

PSY252

Statistická analýza dat v psychologii II

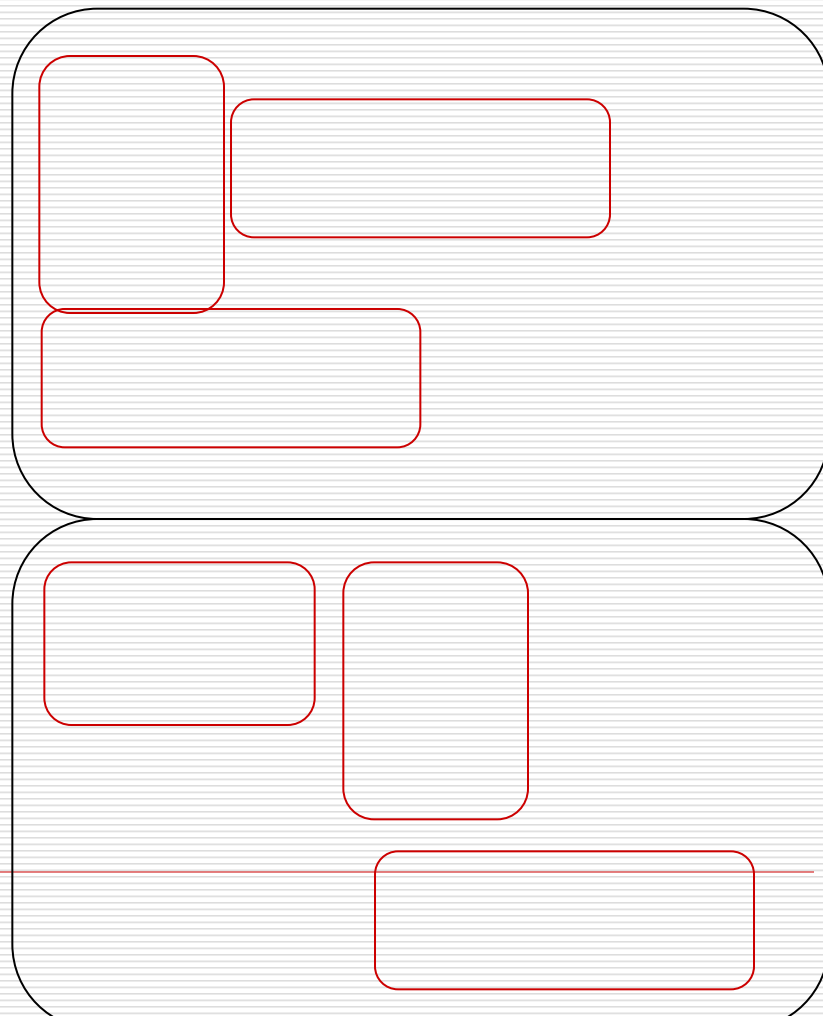
Víceúrovňový lineární model (multilevel, hierarchical, mixed, random-coefficients model)

Víceúrovňová data

ID	Třída	Výkon
100	1	11
...	1	20
120	1	31
121	2	40
..	2	52
150	2	63
151	3	20
...	3	40
180	3	30
181	4	100

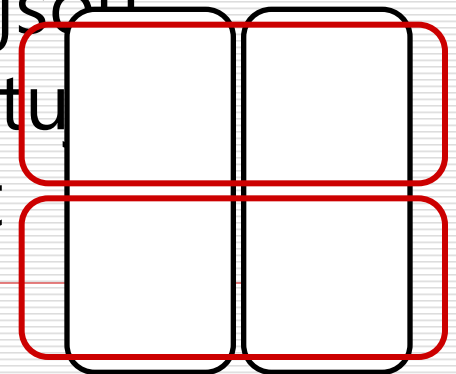
Víceúrovňová data

ID	Třída	Výkon
100	1	11
...	1	20
120	1	31
121	2	40
..	2	52
150	2	63
151	3	20
...	3	40
180	3	30
181	4	100



Víceúrovňová data – vnořené (nested) faktory

- Určité úrovně faktorů nižší úrovně se vyskytují pouze v jediné úrovni faktorů vyšší úrovně
 - Proto též hierarchická data
 - Konkrétní třída je jen v jedné škole, žák je členem jen jedné třídy
- Protikladem pro vnořené faktory jsou zkřížené (crossed) faktory – vyskytují se všechny kombinace jejich hodnot

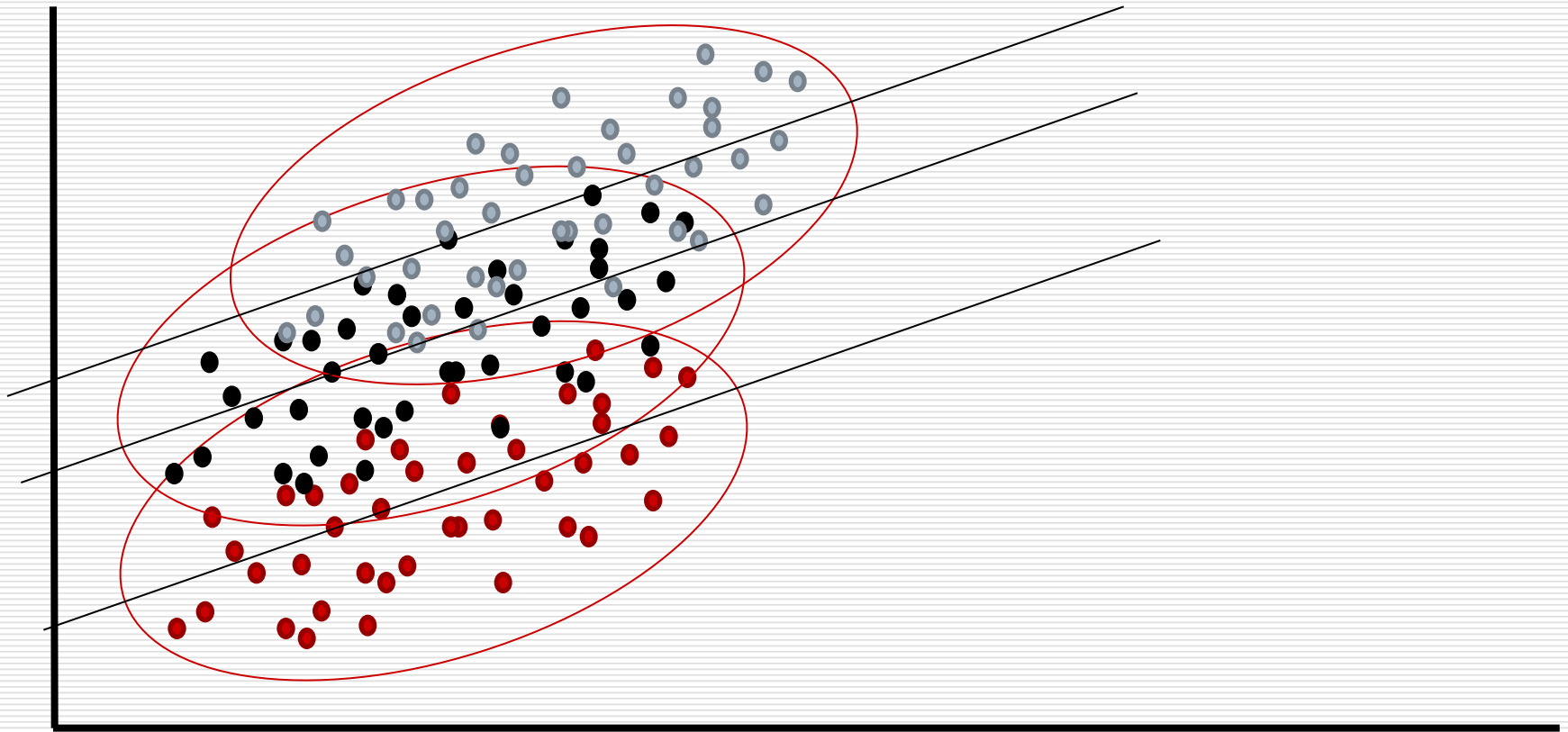


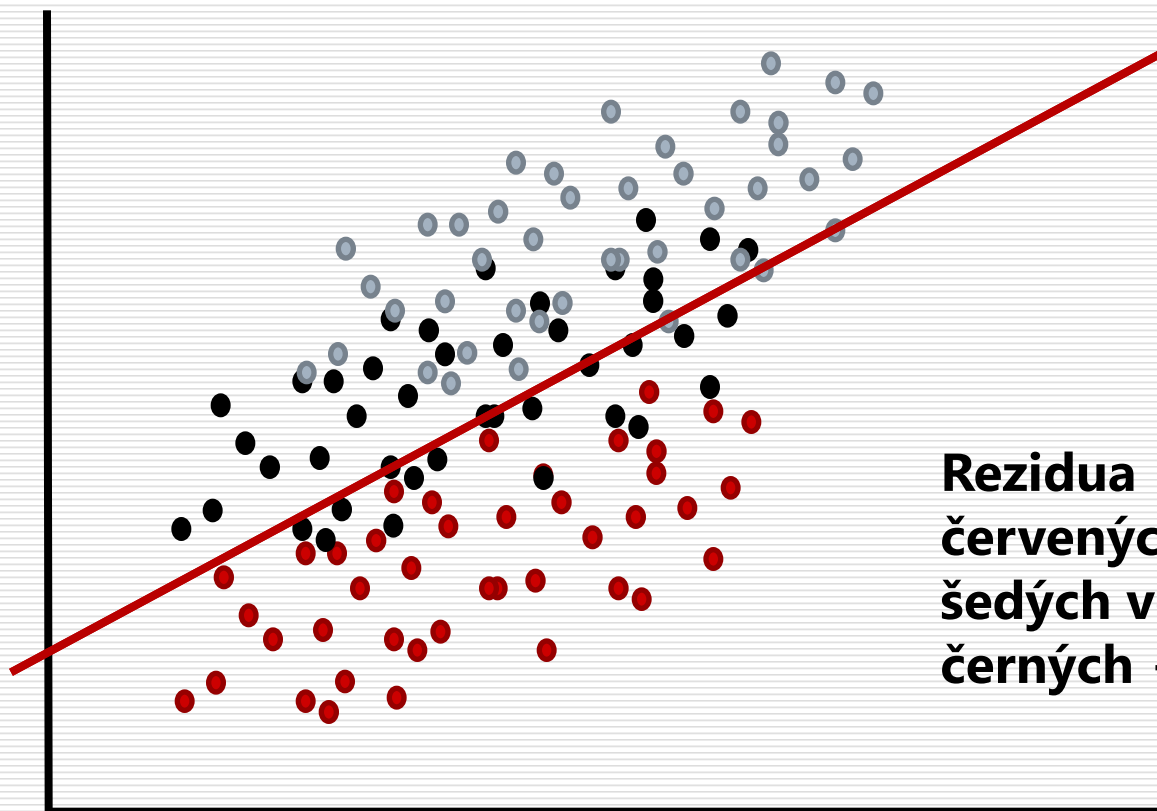
Příklady víceúrovňových dat

- Žáci(L1) ve třídách (L2) ve školách (L3) v okresech (L3) ...
 - Účastníci experimentu (L1) testování po skupinkách (L2), popř. na různých místech (L2 či L3)
 - ...
 - Opakovaná měření (L1) týchž lidí (L2)
-

Víceúrovňovost způsobuje závislost reziduí

- Pokud proměnná definující skupiny na vyšší úrovni jakkoli souvisí s modelovanou charakteristikou, její ignorování způsobuje to, že rezidua lidí ve skupině si budou podobnější než rezidua lidí napříč skupinami.
-





Rezidua
červených jsou většinou záporná,
šedých většinou kladná,
černých +-

Víceúrovňový model zohledňuje závislost reziduí danou členstvím ve skupinách

$$Y_i = b_0 + b_1 X_i + e_i$$

$$Y_{ij} = b_{0j} + b_{1j} X_{ij} + e_{ij}$$

$$b_{0j} = b_0 + u_{0j}$$

<1. úroveň>

<2. úroveň>

Průsečík ve skupině j

Průměrný průsečík

Odchylka průsečíku skupiny j od průměrného průsečíku

Odchylky rozptyl

b_0 se stává náhodným koeficientem (random coefficient)

Víceúrovňový model zohledňuje závislost reziduí danou členstvím ve skupinách

$$Y_{ij} = b_{0j} + b_{1j}X_{ij} + e_{ij} \quad \langle 1. \text{ úroveň} \rangle$$

$$b_{0j} = b_0 + u_{0j} \quad \langle 2. \text{ úroveň} \rangle$$

Alternativně (dosazením sloučeno)

$$Y_{ij} = (b_0 + u_{0j}) + b_{1j}X_{ij} + e_{ij}$$

$$Y_{ij} = b_0 + b_{1j}X_{ij} + (e_{ij} + u_{0j})$$

Random-intercepts model

$$Y_{ij} = b_0 + b_{1j}X_{ij} + (e_{ij} + u_{0j})$$

Y predikovaná proměnná

Efekty (fixed effects)

b_0 průměrný průsečík napříč skupinami

b_{1j} efekt pro všechny skupiny (*není random*)

Struktura reziduí (kovarianční parametry)

$\text{Var}(u_{0j})$ rozptyl průsečíků, $u_{0j} \sim N(0, \sigma^2_{u0})$

$\text{Var}(e_{ij})$ rozptyl reziduí, $e_{ij} \sim N(0, \sigma^2_e)$

Model má 4 odhadované parametry.

Příklad – Skotské zkoušky

Liší se holky a kluci ve výsledku testů?

- Ano, $m_B - m_G = -5,5$ ($t(1903) = 5,57$, $d \approx 0,25$)
 - Jenže různé školy se liší jednak průměrnou výkonností, tak zastoupením pohlaví.
 - Pokud by náhodou bylo ve škole s vysokou výkonností více kluků, mohli by kluci vyjít lépe jen díky tomu.
 - Navíc, Durbin-Watson = 1,4
-

Příklad – Skotské zkoušky

Liší se holky a kluci ve výsledku testů?

- Multilevel model, kde je zohledněno to, jaké školy žáci pochází
- Random-intercept model = předpokládáme, že
 - školy se liší průměrnou výkonností v testu (random Intercept)
 - rozdíl mezi pohlavími je ve všech školách stejný (fixed Slope)
 - ID jsou vnořena do škol

$$Test_{ij} = b_{0š} + b_1 Gender_i + e_{iš} \quad \langle 1. \text{ úroveň} \rangle$$

$$b_{0š} = b_0 + u_{0š} \quad \langle 2. \text{ úroveň} \rangle$$

Nepodmíněný model průměrů

Unconditional means model, variance components

- Model bez prediktorů zohledňující strukturu dat
- Pouze dělí rozptyl na reziduální rozptyl a rozptyl průměrů skupin
 - ICC=rozptyl průměrů/(rozptyl průměrů+reziduální rozptyl)
 - ICC= jaká část rozptylu výkonů je vysvětlitelná pouze rozdíly mezi školami?
- 3 parametry – průměrný průměr škol (b_0), rozptyl průměrů škol, rozptyl reziduí (variabilita uvnitř škol)

$$Test_{ij} = b_{0\check{s}} + e_{i\check{s}} \quad \langle 1. \text{ úroveň} \rangle$$

$$b_{0\check{s}} = b_0 + u_{0\check{s}} \quad \langle 2. \text{ úroveň} \rangle$$

Random-intercepts model

- Předpokládá,
 - že jednotky vyššího řádu se liší svým průměrem,
 - že průměry mají normální rozložení
 - že efekty jsou stejné (fixed) napříč všemi jednotkami vyššího řádu
 - že rezidua jsou napříč jednotkami vyššího řádu stejná
-

Random-slopes model

- Předpokládá,
 - že všechny jednotky vyššího řádu mají stejný průměr,
 - že efekt prediktoru je v každé jednotce vyššího řádu jiný a
 - že tyto efekty mají nějakou průměrnou hodnotu a nějakou variabilitu
-

Random-slopes model

$$Y_{ij} = b_0 + b_{1j}X_{ij} + e_{ij} \quad <1. \text{ úroveň}>$$
$$b_{1j} = b_1 + u_{1j} \quad <2. \text{ úroveň}>$$

Alternativně (dosazením sloučeno)

$$Y_{ij} = b_0 + (b_1 + u_{1j})X_{ij} + e_{ij}$$

$$Y_{ij} = b_0 + b_1X_{ij} + (e_{ij} + u_{1j})$$

Jen zřídka má smysl předpokládat, náhodné efekty při fixovaných průsečících!

Random intercept and slope model

Předpokládá,

- že jednotky vyššího řádu mají různé průměry(průsečíky),
 - že efekt prediktoru je v každé jednotce vyššího řádu jiný,
 - že tyto průsečíky i efekty mají nějakou průměrnou hodnotu a nějakou variabilitu napříč skupinami,
 - že reziduální rozptyl je napříč skupinami konstatní.
 - Lze uvažovat i to, že mezi hodnotou průsečíku a efektu je nějaká korelace.
-

Random intercept and slope model

$$Y_{ij} = b_{0j} + b_{1j}X_{ij} + e_{ij} \quad \langle 1. \text{ úroveň} \rangle$$

$$b_{0j} = b_0 + u_{0j} \quad \langle 2. \text{ úroveň} \rangle$$

$$b_{1j} = b_1 + u_{1j} \quad \langle 2. \text{ úroveň} \rangle$$

Alternativně (dosazením sloučeno)

$$Y_{ij} = b_0 + b_1X_{ij} + (e_{ij} + u_{0j} + u_{1j})$$

Random intercept and slope model

$$Y_{ij} = b_0 + b_1 X_{ij} + (e_{ij} + u_{0j} + u_{1j})$$

Efekty (fixed effects)

b_0 průměrný průsečík napříč skupinami

b_1 průměrný efekt pro všechny skupiny

Struktura reziduí (kovarianční parametry)

$\text{Var}(u_{0j})$ rozptyl průsečíků, $u_{0j} \sim N(0, \sigma^2_{u0})$

$\text{Var}(u_{1j})$ rozptyl efektů, $u_{1j} \sim N(0, \sigma^2_{u1})$

$\text{Var}(e_{ij})$ rozptyl reziduí, $e_{ij} \sim N(0, \sigma^2_e)$

Model má 5 odhadovaných parametrů.

Šestý $\text{Cov}(u_{0j}, u_{1j})$ kovariance průsečíků s efekty

Příklad – Skotské zkoušky

Liší se holky a kluci ve výsledku testů?

□ Zvažme, zda se mohou lišit i efekty napříč školami

$$Test_{i\check{s}} = b_0 + b_1 Gender_{i\check{s}} + (e_{ij} + u_{0\check{s}} + u_{1\check{s}})$$

Příklad – Skotské zkoušky

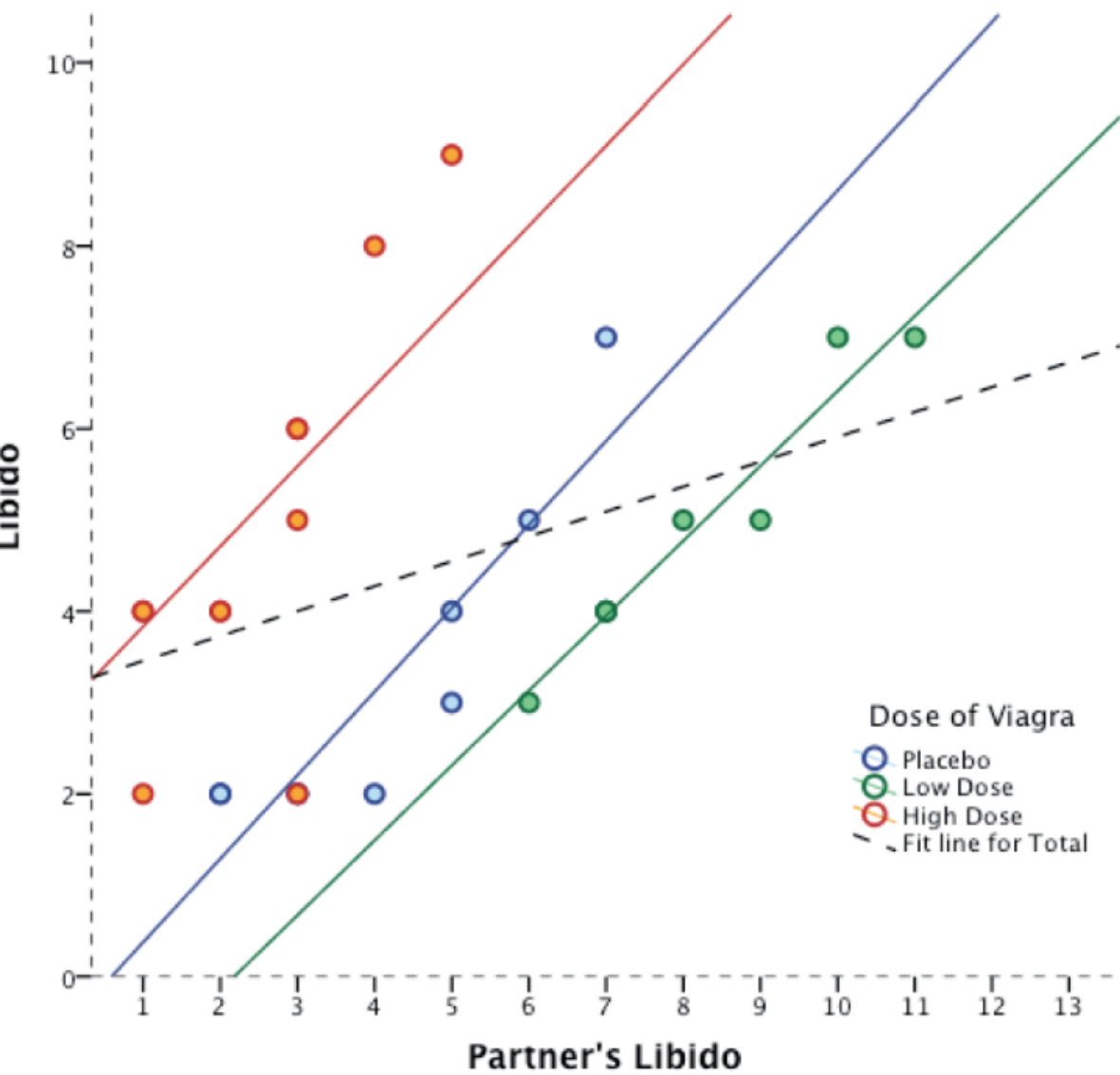
$$\square Test_i = (79 \pm 11) - (3,9 \pm 1,8) Gender_i \pm 17,8$$

Jsou-li kluci 0 a holky 1, pak...

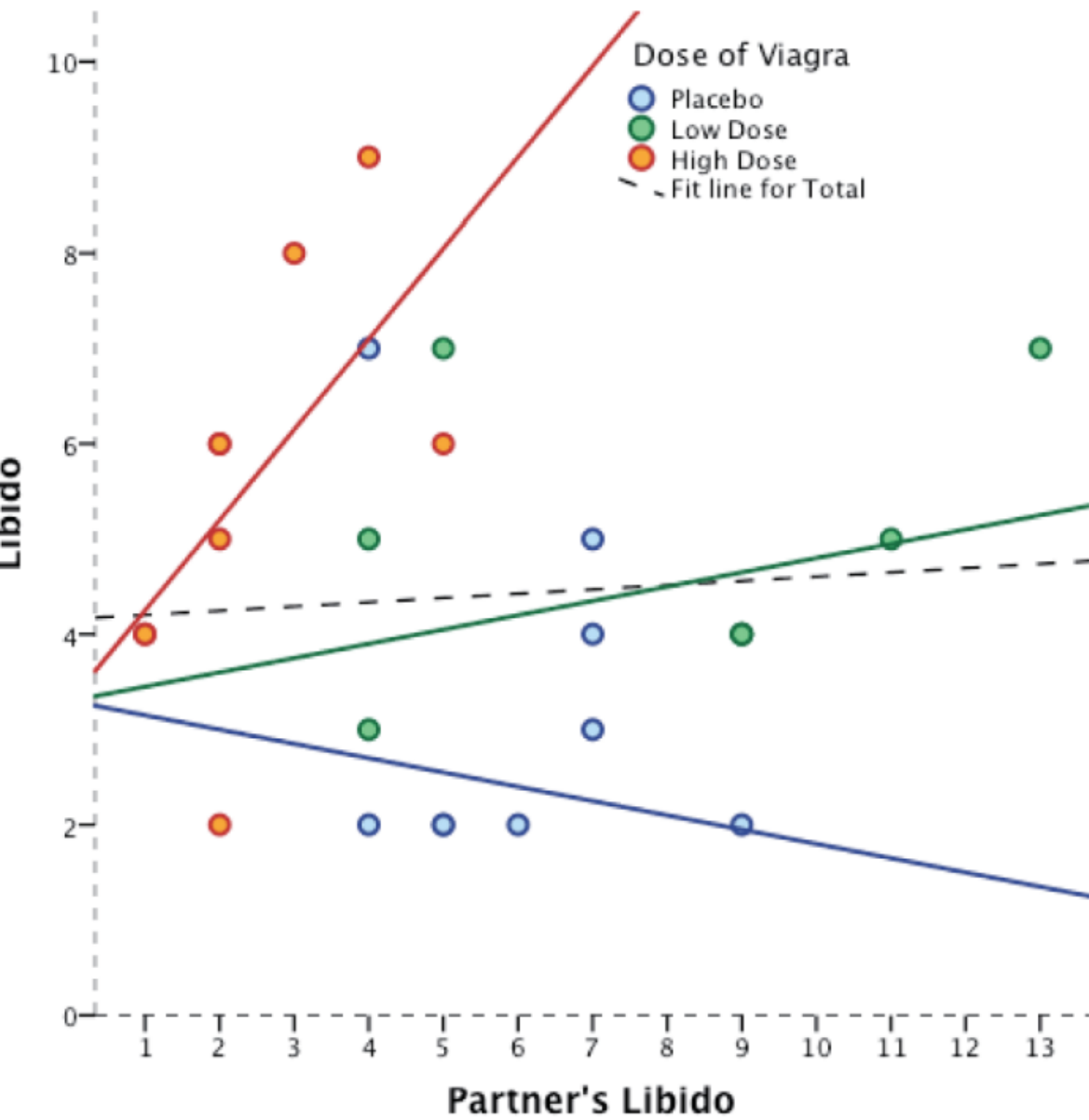
- \square Výkon průměrného kluka v průměrné škole je 79, přičemž školy se liší tak, že výkony průměrných kluků mají $SD=11$.
 - \square Průměrná holka má v průměrné škole o 3,9 bodu míň.
 - \square I když napříč školami mají rozdíly mezi průměrnou holkou a průměrným klukem $SD=1,8$, rozptyl efektů není signifikantně odlišný od 0.
-

Shrnutí

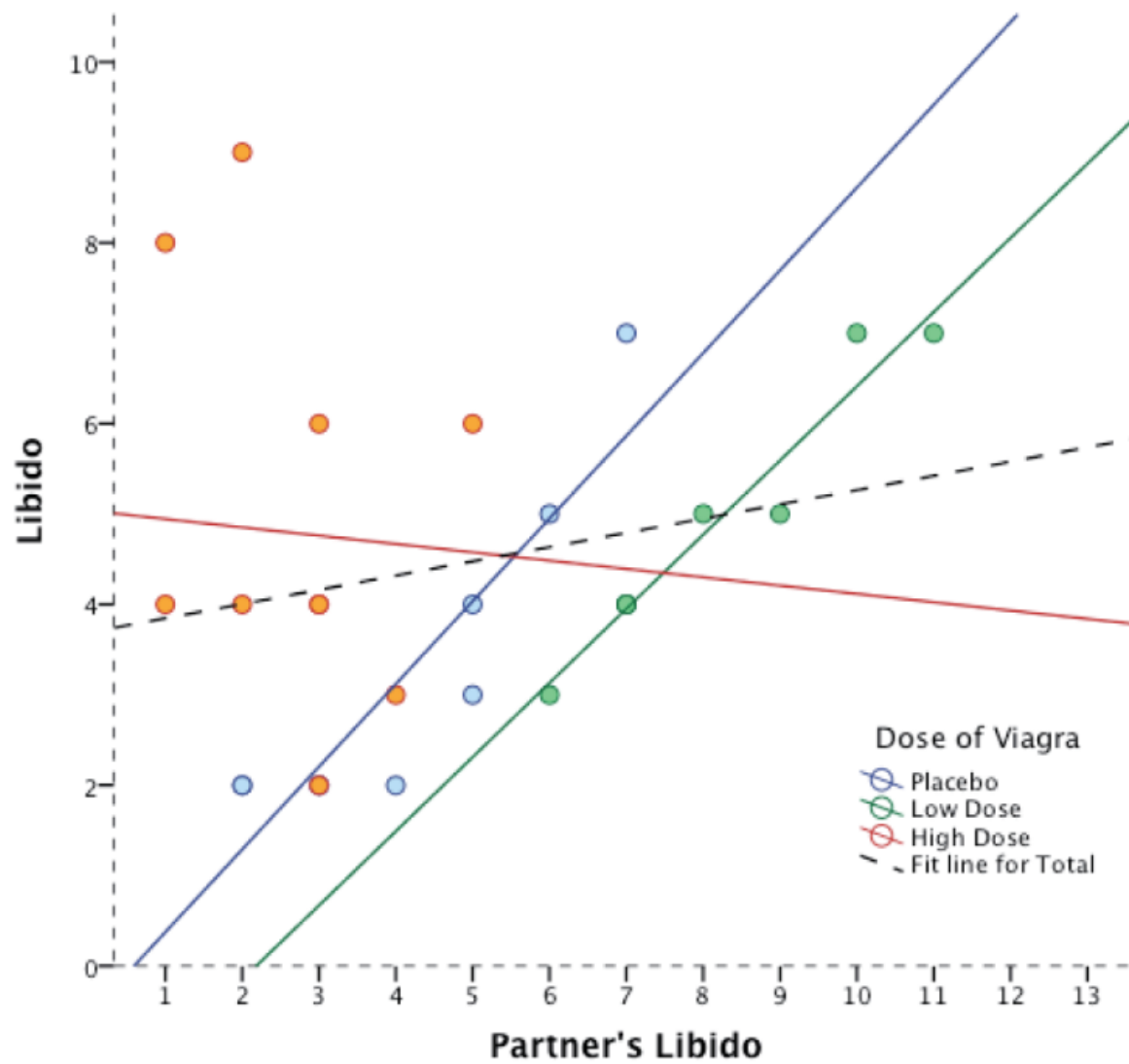
- ❑ Multilevel modely nám umožňují modelovat to, že některé parametry regresního modelu se mohou pro různé skupiny lišit.
- ❑ Od moderace se to liší tím, že různost parametrů má podobu normálního rozložení. Nezajímáme se o hodnoty pro jednotlivé skupiny – ze vzorku skupin usuzujeme na populaci skupin
- ❑ S tím je spojen předpoklad, že vzorek jednotek druhé úrovně (skupin) je reprezentativním vzorkem populace skupin



**Random Intercept,
Fixed Slope**



**Fixed Intercept,
Random Slope**



**Random Intercept,
Random Slope**

Shoda modelu s daty

- Podobně jako u logistické regrese vyjadřují celkový fit modelu informační kritéria založená na $-2LL$
 - AIC, AICC, CAIC, BIC
 - Vnořené modely lze srovnávat LRT – rozdíl $-2LL$ dvou vnořených modelů má chí-kvadrát rozložení s df rovným rozdílu v počtu parametrů mezi srovnávanými modely (*nefunguje s REML*)
-

Typy kovariančních struktur

- Ve výše popsaných modelech jsou smysluplné jen 2 volby a hraje to roli, jen, když máme v modelu více než 1 náhodný koeficient
- VC – Variance components – náhodné koeficienty nekorelují
- UN – Unstructured – náhodné koeficienty mohou korelovat

Čtěte opatrně, Andy tu nejistě mlží.

Předpoklady

- Jako lineární regrese
 - Je-li závislost reziduí modelovatelná (=je to skupinami), vyléčí se tím problémem
-

Prediktor na úrovni skupin

- Do modelu lze vložit i prediktor, který vysvětluje rozdíly mezi skupinami.
 - Například „nóblóznost“ spádové oblasti školy – nbrhd
-



Longitudinální, repeated data

1. úroveň: měření

2. úroveň: jednotlivec

- Čas, či pořadí měření je proměnnou na 1. úrovni.
 - Čas může nabývat různé hodnoty pro různé lidi v různé časy měření
- Charakteristiky jednotlivců jsou proměnnými na 2. úrovni.

LATENT GROWTH-CURVE MODELING

ŠIROKÁ

VS.

DLOUHÁ DATA

ID	EDA klid	EDA stres1	EDA stres2
101A	1	2	3
102A	4	5	6
...			
199A	5	3	5

ID	Stres	EDA
101A	Klid	1
101A	Stres1	2
101A	Stres2	3
102A	Klid	4
102A	Stres1	5
102A	Stres2	6
...		
199A	Klid	5
199A	Stres1	3
199A	Stres2	5

Převod širokých dat na dlouhá a zpět

- SPSS >> Data >> Restructure
(VARSTOCASES)
-