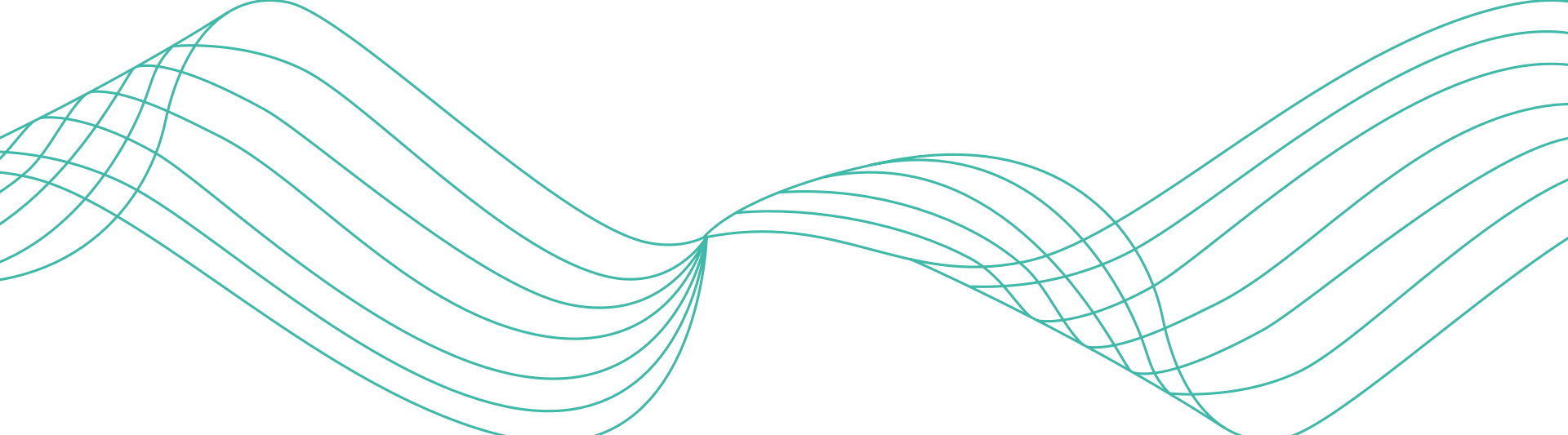




Vanliga statistiska misstag som hittas i medicinska vetenskapliga artiklar

Per Liv

per.liv@regionvasterbotten.se



Kliniska Studier
Sverige
Forum Norr



Registercentrum **Norr**

Statistikernätverket inom Forum Norr



Kliniska Studier
Sverige
Forum Norr

Region Norrbotten

Kontaktperson: Robert Lundqvist

Robert.lundqvist@norrboten.se

Region Västernorrland

Kontaktperson: Erling Englund

Erling.englund@rvn.se

Region Jämtland Härjedalen

Kontaktperson: Anna Lindam

Anna.lindam@regionjh.se

Region Västerbotten

Registercentrum Norr

Kontaktpersoner:

Gabriel Granåsen

Per Liv

Simon Vallin

Henrik Holmberg

Therese Kellgren

Christel Häggström

Oscar Öhman

statistik.rcnorr@regionvasterbotten.se

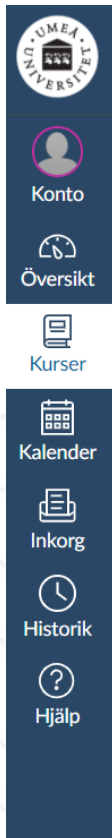


Registercentrum **Norr**

Statistikernätverket inom Forum Norr

Ny CANVAS-sida: <https://www.canvas.umu.se/courses/2600>

Länk dyker upp på Forum Norrs samt Registercentrum Norrs webbsida



Personer

Sidor

Moduler

Chat

UmU Play

UmU Play-My Media

▼ Välkommen till Forum Norrs statistikernätverks canvassida

 Vad är Forum Norrs statistikernätverk?

 Vart hör jag av mig om jag söker statistiskt forskningsstöd?

▼ Program digitala statistikseminarier våren 2021

 Program våren 2021

 Önskemål på aktiviteter och teman på framtida seminarier

► Presentationsmaterial från föreläsningar



Registercentrum Norr

Målet

- Dela med mig av några misstag (ur en statistikers synvinkel)
- Erbjud alternativ/lösningar som kan göra analyserna effektivare
- Dela lite lästtips



Agenda

- "Absence of evidence is not evidence for absence"
- Kategorisering av kontinuerliga variabler
- Inte utvärdera linjäritet vid regressionsanalys
- Jämföra grupper vid baseline i randomiserade studie med hjälp av statistiska test
- Missförstå normalfördelningens betydelse
- Underrapportering av metod
- Underrapportering av resultat



Misstag 1

“Absence of evidence is not evidence of absence”



ORIGINAL ARTICLE

Levosimendan in Patients with Left Ventricular Dysfunction Undergoing Cardiac Surgery

R.H. Mehta, J.D. Leimberger, S. van Diepen, J. Meza, A. Wang, R. Jankowich, R.W. Harrison, D. Hay, S. Femes, A. Duncan, E.G. Soltesz, J. Lubner, S. Park, M. Argenziano, E. Murphy, R. Marcel, D. Kalavrouziotis, D. Nagpal, J. Bozinovski, W. Toller, M. Heringlake, S.G. Goodman, J.H. Levy, R.A. Harrington, K.J. Anstrom, and J.H. Alexander, for the LEVO-CTS Investigators*

RESULTS

A total of 882 patients underwent randomization, 849 of whom received levosimendan or placebo and were included in the modified intention-to-treat population. The four-component primary end point occurred in 105 of 428 patients (24.5%) assigned to receive levosimendan and in 103 of 421 (24.5%) assigned to receive placebo (adjusted odds ratio, 1.00; 99% confidence interval [CI], 0.66 to 1.54; P=0.98). The two-component primary end point occurred in 56 patients (13.1%) assigned to receive levosimendan and in 48 (11.4%) assigned to receive placebo (adjusted odds ratio, 1.18; 96% CI, 0.76 to 1.82; P=0.45). The rate of adverse events did not differ significantly between the two groups.

CONCLUSIONS

Prophylactic levosimendan did not result in a rate of the short-term composite end point of death, renal-replacement therapy, perioperative myocardial infarction, or use of a mechanical cardiac assist device that was lower than the rate with placebo among patients with a reduced left ventricular ejection fraction who were undergoing cardiac surgery with the use of cardiopulmonary bypass. (Funded by Tenax Therapeutics; LEVO-CTS ClinicalTrials.gov number, NCT02025621.)



Registercentrum **Norr**

Vad skulle dom gjort?

Att en effekt inte är statistiskt signifikant **är inte en slutsats**

Studien är uppenbart underdimensionerad

Konfidensintervall: 0.66 – 1.56

Data från studien är kompatibelt med att levosimendan:

- kan sänka oddset med 44% jämfört med placebo
- öka oddset med 56% jämfört med placebo

Studien är snarast inkonklusiv.



Fotnot

Impact factor NEJM: 74,7

Lästips

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *Bmj*, 311(7003), 485.



Frågor?



Registercentrum **Norr**

Misstag 2

Kategorisering av kontinuerliga variabler



Kategorisering av kontinuerliga variabler är att slänga bort information.

Kan (undvikas att) göras på:

- Utfallsvariabler
- Prediktorer

Stor powerförlust



Exempel på kategorisering av utfallsvariabel

Primary Care

Antidepressant drugs and generic counselling for treatment of major depression in primary care: randomised trial with patient preference arms

BMJ 2001 ; 322 doi: <https://doi-org.proxy.ub.umu.se/10.1136/bmj.322.7289.772> (Published 31 March 2001)

Cite this as: BMJ 2001;322:772

Article

Related content

Metrics

Responses

Clair Chilvers, professor (Clair.Chilvers@doh.gsi.gov.uk)^a, Michael Dewey, senior lecturer^a, Katherine Fielding, lecturer^a, Virginia Gretton, research assistant^a, Paul Miller, lecturer in health economics^a, Ben Palmer, research associate^a, David Weller, professor^c, Richard Churchill, lecturer^b, Idris Williams, professor^b, Navjot Bedi, specialist registrar in psychiatry^d, Conor Duggan, professor^e, Alan Lee, consultant psychiatrist and special senior lecturer^f, Glynn Harrison, professor^g

for the Counselling versus Antidepressants in Primary Care Study Group.



Registercentrum Norr

Study size

We originally planned to use a binary outcome: recovery or not, based on change in Beck score. With a recovery rate of 50% in one group and 40% in the other, two sided significance of 0.05, and power of 0.90, we required 400 patients in each arm. Because recruitment was slow, we resized the study using the difference in mean Beck scores at 12 months as the outcome. Based on a clinically important difference between the groups in mean Beck scores of 5 points, and assuming a standard deviation of 8.3 with two sided significance of 0.05, we required 44 patients per arm for power of 80% and 60 per arm for power of 90%.

400 eller 60 patients i varje behandlingsarm?

Vilket skulle EPM föredra?



Vad kan man göra istället?

Motsättning kan finnas mellan statistisk effektivitet av utfalls variabel och klinisk betydelse

T.ex. :

KOL: $FEV1/FVC < 0,7$

Diabetes typ 2: fasteglukos ≥ 7.0 mmol

Överväg analys av rådata utan kategorisering
Inga responder/non-responder-analyser



Kategorisering av prediktorer

Vissa variabler kategoriseras väldigt ofta, t.ex. BMI eller ålder

Dessa har sällan linjära effekter

Är detta ett argument för kategorisering? Nja...

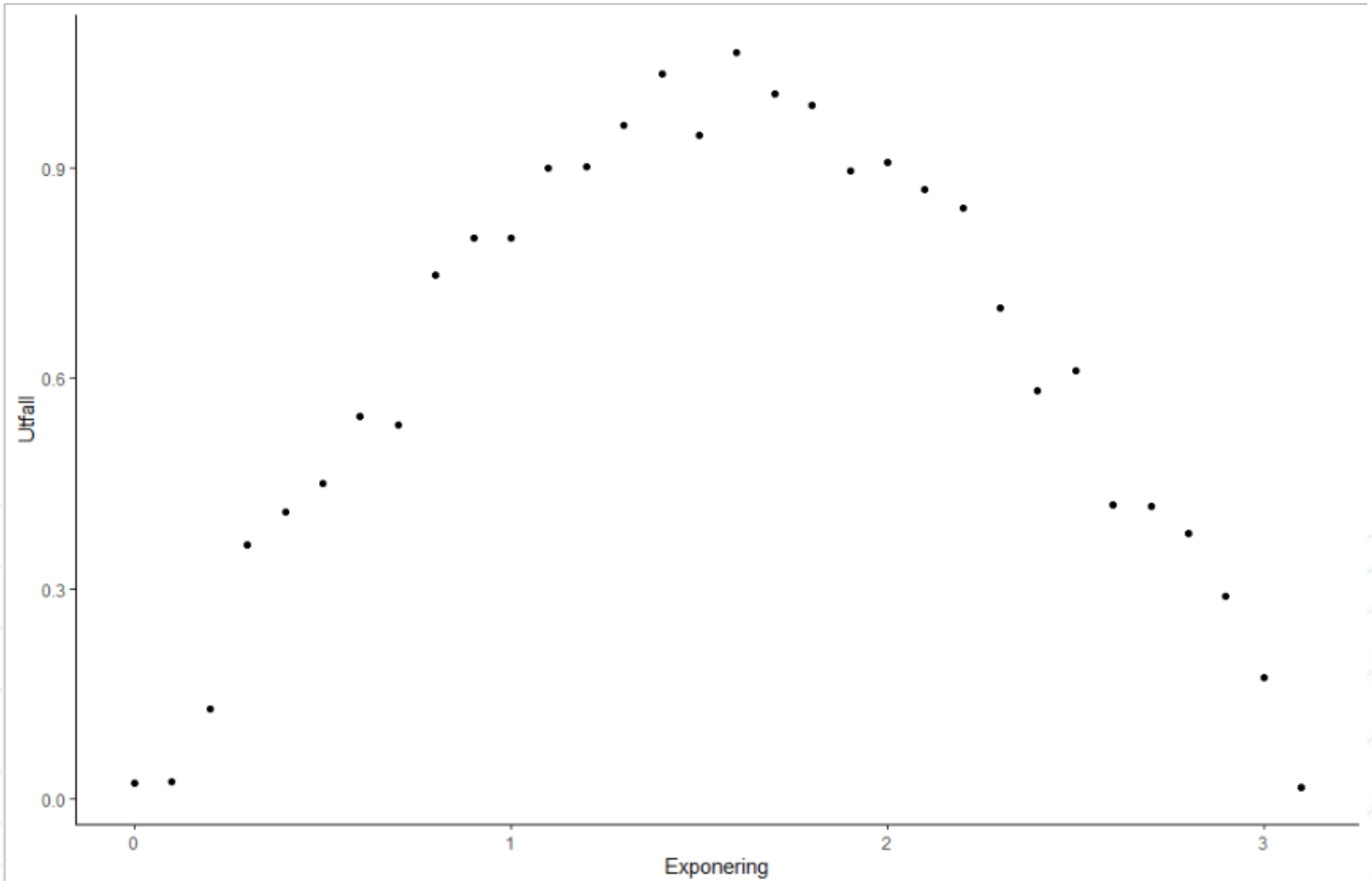


Misstag 3

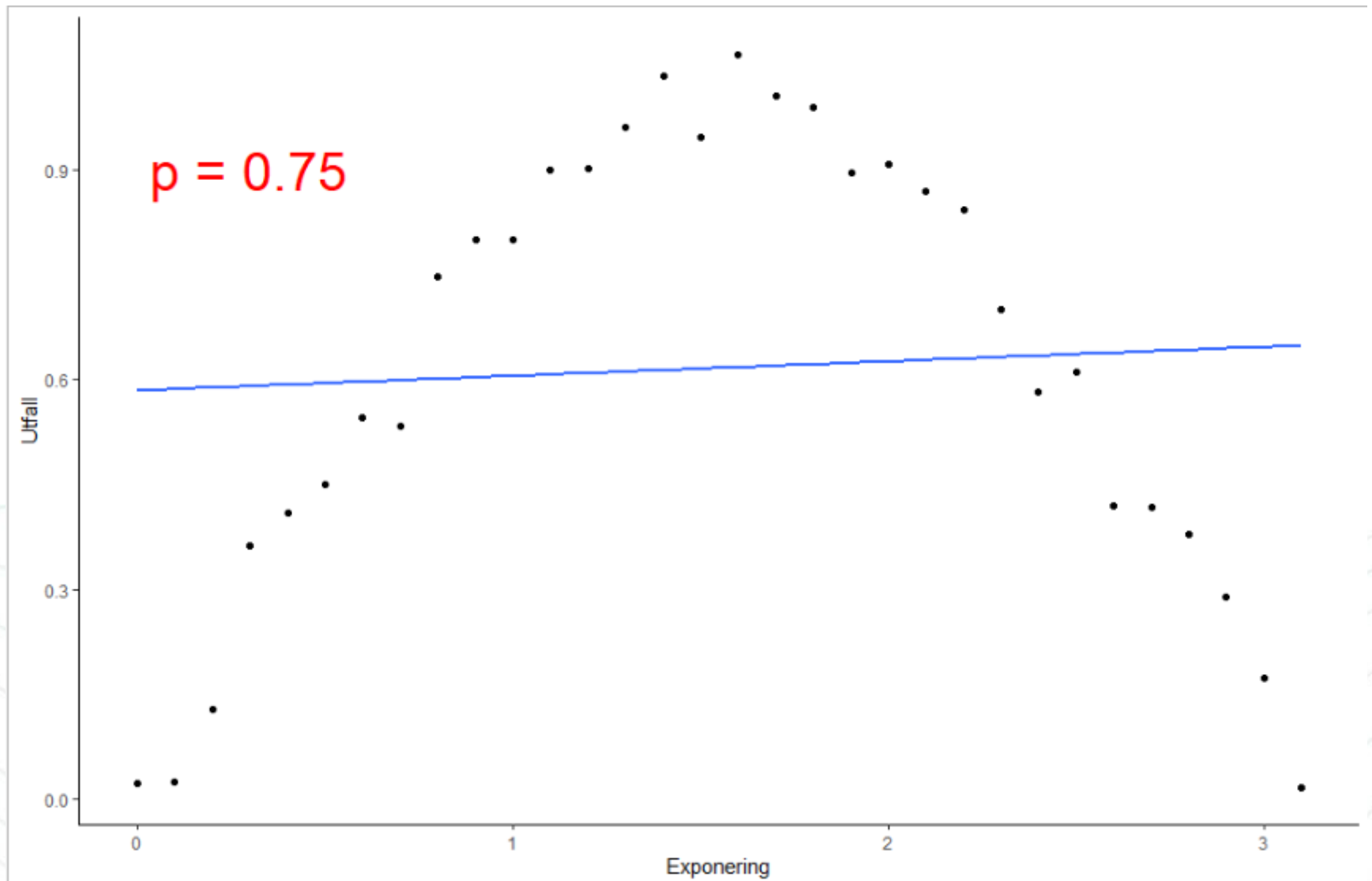
***Inte utvärdera linjäritet vid
regressionsanalys***



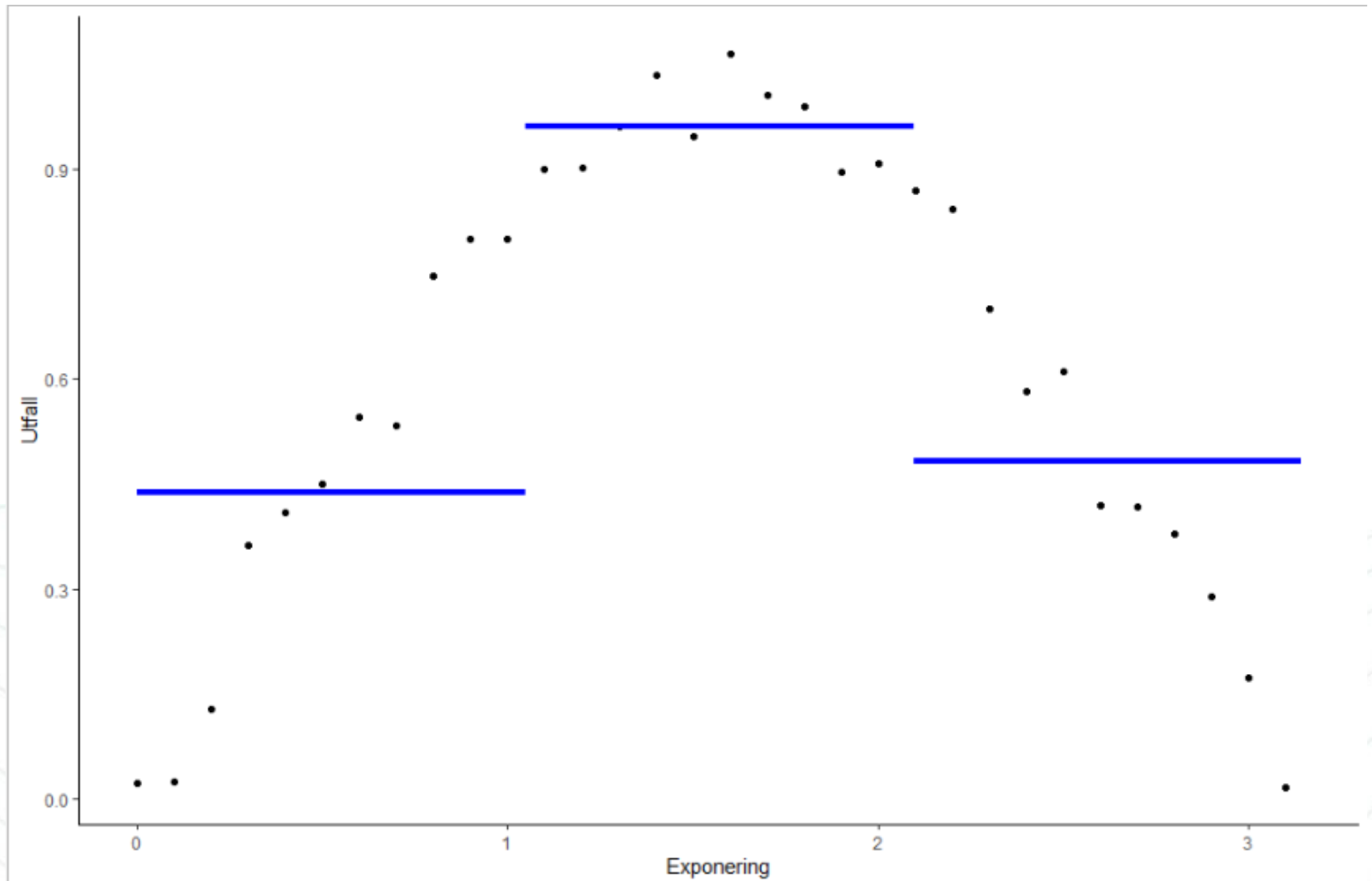
Hypotetiskt data



Linjär effekt

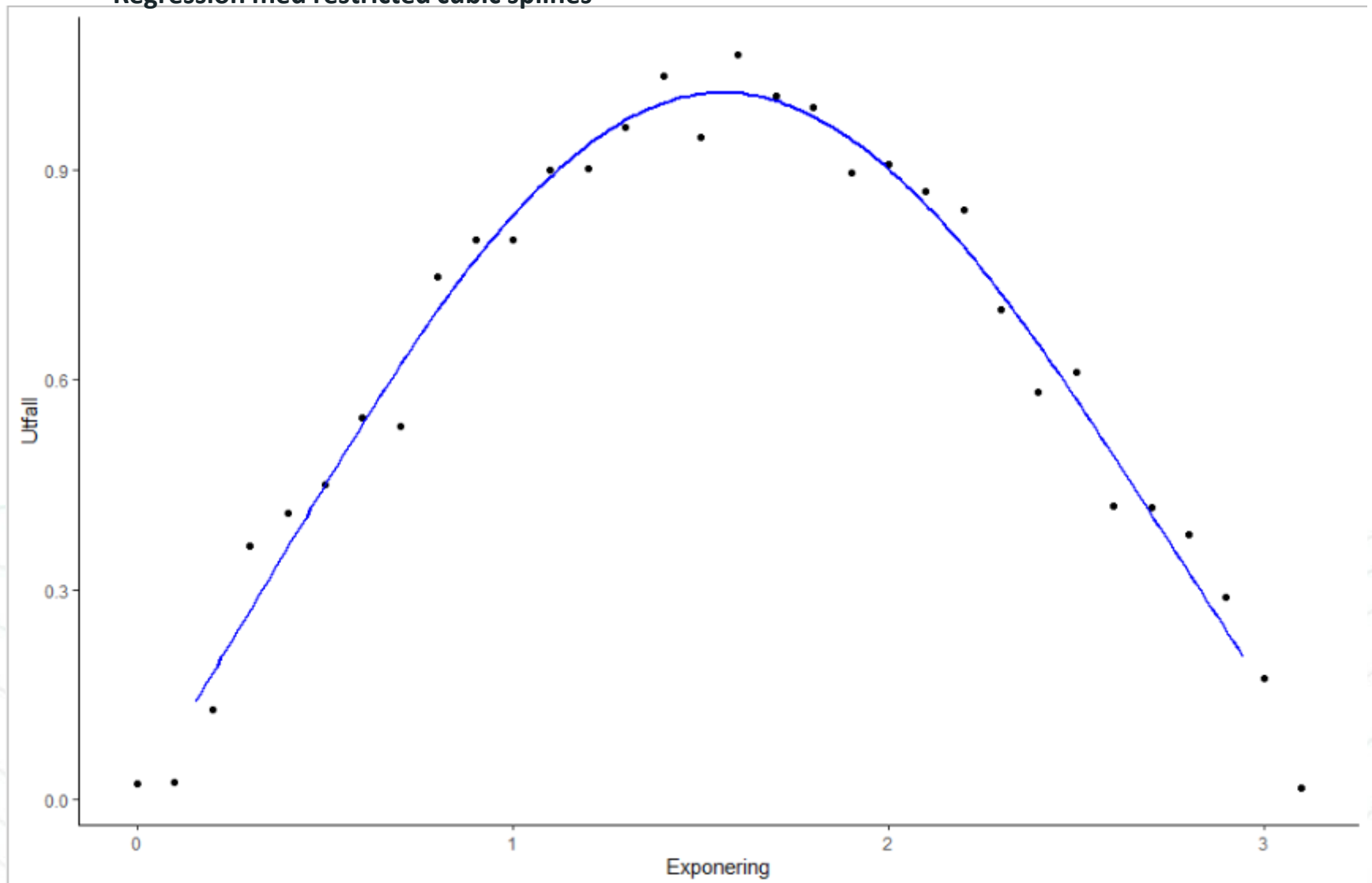


Kategoriserad i tre lika stora kategorier



Icke-linjär effekt

Regression med restricted cubic splines



Vad ska man göra?

- Undvika kategoriseringar
- Försök värdera linjäritetsantagande om linjär effekt antas
- Modellera kontinuerliga prediktorer som icke-linjära effekter när linjäritet inte kan antas
 - t.ex. genom s.k. splines



Lästips

Bennette, C., & Vickers, A. (2012). Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC medical research methodology*, 12(1), 1-5.

Gauthier, J., Wu, Q. V., & Gooley, T. A. (2020). Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians.



Frågor?



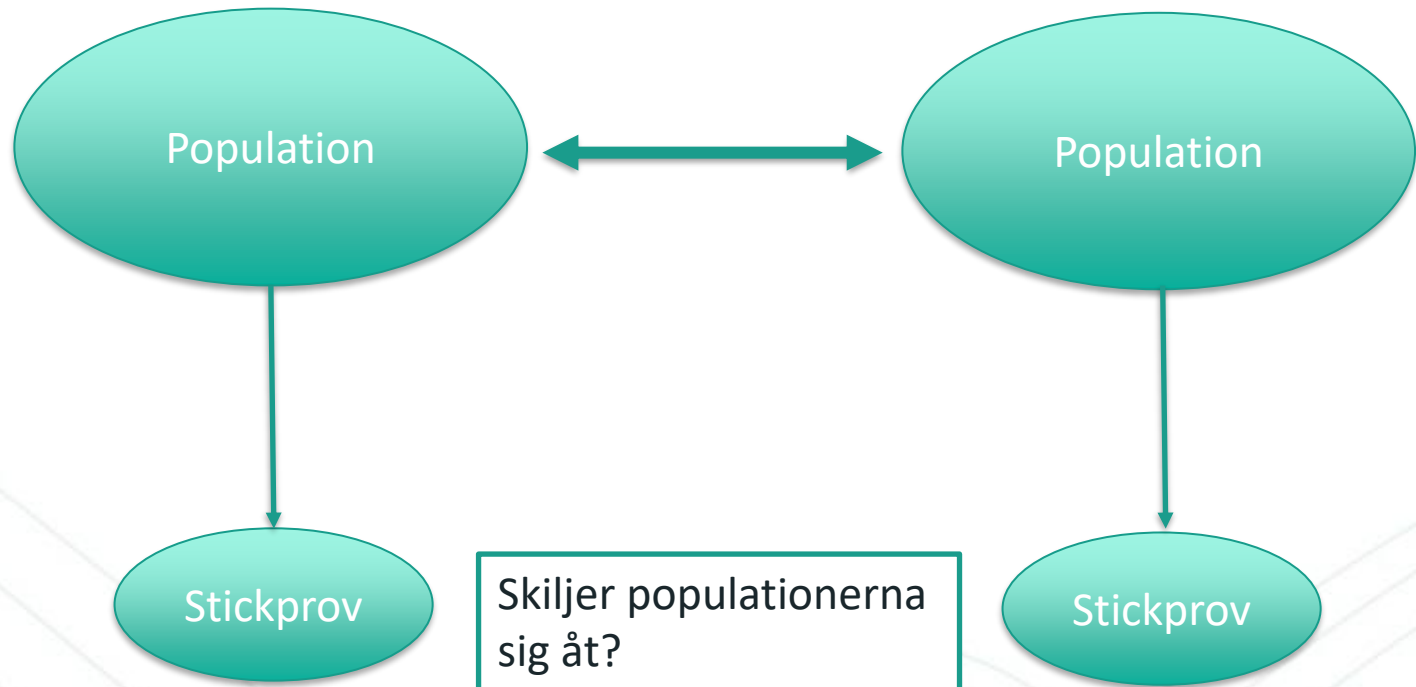
Registercentrum **Norr**

Misstag 4

*Jämföra grupper vid baseline i randomiserade studie med
hjälp av statistiska test
(och använda det som skäl för att justera modellen)*



Principen bakom statistisk hypotesprövning



Randomiserad kontrollerad prövning

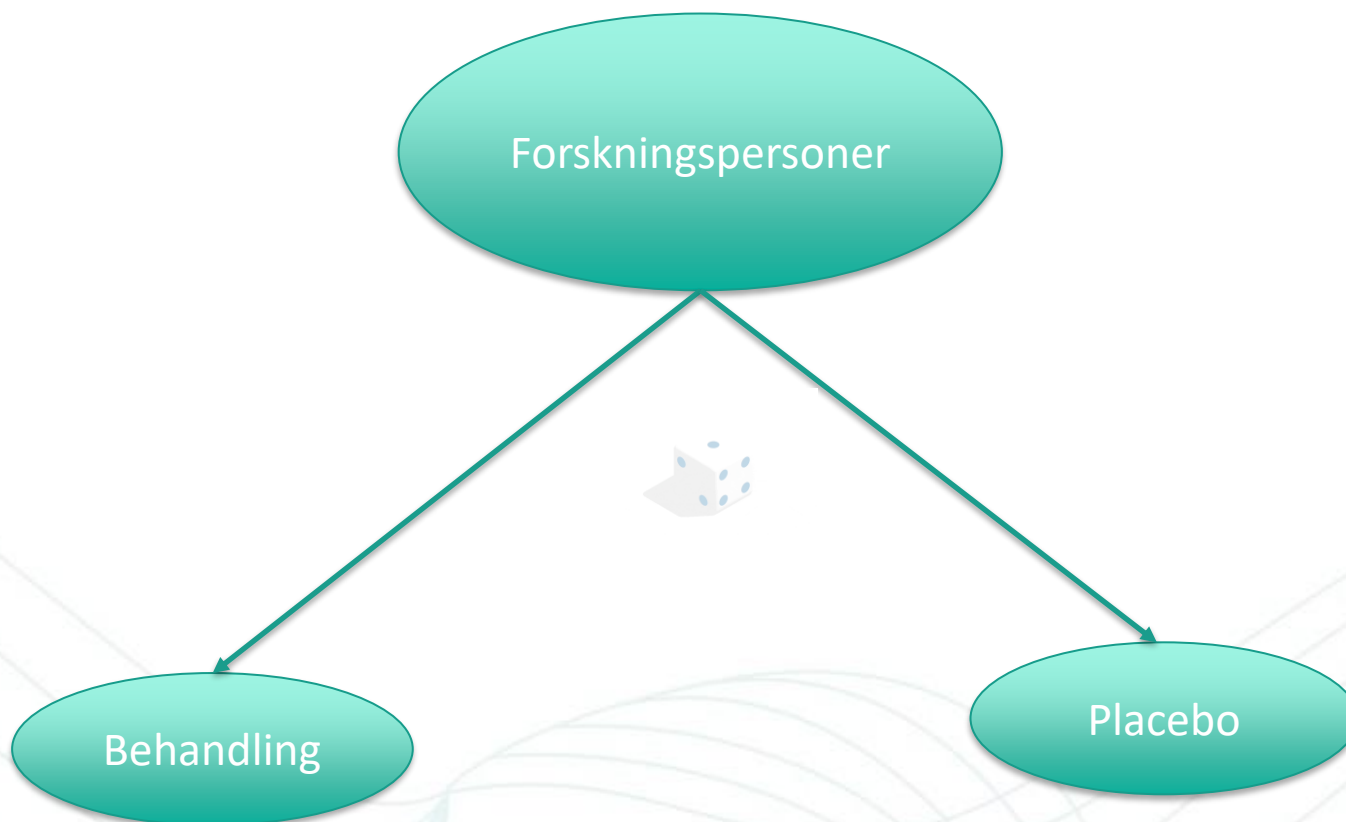


TABLE 1 Participant Demographics and Baseline Characteristics

	Individual-ESI (N = 42)	Group-ESI (N = 40)	p	Hedges's g
Demographics				
Age	19.64 (1.93)	19.58 (1.42)	.86	0.04
Ethnicity			.58	—
White, %	73.8	72.5	—	—
Black, %	7.1	10.0	—	—
Other, %	14.3	15.0	—	—
Gender, % male	81.0	92.5	.13	—
Maternal age	31.98 (5.74)	31.71 (5.44)	.83	0.05
Maternal education	15.64 (2.07)	15.51 (2.26)	.80	0.06
Communication and Symbolic Behavior Scales				
Social Composite	37.02 (17.25)	39.49 (21.11)	.57	0.13
Speech Composite	8.61 (2.32)	6.50 (10.03)	.42	0.18
Symbolic Composite	24.00 (16.60)	21.54 (15.15)	.49	0.15
Autism Diagnostic Observation Schedule				
Social Affect	13.50 (4.28)	14.43 (3.86)	.47	0.17
Restricted, Repetitive Behavior	3.05 (1.50)	2.85 (1.55)	.66	0.10
Vineland Adaptive Behavior Scales				
Communication	78.83 (13.06)	79.79 (13.51)	.98	0.01
Daily Living	86.60 (10.98)	87.42 (11.97)	.97	0.01
Socialization	84.55 (8.77)	87.21 (9.62)	.52	0.15
Motor	94.55 (8.85)	95.34 (11.66)	.72	0.08
Adaptive Behavior Composite	83.98 (8.93)	85.03 (11.37)	.65	0.07
Mullen Scales of Early Learning				
Visual Reception	42.07 (13.01)	40.42 (10.44)	.45	0.17
Fine Motor	46.20 (11.59)	42.48 (12.65)	.15	0.33
Receptive Language	29.27 (12.34)	31.35 (12.61)	.39	0.19
Expressive Language	29.61 (11.22)	28.68 (10.95)	.66	0.10
Early Learning Composite	75.56 (16.68)	74.05 (16.86)	.69	0.17

P and Hedges's g values refer to comparison of individual-ESI and group-ESI conditions. —, not applicable.

“Performing a significance test to compare baseline variables is to assess the probability of something having occurred by chance when we know that it did occur by chance”
Douglas Altman

- RCT kräver inte lika grupper för att vara giltiga
- Justera för prediktiva kovariat är gynnsamt oavsett om signifikanta skillnader finns mellan grupperna



Vad ska man göra istället?

- Inte presentera p-värden i tabell 1
 - CONSORT-statements guideline påpekar det olämpliga
 - (Även STROBE avråder)
- Justera för viktiga prediktorer oavsett om signifikant skillnad
 - Prespecificera dessa



Lästips

Altman, D. G. (1985). Comparability of randomised groups. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 34(1), 125-136.

De Boer, M. R., Waterlander, W. E., Kuijper, L. D., Steenhuis, I. H., & Twisk, J. W. (2015). Testing for baseline differences in randomized controlled trials: an unhealthy research behavior that is hard to eradicate. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 1-8.

Lingsma, H., Rozenbeek, B., & Steyerberg, E. (2010). Covariate adjustment increases statistical power in randomized controlled trials. *Journal of clinical epidemiology*, 63(12), 1391.



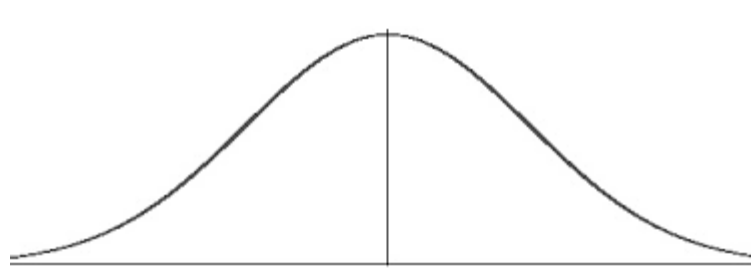
Frågor?

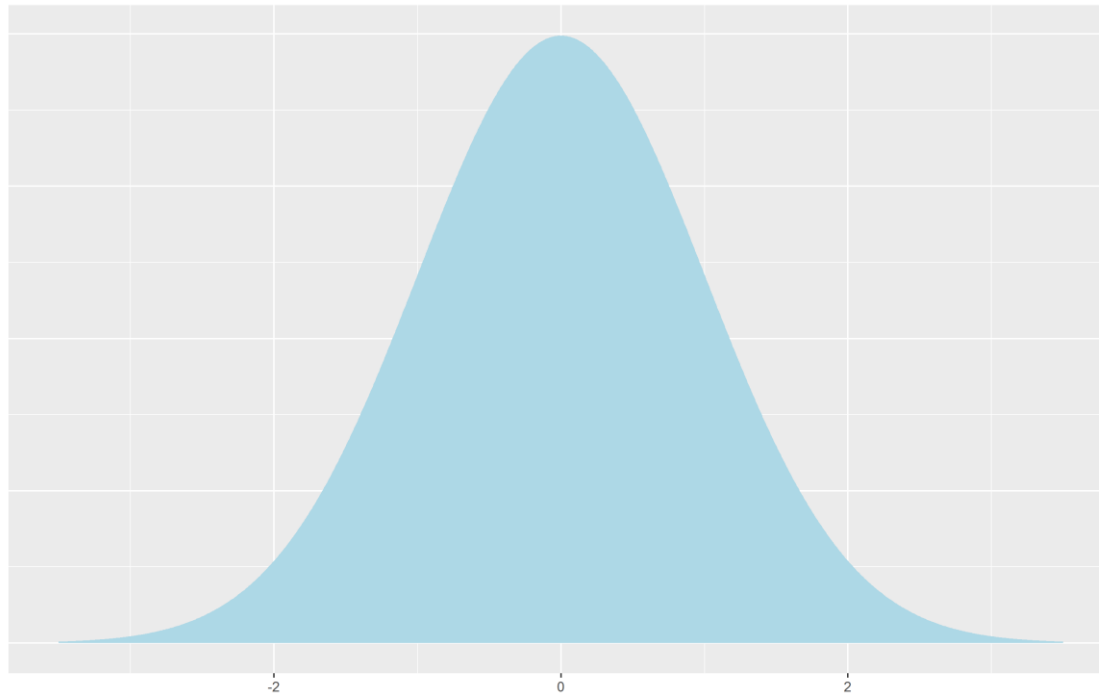


Registercentrum **Norr**

Misstag 5

Missförstå normalfördelningens betydelse





$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

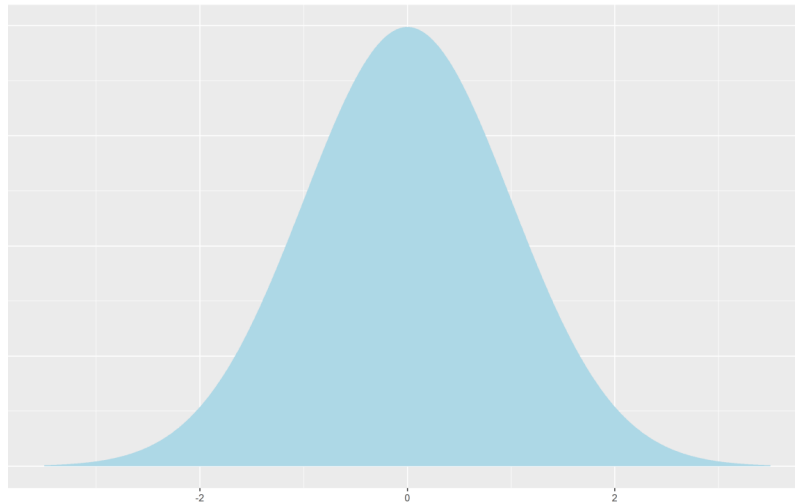
Fördelningens utseende bestäms helt av två **parametrar**:

- μ - medelvärde
- σ - standardavvikelse



Normalfördelning

Otroligt viktigt och inte sällan oviktigt



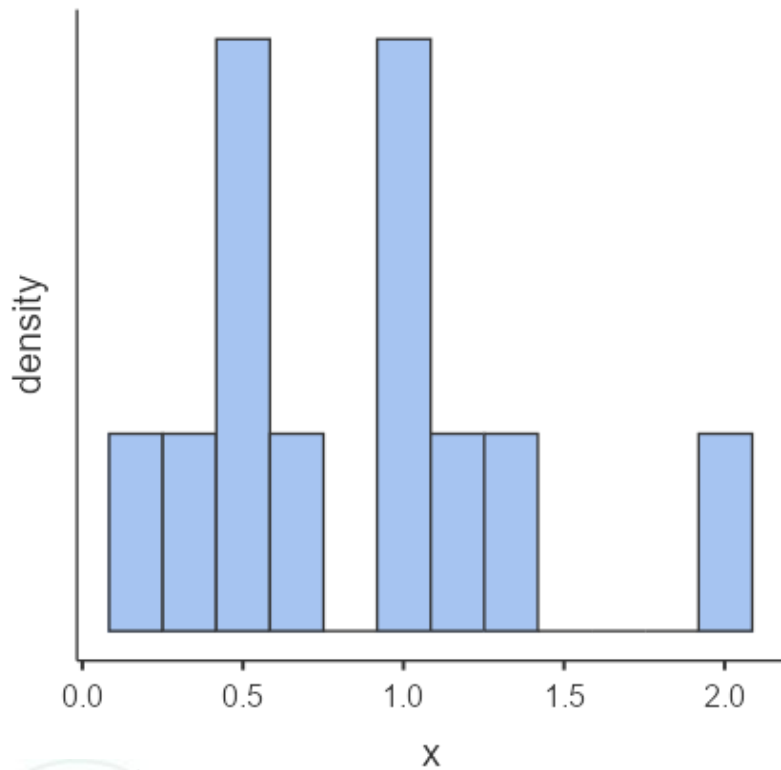
Dom flesta statistiker är överens om att ofta fungerar parametrisk statistik bra vid måttliga avvikelser från normalfördelning

Vad är måttliga avvikelser?

Gissa normalfördelningen



Registercentrum **Norr**

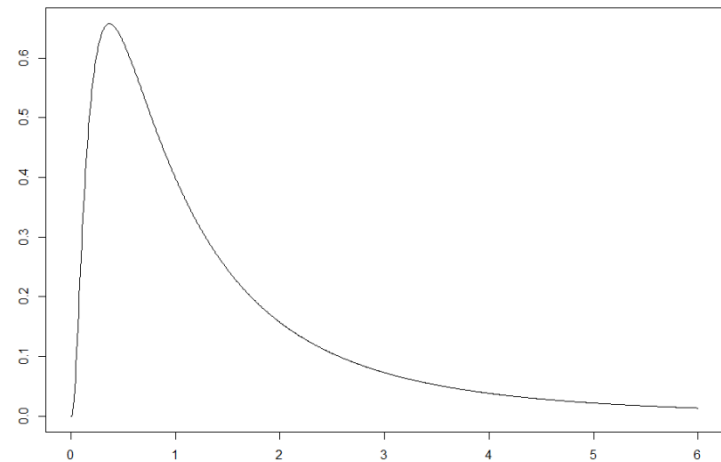
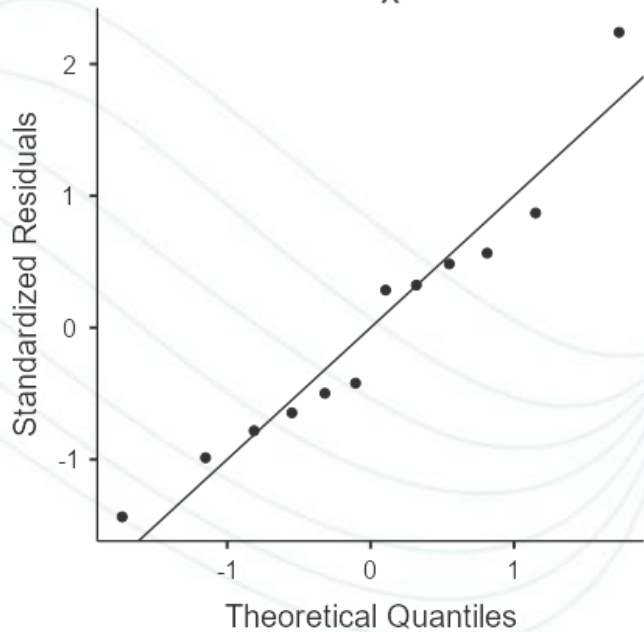


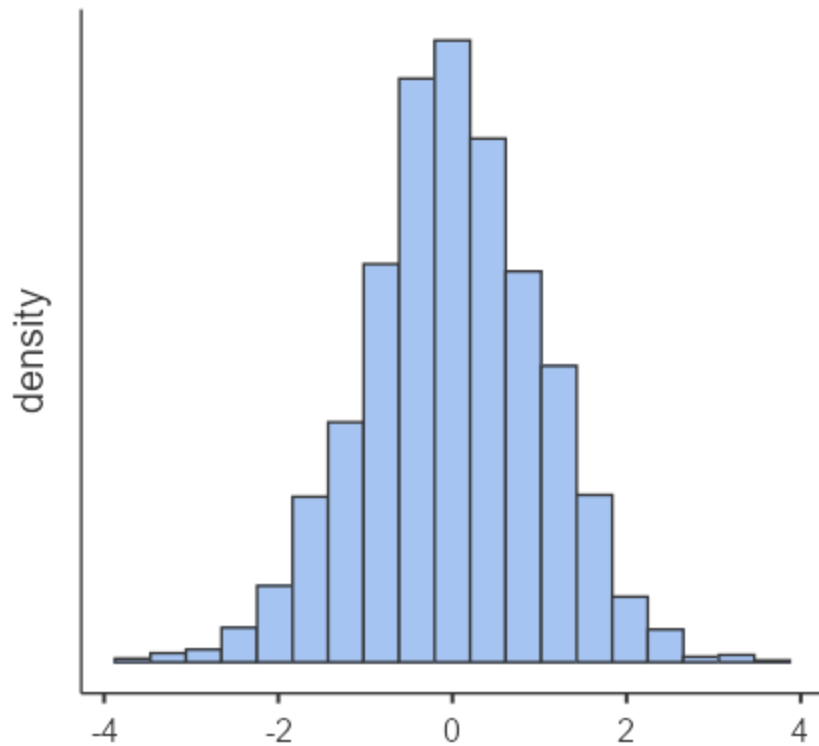
$$n = 12$$

$$p = 0.55$$

Normalfördelat?

Nej!



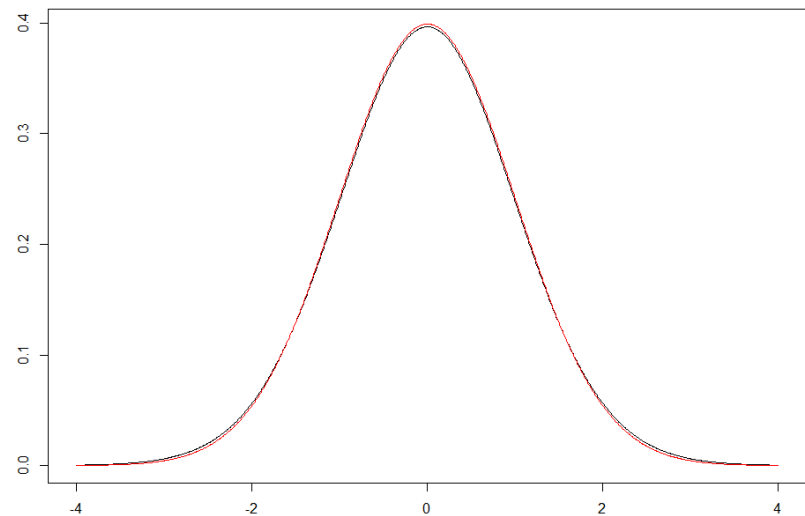
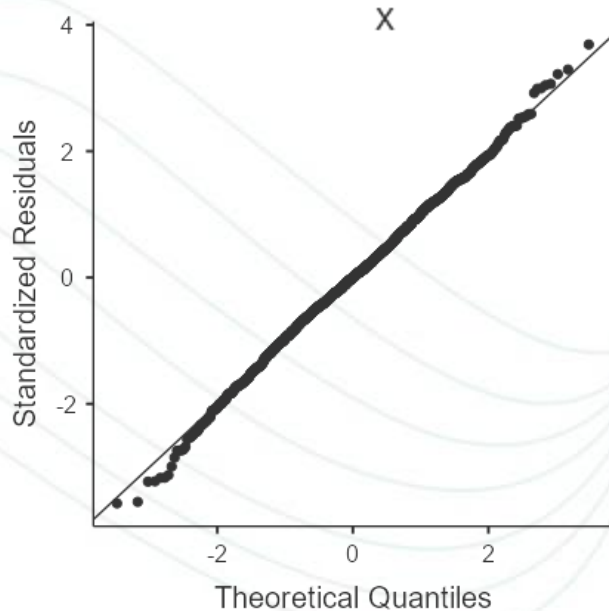


$n = 2000$

$p = 0.04$

Normalfördelat?

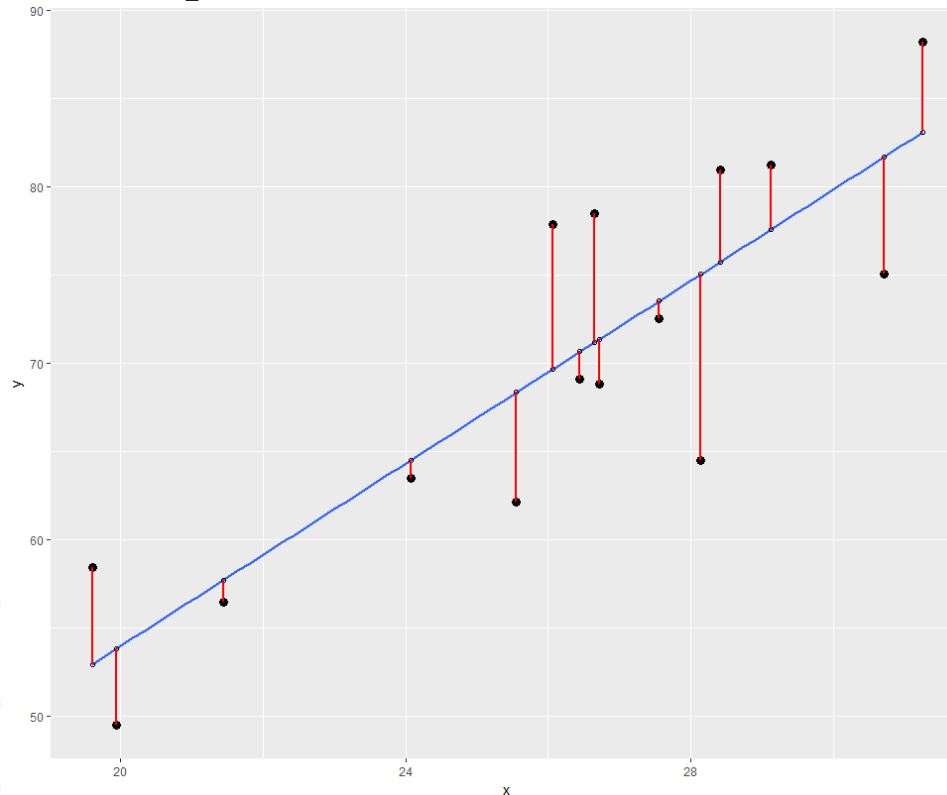
Nej!



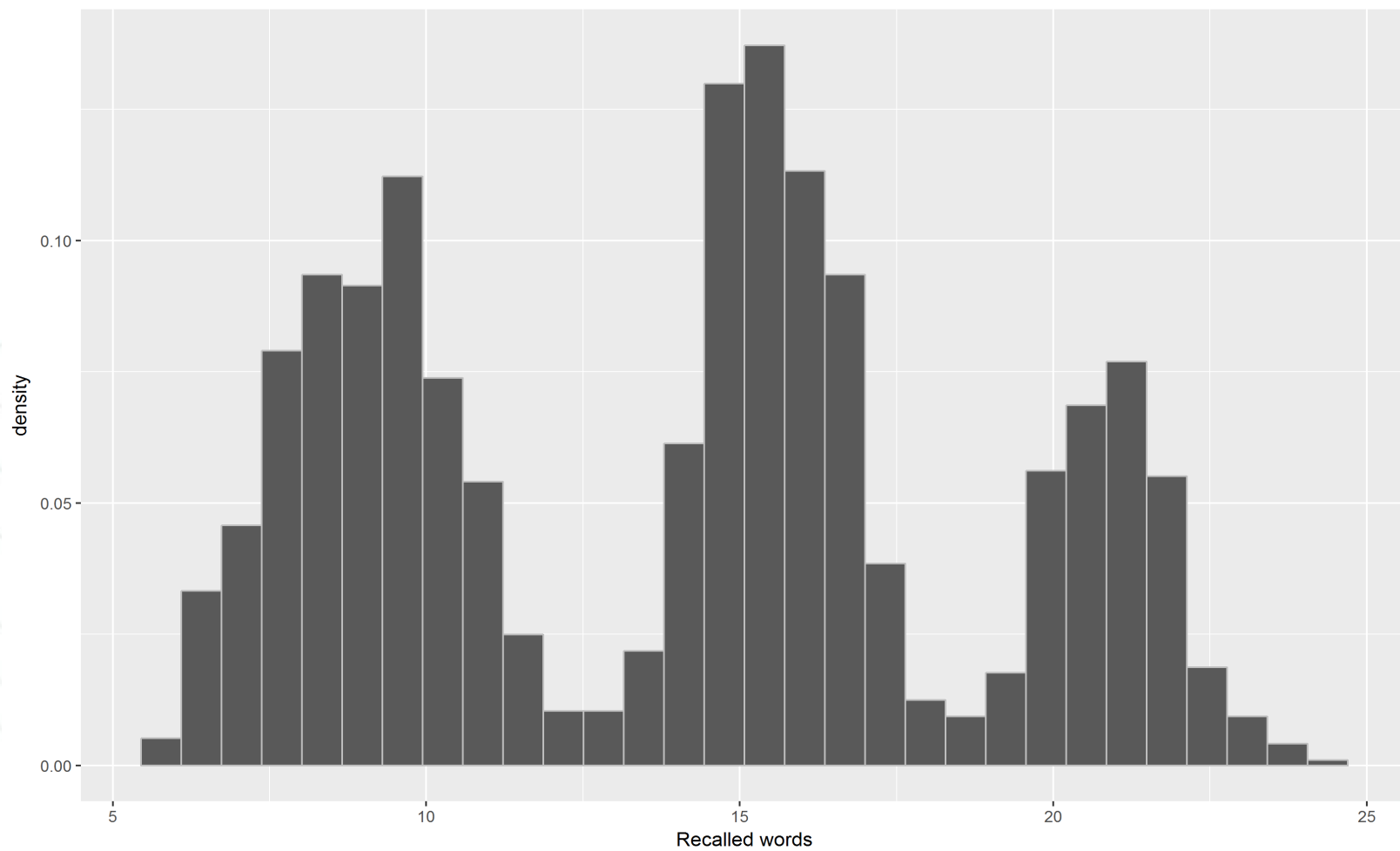
I linjär regression

Residualerna antas vara normalfördelade

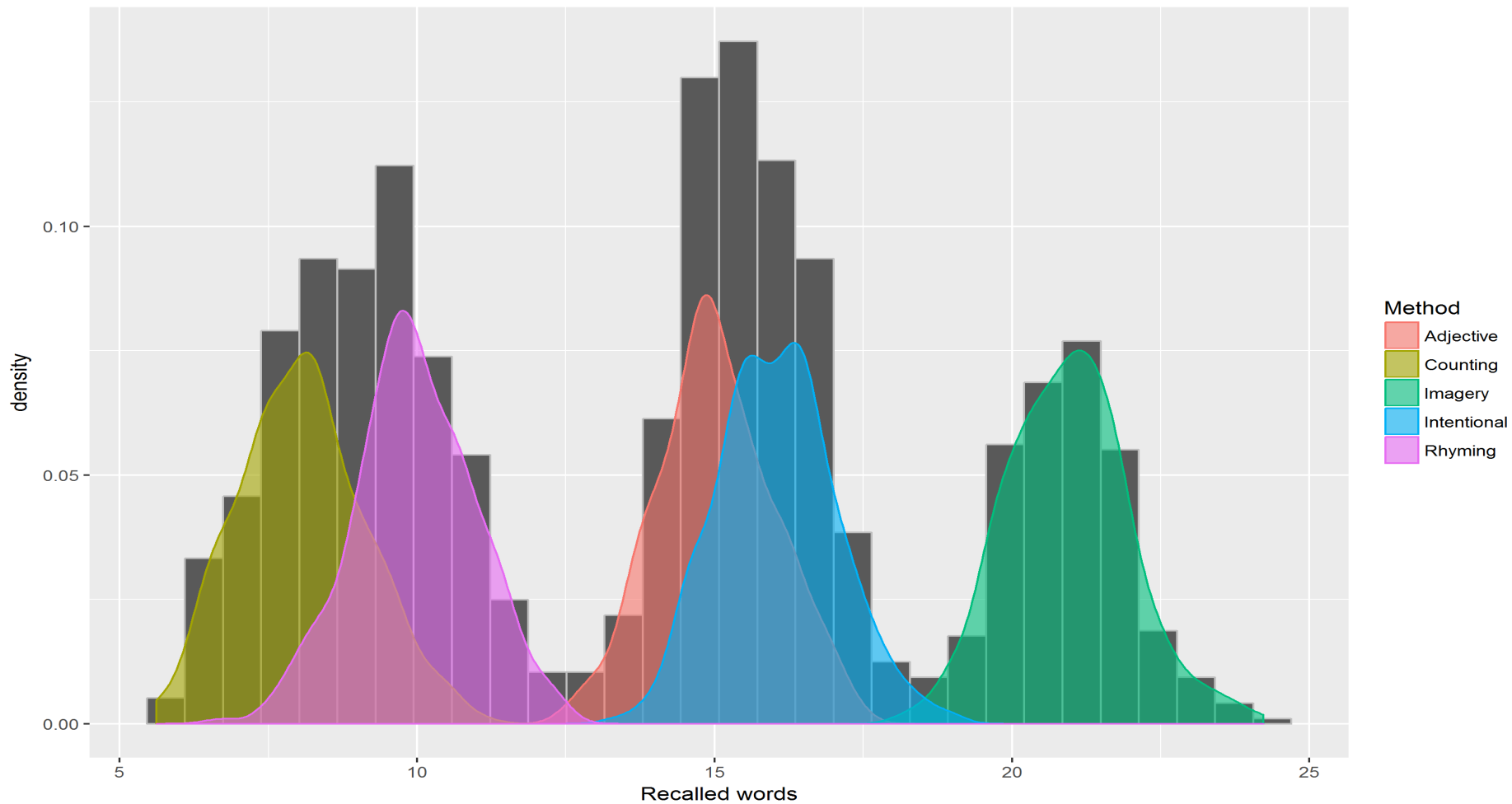
$$y = a + bx + \varepsilon$$



Inte y eller x



Inte variablerna som ska normalfördelade utan residualerna



Vad ska man göra?

Osäker? Prata med statistiker

Förvänta dig inte att alla statistiker kommer ge samma svar

Undvik normalfördelningstest för att utvärdera normalfördelning

- bedöm grafiskt utifrån hur stor avvikelser är
 - qq-plottar
 - histogram



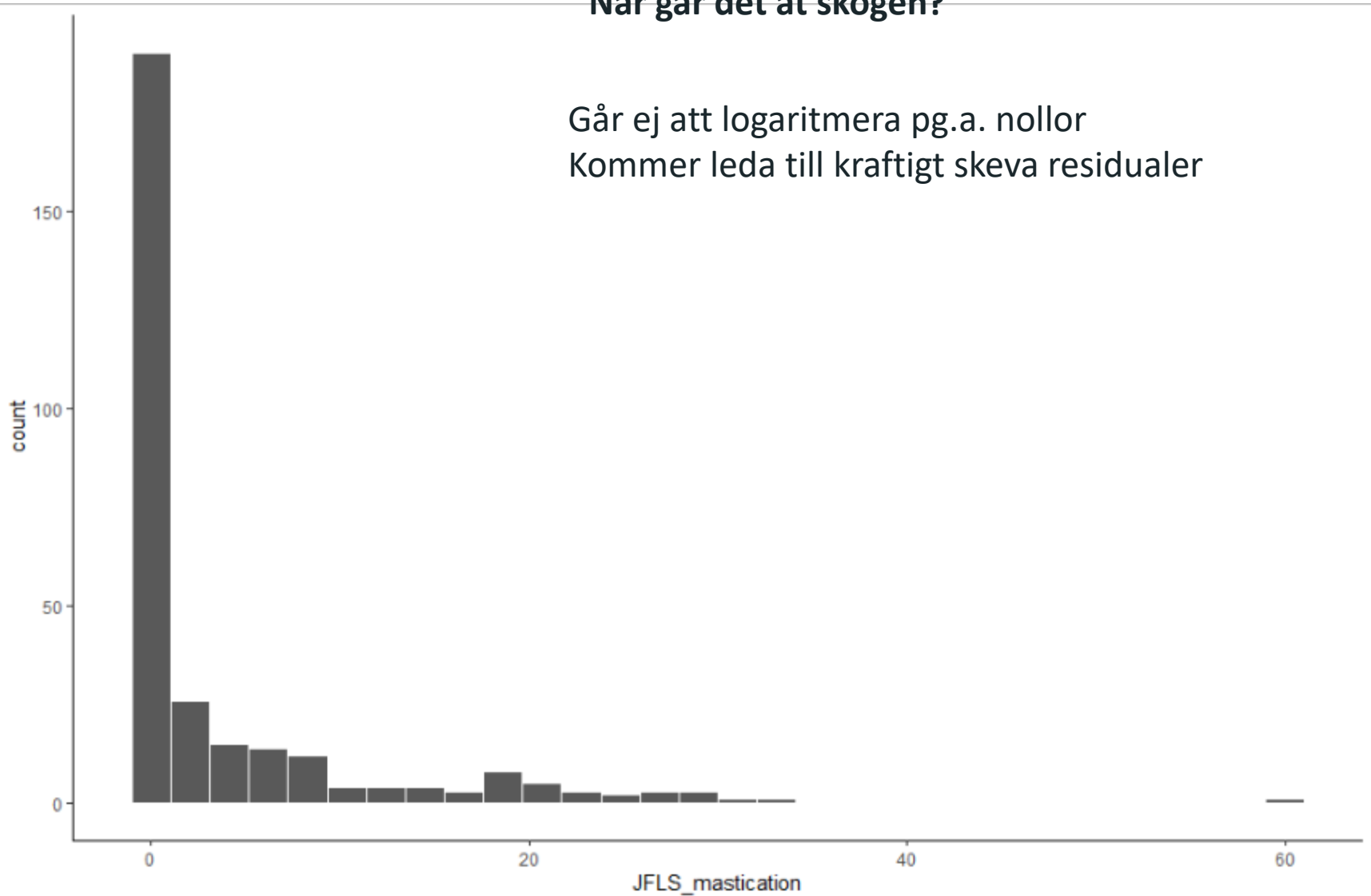
Vad ska man göra om normalfördelningsantagandet inte uppfylls?

- Icke-parametrisk statistik
- Transformation, t.ex. logaritmera utfallsvariabeln



När går det åt skogen?

Går ej att logaritmera pg.a. nollor
Kommer leda till kraftigt skeva residualer



Vad ska man göra om normalfördelningsantagandet inte uppfylls?

- Icke-parametrisk statistik
- Transformation, t.ex. logaritmera utfallsvariabeln
- **Semi-parametriska modeller**
 - **T.ex. ordinal regression**



Frågor?



Registercentrum **Norr**

Misstag 6

Underrapportering av metod



Stephen John Senn @stephensenn · 5 aug. 2015



Dear Medic, telling me you used ANOVA to analyse the data is as informative as telling your colleague you treated the patient with a pill



8



12



Registercentrum Norr

I artikel på 3500 ord måste det väl finnas plats för att kunna beskriva den statistiska metod tillräckligt för replikering, eller?



Vad ska man göra?

- Beskriv din statistiska modell: beroende variabler och oberoende variabler samt modellens struktur
- Låt syftet med analysen framgå
 - Undvik:
 - "Data was analysed using..."
 - "Mann-Whitney U-test was used when appropriate"



Misstag 7

Underrapportering av resultat



Underrapportering av resultat

Ett p-värde - inte ett fullständigt resultat

Hur mycket skiljer sig grupperna åt?

Osäkerheten i denna skillnad? (konfidensintervall)



Lästips

Lang, T. A., Lang, T., & Secic, M. (2006). *How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers*. ACP Press.



Frågor?



Registercentrum **Norr**

Hjälp med ansökningar till nationella forskningsfinansiär

ALF-kommitténs riktade satsning statistikstöd möjliggör att vi hjälpa kostnadsfritt

Hör av er till:

statistik.rcnorr@regionvasterbotten.se

Gärna med god framförhållning så att vi får vara med tidigt i planeringen. 😊



Registercentrum **Norr**