Limperg Instituut

Statistical engineering, transparant en efficiënt controleren

Ed Broeze (Vrije Universiteit)
Wouter Gerards (Agentschap SZW)

Limperg Instituut

Inleiding

1. ESF uitvoering/achtergrond
   1. ESF in Nederland
   2. Europese controle structuur
   3. Doel van het Agentschap

2. Technische informatie methodiek
   1. Voor het trekken van de steekproef
   2. Declaratie en Steekproef
   3. Controle
   4. Op te hogen en bepaling correctie

3. Karakteristieken
   1. Beta verdeling
   2. Onzekerheidgrenzen
   3. Trekken steekproef

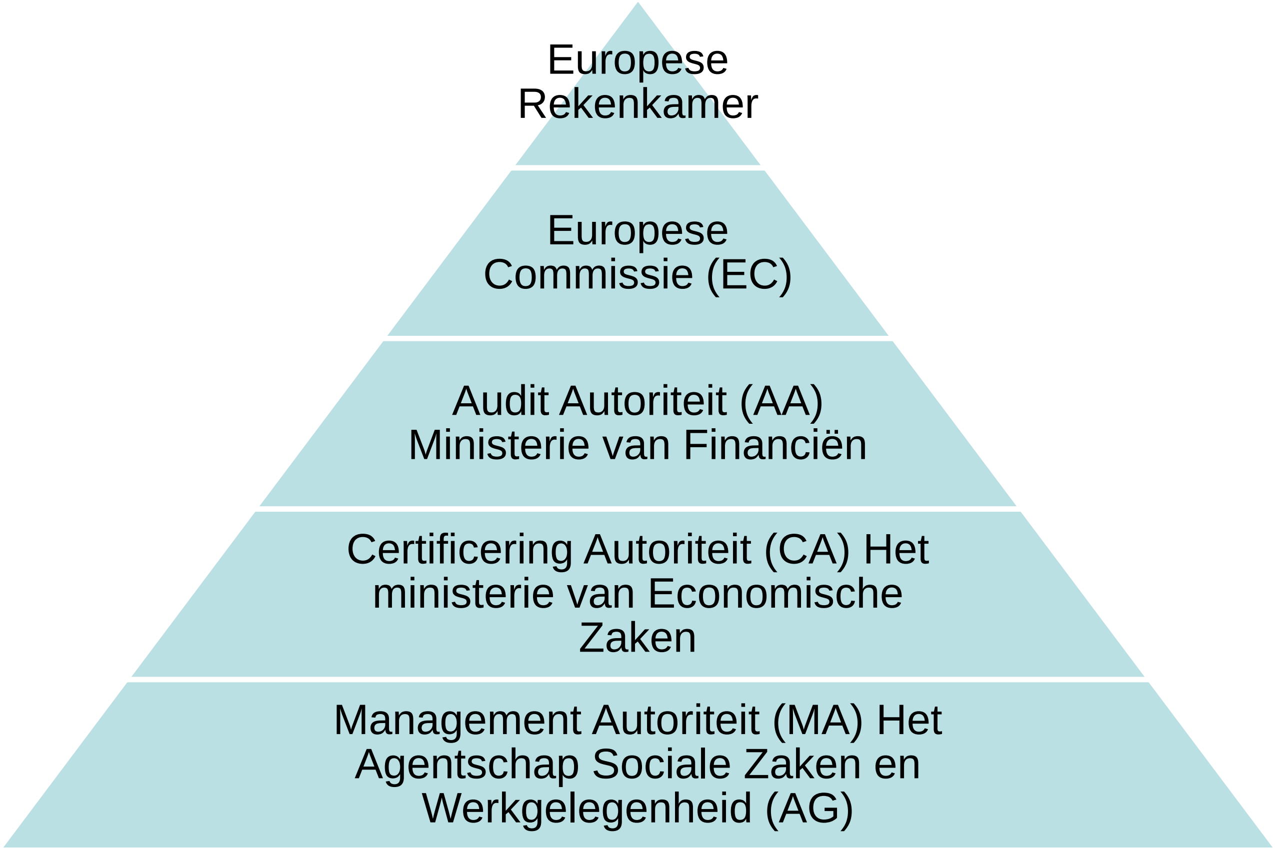
4. (Bredere) Toepasbaarheid en aannamen

Limperg Instituut

1.1Europees Sociaal Fonds in Nederland

- Het Europees Sociaal Fonds (ESF) verbetert de kansen van mensen op de arbeidsmarkt.
- Nederland heeft van 2007 tot en met 2013 een bedrag van € 830 miljoen te verdelen.
- Bedrijven, gemeentes en het UWV voeren projecten uit.

1.2 Audit; Simpele Piramide



1.3 Doel van de MA; Het Agentschap

- Het AG controleert declaraties voor projecten vanuit ESF. Vervolgens bundelt zij deze en declareert deze aan Brussel.

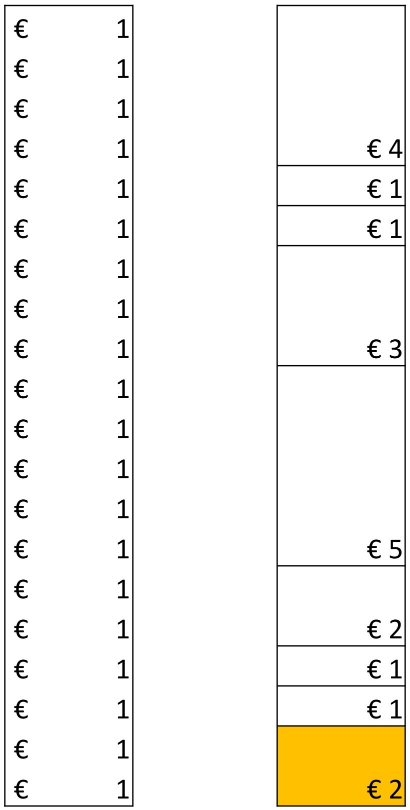
- Met 95% zekerheid kunnen stellen dat de geschatte fout niet meer dan 2% afwijkt van de werkelijke fout.

- Maximale fout (V_{95})- puntschatting (V) <2% van de declaratie (M1)

1.3.1Geld Eenheden Steekproef/ Monetary Unit Sample

- Doel AGSZW bereikt mbt GES/MUS
- Wijkt af van klassieke en meest voorkomende vorm;
  - Doel is schatting van de fout en niet afgeven goedkeurende verklaring!
  - Twee statistische grootheden;
    1. Gezien deel
    2. Niet gezien deel
- Evaluatie met Beta verdeling
- Hoog percentage gezien (50%+)
  - Goed toepasbaar op declaraties met verdichte posten!

2.1 Technische informatie; voor het trekken van de steekproef



1 kritische post van 2 euro (K) wordt verwijderd; M1-K=M= Massa gecontroleerd met steekproef

Declaratie (M1) 20 euro's 9 posten

Limperg Instituut

2.2 Technische informatie; Declaratie en Steekproef

- Geld Eenheden steekproef (GES of MUS)

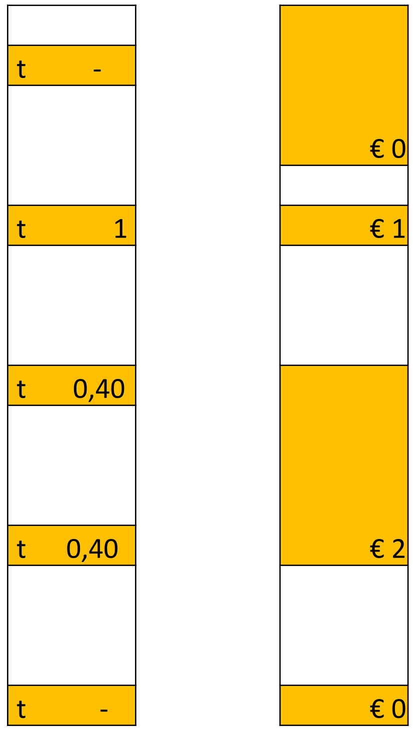
€ 1		€ 1	
€ 1		€ 1	
€ 1		€ 1	
€ 1	€ 4	€ 1	€ 4
€ 1	€ 1	€ 1	€ 1
€ 1	€ 1	€ 1	€ 1
€ 1		€ 1	
€ 1		€ 1	
€ 1	€ 3	€ 1	€ 3
€ 1		€ 1	
€ 1		€ 1	
€ 1		€ 1	
€ 1		€ 1	
€ 1	€ 5	€ 1	€ 5
€ 1		€ 1	
€ 1	€ 2	€ 1	€ 2
€ 1	€ 1	€ 1	€ 1
€ 1	€ 1	€ 1	€ 1

Declaratie (M)18 euro's 8 posten

Steekproef 5 euro's 4 posten

2.3 Technische informatie; Controle

Gemiddelde Taint;
1,8/5=0,36

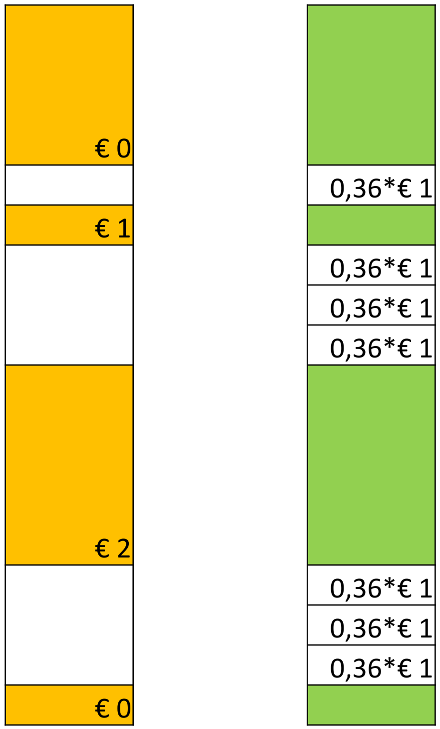


3 euro is fout in de posten; bekende fout

Controle, Fouten in steekproef; Taints

Limperg Instituut

2.4 Technische informatie; Ophogen en bepaling fout (correctie)



Gemiddelde Taint;
1,8/5=0,36

Op te hogen fout=
€7*0,36= €2,52

€ 3 euro bekende fout

Totale Correctie=
€3+€2,52=€5,52

Onbekende massa; 7 euro's

Limperg Instituut

3.1a Karakteristieken; **beta verdeling**

- De zekerheidsgrenzen worden bepaald a.h.v de Beta inverseverdeling.
- De Betaverdeling: kan als kansverdeling voor de gemiddelde taint (T_{gem}) in de declaratie; heeft soortgelijke karakteristieken als de Stringerbound maar rekent veel gemakkelijker (fout in euro's= T_{gem}*omvang declaratie) .
- Betaverdeling is gebaseerd op 2 parameters, die gekoppeld zijn aan steekproef: a en b; a staat voor kans per steek op een fout , a+b voor steekproefomvang

Limperg Instituut

3.1b Karakteristieken; **beta verdeling**

- In (Bayesiaanse) kansrekening horen betaverdeling en binomiale verdeling, wiskundig sluitend, bij elkaar (zijn ‘geconjugeerd’), als volgt:
- Als beta(a_1;b_1) een apriori verdeling (prior) is voor p en als in n trekkingen uit een binomiale verdeling met succeskans diezelfde p, k successen worden gevonden, dan is de aposteriori verdeling (posterior) voor p wéér beta, met parameters
  a_2=a_1+k; b_2=b_1+n-k.
- In ons voorbeeld en in controlepraktijk Agentschap gaan we niet uit van voorinformatie; daarbij past een beta(1;1) verdeling als prior; als in voorbeeld 1 hele fout was gevonden (in 5 trekkingen) dan wordt posterior beta (2;5); als 2 hele fouten waren gevonden, dan wordt posterior beta(3;4). Tot zover wiskundig sluitend. Maar er is bij elkaar 1,8 fouten gevonden. Kan simpel in betaverdeling worden verwerkt tot posterior beta(2,8; 4,2).
- Met excelfunctie beta-invers kan hieruit de 95% boven-zekerheidsgrens voor p worden gevonden: p_{bov}=beta.inv(0,95;2,8;4,2)=0,7 (dus bij t_{gem}=0,36).

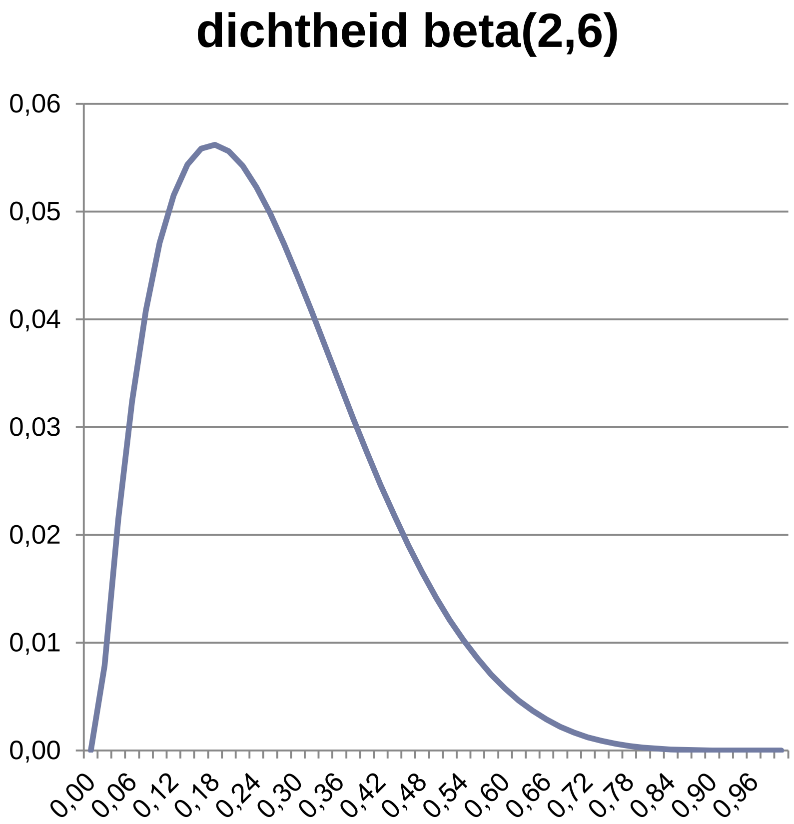
3.1c karakteristieken; zekerheidsgrenzen in **beta verdeling**

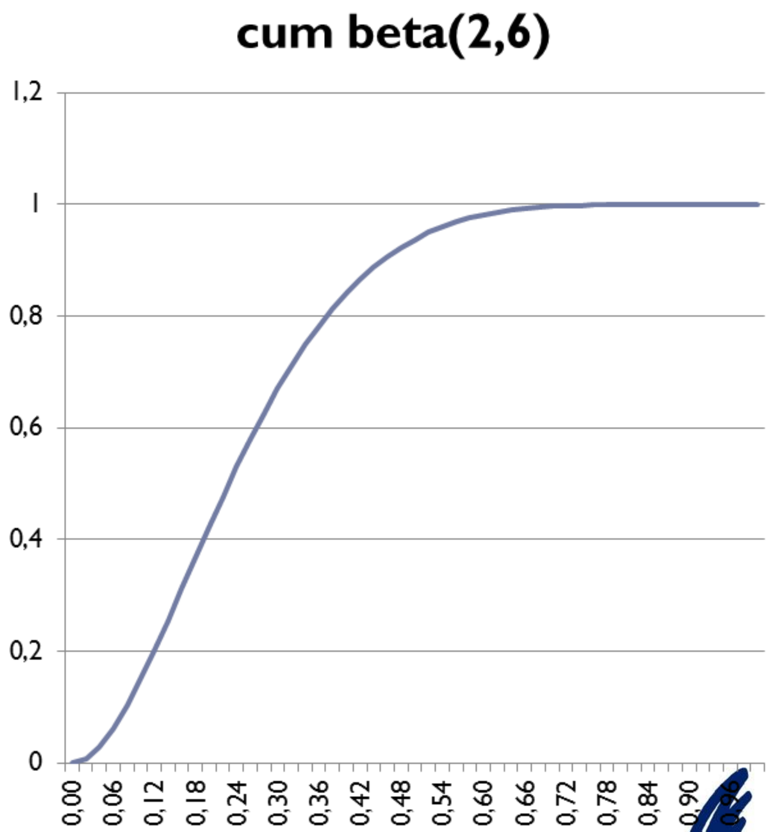
Algemener:

- Steekproef van omvang n (euro's); levert een gemiddelde taint, t_{gem} op; leidt bij noninformatieve prior (beta(1, 1)tot posterior voor p, oftewel T_{gem}: beta met parameters $a=1+n*t_{gem}$, $b=1+n-n*t_{gem}$ **)
- Zekerheidsgrens: T_{bov} voor gemiddelde taint T_{gem}: uit betaverdeling, zo, dat geldt $P(T_{gem}>T_{bov})=0,05$ (dus zekerheid van 95%).
- Deze grens wordt door Excel geleverd mbv de BETA.INV functie.

Limperg Instituut

3.1d Karakteristieken; **beta verdeling**





Limperg Instituut

3.2 Karakteristieken; **onnauwkeurigheid**

- De controle levert de waarde van twee statistische grootheden op:
  - De bekende fout (geziene deel)
  - De (geschatte) taint
- Over het geziene is er geen sprake van statistische onnauwkeurigheid of onzekerheid en nemen we aan dat er geen meetfouten zijn gemaakt.
- Onder aanname dat verdeling van taints in niet geziene deel overeenkomt met verdeling in gehele populatie is onnauwkeurigheid gemiddelde van geziene (60%) en niet geziene deel (40%). 60%*0+40%*4%=1.6% (zie ook 4b, 4c)
- ‘Onnauwkeurigheidsversoepeler ‘=**1/(1-percentage gezien)**

3.3Karakteristieken; **trekken steekproef/ gebruik excel**

Ex Ante aannamen: De fout is tussen 0% en 100%

- Stap 1 Declaratie wordt in Excel format geplakt.

- Stap 2 voor n steken wordt gecontroleerd of met 95% zekerheid de onnauwkeurigheid wordt overschreden van 2% voor alle fouten tussen de 0% en 100%.

- Stap 2 wordt herhaald met een hogere n totdat de onnauwkeurigheideis voldaan is.

- Stap 3 steekproef wordt vastgezet.

Limperg Instituut

4a (Bredere) Toepasbaarheid

- Methodiek heeft experimentele elementen (onnauwkeurigheidsverssoepeler)
- Testen voor gebruik Agentschap SZW
  - Vergelijking klassiek ←→ methodiek
  - Vergelijking integraal ←→ methodiek
  - Top stratum ←→ methodiek
- Raison d’être experimentele elementen.
  - Toepasbaarheid
  - Efficiëntie
- Opletten bij bredere toepassing!

Limperg Instituut

4b Meer over aanname

- Aanname voor onnauwkeurigheidsverruimer is intuïtief aannemelijk; gaat er vanuit dat alle taints door zelfde administratieve proces worden voortgebracht.
- Is echter ook aanvechtbaar, bijvoorbeeld omdat het niet gezien deel slechts een (klein) deel van de populatie vertegenwoordigt, waardoor er relatief makkelijk afwijkende verdeling kan ontstaan
- Lastig om er dingen over te bewijzen, omdat er veel afhankelijkheden zijn (n van de procedure in vorige sheets, zeker getrokken posten van n, hiervan weer overig geziene deel, daarvan weer niet geziene deel)
- Maar wel toetsen op praktische werking mogelijk

4c Meer over aanname

Toetsen op goede werking:

- Correlatie postgrootte ~ taint is positief: wijst op geen grotere variatie in niet geziene deel en geen onderschatting fout.
- Nog te doen: verschil gemiddelde taint in posten > interval vs gemiddelde taint in kleinere posten

Limperg Instituut

4d Conclusie

Conclusie: ‘Work in progress’: toepasbaarheid voor Agentschap bevestigd; doorgaand onderzoek naar grenzen en/of voorwaarden toepasbaarheid, per geval en globaal

Vragen?

Contact:

Ed Broeze
g.b.broeze@vu.nl

Wouter Gerards
http://nl.linkedin.com/pub/wouter-gerards/23/980/77a