

Forskningstræning

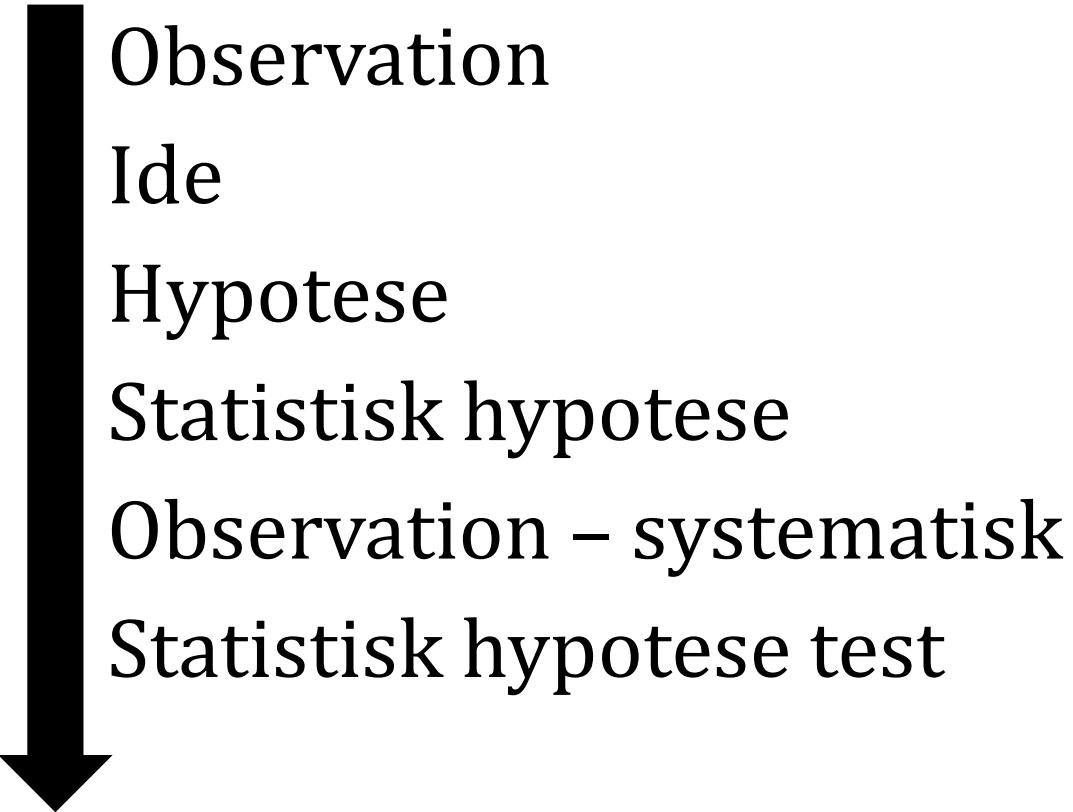
Statistik og forskningsmetodik

2025

Program - Del 1

- Dag 2
 - 9.15-10.00
 - Repetition
 - Regression

Naturvidenskabelig metode



Tese



Statistik skal hjælpe os med at finde ud af, om den forskel vi observerer, er kausalt forbundet med en intervention/proces, eller et resultat af tilfældig variation.



Take-home Forsknings metodologi 1

Ethvert forsøg skal baserer sig på et videnskabeligt solidt grundlag.

- Statistik skal hjælpe os med at finde ud af, om den forskel vi observerer, er kausalt forbundet med en intervention/proces, eller et resultat af tilfældig variation.



Statistik

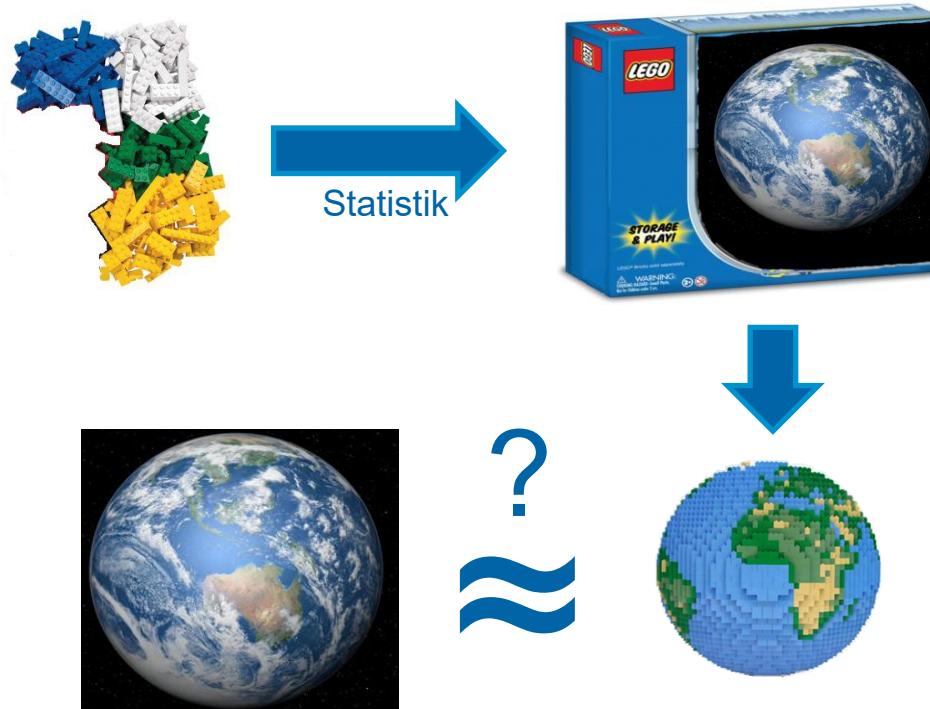
- Median
- Gennemsnit
- Spredning
- SE
- SEM/SD
- Korrelation/sammenhæng
- Tid til hændelse modeller
- Generaliserede liniære modeller
 - Multivariat analyser

Regression

- Regression er en modelering af data.
 - Lineær
 - Cox
 - tid-til hændelse regression

Modeller forklaret med Lego

- Modeller laves på basis af observationer



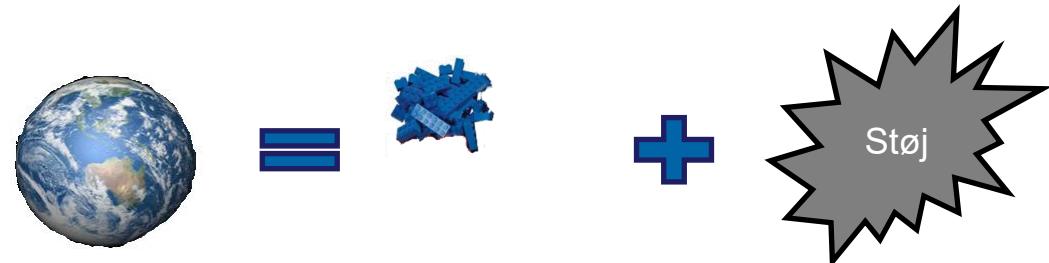
Modeller forklaret med Lego

- Modeller laves på basis af observationer



Modeller forklaret med Lego

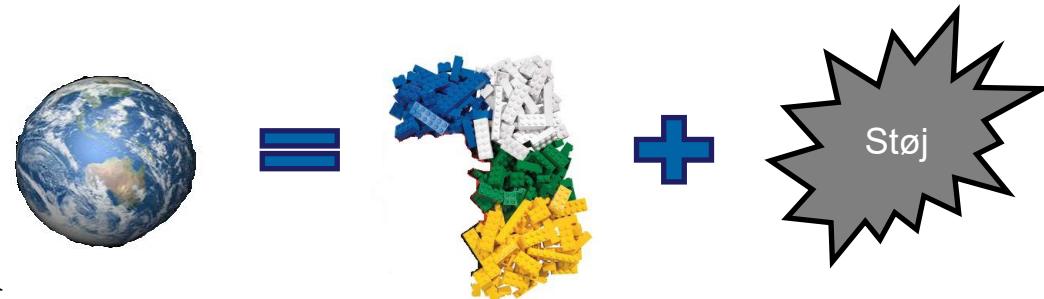
- Forklare variationer i data
 - **Outcome = kendt faktor + støj**



- Univariat
 - Forklarer typisk kun lidt variation
 - Selvom p-værdien er lille

Modeller forklaret med Lego

- Forklare variationer i data
 - **Outcome = kendt faktor + støj**



- Multivariat model
 - Overfitting
 - Kan være vanskelige at tolke
 - Komplekse sammenhænge
 - Hvilke variabler og hvor mange?

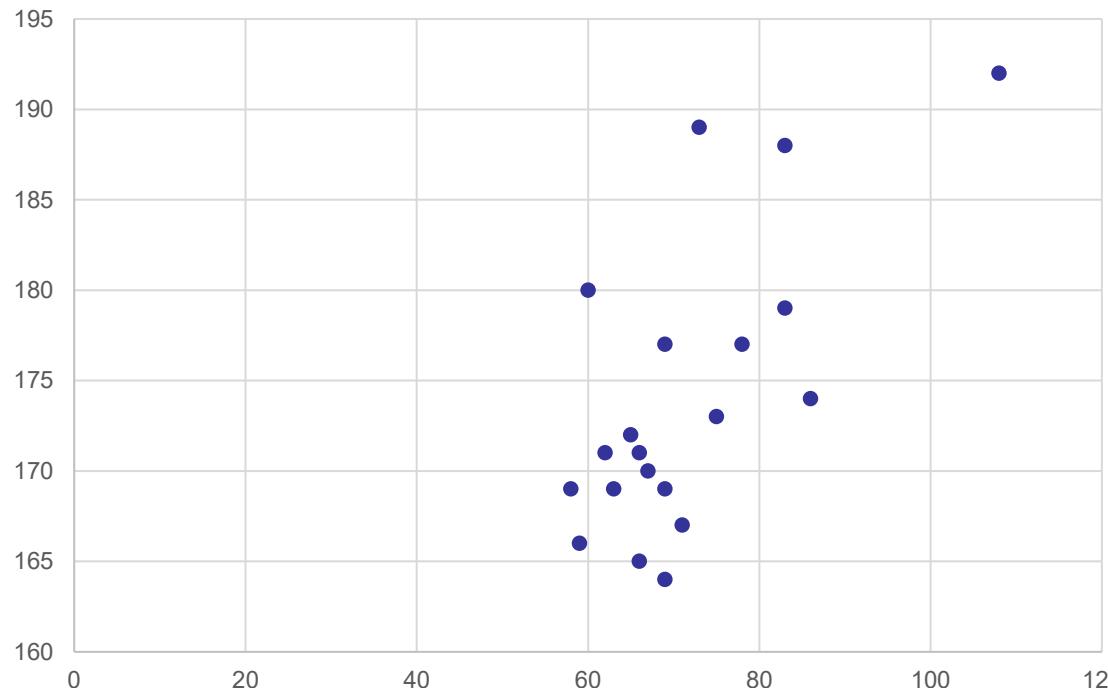
Lineær regression

- Vi undersøger sammenhængen mellem
 - højde og vægt
- Tegn en graf over højde/vægt fra data

ID	Height	Gender	Weight	Blond
1	169	0	58	0
2	171	0	62	1
3	167	0	71	0
4	170	0	67	1
5	177	0	78	0
6	164	0	69	1
7	165	0	66	0
8	169	0	63	1
9	171	0	66	0
10	169	0	69	1
11	180	1	60	0
12	172	1	65	1
13	174	1	86	0
14	179	1	83	1
15	188	1	83	0
16	173	1	75	1
17	177	1	69	0
18	189	1	73	1
19	166	1	59	0
20	192	1	108	1

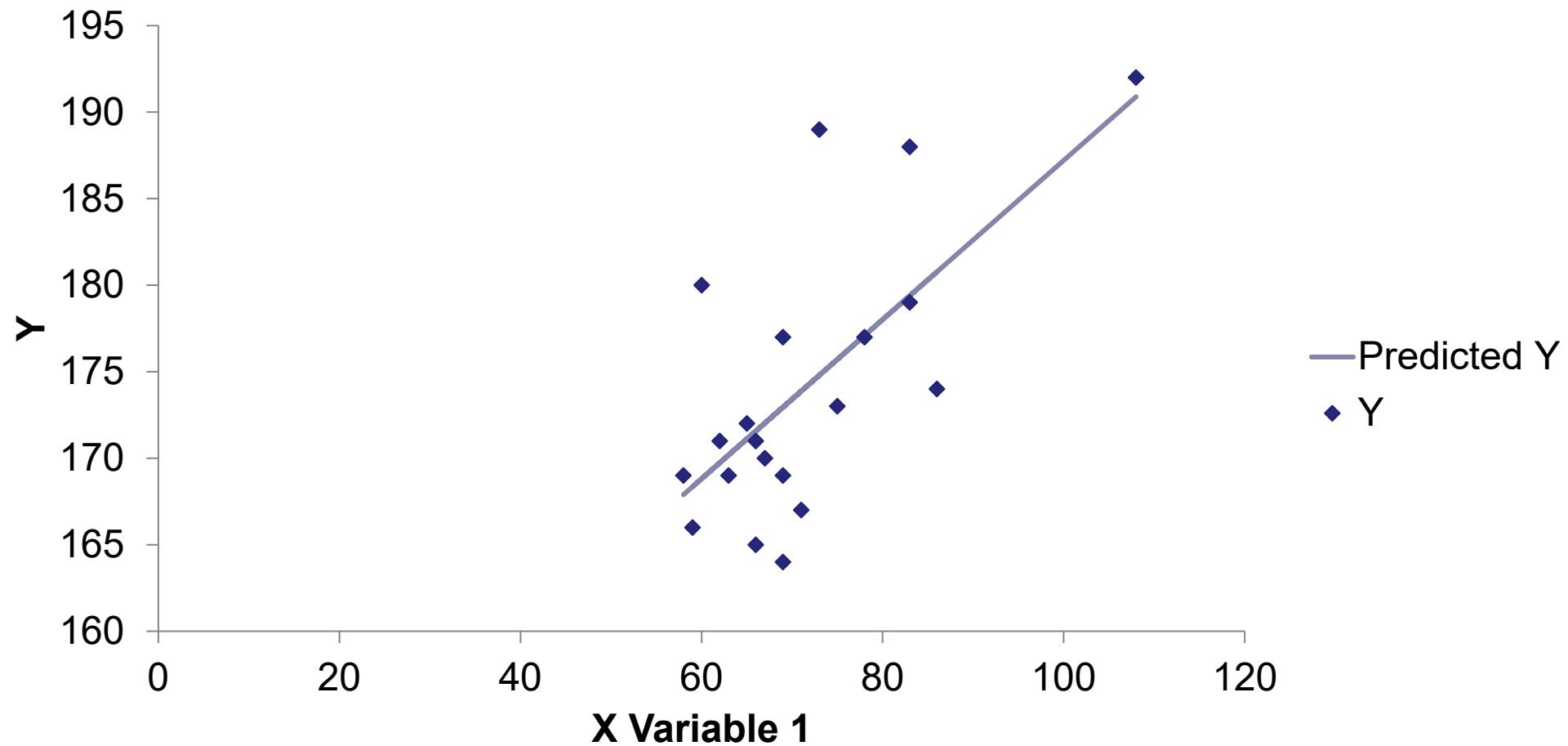
Lineær regression

- Tegn en graf over højde/vægt fra data

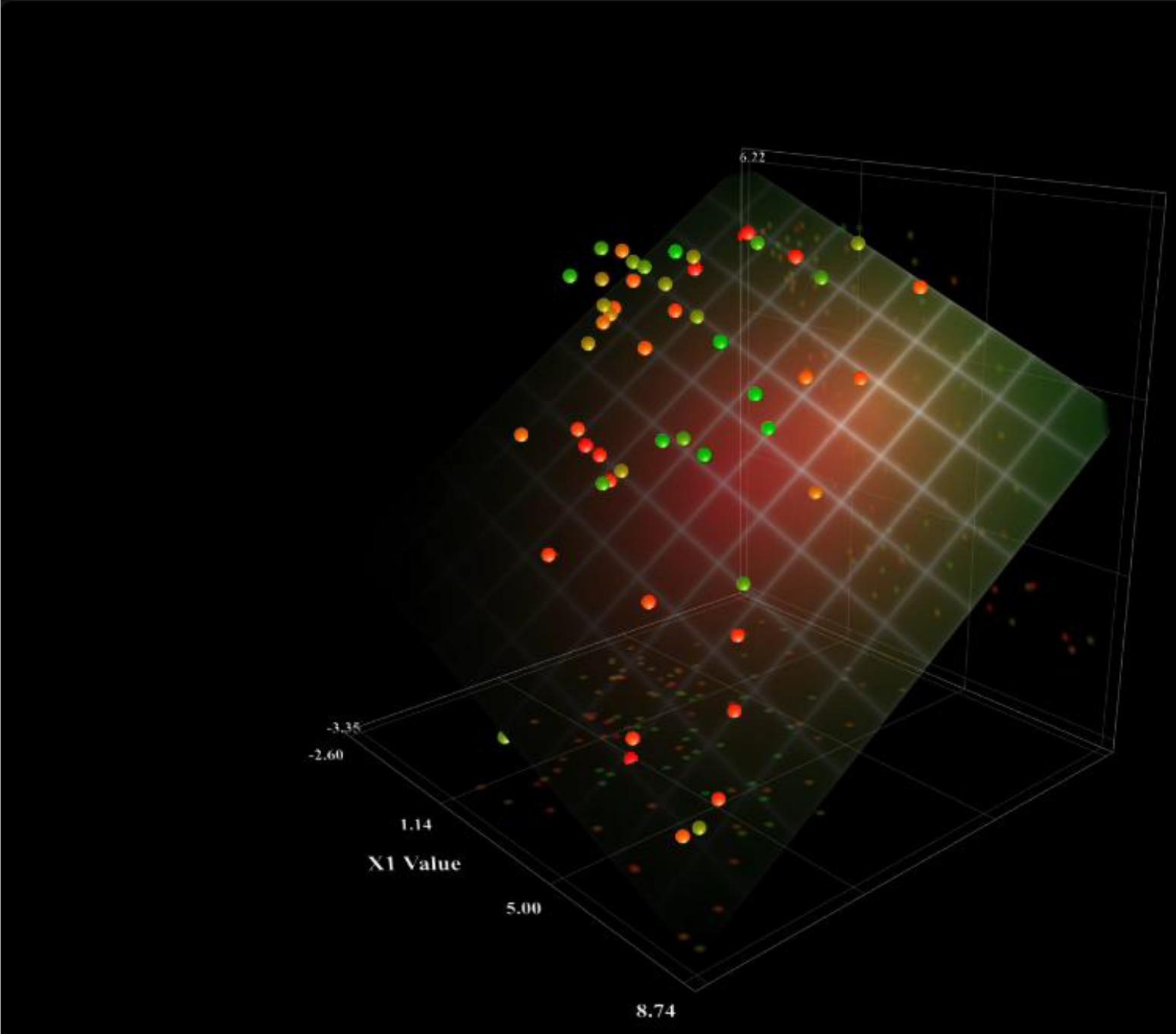


ID	Height	Gender	Weight	Blond
1	169	0	58	0
2	171	0	62	1
3	167	0	71	0
4	170	0	67	1
5	177	0	78	0
6	164	0	69	1
7	165	0	66	0
8	169	0	63	1
9	171	0	66	0
10	169	0	69	1
11	180	1	60	0
12	172	1	65	1
13	174	1	86	0
14	179	1	83	1
15	188	1	83	0
16	173	1	75	1
17	177	1	69	0
18	189	1	73	1
19	166	1	59	0
20	192	1	108	1

- Hvor den samlede afstand er mindst



Multivariabel regression

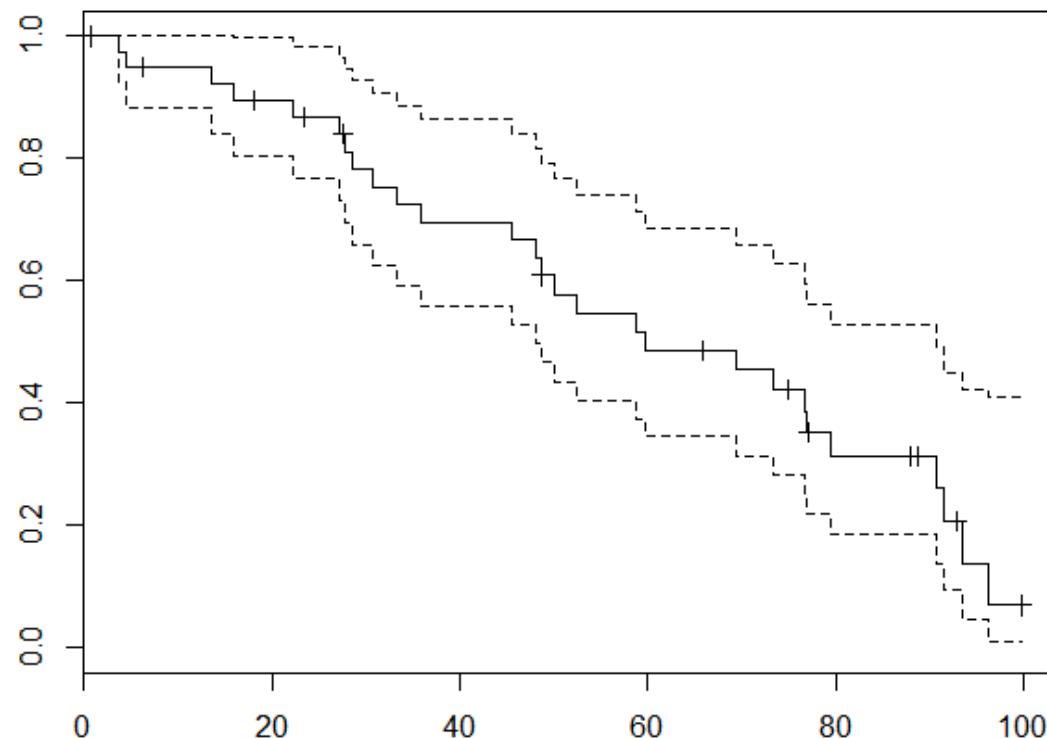


Take-home

- Regression
 - De fleste anvender en generaliseret liniær model (GLM)
 - Minimerer afstanden af observation til de forventede.
 - Residualer
 - Modellen vurderes ud fra residualerne

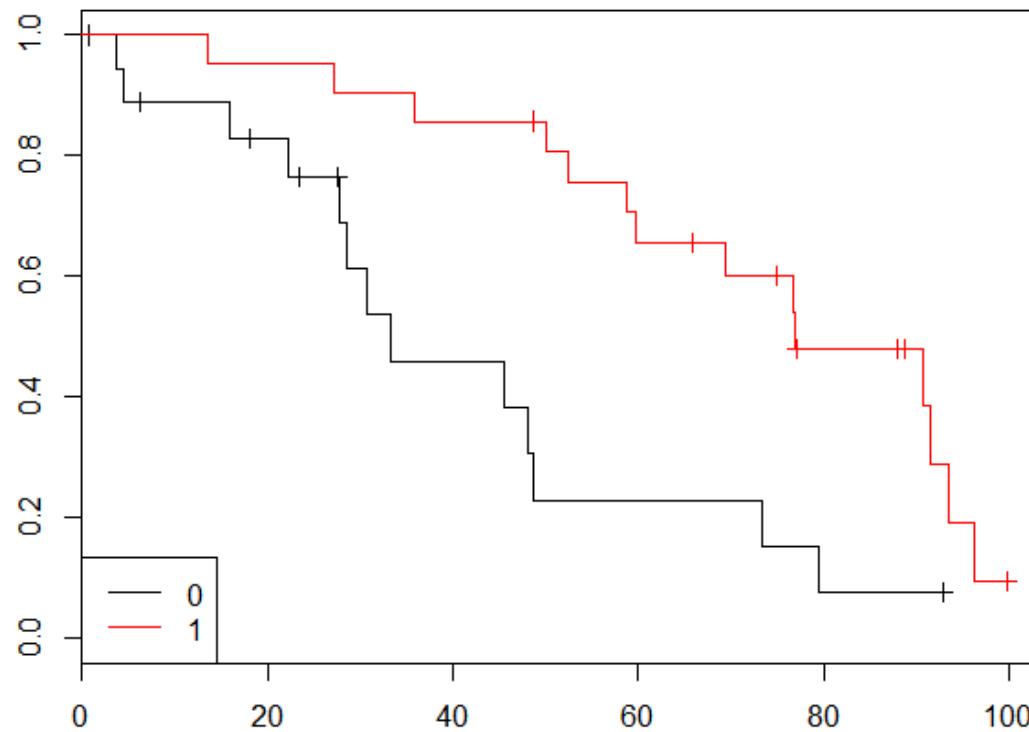
Regression

- Cox-regression
 - Risiko for at dø per tidsenhed
 - Hazard
 - Gruppe mod Gruppe
 - Ratio



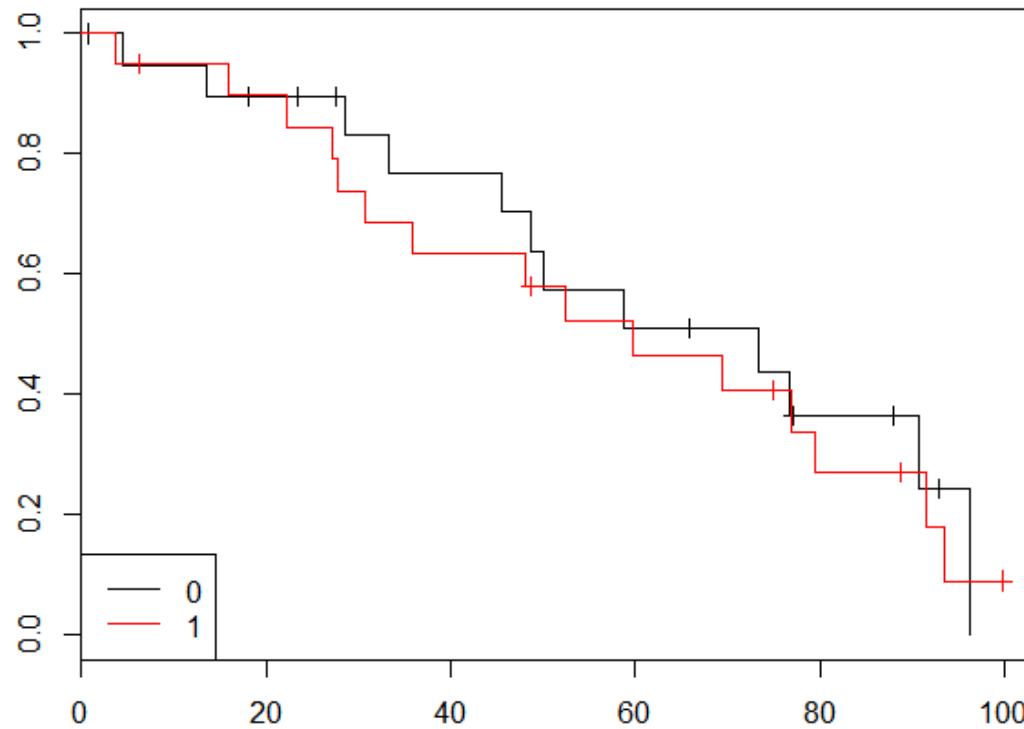
Cox-regression

- Factor 0
 - Hazard ratio:
 - HR: 0,31



Cox-regression

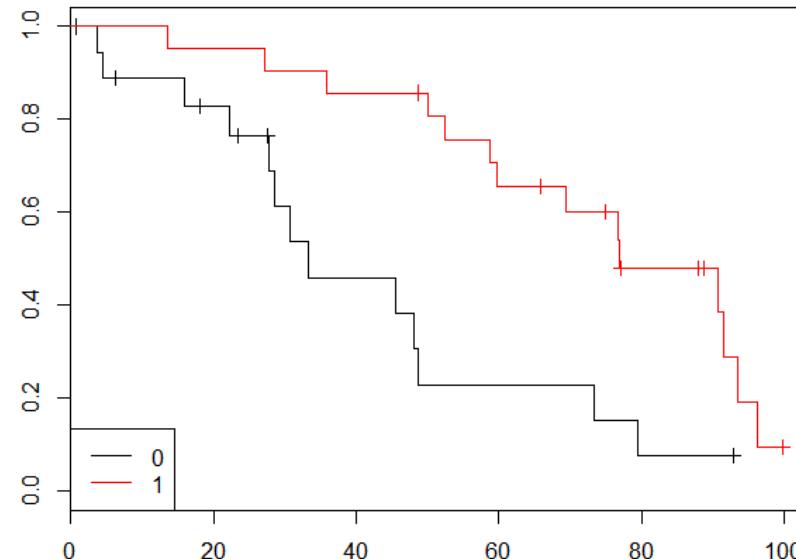
- Factor 1
 - Hazard ratio:
 - HR: 1,1



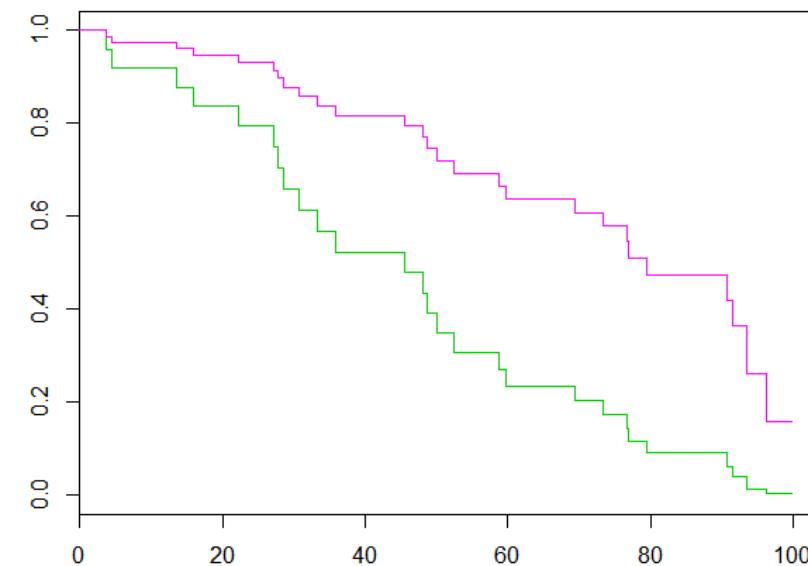
Factor0

Model vs. Observeret

Observeret



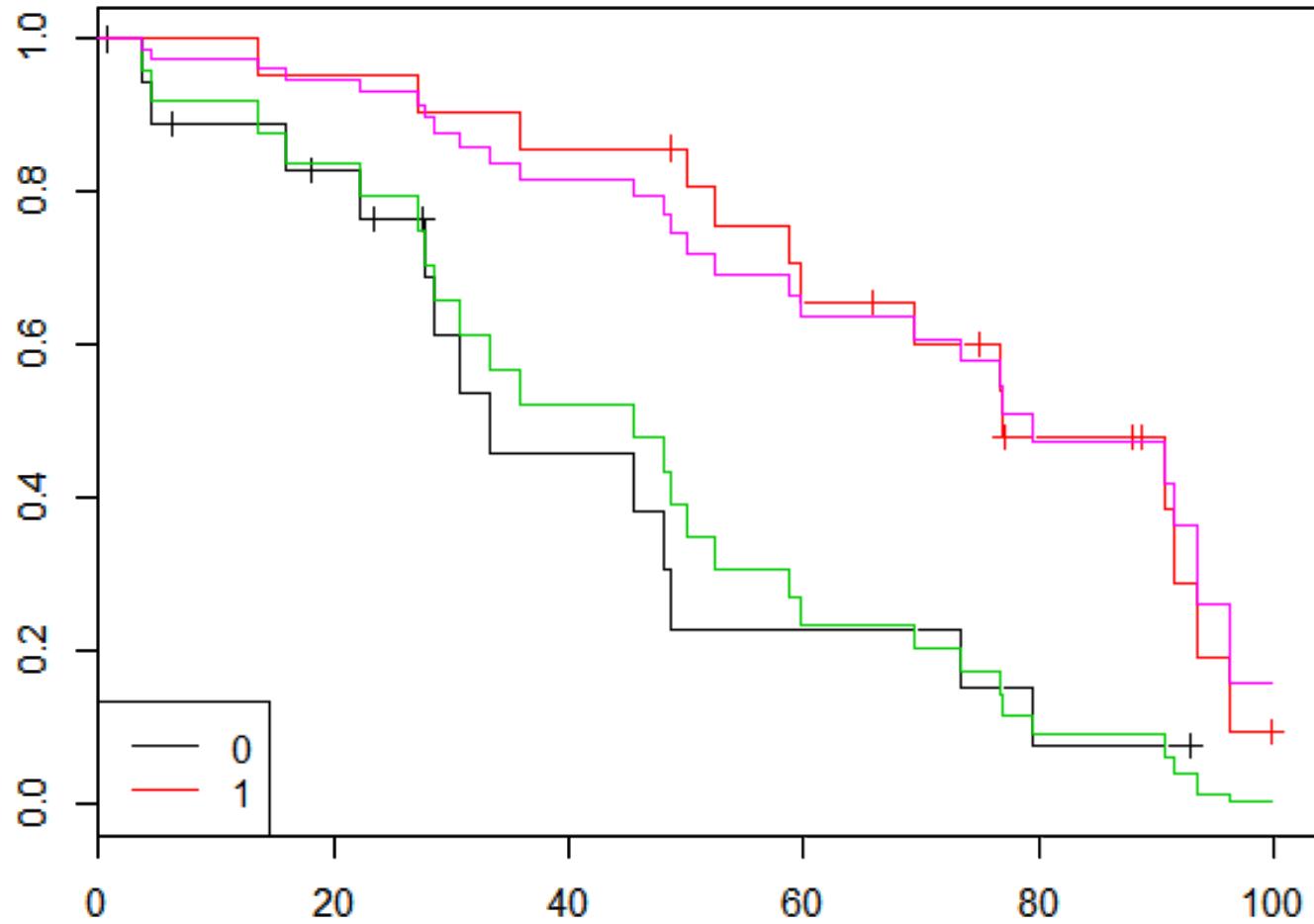
Model
HR: 0,33



Hvad er forskellen?

Factor0

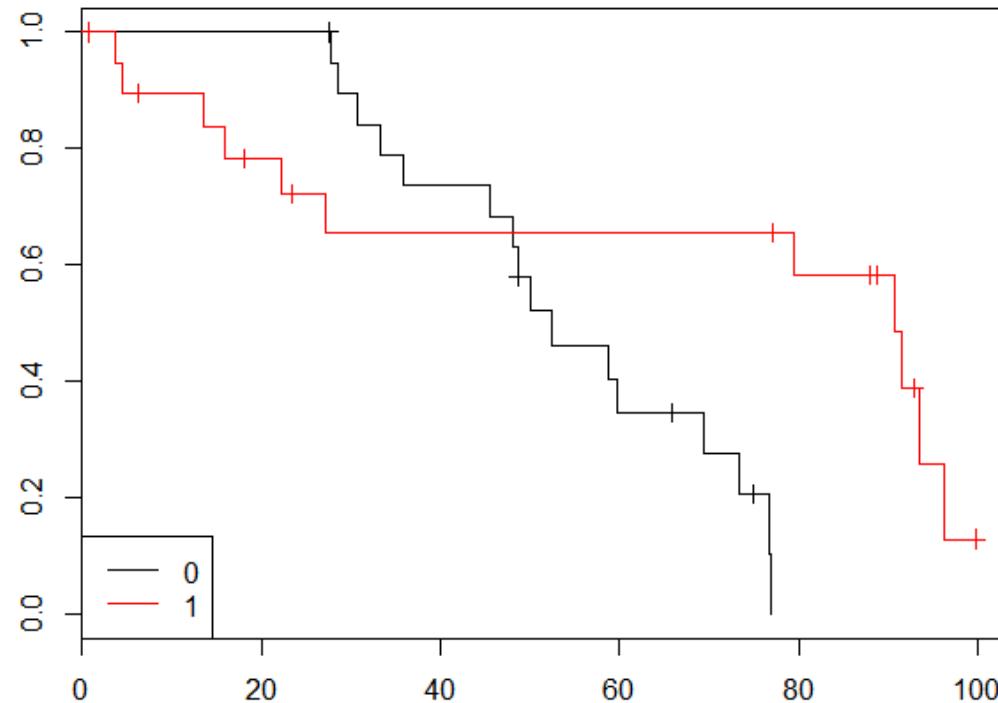
Model vs. Observeret



Cox-regression

- Udfordringer
 - Tolk ratioet

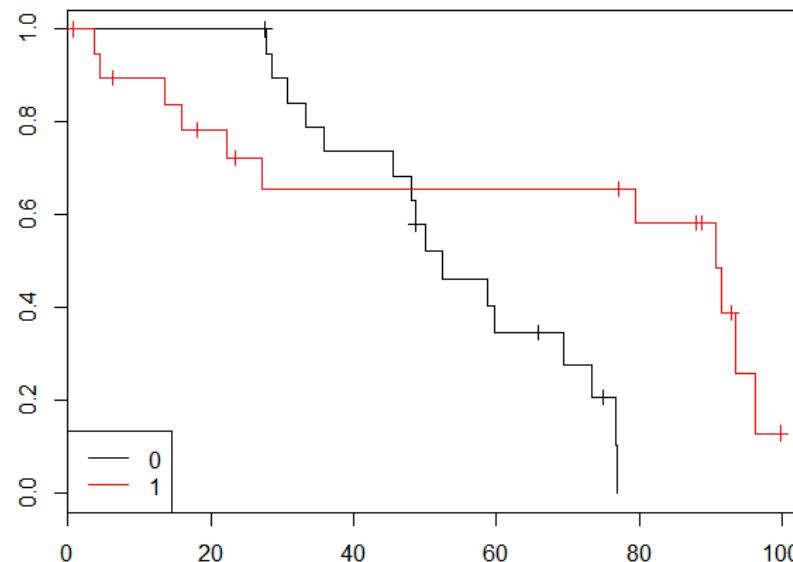
HR: 0,33 (0,12-0,9)



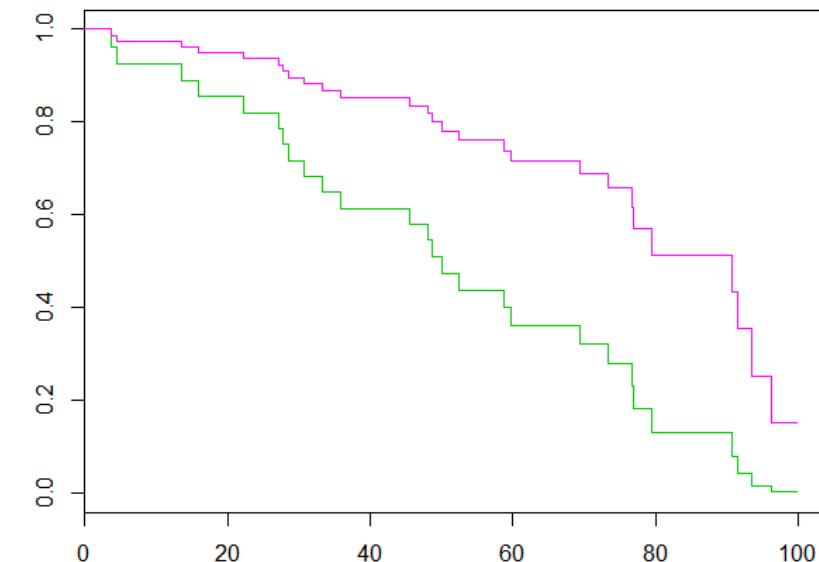
Factor2

Model vs. Observeret

Observeret

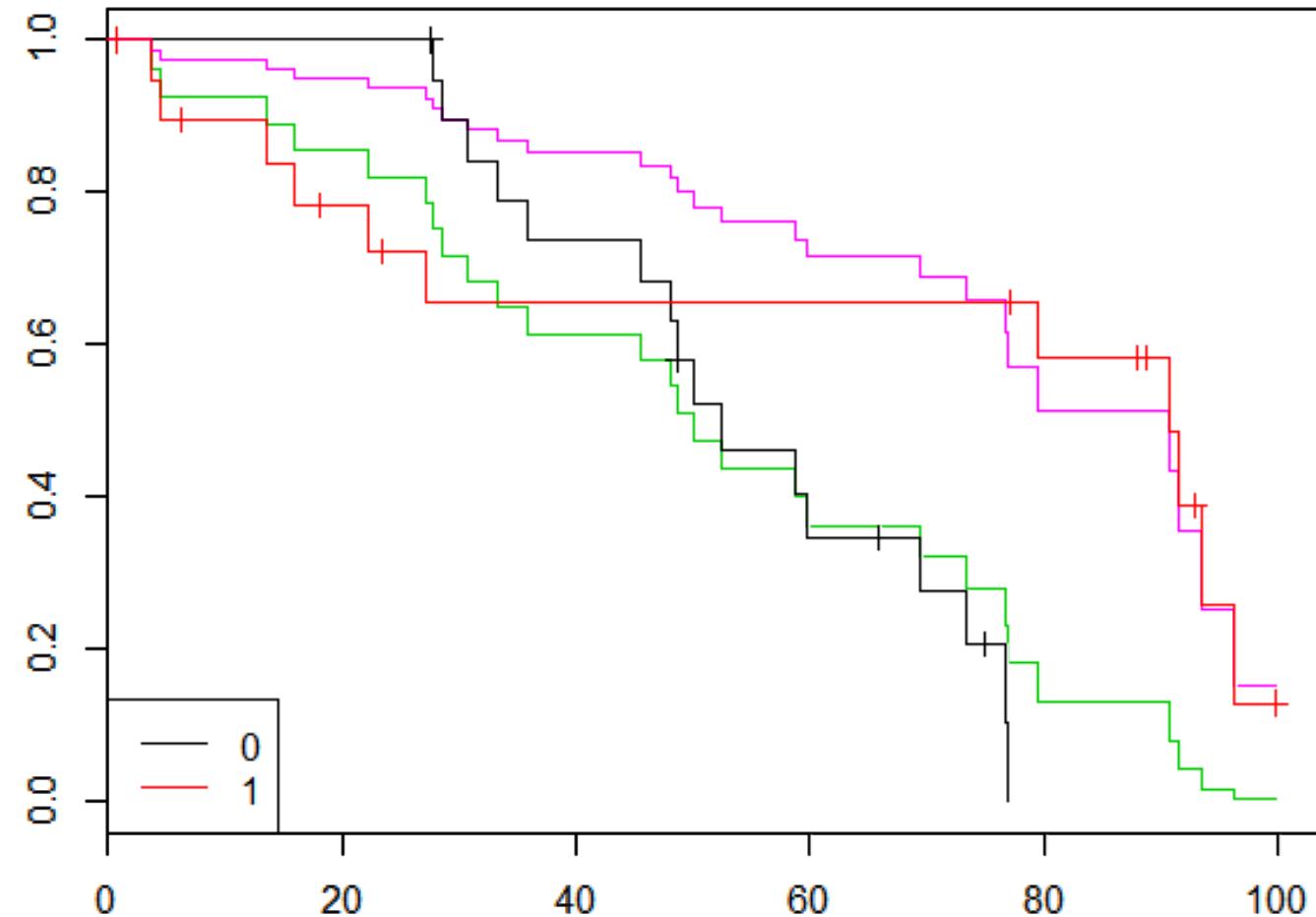


Model
HR: 0,33



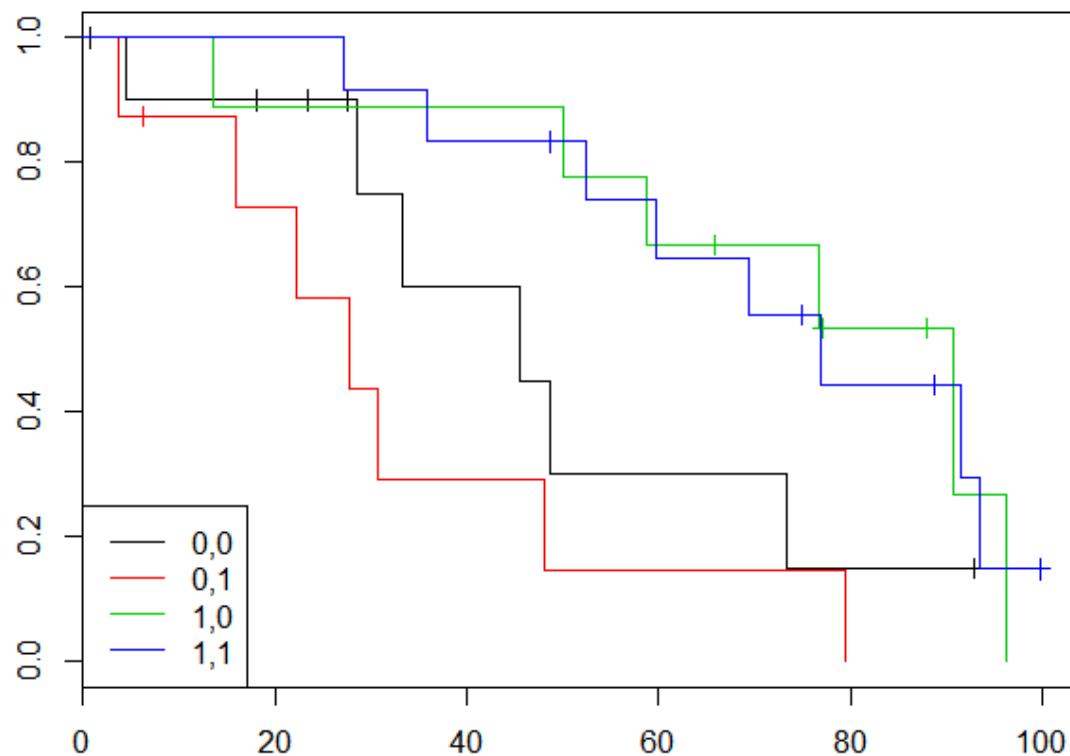
Factor2

Model vs. Observeret



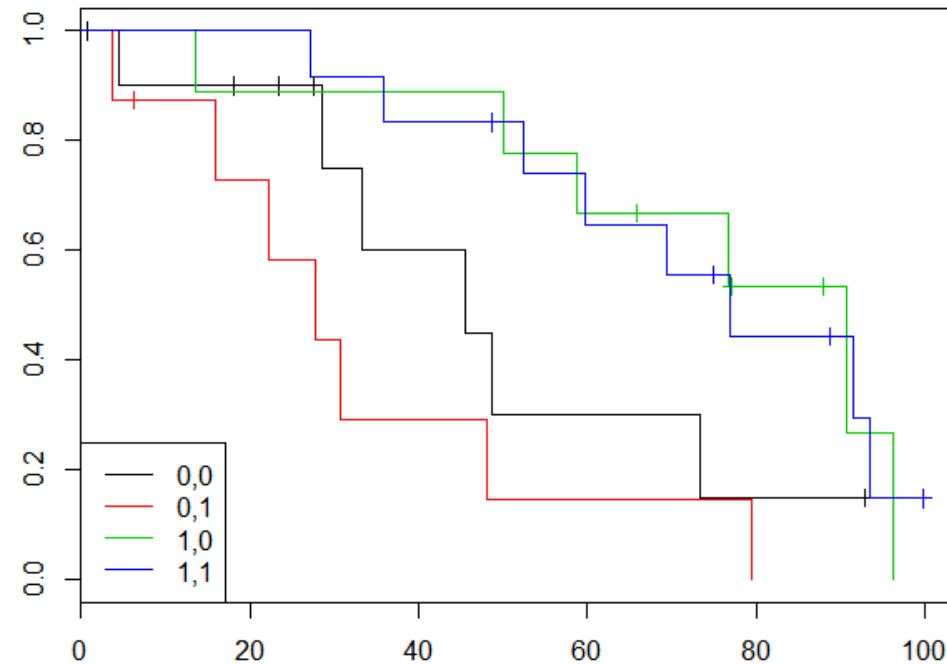
Flere Grupper

- Factor 0 + Factor 1



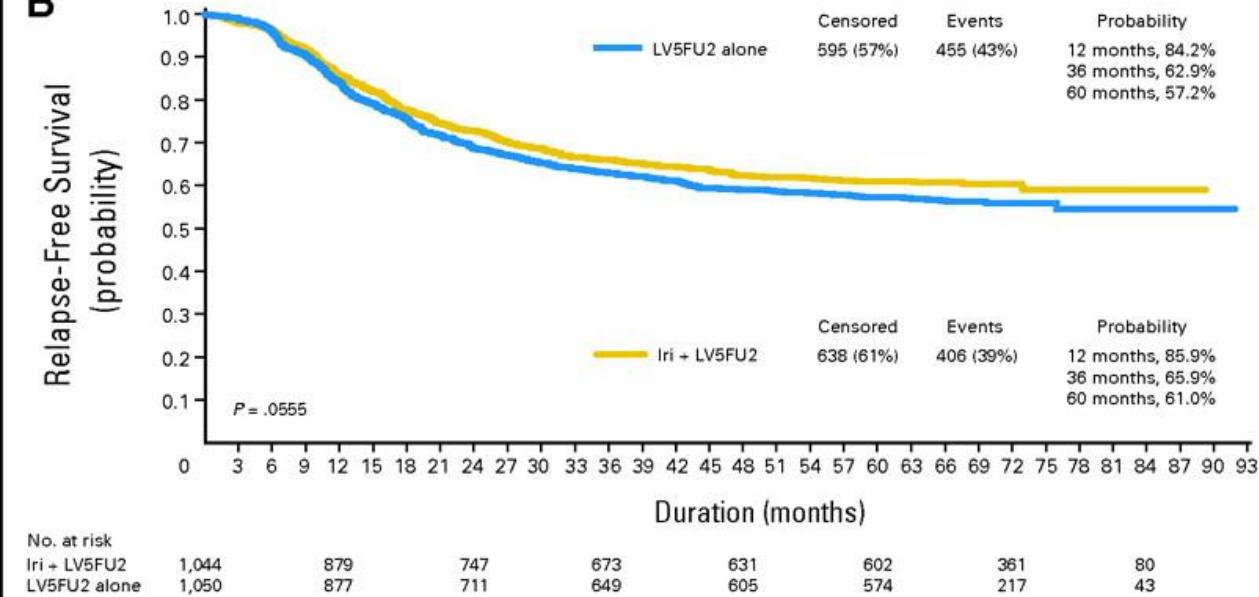
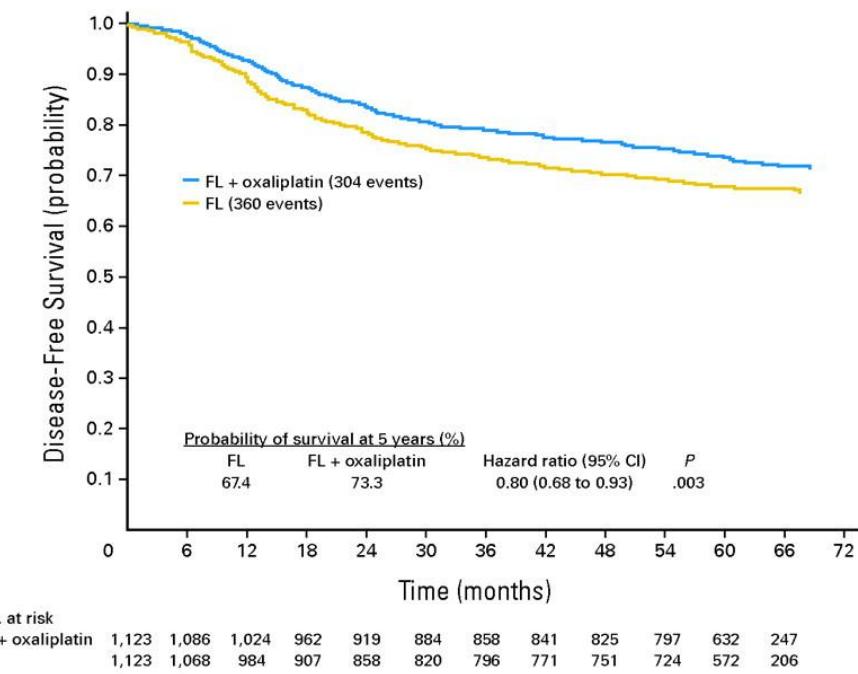
Flere Grupper

- Teknik
 - Sætte Factor til dens gennemsnit
 - Bestemmer
 - Factor0
 - Factor1
 - Hver for sig
- Factor 1&2
 - Skal være uafhængige
 - Vi tager interaktioner på del 2



Event definitioner

	OS	DFS	RFS	TTR	CSS	C-CSS
anastomotic relapse	red	green	green			light green
metastatic relapse	red	green	green		light green	light green
2 nd primary CRC	red	light green	pink	pink	pink	light green
2 nd primary non-colorectal cancer	red	white	red	red	red	red
death with evidence of recurrence	green	green	green	green	green	green
death without evidence of recurrence	green	light green	pink	red	red	red
death related to protocol treatment	green	green	white	pink	light green	white
death from colon cancer cause	green	green	green	green	green	green
death from non-colon cancer cause	green	light green	pink	red	pink	red
death from unknown cause	green	light green	white	pink	white	pink
death from any cause	green	light green	pink	red	pink	pink
lost of follow-up	pink	pink	red	red	red	pink

B**A**

Status

- Du skulle nu gerne have en forståelse for:
 - Stikprøver
 - Variation
 - Styrke-beregninger
 - t-test
 - Chi2 test
 - Log-rank test
 - Konceptet
 - Observeret vs. Forventet
 - Regression
 - Liniær
 - Tid-til-hændelse
 - Basal multivariat analyse

Studie designs/tests

- Korrelationer/sammenhæng
- Sensitivitet/Specifitet/AUC analyser
- Sammenligninger
 - Kontinuerlige
 - Kategoriske
- Tid-til-event statistik
- Regressioner
 - Logistisk
 - Cox

Osv.

Valg af statistisk metode

- Hvad er response?
 - Overlevelse
 - Rask/syg
 - Lyser lidt/lyser meget
 - Sammenhæng mellem 2 undersøgelser
 - Ændring

A photograph of a forest floor covered in moss and fallen leaves. In the background, tall, thin trees stand in a dense cluster. A large tree trunk is visible on the right side.

Dyb indånding

Ingen Panik

Datatyper

- Kontinuerlige variabler
 - 1,2,3,..... ∞
- Kategoriske variabler
 - Let, Moderat, Svær

Standard error of the mean

- Et mål for præcisionen

$$SEM = \frac{SD}{\sqrt{n}}$$

Dyb indånding

Ingen Panik

Konsekvenser

- Variation i måling
 - Stor
 - SEM højere
 - Lille
 - SEM mindre
- Antal patienter
 - Stort
 - SEM falder
 - Lille
 - SEM stiger

Take-home

- Normafordeling er et centralt princip
- Fordelingen kan beskrives udfra
 - Gennemsnit og spredning(SD)
- En stikprøve bruges til at estimere den "ægte" værdi
 - Stikprøvens præcision øges ved at lave den større

Kategoriske variabler

- Anvendes ved ikke kontinuerlige data.
 - Eksempler
 - Performance status
 - TNM
 - Køn
 - Beskrives ofte som en fraktion/procent

Test

- Vi vil gerne vide om mænd har højere stadie end kvinder ved diagnosen.

	T1-T2	T3-T4	
Mand	134	93	
Kvinde	32	14	

Test

- Vi vil gerne vide om mænd har højere stadie end kvinder ved diagnosen.
 - 2x2 table
 - Chi² test

	T1-T2	T3-T4	
Mand	134	93	227
Kvinde	32	14	46
	166	107	273

Baggrund for test

- Observeret vs. Forventet
- Hvis to grupper er ens vil fordelingen af variablen også være ens.
 - Hvis fordelingen er anderledes end forventet
 - Efter kompenstation for usikkerhed
 - Bestemt af størrelsen på stikprøven
 - Er der en signifikant forskel

Test

- Vi vil gerne vide om mænd har højere stadie end kvinder ved diagnosen.
 - 2x2 tabel
 - Er det en forskel?
 - Chi² test

	T1-T2	T3-T4	
Mand			
Kvinde			
	166	107	273

Test

- Vi vil gerne vide om mænd har højere stadie end kvinder ved diagnosen.
 - 2x2 tabel
 - Er det en forskel?
 - Chi² test

	T1-T2	T3-T4	
Mand			
Kvinde			
	61%	39%	273

Test

- Vi vil gerne vide om mænd har højere stadie end kvinder ved diagnosen.
 - 2x2 tabel
 - Er det en forskel?
 - Chi² test

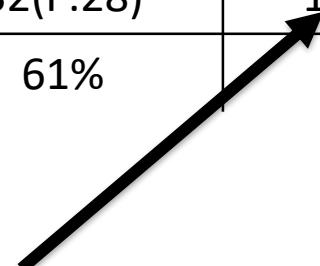
	T1-T2	T3-T4	
Mand	138	88	227
Kvinde	28	18	46
	61%	39%	273

Forventet hvis der var ens fordeling af T1-2 og T3-4

Test

- Vi vil gerne vide om mænd har højere stadie end kvinder ved diagnosen.
 - 2x2 tabel
 - Er det en forskel?
 - Chi² test

	T1-T2	T3-T4	
Mand	134(F:138)	93(F:88)	227
Kvinde	32(F:28)	14(F:18)	46
	61%	39%	273



Under 10 så Fischer's eksakte test

To ting

- En forventning til hvordan verden opfører sig
 - Hypotesen
- Statistik skal hjælpe os med at finde ud af, om den forskel vi observerer, er kausalt forbundet med en intervention/proces, eller et resultat af tilfældig variation.
 - Nul-hypotesen

