

Interaktioner og modelspecifikation

Metode 3

Link til slides: kortlink.dk/2acdr

Søren Damsbo-Svendsen

sdas@ifs.ku.dk

Institut for Statskundskab

Københavns Universitet

Uge 9

Recap fra sidste gang

- (Sproglige) repræsentationer er væsentlige for politik!
- Diskurser er (stabile og ustabile) strukturer, der organiserer og skaber mening og identitet via links og differentieringer - "*a system for the formation of statements*"
- I diskursanalyse bevæger vi os "metodisk" fra grundig nærlæsning af tekster til afdækning af diskurser og deres repræsentation af identiteter og politikker
- Vi mødte "Diana", hvis identitet (links og differentieringer) varierede markant afhængigt af det tidslige rum (i Uganda, på asylcentret i DK, efter mødet med LGBT Asylum i DK) og mellem de to tekster (debatindlæg og forvaltningsafgørelse)

Oversigt

Uge	Holdtime	Emne	Øvelsesopgave
6		Kvantitativ indholdsanalyse	1
7		Diskursanalyse I	2
8	25-26 februar 2021	Diskursanalyse II	3
9	04-05 marts 2021	Interaktioner og modelspecifikation	4
10	11-12 marts 2021	Logistisk regression	5
11	18-19 marts 2021	Multilevel analyse	6
12		Kausal inferens I: Kausalitet og instrumentvariable (IV)	7
13		Påskeferie	
14	08-09 april 2021	Kausal inferens II: Paneldata	8
15	15-16 april 2021	Kausal inferens III: Eksperimentelle designs	9
16		Kausal inferens IIII: Regression Discontinuity (RD)	10
17		Social Data Science: Big Data	11
18		Process tracing	12
19		Kriterier for god videnskab	
20		Opsamling/spørgetime	
23		Aflevering af skriftlig hjemmeopgave (07 juni 2021)	

Dagens program og formål

Dagens program

1. Recap på ugens forelæsning om interaktion
2. Masser af tid til praktiske eksempler og øvelsesopgaver

Dagens formål

- At forstå, hvad man bruger en **interaktionsmodel** til
- Hvilke typer spørgsmål hjælper den med at besvare?
- At lære, hvordan man **laver** en interaktionsmodel i Stata
- At træne os i at **fortolke** interaktionsmodeller pba. Stata output

Side note: Timen hedder "og modelspecifikation", fordi interaktionsanalyse foretages ved at specificere sin regressionsmodel på en bestemt måde

Pensum

- Kellstedt, P. M., Whitten, G. D. (2018). The Fundamentals of Political Science Research. 3rd edition. Cambridge University Press. Kapitel 11.3
- Sønderskov, Kim Mannemar (2014). Stata - en praktisk introduktion (2. udg.). København: Hans Reitzels forlag. Kap. 10.

Lidt indledende begrebsafklaring

- **Interaktion:** *X interagerer med Z i sin påvirkning af Y*
- **Moderering:** Effekten af X på Y *modereres af Z*
- **Betingede effekter:** Effekten af X på Y er betinget af (afhænger af) Z
- **Heterogene effekter:** Effekten af X på Y er *heterogen på tværs af Z* (de værdier Z antager - variationsbredden) \neq homogen

De beskriver alle det samme: Vi undersøger X og Z's fælles indflydelse på Y (matematisk set multiplicerer vi de to variable, X og Z)

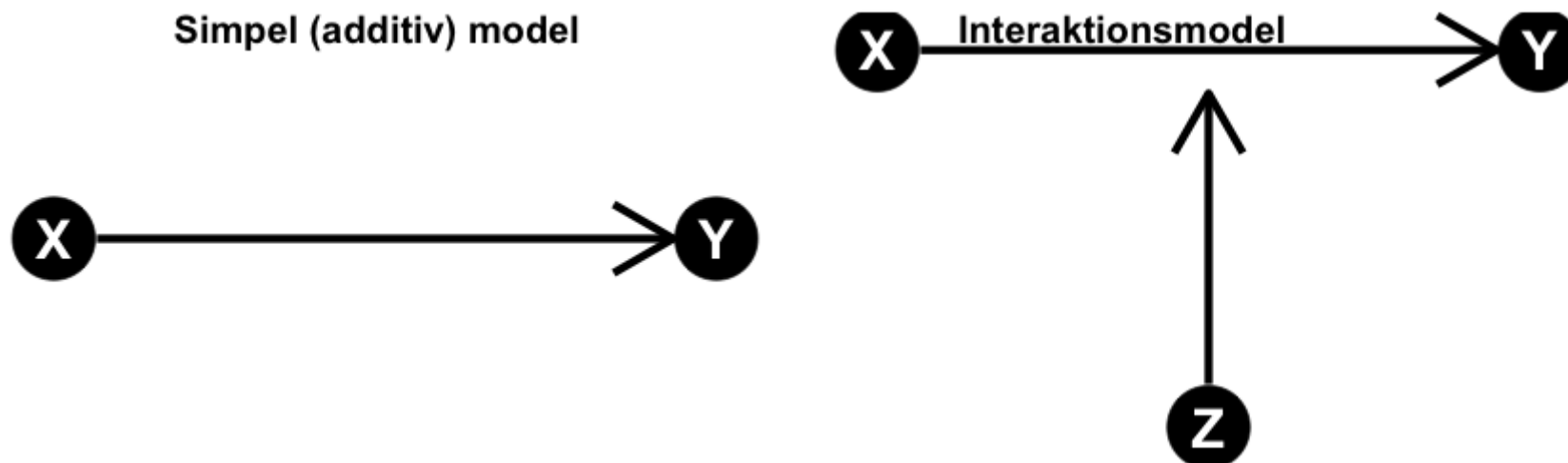
- Derudover en lille reminder om variables måleniveauer:
 - intervalskaleret \approx kontinuert \approx metrisk $\{1,2,3,4,5,6,7,8,9,10,\dots\}$
 - binær \approx dikotom \approx dummy $\{0,1\}$
 - kategorisk \approx faktorvariabel $\{"a","b","c"\}$

Lidt indledende begrebsafklaring

Kausalitet

- Ordet **"effekt"** implicerer kausalitet → at en uafhængig variabel kausalt forårsager en ændring i en afhængig variabel.
- Når vi snakker interaktion i dag bruger vi ordet "effekt" i flæng uden rigtigt at forholde os til, om kriterierne for kausalitet er opfyldt (ofte er de ikke). Mao. skyder vi spørgsmål om kausalitet lidt til hjørne og antager simpelthen at **"X påvirker Y"** - for bedre at kunne lære om interaktion med **Z**
- Vi vender tilbage til spørgsmålet om kausalitet senere

Et første blik på interaktion



- Indtil videre har I lært (med lineær regression) at undersøge, hvordan en afhængig variabel (Y) påvirkes af en eller flere uafhængige variable (X). Dette kaldes en "**additiv model**"
- Men ofte afhænger effekten af X på Y af en tredjevariabel Z. Dette undersøges med en **interaktiv model**
 - Eksempel: Corona-udbetalte feriepenge (X) øger fattige personers (Z = fattig) privatforbrug (Y), men påvirker ikke rige personers (Z = rig) forbrug (Y)
 - Så siger man, at formue (Z) **modererer** effekten af $X \rightarrow Y$
 - Og der er **heterogene effekter** i den forstand, at X påvirker Y forskelligt afhængigt af Z (ikke kun én homogen effekt)

Kort diskussion

- Kort break-out (2 min.)
- Hvad er en interaktion(sanalyse)?
- Hvad er særligt svært at forstå rigtigt ved interaktion (uagtet om I faktisk har forstået det)?

Additive modeller

- En lineær regressionsmodel uden interaktion kaldes en "additiv model", fordi de uafhængige variables indflydelse blot lægges sammen (+) for at forklare den afhængige variabel:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon$$

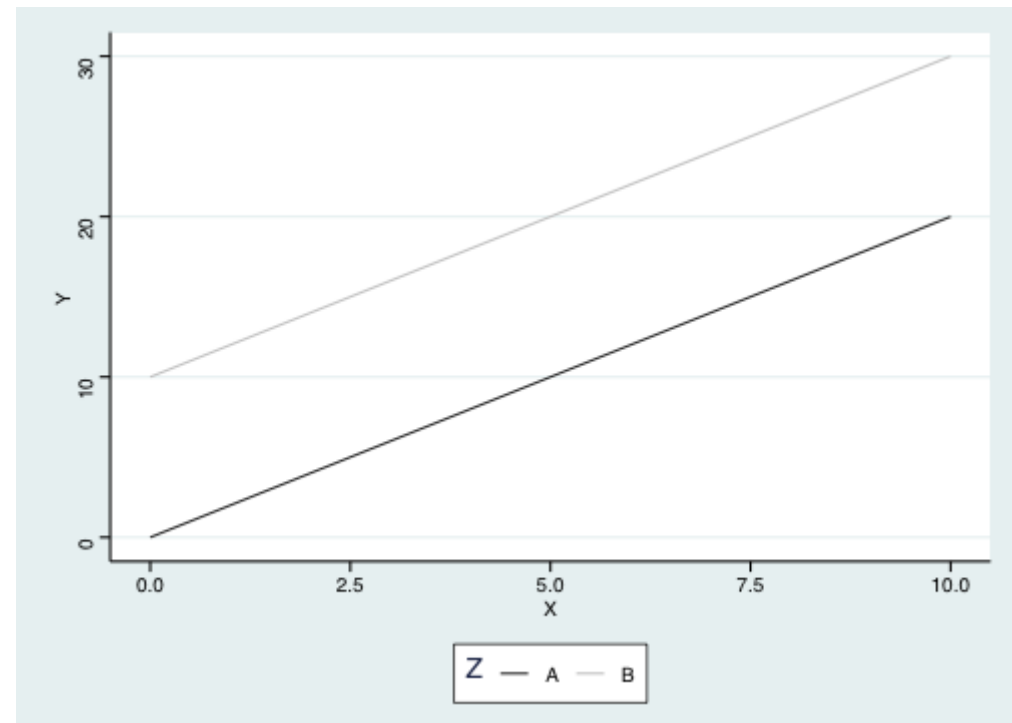
- Antagelse: Effekten af X på Y er konstant (effekten af X_1 er β_1) uanset hvilke værdier de øvrige uafhængige variable i modellen har
- Med andre ord estimeres en **gennemsnitseffekt** af X for alle enheder, f.eks. den gennemsnitlige effekt af feriepenge (X) på privatforbrug (Y) på tværs af rige og fattige (Z)

Additive modeller

Hvis vi har en model med to uafhængige variable, X og Z, kunne den additive model se således ud:

$$Y = \alpha + \beta_1 X + \beta_2 Z + \epsilon$$

- Vi har kontrolleret for Z, hvilket gør, at Y *ikke* er det samme i alle Z-grupper (pga. β_2), men
- **effekten af X på Y** (dvs. $\beta_1 \rightarrow$ hældningen!) **er den samme for begge grupper**, dvs. en gennemsnitseffekt



Interaktive modeller

"Interaktion refererer til en situation, hvor størrelsen på effekten af en variabel [X] på en anden variabel [Y] afhænger af en tredje variabel [Z] ... Z modererer eller betinger X's effekt, og at X og Z interagerer"

-tilpasset fra Sønderskov (2014), s. 233

Hvorfor interaktion?

- Når effekten af X interagerer med en tredjevariabel Z, kan det være **misvisende at se på gennemsnitlige effekter** (dvs. lave alm. lineær regression uden interaktion → additiv model)
- Derudover gør interaktionsanalyse os i stand til at besvare **andre (vigtige) spørgsmål/hypoteser**:
 - modererer Z mon $X \rightarrow Y$?
 - afhænger $X \rightarrow Y$ af Z?
 - er X's påvirkning af Y forskellig for forskellige niveauer af Z?

Matematikken

Tre variable X , Y og $Z \rightarrow$ vi kan bruge eksemplet med feriepenge (X), forbrug (Y) og formue (Z)

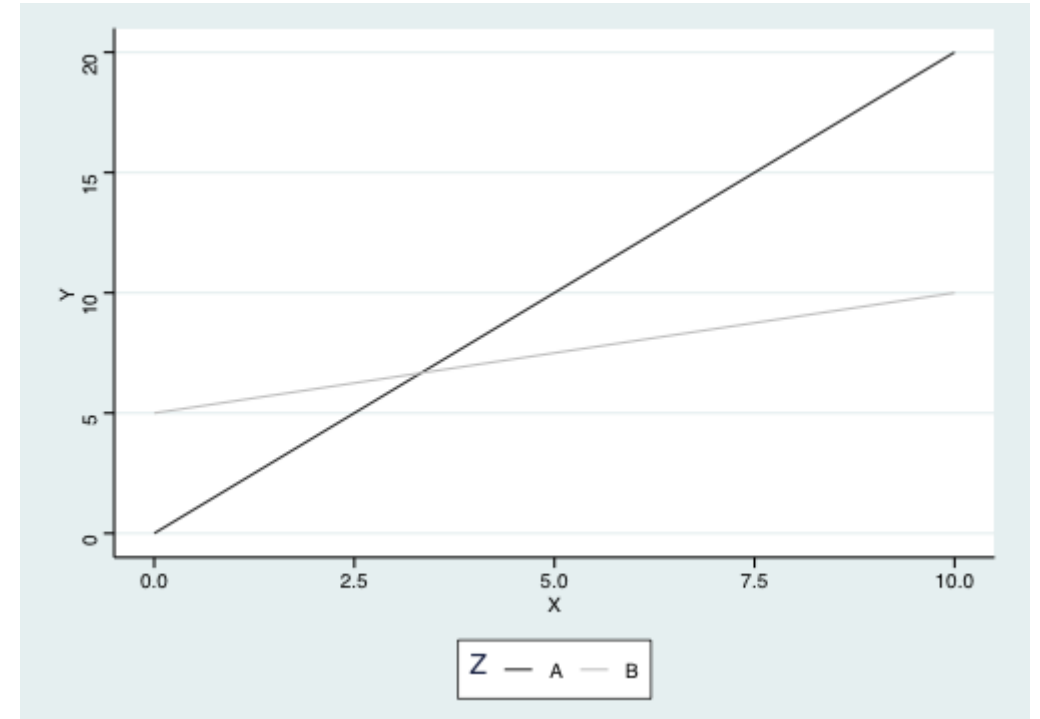
Type	Model
Additiv model	$Y = \alpha + \beta_1 X + \beta_2 Z$
Interaktionsmodel	$Y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 XZ$

- **Hvad er forskellen?**
- Hvad er konstituerende led (hovedled) og hvad er interaktionsled?
- Vi kommer tilbage til fortolkning om lidt

Varierende hældning = varierende effekt

Interaktiv model

- Hældningen for de to grupper varierer:
 - Hældningen for $Z = \text{"A"}$ er 2
 - Hældningen for $Z = \text{"B"}$ er 0.5
- Hældningen for $Z = \text{"B"}$ kan også udtrykkes som forskellen fra A, dvs. **-1.5**, fordi $0.5 - 2 = -1.5$
- Det er sådan Stata ville præsentere interaktionseffekten for os
- I Stata tester vi, om der statistisk signifikant forskel på hældningerne 2 og 0.5 → om differencen (-1.5) er signifikant forskellig fra nul



Hvordan laver man interaktion i Stata?



Hvordan laver man interaktion i Stata?

Interaktive modeller **implementeres nemt i Stata** i form af en mindre tilføjelse til den lineære regression

Interaktion mellem to uafhængige variable specificeres ved hjælp **#** a la:

```
regress y c.x i.z c.x#i.z
```

... hvilket vi ofte forkorter til:

```
regress y c.x##i.z
```

- I eksemplet er x kontinuert (derfor "c.") og z er kategorisk (derfor "i.")
- Når man bruger **##** i stedet for **#** inkluderer Stata automatisk de konstitutive led (hovedleddene), dvs. *c.x* og *i.z*, foruden interaktionsleddet (andenordensleddet) *c.x#i.z*

Hvordan laver man interaktion i Stata?

Når man laver interaktion er det særligt vigtigt at være opmærksom på - og specificere! - variablenes måleniveauer:

- Kategorisk variabel eller dummy: **i.variabelnavn**
- Kontinuert (intervalskaleret) variabel: **c.variabelnavn**
- Man kan aldrig være for eksplicit med brug af "i." og "c."
- Måleniveauer er afgørende for fortolkningen → det ser vi i øvelsesopgaverne om lidt

Det giver altså:

- Når X er kontinuert og Z er kategorisk/binær: **reg y c.x##i.z**
- Når både X og Z er kontinuert: **reg y c.x##c.z**
- Når både X og Z er kategorisk/binær: **reg y i.x##i.z**

Hvorfor ## og ikke bare # i Stata?

x##z giver os automatisk hovedled *og* interaktionsled, men hvorfor ikke bare køre med # (uden hovedled)?

Fordi det i mange tilfælde (afhængigt af måleniveau) implicit antager, at begge hovedled er lig med 0, hvilket:

- kan medføre et biased estimat af interaktionen
- med sikkerhed medfører forvirring og en anden tolkning
- sjældent er nyttigt

Derfor → inkluder hovedled og brug ##

Fortolkning af interaktionsleddet

Når vi fortolker et regressionsoutput med interaktion er det væsentlige at fortolke koefficienten for interaktionsleddet:

- **Fortegn:** Retning som hovedeffekten ($X \rightarrow Y$) bevæger sig, når den modererende variabel (Z) ændres med $+1 \rightarrow$ *større eller mindre?*
- **Størrelse:** *Hvor meget* hovedeffekten ($X \rightarrow Y$) ændrer sig, når den modererende variabel (Z) vokser med $+1$
- **Beta-koefficienten for X-hovedleddet:** Hvordan påvirker $X \rightarrow Y$, når den modererende variabel er lige præcis $Z = 0$ - kan ses som "udgangspunktet" for effekten
 - Det er en **marginal effekt** \rightarrow effekten for en bestemt gruppe
 - Har ikke altid en meningsfuld tolkning (Stata og *margins* hjælper os om lidt)
- **Beta-koefficienten for interaktionsleddet ($X * Z$):** Den "ekstra effekt", der lægges til hovedleddets koefficient (eller trækkes fra) for at få de **marginale effekter af $X \rightarrow Y$** for andre niveauer af Z end 0

Fortolkning af interaktionsleddet

Den konkrete fortolkning afhænger af variablenes måleniveau:

- Hvis Z er **kontinueret**: *Hvor meget vokser effektstørrelsen, når Z vokser med +1?*
- Hvis Z er **binær**: *Hvad er forskellen i effektstørrelse mellem de to Z-grupper?*
- Hvis Z er **kategorisk** med mere end to kategorier:
 - Stata indsætter automatisk en serie af (antal kategorier - 1) binære variable, som sammenligner hvert niveau af Z med det laveste niveau af Z (referencekategorien)
 - Der er i princippet et interaktionsled ($X * Z$) for hver af disse "sammenligninger"
 - Interaktions-koefficienterne giver her forskellene i effektstørrelse mellem Z-gruppe 1 og 2, Z-gruppe 1 og 3, Z-gruppe 1 og 4, osv.
- **Den mest tricky variant - by far! - er, når både X og Z er kategoriske.** Det følger fuldstændigt samme princip, så det er altså ikke umuligt. Men tungen skal holdes lige i munden pga. mere end én "hovedeffekt" ($X \rightarrow Y$), som så tilmed varierer forskelligt for forskellige niveauer af Z. Den springer vi over for nu.

Marginale effekter

- Hvad er marginale effekter?
- Når vi beregner den betingede effekt af $X \rightarrow Y$ for forskellige niveauer af Z , taler vi om **marginale effekter**
- Kan forstås som effekten ($X \rightarrow Y$) for forskellige grupper eller typer af observationer
- Beregnes i **Stata**, efter vi har kørt regressionen, ved hjælp af *margins*. Den nyttigste kommando er:

```
margins, dydx(x) at(z = (1 2 3 4 5 6 7 8 9 10))
```

... hvor X er vores kausale variabel, Z er en modererende tredjevariabel og 1-10 er de niveauer af Z , som vi ønsker at beregne den **marginale effekt** af $X \rightarrow Y$ for

- **at(z = (1 2 3 4 5 6 7 8 9 10))** kan også skrives som **at(z = (1/10))** eller **at(z = (1(1)10))**
- Vær varsom med at *ekstrapolere* marginale effekter til subjekter, der ikke findes (fx en -200-årig)

Det hjælper altid at **visualisere de marginale effekter med marginsplot** (efter *margins*)

Eksempler

Problemstilling og hypotese er styrende for fortolkning af interaktion

1. Modererer kvaliteten af statslige institutioner (Z) ulandsbistands (X) effekt på økonomisk vækst (Y)?
(Sønderskov, 2014: 234ff)
2. Does the effect of national identification (X) on political trust (Y1) and participation (Y2) differ between natives and immigrants (Z)? (Rapp, 2018)
3. Min egen forskning: **Temperaturer (X) påvirker klimaholdninger (Y), men afhænger effekten af (Z1) uddannelsesniveau, (Z2) partitilhørsforhold eller (Z3) nyhedsforbrug?** (Damsbo-Svendsen, 2020)



Temperaturer (X) påvirker klimaholdninger (Y), men **afhænger effekten af (Z1) uddannelsesniveau, (Z2) partitilhørsforhold eller (Z3) nyhedsforbrug?**

Table 7: The full Local Warming Effect model where temperature is interacted with level of education, party identification, and news exposure.

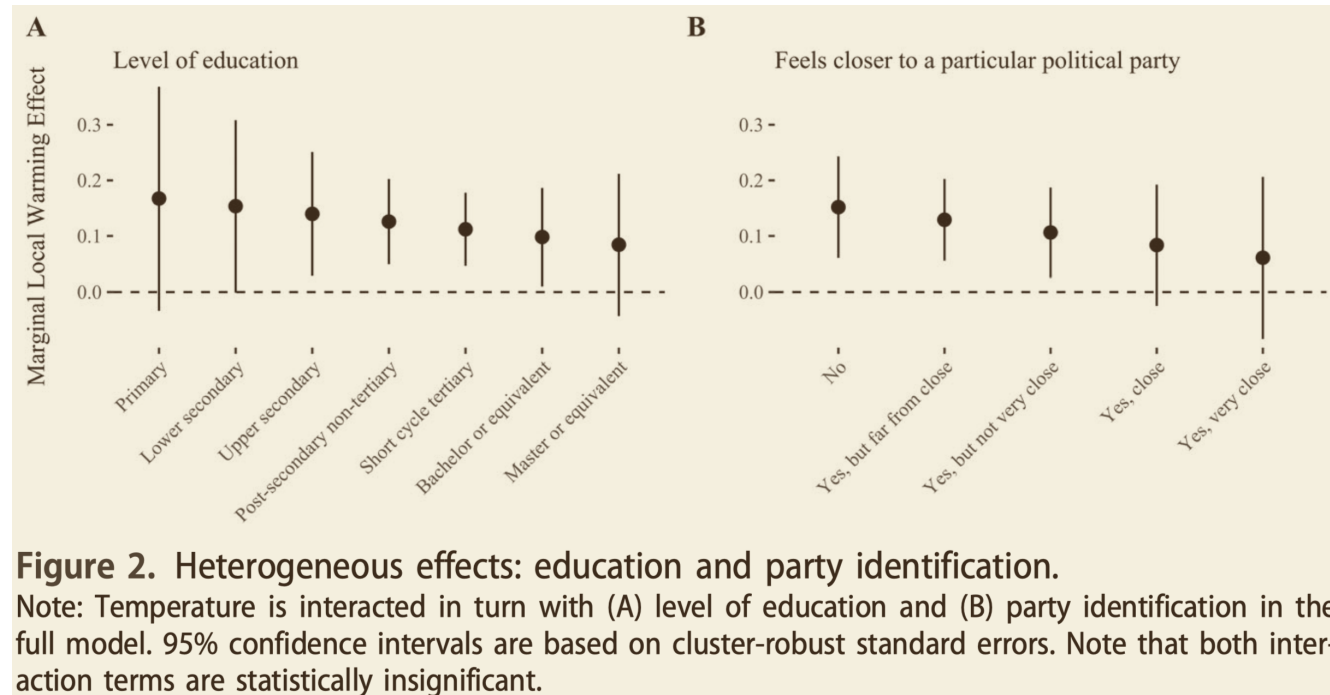
	Climate opinion index		
	(1)	(2)	(3)
Temperature	0.18 (0.13)	0.15*** (0.05)	−0.01 (0.14)
News exposure			4.99*** (0.77)
Education	1.32*** (0.10)	1.31*** (0.10)	1.26*** (0.10)
Ideology	−0.82*** (0.14)	−0.82*** (0.14)	−0.82*** (0.14)
Party ID	1.26*** (0.10)	1.26*** (0.10)	1.18*** (0.11)
Age	−0.08*** (0.02)	−0.08*** (0.02)	−0.10*** (0.02)
Gender (ref. = Male)	0.90** (0.41)	0.90** (0.41)	1.04*** (0.40)
Income	0.32*** (0.09)	0.32*** (0.09)	0.32*** (0.08)
Religiosity	−0.16 (0.15)	−0.16 (0.15)	−0.13 (0.14)
Ethnic minority (ref. = No)	−0.88 (0.97)	−0.88 (0.97)	−0.82 (0.98)
Temperature × Education	−0.01 (0.03)		
Temperature × Party ID		−0.02 (0.02)	
Temperature × News exposure			0.26 (0.26)
Constant	65.95*** (1.66)	65.94*** (1.64)	63.94*** (1.65)
Month + day of week FEs	+	+	+
Demographic controls	+	+	+
Countries in sample	12	12	12
Observations	22,850	22,850	22,598
R ²	0.17	0.17	0.17

*p<0.1; **p<0.05; ***p<0.01

Note: Full model with interactions between temperature and
(1) education, (2) party identification, and (3) news exposure.

All models are OLS regressions with country, month, and day of week fixed effects
and country-wise cluster-robust standard errors (in parentheses).

Hvordan ser de **marginal effekter** ud?



- "Figure 2(A) illustrates marginal effects for each education level. The trend is downward sloping, but **the interaction term is indistinguishable from zero**. The evidence therefore suggests that temperatures affect everyone equally regardless of education level."
- "Figure 2(B) shows that marginal effects are stronger for people who do not identify with a party, but this interaction term, too, is **statistically insignificant** meaning that the evidence of party identification as a moderating factor is also very weak."

Øvelsesopgaver i Stata

Øvelsesopgave 4

- Vi laver opgaverne sammen "på tavlen":
 - (*Kort*) hvad de rigtige Stata-kommandoer er
 - (*Vigtigst*) hvordan resultaterne fortolkes → hvad er svaret på spørgsmålene?
-

Prøv selv:

- Link til **slides**: <http://kortlink.dk/2acdr>
- Link til **data** (*states.dta* på Absalon): https://absalon.ku.dk/files/4521284/download?download_frd=1
- Link til "tom" **do-fil** med spørgsmål: <http://kortlink.dk/2aphr>

En sidste pointe

Valget af uafhængig (X) og modererende (Z) variabel er et teoretisk valg - et spørgsmål om synsvinkel

Hvorfor? (hint: $\beta_1 * X * Z = \beta_1 * Z * X$)

Matematisk set er interaktionen symmetrisk (Stata skelner ikke mellem $X * Z$ og $Z * X$):

- Interaktionsleddet fortæller, hvordan $X \rightarrow Y$ er forskellig afhængigt af Z , men **også hvordan $Z \rightarrow Y$ afhænger af X**
- Op til os (vores argument, problemformulering, forskningsdesign, hypotese), hvordan vi fortolker og præsenterer resultaterne
- Sjældent i praksis, at begge har en kausal fortolkning (kræver meget af vores forskningsdesign)

Summa summarum:

Standard practice og hjælp til sig selv at fokusere på **én påstået kausaleffekt af X , som potentialt afhænger af Z** (på en ikke-kausal måde)

Opsummering

- Interaktion er et genialt redskab til værktøjskassen, fordi der kan gemme sig rigtig meget information i interaktive modeller
- Kan bidrage med teoretisk relevant viden om, hvorvidt en tredjevariabel Z modererer effekten af X på Y
- Kan kaste lys over, hvordan en effekt varierer på tværs af forskellige typer subjekter med forskellig Z (marginale effekter)
- Men man skal ofte holde tungen lige i munden, når man fortolker interaktion(seffekter)/moderation/marginale effekter
 - "prisen" for nuanceringen af kausale påstande er en mere kompliceret model (og større risiko for *p-hacking*)

Næste gang

- Logistisk regression (logit)
- En ikke-lineær generalisering af OLS (lineær regression), som bruges, når den **afhængige variabel er binær**
 - 0/1, ja/nej, succes/fiasko
- En række fordele ift. OLS, når vi har et binært outcome - men også ulemper
- Et par ord om eksamen

Tak for i dag!