



L08

Liší se vztah délky a zadní šířky krunýře mezi samicemi a samci?

Interakce prediktorů

Ilustrace vytvořená pomocí AI.



cuni-natur-biostatistics.github.io/materialy.html

Co vlastně na krabovi měříme?



Ilustrace vytvořená pomocí AI.

Délka krunýře a zadní šířka krunýře jsou dvě měření téhož kraba.

Co si pamatujete z minulé lekce?

Odpovězte samostatně na tři krátké otázky. Společné odpovědi projdeme po hlasování.

Co znamená záporný koeficient roční teploty v modelu s roční i maximální teplotou?

Co znamená sklon v modelu se dvěma teplota

tloušťka listu ~ roční teplota + maximální teplota

roční teplota

-0.00528 mm / °C

při stejné

maximální teplotě

Každý koeficient drží ostatní prediktory stejné.

Celkovou změnu tloušťky listu při zvýšení obou teplot o 1 °C.

Rozdíl průměrné roční teploty mezi původními a zavlečenými populacemi.

Vztah roční teploty k tloušťce listu bez ohledu na maximální teplotu.

Změnu odhadované tloušťky listu při zvýšení roční teploty o 1 °C, když maximální teplotu držíme stejnou.

SPRÁVNÁ ODPOVĚĎ

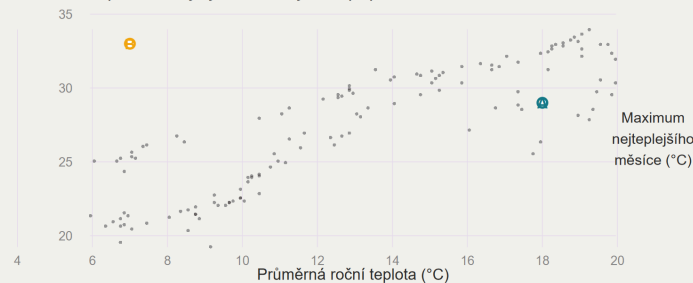
: Změnu odhadované tloušťky listu při zvýšení roční teploty o 1 °C, když maximální teplotu držíme stejnou.

V modelu s více prediktory čteme každý sklon při stejné hodnotě ostatních prediktorů. Koeficient proto není totožný s jednoduchým jednoprediktorovým sklonem.

Kterému odhadu z modelu se dvěma teplotami bychom měli více důvěřovat?

Která kombinace teplot má oporu v datech?

Jeden bod představuje jednu zdrojovou populaci.



¹⁵ A leží v pozorovaném mraku · B kombinuje hodnoty mimo něj

Odhadu pro kombinaci roční a maximální teploty, která leží uvnitř pozorovaného mraku bodů.

Odhadu pro libovolnou kombinaci uvnitř rozsahu každé teploty, i když se tato dvojice v datech nevyskytuje.

Odhadu pro kombinaci nejnižší roční a nejvyšší maximální teploty, protože využívá oba extrémy.

Všem odhadům stejně, protože lineární model umí spočítat každou číselnou kombinaci.

SPRÁVNÁ ODPOVĚĎ

: Odhadu pro kombinaci roční a maximální teploty, která leží uvnitř pozorovaného mraku bodů.

Model výpočet provede i mimo společně pozorované kombinace, ale takový odhad je extrapolace. Opora je silnější uvnitř mraku skutečných kombinací prediktorů.

Prohodíme pořadí dvou korelovaných teplot ve formuli. Co se může změnit?

Stejný model, jiné pořadí sekvenčních testů

Koeficienty a předpovědi zůstávají stejné.

roční + maximální

sekvenční SS
roční: 0.2017
maximální: 0.0133

maximální + roční

sekvenční SS
maximální: 0.1968
roční: 0.0182

anova() čte členy zleva doprava

Koeficienty modelu i jeho předpovědi.

Koeficienty se nezmění, ale sekvenční tabulka z anova() může přidělit sdílenou informaci jinak.

Pouze počet pozorování v modelu.

Nic; koeficienty, předpovědi i každý řádek sekvenční anova() musí zůstat stejné.

SPRÁVNÁ ODPOVĚĎ

: Koeficienty se nezmění, ale sekvenční tabulka z anova() může přidělit sdílenou informaci jinak.

Pořadí prediktorů nemění fit ani koeficienty stejného modelu. Sekvenční anova() však testuje členy v pořadí ve formuli, takže u korelovaných prediktorů se její řádky mohou změnit.

Výsledky učení

Po absolvování této přednášky budete schopni:

- rozpoznat biologickou otázku, ve které vztah jednoho prediktoru k odezvě **závisí na druhém prediktoru**;
- fitovat model s interakcí a zobrazit jeho **předpovědi** pro kombinace číselných i skupinových prediktorů;
- vyložit koeficienty modelu s interakcí jako **podmíněné** a porovnávat skupiny nebo různé hodnoty prediktoru pomocí předpovědí;
- vysvětlit interakci dvou skupinových prediktorů jako **rozdíl dvou rozdílů**;
- omezit závěr na pozorovaný rozsah dat a na souvislost v pozorovacích datech.

**Platí rovnoběžné přímky i pro
kraby?**

Proč by se samice a samci mohli lišit tvarem krunýře?

CO VÍME O KRABECH

- Zadní část krunýře leží zhruba nad zadečkem.
- Samice pod zadečkem nosí vajíčka.
- Samice a samci se liší poměrem zadní šířky a délky krunýře.

Bude rozdíl v zadní šířce mezi stejně dlouhou samicí a samcem stejný u malých i velkých krabů?

Jeden bod představuje jednoho kraba

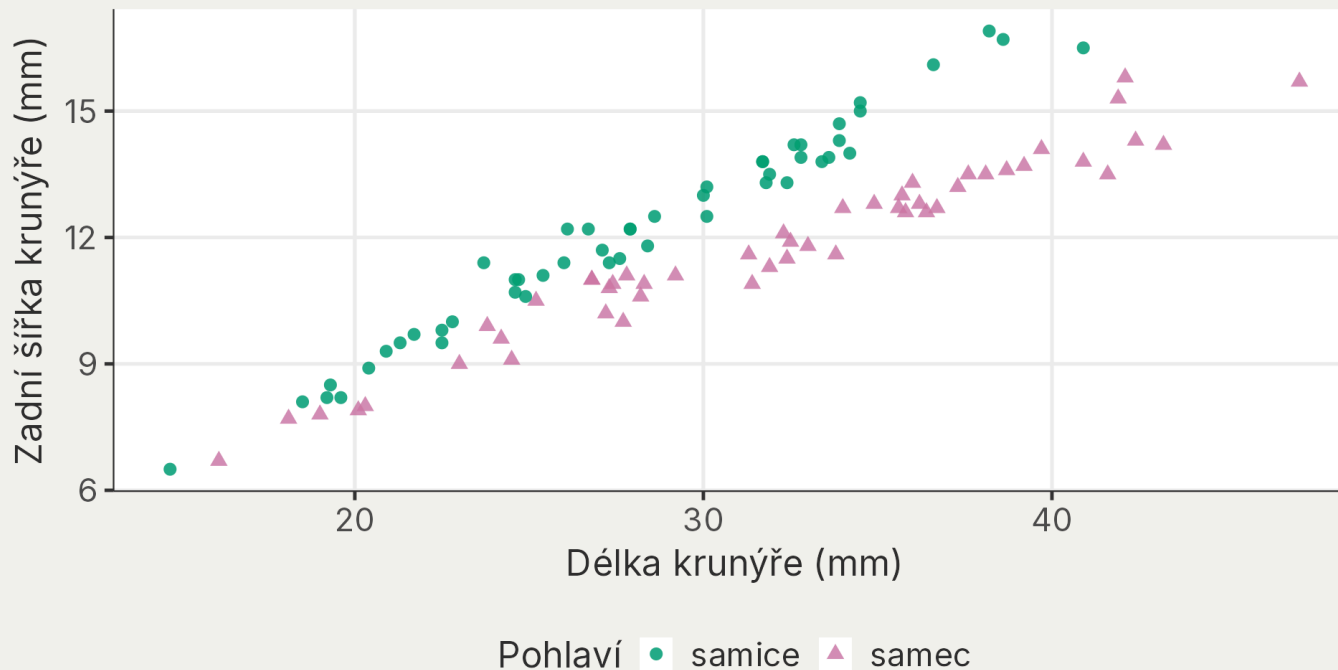
```
1 data_krabi_vse <-  
2   read.csv(file = "data/krabi.csv")  
3  
4 # Modrá forma, pohlaví, délka a zadní šířka  
5 data_krabi <-  
6   data_krabi_vse[  
7     data_krabi_vse$sp == "B",  
8     c("sex", "CL", "RW")  
9   ]  
10 names(data_krabi) <-  
11   c("pohlavi", "delka_krunyre_mm",  
12     "zadni_sirka_mm")  
13 data_krabi$pohlavi <-  
14   factor(  
15     x = data_krabi$pohlavi,  
16     levels = c("F", "M"),  
17     labels = c("samice", "samec")  
18   )  
19 head(x = data_krabi, n = 3)
```

	pohlavi	delka_krunyre_mm	zadni_sirka_mm
1	samec	16.1	6.7
2	samec	18.1	7.7
3	samec	19.0	7.8

100 KRABŮ

-  modrá barevná forma
- 50 samic · 50 samců
- délka krunýře v mm
- zadní šířka v mm

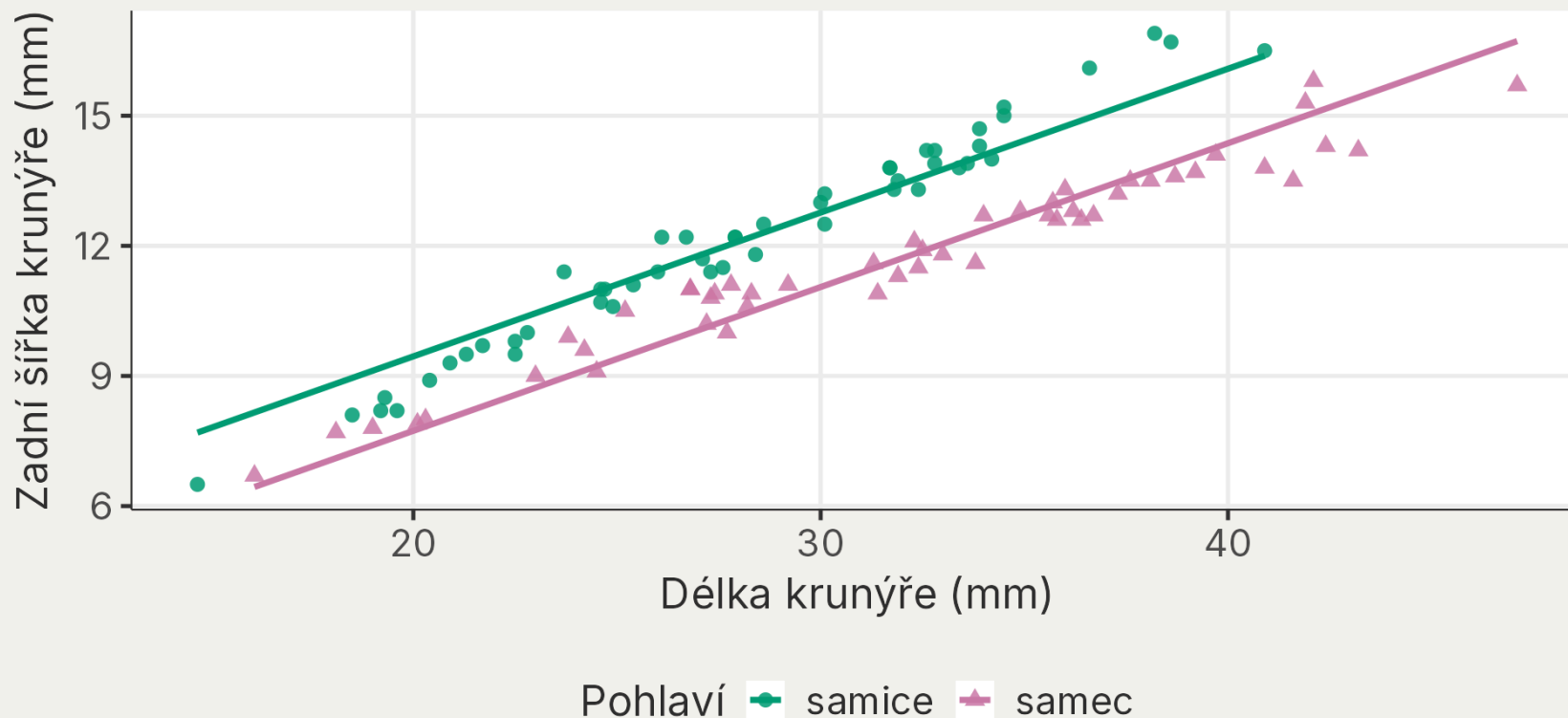
Co naznačují měření krabů?



1. Jaký je směr vztahu?
2. Kde se skupiny překrývají?
3. Vypadají oba sklony stejně?

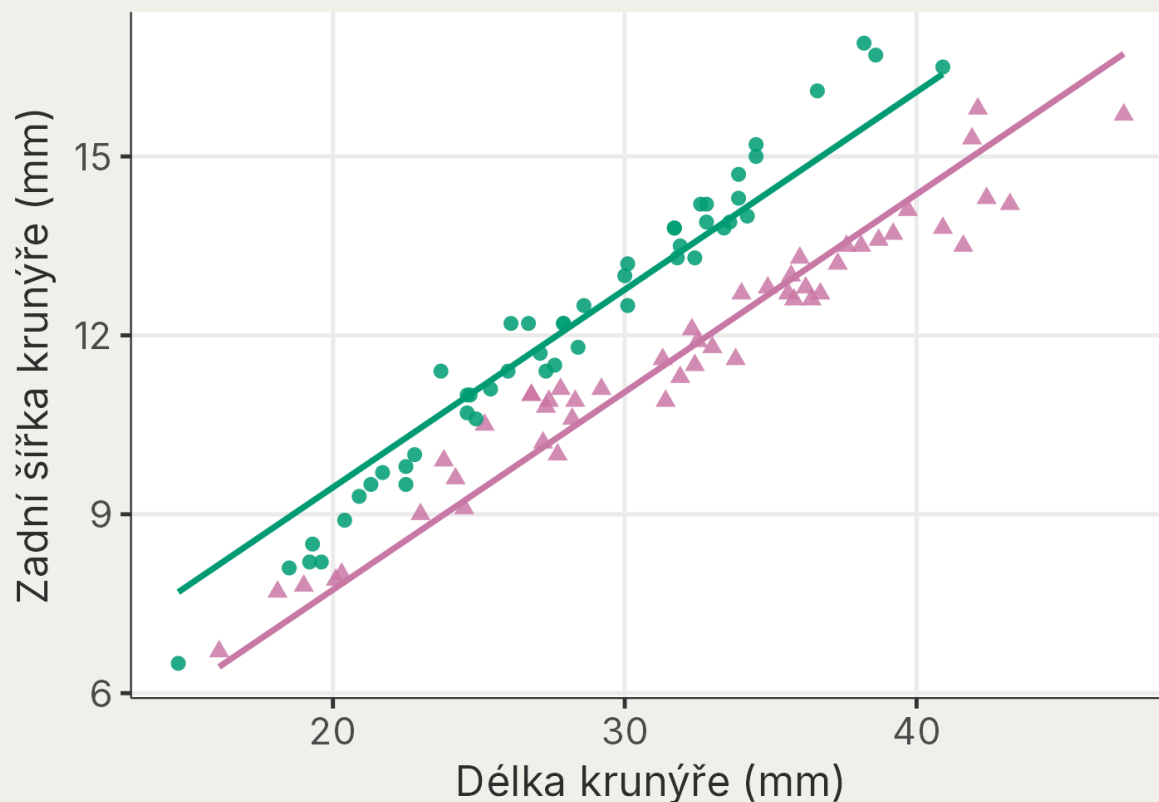
Aditivní model kreslí rovnoběžné přímky

```
1 mod_rovnobezne <-  
2   lm(  
3     formula = zadni_sirka_mm ~ delka_krunyre_mm + pohlavi,  
4     data = data_krabi  
5   )
```



+ dovoluje posun mezi skupinami, ale ukládá jim společný sklon.

Zůstává rozdíl mezi pohlavími při každé délce stejný?

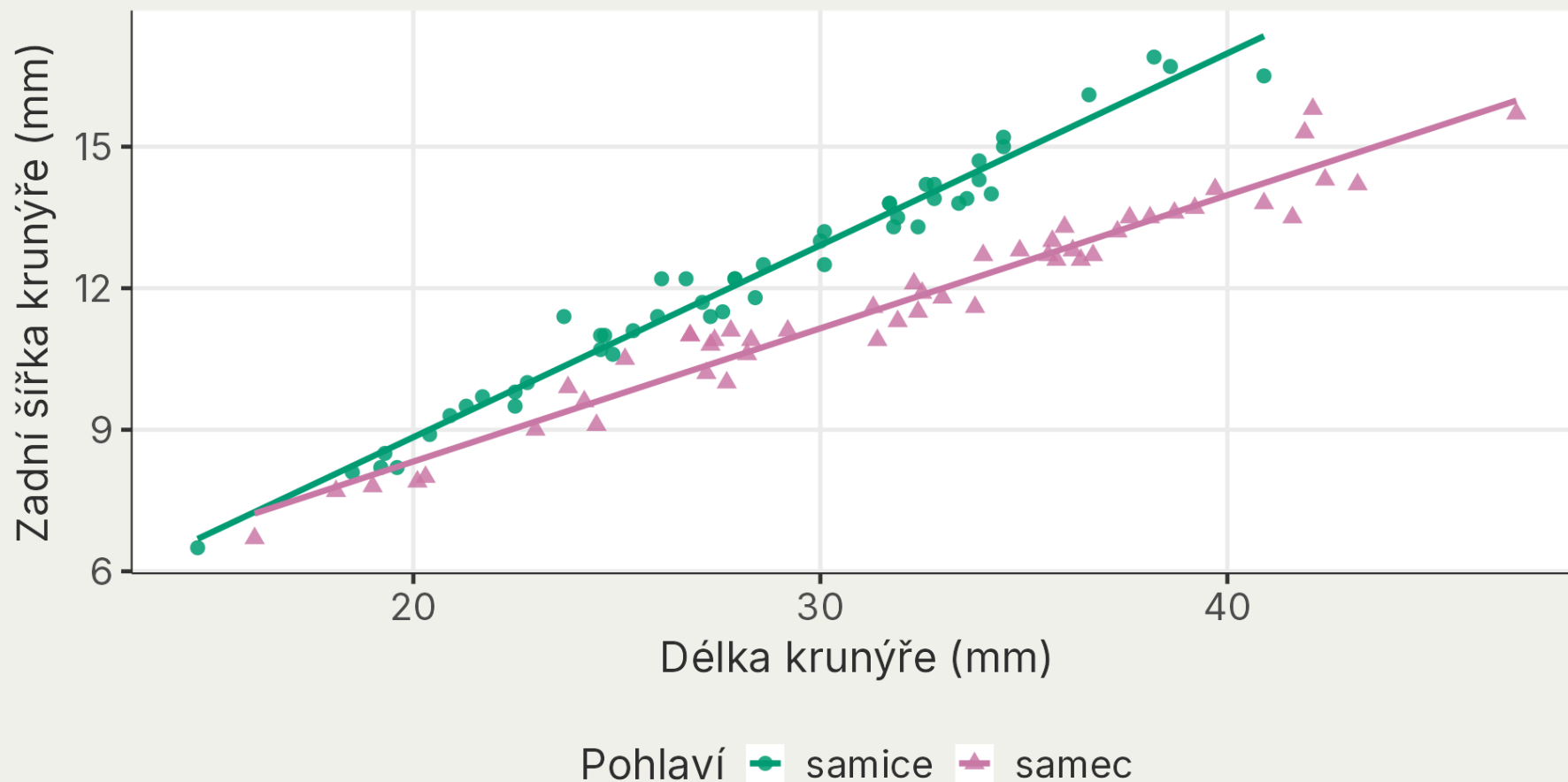


Pohlaví  samice  samec

VE DVOJICI

Porovnejte svislou vzdálenost přímek u kratších a delších krabů. Sedí přímky k bodům i na koncích rozsahu?

Data naznačují různé sklony



Svislá vzdálenost mezi přímkami se s délkou mění.

Když se rozdíl mění, mluvíme o interakci

ADITIVNÍ MODEL

Rozdíl mezi pohlavími je **při každé délce stejný**.

Rovnoběžné přímky

MODEL S INTERAKCÍ

Rozdíl mezi pohlavími se **mění s délkou**.

Různé sklony

Interakce: vztah délky a zadní šířky **závisí na pohlaví**.

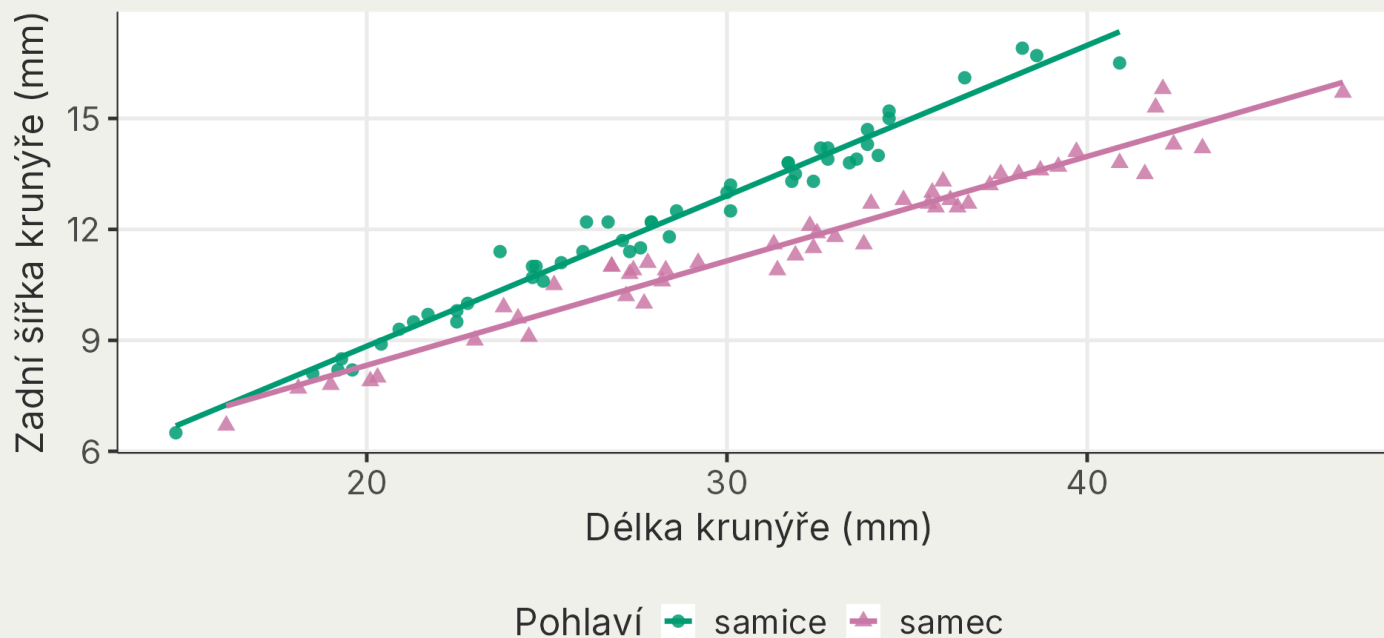
Jak zapsat model s interakcí?

1. `zadni_sirka_mm ~ delka_krunyre_mm + pohlavi`
2. `zadni_sirka_mm ~ delka_krunyre_mm + pohlavi + delka_krunyre_mm:poohlavi`
3. `zadni_sirka_mm ~ delka_krunyre_mm * pohlavi`

`x * y` je zkratka pro `x + y + x:y`; dvojtečka označuje interakční člen.

Model s interakcí dovoluje dvěma pohlavím různé sklony

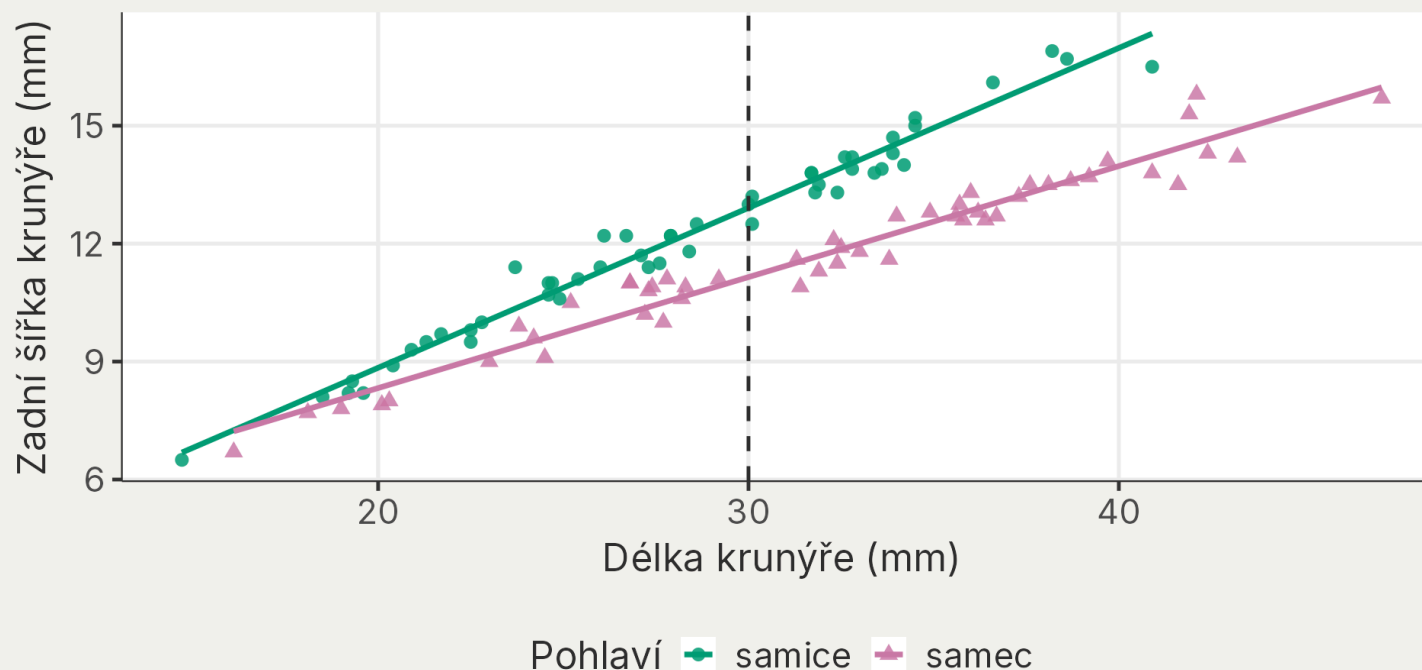
```
1 mod_interakce <-  
2   lm(  
3     formula = zadni_sirka_mm ~ delka_krunyre_mm * pohlavi,  
4     data = data_krabi  
5   )
```



Každou přímkou kreslíme jen přes délky zastoupené v její skupině.

Co model odhaduje pro dva stejně dlouhé kraby?

```
1 # Společná délka uvnitř rozsahu obou pohlaví  
2 delka_porovnani_mm <- 30
```



Která přímka je při 30 mm výše? O kolik přibližně?

Dva odhady při stejné délce krunýře

```
1 data_porovnani <-  
2   data.frame(  
3     delka_krunyre_mm = rep(x = delka_porovnani_mm, times = 2),  
4     pohlavi = c("samice", "samec")  
5   )  
6  
7 data_porovnani$odhad_sirky_mm <-  
8   predict(  
9     object = mod_interakce,  
10    newdata = data_porovnani  
11  )  
12  
13 data_porovnani
```

	delka_krunyre_mm	poohlavi	odhad_sirky_mm
1	30	samice	12.91000
2	30	samec	11.14947

Samec má při 30 mm užší zadní část krunýře

● Samice

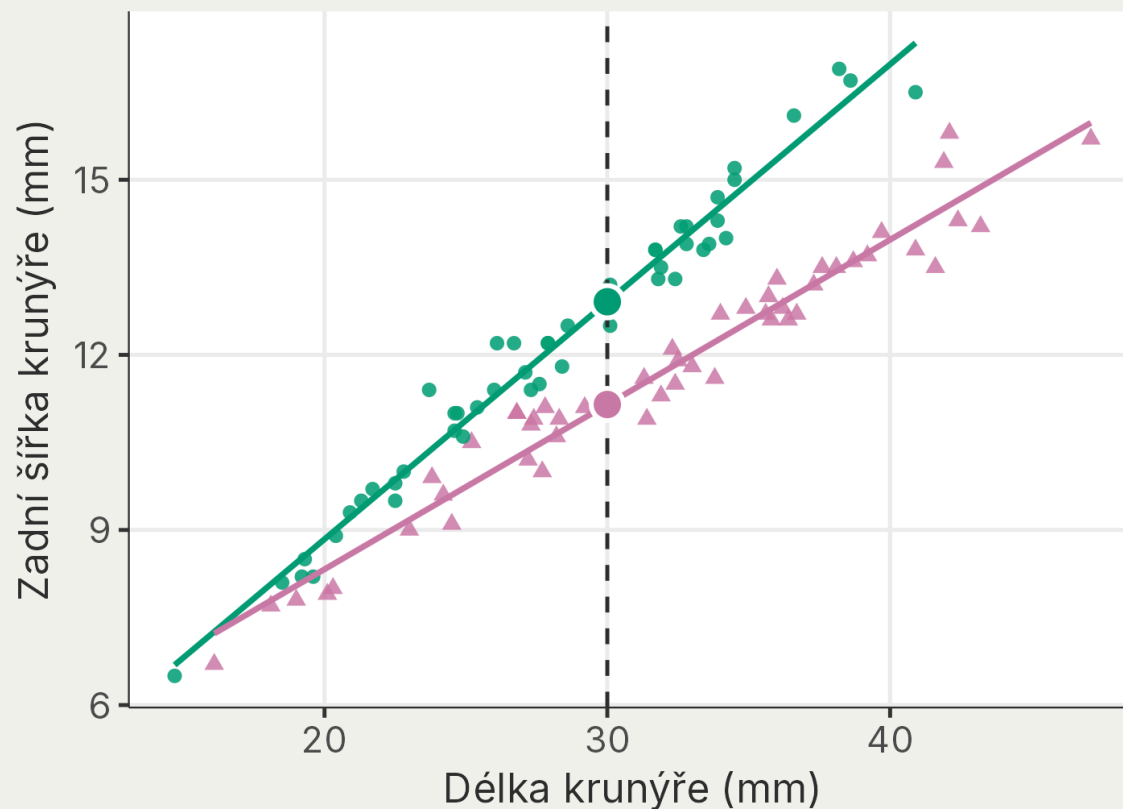
12,91 mm

▲ Samec

11,15 mm

Samec – samice

-1,76 mm



Pohlaví ● samice ▲ samec

Co už můžeme říci o stejně dlouhých krabech?

CO VÍME

- Zadní šířka roste s délkou u obou pohlaví.
- Samci mají pomalejší sklon.
- Při 30 mm má samec zadní šířku o 1,76 mm menší.

CO JEŠTĚ NEVÍME

- Jak tento výsledek vyjadřují koeficienty modelu?
- Jak přesně známe rozdíl sklonů?

Koeficienty interakčního modelu

Jaké jsou koeficienty v `summary()` ?

```
1 souhrn_interakce <-  
2   summary(object = mod_interakce)  
3  
4 round(  
5   x = souhrn_interakce$coefficients,  
6   digits = 3  
7 )
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.708	0.294	2.409	0.018
delka_krunyre_mm	0.407	0.010	39.742	0.000
pohlavisamec	1.973	0.400	4.928	0.000
delka_krunyre_mm:pohlavisamec	-0.124	0.013	-9.450	0.000

Pro výklad nám stačí sloupec **Estimate**.

První dva koeficienty popisují referenční samice

(Intercept)

0,708 mm

Odhad pro samici při délce **0 mm**
(hypotetický krab).

delka_krunyre_mm

+0,407 mm/mm

Sklon samičí přímky: samice delší o 1 mm má odhadem širší zadní část.

Koeficient pohlaví porovnává skupiny při délce 0 mm

pohlavísamec = **+1,973 mm**

Rozdíl **samec – samice**, když `delka_krunyre_mm = 0`.

Není to:

- rozdíl při každé délce;
- rozdíl při 30 mm;
- průměrný rozdíl mezi všemi samci a samicemi.

Biologicky smysluplné srovnání získáme z předpovědí uvnitř dat.

Interakční koeficient mění sklon samců vůči samicím

SKLON SAMIC

0,4067 mm/mm

ROZDÍL SKLONŮ SAMEC – SAMICE

-0,1245 mm/mm

$$\text{sklon samců} = 0,4067 + (-0,1245) \approx 0,282 \text{ mm/mm}$$

95% interval spolehlivosti rozdílu sklonů: -0,151 až -0,098 mm/mm

Rozdíl při 30 mm skládají dva známé příspěvky

Rozdíl **samec – samice při 30 mm** tvoří:

1. rozdíl mezi pohlavími při délce 0 mm: **+1,9731 mm**;
2. 30 mm × změna rozdílu o **-0,1245 mm** na každý milimetr délky.

$$\text{rozdíl při 30 mm} = 1,9731 + 30 \times (-0,1245) \approx -1,76 \text{ mm}$$

Obecně pro délku x , kde β označuje koeficient modelu:

$$\text{rozdíl}(x) = \beta_{\text{samec}} + x \beta_{\text{délka:samec}}$$

Platí koeficient **pohlavísamec** při každé délce?

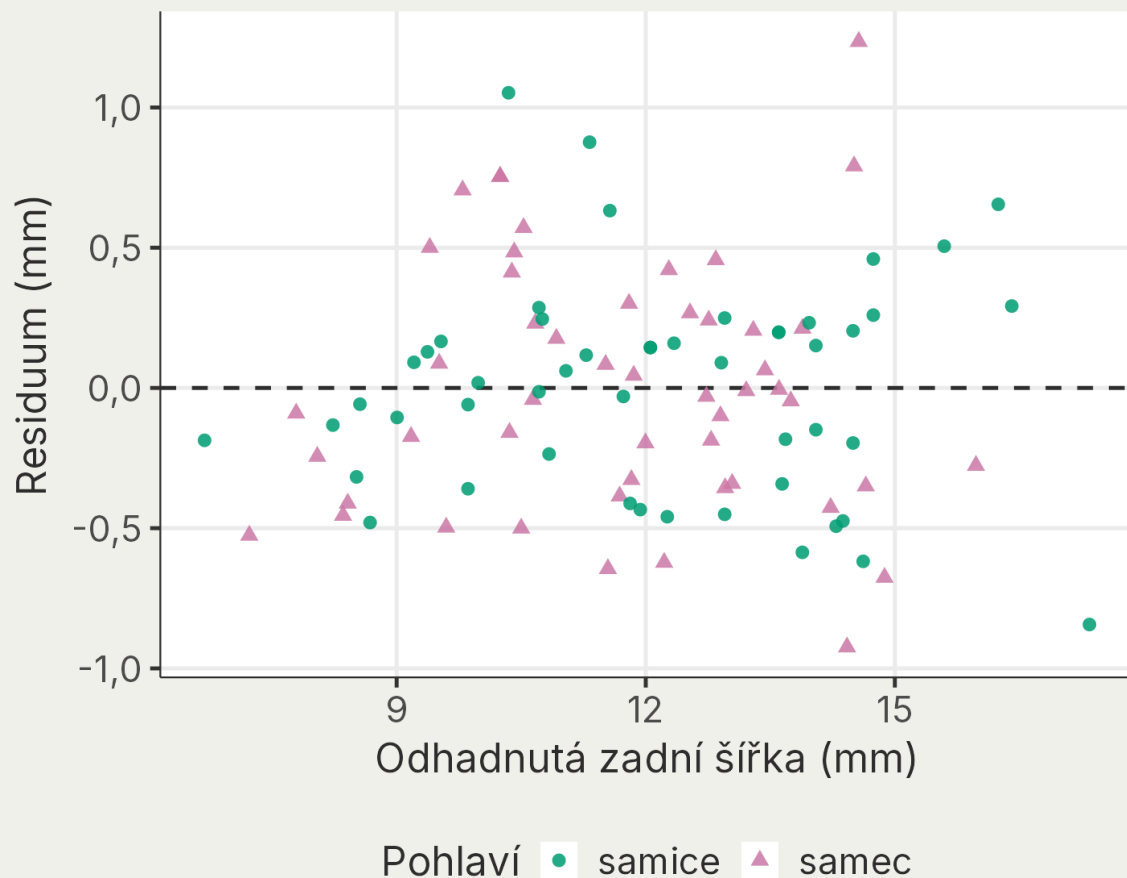
A) Ano, při každé délce.

B) Ne, pouze při délce 0 mm.

C) Ne, pouze při délce 30 mm.

Při jiné délce musíme přidat příspěvek **interakčního koeficientu**.

Jak dobře přímky popisují změřené kraby?



Residua leží kolem nuly **bez oblouku** a s podobným rozptylem u malých i velkých krabů.

Graf neověří příčinu ani platnost mimo pozorované délky.

Co víme o rozdílu mezi samicemi a samci modré formy?

Bude rozdíl v zadní šířce mezi stejně dlouhou samicí a samcem stejný u malých i velkých krabů?

CO DATA UKAZUJÍ

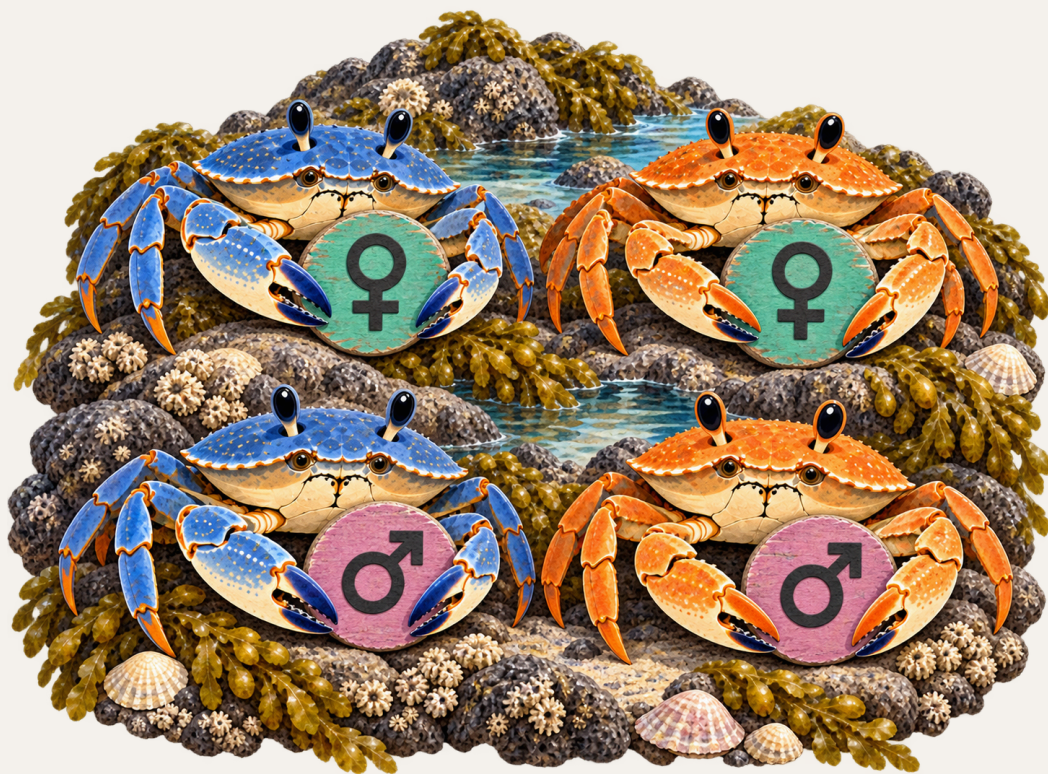
- Sklon samic 0,407, samců 0,282 mm/mm.
- Rozdíl samec – samice se mění o **-0,124 mm** na 1 mm délky (95% interval spolehlivosti -0,151 až -0,098).
- Při 30 mm: samec – samice **-1,76 mm**.

CO ZŮSTÁVÁ OTEVŘENÉ

- Proč se pohlaví liší: data jsou pozorovací.
- Platí to i pro kraby mimo tento vzorek?
- A pro kraby menší nebo větší, než jsme změřili?

Když jsou oba prediktory skupiny

Liší se rozdíl mezi pohlavími u modré a oranžové formy?



Ilustrace vytvořená pomocí AI.

ODEZVA

Délka krunýře (mm)

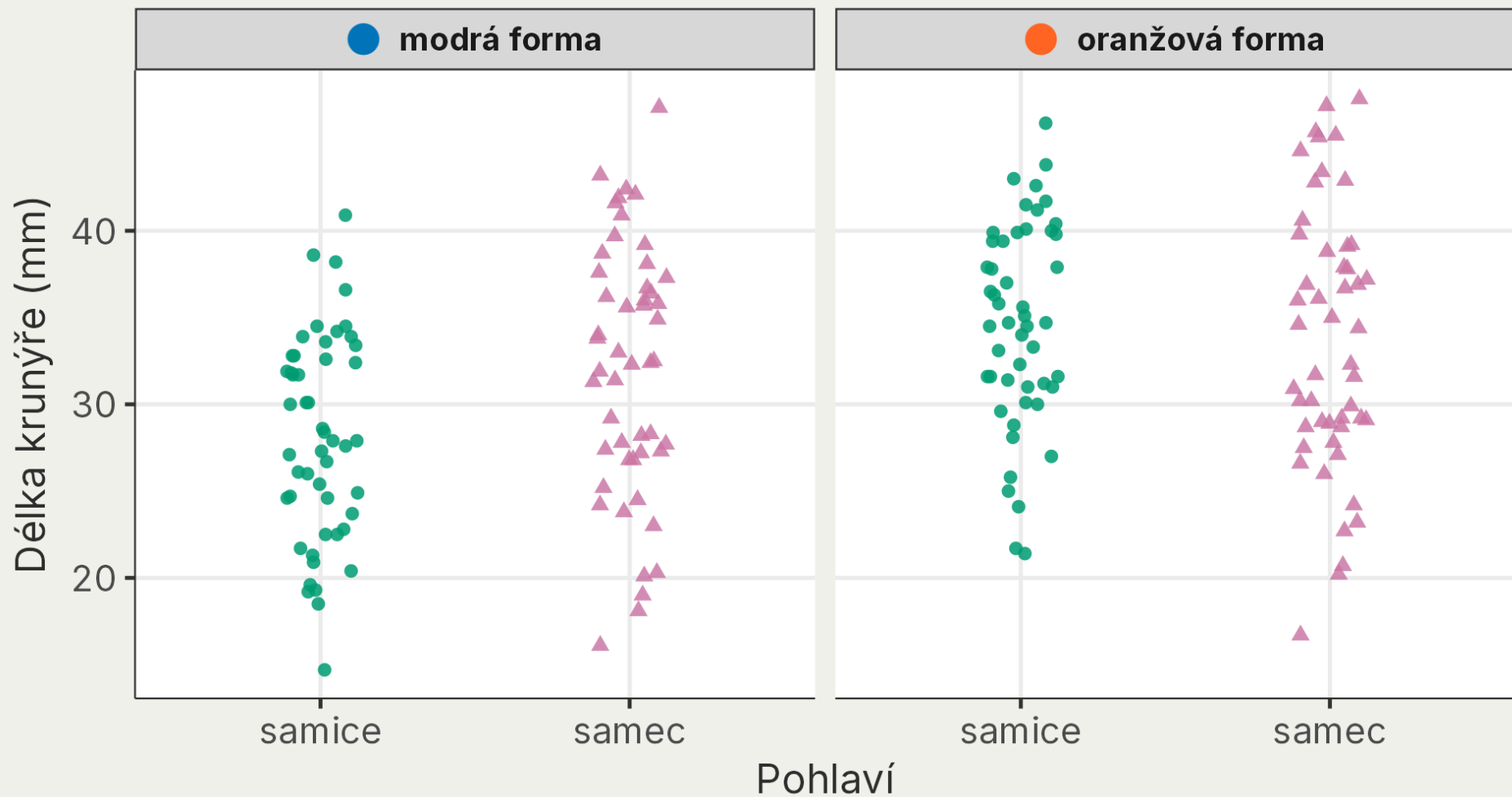
DVA SKUPINOVÉ PREDIKTORY

Barevná forma · pohlaví

Mění se rozdíl **samec – samice** mezi barevnými formami?

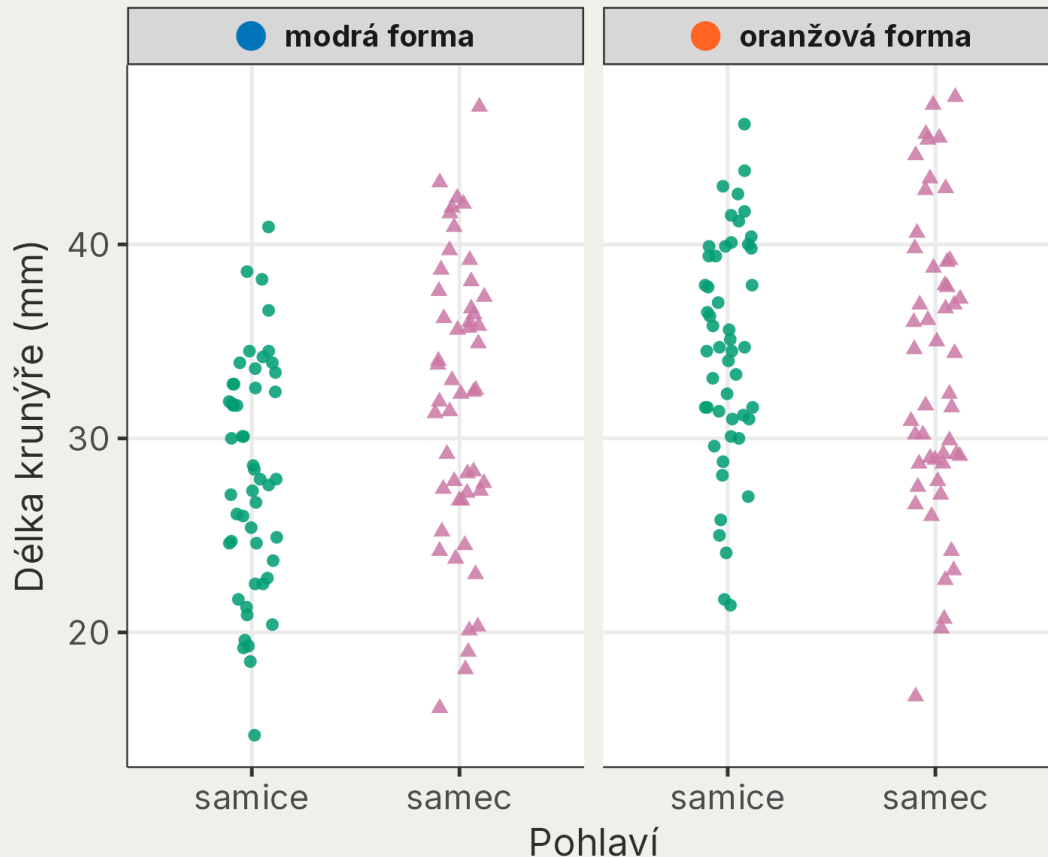
Do každé skupiny vybrali autoři stejný počet krabů, takže průměrné délky závisejí i na výběru.

Jaké délky vidíme ve čtyřech skupinách?



Každý bod je jeden krab. Autoři vybrali do každé skupiny 50 krabů.

Je rozdíl mezi samci a samicemi u obou forem stejný?

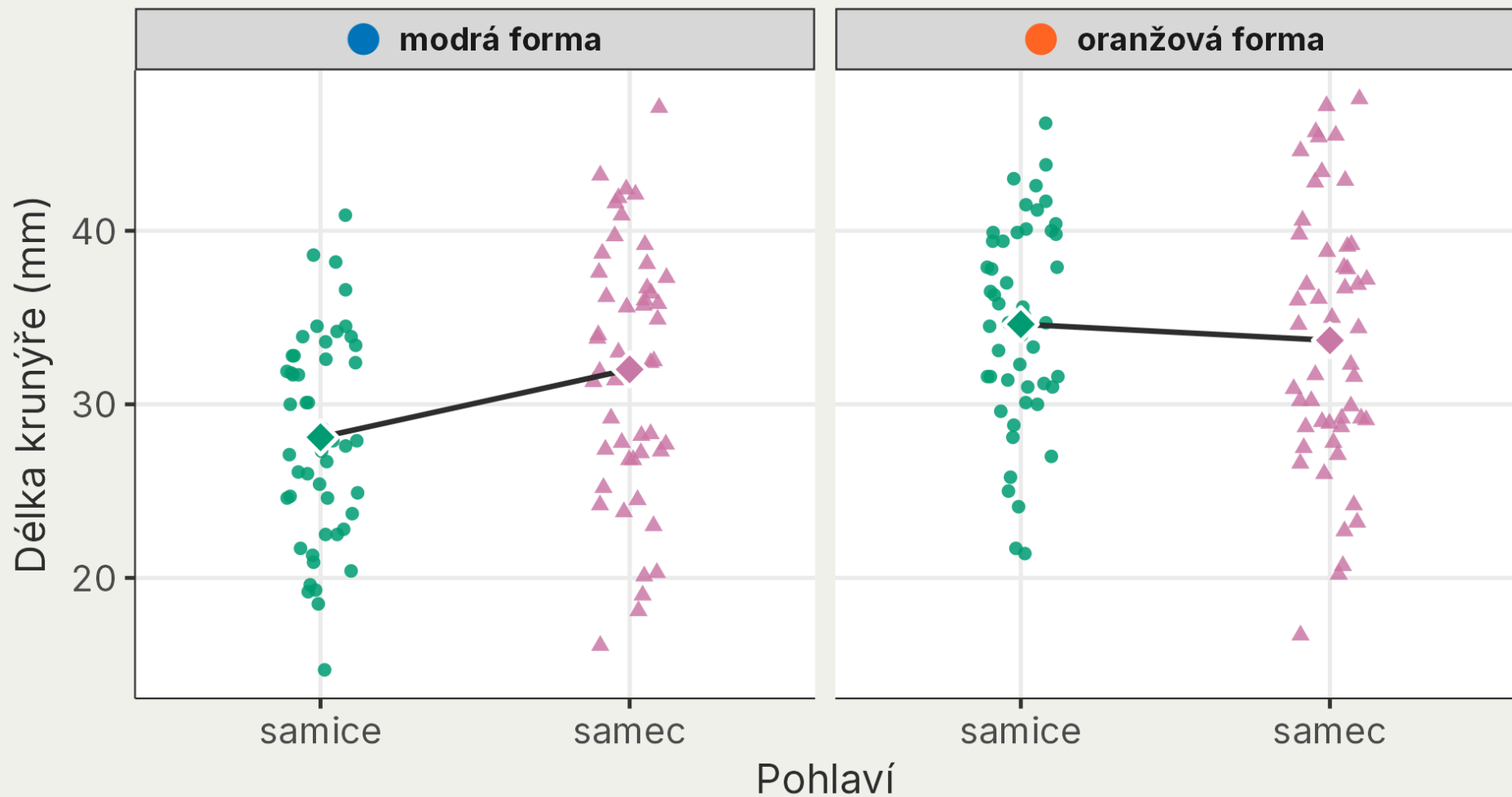


MINUTA BEZ HOVORU

Napište pro každou formu:

- kdo má vyšší průměr;
- zda čekáte podobně velký rozdíl;
- jednu větu, kterou z bodů říci nemůžeme.

Průměry ukazují dva různé rozdíly



Body, osy i náhodné posunutí jsou stejné. Přibyly pouze průměry a spojnice.

Jak velký je rozdíl samec minus samice v každé formě?

● MODRÁ FORMA

$$32,01 - 28,10 = +3,91 \text{ mm}$$

● ORANŽOVÁ FORMA

$$33,69 - 34,62 = -0,93 \text{ mm}$$

U modré formy jsou samci v průměru delší, u oranžové o něco kratší.

Interakce je rozdíl dvou rozdílů

Rozdíl dvou rozdílů Δ = rozdíl u  oranžové formy – rozdíl u  modré formy

$$\Delta = (-0,93) - (+3,91) = -4,84 \text{ mm}$$

Rozdíl mezi pohlavími se při přechodu od modré k oranžové formě změní o **-4,84 mm**.

forma * pohlavi zapíše všechny čtyři kombinace

forma + pohlavi + forma:poohlavi = forma * pohlavi; referenční kombinace je modrá forma a samice.

```
1 mod_formy <-  
2   lm(  
3     formula = delka_krunyre_mm ~ forma * pohlavi,  
4     data = data_formy  
5   )  
6  
7   round(  
8     x = summary(object = mod_formy)$coefficients,  
9     digits = 3  
10  )
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	28.102	0.950	29.584	0.000
formaoranžová	6.516	1.343	4.851	0.000
poohlavisamec	3.912	1.343	2.912	0.004
formaoranžová:poohlavisamec	-4.842	1.900	-2.549	0.012

Který údaj je rozdíl samec – samice v oranžové formě?

Koeficient	Odhad (mm)
(Intercept)	28,102
formaoranžová	+6,516
pohlavísamec	+3,912
formaoranžová:pohlavísamec	-4,842

A) pohlavísamec

B) formaoranžová:pohlavísamec

C) pohlavísamec +
formaoranžová:pohlavísamec

D) formaoranžová

$+3,91 + (-4,84) = -0,93 \text{ mm}$: stejný rozdíl, jaký jsme spočítali z průměrů oranžové formy.

Co znamená každý ze čtyř koeficientů?

(Intercept): ● modrá samice

28,10 mm

pohlavisamec: samec – samice v

● modré formě

+3,91 mm

formaoranžová: ● oranžová – ●
modrá mezi samicemi

+6,52 mm

formaoranžová:pohlavisamec:
změna rozdílu mezi formami

–4,84 mm

95% interval spolehlivosti interakce: –8,59 až –1,10 mm

Odhad pro každou ze čtyř skupin

```
1 prumery_formy <-  
2   emmeans::emmeans(  
3     object = mod_formy,  
4     specs = ~ pohlavi | forma  
5   )  
6  
7 prumery_formy
```

forma = modrá:

poohlavi	emmean	SE	df	lower.CL	upper.CL
samice	28.1	0.95	196	26.2	30.0
samec	32.0	0.95	196	30.1	33.9

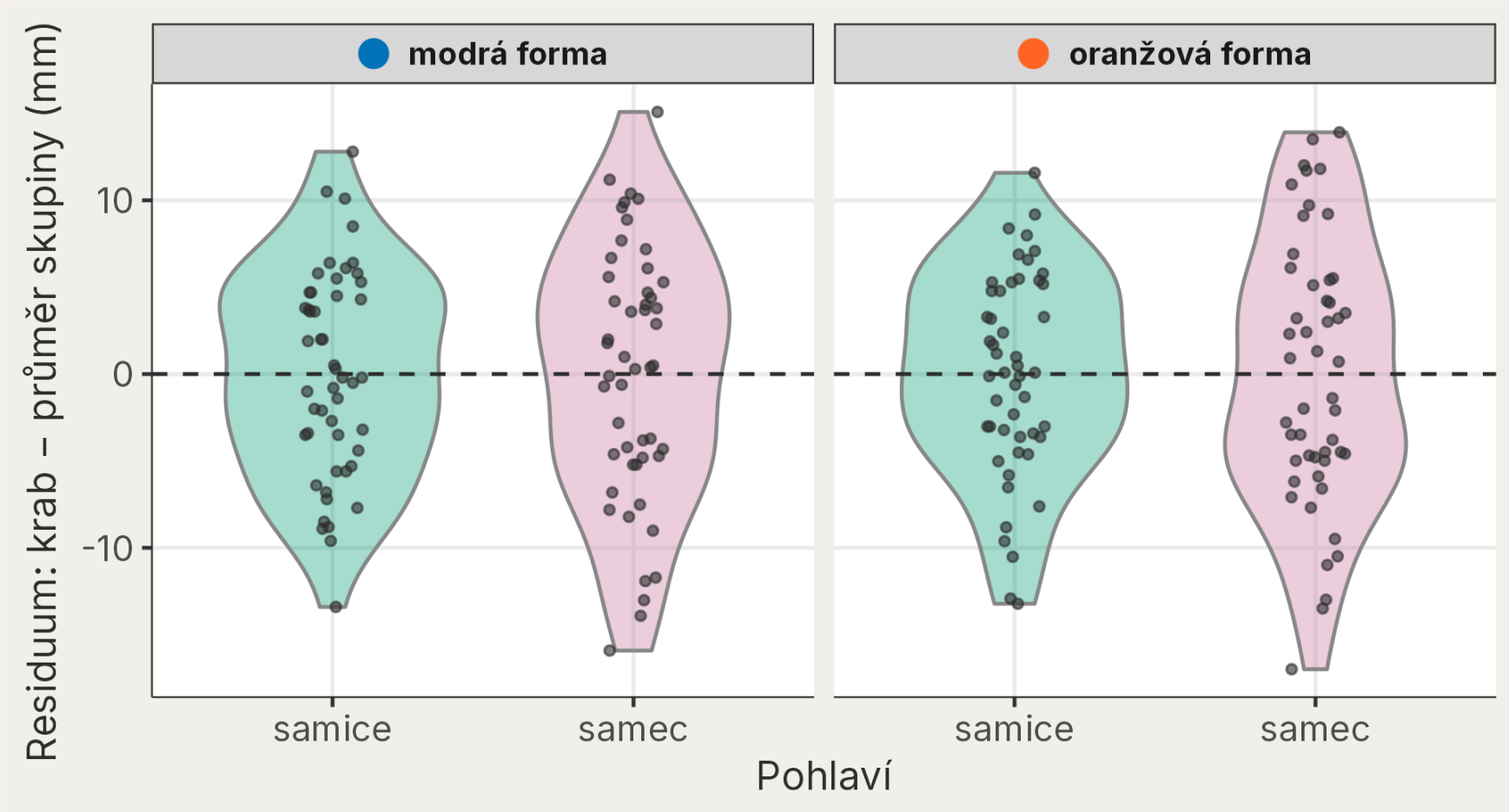
forma = oranžová:

poohlavi	emmean	SE	df	lower.CL	upper.CL
samice	34.6	0.95	196	32.7	36.5
samec	33.7	0.95	196	31.8	35.6

Confidence level used: 0.95

Model má jeden parametr pro každou kombinaci, proto odhady odpovídají pozorovaným skupinovým průměrům.

Residua jsou odchylky jednotlivých krabů od skupinových průměrů



Residuum = délka kraba - průměr jeho skupiny.

Rozdíl mezi pohlavími závisí na barevné formě

- ● Modrá forma: samec – samice = **+3,91 mm**
- ● Oranžová forma: samec – samice = **-0,93 mm**
- Změna rozdílu: **-4,84 mm**; 95% interval spolehlivosti -8,59 až -1,10 mm

Průměrné délky závisejí i na tom, které kraby autoři vybrali. Výsledek popisuje tento soubor, ne velikost krabů v přírodě.

Když jsou oba prediktory čísla

Přibude vysokému stromu s každým centimetrem průměru víc dřeva?

STŘEMCHA POZDNÍ

- severoamerický, u nás invazní strom
- objem dřeva z **průměru kmene** a **výšky**
- 31 pokácených stromů, z amerických jednotek převedeno na cm, m a m³

VE DVOJICI

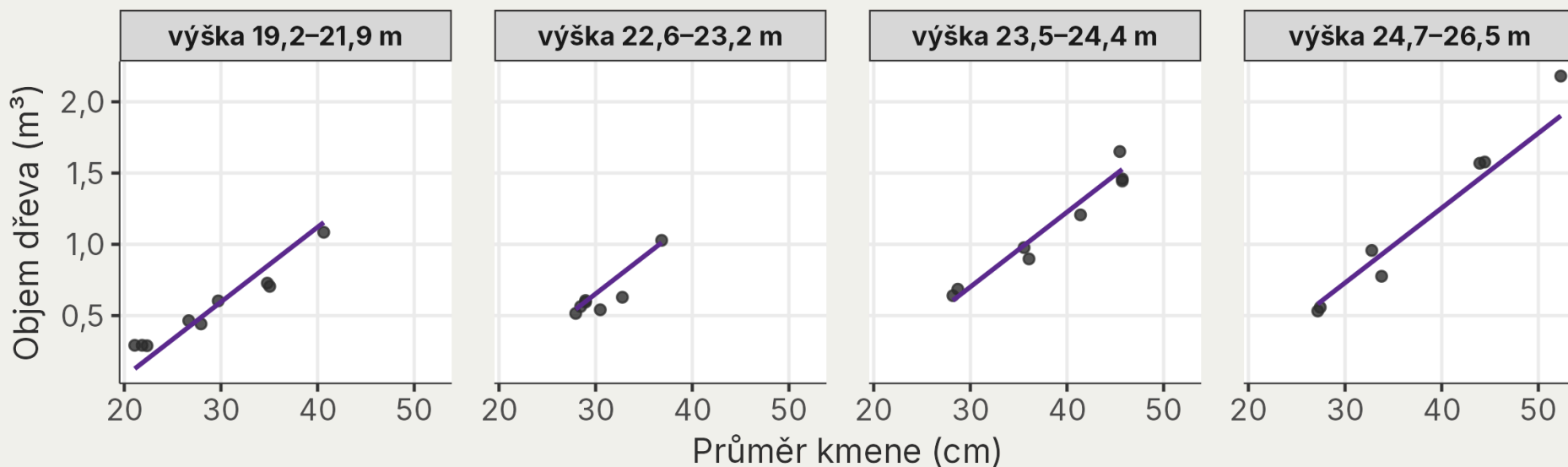
Dva stejně tlusté kmeny ztloustnou o 1 cm. Jeden strom je nízký, druhý vysoký.

Kterému přibude **víc dřeva**?

Vysokému: nová vrstva dřeva obalí delší kmen.

Vystihne společný sklon nízké i vysoké stromy?

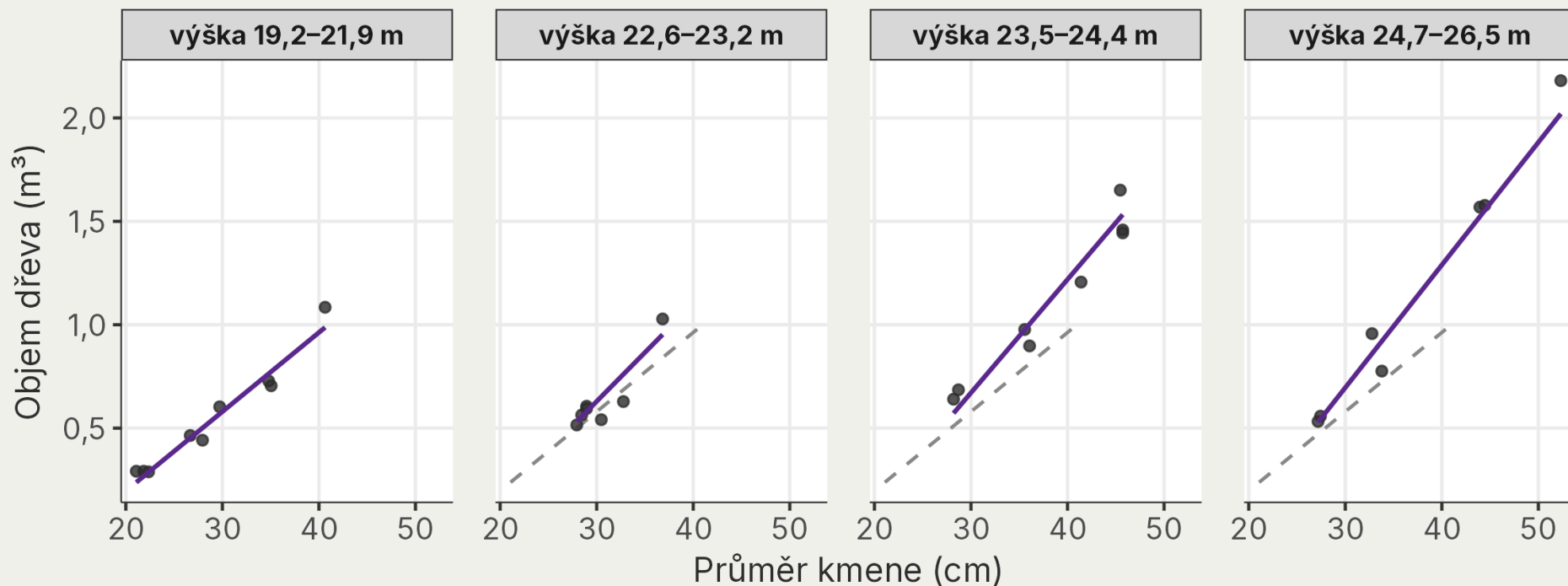
```
1 mod_aditivni <-  
2   lm(  
3     formula = objem_m3 ~ prumer_cm + vyska_m,  
4     data = data_stromy  
5   )
```



Ve dvojici: kde leží tenké a tlusté kmeny v krajních panelech?

Panel = výšková čtvrtina stromů · přímka = model při střední výšce panelu

Mění se sklon přímky s výškou stromu?



Čím vyšší strom, tím strmější přímka.

Šedá přerušovaná = přímka nejnižšího panelu · sklon = m³ na 1 cm průměru

prumer_cm * vyska_m: sklon průměru závisí na výšce

```
1 mod_stromy <-  
2   lm(  
3     formula = objem_m3 ~ prumer_cm * vyska_m,  
4     data = data_stromy  
5   )  
6 round(x = coef(object = mod_stromy), digits = 5)
```

(Intercept)	prumer_cm	vyska_m	prumer_cm:vyska_m
1.96508	-0.06528	-0.12050	0.00493

sklon průměru = prumer_cm + prumer_cm:vyska_m × výška

Výška stromu

21 m

23 m

25 m

Sklon průměru (m³ na 1 cm)

0,038

0,048

0,058

Pro který strom platí koeficient `prumer_cm`?

`prumer_cm` = **-0,065 m³ na 1 cm**

- A)** Pro každý změřený strom.
- B)** Jen pro nejnižší stromy v datech.
- C)** Pro žádný: tlustší kmen má méně dřeva.
- D)** Jen pro strom vysoký 0 m.

Takový strom v datech není. Sklon průměru proto uvádíme **při zvolené výšce** uvnitř dat.

Co víme o stromech?

Přibude vysokému stromu s každým centimetrem průměru víc dřeva?

CO DATA UKAZUJÍ

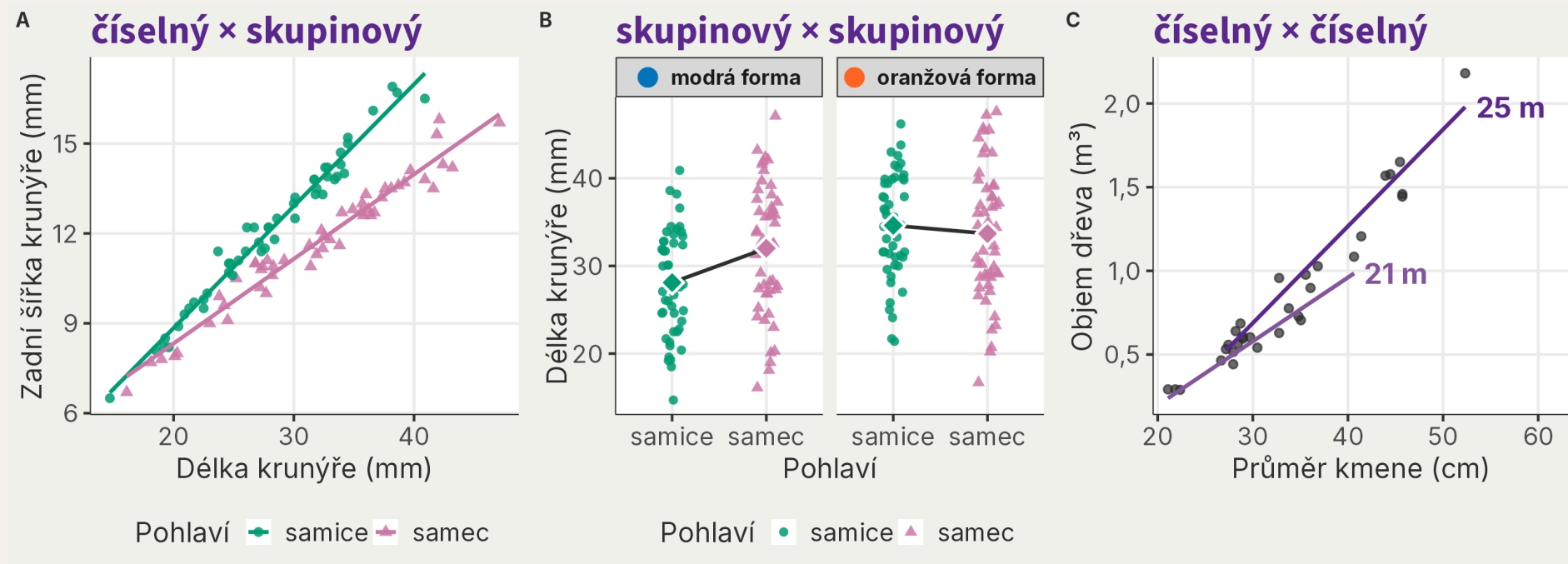
- Kmen 36 cm místo 28 cm podle modelu: $+0,31 \text{ m}^3$ při 21 m, $+0,46 \text{ m}^3$ při 25 m.
- Sklon průměru (m^3 na 1 cm) roste o $0,0049$ na každý 1 m výšky (95% interval spolehlivosti 0,0031 až 0,0068).

CO ZŮSTÁVÁ OTEVŘENÉ

- 31 stromů bez popsání výběru.
- Rezidua ani vliv jednotlivých stromů jsme nekontrolovali.
- Nepopsal by vztah lépe zakřivený model?

Jedna myšlenka, tři obrazy

Interakce mění sklon nebo skupinový rozdíl



Co se mění? Rozdíl mezi přímkami podél osy délky.

Co se mění? Rozdíl mezi pohlavími podle formy.

Co se mění? Sklon průměru s výškou stromu.

Co musíme doplnit, aby tvrzení o efektu bylo úplné?

NEÚPLNÉ

„Délka má efekt.“

„Samci jsou větší.“

„Průměr kmene má efekt.“

OPRAVTE TVRZENÍ

- Pro které pohlaví?
- Při jaké délce?
- V které barevné formě?
- Větší v jakém znaku?
- Při jaké výšce?

V modelu s interakcí vždy uvádíme hodnotu nebo úroveň druhého prediktoru.

Jak číst model s interakcí?

1. OTÁZKA

Mění se **vztah nebo rozdíl** podle druhého prediktoru?

2. DATA A PŘEDPOVĚDI

Ukažte body a **předpovědi pro konkrétní kombinace**.

3. KOEFICIENTY

Hlavní koeficient platí v **referenční skupině nebo při nule** druhého prediktoru.

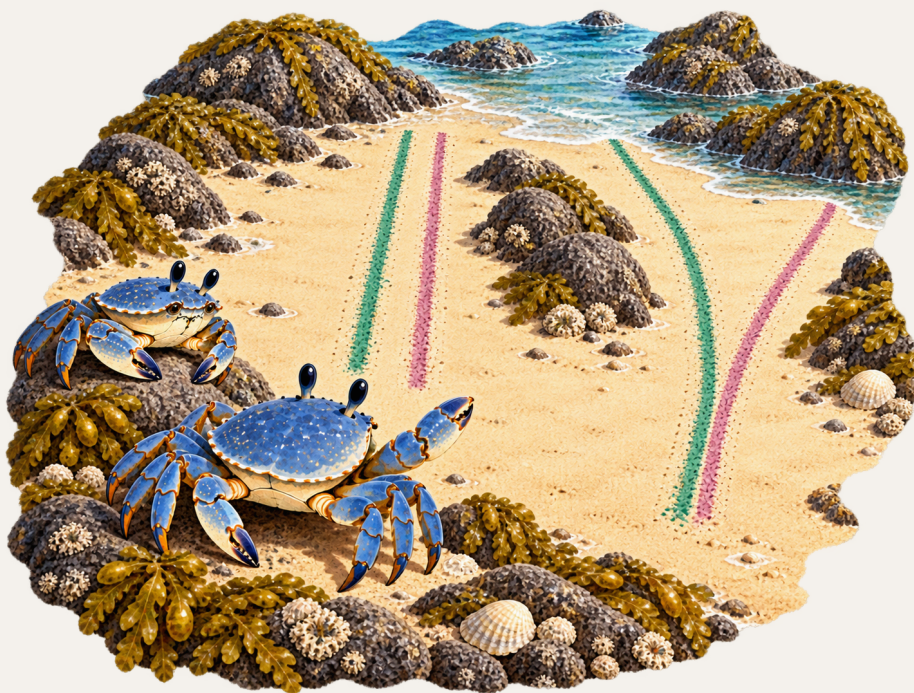
4. ZÁVĚR

Jen v rozsahu dat a jako souvislost, ne příčina.

Co si odnést

- Interakci potřebujeme, když vztah jednoho prediktoru **závisí na druhém**; podpořit ji má biologická otázka i graf.
- $x * y$ je zkratka pro $x + y + x:y$.
- Hlavní koeficienty jsou **podmíněné**; skupiny porovnáváme pomocí předpovědí při zvolené hodnotě.
- U dvou skupinových prediktorů je interakce **rozdíl dvou rozdílů**.
- U dvou číselných prediktorů se **sklon jednoho plynule mění** s hodnotou druhého.
- Závěr omezujeme na pozorovaný rozsah a na souvislost v pozorovacích datech.

Který biologicky smysluplný model máme zvolit?



Ilustrace vytvořená pomocí AI.

Aditivní kandidát

```
zadni_sirka_mm ~  
delka_krunyre_mm + pohlavi
```

Stálý rozdíl

Interakční kandidát

```
zadni_sirka_mm ~  
delka_krunyre_mm * pohlavi
```

Měnící se rozdíl

Jak mezi kandidáty vybírat, ukáže následující lekce.

Kde byste čekali interakci vy?

Uved'te biologickou otázku, u které byste čekali, že vztah dvou znaků závisí na skupině nebo na hodnotě jiného znaku.

Jak by vypadal graf?