

Hvem er Institutt for Psykologis beste p-hacker?

Ole Fredrik Borgundvåg Berg
NTNU

Replikasjonsgraden i psykologi er lav. Den mest kjente studien på området kommer fra Open Science Collaboration, som fant at når de prøvde å replikere 97 ulike psykologistudier, så klarte de kun å få signifikante resultater 36% av replikasjonene (Open Science Collaboration, 2015). Videre var effektstørrelsen i replikasjonene omtrent halvparten av de effektstørrelsen i de opprinnelige studiene. En annen studie med kun 28 replikasjoner fant en replikasjonsrate på 54% (Klein mfl., 2018). Det er ulike grunner til at replikasjonsraten er lav, og noe kan skyldes ting som publikasjonsbias, p-hacking og liknende. Ved å se på p-verdiene i artiklene til ulike forskere kan man prøve å finne i hvilken grad p-hacking forekommer og man kan prøve å estimere i hvilken grad resultatene er forventet å replikere. I denne artikkelen skal jeg se på p-verdiene til ulike forskere ved Institutt for Psykologi.

Hvoran lese grafene

Grafene er histogrammer, der dataene er z-verdier fra artiklene de har publisert. Om andre statistiske tester enn z-test er brukt, er test-statistikken omgjort til en z-verdi ved å ta z-verdien med samme p-verdi. I grafen er det en rød linje som går ved $z = 1.96$. Dette er fordi en z-verdi på 1.96 eller høyere fører til en p-verdi på 0.05 eller lavere. Med andre ord, signifikante resultater er til høyre for linja og null-resultater er til venstre for linja. Det er tre tall i grafen. «Observed discovery rate» (ODR) er andelen av de publiserte resultatene som er signifikante. «Expected discovery rate» (EDR) er den forventede andelen signifikante resultater du vil få om du kjører alle forsøkene til forskeren på nytt. Man skulle kanskje tro at dette ville være det samme som ODR, men forskjellen er at EDR tar hensyn til seleksjonseffekter. Med spesifikt, er ODR basert på de publiserte resultatene, mens EDR tar baserer seg på alle resultatene, også de null-resultatene modellen mener finnes men ikke er blitt publisert (see avsnittet om publikasjonsbias). «Expected replication rate» (ERR) er forventet andel signifikante resultater man vil få dersom man replikerer alle de signifikante resultatene på nytt. Det er også det som tilsvarer «replikasjonsrate» slik det er referert i første avsnitt. Mer detaljer om disse verdiene kan man finne i Bartoš og Schimmack (2022).

Alexander Olsen

[Figur 1 about here.]

Alexander Olsen har gjort en god jobb med p-hacking og publiseringsbias. Vi ser en klar diskontinuitet ved $z = 1.96$. En forventet replikasjonsrate på 68% er ganske OK for psykologi.

Andrea Melanie Kessler

[Figur 2 about here.]

Andrea Melanie Kessler merker seg ved å ha en høy forventet replikasjonsrate på 90%. Hun har heller ingen diskontinuetet ved $z = 1.96$ noe som forteller oss at p-hacking eller publiseringsbias er usannsynlig.

Audrey Van der Meer

[Figur 3 about here.]

Audrey Van der Meer har kun en liten diskontinuetet ved $z = 1.96$, noe som tyder på lite men noe p-hacking/publiseringsbias. Hun har også en forventet replikasjonsrate på 70%, noe som er ganske OK. Det skal dog sies at vi fant noen p-verdier i den ene artikkelen hennes som ikke samsvarte med t-verdiene (Kayed & Van der Meer, 2009). Om t-verdiene var riktige, er noen av resultatene som er rapportert som signifikante ikke faktisk signifikante.

Audun Havnen

[Figur 4 about here.]

Audun Havnen har en noe lav replikasjonsrate på bare 58%, samt noe diskontinuetet ved $z = 1.96$. Dette tyder på litt p-hacking/publiseringsbias, men ikke så mye.

Beate Wold Hygen

[Figur 5 about here.]

Beate Wold Hygen har en diskontinuetet ved $z = 1.96$, men her med motsatt fortegn! Ingen p-hacking her, om noe så tyder ting på det motsatte. En forventet replikasjonsrate på 66% er også ganske bra. God rapportering av null-resultater!

Christian Kløckner

[Figur 6 about here.]

Christian Kløckner har en solid z-verdi-graf. Ingen diskontinuetet og mange null-resultater. En høy forventet replikasjonsrate på hele 89% er også til å legge merke til.

Dawn Marie Behne

[Figur 7 about here.]

Dawn Marie Behne har ingen diskontinuetet ved $z = 1.96$ og en helt OK forventet replikasjonsrate på 72%. Ingen tegn til p-hacking eller publiseringsbias!

Frederick Anyan

[Figur 8 about here.]

Frederick Anyan har en helt OK z-verdi-graf. Ingen store diskontinueteter og en forventet replikasjonsrate på 73%.

Frederikus Van der Weel

[Figur 9 about here.]

Frederick Van der Weel har den største diskontinueteten av alle i sin z-verdi-graf. Det snek seg dog med noen få null-resultater innimellom. Forventet replikasjonsraten på 70% er ganske OK, da.

Gerit Pfuhl

[Figur 10 about here.]

Gerit Pfuhl har klart å unngå diskontinueteter i sin graf og har mange null-resultater. En forventet replikasjonsrate på 67% er ganske grei. Man begynner dog å se konturene av en gausskurve i z-verdi-grafen hennes, noe som gjør at man kan begynne å lure på om det er nullhypotesen som er i grafen her.

Hermundur Sigmundsson

[Figur 11 about here.]

Hermundur Sigmundsson har noe diskontinuetet ved $z = 1.96$, men det er også en god del null-resultater. Den forventede replikasjonsraten på 70% er også ganske OK.

Ida Emilia Brunner

[Figur 12 about here.]

Ida Emilia Brunner har en ganske høy forventet replikasjonsrate på 79% og litt diskontinuetet ved $z = 1.96$, men det er også en god del null-resultater.

Ingvild Saksvik-Lehouillier

[Figur 13 about here.]

Ingvild Saksvik-Lehouillier sin forskning vil replikeres i 84% av tilfellene i følge modellen. Skal sies at det er en viss diskontinuetet ved $z = 1.96$, noe som gjør at det kan være noe p-hacking eller publiseringsbias.

Jolene Van der Kaap-Deeder

[Figur 14 about here.]

Jolene Van der Kaap-Deeder har en veldig gausskurve-form på funnene hennes på høyresiden av $z = 1.96$, men denne stopper helt opp. Dette kan tyde på en del manglende null-resultater, men det er også en del nullresultater i grafen hennes. Den forventede replikasjonsraten på 64% er noe lavt i forhold til andre her.

Lars Wichstrøm

[Figur 15 about here.]

Lars Wichstrøm har en helt OK z -verdi-graf. Ingen store diskontinuiteter og en replikasjonsrate på 71%. Har med en grei andel nullresultater.

Leif Edward Ottesen Kennair

[Figur 16 about here.]

Leif Edward Ottesen Kennair har en viss diskontinuitet ved $z = 1.96$, men den er ikke så stor. Likevel, hvis man ser på trenden på høyresiden av $z = 1.96$ så kan man tro at trenden tilsier at det er en del manglende nullresultater her. Forventet replikasjonsrate er på relativt greie 74%, såvidt under (men ikke signifikant forskjellig fra) hans store forbilde David M. Buss på 76% (Schimmack, 2022b).

Magne Arve Flaten

[Figur 17 about here.]

Magne Arve Flaten har en fin gausskurve i sin z -verdi-graf, bare at den er sentrert rundt $z \approx 2.3$ i stedet for $z = 0$ som ved nullhypotesen. Diskontinuiteten ved $z = 1.96$ er ikke så stor at noe nødvendigvis er galt her. I praksis er det to muligheter her; enten så er han veldig god på styrkeanalyse eller så er det noen manglende nullresultater her. Forventet replikasjonsrate er på 58%, noe som ganske lavt i forhold til kollegaer.

Magnus Alm

[Figur 18 about here.]

Magnus Alm har en helt OK z -verdi-graf. Noe få z -verdier her noe som gjør det vanskelig å si noe sikkert. Ingen store diskontinuiteter, og en ganske grei forventet replikasjonsrate på 67%. En slags kontur av en gausskurve kan skimtes.

Matthias Mittner

[Figur 19 about here.]

Matthias Mittner har en ganske flat z -verdi-graf. Ingen store diskontinuiteter, men kan se ut som nivået på nullresultater er noe lavere enn signifikante resultater. Mulig noen manglende nullresultater. Forventet replikasjonsrate er på 77% som er ganske bra.

Mons Bendixen

[Figur 20 about here.]

Mons Bendixen har noe diskontinuitet ved $z = 1.96$. På høyresiden av $z = 1.96$ er det konturene til en gausskurve, men den forsvinner på venstresiden, noe som kan tyde på noen manglende nullresultater her. Forventet replikasjonsrate på 78% må sies å være ganske bra.

Nicholas Hagen Kirkerud

[Figur 21 about here.]

Nicholas Haugen Kirkerud har en ganske flat z -verdi-graf, men ganske få z -verdier her, så det er vanskelig å si noe sikkert. Forventet replikasjonsrate på 90% er veldig bra!

Nunne Englund

[Figur 22 about here.]

Nunne Englund har også en ganske flat z -verdi graf med få z -verdier, noe som skaper noe usikkerhet her. Skal sies at forventet replikasjonsrate er på veldig solide 89%!

Odin Hjemdal

[Figur 23 about here.]

Odin Hjemdal har en grei replikasjonsrate på 79%. Ingen store diskontinueteter og grafen er relativt flat. Det skal sies at nivået på venstresiden av $z = 1.96$ er noe lavere enn på høyresiden, noe som kan tyde på noen manglende nullresultater.

Patrick A. Vogel

[Figur 24 about here.]

Patrick A. Vogel har ingen diskontinuitet ved $z = 1.96$ og har en liten tendens til en gausskurve. Forventet replikasjonsrate på 64% er noe lavt i forhold til andre her, men også lite som tyder på p -hacking/publiseringsbias.

Roger Hagen

[Figur 25 about here.]

Roger Hagen har noe diskontinuitet ved $z = 1.96$ som tyder på noen manglende nullresultater. Forventet replikasjonsgrad på 74% er ganske grei.

Stian Solem

[Figur 26 about here.]

Stian Solem har en ganske god forventet replikasjonsrate på 81% og heller ingen diskontinuiteter i grafen, noe som tyder på liten grad av p -hacking/publiseringsbias. Dog skal det sies at man ser en tendens til en gausskurve, noe som kan få en til å lure på i hvilken grad det er nullhypotesen vi ser grafen til her.

Tilmann von Soest

[Figur 27 about here.]

Tilmann von Soest har en ganske klar diskontinuitet ved $z = 1.96$, der man ser konturene til en gausskurve på høyresiden av $z = 1.96$, men denne forsvinner helt på venstresiden. Dette tyder på en del manglende nullresultater. Den forventede replikasjonsgraden på 75% er helt OK.

Timo Juhani Lajunen

[Figur 28 about here.]

Timo Juhani Lajunen har en forventet replikasjonsrate på 71%, noe som er ganske OK. Når det kommer til diskontinuiteter, er diskontinuiteten ganske stor ved $z = 1.96$. Man ser konturene til en gausskurve på høyre side av $z = 1.96$, men ingenting på venstresiden av grafen. Det kan tyde på en del manglende nullresultater, og man kan spørre seg om man bare ser nullhypotesen og publiseringsbias her.

Tor Sunde

[Figur 29 about here.]

Man ser ingen diskontinuitet i grafen til Tor Sunde, noe som tyder på liten grad av p-hacking/publiseringsbias. Forventet replikasjonsrate på 63% er litt lavt, dog.

Torbjørn Rundmo

[Figur 30 about here.]

Torbjørn Rundmo har en av de største diskontinuitetene vi har sett ved $z = 1.96$ (kun slått at Frederikus Van der Weel), med mange manglende nullresultater. Mens 94% av hans publiserte resultater var signifikante, antar modellen at kun 25% av de faktisk utførte forsøkene var signifikante. Det skal dog sies at han har en forventet replikasjonsrate på 81%, noe som er ganske bra.

Tore C Stiles

[Figur 31 about here.]

Tore C Stiles har en litt lav forventet replikasjonsrate på 65%. Det er også noe diskontinuitet ved $z = 1.96$ på den måten at det er en trendendring der. Det kan være at det skyldes manglende resultater, men kan også være at det skyldes god styrkeanalyse.

Torun Grøtte

[Figur 32 about here.]

Torun Grøtte har ingen diskontinuiteter i grafen og en god del publiserte nullresultater. På en annen side er den forventede replikasjonsraten på 55% ganske lav i forhold til kollegaer.

Trond Nordfjærn

[Figur 33 about here.]

Trond Nordfjærn har lite diskontinuitet i grafen sin og en ganske god forventet replikasjonsrate på 84%.

Trond Viggo Grøntvedt

[Figur 34 about here.]

Trond Viggo Grøntvedt har en ganske grei forventet replikasjonsrate på 67% og en god del publiserte nullresultater. Likevel er det en viss diskontinuitet ved $z = 1.96$ ved at trenden snur, noe som kan tyde på noen manglende nullresultater.

Ute Gabriel

[Figur 35 about here.]

Ute Gabriel har en relativt flat graf, og ingen synlig diskontinuitet. Forventet replikasjonsrate på 76% er også ganske bra.

Vera Skalicka

[Figur 36 about here.]

Vera Skalicka har ikke så mange resultater så vanskelig å si så mye ut fra grafen. Er en god del nullresultater, så lite tyder på p-hacking eller publikasjonsbias. En forventet replikasjonsrate på 65% er ganske OK.

Xi Chu

[Figur 37 about here.]

Xi Chu har en noe lav forventet replikasjonsrate på 63%. Det er noe diskontinuitet ved $z = 1.96$, selv om det også er publisert en god del nullresultater.

En kommentar om forventet replikasjonsrate

Som vi så i første avsnitt er den empiriske replikasjonsraten i psykologi lav (et sted rundt 50% eller mindre). Likevel har de aller fleste som er nevnt ovenfor her en replikasjonsrate på greit over 50%! Derfor kan det være verdt å merke seg at sammenligning mellom forventet replikasjonsrate og empirisk replikasjonsrate tilsier at forventet replikasjonsrate slik som den er regnet ut her er ofte høyere enn replikasjonsraten i praksis, så det er fort at forskerne over vil replikeres sjeldnere enn det statistikken over tilsier (Bartoš & Schimmack, 2022; Schimmack, 2021). Schimmack (2022a) bruker et gjennomsnitt mellom ERR og EDR kalt «Actual Replicability Prediction» (ARP) som sin beste prediksjon på utfall av replikasjonsstudier. Som man kan se fra grafene over vil et ARP være noe lavere enn ERR som er brukt som forventet replikasjonsrate i denne artikkelen.

Så hvem vant?

Det var mange her som har gjort en god innsats i å skulle bli instituttets beste p-hacker. Likevel var det ingen som hadde en så god diskontinuetet på $z = 1.96$ som Frederikus Van der Weel. Han vant dermed konkurransen om å være instituttets beste p-hacker. Gratulerer!

Mangler vi din yndlingsforsker?

Mangler vi din yndlingsforsker (deg selv?) i denne artikkelen? Det er i såfall fordi vi ikke hadde nok p-verdier til å kunne få noe nyttig ut av en slik graf.

Disclaimer

P-verdiene ble hentet ut ved hjelp av et dataprogram, det fører til at noe av p-verdiene fra artikler mangler siden programmet ikke klarer å hente ut alle p-verdiene. Det kan være at noen er hentet ut feil. En diskontinuetet i en z-verdi-graf trenger ikke å bety at man har gjort noe uetisk i forskningen sin eller at man har drevet med p-hacking, men heller at man har større sjanse for å nevne statiske resultater på signifikante resultater, mens nullresultater beskrives kvalitativt og vil dermed ikke bli regnet med her. Den forventede replikasjonsraten er et statistisk verdi utregnet fra p-verdiene, og er ikke sikkert samsvarer med en faktisk replikasjonsrate på studiene. Når forventet replikasjonsrate er nevnt hos forskere i artikkelen er det fokus på den estimerte verdien, men konfidensintervallet er stort. Mange med «gode» forventede replikasjonsrater har også ganske «dårlige» forventede replikasjonsrater i 95% konfidensintervallet sitt og vica versa. Så den kvalitative beskrivelsen av replikasjonsratene i teksten er ofte ikke statistisk signifikant i seg selv. Om noen av forskerene har gitt ut få artikler kan også det føre til at z-verdi-grafen ikke er særlig god.

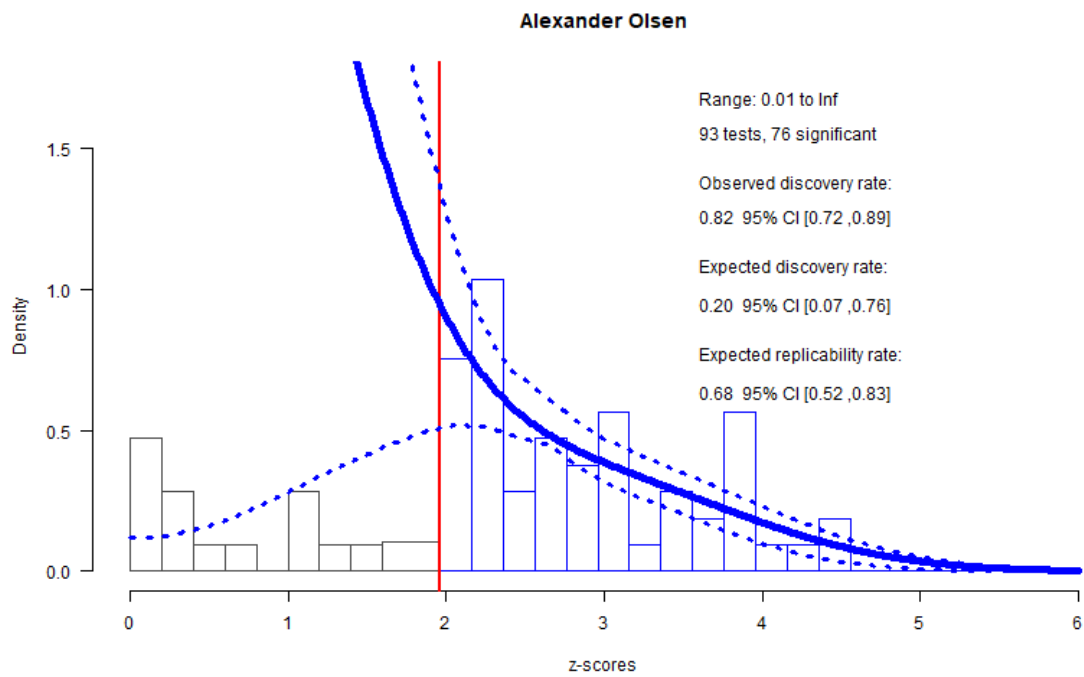
Referanser

- Bartoš, F., & Schimmack, U. (2022). Z-Curve 2.0: Estimating Replication Rates and Discovery Rates. *Meta-Psychology*, 6. <https://doi.org/10.15626/MP.2021.2720>
- Kayed, N. S., & Van der Meer, A. L. H. (2009). A longitudinal study of prospective control in catching by full-term and preterm infants. *Experimental Brain Research*, 194(2), 245–258. <https://doi.org/10.1007/s00221-008-1692-2>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490. <https://doi.org/10.1177/2515245918810225>
- Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Schimmack, U. (2021). Predicting Replication Outcomes: Prediction Markets vs. R-Index. Hentet 22. januar 2023, fra <https://replicationindex.com/2021/05/16/pmvsrindex/>
- Schimmack, U. (2022a). 2022 Replicability Rankings of Psychology Journals. Hentet 22. januar 2023, fra <https://replicationindex.com/2022/01/26/rr21/>

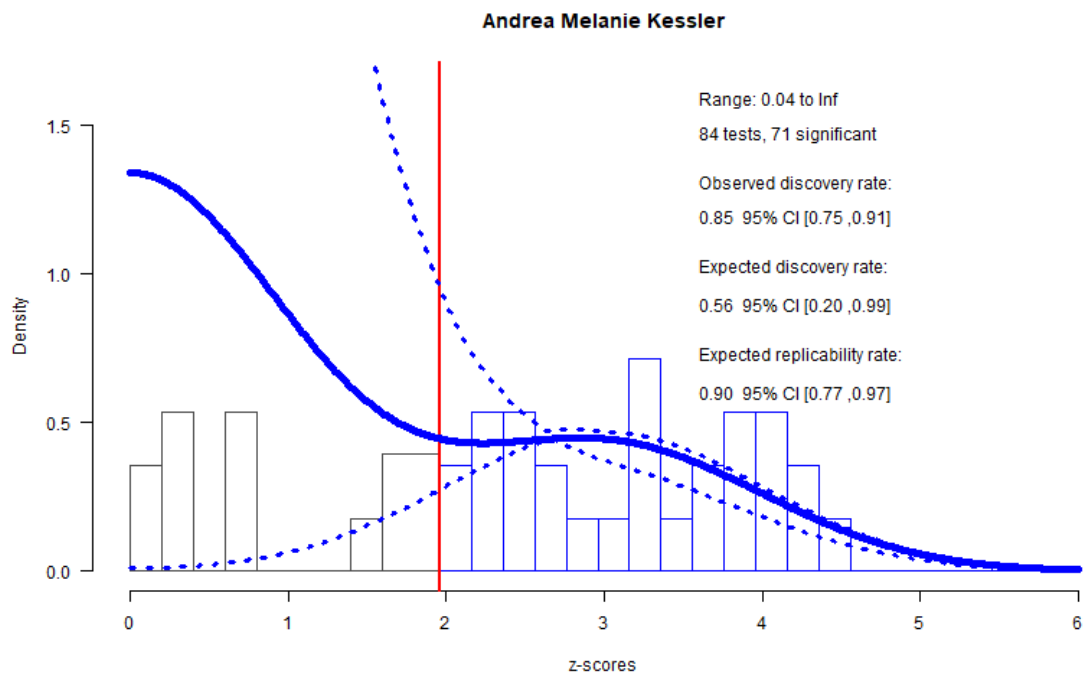
Schimmack, U. (2022b). 2021 Replicability Report for the Psychology Department at U Texas – Austin. Hentet 22. januar 2023, fra <https://replicationindex.com/2022/03/13/rr22-utexas-austin/>

Figurer

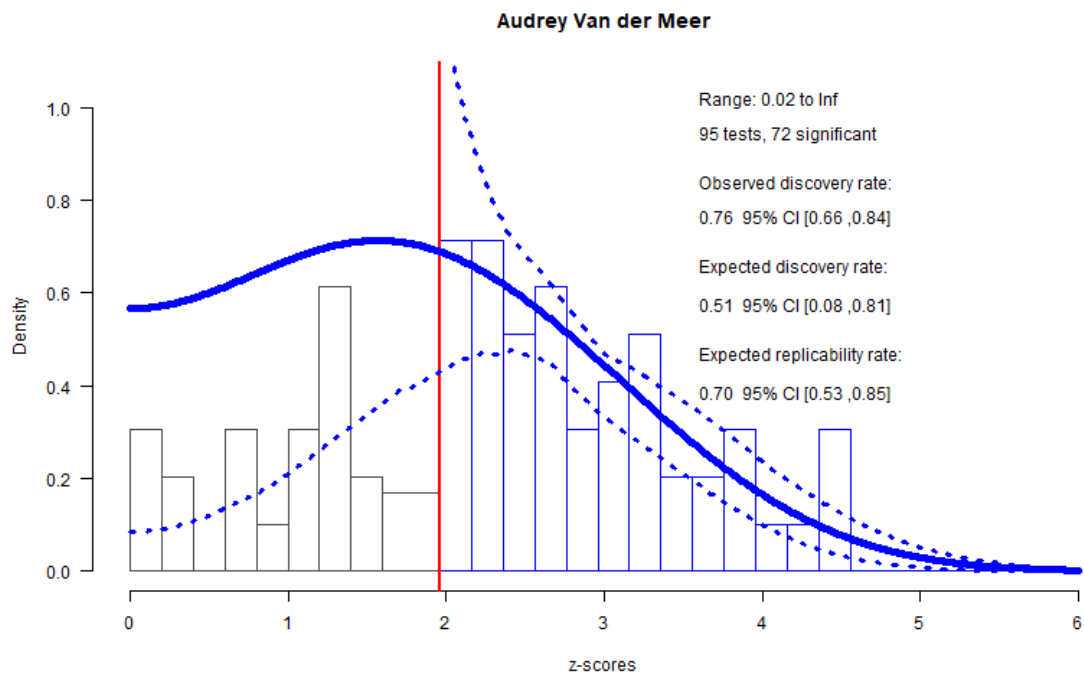
1	z-verdier for Alexander Olsen	11
2	z-verdier for Andrea Melanie Kessler	12
3	z-verdier for Audrey Van der Meer	13
4	z-verdier for Audun Havnen	14
5	z-verdier for Beate Wold Hygen	15
6	z-verdier for Christian Kløckner	16
7	z-verdier for Dawn Marie Behne	17
8	z-verdier for Frederick Anyan	18
9	z-verdier for Frederikus Van der Weel	19
10	z-verdier for Gerit Pfuhl	20
11	z-verdier for Hermundur Sigmundsson	21
12	z-verdier for Ida Emilia Brunner	22
13	z-verdier for Ingvild Saksvik-Lehouillier	23
14	z-verdier for Jolene Van der Kaap-Deeder	24
15	z-verdier for Lars Wichstrøm	25
16	z-verdier for Leif Edward Ottesen Kennair	26
17	z-verdier for Magne Arve Flaten	27
18	z-verdier for Magnus Alm	28
19	z-verdier for Matthias Mittner	29
20	z-verdier for Mons Bendixen	30
21	z-verdier for Nicholas Hagen Kirkerud	31
22	z-verdier for Nunne Englund	32
23	z-verdier for Odin Hjemdal	33
24	z-verdier for Patrick A. Vogel	34
25	z-verdier for Roger Hagen	35
26	z-verdier for Stian Solem	36
27	z-verdier for Tilmann von Soest	37
28	z-verdier for Timo Juhani Lajunen	38
29	z-verdier for Tor Sunde	39
30	z-verdier for Torbjørn Rundmo	40
31	z-verdier for Tore C Stiles	41
32	z-verdier for Torun Grøtte	42
33	z-verdier for Trond Nordfjærn	43
34	z-verdier for Trond Viggo Grøntvedt	44
35	z-verdier for Ute Gabriel	45
36	z-verdier for Vera Skalicka	46
37	z-verdier for Xi Chu	47

**Figur 1**

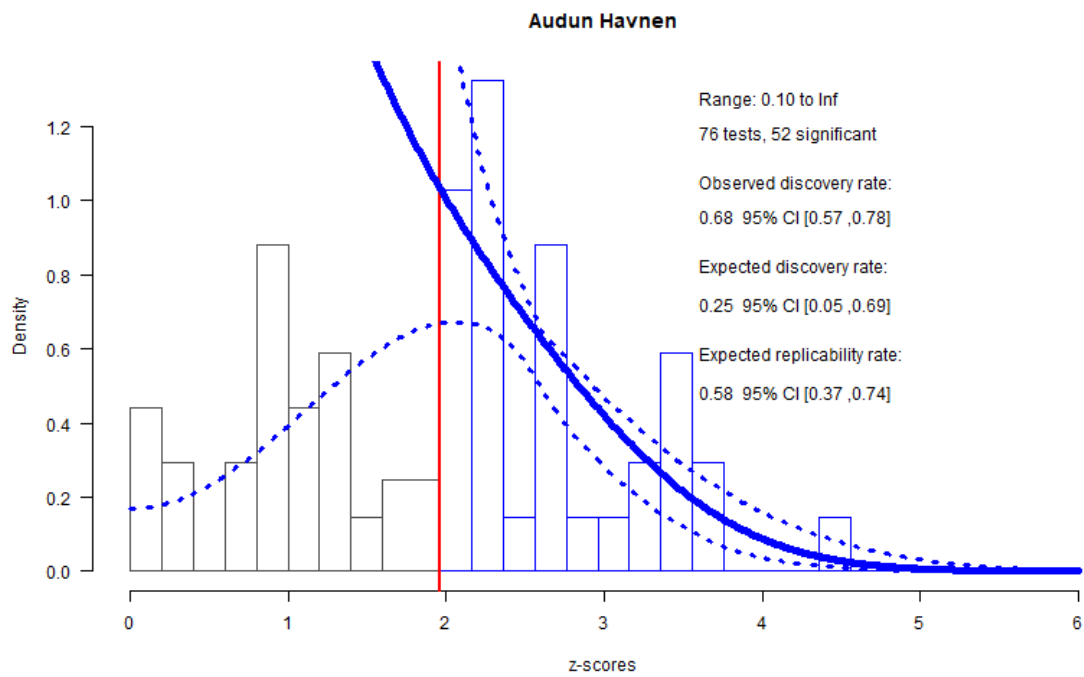
z-verdier for Alexander Olsen

**Figur 2**

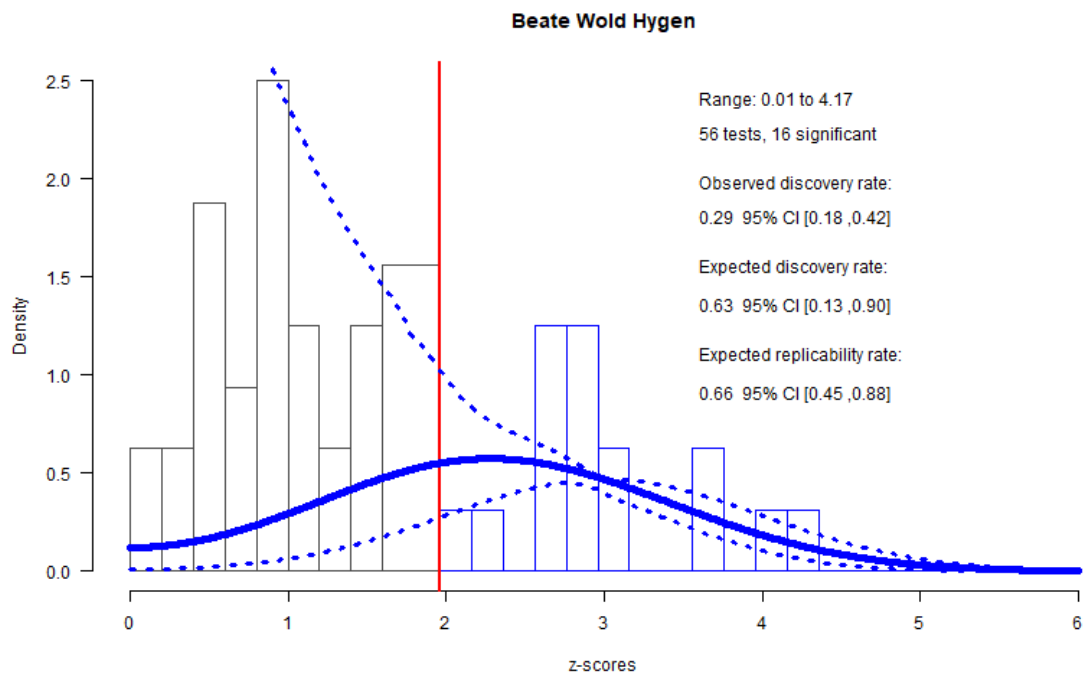
z-verdier for Andrea Melanie Kessler

**Figur 3**

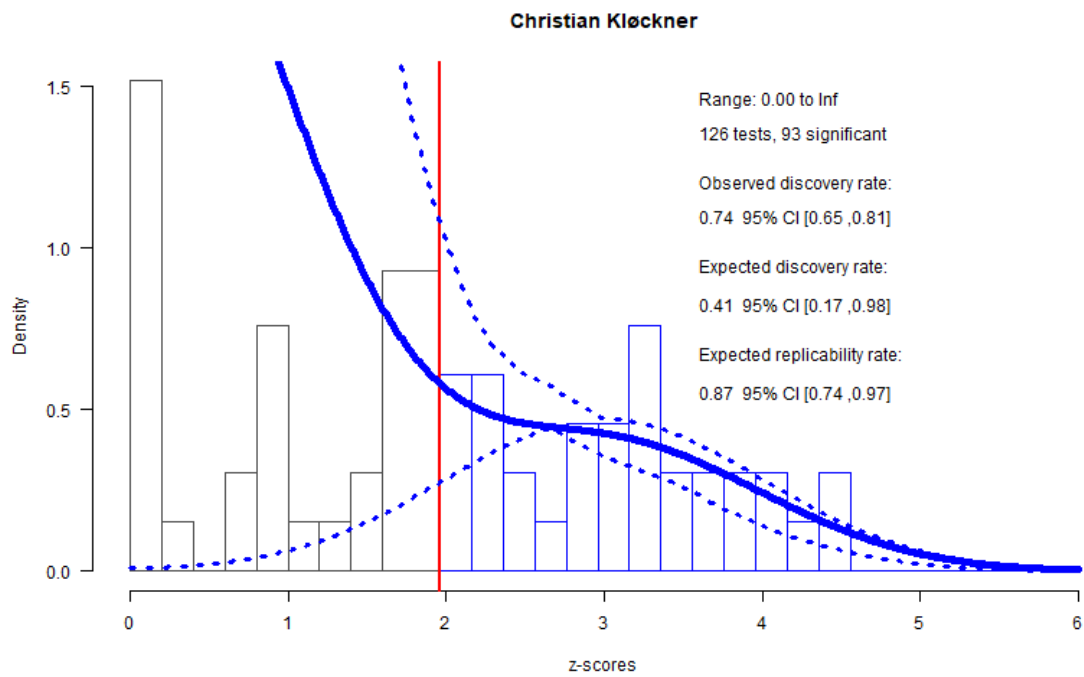
z-verdier for Audrey Van der Meer

**Figur 4**

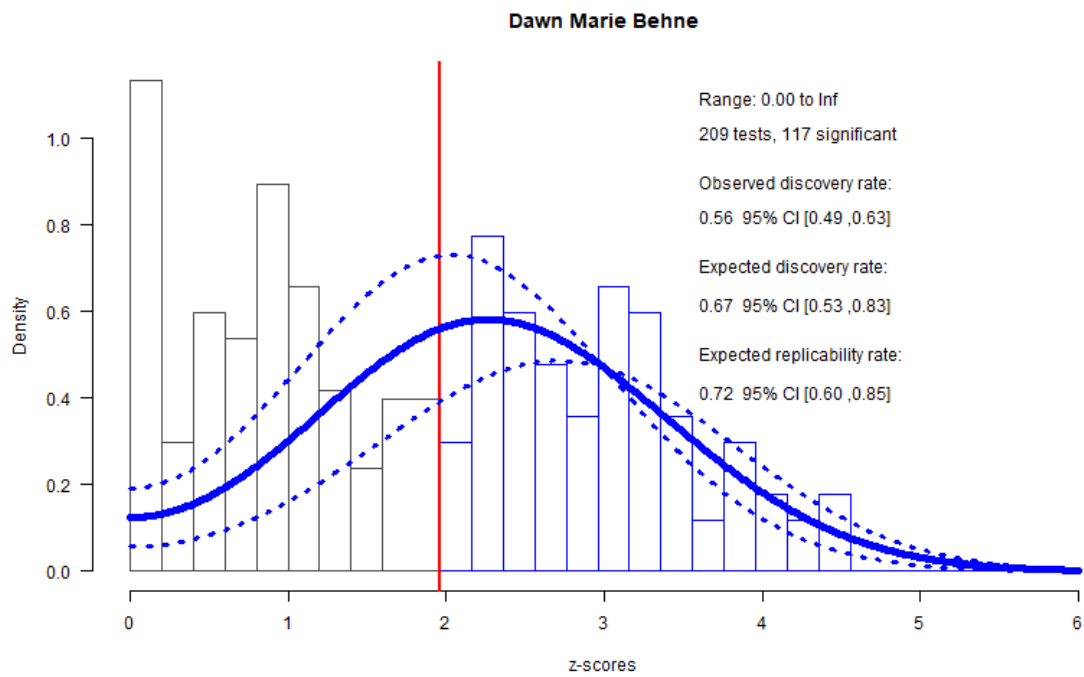
z-verdier for Audun Havnen

**Figur 5**

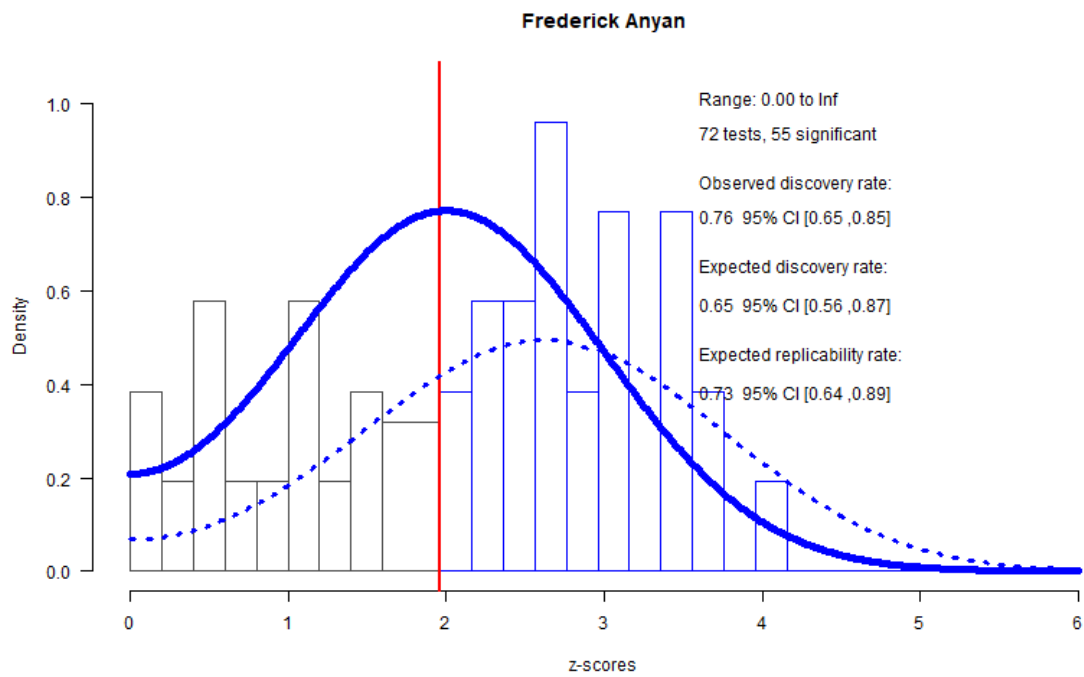
z-verdier for Beate Wold Hygen

**Figur 6**

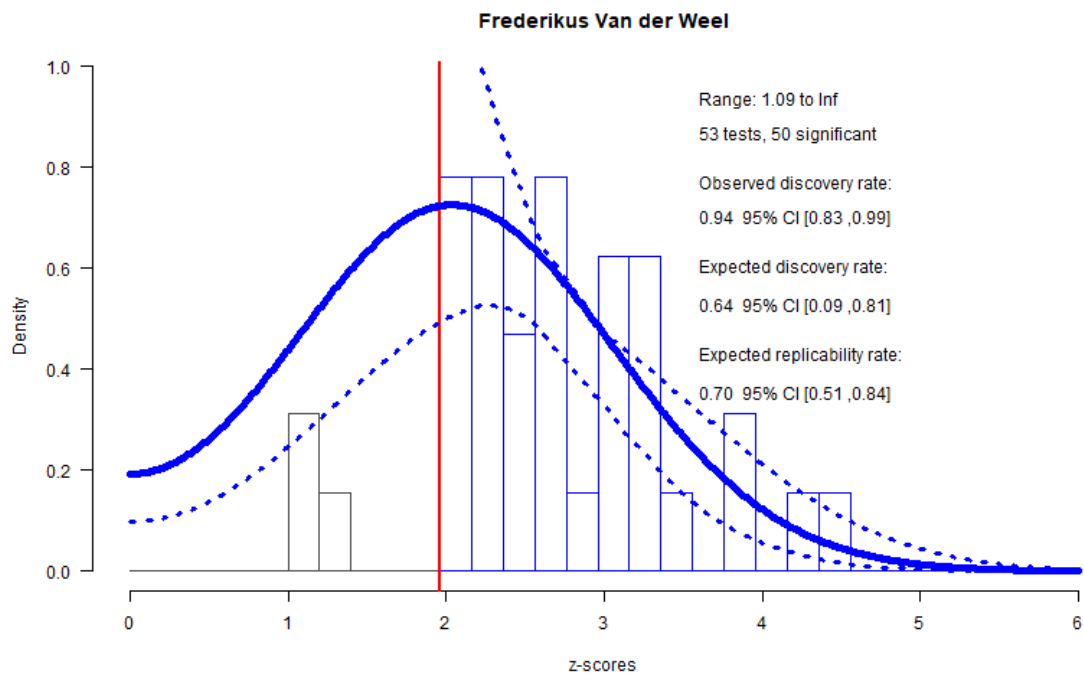
z-verdier for Christian Kløckner

**Figur 7**

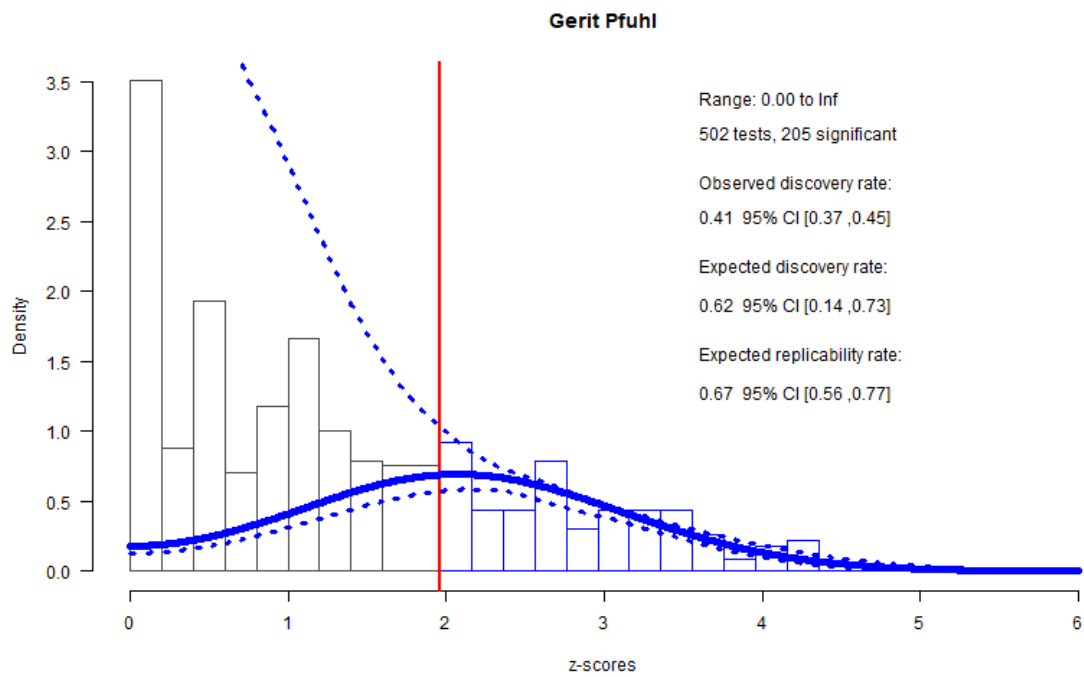
z-verdier for Dawn Marie Behne

**Figur 8**

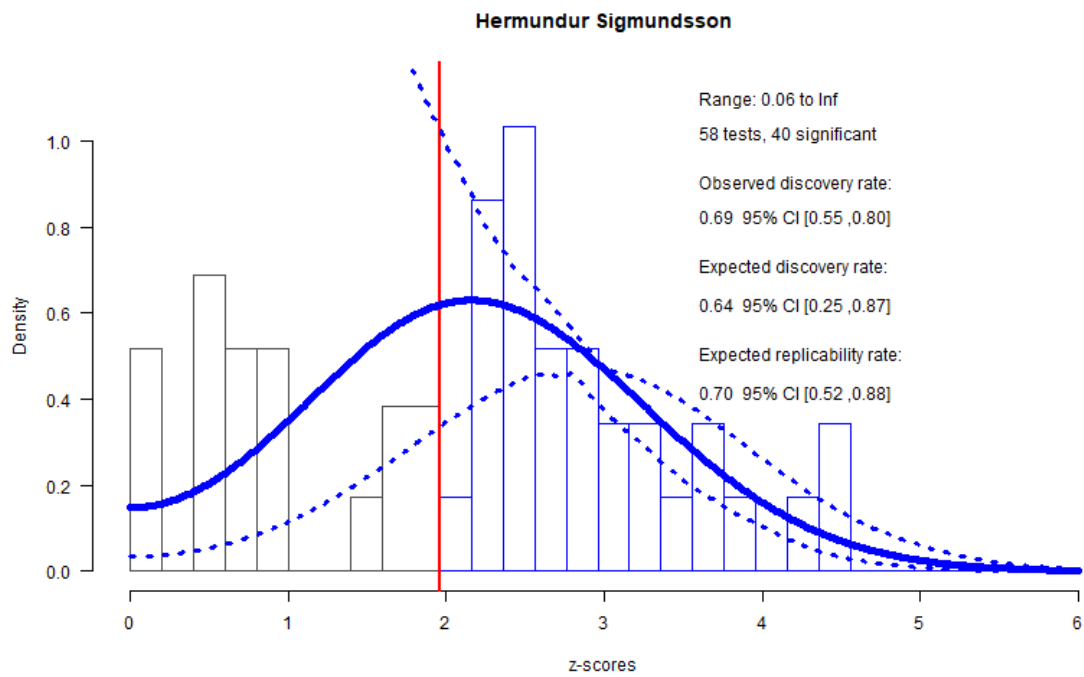
z-verdier for Frederick Anyan

**Figur 9**

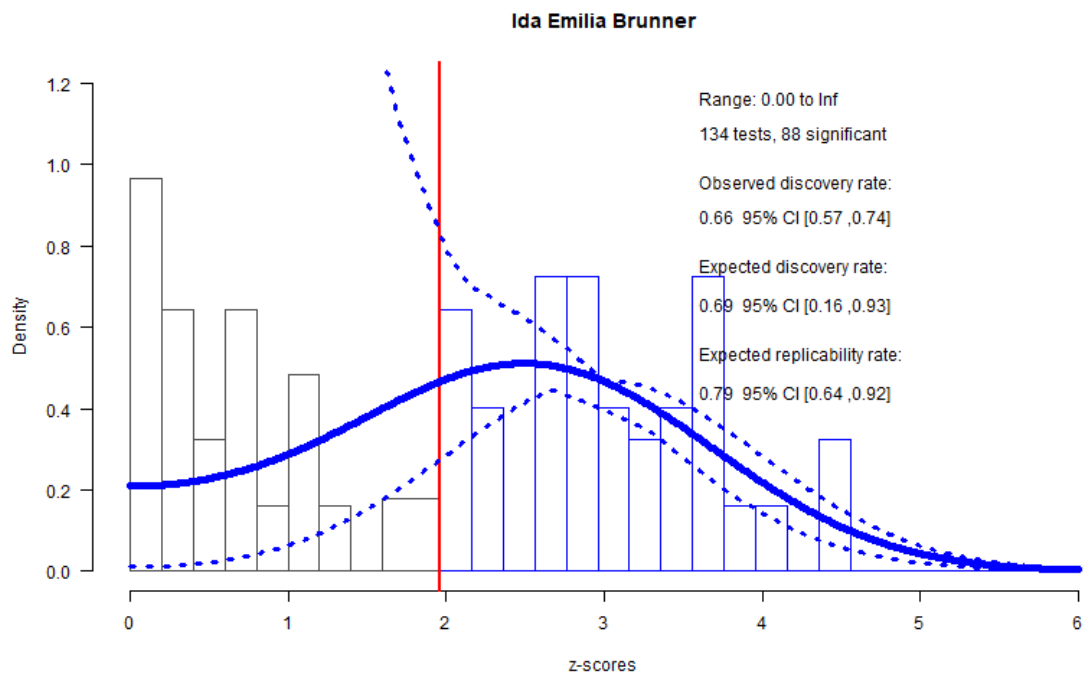
z-verdier for Frederikus Van der Weel

**Figur 10**

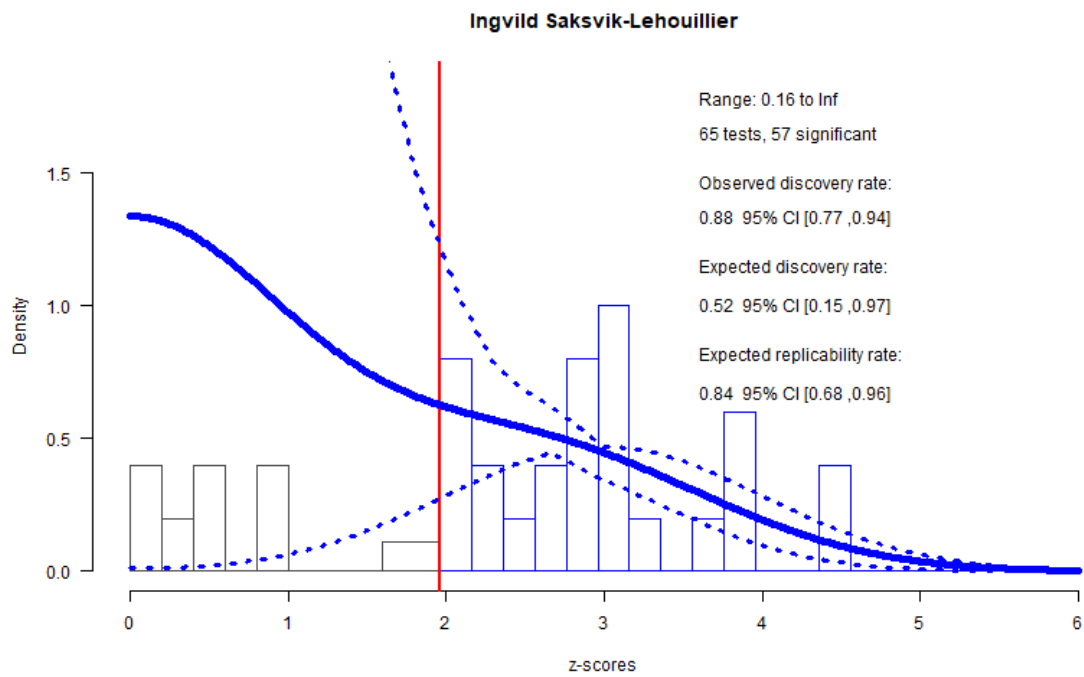
z-verdier for Gerit Pfuhl

**Figur 11**

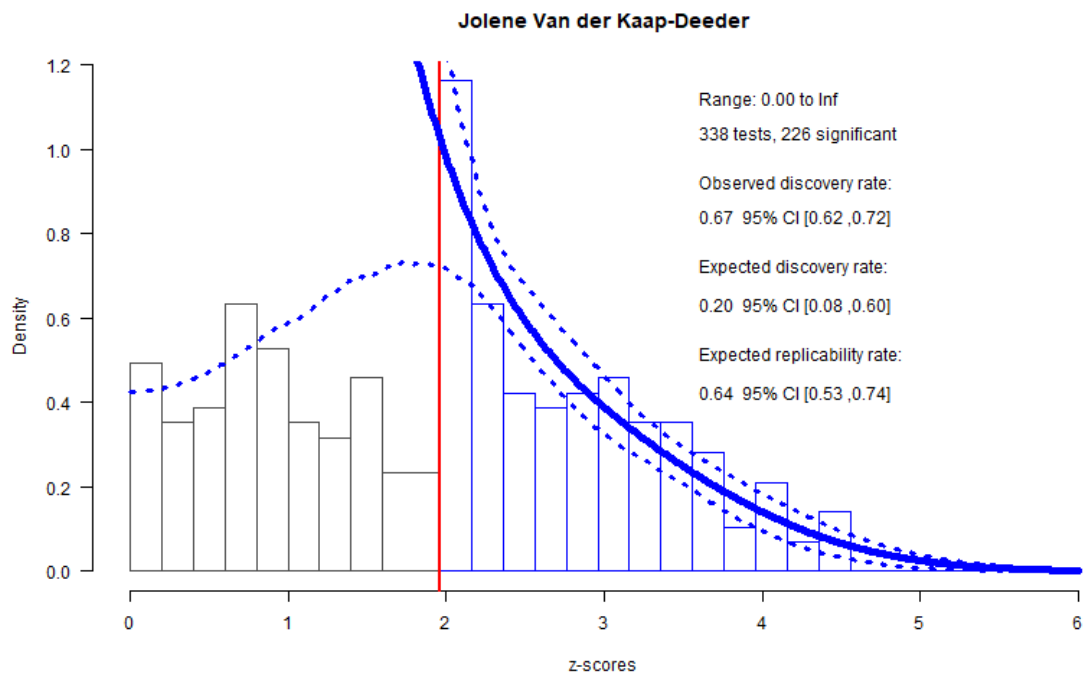
z-verdier for Hermundur Sigmundsson

**Figur 12**

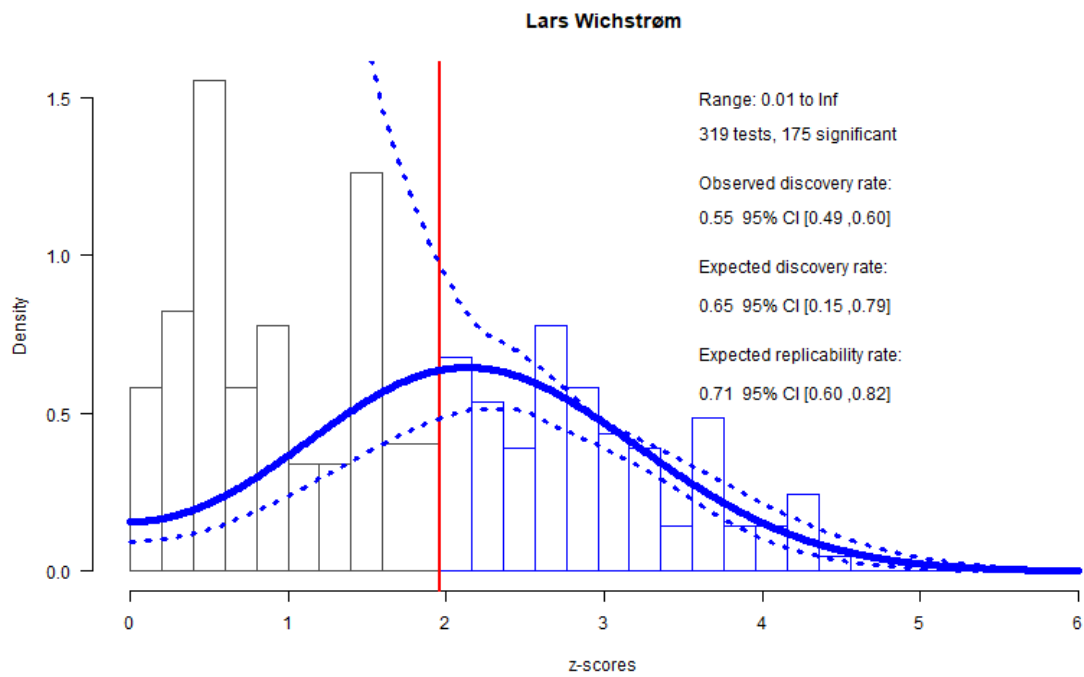
z-verdier for Ida Emilia Brunner

**Figur 13**

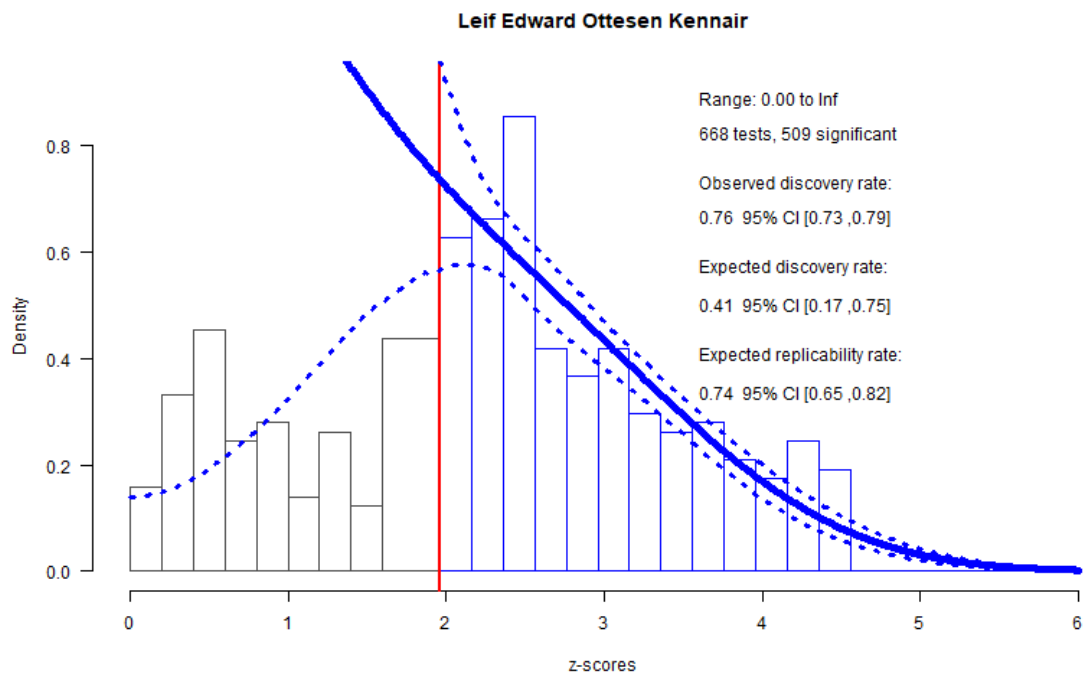
z-verdier for Ingvild Saksvik-Lehouillier

**Figur 14**

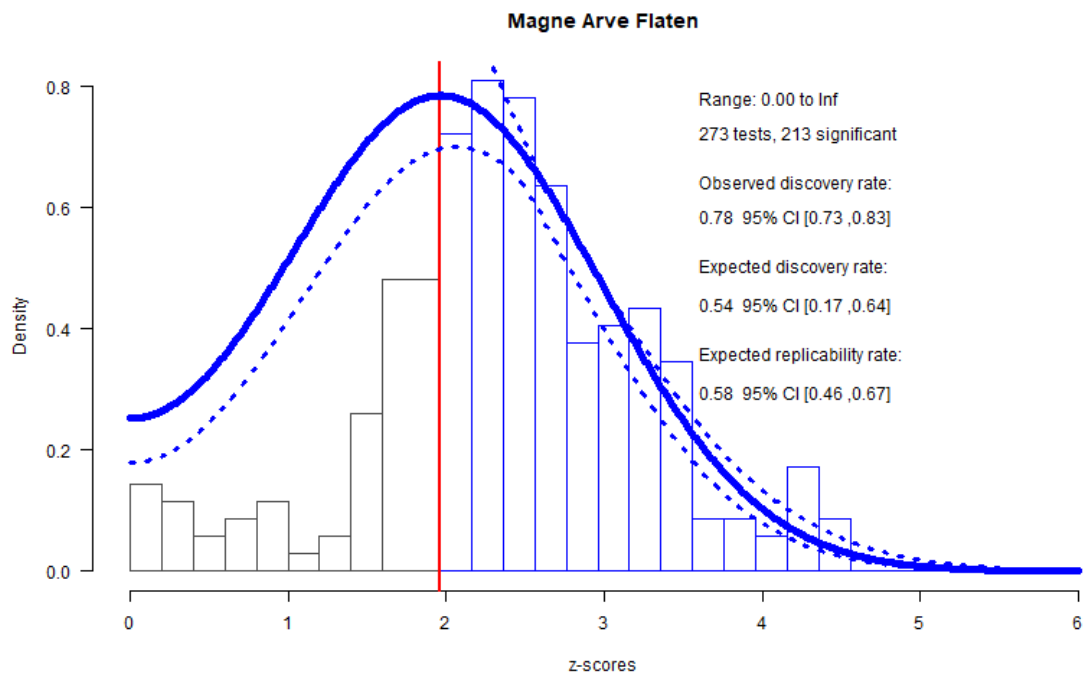
z-verdier for Jolene Van der Kaap-Deeder

**Figur 15**

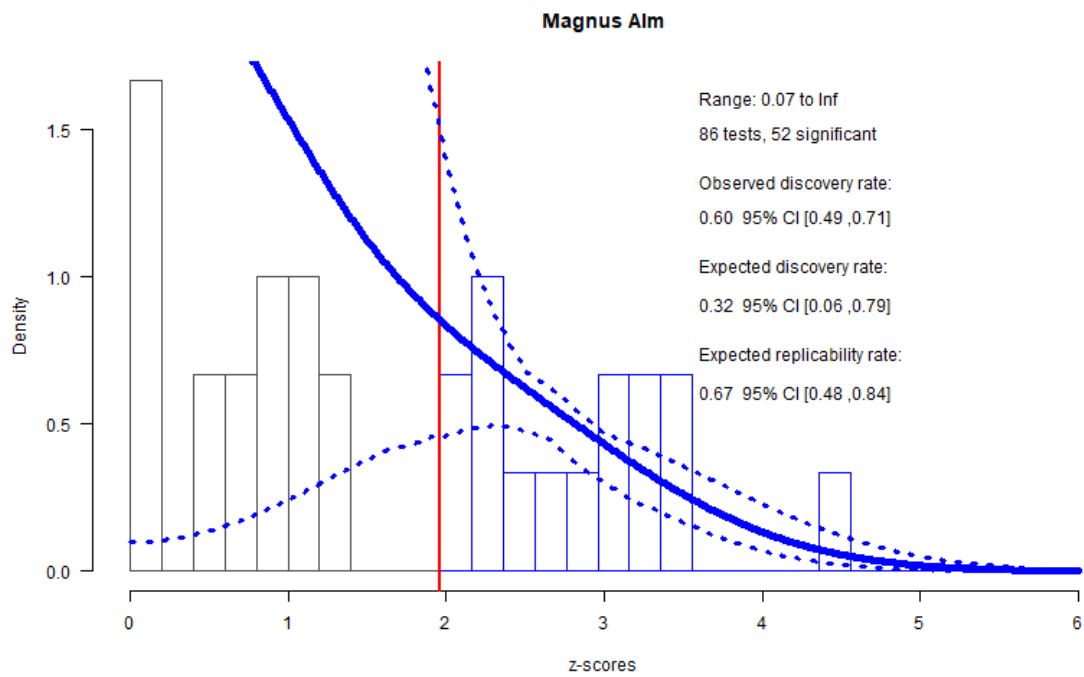
z-verdier for Lars Wichstrøm

**Figur 16**

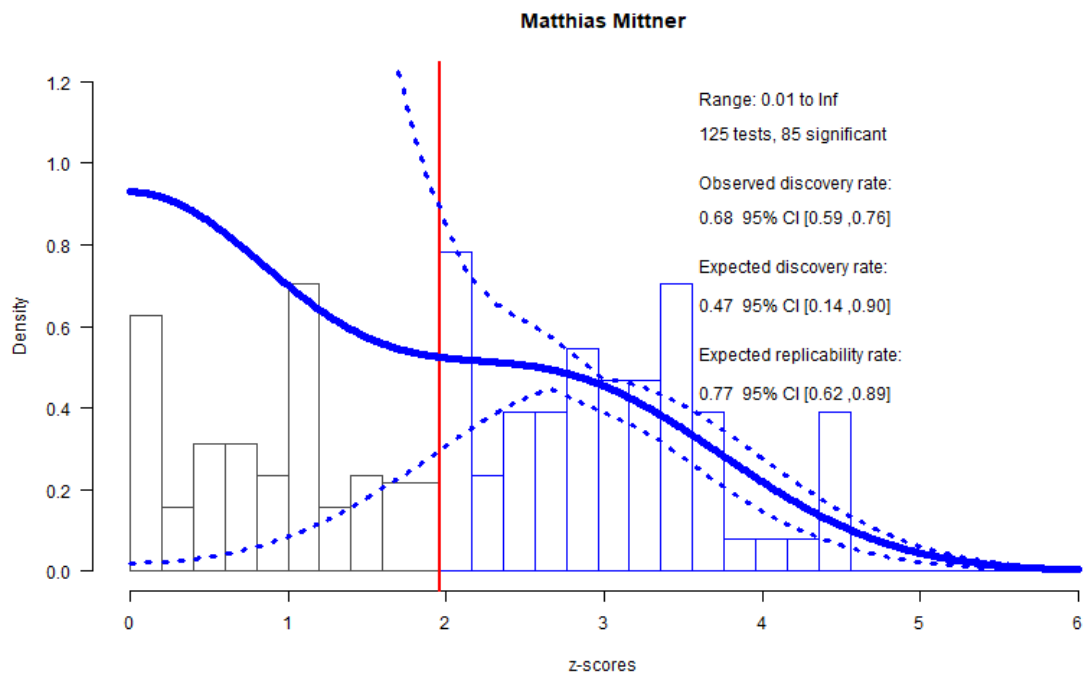
z-verdier for Leif Edward Ottesen Kennair

**Figur 17**

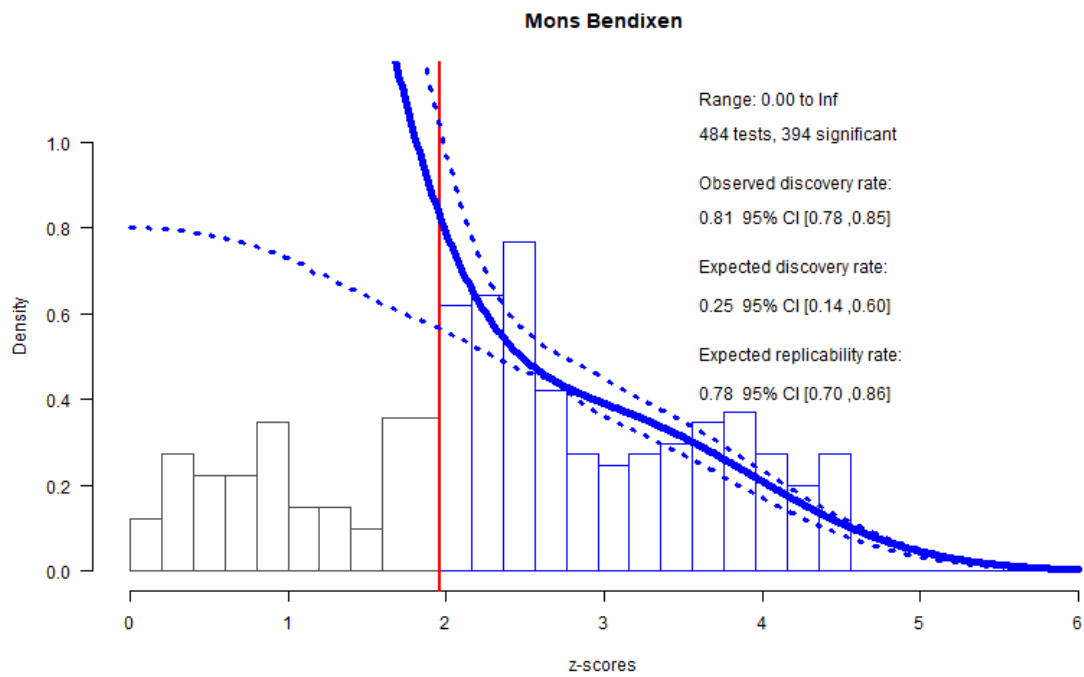
z-verdier for Magne Arve Flaten

**Figur 18**

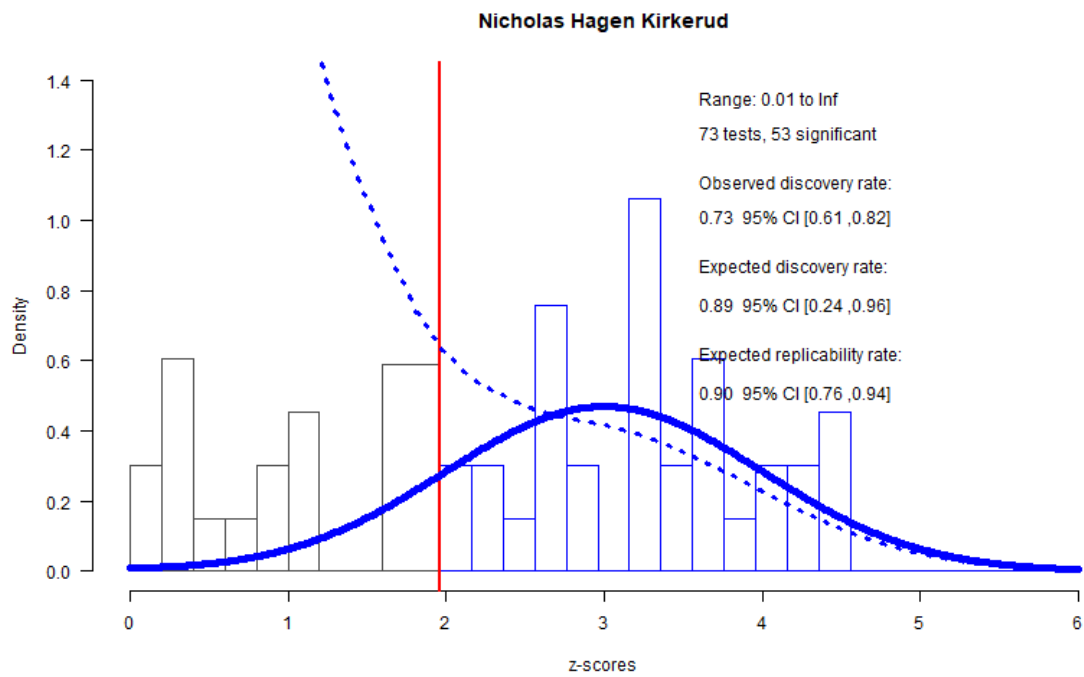
z-verdier for Magnus Alm

**Figur 19**

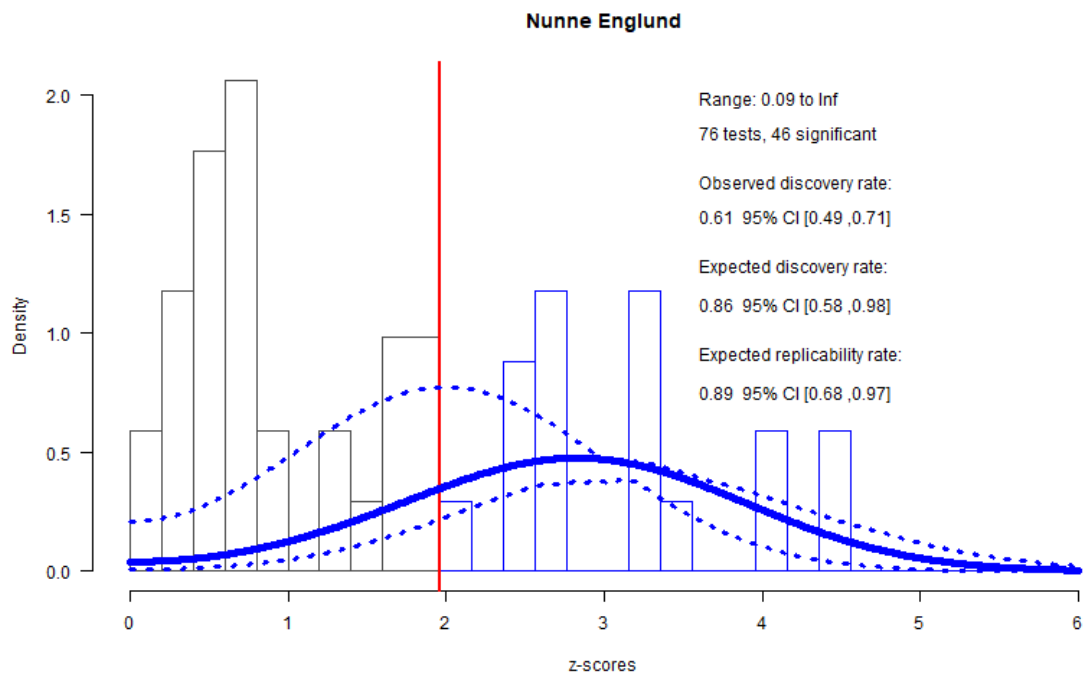
z-verdier for Matthias Mittner

**Figur 20**

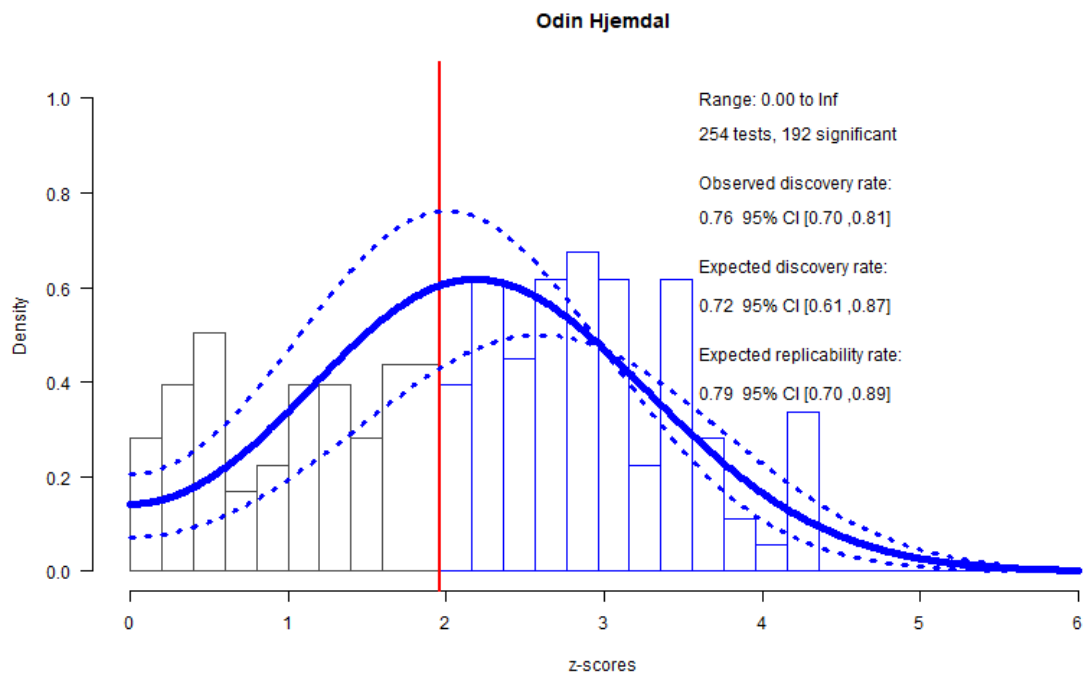
z-verdier for Mons Bendixen

**Figur 21**

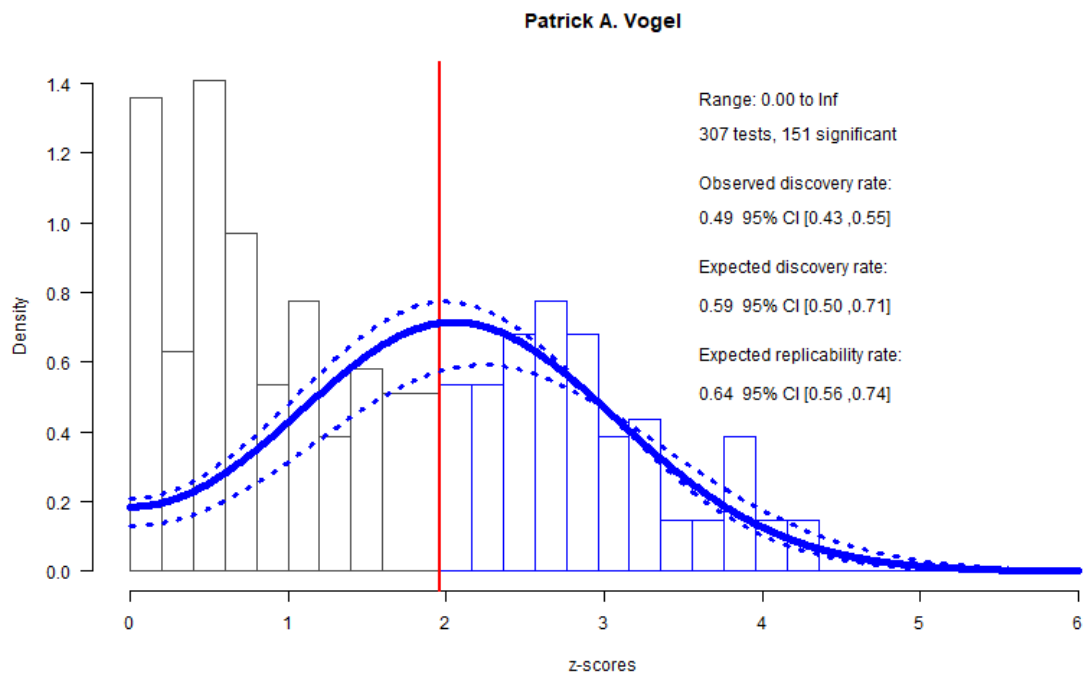
z-verdier for Nicholas Hagen Kirkerud

**Figur 22**

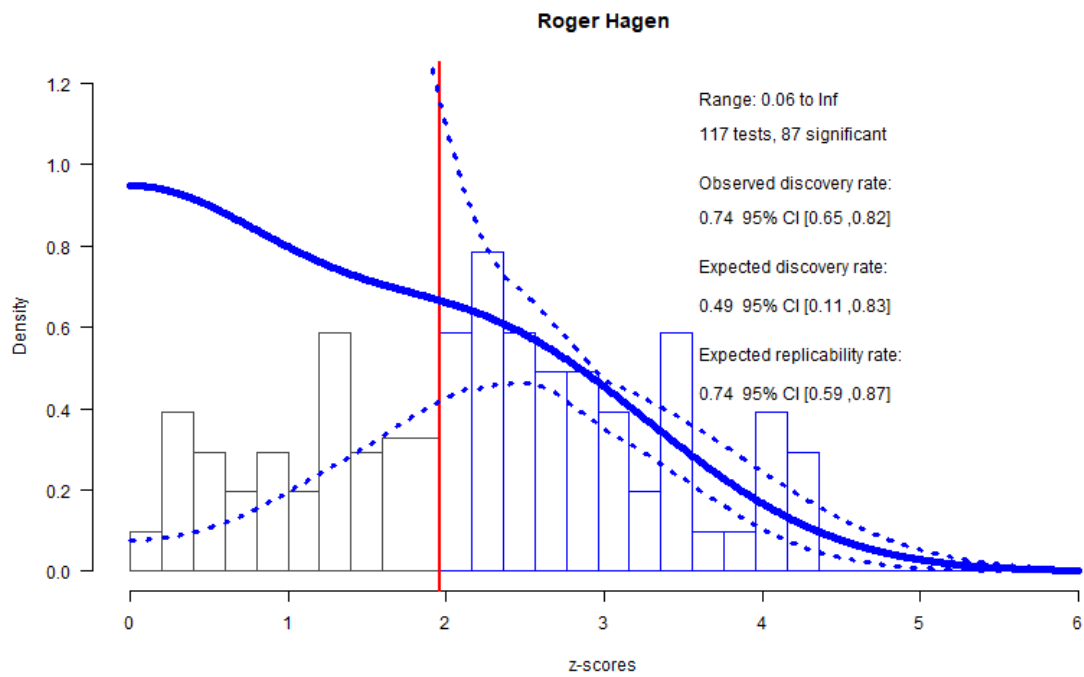
z-verdier for Nunne Englund

**Figur 23**

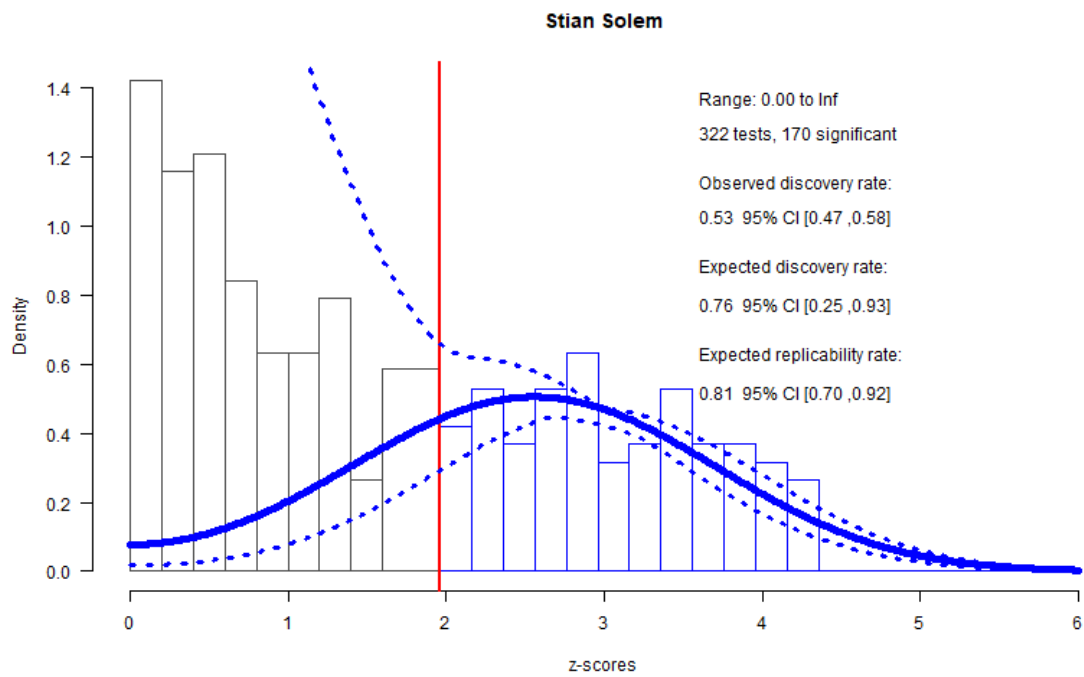
z-verdier for Odin Hjemdal

**Figur 24**

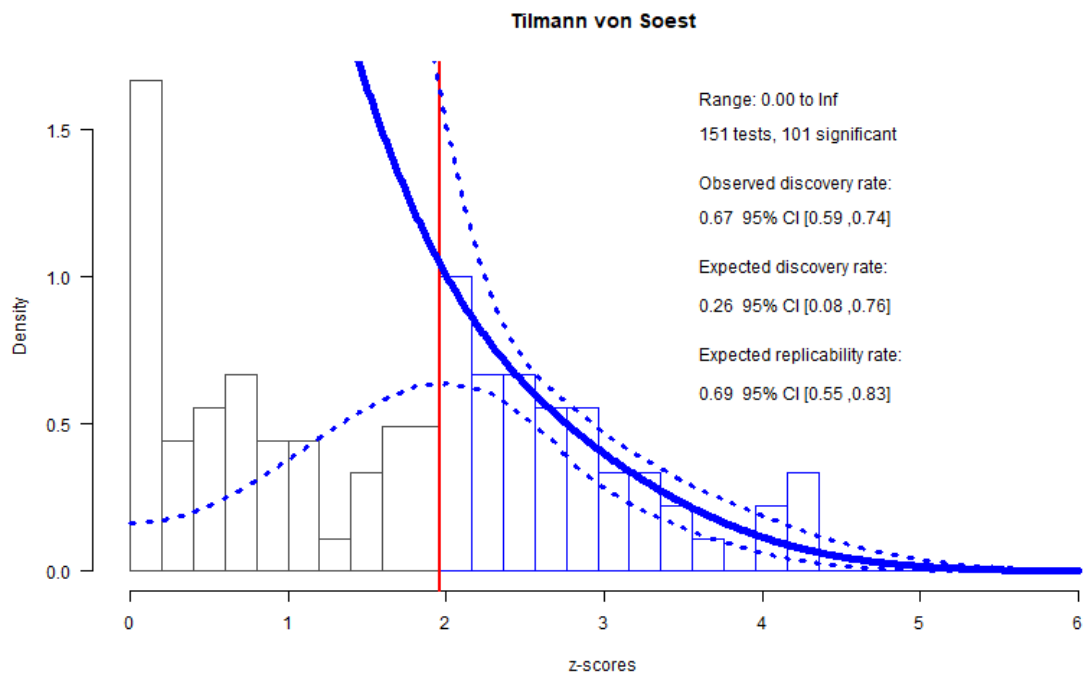
z-verdier for Patrick A. Vogel

**Figur 25**

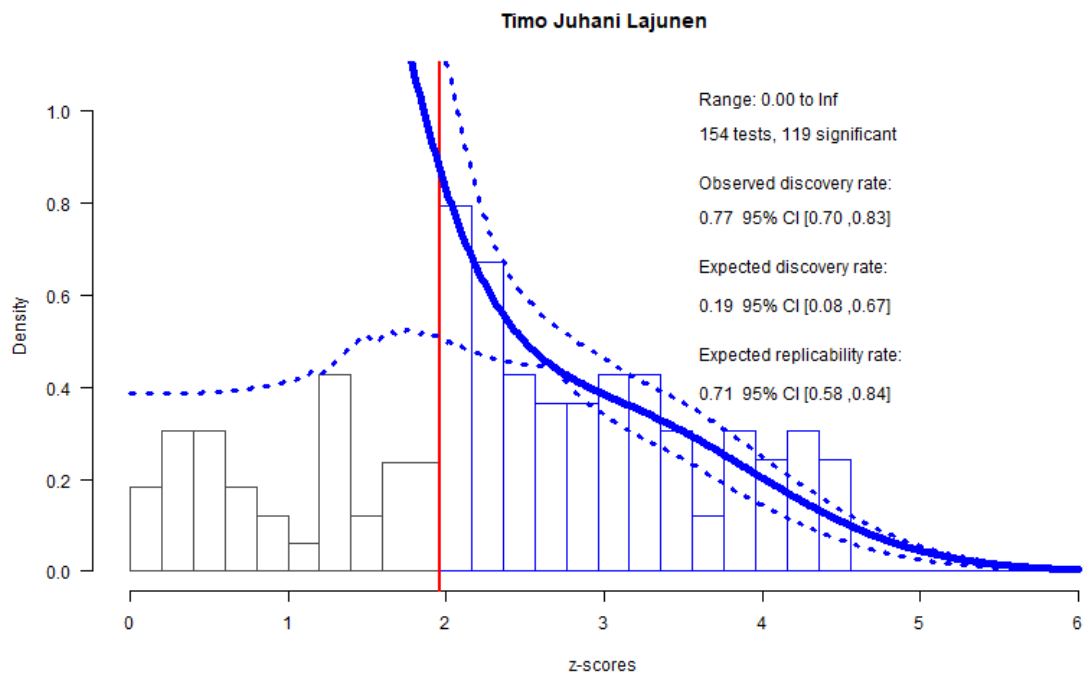
z-verdier for Roger Hagen

**Figur 26**

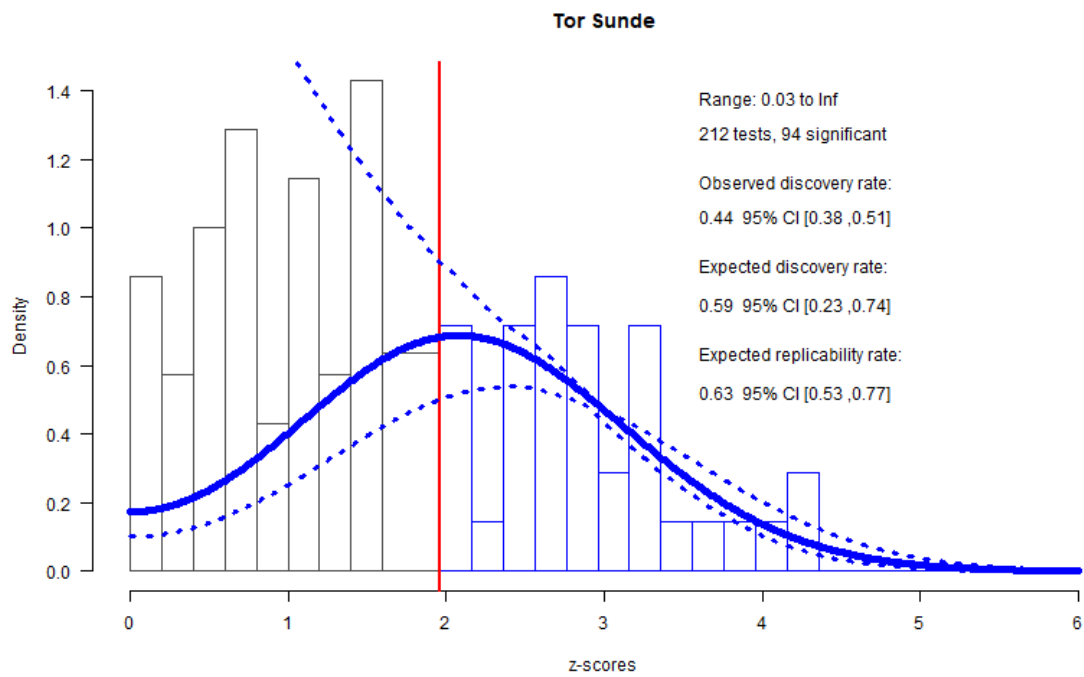
z-verdier for Stian Solem

**Figur 27**

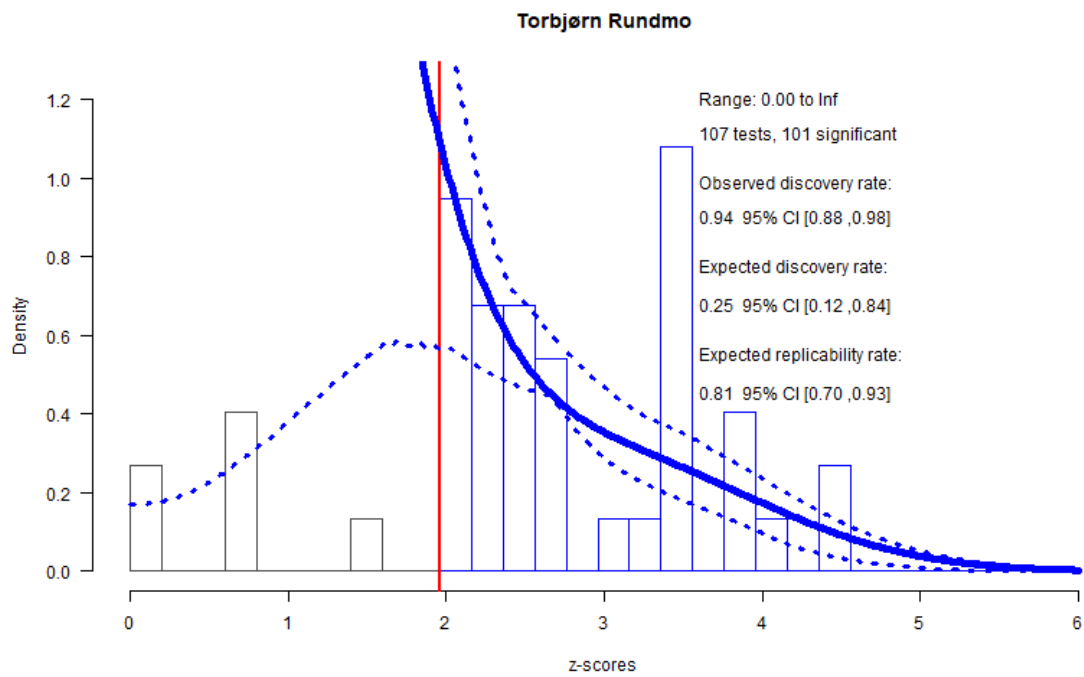
z-verdier for Tilman von Soest

**Figur 28**

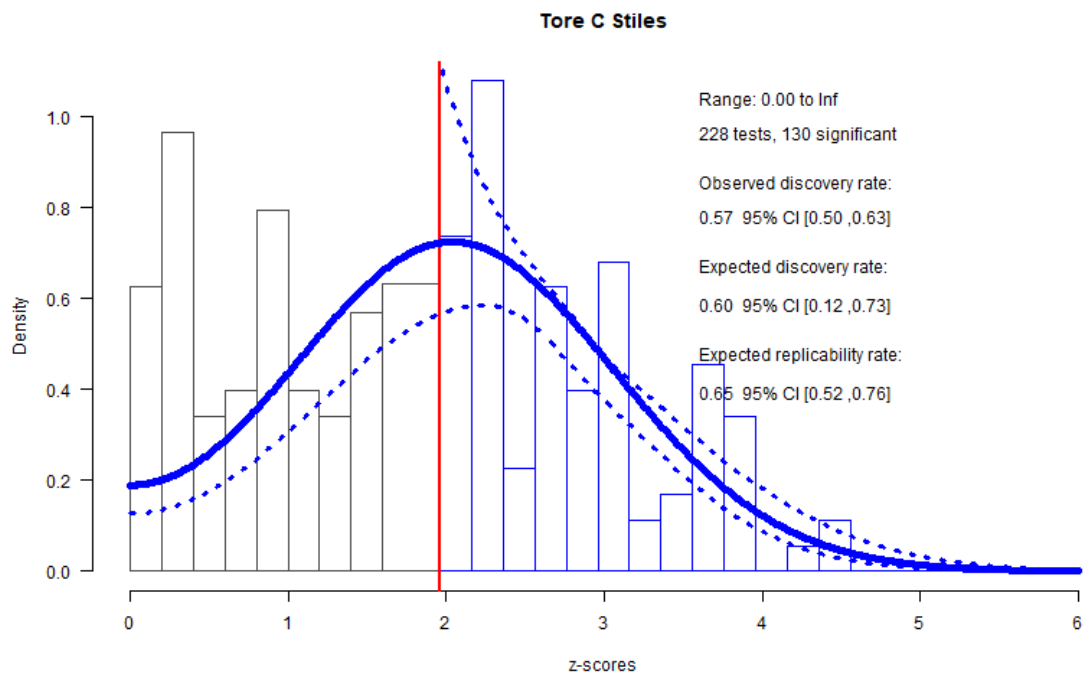
z-verdier for Timo Juhani Lajunen

**Figur 29**

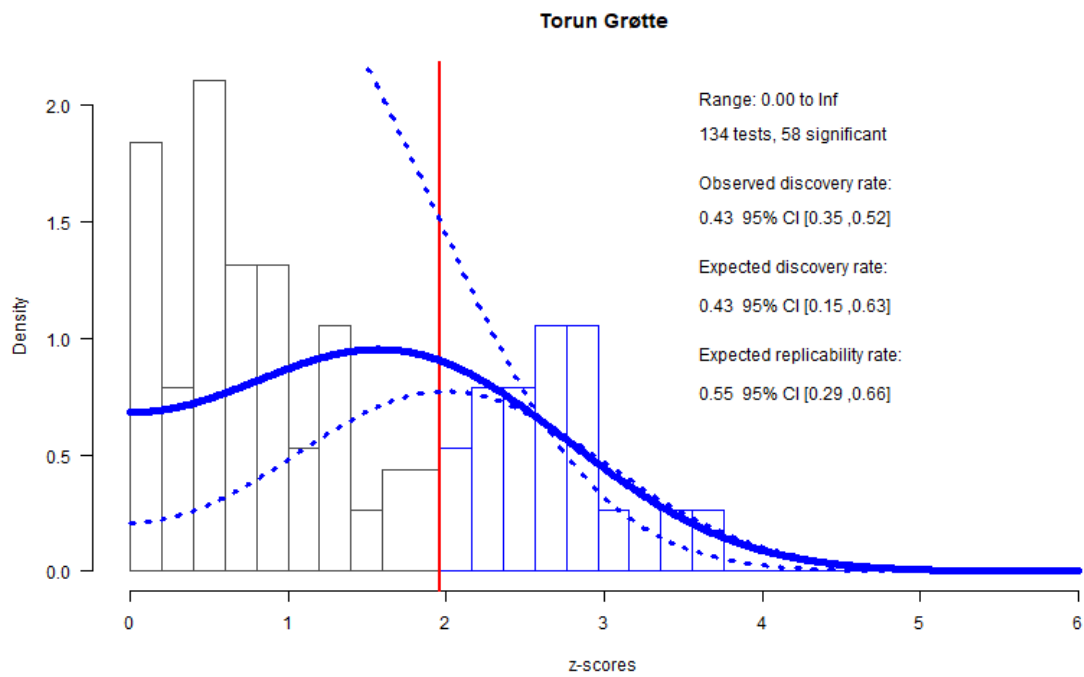
z-verdier for Tor Sunde

**Figur 30**

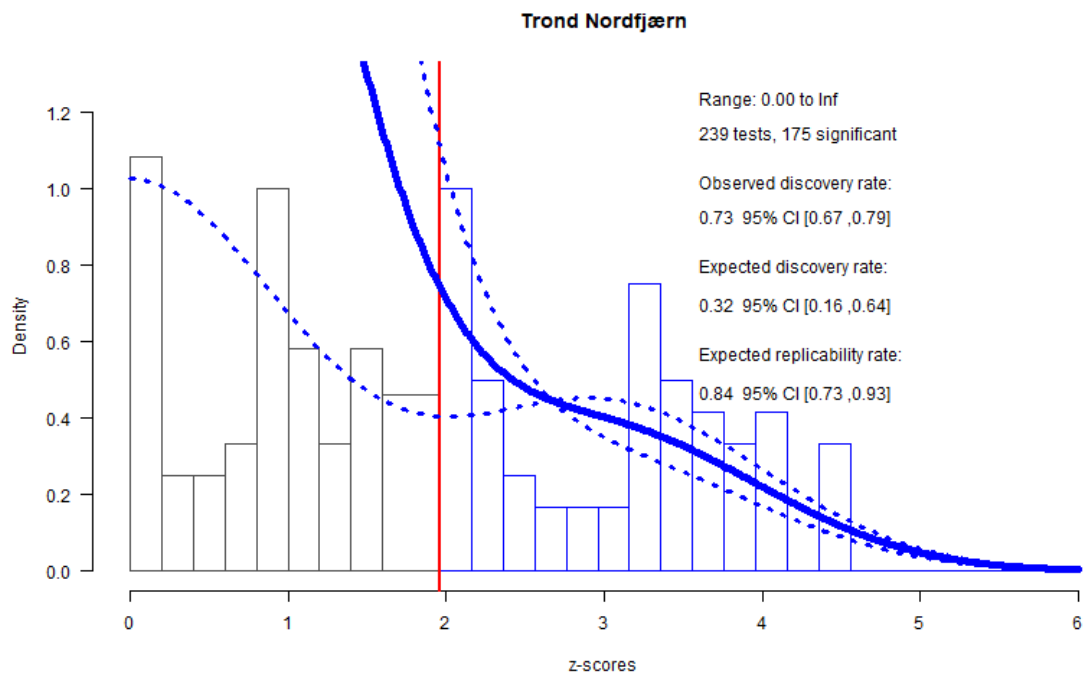
z-verdier for Torbjørn Rundmo

**Figur 31**

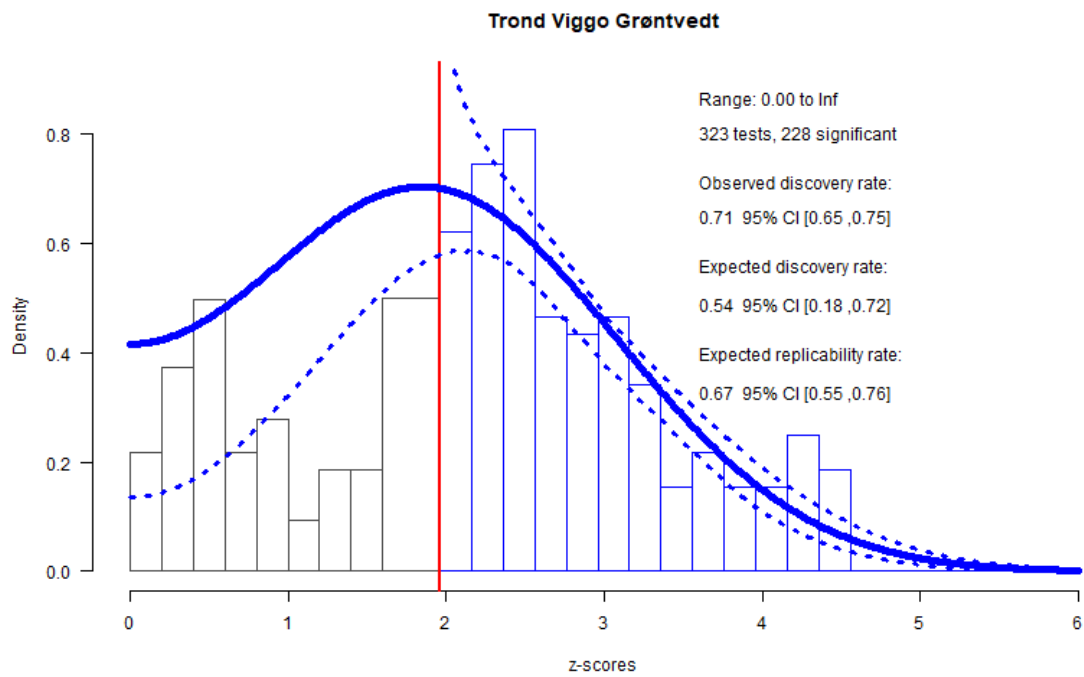
z-verdier for Tore C Stiles

**Figur 32**

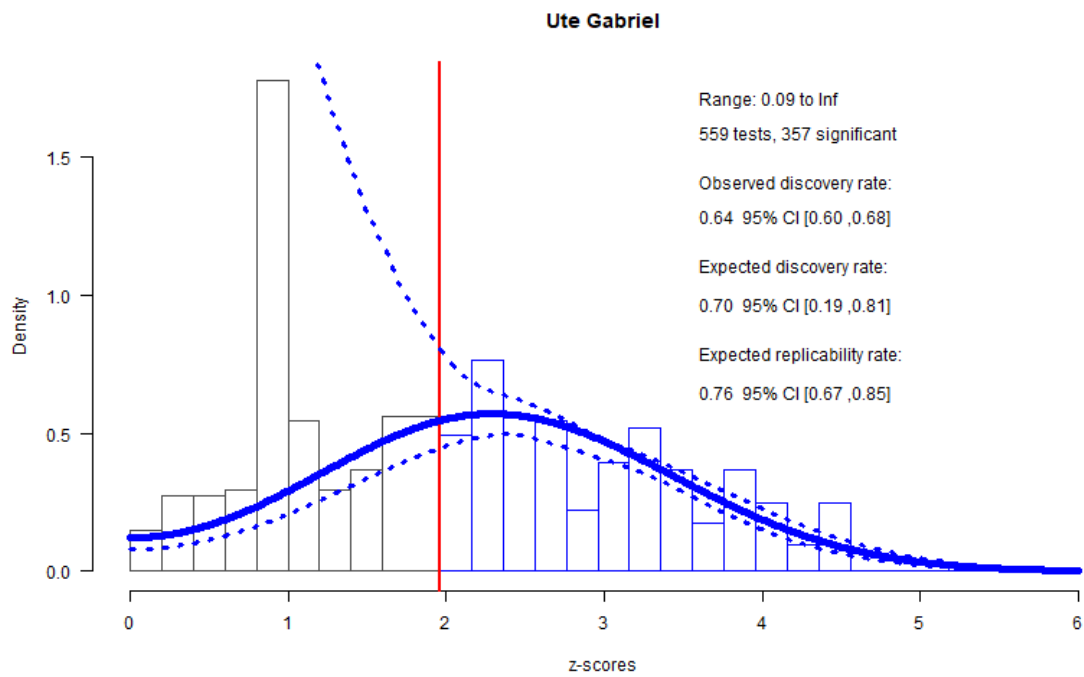
z-verdier for Torun Grøtte

**Figur 33**

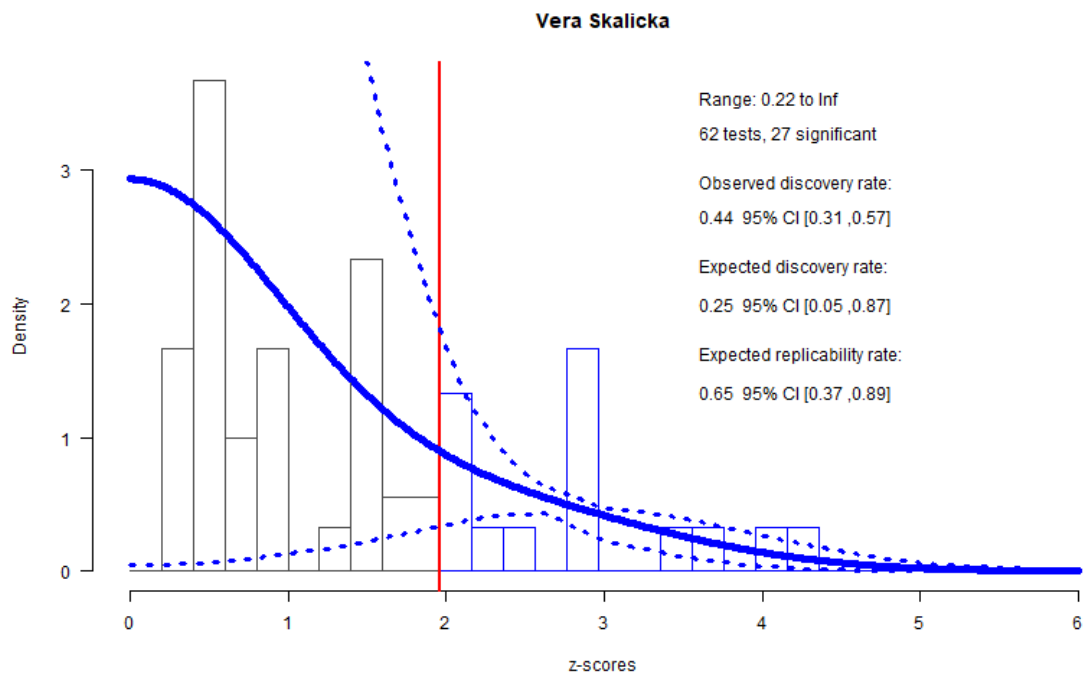
z-verdier for Trond Nordfjærn

**Figur 34**

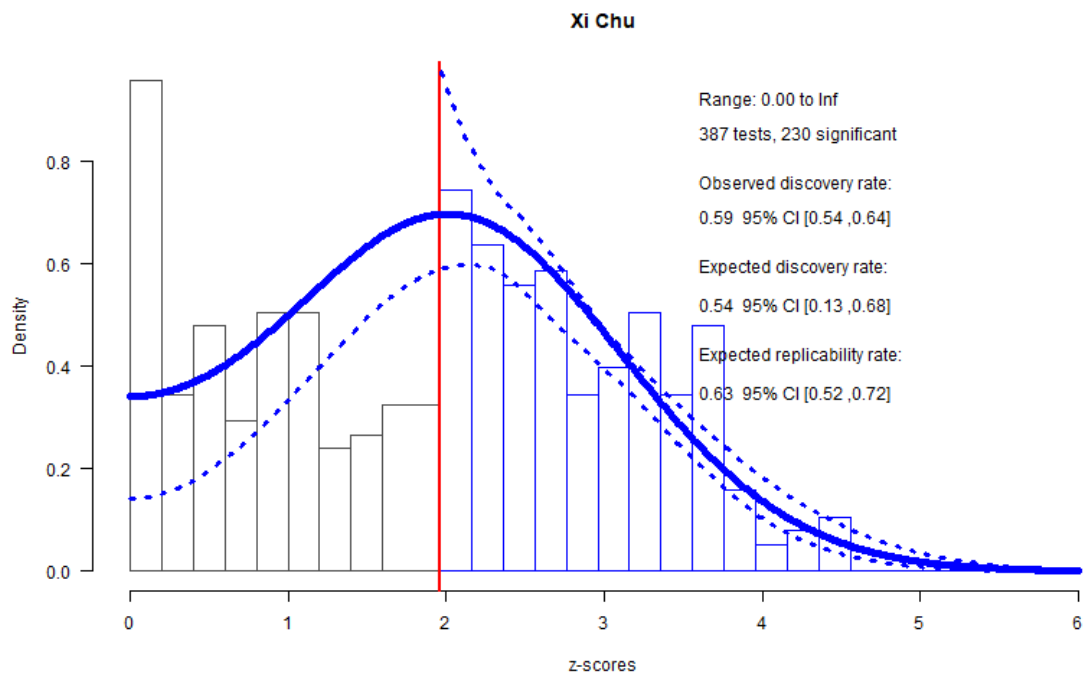
z-verdier for Trond Viggo Grøntvedt

**Figur 35**

z-verdier for Ute Gabriel

**Figur 36**

z-verdier for Vera Skalicka

**Figur 37**

z-verdier for Xi Chu