 0.0$ i $\hat{\delta}_2 = \hat{\delta}_2^E - \hat{\delta}_2^C = -1.282$.

Kada su promatranja normalno distribuirana, $\hat{\delta}_2$ je procjena od

$$\frac{\mu_y^E - \mu_x^E}{\sigma_y^E} - \frac{\mu_y^C - \mu_x^C}{\sigma_y^C} \tag{41}$$

gdje su σ_y^E i σ_y^C završna mjerenja standardne devijacije eksperimentalne i kontrolne grupe.

2.3.4. Procjene temeljene na povećanjima rezultata u eksperimentalnoj grupi u odnosu na kontrolnu grupu

Drugačiji pristup je u upotrebi proporcija rezultata u kontrolnoj grupi koji su se povećali manje nego što je očekivano u eksperimentalnoj grupi. Ova ideja je ilustrirana grafički na Slici 3.

Određite medijan rezultata inicijalnih i završnih mjerenja u eksperimentalnoj grupi, a na podacima kontrolne grupe nacrtajte pravac M od ishodišta (0,0) kroz točku $\left[Med\left(x^E\right), Med\left(y^E\right)\right]$. Jednadžba pravca M je

$$y^C = \frac{Med\left(y^E\right)}{Med\left(x^E\right)} x^C \tag{42}$$

Proporcija $\hat{q}_1$ točke iznad pravca M rezentira proporcije ispitanika u kontrolnoj grupi čiji su rezultati manje povećani nego oni od njihovih „dvojnika“ u eksperimentalnoj grupa. Metoda procjene veličine učinka $\hat{y}_1$ je dana sa

$$\hat{y}_1 = \Phi^{-1}\left(\hat{q}_1\right) \tag{43}$$

Za podatke u Tablici 5. je $\hat{q}_1 = \frac{2}{20} = 0.10$ i $\hat{y}_1 = \Phi^{-1}(0.10) = -1.282$. Ako su rezultati normalno distribuirani, $\hat{y}_1$ daje procjenu $\left(\mu_y^E - \mu_y^C\right)/\sigma_y^C$.

Simetrična verzija se dobije crtajući na grafičkom prikazu eksperimentalnih podataka pravac između ishodišta (0,0) i točke $\left[Med\left(x^C\right), Med\left(y^C\right)\right]$ reprezentirajući medijane rezultata završnih mjerenja u kontrolnoj grupi. Definiramo $\hat{q}_2$ kao proporciju rezultata u eksperimentalnoj grupi koji leže iznad pravca, za koje je

$$y^E > \frac{Med\left(y^C\right)}{Med\left(x^C\right)} x^E \tag{44}$$

Metoda procjene veličine učinka $\hat{\gamma}_2$ je dana sa

$$\hat{\gamma}_2 = \Phi^{-1}(\hat{q}_2) \tag{45}$$

Za podatke u Tablici 5. je $\hat{q}_2 = \dfrac{2}{20} = 0.10$ i $\hat{\gamma}_2 = \Phi^{-1}(0.10) = -1.282$. Ako su rezultati normalno distribuirani tada $\hat{\gamma}_2$ daje procjenu $\left(\mu_y^E - \mu_y^E\right)/\sigma_y^E$.

2.3.5. Neparametrijske metode procjene koje uključuju samo završne rezultate

Ako su slučajnim odabirom ispitanici podijeljeni u kontrolnu i eksperimentalnu grupu, tada očekujemo da distribucija rezultata inicijalnih mjerenja bude slična u obje grupe. U ovom slučaju mogu se koristiti samo rezultati završnih mjerenja. Izračunavamo proporciju $\hat{q}_1^*$ rezultata kontrolne grupe koji su manji od medijana eksperimentalne grupe (ispod pravca M^*) i transformiramo proporciju u procjenu $\hat{\gamma}_1^*$ veličine učinka

$$\hat{\gamma}_1^* = \Phi^{-1}(\hat{q}_1^*) \tag{46}$$

Proučavajući Tablicu 5., vidimo da je $\hat{q}_1^* = \dfrac{4}{20} = 0.20$ za podatke Kraemer-Andrews, što odgovara $\hat{\gamma}_1^* = \Phi^{-1}(0.20) = -1.038$. Kada su promatranja nezavisno normalno distribuirana, $\hat{\gamma}$ je metoda procjene od $\left(\mu_y^E - \mu_y^C\right)/\sigma_y^C$, što je razlika aritmetičkih sredina standardizirana standardnom devijacijom σ_y^C rezultata kontrolne grupe.

Simetrična metoda procjene se dobije koristeći proporciju $\hat{q}_2^*$ od rezultata završnih mjerenja u eksperimentalnoj grupi koji premašuju rezultat medijana kontrolne grupe. Proporcija $\hat{q}_2^*$ je transformirana u metodu procjene veličine učinka kao

$$\hat{\gamma}_2^* = \Phi^{-1}(\hat{q}_2^*) \tag{47}$$

Kraemer-Andrews podaci u Tablici 5. imaju $\hat{q}_2^* = \dfrac{2}{20} = 0.10$, što odgovara $\hat{\gamma}_2^* = \Phi^{-1}(0.10) = -1.282$.

Kada su promatranja nezavisno normalno distribuirana, $\hat{\gamma}_2^*$ je metoda procjene od $\left(\mu_y^E - \mu_y^C\right)/\sigma_y^E$, što je razlika aritmetičkih sredina eksperimentalne i kontrolne grupe standardizirane pomoću standardne devijacije σ_y^E rezultata eksperimentalne grupe.

Dakle, ako su rezultati normalno distribuirani i standardne devijacije eksperimentalne i kontrolne grupe jednake, tada i $\hat{\gamma}_1^*$ i $\hat{\gamma}_2^*$ imaju procjene jednake veličine. Kada su podaci inicijalnih i završnih mjerenja dostupni, radije ćemo koristiti $\hat{\gamma}_1$ nego $\hat{\gamma}_2$, a kada su samo podaci završnih mjerenja dostupni koristit ćemo $\hat{\gamma}_1^*$.

2.3.6. Odnosi između metoda procjene

Sedam neparametrijskih metoda procjena veličine učinka su različite u smislu da različito upotrebljavaju podatke i procjenjuju neznatno različite parametre kada su podaci normalno distribuirani. Proporcije koje su upotrebljavane u izračunima svake od metoda procjene su dani u Tablici 6. zajedno sa procijenjenom veličinom učinka, pod pretpostavkom da su promatranja normalno distribuirana. Primijetimo da $\hat{\gamma}_1$ i $\hat{\gamma}_1^*$ imaju procjenu jednake veličine, kao i $\hat{\gamma}_2$ i $\hat{\gamma}_2^*$. Kada je $\sigma_y^E = \sigma_y^C$, četiri metode procjene $\hat{\gamma}_1$, $\hat{\gamma}_1^*$, $\hat{\gamma}_2$ i $\hat{\gamma}_2^*$ imaju identičnu veličinu procjene. Slično, $\hat{\delta}_1$ i $\hat{\delta}_2$ imaju jednake parametre procjene ako je $\sigma_x^E = \sigma_y^C$ i $\sigma_y^E = \sigma_y^C$.

Malo je poznato o svojstvima uzoraka kod neparametrijskih metoda procjene veličine učinka. Oni su vjerojatno manje učinkoviti nego njihovi parametrijski dvojnici. Neparametrijske metode procjene bi trebale biti korištene samo kada imamo razloga vjerovati da su parametrijske pretpostavke prekršene. Neparametrijke metode procjene veličine učinka su izračunate samo kada su sva originalna promatranja dostupna. Ovo ozbiljno ograničava primjenjivost ovih metoda procjene u praksi, otkad su neobrađeni podaci rijetko dostupni u meta-analizama.

Tablica 6. Kratak pregled neparametrijskih metoda procjene, [HEDG1985]

Metoda procjene	Proporcije točaka za izračunavanje	Procjenjenost pod normalnim pretpostavkama
$\hat{\delta}_1^C$	$x^C < Med(y^C)$	$(\mu_y^C - \mu_x^C)/\sigma_x^C$
$\hat{\delta}_1^E$	$x^E < Med(y^E)$	$(\mu_y^E - \mu_x^E)/\sigma_x^E$
$\hat{\delta}_1 = \hat{\delta}_1^E - \hat{\delta}_1^C$		$\delta_1^E - \delta_1^C$
$\hat{\delta}_2^C$	$y^C > Med(x^C)$	$(\mu_y^C - \mu_x^C)/\sigma_y^C$
$\hat{\delta}_2^E$	$y^E > Med(x^E)$	$(\mu_y^E - \mu_x^E)/\sigma_y^E$
$\hat{\delta}_2 = \hat{\delta}_2^E - \hat{\delta}_2^C$		$\delta_2^E - \delta_2^C$
$\hat{\delta}_3^C$	$y^C > x^C$	$(\mu_y^C - \mu_x^C)/\sigma_*^C$
$\hat{\delta}_3^E$	$y^E > x^E$	$(\mu_y^E - \mu_x^E)/\sigma_*^E$
$\hat{\delta}_3 = \hat{\delta}_3^E - \hat{\delta}_3^C$		$\delta_3^E - \delta_3^C$
$\hat{\gamma}_1$	$y^C < \frac{Med(y^E)}{Med(x^E)} x^C$	$(\mu_y^E - \mu_y^C)/\sigma_y^C$
$\hat{\gamma}_2$	$y^E > \frac{Med(y^C)}{Med(x^C)} x^E$	$(\mu_y^E - \mu_y^C)/\sigma_y^E$
$\hat{\gamma}_1^*$	$y^C < Med(y^E)$	$(\mu_y^E - \mu_y^C)/\sigma_y^C$
$\hat{\gamma}_2^*$	$y^E > Med(y^C)$	$(\mu_y^E - \mu_y^C)/\sigma_y^E$

3. Parametrijske procjene veličine učinka iz niza eksperimenata

Ako se iz niza od *k* istraživanja s razlogom može očekivati da imaju zajedničku veličinu učinka δ, tada trebamo izvući informacije iz svih istraživanja s ciljem da dobijemo kombiniranu procjenu od δ. Postoje načini kako to napraviti. Ovdje prikazujemo nekoliko metoda za dobivanje zajedničke metode procjene od δ iz niza studija s umjerenim do velikim uzorcima.

Jedna od jednostavnijih zajedničkih procjena je prosjek procjena dobivenih iz pojedinačnog istraživanja, a zbog jednostavnosti, ova metoda je često korištena. Ukoliko istraživanja nemaju zajedničku veličinu učinka, onda bi trebao postojati barem težinski faktor. Ovo postaje jasnije kada primijetimo da varijanca metode procjene ovisi o veličini uzorka, tako da su procjene iz istraživanja sa većim uzorkom mnogo preciznije od onih s manjim uzorkom.

Dvije metode su prezentirane za dobivanje optimalne kombinacije procjene veličine učinka iz niza studija:
(i) težinska linearna kombinacija metoda procjene iz različitih istraživanja
(ii) metoda procjene maksimalnih vjerojatnosti.

Oba dvije metode su pokazale da imaju jednaku asimptotsku distribuciju. Obje metode predlažu korištenje transformacije procjene veličine učinka. Metoda temeljena na linearnim kombinacijama je jednostavnija, te zahtjeva manje izračunavanja.

3.1. Model i notacija

Statistička svojstva procedure za kombiniranje rezultata iz niza eksperimenata ovise o strukturnom modelu rezultata eksperimenta. Strukturni model zahtjeva da je svaki eksperiment temeljen na *skali odaziva* (eng. response scale) iz prikupljenih poopćenih mjera, tj. svaka skala odaziva je linearna transformacija skala odziva s jediničnom varijancom unutar grupa.

Populacijski parametar označavamo sa δ, i to je nepristrana procjena pomoću d. Glass procjenjuje δ za svako istraživanje, te kombinira procjene kroz istraživanja [GLAS1976]. Statističke analize u takvim istraživanjima tipično uključuju upotrebu t- ili F-testa za testiranje razlika između srednjih vrijednosti grupa. Ukoliko su pretpostavke za valjanost t-testa ispunjene, moguće je izvesti točnu procjene od δ. Počinjemo eksplicitnim prikazivanjem pretpostavki. Pretpostavimo da su podaci koji su proizašli iz niza od *k* nezavisnih istraživanja, u kojem svako istraživanje uspoređuje eksperimentalnu grupu (E) s kontrolnom grupom (C), kako je ispod prikazano:

BAZA PODATAKA:

Studija	Promatranja	
	Eksperimentalna	Kontrolna
1	$Y_{11}^E,...Y_{1n_1^E}^E$	$Y_{11}^C,...Y_{1n_1^C}^C$
$\vdots$	$\vdots$	$\vdots$
k	$Y_{k1}^E,...Y_{kn_k^E}^E$	$Y_{k1}^C,...Y_{kn_k^C}^C$

PARAMETRI:

Studija	Eksperimentalna		Kontrolna		Veličina učinka
	Srednja vrijednost	Varijanca	Srednja vrijednost	Varijanca	
1	μ_1^E	σ_1^2	μ_1^C	σ_1^2	$\delta_1 = \left(\mu_1^E - \mu_1^C\right)/\sigma_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	μ_k^E	σ_k^2	μ_k^C	σ_k^2	$\delta_k = \left(\mu_k^E - \mu_k^C\right)/\sigma_k$

Pretpostavimo da su za i-to istraživanje eksperimentalna promatranja $Y_{i1}^E,...Y_{in_i^E}^E$ normalno distribuirana sa zajedničkom srednjom vrijednošću μ_i^E i zajedničkom varijancom σ_i^2. Slično, promatranja u kontrolnoj grupi $Y_{i1}^C,...Y_{in_i^C}^C$ su normalno distribuirana sa zajedničkom srednjom vrijednošću μ_i^C i zajedničkom varijancom σ_i^2. Ispuštajući eksponent E ili C na varijancama, prešutno smo pretpostavili da su varijance eksperimentalne i kontrolne grupe, za i-ti eksperiment, jednake. Ili kraće,

$$\begin{aligned} Y_{ij}^E &\approx N\left(\mu_i^E, \sigma_i^2\right), \qquad j = 1,...,n_i^E, \quad i = 1,...,k \qquad \text{i} \\ Y_{ij}^C &\approx N\left(\mu_i^C, \sigma_i^2\right), \qquad j = 1,...,n_i^C, \quad i = 1,...,k \end{aligned} \tag{48}$$

U ovoj notaciji, veličina učinka δ_i za i-to istraživanje je definirana kao

$$\delta_i = \left(\mu_i^E - \mu_i^C\right)/\sigma_i \tag{49}$$

Iako se vrijednosti populacijskih srednjih vrijednosti i standardne devijacije mogu promijeniti pod utjecajem linearne transformacije promatranja, veličina učinka δ ostaje neizmijenjena. Obratno, ako dva mjerenja nisu linearno povezana, tada će populacija rezultata testova, generalno, rezultirati različitim veličinama učinka.

Pretpostavimo da je svaki eksperiment u nizu replikacija drugog, tj. eksperimenti mjere jednake izlazne konstrukcije i razlikuju se samo po skali odaziva i veličini uzorka. Populacijska vrijednost učinka eksperimentalnog faktora će biti identična ako su sve studije upotrebljavale jednaku skalu odaziva za zavisnu varijablu. Ova pretpostavka znači da je $\delta_1 = \delta_2 = ... = \delta_k = \delta^*$.

3.2. Težinske linearne kombinacije procjena

Kada niz od k nezavisnih istraživanja dijeli zajedničku veličinu učinka δ, prirodna procjena od δ je udruživanje procjena iz svake od studija. Ako je veličina uzorka studija različita, tada su metode procjene iz većih studija mnogo preciznije, nego one iz manjih studija. U ovom slučaju s razlogom dajemo veću težinu (eng. weight) preciznijim metodama procjene kod udruživanja. Ovo dovodi do uvođenja težinske metode procjene dane formulom:

$$d_W = w_1 d_1 + ... + w_k d_k \tag{50}$$

gdje su $w_1,...,w_k$ nenegativni težinski faktori.

3.2.1. Procjenjivanje težine

Težine koja minimiziraju varijancu od d_W daju težinu koja je obrnuto proporcionalna varijanci u svakom istraživanju. Intuitivno je jasno da manje varijance daju veće težine. Dakle,

$$w_i = \frac{1}{\sigma^2(d_i)} / \sum_{j=1}^{k} \frac{1}{\sigma^2(d_j)} \tag{51}$$

gdje je $\sigma^2(d_i)$ varijanca od d_i. Zbog primjene teorije na velikim uzorcima, težine su

$$w_i = \frac{1}{\sigma_\infty{}^2(d_i)} / \sum_{j=1}^{k} \frac{1}{\sigma_\infty{}^2(d_j)} \tag{52}$$

gdje je $\sigma_\infty{}^2(d_i)$ varijanca velikog uzorka , dana u izrazu (20).

Praktični problem koji se javlja kod izračuna najpreciznije težinske procjene je taj da i-ta težina zavisi o varijanci d_i, koja zavisi o nepoznatom parametru δ, kojeg moramo procijeniti. Da bi riješili ove poteškoće, upotrebljavati ćemo težine koje su temeljene na aproksimaciji varijanci i ne ovise o δ. Ovaj način rezultira zajedničkom metodom procjene koja je nepristrana, i često manje precizna nego kada se koristi optimalna težina. Npr. aproksimativna težina je dana sa

$$w_i \cong \frac{\tilde{n}_i}{\sum_{j=1}^{k} \tilde{n}_j} \tag{53}$$

gdje je $\tilde{n}_i = n_i^E n_i^C / (n_i^E + n_i^C)$. Ovako izvedena težina je blizu optimalne kada je δ blizu nule i kad su $\tilde{n}_i$ veliki.

Ukoliko je procjena od δ dostupna (nije nula), tada težina može biti procijenjena unošenjem vrijednosti od δ u formulu (20) za varijancu d_i i koristeći formulu (52) za w_i. Općenito, rezultat će biti nepristrana zajednička metoda procjene od δ, što je neznatno manje precizno nego najpreciznija težinska metoda procjene.

Težinska metoda procjene od δ temeljena na upotrebljavanju procjene uzorka od δ za izračunavanje težina za svaku studiju, je dana sa

$$d_+ = \sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2(d_i)} / \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)} \tag{54}$$

gdje je $\hat{\sigma}^2(d)$ definirana u izrazu (21). Metoda procjene d_+ je dobivena izračunavanjem težine koristeći d_i za procjenjivanje δ_i. Također, d_i je nepristrana, a d_+ nije. Pristranost od d_+ je malena kod većih uzoraka i teži nuli kako se veličina uzorka povećava.

Metoda procjene d_+ se može modificirati, zamjenjujući d_i sa d_+, u izrazu za $\hat{\sigma}^2(d_i)$ i ponavljanjem. Definirajmo $d_+^{(0)}$ da bude d_+, za dobivanje prve iteracije:

$$d_+^{(1)} = \sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2\left(d_i / d_+^{(0)}\right)} / \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2\left(d_i / d_+^{(0)}\right)} \tag{55}$$

I sada definiramo $d_+^{(2)}$ sa

$$d_+^{(2)} = \sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2\left(d_i / d_+^{(1)}\right)} / \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2\left(d_i / d_+^{(1)}\right)} \tag{56}$$

gdje je $\hat{\sigma}^2(d_i / X)$ je procjena od $\hat{\sigma}_\infty{}^2(d_i)$ dobivena zamjenjujući d sa X i izrazu (21) za varijancu od d. Za bilo koju razinu l, procjena sljedeće razine je dobivena iz

$$d_+^{(l+1)} = \sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2\left(d_i / d_+^{(l)}\right)} / \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2\left(d_i / d_+^{(l)}\right)}, \qquad l = 0,1,2,... \tag{57}$$

Iterirana metoda procjene $d_+^{(l)}$ teži da bude manje pristrana nego $d_+^{(l-1)}$. Ako je veličina učinka homogena kroz eksperimente, proces iteracije obično ne mijenja mnogo procjene.

Asimptotska distribucija od d_+ se može koristiti za dobivanje velikih uzoraka intervala pouzdanosti za δ, temeljenih na d_+. Definicija velikih uzoraka u ovom slučaju znači da individualna veličina uzorka n_i^E i n_i^C, $i = 1,...,k$, raste po istoj stopi, i da je $N = \sum\left(n_i^E + n_i^C\right)$ velik.

3.2.3. Preciznost aproksimacije velikog uzorka za distribuciju težinske metode procjene veličine učinka

Preciznost intervala pouzdanosti i testa hipoteze da $\delta^* = 0$ ovisi o preciznosti normalne aproksimacije velikog uzorka za distribuciju od d_+. Distribucija metode procjene d_+ je kombinacija od $d_1,...,d_k$. Preciznost normalnih aproksimacija velikih uzoraka prema distribuciji od d_i sugerira da će aproksimacija za distribuciju od d_+ biti razumno precizna u mnogim situacijama. Za očekivati je da će d_+ imati neznatno negativnu pristranost za $\delta > 0$, budući da su kolebanja u metodama procjene negativno korelirana sa kolebanjima u procjenama težine.

Preciznost aproksimacije velikog uzorka za distribucije od d_+ i $d_+^{(1)}$ je istraživana od strane Monte Carlo metoda za dva i pet nezavisnih istraživanja. Rezultati za $k = 2$ i 5, neke kombinacije od veličine uzorka $n^E = n^C = 10,20,30,40$ i 50, i veličine učinka $\delta = 0.25,0.50,1.00,$ i 1.50 su dani u Tablicama 7. i 8.

Tablica 7. Mali uzorak preciznosti intervala pouzdanosti za δ temeljen na normalnoj aproksimaciji distribucije od d_+ sa 2000 replikacija, [HEDG1985]

Sample sizes	δ	Mean of d_+	Variance of d_+	Proportion of confidence intervals containing δ with nominal significance level 0.60	0.70	0.80	0.90	0.95	0.99
				$k = 2$					
(10, 10)	0.25	0.243	0.096	0.620	0.721	0.821	0.920	0.964	0.994
(10, 20)	0.25	0.252	0.064	0.628	0.731	0.824	0.915	0.959	0.992
(10, 50)	0.25	0.245	0.033	0.616	0.719	0.819	0.909	0.949	0.991
(50, 50)	0.25	0.250	0.020	0.596	0.705	0.802	0.910	0.957	0.993
(10, 10)	0.50	0.485	0.097	0.620	0.731	0.826	0.912	0.959	0.992
(10, 20)	0.50	0.484	0.067	0.611	0.712	0.816	0.903	0.960	0.992
(10, 50)	0.50	0.487	0.035	0.589	0.684	0.791	0.896	0.948	0.990
(50, 50)	0.50	0.490	0.021	0.601	0.710	0.811	0.901	0.945	0.990
(10, 10)	1.00	0.959	0.111	0.613	0.709	0.808	0.908	0.946	0.992
(10, 20)	1.00	0.976	0.072	0.625	0.719	0.808	0.905	0.951	0.990
(10, 50)	1.00	0.985	0.038	0.598	0.693	0.790	0.904	0.951	0.991
(50, 50)	1.00	0.991	0.022	0.610	0.712	0.810	0.903	0.954	0.993
(10, 10)	1.50	1.458	0.122	0.591	0.706	0.814	0.917	0.959	0.992
(10, 20)	1.50	1.474	0.087	0.595	0.690	0.803	0.909	0.951	0.990
(10, 50)	1.50	1.487	0.042	0.595	0.694	0.804	0.902	0.950	0.992
(50, 50)	1.50	1.493	0.025	0.587	0.694	0.791	0.904	0.953	0.992
				$k = 5$					
(10, 10, 10, 10, 10)	0.25	0.242	0.037	0.625	0.718	0.821	0.922	0.964	0.993
(10, 10, 10, 50, 50)	0.25	0.247	0.015	0.621	0.715	0.818	0.904	0.954	0.991
(20, 20, 20, 20, 20)	0.25	0.250	0.021	0.601	0.702	0.798	0.902	0.946	0.987
(50, 50, 50, 50, 50)	0.25	0.249	0.008	0.607	0.704	0.807	0.908	0.959	0.990
(10, 10, 10, 10, 10)	0.50	0.484	0.039	0.625	0.718	0.821	0.914	0.955	0.991
(10, 10, 10, 50, 50)	0.50	0.487	0.015	0.597	0.701	0.809	0.905	0.951	0.990
(20, 20, 20, 20, 20)	0.50	0.487	0.020	0.606	0.698	0.800	0.905	0.960	0.991
(50, 50, 50, 50, 50)	0.50	0.493	0.008	0.613	0.721	0.819	0.904	0.945	0.988
(10, 10, 10, 10, 10)	1.00	0.953	0.042	0.605	0.714	0.813	0.905	0.951	0.989
(10, 10, 10, 50, 50)	1.00	0.980	0.017	0.612	0.708	0.798	0.897	0.947	0.991
(20, 20, 20, 20, 20)	1.00	0.975	0.021	0.611	0.711	0.807	0.913	0.956	0.989
(50, 50, 50, 50, 50)	1.00	0.990	0.009	0.598	0.700	0.792	0.901	0.952	0.989
(10, 10, 10, 10, 10)	1.50	1.438	0.050	0.585	0.692	0.792	0.895	0.941	0.984
(10, 10, 10, 50, 50)	1.50	1.476	0.019	0.586	0.685	0.787	0.896	0.953	0.992
(20, 20, 20, 20, 20)	1.50	1.472	0.025	0.605	0.695	0.796	0.898	0.946	0.991
(50, 50, 50, 50, 50)	1.50	1.489	0.010	0.602	0.706	0.803	0.905	0.953	0.990

Tablica 8. Mali uzorak preciznosti intervala pouzdanosti za δ temeljen na normalnoj aproksimaciji distribucije od $d_{+}^{(1)}$, [HEDG1985]

Sample sizes	δ	Mean of $d_{+}^{(1)}$	Variance of $d_{+}^{(1)}$	Proportion of confidence intervals containing δ with nominal significance level 0.60	0.70	0.80	0.90	0.95	0.99
				$k = 2$					
(10, 10)	0.25	0.250	0.102	0.605	0.705	0.803	0.912	0.957	0.992
(10, 20)	0.25	0.256	0.066	0.618	0.716	0.815	0.910	0.952	0.992
(10, 50)	0.25	0.247	0.033	0.610	0.714	0.814	0.906	0.947	0.988
(50, 50)	0.25	0.251	0.020	0.595	0.701	0.798	0.909	0.956	0.992
(10, 10)	0.50	0.498	0.103	0.613	0.716	0.811	0.902	0.954	0.989
(10, 20)	0.50	0.491	0.069	0.604	0.704	0.808	0.897	0.956	0.992
(10, 50)	0.50	0.490	0.036	0.586	0.678	0.783	0.897	0.947	0.988
(50, 50)	0.50	0.493	0.021	0.601	0.708	0.806	0.902	0.944	0.988
(10, 10)	1.00	0.986	0.119	0.604	0.698	0.791	0.898	0.939	0.989
(10, 20)	1.00	0.992	0.075	0.620	0.713	0.803	0.894	0.948	0.990
(10, 50)	1.00	0.994	0.039	0.593	0.685	0.788	0.903	0.949	0.989
(50, 50)	1.00	0.996	0.022	0.604	0.708	0.811	0.903	0.954	0.992
(10, 10)	1.50	1.499	0.133	0.590	0.688	0.795	0.904	0.952	0.990
(10, 20)	1.50	1.499	0.091	0.588	0.688	0.794	0.901	0.943	0.988
(10, 50)	1.50	1.499	0.044	0.590	0.691	0.796	0.898	0.947	0.989
(50, 50)	1.50	1.501	0.026	0.591	0.690	0.791	0.900	0.954	0.991
				$k = 5$					
(10, 10, 10, 10, 10)	0.25	0.252	0.041	0.600	0.689	0.799	0.909	0.957	0.990
(10, 10, 10, 50, 50)	0.25	0.251	0.015	0.609	0.710	0.812	0.900	0.947	0.989
(20, 20, 20, 20, 20)	0.25	0.255	0.022	0.595	0.694	0.785	0.894	0.940	0.984
(50, 50, 50, 50, 50)	0.25	0.250	0.008	0.606	0.697	0.802	0.903	0.956	0.989
(10, 10, 10, 10, 10)	0.50	0.504	0.043	0.605	0.697	0.794	0.896	0.942	0.986
(10, 10, 10, 50, 50)	0.50	0.494	0.016	0.591	0.693	0.805	0.903	0.947	0.987
(20, 20, 20, 20, 20)	0.50	0.497	0.021	0.591	0.686	0.793	0.897	0.951	0.989
(50, 50, 50, 50, 50)	0.50	0.497	0.008	0.603	0.710	0.813	0.903	0.946	0.988
(10, 10, 10, 10, 10)	1.00	0.994	0.047	0.596	0.696	0.794	0.898	0.944	0.987
(10, 10, 10, 50, 50)	1.00	0.995	0.018	0.611	0.707	0.791	0.893	0.947	0.993
(20, 20, 20, 20, 20)	1.00	0.994	0.022	0.614	0.709	0.806	0.906	0.954	0.989
(50, 50, 50, 50, 50)	1.00	0.997	0.009	0.595	0.693	0.790	0.899	0.954	0.988
(10, 10, 10, 10, 10)	1.50	1.499	0.058	0.582	0.669	0.773	0.878	0.938	0.982
(10, 10, 10, 50, 50)	1.50	1.500	0.020	0.579	0.681	0.793	0.897	0.947	0.991
(20, 20, 20, 20, 20)	1.50	1.503	0.027	0.585	0.697	0.797	0.894	0.948	0.987
(50, 50, 50, 50, 50)	1.50	1.501	0.010	0.610	0.710	0.800	0.901	0.952	0.992

Ove veličine uzoraka i veličine učinka su izabrane jer su to tipične vrijednosti koje susrećemo u većini meta-analiza. Dvije tisuće replikacija za svaku strukturu veličine uzorka su generirane za svaku populacijsku veličinu učinka. Aproksimacija velikog uzorka je upotrebljena za izračunavanje intervala pouzdanosti od za δ. Srednje vrijednosti i varijance procjena, kao i empirijske proporcije intervala pouzdanosti, koje sadrže δ su izračunate za svaku veličinu uzorka i za svaku veličinu učinka. Ovi rezultati dokazuju da je aproksimacija velikih uzoraka za distribuciju metoda procjena razumno precizna za područje gdje je δ ispitivana, čak i kada sve studije imaju veličinu uzorka 10 po grupi. Preciznost aproksimacija teži poboljšanju kako se veličina uzorka povećava. Kao što je za očekivati, metoda procjene d_{+} ima neznatno negativnu pristranost, koja teži prema niskoj procjeni δ.

3.3. Procjene maksimalnih vjerojatnosti veličine učinka iz niza eksperimenata

Metoda maksimalnih vjerojatnosti je opća metoda za dobivanje procjena parametara koji daju specifični statistički model. Ima teorijsku prednost kod iskorištavanja procjena, koje se približavaju pravoj vrijednosti parametra kako se veličina uzorka povećava. Niti jedna druga metoda procjene nije preciznija kod velikih uzoraka kao što je metoda procjene maksimalnih vjerojatnosti.

Ako k eksperimenata dijeli zajedničku veličinu učinka δ, tada je, $\delta_1 = \delta_2 = ... = \delta_k = \delta^*$. Tada se može dobiti procjena maksimalne vjerojatnosti od δ^*. Metoda procjene maksimalnih vjerojatnosti $\hat{\delta}$ temeljena na promatranim veličinama učinka $g_1,\ldots,g_k$ je rješenje jednadžbe

$$A\hat{\delta} + B_1\sqrt{\hat{\delta}^2 + c_1} + \ldots + B_k\sqrt{\hat{\delta}^2 + c_k} = 0 \tag{58}$$

gdje je

$$A = \tilde{n}_1(2 - L_1) + \ldots + \tilde{n}_k(2 - L_k), \qquad B_i = (sign g_i)\tilde{n}_i L_i, \qquad \tilde{n}_i = n_i^E n_i^C / N_i, \qquad N_i = n_i^E + n_i^C,$$

$$L_i = \tilde{n}_i g_i^2 / (\tilde{n}_i g_i^2 + N_i - 2), \quad c_i = 4N_i / \tilde{n}_i L_i, \qquad i = 1,\ldots,k.$$

Kako su izrazi $A, B_1,\ldots,B_k, c_1,\ldots c_k$ određeni sa $g_1,\ldots,g_k$, jednadžba (58) može biti riješena za $\hat{\delta}$, procjenjene maksimalne vjerojatnosti od δ^*. Općenito, kada je $k > 2$, eksplicitna formula za $\hat{\delta}$ se ne može dobiti. Ali numerička rješenja od (58) mogu biti realizirana na računalu bez nekih velikih poteškoća. Svojstva velikog uzorka od $\hat{\delta}$ i d_+ su jednaka, pa tako $\hat{\delta}$ ima približno normalnu distribuciju na velikim uzorcima sa srednjom vrijednošću od δ i varijancom

$$\hat{\sigma}^2(\hat{\delta}) = \left(\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)}\right)^{-1} \tag{59}$$

gdje je $\hat{\sigma}^2(d_i)$ dana formulom (21). Mnogo je teže izračunati $\hat{\delta}$ nego d_+, zato je malo razloga za preporučiti korištenje procjenjene maksimalne vjerojatnosti.

3.4. Procjene veličine učinka temeljene na transformiranim procjenama

Jedan od problema kod kombiniranja procjena veličine učinka iz k studija je taj da optimalna linearna kombinacija procjena ovisi o varijancama od $d_1,\ldots,d_k$, koje ovise o nepoznanici δ. Dosada smo ovaj problem rješavali upotrebom procjena od δ da bi procijenili optimalne težine za kombiniranje $d_1,\ldots,d_k$. Druga mogućnost je transformiranje procjena koja rezultira transformiranom slučajnom varijablom koja ima varijance neovisne o δ. Transformirane metode procjene se mogu kombinirati na jednostavan način tako da stvaraju zajedničku procjenu u transformiranoj metrici. Invertiranjem transformacije, zajednička procjena rezultira procjenom od δ^* (Vidi 2.1.4.). Primijetimo da općenito transformacija dana u 2.1.4. ovisi izričito o omjeru n^E / n^C. Kako bi upotrebljavali istu transformaciju za svako istraživanje, omjer n_i^E / n_i^C mora biti jednak za svako istraživanje.

Ako je svaki eksperiment uravnotežen u smislu da su eksperimentalna i kontrolna grupa unutar istraživanja jednake, $n_j^E = n_j^C = n_j$, $j = 1,\ldots,k$, tada je transformacija stabiliziranja varijance (eng. variance-stabilizing transformation) za d dana sa

$$h(d) = \sqrt{2}\sinh^{-1}\left(d / 2\sqrt{2}\right) \tag{60}$$

Neka su $h_1 = h(d_1),\ldots,h_k = h(d_k)$ transformirane procjene i $\eta = h(\delta)$ parametar transformirane veličine učinka, te pretpostavimo da je jednak za sva istraživanja. Tada svaka od transformiranih procjena h_i ima aproksimativno normalnu distribuciju sa srednjom vrijednošću od η i varijancu od $1/(2n_i)$. Stoga je težinska linearna kombinacija procjena od η s najmanjom varijancom

$$h_+ = 2\sum_{i=1}^{k}\frac{n_i h_i}{N} \tag{61}$$

gdje je $N = 2\sum n_i$ ukupna veličina uzorka. Procjena od η je transformirana u procjenu od δ^* pomoću inverzne transformacije $\hat{\delta}^* = 2\sqrt{2}\sinh\left(h_+ / \sqrt{2}\right)$. Štoviše, h_+ će i sama biti normalno distribuirana sa varijancom koja otprilike iznosi

$$\sigma_\infty^2(h_+) = 1/N \tag{62}$$

Prema tome, $100(1-\alpha)$-postotni interval pouzdanosti (η_L, η_U) za η je

$$\eta_L = h_+ - C_{\alpha/2}/\sqrt{N}, \qquad \eta_U = h_+ + C_{\alpha/2}/\sqrt{N}, \tag{63}$$

gdje je $C_{\alpha/2}$ dvosmjerna kritična vrijednost za standardnu normalnu distribuciju. Interval pouzdanosti je invertiran da bi dobili interval pouzdanosti (δ_L, δ_U) za δ iz

$$\delta_L = 2\sqrt{2}\sinh\left(\eta_L / \sqrt{2}\right), \qquad \delta_U = 2\sqrt{2}\sinh\left(\eta_U / \sqrt{2}\right). \tag{64}$$

3.5. Ispitivanje homogenosti veličine učinka

Prije udruživanja procjena veličine učinka iz niza od k istraživanja, važno je utvrditi da li ta istraživanja mogu dijeliti zajedničku veličinu učinka. Statističko testiranje homogenosti veličine učinka je u biti test hipoteze $H_0: \ \delta_1 = \delta_2 = \ldots = \delta_k$ protiv alternative gdje se barem jedna veličina učinka δ_i razlikuje od ostalih. Nije poznato kako se testira ova hipoteza na malom uzorku. Jedan od načina testiranja ove hipoteze na velikom uzorku je sljedeći:

$$Q = \sum_{i=1}^{k} \frac{\left(d_i - d_+\right)^2}{\hat{\sigma}^2\left(d_i\right)} \tag{65}$$

gdje je d_+ metoda procjene težine veličine učinka dana u (54), i $\hat{\sigma}^2\left(d_i\right)$ je dana sa formulom (21). Statistički test Q je suma kvadrata d_i težinske srednje vrijednosti d_+, gdje je i-ti kvadrat težina recipročne vrijednosti procijenjene varijance od d_i.

Tablica 9. Statistički test homogenosti *Q* temeljen na simulaciji sa 2000 replikacija, [HEDG1985]

Sample sizes	δ	Mean of Q	Variance of Q	Proportion of test statistics exceeding the nominal significance value					
				0.40	0.30	0.20	0.10	0.05	0.01
$k = 2$									
(10, 10)	0.25	0.988	2.241	0.380	0.286	0.195	0.100	0.052	0.011
(10, 20)	0.25	0.992	2.154	0.402	0.290	0.194	0.090	0.044	0.011
(10, 50)	0.25	0.980	2.086	0.389	0.284	0.190	0.102	0.051	0.009
(50, 50)	0.25	0.951	1.937	0.383	0.274	0.184	0.089	0.048	0.012
(10, 10)	0.50	0.954	1.825	0.386	0.289	0.186	0.092	0.045	0.008
(10, 20)	0.50	0.907	1.756	0.366	0.262	0.178	0.083	0.041	0.007
(10, 50)	0.50	0.915	1.742	0.369	0.271	0.179	0.094	0.045	0.008
(50, 50)	0.50	1.043	2.130	0.412	0.305	0.199	0.114	0.058	0.012
(10, 10)	1.00	1.006	2.205	0.393	0.302	0.199	0.098	0.049	0.013
(10, 20)	1.00	0.953	1.730	0.389	0.291	0.192	0.090	0.042	0.008
(10, 50)	1.00	1.015	1.956	0.418	0.313	0.201	0.093	0.051	0.011
(50, 50)	1.00	0.990	2.056	0.390	0.288	0.194	0.099	0.055	0.007
(10, 10)	1.50	1.034	2.107	0.403	0.311	0.208	0.104	0.056	0.010
(10, 20)	1.50	0.992	1.944	0.394	0.304	0.202	0.095	0.046	0.010
(10, 50)	1.50	0.992	1.841	0.400	0.301	0.206	0.098	0.043	0.006
(50, 50)	1.50	1.059	2.383	0.402	0.295	0.211	0.109	0.058	0.011
$k = 5$									
(10, 10, 10, 10, 10)	0.25	3.828	7.327	0.378	0.282	0.184	0.085	0.040	0.009
(10, 10, 10, 50, 50)	0.25	3.788	7.602	0.377	0.274	0.179	0.079	0.042	0.010
(20, 20, 20, 20, 20)	0.25	3.896	8.199	0.392	0.285	0.184	0.091	0.048	0.011
(50, 50, 50, 50, 50)	0.25	3.858	7.297	0.397	0.298	0.191	0.092	0.035	0.005
(10, 10, 10, 10, 10)	0.50	3.846	7.490	0.377	0.287	0.177	0.090	0.048	0.008
(10, 10, 10, 50, 50)	0.50	3.900	7.883	0.389	0.290	0.189	0.095	0.045	0.009
(20, 20, 20, 20, 20)	0.50	3.900	7.320	0.395	0.288	0.190	0.093	0.048	0.007
(50, 50, 50, 50, 50)	0.50	3.992	8.078	0.394	0.299	0.198	0.097	0.045	0.012
(10, 10, 10, 10, 10)	1.00	3.933	8.035	0.397	0.293	0.192	0.087	0.044	0.013
(10, 10, 10, 50, 50)	1.00	3.950	7.807	0.394	0.295	0.187	0.092	0.047	0.011
(20, 20, 20, 20, 20)	1.00	3.821	7.489	0.361	0.268	0.183	0.091	0.050	0.007
(50, 50, 50, 50, 50)	1.00	3.926	7.729	0.395	0.289	0.195	0.086	0.043	0.010
(10, 10, 10, 10, 10)	1.50	3.932	7.292	0.396	0.300	0.205	0.093	0.041	0.009
(10, 10, 10, 50, 50)	1.50	4.068	8.362	0.412	0.308	0.211	0.106	0.057	0.010
(20, 20, 20, 20, 20)	1.50	4.019	8.024	0.400	0.304	0.195	0.099	0.054	0.009
(50, 50, 50, 50, 50)	1.50	4.121	8.784	0.412	0.316	0.222	0.112	0.054	0.013

Kada svako istraživanje ima veliki uzorak, asimptotska distribucija od Q se može upotrijebiti kao osnova za aproksimativni test homogenosti od δ_i. Ukoliko svih k istraživanja imaju jednaku veličinu učinka populacije (tj. ako je H_0 istina), tada statistički test Q ima asimptotsku χ^2 (hi-kvadrat) distribuciju sa $k-1$ stupnjeva slobode. Dakle, ako dobivena vrijednost od Q premašuje $100(1-\alpha)$-postotnu kritičnu vrijednost od hi-kvadrat distribucije sa $k-1$ stupnjeva slobode, odbacujemo hipotezu da su δ_i jednaki. Ako odbacimo ovu nul hipotezu, možemo se odlučiti da ne udružujemo procjene od δ, jer ne procjenjuju isti parametar.

Test homogenosti veličine učinka na malom uzorku ovisi o preciznosti statističke asimptotske distribucije u konačnim uzorcima. Rezultati za k = 2 i 5, i za kombinacije veličine uzoraka $n^E = n^C$ = 10,20,30,40 i 50, su dani u Tablici 9. Dvije tisuće replikacija su generirane za svaku veličinu uzorka i za svaku veličinu učinka δ = 0.25,0.50,1.00, i 1.50. Statistička homogenost Q je izračunata za svaku replikaciju i proporciju od statističkog testa koje premašuju različite kritične vrijednosti.

3.5.1. Drugi načini za testiranje homogenosti veličine učinka

Alternativa statističkom testu homogenosti Q je test omjera vjerojatnosti (eng. likelihood ratio test). On uključuje prilično teško izračunavanje procjene maksimalnih vjerojatnosti od δ iz k eksperimenata. Zbog ovih poteškoća, preporuča se korištenje Q-statistike, koja je lakša za izračunavanje i intuitivno mnogo privlačnija nego test omjera vjerojatnosti.

Ipak, postoje i druge metode. Sljedeći test za homogenost veličine učinka koristi činjenicu da je jednostrana p-vrijednost funkcija veličine uzorka i veličine učinka. Ako je p_i jednostrana p-vrijednost povezana sa t-testom za razlike između srednjih vrijednosti od i-tog eksperimenta, i ako su veličine uzorka n_i^E i n_i^C relativno velike, tada $z_i = \Phi^{-1}(p_i)$ ima približno normalnu distribuciju jedinice sa srednjom vrijednošću

$$\lambda_i = \delta_i \sqrt{n_i^E n_i^C / \left(n_i^E + n_i^C\right)} \tag{66}$$

Ako su veličine uzorka eksperimentalne i kontrolne grupe u k istraživanja jednake, tj. $n_1^E = \ldots = n_k^E$ i $n_1^C = \ldots = n_k^C$ uvjet $\lambda_1 = \ldots = \lambda_k$ je ekvivalentan sa $\delta_1 = \ldots = \delta_k$, tako da možemo testirati homogenost od λ_i, umjesto od δ_i. Kako bi testirali hipoteze $\lambda_1 = \ldots = \lambda_k$, izračunamo statistički podatak

$$Q_3 = \sum_{i=1}^{k} (z_i - \bar{z})^2 \tag{67}$$

gdje je $\bar{z}$ netežinska srednja vrijednost (eng. unweighted mean) od $z_1, \ldots, z_k$. Kada je hipoteza homogenosti istinita, Q_3 ima hi-kvadrat distribuciju sa $k-1$ stupnjeva slobode, i odbacujemo hipotezu za velike vrijednosti od Q_3.

Kada je veličina uzoraka nejednaka ova metoda i dalje potiče test homogenosti od λ_i, ali sada jednakosti od ovih λ ne implicira jednakost od δ. Stoga, test temeljen na Q_3 bi trebao biti korišten samo u slučaju ako postoji zajednička veličina uzorka eksperimentalne grupe i zajednička veličine uzorka kontrolne grupe. Razne analize su pokazale, da istraživanja općenito nemaju zajedničku veličinu uzorka, pa tako ovaj test kojeg smo opisali ima ograničenu primjenu.

3.6. Izračun statistike testa homogenosti

Formule za izračun testa se oslanjaju na jednakost

$$\sum a_i (x_i - x_+)^2 = \sum a_i x_i^2 - \left(\sum a_i x_i\right)^2 / \sum a_i \tag{68}$$

gdje je $x_+ = \sum a_i x_i / \sum a_j$. Primjenjujući ovu jednakost sa $a_i = 1/\hat{\sigma}^2(d_i)$, $x_i = d_i$ i $x_+ = d_+$ dobivamo formulu za izračun homogenosti statističkih vrijednosti Q,

$$Q = \sum_{i=1}^{k} \frac{d_i^2}{\hat{\sigma}^2(d_i)} - \left(\sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2(d_i)} \right)^2 / \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)} \tag{69}$$

Prednost ove formule je da Q može biti izračunat iz suma od triju varijabli: $1/\hat{\sigma}^2(d_i)$, $d_i / \hat{\sigma}^2(d_i)$ i $d_i^2 / \hat{\sigma}^2(d_i)$.

Srednja težinska vrijednost d_+ i njena varijanca $\hat{\sigma}^2(d_+)$ su također izračunate iz sume dviju ovih varijabli: $1/\hat{\sigma}^2(d)$ i $d/\hat{\sigma}^2(d)$. Dakle, direktna kalkulacija ovih statističkih vrijednosti je dobivena iz vrijednosti triju varijabli za svako istraživanje.

3.7. Procjena veličine učinka za male uzorke

Rezultati simulacijski istraživanja pokazuju da je teorija velikih uzoraka razumno precizna kada su veličine učinka manje od 1.5 u apsolutnoj jačini i veličine uzorka u svakoj grupi nisu manje od 10. Teorija velikih uzoraka nije razmatrana za uzorke manje od 10, te se pretpostavlja da nije vrlo precizna. Jedna situacija koja se pojavila u praksi je ta da podaci koji su dostupni istraživačima sadrže *veliki broj* istraživanja, gdje je svako istraživanje sa malim uzorkom. U ovom slučaju korisno je razmotriti različite verzije teorije velikih uzoraka - onih u kojima broj istraživanja k postaje velik. Veliki uzorak normalne aproksimacije u ovoj teoriji je primjenjiv kad god je broj istraživanja k velik. Primijetimo da dvije različite vrste asimptotske teorije ne daju jednake rezultate, tj. oni su veoma blizu kada su i broj istraživanja i veličine uzorka velike.

Kada su veličine uzorka n_i^E i n_i^C od i-tog istraživanja mali, pristranost u metodi procjene $\hat{\delta}_i$ i g_i nije zanemariva. Npr., ako je $n_i^E = n_i^C = 3$, pristranost od g_i je preko 25 posto. Nije nužno da ova pristranost nestane kada je izračunat prosjek velikog broja metoda procjene, jer pristranosti sve mogu biti u istom smjeru. Ako je $n_i^E = n_i^C = 3$ za niz od k istraživanja, srednja vrijednost $\bar{g}$ od k metoda procjena će imati pristranost od 25 posto bez obzira o veličini k. Iako varijanca od $\bar{g}$ teži prema nuli kako se k povećava, srednja vrijednost $\bar{g}$ teži prema pogrešnoj vrijednosti, $1.25\delta^*$, umjesto od δ^* kako $k \to \infty$. Ovo ističe potrebu za upotrebom nepristrane metode procjene d_i kada istraživanja imaju jako male uzorke.

Jedna od najjednostavnijih metoda za procjenjivanje veličine učinka iz niza istraživanja s malim uzorcima je korištenje težinske srednje vrijednosti. Težinska srednja vrijednost sa najmanjom varijancom je dana sa

$$\bar{d}_W = w_1 d_1 + \ldots + w_k d_k \tag{70}$$

gdje su optimalne težine w_i procijenjene sa

$$w_i = \frac{1}{v_i(\bar{d})} / \sum_{j=1}^{k} \frac{1}{v_j(\bar{d})} \tag{71}$$

$\bar{d}$ je netežinska srednja vrijednost od $d_1, \ldots, d_k$, i

$$v_i(\bar{d}) = a_i + b_i \bar{d}^2 \tag{72}$$

gdje je

$$\begin{aligned} a_i &= (N_i - 2)[J(N_i - 2)]^2 / [\tilde{n}_i (N_i - 4)], \\ b_i &= \{(N_i - 2)[J(N_i - 2)]^2 - (N_i - 4)\} / (N_i - 4), \end{aligned} \tag{73}$$

i $J(m)$ je dan u Tablici 2.

Primijetimo da je $v_i(\delta)$ točno varijanca od d_i i da su težine $w_1, \ldots, w_k$ dobivene iz $v_i(\bar{d})$ izračunate korištenjem $\bar{d}$ za procjenu δ.

Analiza distribucije od $\bar{d}$ pokazuje da se približava normalnoj kako se broj studija povećava, sa srednjom vrijednošću od δ^* i procijenjenom varijancom

$$v = \left(k \sum_{i=1} \frac{1}{v_i(\bar{d})} \right)^{-1} \tag{74}$$

Dakle, možemo dobiti $100(1-\alpha)$-postotni interval pouzdanosti (δ_L, δ_U) za δ

$$\delta_L = \bar{d}_W - C_{\alpha/2}\sqrt{v}, \qquad \delta_U = \bar{d}_W + C_{\alpha/2}\sqrt{v}, \tag{75}$$

gdje je $C_{\alpha/2}$ dvosmjerna kritična vrijednost standardne normalne distribucije.

3.8. Utjecaj pogrešaka u mjerenju i nevaljanosti

Pretpostavka da su $\delta_1, \ldots, \delta_k$ identični je izvedena iz namjere da eksperimentalni faktor proizvodi jednaki učinak u svakom istraživanju. Ako su mjerni uređaji (testovi) upotrebljeni u istraživanjima identični, tada se izvorni (neobrađeni) podatci iz istraživanja mogu udružiti direktno, te analizirati koristeći standardne metode. Jako malo grupa istraživanja koriste točno isti mjerni uređaj u svakom istraživanju. Mnogo realnija je pretpostavka da istraživanja koriste različite, ali linearno izjednačene testove. Ako različita istraživanja imaju jednak učinak, tada će veličina učinka biti jednaka kada je populacija rezultata reprezentirana na različitim, ali linearno izjednačenim testovima. Stoga, ako različita istraživanja imaju isti učinak, i koriste različite (ali linearno izjednačene) testove, veličina učinka populacije će biti jednaka.

Tvrdnja o savršenoj linearnoj izjednačenosti između testova koji mjere istu karakteristiku ne mora uvijek biti održiva. Dva su faktora koja mogu dovesti do kršenja tvrdnje:
(i) pogreška mjerenja
(ii) nevaljanost

Standardna razlika aritmetičkih sredina δ definirana u 2.1. je mjera jačine učinka eksperimentalnog faktora uspoređena sa varijabilnošću između dvije grupe u eksperimentu. Implicitna tvrdnja je da se varijabilnost unutar eksperimentalne i kontrolne grupe temelji na stalnim razlikama između ispitanika. Ako mjera odziva nije savršeno pouzdana, tj. ako postoje pogreške mjerenja (eng. error of measurement), tada mjerna pogreška također doprinosi varijabilnosti unutar grupa. Mjerna pogreška zbog toga mijenja populacijsku vrijednost standardne razlike aritmetičkih sredina. Ukoliko nam je cilj procjena vrijednosti δ na temelju standardne razlike aritmetičkih sredina kada pogreške mjerenja nisu prisutne, tada su potrebne neke procedure za ispravljanje mjernih pogrešaka.

Strukturni modeli (48) i (76) ne dopuštaju mogućnost da neke mjere odziva imaju jedinstvene faktore). Npr., neki eksperimenti koriste skupocjene standardizirane testove za mjerenje postignuća u čitanju, dok ostale studije koriste lokalno razvijene testove koji su u korelaciji sa standardiziranim testovima. Ukoliko lokalno razvijeni testovi imaju jedinstvene faktore, oni neće biti potpuno valjana mjerenja za postignuća u čitanju, kao mjerenja dobivena standardiziranim testom

4. Procjene veličine učinka kada nisu promatrani svi rezultati istraživanja

Često postoji razlog za vjerovati da nereprezentativni uzorci prevladavaju u kvantitativnom istraživanju. Pristranost vezana uz uzorak (eng. sampling bias) proizlazi iz "predrasuda protiv nul hipoteza" u smislu da samo statistički značajna istraživanja objavljena, i procjena veličina učinka koja se podudara sa statistički neznačajnim razlikom aritmetičkih sredina može biti nedostupna za uključivanje u meta-analizu. Ovdje se raspravlja o problemu koji je stvoren cenzuriranjem procjena veličine učinka koje odgovaraju statistički neznačajnim rezultatima. Prikazuju se neki dokazi za postojanje pristranosti vezane uz uzorak, statistički model za istraživanje učinaka od pristranosti vezane uz uzorak, te posljedice takvih pristranosti. Nadalje se prikazuje procjena maksimalnih vjerojatnosti veličine učinka na temelju modela pristranosti vezane uz uzorak, kombinacija procjena iz nekoliko eksperimenata na temelju modela pristranosti vezane uz uzorak, te razmatranja načina u kojima dobiveni rezultati mogu biti korišteni za izvlačenje zaključaka kada je moguće da postoji pristranost vezana uz uzorak.

4.1. Pristranost vezana za uzorak

Izračun procjene veličine učinka zahtjeva srednjih vrijednosti i standardnih devijacija uzorka ili vrijednost *t*- ili *F*- testa za razliku između srednjih vrijednosti. Autori ponekad ne izvijeste je li statističke testa ili rezultate deskriptivne statistike, ili jednostavno navedu da "nema statistički značajne razlike". Jedna od posljedica ovakve prakse je da su statistički neznačajni rezultati koji trebaju za izračun procjena veličine učinka, ponekad selektivno izostavljeni u izvještajima o istraživanju.

Drugi potencijalni izvor pristranosti u procjenama veličine učinka proizlazi iz tendencije urednika časopisa da ne objave istraživanja koja nisu donijela statistički značajne rezultate. Sterling (1959) je raspravljao o tome kako su neznačajni rezultati rijetko objavljivani, te da je objavljena literatura puna pogrešaka tipa I.

4.2. Posljedice promatranja samo značajnih veličina učinka

Ako za procjenu veličine učinka nije moguće uzeti i istraživanja sa statistički neznačajnim rezultatima, tada će uzorak procjene veličine učinka biti pristran. Kako rezultati *t*- testa povezani s veličinom učinka, istraživanja koja daju statistički značajne rezultate, imaju veličine učinka koje su veće u apsolutnoj jačini.

Ekstremni oblik pristranosti prema značajnim rezultatima su istražili Lane i Dunlap, [LANE1978] i Hedges, [HEDG1984] u slučaju kada su promatrana samo istraživanja koja donose statistički značajne rezultate. Lane i Dunlap su simulirali rezultate velikog broja eksperimenata s eksperimentalnom i kontrolnom grupom, te su odabirali samo eksperimente koji donose statistički značajne srednje vrijednosti, [LANE1978]. Kao što je bilo za očekivati, otkrili su da srednja vrijednost procijenjena iz eksperimenata donosi značajne rezultate precjenjuje srednju vrijednost populacije.

Model koji je ovdje predstavljen uključuje precizno cenzuriranje svih rezultata koji nemaju značajnost za α =0.05. Pretpostavimo da procjena veličine učinka proizlazi iz povezanosti sa eksperimentima koji imaju zajedničku veličinu učinka $n^E = n^C = n$. Označimo da je populacija i uzorak veličina učinka (standardizirana razlika aritmetičkih sredina) dana sa δ i g kao u poglavlju 2. Po potrebi su indeksi korišteni za razlikovanje parametara i procjena različitih istraživanja. Nadalje, tvrdimo da je g promatran samo ako je F-test za razliku srednjih vrijednosti statistički značajan na nekom unaprijed postavljenom nivou značajnosti (eng. significance level) α. Odnosno, uzorak veličine učinka je promatran samo ako je $g^2 > 2F(\alpha, n)/n$, gdje $F(\alpha, n)$ je $100(1-\alpha)$-postotna točka od *F*-distribucije sa 1 i $2n-2$ stupnjeva slobode. Ukoliko nije drugačije naznačeno, nivo značajnosti je α =0.05. Sa g^* označimo promatranu veličinu učinka. Primijetimo da u ovom modelu nema vrijednosti od g^* koje su manje od $\sqrt{2F(\alpha, n)/n}$, budući da odgovarajuće razlike srednjih vrijednosti nisu statistički značajne.

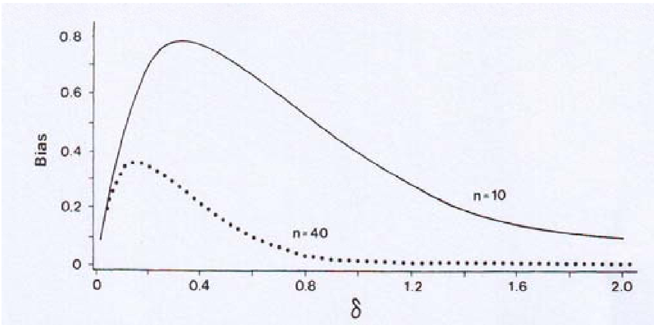
Statistički podatak g^* odgovara procjeni veličine učinka kada su samo statistički značajni rezultati promatrani. Želimo dobiti distribuciju od g^*, s ciljem istraživanja svojstava uobičajenih metoda procjene veličine učinka kada su ograničene samo na značajne rezultate. Primijetimo da g^* cenzurira sve vrijednosti unutar intervala $\left(-\sqrt{2F(\alpha, n)/n}, \sqrt{2F(\alpha, n)/n}\right)$.

Očekivana vrijednost i varijanca od g^* su dobivene pomoću numeričkih integracija, i dane su u Tablici 12. za $\delta = 0.25, 0.50, 0.75, 1.00$ i 1.50 i n =10,20,30,40, i 50. Kao što je za očekivati, pristranost od g^* (metoda procjene od δ) teži povećanju za mali n i male (ali ne nula) vrijednosti od δ. Pristranost brzo raste kako se δ povećava, tako da je pristranost manja od 10 posto za $\delta \geq 1.0$ i $n \geq 20$ po grupi. Za manji δ, pristranost može biti prilično velika. Za $\delta = 0.5$ i $n = 15$ pristranost u g^* je približno 100 posto, a za $\delta = 0.25$ pristranost od g^* je preko 200 posto čak i ako je $n = 40$ po grupi.

Apsolutna pristranost $\left|E(g^*) - \delta\right|$ teži nuli kako se δ približava nuli, i također teži nuli za veliki δ. Apsolutna pristranost ima maksimum između $\delta = 0.3$ (za $n = 10$) i $\delta = 0.2$ (za $n = 40$), te on je najveći za male, ali različite od nule, veličine učinka. Relativna pristranost $E(g^*)/\delta$ se povećava kako se δ približava nuli, ali teži jedinstvenosti kako se $|\delta|$ povećava. Pristranost od g^* kao funkcije od δ je prikazana na Slici 4.

Tablica 10. Očekivane vrijednosti $E(g^*)$ i varijance $V(g^*)$ od veličine učinka g^* kada su samo značajni rezultati promatrani, [HEDG1985]

		δ									
		0.25		0.50		0.75		1.00		1.50	
n		$E(g^*)$	$V(g^*)$	$E(g^*)$	$V(g^*)$	$E(g^*)$	$V(g^*)$	$E(g^*)$	$V(g^*)$	$E(g^*)$	$V(g^*)$
10	Exact	1.01	0.44	1.22	0.11	1.30	0.10	1.39	0.13	1.67	0.24
	Simulated	0.99	0.48	1.23	0.11	—	—	1.39	0.14	1.67	0.24
20	Exact	0.77	0.08	0.87	0.04	0.96	0.06	1.10	0.08	1.53	0.14
	Simulated	0.77	0.08	0.87	0.04	—	—	1.09	0.08	1.54	0.14
30	Exact	0.65	0.03	0.73	0.03	0.85	0.05	1.03	0.07	1.52	0.09
	Simulated	0.65	0.03	0.73	0.03	—	—	1.03	0.07	1.53	0.09
40	Exact	0.57	0.02	0.65	0.02	0.80	0.04	1.01	0.06	1.52	0.07
	Simulated	0.57	0.02	0.65	0.02	—	—	1.01	0.06	1.52	0.07
50	Exact	0.51	0.01	0.60	0.02	0.77	0.04	1.00	0.05	1.50	0.05
	Simulated	0.52	0.01	0.60	0.02	—	—	1.00	0.05	1.51	0.06



Slika 4. Pristranost promatrane veličine učinka g^* kao funkcije od populacijske veličine učinka δ, [HEDG1985]

Mnoge kvantitativne sinteze istraživanja rado uključuju istraživanja sa veličinom uzorka manjom od 40 po grupi i veličinom učinka manjom od 1.00. Velika pristranost od g^* za mali n i mali δ mogu dovesti do značajnih precjenjivanja veličine učinka, ako neznačajni rezultati nisu izviješteni.

4.3. Procjena veličine učinka jednog istraživanja na temelju samo značajnih rezultata

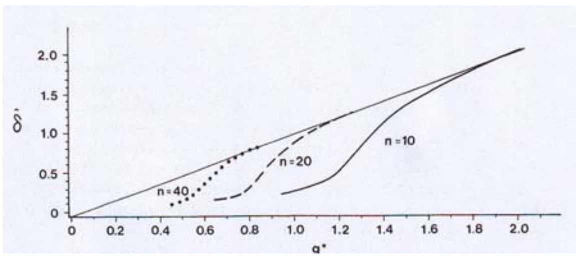
Rezultati koji su navedeni upućuju da upotreba od g^* (primjer procjene veličine učinka kada su samo statistički značajni rezultati promatrani) kao metode procjene od δ nas može dovesti do ozbiljne pristranosti. Kada su veličina uzorka i veličina učinka ograničeni, pristranost od g^* može biti toliko značajna da utječe na zaključke, te može dovesti do problema u analizi i interpretaciji temeljenoj na veličini učinka.

Nadalje razmatramo metode za procjenjivanje veličine učinka iz jednog istraživanja, gdje su promatrane samo statistički značajne razlike srednjih vrijednosti. Metoda uključuje procjenu maksimalnih vjerojatnosti od δ temeljenoj na distribuciji od g^*. Ovaj načina implicitno, kroz distribuciju od g^*, radi ispravke za cenzuriranje vrijednosti od g^* koje odgovaraju statistički neznačajnim srednjim vrijednostima. Prvo izvodimo metodu procjene maksimalnih vjerojatnosti od δ temeljenoj na g^* vrijednosti iz jednog eksperimenta. Nakon toga dobivamo njenu preciznu distribuciju numerički, i koristimo distribuciju za istraživanje pristranosti metode procjene.

4.3.1. Procjena veličine učinka

Metoda procjene maksimalnih vjerojatnosti $\hat{\delta}$ od veličine učinka δ je dobivena pomoću maksimiziranja vjerojatnosti od g^* kao funkcije od δ. Nakon toga pronalazimo istu promatranu vrijednost standardne razlike aritmetičkih sredina koja nas vodi do procjene manjih maksimalnih vjerojatnosti iz modela gdje su promatrane samo značajni rezultati u model gdje su svi rezultati promatrani. Kada razmotrimo sve rezultate, procjena maksimalnih vjerojatnosti od δ je približno jednaka promatranoj vrijednosti standardne razlike aritmetičkih sredina. Za očekivati je da procjenjena maksimalna vjerojatnost od δ bazirana na g^* bude manja nego g^*, posebno kada je g^* sam po sebi malen.

Slika 5. je grafički prikaz od $\hat{\delta}$ kao funkcije od g^* za n =10,20, i 40. U svakom se slučaju referentni pravac koji je odgovarajući za $g^* = \hat{\delta}$ također pojavljuje. Zbog simetrije su prikazane samo pozitivne vrijednosti od g^*. Odnosno, ako g^* odgovara procjeni $\hat{\delta}_0$ tada $-g^*$ odgovara procjeni $-\hat{\delta}_0$.



Slika 5. Metoda procjene maksimalnih vjerojatnosti $\hat{\delta}$ od veličine učinka kao funkcija promatrane veličine učinka g^*, [HEDG1985]

Male vrijednosti od g^* vode prema puno manjim vrijednostima od $\hat{\delta}$. Ovo je ilustrirano na slici 5., gdje funkcija koja povezuje g^* sa $\hat{\delta}$ leži ispod referentne linije (eng. reference line) za male g^*. Jedna interpretacija je da samo značajne vrijednosti od g^* teže da budu pridružene malim vrijednostima od δ, te je nakon više cenzururanih vrijednosti u ovom slučaju, metoda procjene maksimalnih vjerojatnosti manja.

Velike vrijednosti od g^* vode prema vrijednostima od $\hat{\delta}$ koje su gotovo identične g^*. Ovo je prikazano na slici 5., gdje funkcija koja povezuje g^* sa $\hat{\delta}$ ne odstupa u velikoj mjeri od referentne linije za velike vrijednosti od g^*. Interpretacija ovog otkrića je ta da velike vrijednosti od g^* proizlaze iz velikih vrijednosti od δ, i potom se cenzuriranje rijetko događa, tako da je g^* dobra procjena od δ.

Primijetimo da $\hat{\delta}$ opada kako g^* teži minimalnim vidljivim vrijednostima različitima od nule, ali minimum od $\hat{\delta}$ (za pozitivne g^*) nije nula. Za $n = 40$ ova minimalna vrijednost je $\hat{\delta}$=0.111, dok je za $n = 10$ minimalna vrijednost $\hat{\delta}$=0.242.

Tablica 11. Metoda procjene maksimalnih vjerojatnosti $\hat{\delta}$ kao funkcija veličine uzorka po grupi *n* i promatrane veličine učinka g^*, [HEDG1985]

	n																
	10	12	14	16	18	20	25	30	35	40	45	50	60	70	80	90	100
g_{Min}	0.940	0.847	0.777	0.722	0.677	0.640	0.569	0.517	0.477	0.445	0.419	0.397	0.362	0.334	0.312	0.294	0.279
δ_{Min}	0.242	0.216	0.197	0.183	0.171	0.161	0.142	0.129	0.120	0.111	0.108	0.101	0.092	0.085	0.079	0.074	0.070
g^*																	
0.30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.075	0.080
0.35	—	—	—	—	—	—	—	—	—	—	—	—	—	0.090	0.100	0.116	0.145
0.40	—	—	—	—	—	—	—	—	—	—	—	0.100	0.112	0.135	0.186	0.253	0.299
0.45	—	—	—	—	—	—	—	—	—	0.113	0.122	0.133	0.179	0.273	0.339	0.378	0.401
0.50	—	—	—	—	—	—	—	—	0.132	0.146	0.168	0.210	0.337	0.405	0.442	0.463	0.476
0.55	—	—	—	—	—	—	—	0.147	0.169	0.210	0.291	0.370	0.458	0.498	0.519	0.531	0.539
0.60	—	—	—	—	—	—	0.159	0.187	0.249	0.361	0.445	0.495	0.546	0.571	0.584	0.591	0.595
0.65	—	—	—	—	—	0.166	0.196	0.269	0.411	0.504	0.556	0.588	0.619	0.634	0.642	0.646	0.648
0.70	—	—	—	—	0.182	0.195	0.261	0.430	0.548	0.608	0.643	0.663	0.682	0.692	0.696	0.698	0.699
0.75	—	—	—	0.196	0.212	0.239	0.401	0.573	0.652	0.693	0.715	0.729	0.740	0.746	0.748	0.749	0.750
0.80	—	—	0.209	0.228	0.259	0.317	0.563	0.682	0.737	0.764	0.780	0.789	0.795	0.798	0.800	0.800	0.800
0.85	—	0.218	0.237	0.273	0.342	0.469	0.687	0.771	0.810	0.829	0.839	0.845	0.847	0.849	0.850	0.850	0.850
0.90	—	0.243	0.279	0.349	0.498	0.631	0.788	0.847	0.875	0.888	0.895	0.898	0.899	0.900	0.900	0.900	0.900
0.95	0.246	0.277	0.340	0.494	0.659	0.757	0.872	0.916	0.935	0.945	0.948	0.950	0.949	0.950	0.950	0.950	0.950
1.00	0.272	0.324	0.455	0.662	0.789	0.862	0.947	0.979	0.991	0.998	1.000	1.001	1.000	1.000	1.000	1.000	1.000
1.05	0.305	0.394	0.626	0.801	0.896	0.951	1.015	1.038	1.047	1.050	1.051	1.052	1.050	1.050	1.050	1.050	1.050
1.10	0.347	0.523	0.782	0.916	0.987	1.030	1.077	1.094	1.100	1.101	1.102	1.102	1.100	1.100	1.100	1.100	1.000
1.15	0.409	0.695	0.910	1.013	1.068	1.100	1.136	1.148	1.151	1.153	1.153	1.152	1.150	1.150	1.150	1.150	1.150
1.20	0.512	0.852	1.018	1.098	1.141	1.166	1.193	1.201	1.202	1.203	1.203	1.203	1.200	1.200	1.200	1.200	1.200
1.25	0.672	0.981	1.111	1.175	1.209	1.228	1.247	1.252	1.253	1.254	1.253	1.253	1.250	1.250	1.250	1.250	1.250
1.30	0.844	1.092	1.196	1.246	1.272	1.286	1.300	1.304	1.304	1.303	1.303	1.303	1.300	1.300	1.300	1.300	1.300
1.35	0.990	1.189	1.272	1.311	1.332	1.342	1.353	1.354	1.354	1.353	1.353	1.353	1.350	1.350	1.350	1.350	1.350
1.40	1.116	1.276	1.343	1.374	1.389	1.398	1.403	1.404	1.404	1.403	1.403	1.403	1.400	1.400	1.400	1.400	1.400
1.45	1.225	1.355	1.409	1.433	1.445	1.450	1.455	1.454	1.454	1.453	1.454	1.453	1.450	1.450	1.450	1.450	1.450
1.50	1.322	1.429	1.471	1.490	1.499	1.504	1.505	1.505	1.504	1.504	1.504	1.503	1.500	1.500	1.500	1.500	1.500
1.55	1.409	1.497	1.531	1.545	1.552	1.555	1.556	1.555	1.554	1.554	1.553	1.552	1.500	1.550	1.550	1.550	1.550
1.60	1.490	1.562	1.589	1.600	1.604	1.606	1.606	1.605	1.604	1.604	1.603	1.603	1.600	1.600	1.600	1.600	1.600
1.65	1.565	1.622	1.645	1.654	1.656	1.656	1.656	1.655	1.655	1.654	1.654	1.653	1.650	1.650	1.650	1.650	1.650
1.70	1.635	1.683	1.699	1.705	1.707	1.707	1.706	1.706	1.704	1.705	1.704	1.704	1.700	1.700	1.700	1.700	1.700
1.75	1.701	1.741	1.753	1.757	1.758	1.758	1.756	1.755	1.755	1.755	1.754	1.753	1.750	1.750	1.750	1.750	1.750
1.80	1.764	1.796	1.805	1.808	1.808	1.809	1.807	1.806	1.805	1.804	1.804	1.803	1.800	1.800	1.800	1.800	1.800
1.85	1.825	1.850	1.857	1.859	1.859	1.858	1.857	1.855	1.855	1.854	1.854	1.853	1.850	1.850	1.850	1.850	1.850
1.90	1.884	1.905	1.909	1.910	1.910	1.908	1.906	1.905	1.905	1.905	1.904	1.903	1.900	1.900	1.900	1.900	1.900
1.95	1.942	1.956	1.960	1.960	1.959	1.959	1.956	1.956	1.955	1.954	1.954	1.954	1.950	1.950	1.950	1.950	1.950
2.00	1.998	2.010	2.011	2.011	2.009	2.009	2.007	2.006	2.005	2.004	2.004	2.003	2.000	2.000	2.000	2.000	2.000

Iako metodu procjene maksimalnih vjerojatnosti $\hat{\delta}$ temeljene na g^* nije teško numerički dobiti, izračunavanje zahtjeva specijalizirani računalni program. U tu svrhu prilažemo Tablicu 13. koja daje $\hat{\delta}$ kao funkciju od g^* za vrijednosti od *n* između 10 i 100. Također daje metodu procjene maksimalnih vjerojatnosti od δ za svaku vrijednost od g^*, kao i minimalnu promatranu vrijednost od g^*(pri α =0.05) i minimalnu vrijednost od $\hat{\delta}$ za dani *n*. Primijetimo da su samo pozitivne vrijednosti od g^* i $\hat{\delta}$ prikazane u tablici, a negativne vrijednosti od $\hat{\delta}$ se dobivaju pomoću simetrije. Odnosno, ako g_0 odgovara $\hat{\delta}_0$, tada $-g_0$ odgovara $-\hat{\delta}_0$. Pronalazimo $\hat{\delta}$ za negativne vrijednosti od g^* tako da nađemo $\hat{\delta}$ koja je odgovarajuća za $|g^*|$ te invertirati predznak. Također vrijednosti od g^* veće od 2.00 nisu prikazane, zato što je u tom slučaju $\hat{\delta} = g^*$. Nadalje se bavimo isključivo slučajem kada je $n^E = n^C = n$.

4.3.2. Distribucija vrijednosti maksimalnih vjerojatnosti

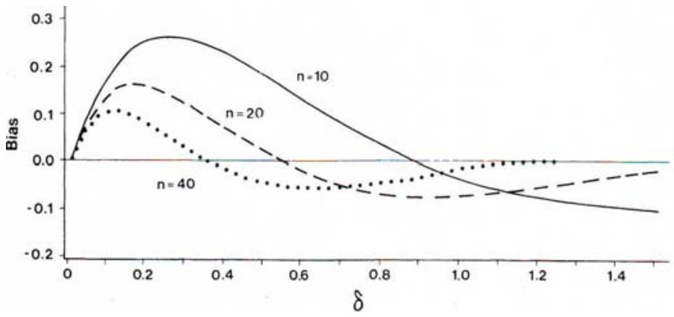
Metoda procjene maksimalnih vjerojatnosti $\hat{\delta}$ od δ je rastuća funkcija od g^*, tako da iz distribucije od g^* možemo dobiti distribuciju od $\hat{\delta}$. Praktični problem je taj da nije poznata eksplicitna matematička forma koja prenosi $\hat{\delta}$ u g^*. Jedan pristup ovom problemu je da koristimo blagu aproksimaciju ove funkcije, tako da je možemo koristiti u izračunavanju.

Funkcija gustoće se može numerički integrirati da nam pruži očekivanja i varijancu od $\hat{\delta}$. Srednja vrijednost i varijanca od $\hat{\delta}$ za različite vrijednosti od δ su dane u Tablici 14. Očekivana vrijednost od $\hat{\delta}$ nije uvijek jednaka δ. Prema tome, $\hat{\delta}$ je pristrana metoda procjene, te granica pristranosti ovisi o *n* i $\hat{\delta}$.

Tablica 12. Srednja vrijednost $E(\hat{\delta})$ i varijanca $V(\hat{\delta})$ metode procjene maksimalne vjerojatnosti $\hat{\delta}$ veličine učinka kada su samo značajni rezultati promatrani, [HEDG1985]

		δ									
		0.25		0.50		0.75		1.00		1.50	
n		$E(\hat{\delta})$	$V(\hat{\delta})$	$E(\hat{\delta})$	$V(\hat{\delta})$	$E(\hat{\delta})$	$V(\hat{\delta})$	$E(\hat{\delta})$	$V(\hat{\delta})$	$E(\hat{\delta})$	$V(\hat{\delta})$
10	Exact	0.52	0.28	0.68	0.27	0.81	0.33	0.96	0.40	1.40	0.52
	Simulated	0.51	0.30	0.70	0.27	—	—	0.97	0.42	1.44	0.54
20	Exact	0.39	0.10	0.52	0.13	0.69	0.18	0.92	0.21	1.48	0.20
	Simulated	0.39	0.10	0.52	0.13	—	—	0.92	0.21	1.51	0.18
30	Exact	0.34	0.07	0.48	0.10	0.68	0.13	0.95	0.13	1.49	0.11
	Simulated	0.34	0.07	0.48	0.10	—	—	0.95	0.13	1.53	0.10
40	Exact	0.31	0.05	0.46	0.08	0.70	0.10	0.98	0.08	1.50	0.08
	Simulated	0.31	0.05	0.46	0.08	—	—	0.98	0.08	1.52	0.07
50	Exact	0.28	0.04	0.44	0.07	0.71	0.07	0.99	0.06	1.49	0.07
	Simulated	0.28	0.04	0.45	0.07	—	—	0.99	0.06	1.51	0.06

Slika 6. je skica pristranosti $E(\hat{\delta}) - \delta$ kao funkcije od δ za *n* = 10,20 i 40. Pregled slike pokazuje da $\hat{\delta}$ precjenjuje δ za male vrijednosti od δ. Za ograničene vrijednosti od δ metoda procjene $\hat{\delta}$ podcjenjuje stvarne vrijednosti, te $\hat{\delta}$ postaje gotovo nepristrana za velike vrijednosti od δ.



Slika 6. Skica pristranosti $E(\hat{\delta}) - \delta$ kao funkcije od δ za n = 10, 20, i 40

U mnogim praktičnim situacijama pristranost može biti znatno manja nego maksimalna pristranost. Za veličine učinka u rangu od 0.5 do 0.8, pristranost nikad ne prelazi 0.18 čak i ako je n malen kao npr. 10. Npr., ako je n = 10 i δ = 0.50, tada je pristranost samo 0.06. Slično, kada je n = 20 i δ = 0.50, tada je pristranost samo 0.02. Kako su nezavisne interpretacije temeljene na sveukupnoj jačini od veličine učinka, relativno mala apsolutna pristranost od $\hat{\delta}$ ne djeluje ozbiljno velik da bi utjecao na nezavisne zaključke, osim ako su i veličina uzorka i veličina učinka prilično mali.

Zanimljivo je usporediti učinkovitost od $\hat{\delta}$ sa primjerom veličine učinka g kada su promatrani rezultati svih studija. Koristeći (8) za izračun varijance od g, vidimo da za δ =0.25 metoda procjene $\hat{\delta}$ ima aproksimativno jednaku varijancu kao i g. Kada su i n i δ veliki, varijanca od $\hat{\delta}$ je također prilično blizu varijance od g. Za posredne okolnosti, kada je δ ograničen da bude velik, ali n nije velik, tada je $\hat{\delta}$ znatno manje učinkovita nego g. Na primjer, kada je n = 20 i δ = 1.0, varijanca od $\hat{\delta}$ je približno dvostruka varijanca od g.

4.4. Procjena veličine učinka iz niza nezavisnih eksperimenata na temelju samo značajnih rezultata

Metoda maksimalnih vjerojatnosti omogućuje alternativu jednostavnim metodama brojanja glasova (eng. vote-counting method), a prednost joj je da se može primijeniti na bilo koju kolekciju istraživanja, uključujući i istraživanja koje nemaju jednake veličine uzorka. Najveći nedostatak ove metode je da zahtjeva posebni računalni program za izračunavanje vjerojatnosti i njenog maksimuma.

Ako je $g^* = (g_1^*, \ldots, g_k^*)$ vektor promatranja veličine učinka iz k nezavisnih eksperimenata sa veličinama uzorka $n_1, \ldots n_k$ po grupi i ako je δ veličina učinka populacije, tada zapis vjerojatnosti $L(g^* / \delta)$ je suma zapisa vjerojatnosti za svaki g_i^*.

Alternativa pronalaženju procjene maksimalnih vjerojatnosti od δ temeljena na svih k eksperimenata je dobivanje procjena maksimalnih vjerojatnosti $\hat{\delta}_1, \ldots, \hat{\delta}_k$ iz svakog eksperimenta posebno, te nakon toga udruživanje ovih procjena. Ovaj način ima nekoliko prednosti:

- nema potrebe za posebnim računalnim programom (Tablica 13. se može koristiti za dobivanje eksplicitnih procjena maksimalnih mogućnosti od δ iz svakog eksperimenta),
- procjene $\hat{\delta}_1, \ldots, \hat{\delta}_k$ se mogu proučavati individualno za otkrivanje potencijalnih vanjskih vrijednosti koje uvelike odstuju od ostalih procjena,
- analiza za linearnu kombinaciju od $\hat{\delta}_i$ vrijednosti je slična analizi uobičajeno korištenoj u kvantitativnim sintezama istraživanja.

Najjednostavnija kombinacija od $\hat{\delta}_1, \ldots, \hat{\delta}_k$ je netežinska srednja vrijednost. Kada eksperimenti imaju različite veličine učinka, netežinska srednja vrijednost neće biti optimalna, budući da eksperimenti sa velikom veličinom uzorka proizvode preciznije procjene od δ i morali bi davati više težine. Za težinske metode procjene od forme

$$\sum_{i=1}^{k} w_i \hat{\delta}_i / \sum_{i=1}^{k} w_i \tag{76}$$

težine koje minimaliziraju varijancu od (76) su proporcionalne recipročnim varijancama temeljenim na uzorku od $\hat{\delta}_i$, tako da je

$$w_i = 1/Var\left(\hat{\delta}_i\right), \qquad i = 1, \ldots, k\,. \tag{77}$$

Težinska metoda procjene će težiti manjoj pristranosti nego netežinska srednja vrijednost, budući da će procjene iz većih eksperimenata (sa odgovarajuće manjom pristranosti) dati veću težinu.

5. Procjenjivanje veličine učinka niza eksperimenata o učinkovitosti sustava xTEx-Sys

U ovom poglavlju će biti provedena meta-analiza na istraživanjima provedenim u akademskoj godini 2005/2006 i 2006/2007 na područnim znanjima: Uvod u računarstvo, Kemija, Fizika – optika, Priroda i društvo, QBASIC – programiranje, i Matematika. Tablica 13. prikazuje istraživanja, te broj ispitanika po grupama. Ukupan broj ispitanika koji su sudjelovali u ovim istraživanjima je 499. Tablica 14. prikazuje detaljne statističke podatke o svakom istraživanju, te dobivenu veličinu učinka svakog eksperimenta o učinkovitosti sustava xTEx-Sys. Cilj meta-analize koju ćemo napraviti nad ovim podacima je primjeniti metode procjene veličine učinka, dobiti prosječnu veličinu učinka svih istraživanja, ispitati homogenost, te na osnovu toga donijeti zaključke o učinkovitosti sustava xTEx-Sys.

Tablica 13. Istraživanja i broj ispitanika po grupama

Red. Br.	Područno znanje	Eksperimentalna grupa	Kontrolna grupa
Akademska godina 2005/2006			
1.	Uvod u računarstvo	40	40
2.	Kemija	20	21
3.	Fizika-optika	40	40
4.	Priroda i društvo	24	24
5.		24	24
6.		20	20
Akademska godina 2006/2007			
7.	Uvod u računarstvo	20	19
8.	QBASIC - programiranje	20	19
9.	Matematika	9	9
10.		9	9
11.		24	24
	$\sum$	250	249
		499	

Tablica 14. Rezultati iz niza eksperimenata o učinkovitosti

Područno znanje	Veličina uzorka	Trajanje	Statistički značajna razlika	Originalni rezultati srednjih vrijednosti i standardna devijacija	Razlike rezultatiaa srednjih vrijednosti i standardna devijacija	Veličina učinka
			Akademska godina 2005/2006			
Uvod u računarstvo	Eksperimentalna grupa: 40 studenata I. god. Kontrolna grupa: 40 studenata I. god.	14 tjedana	α =0.05, df=78			0,16
			chk test 1: t = -0,73, p=0,4676 NO	ctrl: $\bar{X}$ = 40,72, sd = 15,78 exp: $\bar{X}$ = 46,13, sd = 16,80	ctrl: $\bar{X}$ = -9,28, sd = 17,74 exp: $\bar{X}$ =-6,19 , sd =18,97	(0,17)
			chk test 2: t = 2,31, p=0,0235 YES	ctrl: $\bar{X}$ = 54,95, sd = 17,36 exp: $\bar{X}$ = 46,95, sd = 12,80	ctrl: $\bar{X}$ = 4,95, sd = 21,68 exp: $\bar{X}$ =-5,36, sd =17,86	(-0,47)
			final test: t = -3,62, p=0,0005 YES	ctrl: $\bar{X}$ = 37,48, sd = 13,44 exp: $\bar{X}$ = 51,23, sd = 12,30	ctrl: $\bar{X}$ = -12,53, sd =14,32 exp: $\bar{X}$ =-1,09, sd =13,66	(0,81)
Kemija	Eksperimentalna grupa: 20 učenika, 8. raz. O. Š. Kontrolna grupa: 21 učenika, 8. raz. O. Š.	10 tjedana	α =0.05, df= 39 t= -1,81, p= 0,0780 YES	ctrl: $\bar{X}$ = 24,42, sd = 8,16 exp: $\bar{X}$ = 25,70, sd = 5,28	ctrl: $\bar{X}$ = -9,83 , sd = 7,95 exp: $\bar{X}$ = -8,63 , sd = 6,71	0,60
Fizika - optika	Eksperimentalna grupa: 40 učenika, 8. raz. O. Š. Kontrolna grupa: 40 učenika, 8. raz. O. Š.	7 tjedana	α =0.05, df= 78 t=-3,67, p=0,0004 YES	ctrl: $\bar{X}$ = 44,03, sd = 22,89 exp: $\bar{X}$ = 62,28, sd = 20,91	ctrl: $\bar{X}$ = 0,16, sd = 0,24 exp: $\bar{X}$ = 0,34, sd = 0,20	0,75
Priroda i društvo	Eksperimentalna grupa: 24 učenika, 2. raz. O. Š. Kontrolna grupa: 24 učenika, 2. raz. O. Š.	6 tjedana	α =0.05, df= 46 t= -2,88, p= 0,0060 YES	ctrl: $\bar{X}$ = 79,83, sd = 10,58 exp: $\bar{X}$ = 91,00, sd = 13,10	ctrl: $\bar{X}$ = 8,17, sd = 8,30 exp: $\bar{X}$ = 14,83, sd = 7,69	0,80
	Eksperimentalna grupa: 24 učenika, 3. raz. O. Š. Kontrolna grupa: 24 učenika, 3. raz. O. Š.	6 tjedana	α =0.05, df= 46 t= -3,08, p= 0,0035 YES	ctrl: $\bar{X}$ = 86,17, sd = 7,64 exp: $\bar{X}$ = 95,50, sd = 5,32	ctrl: $\bar{X}$ = 8,67 , sd = 7,24 exp: $\bar{X}$ = 15,17 , sd = 7,36	0,83
	Eksperimentalna grupa: 20 učenika, 4. raz. O. Š. Control group: 20 učenika, 4. raz. O. Š.	6 tjedana	α =0.05, df= 38 t= -3,48, p= 0,0013 YES	ctrl: $\bar{X}$ = 84,00, sd =10,24 exp: $\bar{X}$ = 92,83, sd = 7,82	ctrl: $\bar{X}$ = 10,83, sd = 6,57 exp: $\bar{X}$ = 18,17 , sd = 7,98	1,11
			Akademska godina 2006/2007			
Uvod u računarstvo	Eksperimentalna grupa: 20 studenata I. god Kontrolna grupa: 19 studenata I. god	14 tjedana	α =0.05, df=37			0,42
			chk test 1: t = 1,04, p=0,3051 NO	ctrl: $\bar{X}$ = 54,74, sd = 19,62 exp: $\bar{X}$ = 50,30, sd = 18,62	ctrl: $\bar{X}$ = 13,74, sd =19,62 exp: $\bar{X}$ = 7,35 , sd =18,62	(0,33)
			chk test 2: t = -1,11, p=0,2742 NO	ctrl: $\bar{X}$ = 31,89 sd = 23,30 exp: $\bar{X}$ = 42,05, sd = 22,78	ctrl: $\bar{X}$ = -9,11, sd =23,30 exp: $\bar{X}$ = -0,90, sd =22,78	(0,35)
			final test: t = -3,77, p=0,0006 YES	ctrl: $\bar{X}$ = 40,79, sd = 11,79 exp: $\bar{X}$ = 57,20, sd = 12,14	ctrl: $\bar{X}$ = -0,21, sd =11,79 exp: $\bar{X}$ = 14,25, sd =12,14	(1,23)
QBASIC - programiranje	Eksperimentalna grupa: 20 studenata I. god Kontrolna grupa: 19 studenata I. god	14 tjedana	α =0.05, df=37			0,45
			chk test 1: t = -1,27, p=0,2110 NO	ctrl: $\bar{X}$ = 59,94, sd = 27,24 exp: $\bar{X}$ = 61,72, sd = 28,19	ctrl: $\bar{X}$ = 36,53, sd = 25,82 exp: $\bar{X}$ = 46,20, sd = 21,28	(0,37)
			chk test 2: t = -0,94, p=0,3533 NO	ctrl: $\bar{X}$ = 52,72, sd = 23,78 exp: $\bar{X}$ = 51,89, sd = 25,35	ctrl: $\bar{X}$ = 29,37, sd = 22,11 exp: $\bar{X}$ = 35,60, sd = 19,08	(0,28)
			final test: t = -1,83, p=0,0753 YES	ctrl: $\bar{X}$ = 43,61, sd = 23,90 exp: $\bar{X}$ = 49,22, sd = 23,66	ctrl: $\bar{X}$ = 21,74, sd = 15,23 exp: $\bar{X}$ =32,60 , sd = 21,47	(0,71)
Matematika	Eksperimentalna grupa: 9 učenika, 6. raz. O. Š. Kontrolna grupa: 9 učenika, 6. raz. O. Š.	7 tjedana	α =0.05, df=16			1,32
			chk test 1: t = -2,07, p=0,0550 YES	ctrl: $\bar{X}$ = 38,67, sd = 14,19 exp: $\bar{X}$ =54,33 sd = 21,11	ctrl: $\bar{X}$ = -34,33, sd = 11,85 exp: $\bar{X}$ = -19,11, sd = 18,62	(1,28)
			final test: t = -2,33, p=0,0332 YES	ctrl: $\bar{X}$ = 30,00, sd = 15,81 exp: $\bar{X}$ = 49,44, sd = 23,78	ctrl: $\bar{X}$ = -43,00, sd = 13,93 exp: $\bar{X}$ = -24,00, sd = 20,11	(1,36)
	Eksperimentalna grupa: 9 učenika, 8. raz. O. Š. Kontrolna grupa: 9 učenika, 8. raz. O. Š.	7 tjedana	α =0.05, df=16			0,15
			chk test 1: t = -0,30, p=0,7680 NO	ctrl: $\bar{X}$ = 57,22, sd = 24,12 exp: $\bar{X}$ = 58,89, sd = 22,93	ctrl: $\bar{X}$ = 6,67 , sd = 13,78 exp: $\bar{X}$ = 8,44, sd = 10,98	(0,13)
			final test: t = -0,24, p=0,8134 NO	ctrl: $\bar{X}$ = 67,78, sd = 22,93 exp: $\bar{X}$ = 70,00, sd = 30,92	ctrl: $\bar{X}$ = 17,22, sd = 14,12 exp: $\bar{X}$ = 19,55, sd = 25,56	(0,17)
	Eksperimentalna grupa: 24 učenika, 5. raz. O. Š. Kontrolna grupa: 24 učenika, 5. raz. O. Š.	5 tjedana	α =0.05, df=39			0,38
			chk test 1 t = -0,19, p=0,8503 NO	ctrl: $\bar{X}$ = 61,00, sd = 17,07 exp: $\bar{X}$ = 77,40, sd = 16,61	ctrl: $\bar{X}$ = -9,10, sd = 15,20 exp: $\bar{X}$ = -8,05, sd = 18,67	(0,07)
			chk test 2 t = -2,16, p=0,0370 YES	ctrl: $\bar{X}$ = 61,00, sd = 20,78 exp: $\bar{X}$ = 85,00, sd = 13,27	ctrl: $\bar{X}$ = -12,70, sd = 17,46 exp: $\bar{X}$ = 0,14, sd = 20,53	(0,74)
			final test: t = -1,03,	ctrl: $\bar{X}$ = 69,50, sd = 15,26	ctrl: $\bar{X}$ = -5,65, sd = 12,87	(0,34)

5.1. Izračunavanje *metoda procjene* za istraživanje temeljeno na područnom znanju Uvod u računarstvo.

Nepristrana metoda procjene d

Veličina učinka za ovo istraživanje je g =0.16, te $n^E = n^C = 40$. Koristeći (19) i (12) dobivamo:

$$d = J(N-2)g = J(78)*0.16 = 0.9904*0.16 = 0.158.$$

Metoda procjene maksimalne vjerojatnosti za veličinu učinka za ovo istraživanje, koristeći (17) je

$$\hat{\delta} = \sqrt{80/78} * 0.16 = 0.162.$$

Procjena smanjivanjem veličine učinka, koristeći (18), je

$$\tilde{g} = \frac{76}{78} * \frac{0.16}{J(78)} = 0.157.$$

U ovom istraživanju, za $n^E = n^C = 40$, vrijednosti 0.16, 0.158, 0.162, 0.157 za četiri metode procjene se ne razlikuju mnogo. Metode procjene su pouzdane , tako da je $\hat{\delta}^2 \geq g^2 \geq d^2 \geq \tilde{g}^2$.

Iz definicije (19) slijedi da su varijance poredane na isti način:

$$Var(\hat{\delta}) \geq Var(g) \geq Var(d) \geq Var(\tilde{g}).$$

Na ovaj način izračunavamo metode procjene za svako istraživanje posebno, te dobivamo rezultate navedene u tablici ispod.

Tablica 15. Metode procjene

Red. Br.	Područno znanje	g	d	$\hat{\delta}$	$\tilde{g}$
Akademska godina 2005/2006					
1.	Uvod u računarstvo	0.16	0.158	0.162	0.157
2.	Kemija	0.60	0.588	0.615	0.579
3.	Fizika-optika	0.75	0.743	0.759	0.738
4.	Priroda i društvo	0.80	0.787	0.817	0.780
5.		0.83	0.816	0.848	0.807
6.		1.11	1.088	1.139	1.073
Akademska godina 2006/2007					
7.	Uvod u računarstvo	0.42	0.411	0.431	0.406
8.	QBASIC - programiranje	0.45	0.441	0.462	0.435
9.	Matematika	1.32	1.257	1.400	1.213
10.		0.15	0.143	0.159	0.138
11.		0.38	0.374	0.388	0.369

Za sva ostala istraživanja također vrijedi da se metode procjene ne razlikuju mnogo, te je $\hat{\delta}^2 \geq g^2 \geq d^2 \geq \tilde{g}^2$. Zaključujemo da su metode procjene pouzdane i kod ostalih istraživanja, te da za svako vrijedi:

$$Var(\hat{\delta}) \geq Var(g) \geq Var(d) \geq Var(\tilde{g}).$$

5.2. Izračunavanje *intervala pouzdanosti* istraživanja temeljenog na područnom znanju Uvod u računarstvo

Proučavamo veličinu učinka g =0.16 i $n^E = n^C = 40$, te želimo postići interval pouzdanosti za δ. Znamo da je d = 0.158. Koristeći (21) za izračunavanje varijance $\hat{\sigma}^2(d)$ od d, dolazimo do:

$$\hat{\sigma}^2(d) = \frac{40+40}{(40)(40)} + \frac{(0.158)^2}{2(40+40)} = 0.0502.$$

Upotrebom (22) dobivamo da 95-postotni interval pouzdanosti za δ je $0.158 \pm 1.96\sqrt{0.0502}$, tj. $-0.28 \le \delta \le 0.60$.

Na jednak način izračunavamo intervale pouzdanosti za ostala istraživanja.

Tablica 16. Intervali pouzdanosti

Red. Br.	Područno znanje	d	$\hat{\sigma}^2(d)$	Interval pouzdanosti
Akademska godina 2005/2006				
1.	Uvod u računarstvo	0.158	0.0502	$-0.28 \le \delta \le 0.60$
2.	Kemija	0.588	0.1018	$-0.04 \le \delta \le 1.21$
3.	Fizika-optika	0.743	0.0535	$0.29 \le \delta \le 1.20$
4.	Priroda i društvo	0.787	0.8398	$-1.01 \le \delta \le 2.58$
5.		0.816	0.0903	$0.23 \le \delta \le 1.40$
6.		1.088	0.1148	$0.42 \le \delta \le 1.75$
Akademska godina 2006/2007				
7.	Uvod u računarstvo	0.411	0.1048	$-0.22 \le \delta \le 1.05$
8.	QBASIC - programiranje	0.441	0.1051	$-0.19 \le \delta \le 1.08$
9.	Matematika	1.257	0.2661	$-0.24 \le \delta \le 2.27$
10.		0.143	0.2228	$-0.78 \le \delta \le 1.68$
11.		0.374	0.0848	$-0.20 \le \delta \le 0.95$

Jedna od interpretacija po Hedgesu je: *„Aproksimacija distribucije metode procjene veličine učinka za veliki uzorak je ispravna za veličine uzorka koji premašuju 10 ispitanika po grupi, kada je veličina učinka između 0.25 i 1.50.“* [HEDG1985]. U našem slučaju bi to značilo da je aproksimacija distribucije veličine učinka na područnom znanju (3.) Fizika-optika jedina ispravna, dok su Priroda i društvo pod rednim brojem (5.) i (6.) na samoj granici danog uvjeta, pa ih također možemo smatrati ispravnim aproksimacijama.

Ako dobivene rezultate interpretiramo na način da želimo vidjeti koji su od rezultata statistički značajni, onda promatramo one intervale koji ne sadrže nulu. U našem istraživanju to su rezultati dobiveni na istraživanjima na područnom znanju pod rednim brojem (3.), (5.) i (6.). Odnosno, to su statistički značajni rezultati na razini 5%.

5.3. Izračunavanje intervala pouzdanosti za veličine učinka temeljenih na transformacijama

Ovo izračunavanje također prikazujemo na primjeru istraživanja temeljenog na područnom znanju Uvod u računarstvo. Već je poznato da je $g = 0.16$ i $d = 0.158$. Koristeći (23) sa $a = \sqrt{4 + 2(40/40) + 2(40/40)} = \sqrt{8}$ dobivamo:

$$h = h(d) = \sqrt{2}\sinh^{-1}\left(0.158/\sqrt{8}\right) = 0.079\,.$$

Izraz (26) daje granice pouzdanosti η_L i η_U za η:

$$\eta_L = 0.079 - 1.96\sqrt{1/(40+40)} = -0.140 \text{ i } \eta_U = 0.079 + 1.96\sqrt{1/(40+40)} = 0.298\,.$$

Koristeći $h^{-1}(x) = \sqrt{8}\sinh\left(x/\sqrt{2}\right)$, dobivamo:

$$\delta_L = \sqrt{8}\sinh\left(-0.140/\sqrt{2}\right) = -0.28 \text{ i } \delta_U = \sqrt{8}\sinh\left(0.298/\sqrt{2}\right) = 0.60\,,$$

i interval pouzdanosti $-0.28 \le \delta \le 0.60$.

Tablični prikaz intervala pouzdanosti za sva istraživanja:

Tablica 17. Intervali pouzdanosti temeljeni na transformacijama

Red. Br.	Područno znanje	d	a	Interval pouzdanosti
Akademska godina 2005/2006				
1.	Uvod u računarstvo	0.158	$\sqrt{8}$	$-0.28 \le \delta \le 0.60$
2.	Kemija	0.588	2.829	$-0.03 \le \delta \le 1.23$
3.	Fizika-optika	0.743	$\sqrt{8}$	$0.30 \le \delta \le 1.21$
4.	Priroda i društvo	0.787	$\sqrt{8}$	$0.21 \le \delta \le 1.40$
5.		0.816	$\sqrt{8}$	$0.24 \le \delta \le 1.43$
6.		1.088	$\sqrt{8}$	$0.44 \le \delta \le 1.78$
Akademska godina 2006/2007				
7.	Uvod u računarstvo	0.411	2.829	$-0.22 \le \delta \le 1.06$
8.	QBASIC - programiranje	0.441	2.829	$-0.19 \le \delta \le 1.09$
9.	Matematika	1.257	$\sqrt{8}$	$0.30 \le \delta \le 2.35$
10.		0.143	$\sqrt{8}$	$-0.79 \le \delta \le 1.09$
11.		0.374	$\sqrt{8}$	$-0.19 \le \delta \le 0.96$

Dobivene rezultate ćemo interpretirati na način da vidimo koji su od rezultata statistički značajni, odnosno promatramo intervale koji ne sadrže nulu. U našem istraživanju to su rezultati dobiveni na istraživanjima na područnim znanjima pod rednim brojem (3.), (4.), (5.), (6.) i (9.). Odnosno, to su statistički značajni rezultati na razini 5%.

U Tablici 18. su dani dobiveni rezultati iz Tablice 16. i Tablice 17. s ciljem da ih lakše usporedimo.

Tablica 18.

Red. Br.	Područno znanje	d	Interval pouzdanosti	Interval pouzdanosti temeljeni na transformacijama
Akademska godina 2005/2006				
1.	Uvod u računarstvo	0.158	$-0.28 \le \delta \le 0.60$	$-0.28 \le \delta \le 0.60$
2.	Kemija	0.588	$-0.04 \le \delta \le 1.21$	$-0.03 \le \delta \le 1.23$
3.	Fizika-optika	0.743	$0.29 \le \delta \le 1.20$	$0.30 \le \delta \le 1.21$
4.	Priroda i društvo	0.787	$-1.01 \le \delta \le 2.58$	$0.21 \le \delta \le 1.40$
5.		0.816	$0.23 \le \delta \le 1.40$	$0.24 \le \delta \le 1.43$
6.		1.088	$0.42 \le \delta \le 1.75$	$0.44 \le \delta \le 1.78$
Akademska godina 2006/2007				
7.	Uvod u računarstvo	0.411	$-0.22 \le \delta \le 1.05$	$-0.22 \le \delta \le 1.06$
8.	QBASIC - programiranje	0.441	$-0.19 \le \delta \le 1.08$	$-0.19 \le \delta \le 1.09$
9.	Matematika	1.257	$-0.24 \le \delta \le 2.27$	$0.30 \le \delta \le 2.35$
10.		0.143	$-0.78 \le \delta \le 1.68$	$-0.79 \le \delta \le 1.09$
11.		0.374	$-0.20 \le \delta \le 0.95$	$-0.19 \le \delta \le 0.96$

Intervali pouzdanosti, te intervali pouzdanosti veličine učinka temeljeni na transformacijama se ne razlikuju mnogo. Analizirajući oba dva načina dobivanja intervala pouzdanosti vidimo da su istraživanja provedena na područnim znanjima pod rednim brojem (3.), (5.) i (6.) postigla statistički značajne rezultate. Dok su rezultati istraživanja pod rednim brojem (4.) i (9.) statistički značajni kod intervala pouzdanosti temeljenih na transformacijama. Ukoliko želimo biti sigurni jesu li ti rezultati statistički značajni i u kojoj mjeri, trebamo izračunati intervale pouzdanosti temeljem neke drugačije transformacije. Jedna takva je predložena od strane Kraemera (Vidi 2.2.2.), ali je komplicirana za izračunavanje.

5.4. Izračunavanje robusne i neparametrijske procjene veličine učinka

Kod izračunavanja ovih procjena veličina učinka na istraživanjima koja promatramo (Tablica 14.) dolazi do poteškoća. Kako smo ograničeni sa podacima, izračuni se ne mogu napraviti.

Npr. za izračunavanje robusne procjene veličine učinka su potrebna zapažanja u eksperimentalnoj i kontrolnoj grupi za svakog ispitanika, na osnovu kojih se određuju medijani svake grupe, te se dolazi do procjene. Dok su za izračun neparametrijske procjene veličine učinka također potrebna inicijalna i završna mjerenja za svakog pojedinca u grupi, a na osnovu toga se dolazi do potrebnih razlika u grupama, povećanjima u eksperimentalnoj grupi u odnosu na kontrolnu grupa, te završnim rezultatima za svakog ispitanika.

Primjer koji dobro prikazuje izračunavanje neparametrijske procjene veličine učinka je dan u 2.3.2.

5.5. Izračunavanje težinskih linearnih kombinacija procjena

Težinska metoda procjene je dana formulom $d_W = w_1 d_1 + ... + w_k d_k$. U tu svrhu treba izračunati nenegativne težinske faktore w_i pomoću (51), tj. (53). U navedenoj tablici su dane vrijednosti potrebnih parametara.

Tablica 19. Parametri potrebni za izračun težinske metode procjene

Red. Br.	Područno znanje	n^E	n^C	$\widetilde{n}$	w	d	wd
Akademska godina 2005/2006							
1.	Uvod u računarstvo	40	40	20	0.160	0.158	0.02528
2.	Kemija	20	21	10.244	0.082	0.588	0.04822
3.	Fizika-optika	40	40	20	0.160	0.743	0.11888
4.	Priroda i društvo	24	24	12	0.096	0.787	0.07555
5.		24	24	12	0.096	0.816	0.07834
6.		20	20	10	0.080	1.088	0.08704
Akademska godina 2006/2007							
7.	Uvod u računarstvo	20	19	9.744	0.078	0.411	0.03206
8.	QBASIC - programiranje	20	19	9.744	0.078	0.441	0.03439
9.	Matematika	9	9	4.5	0.036	1.257	0.04525
10.		9	9	4.5	0.036	0.143	0.00515
11.		24	24	12	0.096	0.374	0.03590
	$\sum$						0.58606

Suma zadnjeg stupca u tablici je težinska metoda procjene za δ, odnosno $d_W = 0.58606$.
U sljedećoj tablici navodimo parametre koji su potrebni za daljnja izračunavanja.

Tablica 20. Parametri potrebni za izračunavanje d_+

Red. Br.	Područno znanje	n^E	n^C	d	$\hat{\sigma}^2(d)$	$1/\hat{\sigma}^2(d)$	$d/\hat{\sigma}^2(d)$	$d^2/\hat{\sigma}^2(d)$
Akademska godina 2005/2006								
1.	Uvod u računarstvo	40	40	0.158	0.0502	19.920	3.147	0.497
2.	Kemija	20	21	0.588	0.1018	9.823	5.776	3.396
3.	Fizika-optika	40	40	0.743	0.0535	18.692	13.888	10.319
4.	Priroda i društvo	24	24	0.787	0.8398	1.191	0.937	0.738
5.		24	24	0.816	0.0903	11.074	9.036	7.374
6.		20	20	1.088	0.1148	8.711	9.478	10.312
Akademska godina 2006/2007								
7.	Uvod u računarstvo	20	19	0.411	0.1048	9.542	3.922	1.612
8.	QBASIC - programiranje	20	19	0.441	0.1051	9.515	4.196	1.850
9.	Matematika	9	9	1.257	0.2661	3.758	4.724	5.938
10.		9	9	0.143	0.2228	4.488	0.642	0.092
11.		24	24	0.374	0.0848	11.792	4.410	1.649
	$\sum$					108.506	60.156	43.777

U cilju da izračunamo d_+, izraze $\hat{\sigma}^2(d_i)$, $1/\hat{\sigma}^2(d_i)$ i $d/\hat{\sigma}^2(d_i)$ smo prethodno izračunali. Težinska metoda procjene od δ je dana sa (54), te iznosi:

$$d_+ = \sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2(d_i)} / \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)} = \frac{60.156}{108.506} = 0.554,$$

sa procjenjenom varijancom

$$\hat{\sigma}^2(d_+) = \left(\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)} \right)^{-1} = \frac{1}{108.506} = 0.00922.$$

Ako postoji zajednička veličina učinka δ^* za sva istraživanja, tada 95-postotni interval pouzdanosti (δ_L, δ_U) za δ^* je:

$\delta_L = 0.554 - 1.96\sqrt{0.00922} = 0.3658$ i $\delta_U = 0.554 + 1.96\sqrt{0.00922} = 0.7422$, tj. $0.3658 \le \delta^* \le 0.7422$.

Kako interval pouzdanosti (δ_L, δ_U) ne uključuje nulu, možemo reći da je rezultat statistički značajan, i to na razini 5%.

5.6. Izračunavanja maksimalnih vjerojatnosti veličina učinka

Ova metoda se primjenjuje kada k istraživanja dijeli zajedničku veličinu učinka. Ukoliko ne dijeli, može se dobiti transformiranjem procjena. No, općenito transformacija dana u 2.1.4. ovisi izričito o omjeru n^E / n^C. Kako bi upotrebljavali istu transformaciju za svako istraživanje, omjer n_i^E / n_i^C mora biti jednak za svako istraživanje. U našem slučaju taj omjer nije jednak za svako istraživanje, pa nemožemo izračunati maksimalnu vjerojatnost veličine učinka.

S ciljem da pokažemo kako se izračunavanja mogu dobiti, uzet ćemo u obzir samo istraživanja provedena na onim područnim znanjima koja imaju jednak broj ispitanika u eksperimentalnoj i kontrolnoj grupi (Tablica 21.).

Tablica 21. Istraživanja sa svojstvom da je $n_i^E = n_i^C$

Red. Br.	Područno znanje	Eksperimentalna grupa	Kontrolna grupa
Akademska godina 2005/2006			
1.	Uvod u računarstvo	40	40
3.	Fizika-optika	40	40
4.	Priroda i društvo	24	24
5.		24	24
6.		20	20
Akademska godina 2006/2007			
9.	Matematika	9	9
10.		9	9
11.		24	24

S ovakvom restrikcijom smo dobili da je omjer n_i^E / n_i^C jednak za svako istraživanje, čime smo zadovoljili pretpostavke za daljnje izračunavanje. Transformacija stabiliziranja varijance za d je dana sa (60), te na primjeru područnog znanja Uvod u računarstvo iznosi:

$$h(d) = \sqrt{2}\sinh^{-1}\left(d / 2\sqrt{2}\right) = \sqrt{2}\sinh^{-1}\left(0.158 / 2\sqrt{2}\right) = 0.079 .$$

Na isti način dolazimo do h_i za ostala istraživanja, te ih uz ostale parametre koji su potrebni navodimo u Tablici 22.

Tablica 22. Parametri potrebni za izračunavanje metode maksimalnih vjerojatnosti

Red. Br.	Područno znanje	$n_i^E = n_i^C = n_j$	d_i	h_i	$2n_ih_i$	$2n_i(h_i - h_+)^2$
Akademska godina 2005/2006						
1.	Uvod u računarstvo	40	0.158	0.079	6.32	4.122
3.	Fizika-optika	40	0.743	0.367	29.36	0.298
4.	Priroda i društvo	24	0.787	0.389	18.67	0.331
5.		24	0.816	0.403	19.34	0.452
6.		20	1.088	0.531	21.24	2.025
Akademska godina 2006/2007						
9.	Matematika	9	1.257	0.609	10.96	1.653
10.		9	0.143	0.071	1.28	0.994
11.		24	0.374	0.186	8.93	0.691
	$\sum$	190			116.10	10.566

Neka su $h_1 = h(d_1), \dots, h_k = h(d_k)$ transformirane procjene i $\eta = h(\delta)$ parametar veličine učinka, te pretpostavimo da je jednak za sva istraživanja. Stoga je težinska linearna kombinacija procjena od η s najmanjom varijancom dana sa (61) i iznosi:

$$h_+ = 2\sum_{i=1}^{k} \frac{n_i h_i}{N} = \frac{116.10}{190 * 2} = 0.306 .$$

Nadalje, procjena od η je transformirana u procjenu od $\hat{\delta}^*$ pomoću inverzne transformacije $\hat{\delta}^* = 2\sqrt{2}\sinh\left(h_+ / \sqrt{2}\right) = 0.617$.

Prema tome, 95-postotni interval pouzdanosti (η_L, η_U) za η, koristeći (63), je

$$\eta_L = 0.306 - 1.96/\sqrt{380} = 0.205 \quad \text{i} \quad \eta_U = 0.306 + 1.96/\sqrt{380} = 0.407 .$$

Dakle, interval pouzdanosti za δ se dobiva koristeći (64):

$\delta_L = 2\sqrt{2}\sinh\left(0.205/\sqrt{2}\right) = 0.411$ i $\delta_U = 2\sqrt{2}\sinh\left(0.407/\sqrt{2}\right) = 0.825$, pa slijedi da je $0.411 \le \delta \le 0.825$.

Dobiveni interval pouzdanosti za δ ne uključuje nulu, pa je rezultat statistički značajan na razini 5%.

5.7. Prosječna veličina učinka

Ukupna veličina učinka svih dobivenih veličina učinka u meta-analizi se računa "mjerenjem" svake veličine učinka njenom "težinom". To se naziva *prosječna mjerena veličina učinka*, a računa se po formuli:

$$\delta = \frac{\sum w_i d_i}{\sum w_i}. \tag{78}$$

Mi smo za svako istraživanje pokazali kako različitim metodama dolazimo do veličine učinka. Za daljnji izračun koristit ćemo nepristranu metodu procjene veličine učinka d. Već je pokazano kako su metode pouzdane, i ne razlikuju se mnogo međusobno.

Tablica 23. Veličine učinka i težine istraživanja

Red. Br.	**Područno znanje**	d	w	wd
Akademska godina 2005/2006				
1.	Uvod u računarstvo	0.158	0.160	0.253
2.	Kemija	0.588	0.082	0.048
3.	Fizika-optika	0.743	0.160	0.119
4.	Priroda i društvo	0.787	0.096	0.076
5.		0.816	0.096	0.078
6.		1.088	0.080	0.087
Akademska godina 2006/2007				
7.	Uvod u računarstvo	0.411	0.078	0.032
8.	QBASIC - programiranje	0.441	0.078	0.034
9.	Matematika	1.257	0.036	0.045
10.		0.143	0.036	0.005
11.		0.374	0.096	0.036
	$\sum$		0.998	0.813

U Tablici 23. je dan sažetak rezultata svih istraživanja, odnosno veličine učinka i pripadajuće težine, te varijabla koja je potrebna za daljnje računanje. Sada se ukupna veličina učinka izračunava koristeći (78), te prosječna veličina učinka za navedena istraživanja iznosi:

$$\delta = \frac{\sum w_i d_i}{\sum w_i} = \frac{0.813}{0.998} = 0.82\,.$$

Ova prosječna veličina učinka znači da je prosječni učenik (student) 0.82 standardne devijacije uspješniji nego tradicionalno poučavan učenik ili student.

5.8. Ispitivanje homogenosti veličine učinka

Kod homogene distribucije disperzija veličina učinka oko njihovog prosjeka nije veća od očekivane greške vezane uz uzorak. Odnosno, u homogenoj distribuciji se individualna veličina učinka razlikuje od prosječne samo za tu grešku. Za provođenje testa homogenosti izračunava se Q vrijednost koristeći (69), a za izračun su potrebne varijable $1/\hat{\sigma}^2(d)$, $d/\hat{\sigma}^2(d)$ i $d^2/\hat{\sigma}^2(d)$, koje su dane u Tablici 19. Na taj način dobivamo:

$$Q = 43.777 - (60.156)^2 / 108.506 = 10.426 .$$

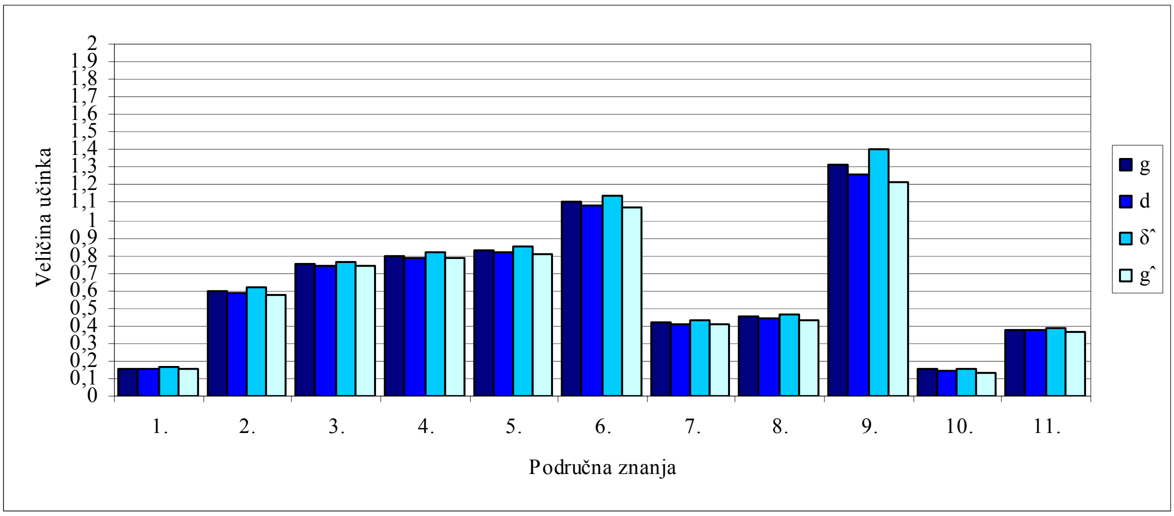
Dobivena vrijednost ne prelazi kritičnu vrijednost χ^2 za $\alpha = 0.05$, koja za 10 stupnjeva slobode iznosi 18.31, [TARA2008]. Nul hipoteza homogenosti se prihvaća, dana distribucija je, dakle, homogena. Nadalje, to znači da varijabilnost u ovom uzorku veličina učinka nije veća nego što se očekuje od greške vezane uz uzorak ispitanika.

5.9. Analiza rezultata meta-analize

U ovu meta-analizu uvrštena su istraživanja provedena na područnim znanjima prikazanim u Tablici 13., a detaljni rezultati su prikazani u Tablici 14. Na temelju tih rezultata primjenili smo teoretski dio ovog rada, te izračunali različitim metodama, različite veličine učinka, koji se u maloj mjeri razlikuju. Nadalje, izračunali smo intervale pouzdanosti, te odredili koji su rezultati statistički značajni na razini 5%. Te u konačnici odredili ukupnu veličinu učinka. Sljedeća tablica daje kratak prikaz pod rednim brojevima područna znanja, da Slika 7. bude jasnija.

Tablica 24. Područna znanja

Red. Br.	Područno znanje
1.	Uvod u računarstvo
2.	Kemija
3.	Fizika-optika
4.	Priroda i društvo
5.	
6.	
7.	Uvod u računarstvo
8.	QBASIC - programiranje
9.	Matematika
10.	
11.	



Slika 7. Veličine učinka odgovarajućih područnih znanja

Kao što smo već naglasili, sada se vidi i iz grafičkog prikaza, da se različite metode procjene za isto istraživanje međusobno mnogo ne razlikuju. Mi smo za dobivanje ukupne veličine učinka koristili nepristranu metodu procjene *d*, te na temelju nje dobili 0.82 sigma.

Sustav xTEx-Sys je pokazao najveću učinkovitost kod 9. područnog znanja Matematika, a nakon toga 6. Prirode i društva, odnosno 1.26 i 1.09 sigma, redom. Dok je najmanju veličinu učinka dalo istraživanje na 10. područnom znanju Matematika koje iznosi 0.14 sigma. Ipak, nije moguće generalizirati po područnim znanjima, na osnovu ovih istraživanja. Broj ispitanika po svakom istraživanju nije prevelik, ali kada bi se u svakoj akademskoj godini istraživalo na istim područnim znanjima, i tako veći niz godina, bilo bi dovoljno rezultata i podataka za opsežniju i precizniju meta-analizu. Ali i uz ovako skroman broj istraživanja sustav xTEx-Sys, kao jedan od sustava e-učenja pokazao se učinkovit, a to potvrđuje i dobivena ukupna veličina učinka, koja u ovoj meta-analizi iznosi 0.82 sigma.

Postoje razne interpretacije veličine učinka. Ovo su neke od njih:

- Prvi način je u izjednačavanju vrijednosti veličine učinka i vrijednosti standardne devijacije za koju je eksperimentalna grupa bolja od kontrolne. U ovom istraživanju bi to značilo da je eksperimentalna grupa bolja od kontrolne za 0.82 standardne devijacije [BLOO1984].
- Kako je veličina učinka ekvivalentna z-vrijednosti standardne normalne distribucije, može se reći da veličina učinka od 0.82 znači da je uspjeh prosječne osobe u eksperimentalnoj grupi bolji od uspjeha 79 % osoba iz kontrolne grupe, što odgovara površini ispod normalne krivulje za z-vrijednost 0.82. U jeziku percentila bi to značilo da je aritmetička sredina eksperimentalne grupe na 79. percentilu kontrolne grupe.

- Sljedeći način interpretiranja je u vjerojatnosti pogađanja iz koje je grupe neka osoba na temelju poznavanja njenog uspjeha. Vjerojatnost u ovom istraživanju je 0.66, što predstavlja otprilike dvije trećine šanse za točnim pogotkom.
- Četvrti način interpretacije veličine učinka je u terminima postotka nepreklapanja rezultata eksperimentalne grupe s rezultatima kontrolne grupe. Veličine učinka 0.82 označava nepreklapanje od otprilike 47.4% dviju distribucija.
- Nadalje, može se interpretirati na način, kako to radi većina analitičara, da je uspješnost tradicionalno poučavane grupe 31%, dok je za eksperimentalnu grupu uspješnost 69%.

6. Zaključak

Kroz ovaj rad objasnili smo ključni korak u meta-analizi, a to je odabir statističke metode procjene veličine učinka i njenu primjenu na daljnju analizu. Pobliže su objašnjene parametrijske i neparametrijske metode, te njihove podvrste. Na kratkim primjerima, kroz tekst, pojašnjene su metode procjene, te je dana međusobna usporedba navedenih metoda, kao i njihove prednosti i nedostatci.

Definiranu smo metodologiju primijenili na istraživanja o učinkovitosti autorske ljuske za izgradnju inteligentnih tutorskih sustava xTEx-Sys. Iako je učinak inteligentnih tutorskih sustava manji od učinka individualnog poučavanja, oni povećavaju učinak i kvalitetu procesa učenja i poučavanja. Jedna od prednosti inteligentnih tutorskih sustava je smanjivanje vremena potrebnog za svladavanje određene nastavne cjeline ili sadržaja, kao i povećanje motivacije kod učenika i studenata. Većina njih smatra da je ovakav način učenja jednostavniji i zanimljiviji, te da je vrednovanje znanja objektivnije. Sustavi omogućavaju učeniku učenje vlastitim tempom, te odmah daju povratnu informaciju vezanu uz ocjenu, kao i ispravak netočnih odgovora na pitanja, što je učenicima od velike važnosti.

Sinteza istraživanja je predstavljena meta-analizom, dakle dobiveni su rezultati kvalitetni onoliko koliko su kvalitetna istraživanja koja ta meta-analiza uključuje. U ovom radu je meta-analiza provedena na istraživanjima na raznim područnim znanjima, kao što su Uvod u računarstvo, Kemija, Fizika, Priroda i društvo, QBASIC i Matematika, a veličina uzorka je ukupno 499 učenika i studenata, tokom dvije akademske godine. Dobivena je ukupna veličina učinka od 0.82 sigma, koja predstavlja značajnu učinkovitost sustava xTEx-Sys naspram tradicionalne nastave. Taj učinak se može interpretirati na razne načine, kao što je prikazano u poglavlju *Analiza rezultata meta-analize*. Dobivena veličina učinka je u skladu sa prijašnjim meta-analizama istraživanja, te se vidi napredak i povećanje učinkovitosti sustava e-učenja.

Na temelju dobivenih rezultata možemo ocijeniti ukupnu veličinu učinka svih istraživanja, ali kako veličine učinka variraju kod određenih područnih znanja, ne možemo iznijeti zaključke za svako područno znanje posebno, jer je veličina uzorka mala. Kada bi imali još rezultata istraživanja iz istih područnih znanja, mogli bi napraviti meta-analizu za svako područje zasebno. Iz takvih rezultata bi se vidjelo u kojim područnim znanjima je učenicima lakše pratiti nastavne sadržaje uz ovakav sustav, a gdje ipak treba više tradicionalnog načina učenja i poučavanja.

U svakom slučaju, inteligentni tutorski sustavi ne mogu „zamijeniti“ uloge nastavnika u potpunosti, ali mogu im olakšati i poboljšati rad s djecom. Znatno bi se olakšalo usvajanje nastavnih sadržaja, više bi motiviralo učenike, ali isto tako nastavnici bi „uštedjeli“ na vremenu, te ostatak vremena bi mogli iskoristiti za kvalitetniju komunikaciju sa učenicima, raspravu, itd. Inteligentne tutorske sustave bi trebalo trajno uključiti u nastavu, kao jednu od metoda učenja i poučavanja, a meta-analize kao ova samo daju smjernice za njihovo daljnje korištenje i unapređivanje.

7. Literatura

[BLOO1984] Bloom, B. S., *The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring*, American Education Research Association, 1984.

[GLAS1976] Glass, G. V.: *Primary, secondary, and meta-analysis of research*, Laboratory of Educational Research, University of Colorado, 3-8, 1976

[HEDG1981] Hedges, L., Distribution Theory for Glass's Estimator of Effect Size and Related Estimators, *Journal of Educational and Behavioral Statistics*, 6, 2, pp. 107-128, 1981.

[HEDG1984] Hedges, L., Estimation of Effect Size Under Nonrandom Sampling: The Effects of Censoring Studies Yielding Statistically Insignificant Mean Differences. 1984.

[HEDG1984] Hedges, L. and Olkin, I. *Nonparametric Estimator of Effect Size in Meta-Analysis,* American Psychological Association, Washington, 1984, pp. 573-580.

[HEDG1985] Hedges, L., Olkin, I.: *Statstical Methods for Meta-Analysis.* New York; Academic Press, 1985, pp. 75-145, 285-307

[JOHN1939] Johnson, N. L., Welch, B. L.: Applications of The Non-central t-distribution, Biometrika 31, pp. 362-389, 1939.

[KENN1997] Kenneth N. T., Randall E. S.:An Evolution of Rosenthal and Rubin's Binomial Effect Size Display, *Journal of Educational and Behavioral Statistics*, pp.109-117, 1997

[KRAE1982] Kraemer , Helena C., Andrews, Gavin: *A nonparametric Technique for Meta-Analysis Effect Size Calculation.* Psychological Bulletin*,* 91 (2), pp 404-412, 1982.

[KRAE1983] Kraemer , Helena C.: *Theory of Estimation and Testing of Effect Sizes: Use in Meta-Analysis,* Journal of Educational an Behavioral Statistics*, 8, 2, pp 93-101, 1983.*

[LANE1978] Lane, D. M, Dunlap, W. P.: Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology*, *31*, 107-112.,1978

[LIPS2001] Mark W. Lipsey*,* David B. Wilson: *Practical meta-analysis.* Sage Publications, Inc., Thousand Oaks, California, 2001.

[ROSE1979] Rosenthal, R., Rubin, Donald B.: *Journal of Applied Social Psychology*, pp. 395-396, 1979.

[SARH1962] Sarhan Ahmed E., Greenberg Bernard G., *Contributions to order statistics*, John Wiley and Sons, Inc., New York, 1962, 482 pp.

[TARA2008] Taradi K. Sunčana, Računalo ("tablica") graničnih hi^2 ili p u hi-kvadrat-raspodjeli, http://www.mef.hr/if/alati/tablice/skripte/hi_kvadrat.htm , (30/08/08.)

[WIKI2008] Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Meta-analysis , (15/09/08).

[WIKI2008] Wikipedija, Slobodna enciklopedija, http://hr.wikipedia.org/wiki/E-u%C4%8Denje#cite_note-1 , (02/09/08).