

En introduktion till

ordinal regression

Per Liv

per.liv@umu.se



Kliniska Studier
Sverige
Forum Norr



Skaltyper



Nominal skala



Ordinal skala



Intervallskala



Kvotskala



Intervallskala eller kvotskala, spelar det någon roll?

DAGENS NYHETER.

E-DN
ARKIVET
KORSORD

KUNDSERVICE
QUIZ
KUNDERBJUDANDEN

Nyheter **Sverige** Världen Ekonomi Kultur Sport Klimatet Ledare DN Debatt ☰ Meny

Sthlm Gbg Skåne Norra Sverige

SVERIGE

Dubbelt så varmt som vanligt i november

PUBLICERAD 2005-11-14

Medeltemperaturen under början av november har varit högre än på länge. Men passa på att njuta under dagen nu blir det snabbt kallare.

Räkneregler för ordinala variabler

- Kan inte räkna summera eller beräkna medelvärden

(“Inte alls” + “Måttligt”) / 2 = “Lite” ???

6. Under de senaste fyra veckorna, i vilken utsträckning har Ditt kroppsliga hälsotillstånd eller Dina känslomässiga problem stört Ditt vanliga umgänge med anhöriga, vänner, grannar eller andra?

Inte alls.....1
Lite.....2
Måttligt.....3
Mycket.....4
Väldigt mycket...5

- Svårigheter att räkna förändringar
 - Vi vet om en patient har förbättrat sig, men inte med hur mycket

Vilken skaltyp?

Borgs RPE-skala

Borg RPE Scale	
Scoring	Level of Exertion
6	No Exertion
7	Extremely Light
8	
9	Very Light
10	
11	Light
12	
13	Somewhat Hard
14	
15	Hard (Heavy)
16	
17	Very Hard
18	
19	Extremely Hard
20	Maximal Exertion

Vilken skaltyp?

VAS-skala



Vilken skaltyp?

”Sum scores” från skattningsinstrument

Exempel

SF-36 fylls i av en fiktiv kvinna. Delskalan ”social funktion” innehåller två frågor och besvaras nedanför:

6. Under de senaste fyra veckorna, i vilken utsträckning har Ditt kroppsliga hälsotillstånd eller Dina känslomässiga problem stört Ditt vanliga umgänge med anhöriga, vänner, grannar eller andra?

Inte alls.....1
Lite.....2
Måttligt.....3
Mycket.....4
Väldigt mycket....5

10. Under de senaste fyra veckorna, hur stor del av tiden har Ditt kroppsliga hälsotillstånd eller Dina känslomässiga problem stört Dina möjligheter att umgås (t ex hälsa på släkt, vänner, etc.)?

Hela tiden.....1
Största delen av tiden....2
En del av tiden.....3
Lite av tiden.....4
Inget av tiden.....5

Icke-parametriska metoder då?

Wilcoxon's rangsummetest

(Mann-Whitney U-test)

- Jämför två oberoende grupper
- Fungera för ordinala variabler
 - Inte optimalt med många "ties"
- Ojusterade analys

Wilcoxon tecken-rangtest

- Jämför parade mätningar
- Beräknar skillnad mellan paren
 - **Inte tillåtet för ordinal variabler**
- Ojusterade analyser

Ordinal regression

- Semi-parametrisk regressionsmetod för ordinal utfallsvariabel
- Finns olika metoder
- Vi ska prata om en specifik:
 - Proportional odds logistic regression
 - Besläktat med Mann-Whitney U test.

Logistisk regression

Modellerar odds för en händelse

$$Odds = \frac{P}{1 - P}$$

Ju högre sannolikhet, desto högre odds

$$\text{Log}(P/(1-P)) = b_0 + b_1x + b_2x_2 + b_3x_3 + b_4x_4$$

Logistisk regression

Regressionskoefficienterna ger oddskvoter

Oddskvot på 0.85:

”Oddset för en hjärt-kärlhändelse efter genomgången behandling är 15% lägre jämfört med oddset efter kontrollbehandling”

Ordinal regression

Modellerar också odds för en händelse

$$Odds = \frac{P}{1 - P}$$

$$\text{Log}(P/(1-P)) = b_0 + b_1x + b_2x_2 + b_3x_3 + b_4x_4$$

Men vilken händelse?

Ordinal regression

Vi modellerar oddset för att göra en högre skattning
(oavsett var på skalan som vi befinner oss)



Ordinal regression



Oddskvot på 0,85:

Oddset att man ska skatta högre än **“inte alls”** är 15%
lägre i behandlingsgrupp jämfört med kontrollgrupp

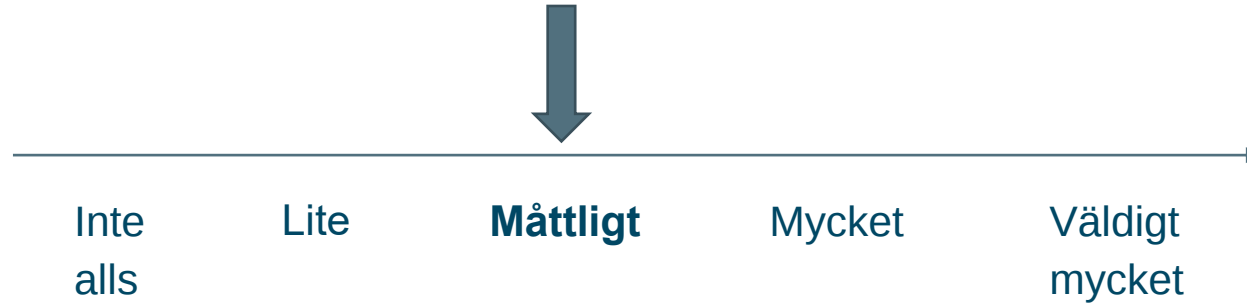
Ordinal regression



Oddsquot på 0,85:

Oddset att man ska skatta högre än **“Lite”** är 15%
lägre i behandlingsgrupp jämfört med kontrollgrupp

Ordinal regression



Oddsquot på 0,85:

Oddset att man ska skatta högre än **“Måttligt”** är 15%
lägre i behandlingsgrupp jämfört med kontrollgrupp

Ordinal regression



Oddsquot på 0,85:

Oddset att man ska skatta högre än **“Mycket”** är 15%
lägre i behandlingsgrupp jämfört med kontrollgrupp

Proportionella oddsantagandet

- Orealistiskt antagande?
- Kan ses som en genomsnittlig effekt över hela skalan
 - Knappast någon nackdel jämfört med att dikotomisera
- Bör undersökas
 - Dikotomisera, gör vanligt logistisk regression och kolla oddskvoterna

Violation of Proportional Odds is Not Fatal

</> Code ▾

2020

ORDINAL

ACCURACY-SCORE

RCT

REGRESSION

HYPOTHESIS-TESTING

METRICS

Many researchers worry about violations of the proportional hazards assumption when comparing treatments in a randomized study. Besides the fact that this frequently makes them turn to a much worse approach, the harm done by violations of the proportional odds assumption usually do not prevent the proportional odds model from providing a reasonable treatment effect assessment.

AUTHOR

Frank Harrell

AFFILIATION

Department of Biostatistics
Vanderbilt University School of Medicine

PUBLISHED

September 20, 2020

Clearly, the dependence of the proportional odds model on the assumption of proportionality can be overstressed. Suppose that two different statisticians would cut the same three-point scale at different cut points. It is hard to see how anybody who could accept either dichotomy could object to the compromise answer produced by the proportional odds model. — Stephen Senn

<https://www.fharrell.com/post/po/>

Kan ordinal regression hjälpa oss att analysera förändring på en ordinal skala

Nej....

Vi kan modellera skillnaden vid en uppföljning givet värdet vid baslinjen, t.ex. vid en RCT.

$$Y_{\text{uppföljning}} = \text{Grupp} + Y_{\text{baslinje}}$$

Svenssons metod – för förändring av ordnade variabler

KLINIK OCH VETENSKAP

Statistisk metod för parade ordinaldata

Vad är behandlingseffekt om patienten blev *bättre*, men ingen vet *hur mycket* bättre?

Elisabeth Svensson, professor i statistik, spec biostatistik, institutionen för ekonomi, statistik och informatik, Örebro universitet
elisabeth.svensson@esi.oru.se

Utvärdering av behandlingseffekt eller annan förändring när datamaterialet består av ordnade kategorier kräver extra uppmärksamhet vid val av statistisk verktygslåda. Det är inte bara bedömningar av kvalitativa variabler som oro, smärta, välbefinnande, kvalitet och funktion på någon skattningsskala som ger upphov till ordnade kategoridata eller ordinaldata, som mätnivån även kallas. Kategoriseringar av mätvärden från standardiserade instrument, som utnyttjar fysikens eller kemins lagar, och sammanfattande nivå- eller stadiindelningar av flera variabler är andra exempel på generering av ordinaldata [1-3]. När kroppslängd till exempel mäts och uttrycks i en standardiserad längdenhet, som meter, då består datamaterialet av kvantitativa kontinuerliga värden, även om de är avrundade, och statistiska metoder för analys av förändring som baseras på differenser, som Students parade t-test och Wilcoxon's teckenrangtest, är möjliga val förutsatt att övriga villkor för testen är uppfylla. Om däremot kroppslängd kategoriseras till »kort«, »normal«, »lång« för sin ålder, så blir valet av statistisk verktygslåda ett annat [2-5].

SAMMANFATTAT

Sammanfattat
Bedömningar med olika typer av skalor och bedömningsformulär ger upphov till ordinaldata. Mätvärdena i denna typ av data representerar enbart en ordningsstruktur och saknar information om storlek och avstånd mellan kategorier.

I artikeln presenteras en statistisk metod, som är speciellt utvecklad för att analysera förändring när datamaterialet består av ordnade kategoridata, men metoden är användbar på alla datamaterial som åtminstone har en ordningsstruktur.

Tack vare metodens speciella rangteknik med parbundna ranger så är det möjligt att statistiskt beskriva och analysera den grupprelaterade förändringen separat från den individuellt avvikande från gruppens förändring.

<https://lakartidningen.se/klinik-och-vetenskap-1/2007/02/statistisk-metod-for-parade-ordinaldata/>

<https://avdic.se/svenssonsmetodsvenska>

När vanlig linjär regression haverar

Vad gör man när normalfördelningsantagandet inte håller?

- Logaritmerar utfallsvariabeln?
 - Hjälper inte alltid
 - Noll går inte att logaritmera
- Icke-parametriskt test?
 - Går inte att göra justering

Men: variabler på intervall- och kvotskala är också ordinala!

Våra moderna datorer ger möjligheter att använda ordinal regression även när vi har 1000-tals olika kategorier på vår skala

Istället för ett ojusterat icke-parametrisk test kan vi använda ordinal regression

Fungerar oavsett vilken fördelning vår utfallsvariabel har

RESEARCH ARTICLE

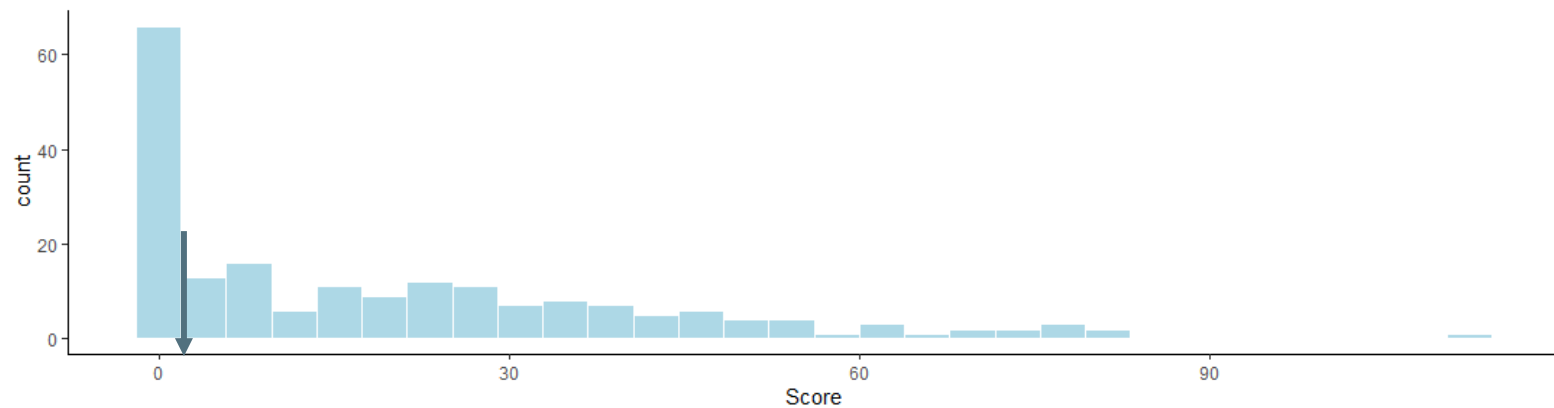
WILEY **Statistics**
in Medicine

Modeling continuous response variables using ordinal regression

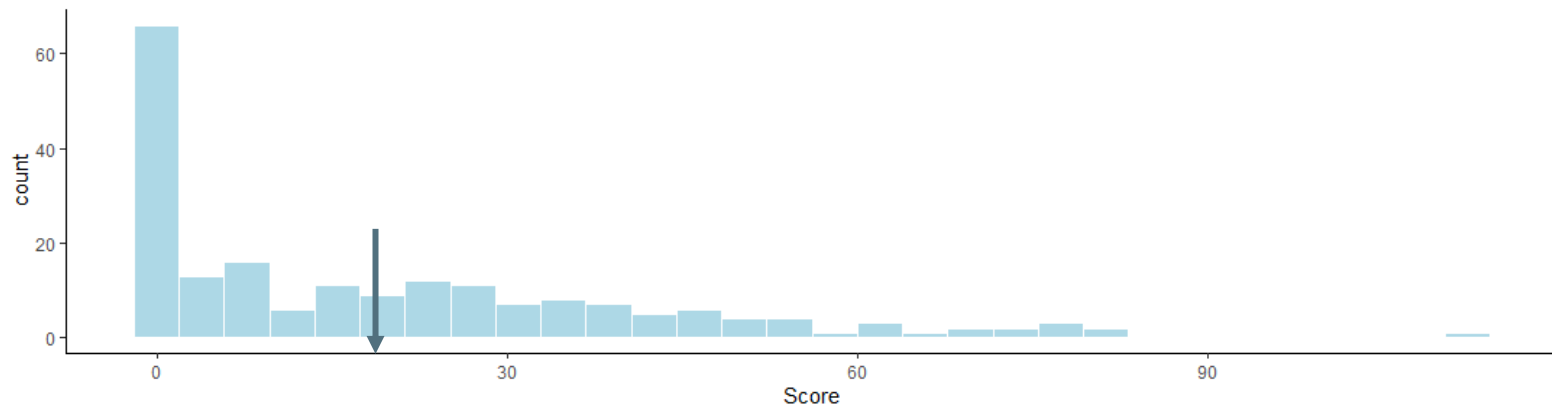
Qi Liu¹  | Bryan E. Shepherd¹ | Chun Li² | Frank E. Harrell Jr.¹ 

Liu, Q., Shepherd, B. E., Li, C., & Harrell Jr, F. E. (2017). Modeling continuous response variables using ordinal regression. *Statistics in medicine*, 36(27), 4316-4335.

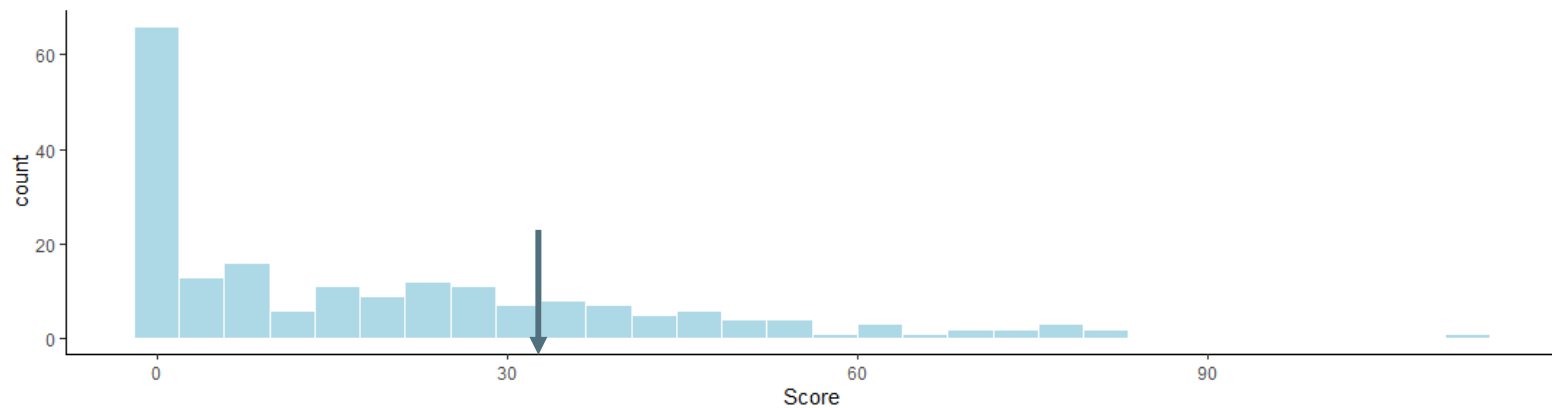
Oddset att man ska skatta högre på skala...



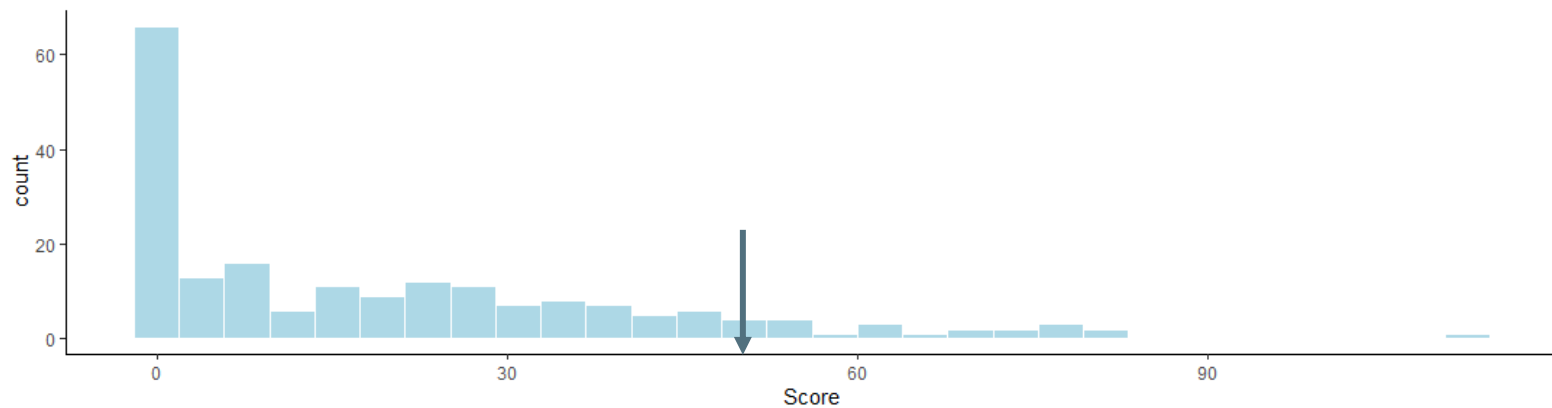
Oddset att man ska skatta högre på skala...



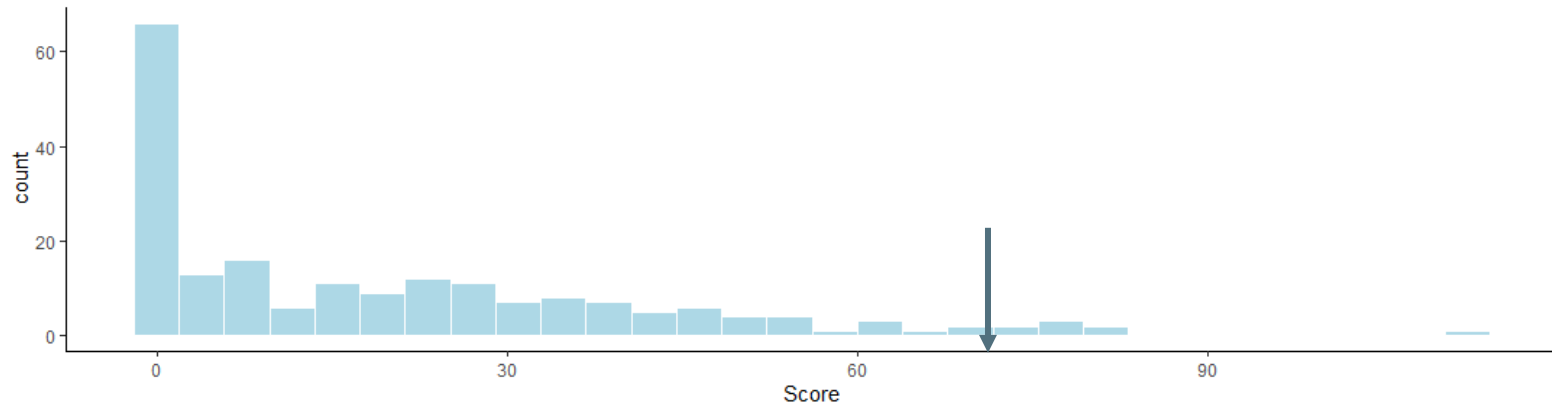
Oddset att man ska skatta högre på skalan...



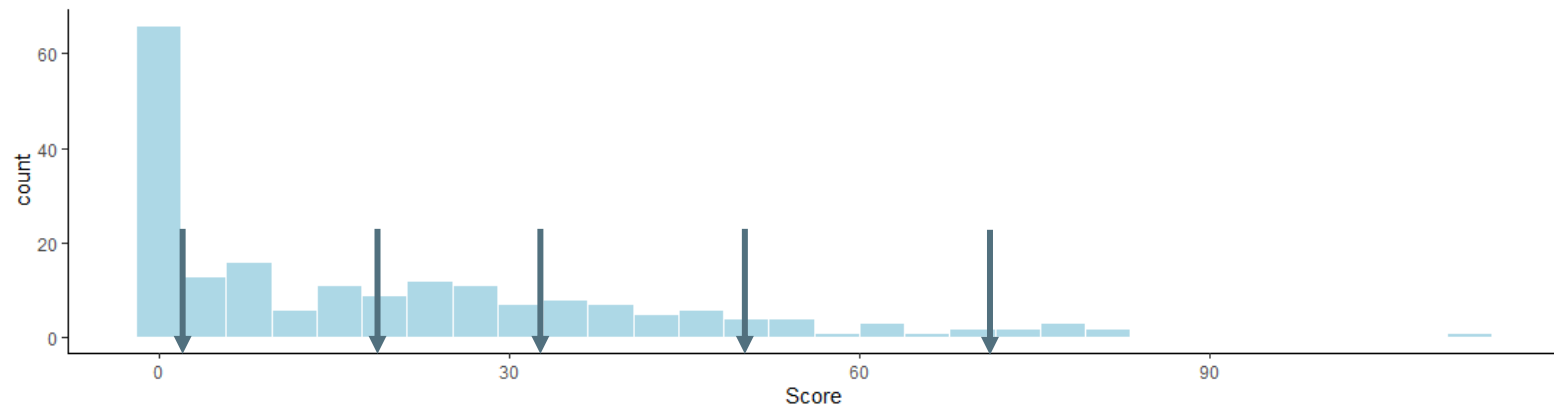
Oddset att man ska skatta högre på skalan...



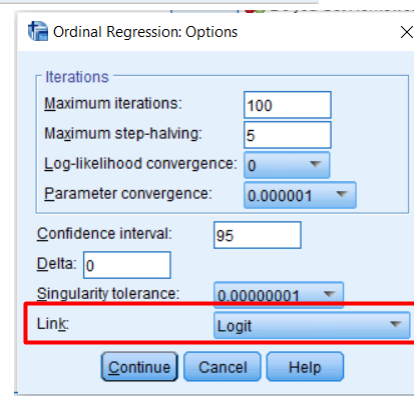
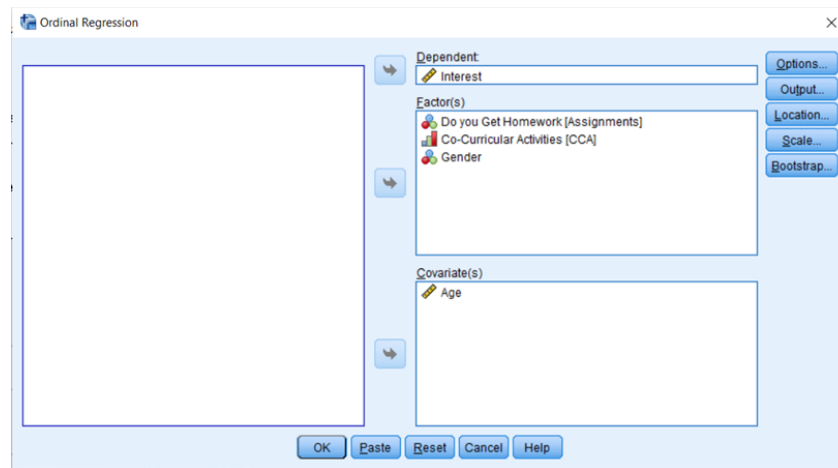
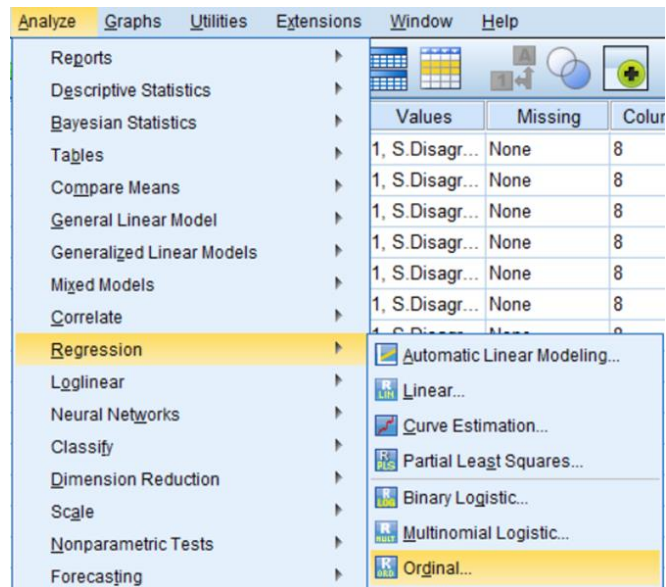
Oddset att man ska skatta högre på skala...



Oddset att man ska skatta högre på skalan...



Ordinal regression i SPSS



Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Interest= 1.00]	4.539	1.520	8.917	1	.003	1.560	7.518
	[Interest= 2.00]	6.427	1.556	17.065	1	.000	3.378	9.476
	[Interest= 3.00]	7.557	1.591	22.562	1	.000	4.439	10.676
	[Interest= 4.00]	9.645	1.658	33.853	1	.000	6.396	12.894
Location	Age	.315	.070	20.444	1	.000	.178	.452
	[Assignments=.00]	-.036	.323	.013	1	.911	-.670	.598
	[Assignments=1.00]	0 ^a	.	.	0	.	.	.
	[CCA=1.00]	-1.330	.461	8.325	1	.004	-2.233	-.426
	[CCA=2.00]	-.100	.349	.082	1	.774	-.784	.584
	[CCA=3.00]	0 ^a	.	.	0	.	.	.
	[Gender=.00]	1.221	.322	14.368	1	.000	.590	1.853
	[Gender=1.00]	0 ^a	.	.	0	.	.	.

Link function: Logit.

a. This parameter is set to zero because it is redundant.

Använd
exponentialfunktionen
för att få oddskvot

Sammanfattning – varför ordinal regression?

Fördelar:

Ordinala eller kontinuerliga utfall

Möjlighet att justera modeller

Inga bekymmer med fördelning

Nackdelar:

Kan vara svårare att förmedla oddskvotens betydelse

Proportionalitetsantagandet

För kontinuerliga variabler:

- Något lägre power
- Kräver mer data



Kliniska Studier
Sverige
Forum Norr

<https://www.canvas.umu.se/courses/2600/>

9/5 kl 12-13 Att strukturera och organisera forskningsdata

Anna Lindam, Region Jämtland Härjedalen/Robert Lundqvist, Region Norrbotten

1/6 kl 12-13 Introduction to Survival Analysis (In English)

Wendy Yi-Ying Wu, Region Västerbotten