

**Bi7491 Regresní modelování**

---

**Případová studie**  
**Vitamin D**  
**(lineární regrese)**

Převzato z knihy:

Andersen, P. K., Skovgaard, L. T. (2010) Regression with Linear Predictors, Springer.

# Úvod

- Nedostatek vitamínu D přispívá k úbytku kostní hmoty – elevace parathormonu, zvyšování resorpce kostní hmoty
- fraktury u starších, poruchy růstu u adolescentů
- Vitamin D
  - tvoří se v kůži působením UV záření
  - přítomen v potravinách: rybí tuk, játra, vejce (D3), houby (D2), ...

# Úvod

- Vyhodnocení dostatku vitamínu D se provádí laboratorním testem sérové koncentrace 25-hydroxyvitamínu D (25OHD)
- Byla provedena mezinárodní studie (Irsko, Polsko, Finsko, Dánsko) u dívek a starších žen
- Data obsahují věk, BMI a stav vitamínu D u 420 dívek a žen

# Cíl

- Popsat dostupnost vitamínu D v různých evropských státech
- Identifikovat faktory, které by mohly vysvětlit rozdíly
  - věk
  - BMI
  - zvyklosti při slunění
  - příjem vitamínu D v potravě

# Příprava dat

- Logaritmická transformace výsledku

# Kauzalita

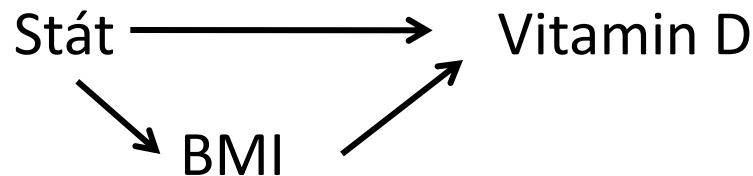
- Zadání úlohy:



- Což platí...
- Čím to lze vysvětlit (co je **mediátor**)?



Pouze vliv BMI?



Něco dalšího?

# Kauzalita

- Přetrvává hodně variability vysvětlené státem (stát zůstává velmi významným prediktorem)
- Musíme vytvořit bohatší model
- věk, zvyky při slunění, příjem vitamínu D – všechny mohou působit jako intermediáta mezi státem a výsledkem

# Cvičení – jak by mohl vypadat modelový diagram?

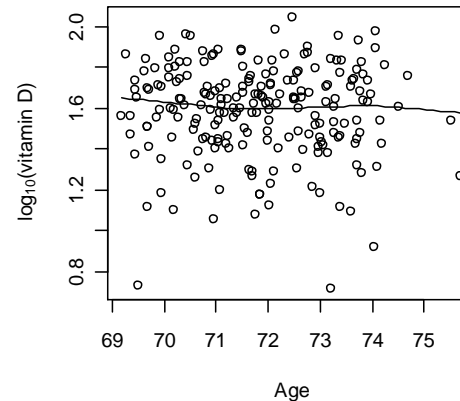
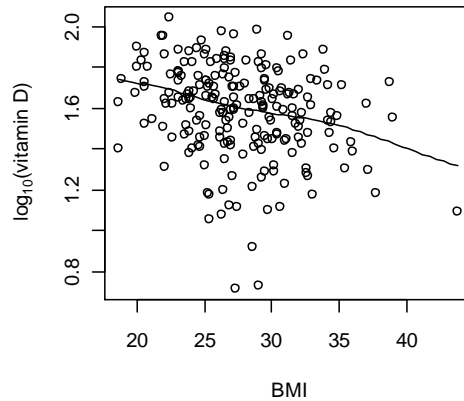




# Spojité prediktory (margin. průzkum)

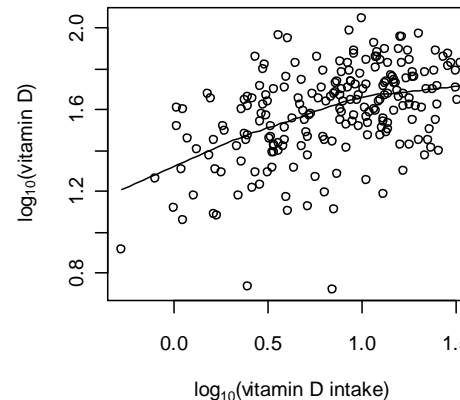
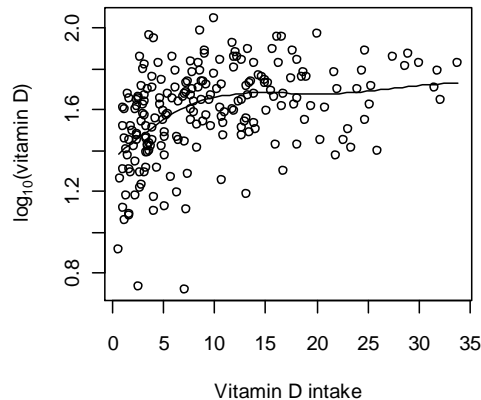
- je třeba rozhodnout o jejich formě v lineárním prediktoru

Linearita



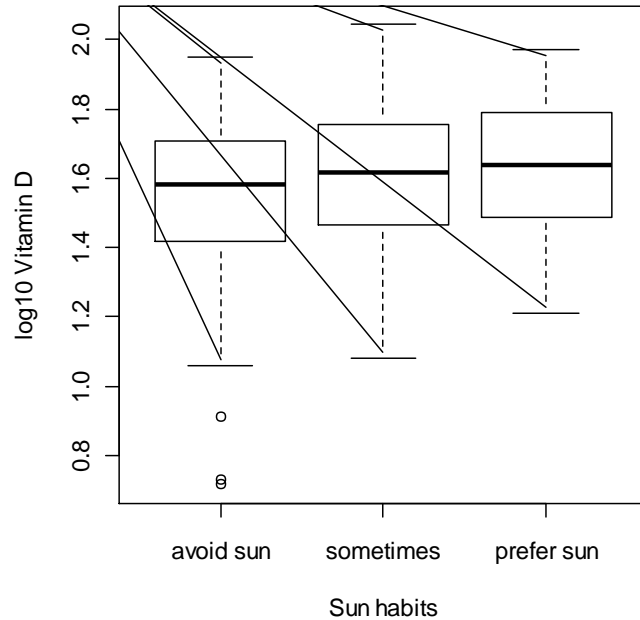
Není efekt  
(malý rozsah)

„Hokejka“

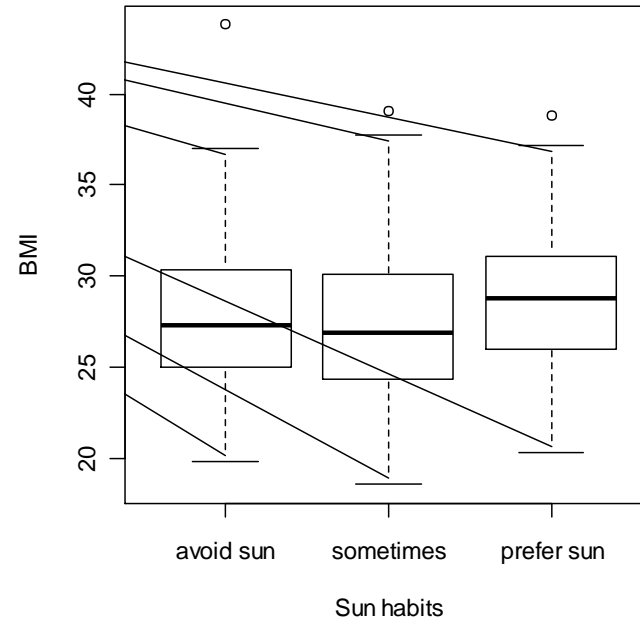


po transformaci  
lineární  
(rozumné i vzhledem  
k jednotce)

# Kategoriální prediktory

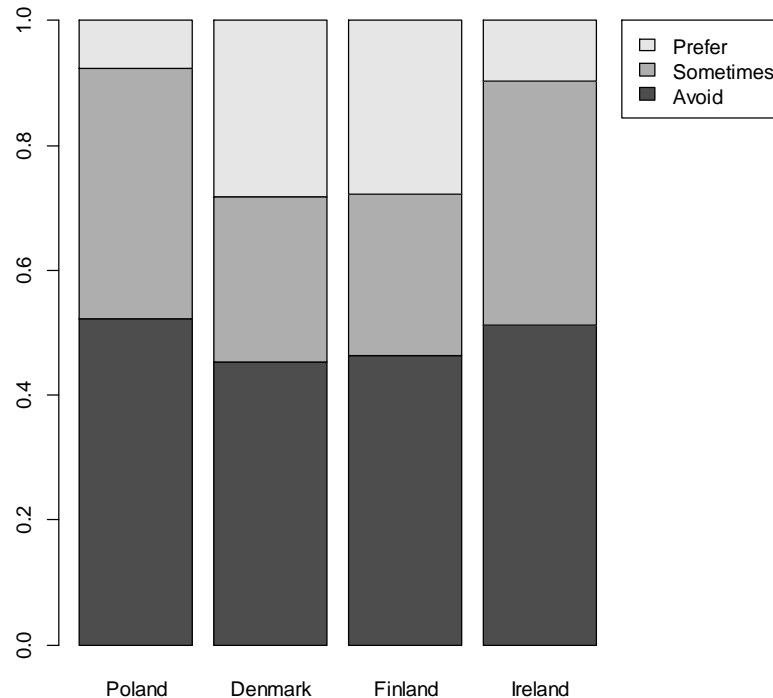


zvyklosti zřejmě  
ovlivňují vitamin D



zvyklosti zřejmě  
nesouvisí s BMI

# Kategoriální prediktory



stát zřejmě  
ovlivňuje  
zvyklosti...

hypotéza může být:

Problém v Polsku je kombinací málo slunění a vysokého BMI?

(protože v Irsku problém s vitamínem D není)

# Model 1

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.849653	0.682145	2.712	0.007269	**
bmi	-0.009777	0.003165	-3.089	0.002289	**
age	-0.003937	0.009481	-0.415	0.678356	nemá účinek
log10(vitdintake)	0.254517	0.035814	7.107	1.95e-11	***
sunexpavoid sun	-0.019531	0.030104	-0.649	0.517204	
sunexpprefer sun	0.042952	0.036842	1.166	0.245039	
countryDenmark	0.096405	0.036678	2.628	0.009230	**
countryFinland	0.076404	0.037275	2.050	0.041670	*
countryIreland	0.140771	0.038994	3.610	0.000385	***

(a jeho odstraněním nedojde k výrazné změně ostatních koeficientů)

# Model 2

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.569491	0.101014	15.537	< 2e-16	***
bmi	-0.009831	0.003156	-3.115	0.002103	**
log10(vitdintake)	0.255116	0.035712	7.144	1.56e-11	***
sunexpavoid sun	-0.021542	0.029653	-0.726	0.468382	
sunexpprefer sun	0.043472	0.036746	1.183	0.238174	
countryDenmark	0.095892	0.036583	2.621	0.009418	**
countryFinland	0.074683	0.036969	2.020	0.044670	*
countryIreland	0.138215	0.038427	3.597	0.000404	***

nemá účinek

# Model 3

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.546103	0.099677	15.511	< 2e-16	***
bmi	-0.009272	0.003134	-2.958	0.003455	**
log10(vitdintake)	0.257662	0.035175	7.325	5.23e-12	***
countryDenmark	0.109021	0.035695	3.054	0.002553	**
countryFinland	0.086036	0.036235	2.374	0.018491	*
countryIreland	0.140712	0.038454	3.659	0.000321	***

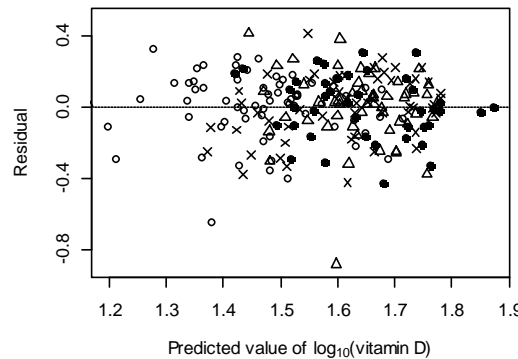
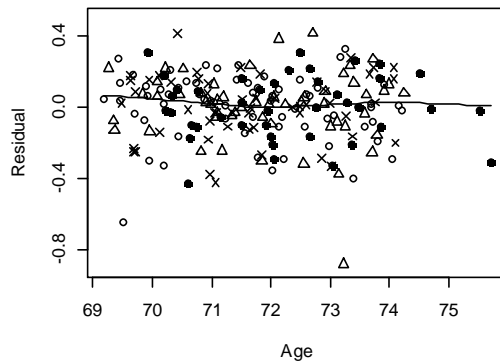
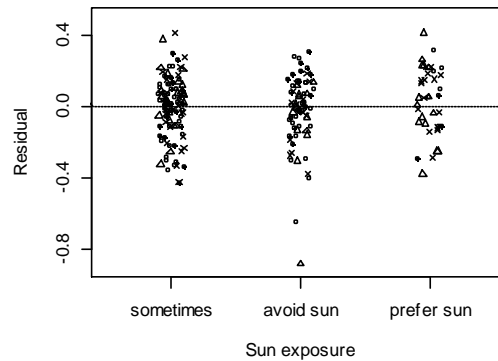
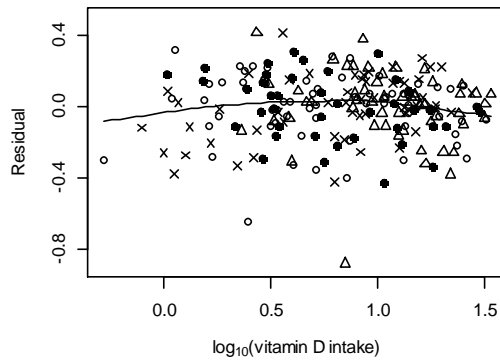
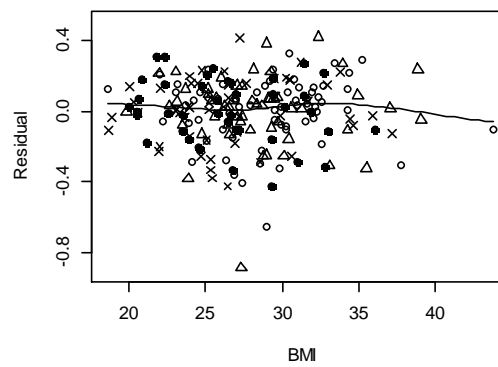
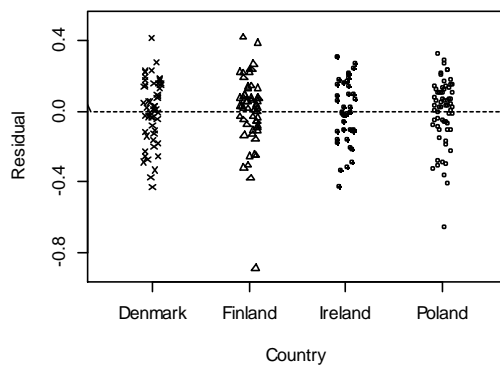
Stát zůstává významným prediktorem

# Analýza reziduí

Model3

vs.

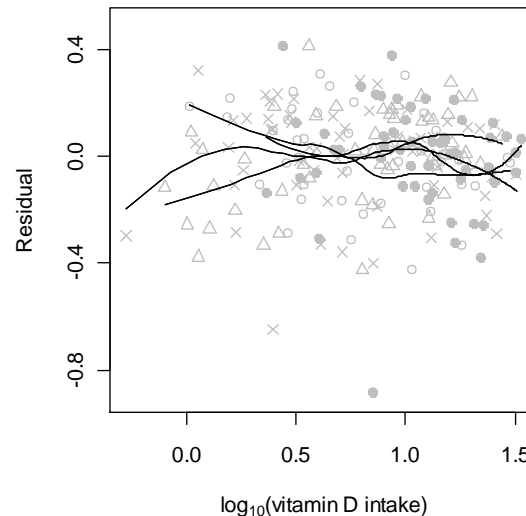
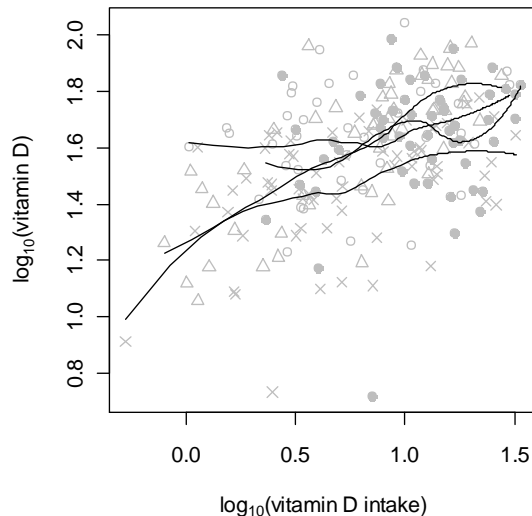
Všechny  
uvažované  
prediktory +  
predikce



žádné zřejmé  
problémy...

# Interakce

- mnoho možných
- jen některé předpokládané
  - stát X slunění – v každé zemi je slunce jinak silné
  - stát X příjem vitamínu D – obsah vitamínu u stejných potravin se může lišit



jde vidět, že  
marginální i  
podmíněné  
závislosti  
neukazují na  
interakce



# Interakce

	Estimate	Std. Error	t value	Pr(> t )	
I(bmi - median.bmi)	-0.009006	0.003070	-2.934	0.00375	**
countryDenmark	1.631813	0.036893	44.231	< 2e-16	***
countryFinland	1.672426	0.039968	41.844	< 2e-16	***
countryIreland	1.620712	0.039459	41.073	< 2e-16	***
countryPoland	1.526577	0.031207	48.918	< 2e-16	***
countryDenmark:I(log10(vitdintake) - median.logintake)	0.436478	0.061658	7.079	2.53e-11	***
countryFinland:I(log10(vitdintake) - median.logintake)	0.132080	0.086855	1.521	0.12995	
countryIreland:I(log10(vitdintake) - median.logintake)	0.114902	0.081007	1.418	0.15766	
countryPoland:I(log10(vitdintake) - median.logintake)	0.245292	0.057448	4.270	3.05e-05	***
countryDenmark:sunexpavoid sun	0.004212	0.061569	0.068	0.94553	
countryFinland:sunexpavoid sun	-0.142507	0.061207	-2.328	0.02092	*
countryIreland:sunexpavoid sun	0.096052	0.061288	1.567	0.11867	
countryPoland:sunexpavoid sun	-0.042359	0.047476	-0.892	0.37337	
countryDenmark:sunexpprefer sun	0.069935	0.059355	1.178	0.24013	
countryFinland:sunexpprefer sun	-0.019098	0.058774	-0.325	0.74557	
countryIreland:sunexpprefer sun	-0.093419	0.099831	-0.936	0.35054	
countryPoland:sunexpprefer sun	0.173724	0.088568	1.961	0.05124	.

sometimes

sklon  
dle  
státu

avoid  
vs.  
sometimes

prefer  
vs.  
sometimes

# Interakce

	Estimate	Std. Error	t value	Pr(> t )	
I(bmi - median.bmi)	-0.009006	0.003070	-2.934	0.00375	**
countryDenmark	1.631813	0.036893	44.231	< 2e-16	***
countryFinland	1.672426	0.039968	41.844	< 2e-16	***
countryIreland	1.620712	0.039459	41.073	< 2e-16	***
countryPoland	1.526577	0.031207	48.918	< 2e-16	***
countryDenmark:I(log10(vitdintake) - median.logintake)	0.436478	0.061658	7.079	2.53e-11	***
countryFinland:I(log10(vitdintake) - median.logintake)	0.132080	0.086855	1.521	0.12995	
countryIreland:I(log10(vitdintake) - median.logintake)	0.114902	0.081007	1.418	0.15766	
countryPoland:I(log10(vitdintake) - median.logintake)	0.245292	0.057448	4.270	3.05e-05	***
countryDenmark:sunexpavoid sun	0.004212	0.061569	0.068	0.94553	
countryFinland:sunexpavoid sun	-0.142507	0.061207	-2.328	0.02092	*
countryIreland:sunexpavoid sun	0.096052	0.061288	1.567	0.11867	
countryPoland:sunexpavoid sun	-0.042359	0.047476	-0.892	0.37337	
countryDenmark:sunexpprefer sun	0.069935	0.059355	1.178	0.24013	
countryFinland:sunexpprefer sun	-0.019098	0.058774	-0.325	0.74557	
countryIreland:sunexpprefer sun	-0.093419	0.099831	-0.936	0.35054	
countryPoland:sunexpprefer sun	0.173724	0.088568	1.961	0.05124	.

sometimes

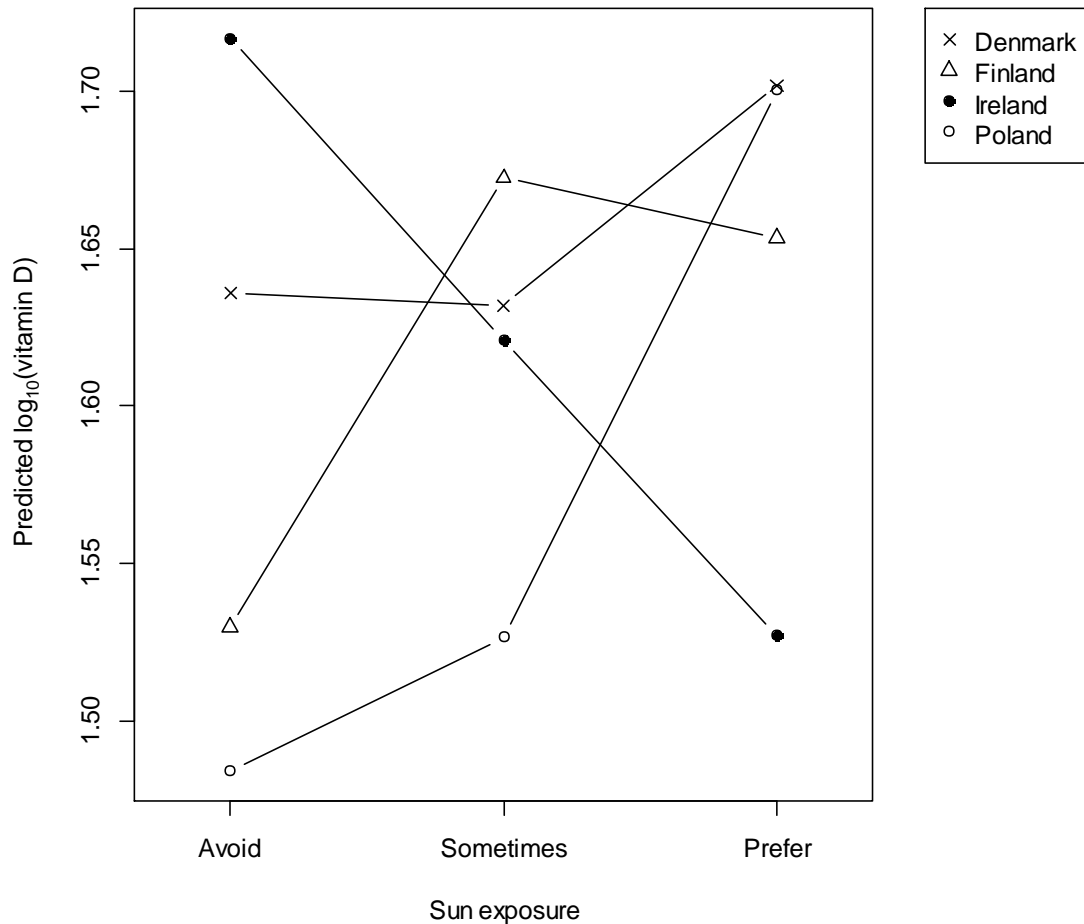
sklon  
dle  
státu

avoid  
vs.  
sometimes

prefer  
vs.  
sometimes

Příjem vitamínu D je silnější prediktor v Dánsku, v Irsku a Finsku vůbec

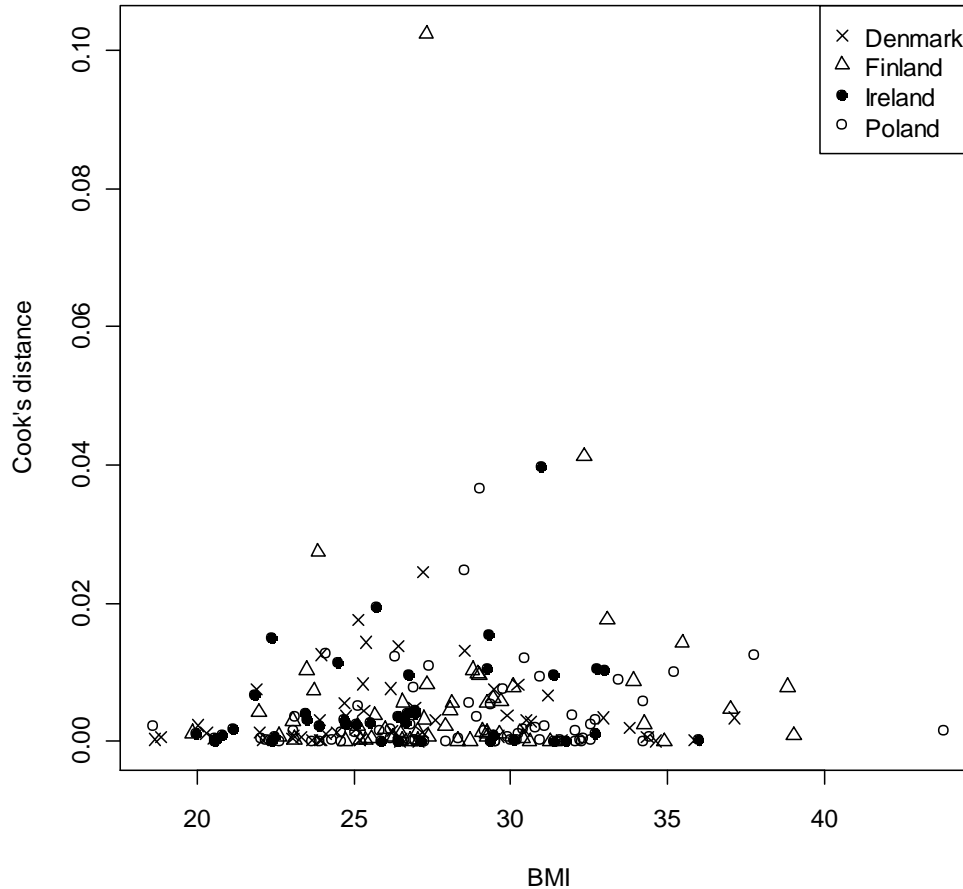
# Interakce



předpokladatelně  
to funguje jenom ve  
Finsku a Polsku

v Irsku úplně naopak,  
zřejmě náhoda, málo  
preferujících žen

# Vlivná pozorování



Finka, která se vyhýbá slunkem...

# Závěry

- Nedokázali jsme zcela vysvětlit rozdíly mezi státy
- Interakce (státy-slunění), která se špatně interpretuje, takže možná lépe nezahrnovat do výsledku
- výsledné prediktory:
  - BMI
  - příjem vitamínu D (v každém státě jiný efekt)
- pozor na logaritmickou škálu

# Finální model

```
lm(formula = log10(vitd) ~ country + bmi + country:log10(vitdintake) -  
1, data = women)
```

	Estimate	Std. Error	t value	Pr(> t )	
countryDenmark	1.523693	0.100271	15.196	< 2e-16	***
countryFinland	1.729285	0.127440	13.569	< 2e-16	***
countryIreland	1.824750	0.111454	16.372	< 2e-16	***
countryPoland	1.579786	0.103125	15.319	< 2e-16	***
bmi	-0.009619	0.003069	-3.134	0.00198	**
countryDenmark:log10(vitdintake)	0.434867	0.061992	7.015	3.32e-11	***
countryFinland:log10(vitdintake)	0.174453	0.087081	2.003	0.04646	*
countryIreland:log10(vitdintake)	0.091818	0.080599	1.139	0.25596	
countryPoland:log10(vitdintake)	0.226312	0.056947	3.974	9.80e-05	***

# Cvičení – interpretace logaritmů

máme-li transformovaný výsledek, jak interpretovat koeficient?  
máme-li transformovaný výsledek i prediktor, jak interpretovat koeficient?



# Interpretace - logaritmy

Logarithmic transformation of outcome:

$$\log f(x) = \alpha + \beta X \quad (A.1)$$

Taking the antilogarithm:

$$f(x) = 10^{\alpha + \beta X} = 10^\alpha 10^{\beta X} = 10^\alpha (10^\beta)^x \quad (A.2)$$

One unit increase in outcome:

$$f(x + 1) = 10^\alpha (10^\beta)^{x+1} = 10^\alpha 10^{\beta x} 10^\beta = f(x) 10^\beta \quad (A.3)$$

Logarithmic transformation of outcome and covariate:

$$\log f(x) = \alpha + \beta \log X \quad (A.4)$$

Taking the antilogarithm:

$$f(x) = 10^{\alpha + \beta \log X} = 10^\alpha 10^{\beta \log X} = 10^\alpha X^\beta \quad (A.5)$$

10-fold increase in outcome:

$$f(10x) = 10^\alpha (10x)^\beta = 10^\alpha 10^\beta x^\beta = f(x) 10^\beta \quad (A.6)$$

50% increase in outcome:

$$f(1.5x) = 10^\alpha (1.5x)^\beta = 10^\alpha 1.5^\beta x^\beta = f(x) 1.5^\beta \quad (A.7)$$

Andersen, 2005



# Interpretace - logaritmy

```
lm(formula = log10(vitd) ~ country + bmi + country:log10(vitdintake) -  
1, data = women)
```

	Estimate	Std. Error	t value	Pr(> t )	
countryDenmark	1.523693	0.100271	15.196	< 2e-16	***
countryFinland	1.729285	0.127440	13.569	< 2e-16	***
countryIreland	1.824750	0.111454	16.372	< 2e-16	***
countryPoland	1.579786	0.103125	15.319	< 2e-16	***
<b>bmi</b>	<b>-0.009619</b>	<b>0.003069</b>	<b>-3.134</b>	<b>0.00198</b>	<b>**</b>
<b>countryDenmark:log10(vitdintake)</b>	<b>0.434867</b>	<b>0.061992</b>	<b>7.015</b>	<b>3.32e-11</b>	<b>***</b>
countryFinland:log10(vitdintake)	0.174453	0.087081	2.003	0.04646	*
countryIreland:log10(vitdintake)	0.091818	0.080599	1.139	0.25596	
countryPoland:log10(vitdintake)	0.226312	0.056947	3.974	9.80e-05	***

**Účinek zvýšení o jednotku BMI: vynásobení  $10^{-0,0096} = 0,978$  ... pokles o 2,2%**

**Účinek zdvojnásobení příjmu v Dánsku: vynásobení  $2^{0,4349} = 1,352$  ... nárůst o 35%**