

Kdy vztah dvou znaků závisí na třetím?

Interakce číselného a skupinového znaku, dvou skupinových i dvou číselných znaků

17. října 2026

Obsah

1	Liší se vztah délky a zadní šířky krunýře mezi samicemi a samci?	2
2	Kdy předpovídají dvě přímky stejný sklon?	2
3	Výsledky učení	2
4	Jeden řádek tabulky představuje jednoho kraba	3
5	Které délky a šířky jsou v datech zastoupené?	5
6	Jak souvisí délka a zadní šířka krunýře?	6
7	Rovnoběžné přímky předpokládají společný sklon	8
8	Mění se rozdíl mezi samicemi a samci s délkou krunýře?	9
9	Model s interakcí dovoluje samicím a samcům různé sklony	10
10	Co model předpovídá pro stejně dlouhou samici a samce?	11
11	Co znamenají koeficienty modelu s interakcí?	13
12	Jak dobře přímky popisují změřené kraby?	17
13	Liší se rozdíl v délce krunýře mezi pohlavími u obou barevných forem?	19
14	Jaké délky krunýře vidíme ve čtyřech skupinách?	21
15	Jak model zachytí změnu rozdílu mezi pohlavími podle formy?	22
16	Co znamenají čtyři koeficienty modelu se dvěma skupinovými znaky?	24
17	Co o jednotlivých krabech čtyři průměry neříkají?	27
18	Přibývá objem dřeva s průměrem kmene stejně rychle u nízkých i vysokých stromů?	28
19	Vystihne společný sklon nízké i vysoké stromy?	30
20	Model s interakcí dovolí, aby se sklon průměru měnil s výškou	31
21	Co znamená koeficient průměru, když se sklon mění s výškou?	33
22	Kdy musíme při výkladu jednoho prediktoru uvést hodnotu druhého?	34
23	Co tedy můžeme říct o krabech a stromech?	35
24	Co bude následovat dál	36
25	Shrnutí	36
25.1	Závěrečná otázka	36
	Slovníček pojmů	36

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

Představte si dva kraby se stejně dlouhým krunýřem, jednoho samce a jednu samici. Musí mít stejně širokou zadní část krunýře? A bude případný rozdíl stejný u kratších i delších krabů? To je otázka **vztahu dvou měřených znaků v různých skupinách**, nikoli otázka, které pohlaví je obecně větší.

Proč bychom rozdíl vůbec čekali? Zadní část krunýře leží zhruba nad zadečkem, který má krab podvinutý pod tělem. Samice pod zadečkem nosí oplozená vajíčka, dokud se z nich nevylíhnou larvy. Campbell a Mahon (1974), z jejichž měření budeme vycházet, zjistili, že se samice a samci pobřežních krabů rodu *Leptograpsus* liší poměrem zadní šířky a délky krunýře. Pokud samice s rostoucí velikostí potřebují pro snůšku relativně více místa, mohl by se rozdíl mezi pohlavími s velikostí krabů měnit. Je to biologické očekávání; zda mu odpovídají i naše data, ukáží až měření.

Hlavní otázka této lekce

Je rozdíl v zadní šířce krunýře mezi stejně dlouhou samicí a samcem stejný u menších i větších krabů?

Zadní šířku budeme odhadovat z délky krunýře a pohlaví. Data jsou pozorovací: samotný rozdíl mezi skupinami nám neřekne, co jej způsobilo.

2 Kdy předpovídají dvě přímky stejný sklon?

V minulé lekci jste viděli model se dvěma prediktory: číselným znakem a skupinou. Jeho dvě předpovědní přímky byly rovnoběžné. Jak se podle takového modelu změní rozdíl mezi skupinami, když se číselný prediktor zvětší?

Vybavte si minulou lekci

Nakreslete dvě rovnoběžné přímky. Vyberte na ose x dvě různé hodnoty. Je svislá vzdálenost mezi přímkami v obou místech stejná?

Možná odpověď

Ano. Rovnoběžné přímky mají společný sklon, takže jejich svislá vzdálenost zůstává stejná.

Zkusme zjistit, zda takové omezení dává smysl i pro kraby.

3 Výsledky učení

Po prostudování této lekce budete schopni:

- rozpoznat biologickou otázku, ve které vztah jednoho prediktoru k odezvě závisí na hodnotě nebo úrovni druhého prediktoru;
- fitovat lineární 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 předpovědi při zvolené hodnotě druhého prediktoru;
- vysvětlit interakci dvou skupinových prediktorů jako rozdíl dvou rozdílů;
- omezit závěr na pozorovaný rozsah dat a na souvislost v pozorovacích datech.

4 Jeden řádek tabulky představuje jednoho kraba

Použijeme měření krabů z práce Campbell a Mahon (1974). Autoři změřili stejný počet samic a samců dvou barevných forem pobřežního kraba, modré a oranžové, které považují za samostatné druhy rodu *Leptograpsus*. Soubor [krabi.csv](#) je součástí materiálů lekce. Pro následující kód jej uložte do složky `data` ve svém projektu v R. Načteme ho a zkontrolujeme názvy a typy sloupců:

```
# Načteme tabulku všech změřených krabů
data_krabi_vse <-
  read.csv(file = "data/krabi.csv")

# Zkontrolujeme názvy a typy sloupců
str(object = data_krabi_vse)
```

```
'data.frame':  200 obs. of  8 variables:
 $ sp   : chr  "B" "B" "B" "B" ...
 $ sex  : chr  "M" "M" "M" "M" ...
 $ index: int   1 2 3 4 5 6 7 8 9 10 ...
 $ FL   : num  8.1 8.8 9.2 9.6 9.8 10.8 11.1 11.6 11.8 11.8 ...
 $ RW   : num  6.7 7.7 7.8 7.9 8 9 9.9 9.1 9.6 10.5 ...
 $ CL   : num  16.1 18.1 19 20.1 20.3 23 23.8 24.5 24.2 25.2 ...
 $ CW   : num  19 20.8 22.4 23.1 23 26.5 27.1 28.4 27.8 29.3 ...
 $ BD   : num  7 7.4 7.7 8.2 8.2 9.8 9.8 10.4 9.7 10.3 ...
```

Sloupec `sp` označuje barevnou formu (B modrá, O oranžová), `sex` pohlaví (F samice, M samec). `CL` je délka krunýře a `RW` šířka jeho zadní části; obě hodnoty jsou v milimetrech. Délka a zadní šířka jsou dvě různé osy téhož krunýře. Ostatní rozměry v této lekci nepotřebujeme.

Pro první otázku vybereme jen kraby **modré formy** a tři potřebné sloupce. Jeden řádek znamená **jednoho změřeného kraba**, ne opakované měření téhož jedince.

```
# Vybereme modrou formu a tři znaky potřebné pro otázku
data_krabi <-
  data_krabi_vse[
```

```

    data_krabi_vse$sp == "B",
    c("sex", "CL", "RW")
  ]

# Přejmenujeme sloupce podle toho, co znamenají
names(data_krabi) <-
  c("pohlavi", "delka_krunyre_mm", "zadni_sirka_mm")

# Pohlaví převedeme na faktor; první úroveň bude referenční skupinou
data_krabi$pohlavi <-
  factor(
    x = data_krabi$pohlavi,
    levels = c("F", "M"),
    labels = c("samice", "samec")
  )

# Každý řádek tabulky patří jednomu krabovi
head(x = data_krabi, n = 6)

```

	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
4	samec	20.1	7.9
5	samec	20.3	8.0
6	samec	23.0	9.0

Doplňující: jak vznikl soubor krabi.csv

Soubor `krabi.csv` je nezměněná kopie tabulky `crabs` z balíčku `MASS`, který se instaluje spolu s R. Přípravený soubor používáme proto, aby všichni pracovali se stejnou verzí dat. Stejnou tabulku si můžete vytvořit sami; kód spusťte v hlavní složce svého projektu:

```

# Tabulka je součástí balíčku MASS
data_krabi_zdroj <-
  as.data.frame(x = MASS::crabs)

# Kontrola: 200 krabů, 8 sloupců a žádné chybějící hodnoty
dim(x = data_krabi_zdroj)
sum(is.na(x = data_krabi_zdroj))

```

```
# Uložíme tabulku do složky data jako soubor CSV
dir.create(
  path = "data",
  showWarnings = FALSE
)
write.csv(
  x = data_krabi_zdroj,
  file = "data/krabi.csv",
  row.names = FALSE
)
```

Funkce `write.csv()` uloží textové sloupce v uvozovkách, takže se soubor může lišit vzhledem, ne hodnotami: po načtení pomocí `read.csv()` dostanete stejnou tabulku. Výuková kopie byla vytvořena z verze 7.3-66 balíčku MASS.

Původ a dosah dat

Campbell, N. A. a R. J. Mahon (1974): A multivariate study of variation in two species of rock crab of the genus *Leptograpsus*. *Australian Journal of Zoology* 22, 417–425. Tabulku šíří balíček MASS (`MASS::crabs`) pod licencí GPL-2 nebo GPL-3; soubor `krabi.csv` je její nezměněná kopie. Autoři vybrali do každé kombinace formy a pohlaví právě 50 krabů. Popis zdroje nedokládá náhodný výběr ze všech krabů; naše závěry proto vztahujeme k těmto měřením, ne automaticky k celé populaci.

5 Které délky a šířky jsou v datech zastoupené?

Před modelem zkontrolujeme typy proměnných, chybějící hodnoty a rozsahy. To nám také ukáže, při jakých délkách má smysl porovnávat předpovědi obou pohlaví.

```
# Ověříme typy sloupců a chybějící hodnoty
str(object = data_krabi)
```

```
'data.frame': 100 obs. of 3 variables:
 $ pohlavi      : Factor w/ 2 levels "samice","samec": 2 2 2 2 2 2 2 2 2 2 ...
 $ delka_krunyre_mm: num 16.1 18.1 19 20.1 20.3 23 23.8 24.5 24.2 25.2 ...
 $ zadni_sirka_mm : num 6.7 7.7 7.8 7.9 8 9 9.9 9.1 9.6 10.5 ...
```

```
colSums(x = is.na(data_krabi))
```

pohlavi	delka_krunyre_mm	zadni_sirka_mm
0	0	0

```
# Rozdělíme kraby podle pohlaví a prohlédneme každou skupinu zvlášť
data_samice <-
  data_krabi[data_krabi$pohlavi == "samice", ]
data_samci <-
  data_krabi[data_krabi$pohlavi == "samec", ]
summary(object = data_samice)
```

pohlavi	delka_krunyre_mm	zadni_sirka_mm
samice:50	Min. :14.70	Min. : 6.50
samec : 0	1st Qu.:23.93	1st Qu.:10.62
	Median :27.90	Median :12.20
	Mean :28.10	Mean :12.14
	3rd Qu.:32.75	3rd Qu.:13.88
	Max. :40.90	Max. :16.90

```
summary(object = data_samci)
```

pohlavi	delka_krunyre_mm	zadni_sirka_mm
samice: 0	Min. :16.10	Min. : 6.70
samec :50	1st Qu.:27.23	1st Qu.:10.65
	Median :32.45	Median :11.70
	Mean :32.01	Mean :11.72
	3rd Qu.:37.15	3rd Qu.:13.28
	Max. :47.10	Max. :15.80

V souboru je 100 krabů, z toho 50 samic a 50 samců. V těchto třech sloupcích chybí 0 hodnot. Délka krunýře samic sahá od 14,7 do 40,9 mm, u samců od 16,1 do 47,1 mm. Obě skupiny tedy mají společný rozsah pozorovaných délek. To nám později umožní porovnat modelové předpovědi pro stejně dlouhou samici a samce bez extrapolace.

6 Jak souvisí délka a zadní šířka krunýře?

Každý bod v následujícím **bodovém grafu (anglicky scatter plot)** představuje jednoho kraba. Vodorovná souřadnice je délka krunýře, svislá zadní šířka. Barva a tvar bodu rozlišují pohlaví, takže graf zůstane čitelný i bez barev.

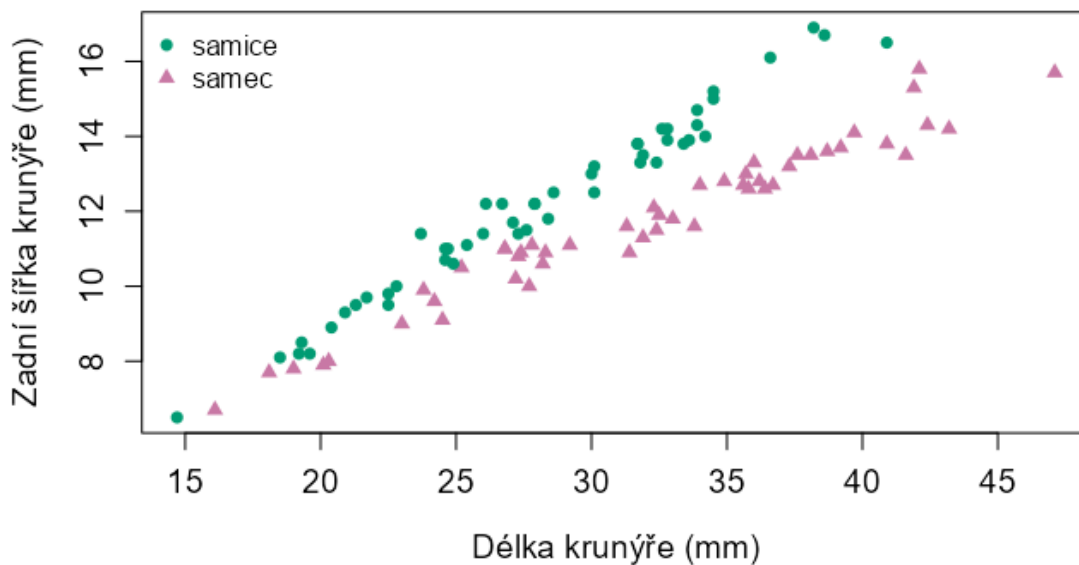
```
# Barvy a symboly přiřazujeme v celé lekci stejným skupinám
vec_barvy_pohlavi <-
```

```

c(samice = "#009E73", samec = "#CC79A7")
vec_symboly_pohlavi <-
  c(samice = 16, samec = 17)

# Jeden bod představuje jednoho změřeného kraba
plot(
  x = data_krabi$delka_krunyre_mm,
  y = data_krabi$zadni_sirka_mm,
  col = vec_barvy_pohlavi[as.character(data_krabi$pohlavi)],
  pch = vec_symboly_pohlavi[as.character(data_krabi$pohlavi)],
  xlab = "Délka krunýře (mm)",
  ylab = "Zadní šířka krunýře (mm)",
  cex.axis = 1.2,
  cex.lab = 1.2
)
legend(
  x = "topleft",
  legend = names(vec_barvy_pohlavi),
  col = vec_barvy_pohlavi,
  pch = vec_symboly_pohlavi,
  bty = "n"
)

```



Co vidíte v grafu?

Když se pohybujete zleva doprava, co se děje se zadní šířkou? Vypadají oba soubory bodů, jako by měly stejný sklon?

Možná odpověď

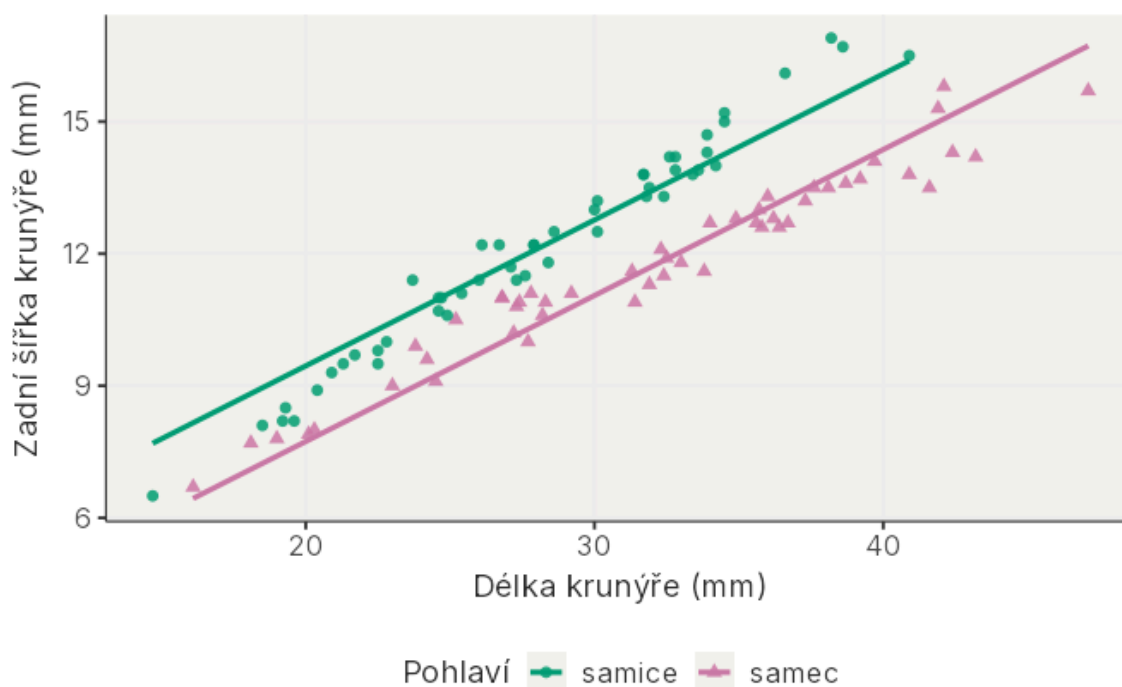
Body naznačují rostoucí vztah u obou pohlaví. Samice mají při podobné délce obvykle větší zadní šířku, ale samotný pohled na body zatím nestačí k přesnému vyčíslení, zda se tento rozdíl s délkou mění.

7 Rovnoběžné přímky předpokládají společný sklon

Začneme známým aditivním modelem. Oba prediktory v něm mohou měnit předpovězenou šířku, ale **sklon vůči délce je pro obě pohlaví společný**.

```
# Aditivní model připouští posun mezi skupinami, ale stejný sklon
mod_rovnobezne <-
  lm(
    formula = zadni_sirka_mm ~ delka_krunyre_mm + pohlavi,
    data = data_krabi
  )
coef(object = mod_rovnobezne)
```

(Intercept)	delka_krunyre_mm	pohlavisamec
2.819657	0.331590	-1.717180



Co předpovídá aditivní model?

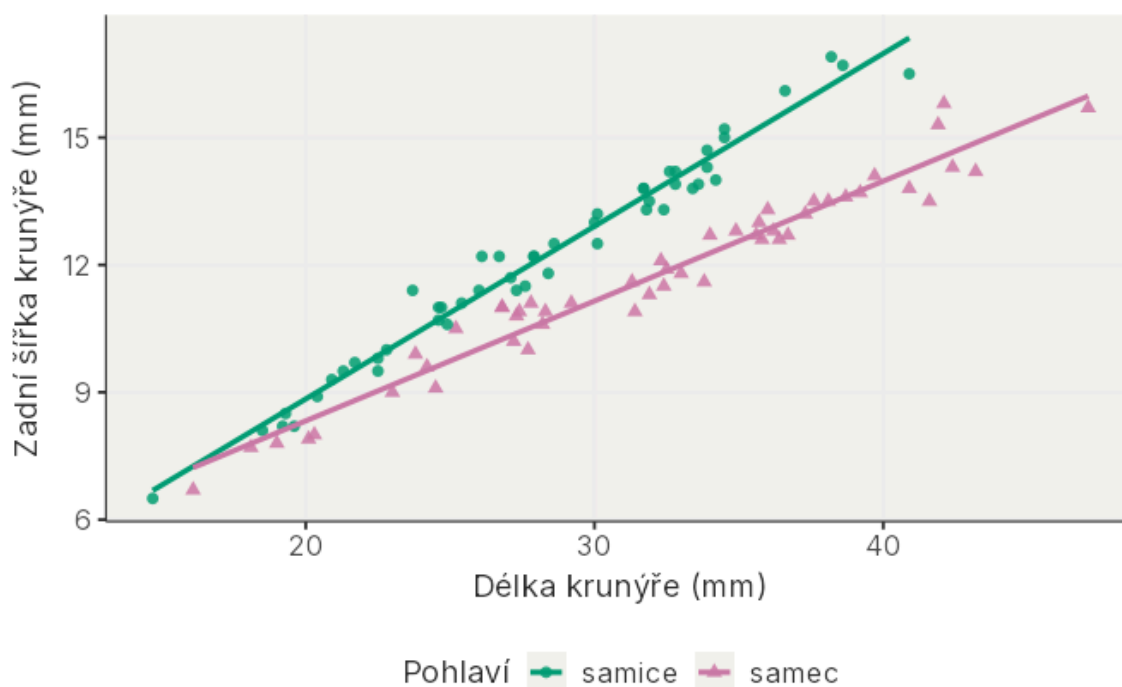
Co by tento model předpověděl o rozdílu v zadní šířce mezi stejně dlouhými kraby u kratších a delších délek z pozorovaného rozsahu?

Možná odpověď

Přímky zůstávají rovnoběžné, takže předpovězený rozdíl je v obou případech stejný. Je to vlastnost modelu, kterou jsme do něj vložili znakem +, ne výsledek samostatného testu této vlastnosti.

8 Mění se rozdíl mezi samicemi a samci s délkou krunýře?

V bodovém grafu se rozdíl mezi skupinami při různých délkách nemusí jevit jako stálý. Na stejné osy proto můžeme orientačně vést přímku každou skupinou zvlášť. Srovnajte jejich směr s rovnoběžnými přímkami předchozího obrázku; body krabů jsou v obou grafech stejné.



Předpovězte tvar

Pokud mají samice při určité společné délce větší předpovězenou zadní šířku a jejich přímka stoupá rychleji, bude rozdíl mezi přímkami u delších krabů menší, stejný, nebo větší?

Možná odpověď

Rozdíl by byl větší. Dvě různě strmé přímky totiž mění vzájemnou svislou vzdálenost s délkou.

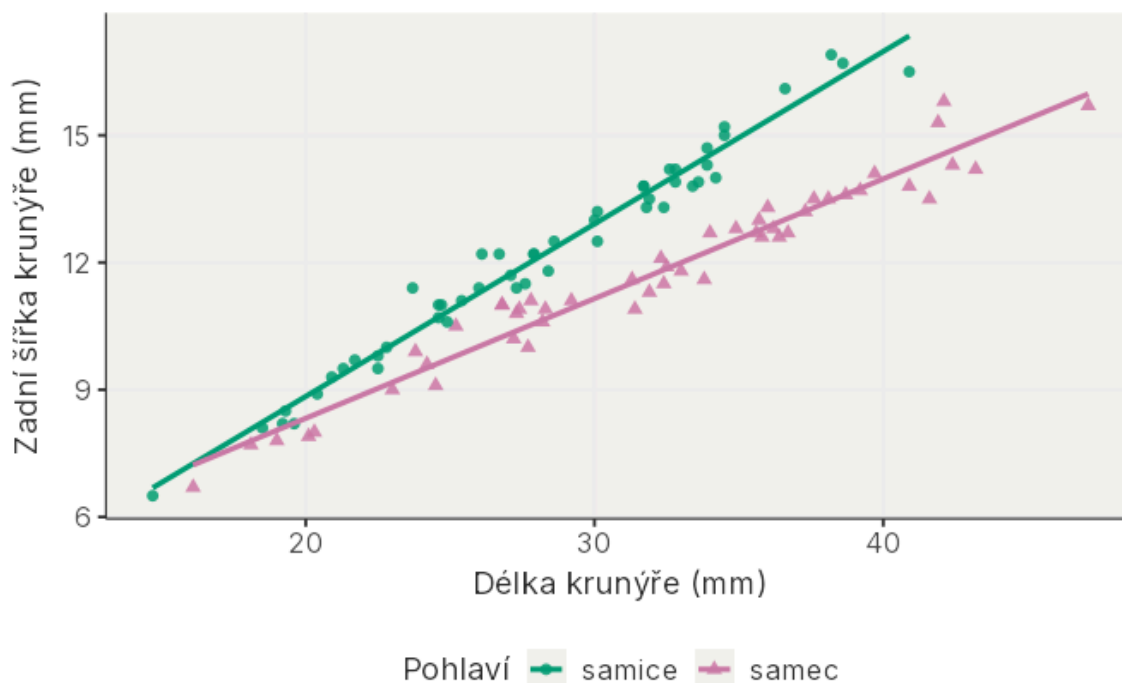
Když se svislá vzdálenost přímek mění s délkou, závisí vztah délky a šířky na pohlaví. Této závislosti říkáme interakce prediktorů. Stejnou myšlenku můžeme vyjádřit jedním modelem, který oba sklony odhadne společně.

9 Model s interakcí dovoluje samicím a samcům různé sklony

Předchozí obrázek ukázal, proč zkusit model se dvěma různými sklony. Délku krunýře v modelu měříme v milimetrech. Zápis `delka_krunyre_mm * pohlavi` dovolí, aby se sklon vůči délce mezi pohlavími lišil. Jednotlivé členy tohoto vzorce rozebereme po konkrétní předpovědi.

```
# Hvězdička připustí odlišné sklony pro samice a samce
mod_interakce <-
  lm(
    formula = zadni_sirka_mm ~ delka_krunyre_mm * pohlavi,
    data = data_krabi
  )
```

Předpovědní přímky kreslíme jen v rozsahu délek, které byly v dané skupině opravdu pozorovány. Za konci těchto úseků bychom už extrapolovali.



Přímka samic zde stoupá strměji než přímka samců. Čáry jsou odhadnuté hodnoty modelu, zatímco body jsou skutečná měření. Svislou vzdálenost přímek vyčíslíme pro jednu společnou délku krunýře.

10 Co model předpovídá pro stejně dlouhou samici a samce?

V grafu obou přímek můžeme porovnat jejich výšku nad stejným místem vodorovné osy. Zvolíme délku, která leží uvnitř pozorovaného rozsahu samic i samců a snadno ji najdeme na ose. Při srovnání pohlaví tak držíme délku krunýře stejnou.

```

# Zvolíme společnou délku uvnitř pozorovaného rozsahu obou pohlaví
delka_porovnani_mm <- 30

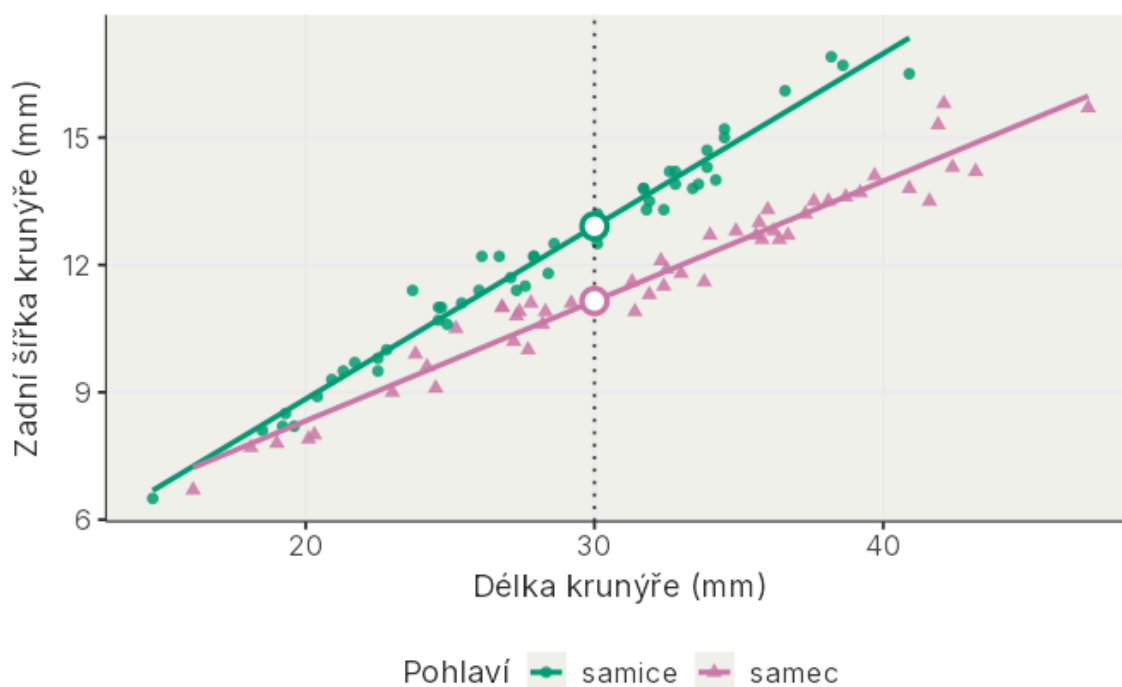
# Připravíme stejnou délku pro samici i samce
data_porovnani <-
  data.frame(
    delka_krunyre_mm = rep(x = delka_porovnani_mm, times = 2),
    pohlavi = c("samice", "samec")
  )

# Model odhadne šířku pro obě kombinace
data_porovnani$odhad_sirky_mm <-
  predict(
    object = mod_interakce,
    newdata = data_porovnani
  )
data_porovnani

```

	delka_krunyre_mm	poohlavi	odhad_sirky_mm
1	30	samice	12.91000
2	30	samec	11.14947

Zvolili jsme **30 mm**. Která přímka je nad touto délkou výše? Oba odhady jsou v grafu označeny prázdnými kruhy. Svislá tečkovaná čára ukazuje místo, kde čteme jejich **svislou vzdálenost**; menší body zůstávají naměřenými kraby.



Rozdíl při zvolené délce

Při délce 30 mm model odhaduje zadní šířku samice 12,91 mm a samce 11,15 mm. **Rozdíl samec minus samice je -1,76 mm.** Samčí předpověď tedy leží o 1,76 mm níže.

Přímky nemají stejný sklon, a proto se jejich svislá vzdálenost s délkou mění. Koeficienty modelu ukážou, o kolik se tento rozdíl změní při porovnání délek vzdálených o 1 mm.

11 Co znamenají koeficienty modelu s interakcí?

Z minulé lekce známe aditivní model se společným sklonem:

```
zadni_sirka_mm ~ delka_krunyre_mm + pohlavi
```

Aby se sklon mohl mezi pohlavími lišit, přidáme interakční člen označený dvojtečkou:

```
zadni_sirka_mm ~ delka_krunyre_mm + pohlavi + delka_krunyre_mm:poohlavi
```

R nabízí pro celý pravý výraz kratší zápis s hvězdičkou:

```
zadni_sirka_mm ~ delka_krunyre_mm * pohlavi
```

Obecně tedy platí $x * y = x + y + x:y$. Výpis `summary()` ukáže čtyři koeficienty: intercept a koeficient pro každý ze tří členů na pravé straně. Jde o tentýž model, který nakreslil obě přímky a vytvořil předpovědi při 30 mm:

```
summary(object = mod_interakce)
```

Call:

```
lm(formula = zadni_sirka_mm ~ delka_krunyre_mm * pohlavi, data = data_krabi)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.92402	-0.34114	-0.01188	0.23472	1.23484

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.70771	0.29380	2.409	0.0179	*
delka_krunyre_mm	0.40674	0.01023	39.742	< 2e-16	***
po hlav isamec	1.97311	0.40042	4.928	3.47e-06	***
delka_krunyre_mm:po hlav isamec	-0.12445	0.01317	-9.450	2.27e-15	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4241 on 96 degrees of freedom

Multiple R-squared: 0.9664, Adjusted R-squared: 0.9654

F-statistic: 921.2 on 3 and 96 DF, p-value: < 2.2e-16

V tabulce **Coefficients** jsou čtyři řádky. Sloupec **Estimate** obsahuje odhady v milimetrech nebo v milimetrech šířky na milimetr délky. Referenční skupinou jsou samice, protože jsme je při tvorbě faktoru **po hlavi** uvedli jako první.

- **(Intercept)**, tedy intercept, je odhad zadní šířky **samice s délkou krunýře 0 mm**: 0,708 mm. Takto malého kraba jsme neměřili; číslo samo o sobě biologicky nevykládáme.
- **delka_krunyre_mm** je sklon **samčí** přímky: při délce o 1 mm větší model odhaduje šířku větší o 0,407 mm. Nejde o růst jednoho kraba v čase, ale o porovnání různě dlouhých jedinců.
- **po hlav isamec** je rozdíl **samec minus samice při délce 0 mm**: +1,973 mm. Není to rozdíl mezi pohlavími při každé délce ani odhad pro některého ze změřených krabů.
- **delka_krunyre_mm:po hlav isamec** je interakční člen: rozdíl **sklon samce minus sklon samice**, -0,124 mm/mm. Záporná hodnota říká, že samčí přímka stoupá pomaleji.

Doplňující: proč s interakcí ponecháváme i oba hlavní členy

Hvězdička rozvine model na **delka_krunyre_mm + po hlavi + delka_krunyre_mm:po hlavi**. První dva členy určují referenční přímku a rozdíl mezi skupinami v referenční hodnotě délky; poslední člen popisuje rozdíl sklonů. Tomuto uspořádání se říká hierarchický princip modelu.

Kdybychom hlavní člen délky nebo pohlaví bez věcného důvodu odstranili, interakční koeficient by zůstal bez přirozeného základu, vůči kterému se interpretuje. Výjimky existují pro úzce formulované modely, ale v úvodní analýze ponecháváme oba hlavní členy a biologickou otázku čteme z jejich společných předpovědí.

Hodnota 0 mm určuje pouze to, co v tomto zápisu znamenají dva koeficienty. Pro srovnání skutečně měřených délek použijeme oba koeficienty sklonu a dosadíme délku z pozorovaného rozsahu. **Rozdíl -1,76 mm jsme už odečetli z předpovědi při 30 mm.** Následující výpočet nepřináší nový výsledek; ukazuje, jak se stejná hodnota sestaví ze čtyř řádků tabulky **Coefficients**. Odhad šířky samice a rozdíl **samec minus samice** mají tuto stavbu:

$$\begin{aligned} \text{odhad šířky samice při zvolené délce} &= \text{odhad šířky samice při 0 mm} \\ &\quad + \text{zvolená délka} \times \text{samičí sklon} \\ \text{rozdíl samec minus samice při zvolené délce} &= \text{rozdíl při 0 mm} \\ &\quad + \text{zvolená délka} \times \text{rozdíl sklonů} \end{aligned}$$

Při 30 mm dosadíme odhady z tabulky **Coefficients**:

$$\begin{aligned} \text{odhad šířky samice} &= 0,7077 + 30 \cdot 0,4067 \approx 12,91 \text{ mm} \\ \text{rozdíl samec minus samice} &= 1,9731 + 30 \cdot (-0,1245) \approx -1,76 \text{ mm} \end{aligned}$$

Ve výpočtu používáme nezaokrouhlené koeficienty modelu; zobrazená čtyři desetinná místa jsou jen pro čitelnost. Po vynásobení se výpočet zkrátí na dva součty:

$$\begin{aligned} \text{odhad šířky samice} &\approx 0,7077 + 12,2023 \approx 12,91 \text{ mm} \\ \text{rozdíl samec minus samice} &\approx 1,9731 + (-3,7336) \approx -1,76 \text{ mm} \end{aligned}$$

Stejný postup lze zapsat stručně pro libovolnou délku:

$$\begin{aligned} \hat{y}_{\text{samice}}(x) &= \beta_0 + \beta_1 x \\ \hat{y}_{\text{samec}}(x) - \hat{y}_{\text{samice}}(x) &= \beta_2 + \beta_3 x \end{aligned}$$

Zde x znamená zvolenou délku krunýře v mm a $\hat{y}$ modelový odhad zadní šířky v mm. β_0 je řádek (**Intercept**) pro samici při 0 mm, β_1 její sklon, β_2 řádek **pohlavisamec** pro rozdíl při 0 mm a β_3 interakční řádek pro rozdíl sklonů.

Odhad šířky samce získáme přičtením druhého výsledku k prvnímu: $12,91 + (-1,76) = 11,15$ mm. To souhlasí s řádkem samce v tabulce předpovědí. Pro biologické porovnání čteme **předpovědi při konkrétní délce** a jejich rozdíl.

Samčí sklon získáme sečtením samičího sklonu a interakčního koeficientu. Vyjde 0,282 mm/mm oproti 0,407 mm/mm u samic. Rozdíl **samec minus samice** se při porovnání o 1 mm delších krabů změnil o -0,124 mm. Pro tento rozdíl sklonů získáme z téhož modelu 95% interval spolehlivosti:

```
# Interval patří rozdílu sklonů samce a samice
confint(
  object = mod_interakce,
  parm = "delka_krunyre_mm:pohlavisamec"
)
```

```
                2.5 %      97.5 %
delka_krunyre_mm:pohlavisamec -0.1505977 -0.0983116
```

Odhadovaný rozdíl sklonů je $-0,124$ mm/mm; interval sahá od $-0,151$ do $-0,098$ mm/mm. Popisuje nejistotu odhadu při předpokladech modelu, nikoli rozsah šířek jednotlivých krabů.

Koeficient pohlaví neplatí při každé délce

Samotný koeficient `pohlavisamec` popisuje rozdíl při **0 mm**, tedy mimo pozorované délky. Při biologickém porovnání vždy uveďte **konkrétní délku** a porovnejte dvě modelové předpovědi při této délce.

Doplňující: co dalšího ukazuje summary()?

Sloupec **Std. Error** popisuje nejistotu jednotlivých koeficientů. **t value** a **Pr(>|t|)** slouží k testování, zda by příslušný *podmíněný* koeficient mohl být nulový při předpokladech modelu. Hvězdičky u p-hodnot nejsou velikostí biologického rozdílu: například test u (**Intercept**) se ptá na nulovou šířku samice při délce 0 mm, což zde není zajímavá biologická otázka.

Část **Residuals** stručně shrnuje odchylky jednotlivých krabů od přímek. **Residual standard error** je reziduální směrodatná odchylka; popisuje typickou velikost těchto odchylek v milimetrech. Údaje **Multiple R-squared**, **Adjusted R-squared** a **F-statistic** zatím nepotřebujeme k výkladu interakce; k porovnávání biologicky zdůvodněných modelů se vrátíme v následující lekci.

Doplňující: modelové průměry při zvolené délce pomocí emmeans

Funkci `emmeans()` znáte z předchozích lekcí. Pracuje se stejným odhadnutým modelem jako naše tabulka předpovědí; nastavíme v ní délku krunýře na 30 mm.

```
# Odhadneme průměrnou zadní šířku obou pohlaví při zvolené délce
prumery_sirka <-
  emmeans::emmeans(
    object = mod_interakce,
    specs = ~ pohlavi | delka_krunyre_mm,
```

```

    at = list(delka_krunyre_mm = delka_porovnani_mm)
  )
prumery_sirka

```

```

delka_krunyre_mm = 30:
  pohlavi emmean      SE df lower.CL upper.CL
samice    12.9 0.0630 96    12.8    13.0
samec     11.1 0.0623 96    11.0    11.3

```

Confidence level used: 0.95

Sloupec `emmean` obsahuje stejné dvě bodové předpovědi jako naše tabulka z `predict()`. Sloupce `lower.CL` a `upper.CL` vymezují modelový interval pro **průměrnou šířku skupiny**, ne rozsah šířek jednotlivých krabů. Chceme-li interval přímo pro rozdíl **samec minus samice**, vypočteme jej zvlášť:

```

# Odečteme samičí odhad od samčího odhadu
rozdily_sirka <-
  emmeans::contrast(
    object = prumery_sirka,
    method = "revpairwise"
  )
confint(object = rozdily_sirka)

```

```

delka_krunyre_mm = 30:
  contrast      estimate      SE df lower.CL upper.CL
samec - samice  -1.76 0.0886 96    -1.94    -1.58

```

Confidence level used: 0.95

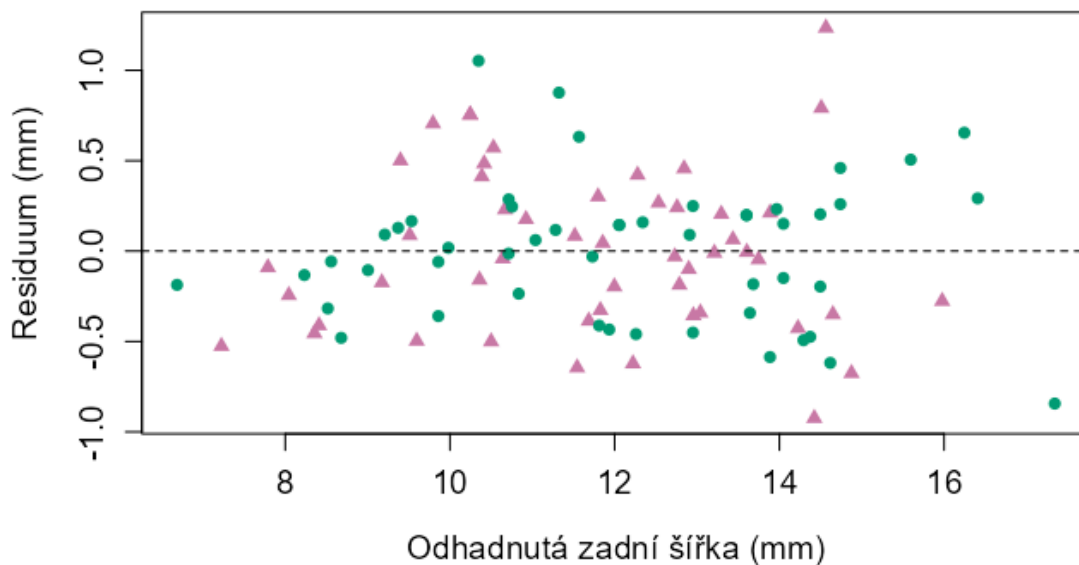
Odhad rozdílu odpovídá dříve vypočteným $-1,76$ mm. **Interval pro rozdíl vzniká z obou předpovědí a jejich společné nejistoty**; nelze jej spolehlivě vyčíst jen z překryvu dvou intervalů pro jednotlivé průměry. Způsob výběru krabů stále omezuje, na koho lze výsledek přenést.

12 Jak dobře přímky popisují změřené kraby?

Přímky procházejí oběma mračny bodů, ale jednotlivé šířky se od nich odchylují. Pro každého kraba je residuum rozdíl mezi naměřenou a odhadnutou šířkou. Graf residuí proti odhadům pomůže odhalit zakřivení nebo změnu rozptylu, které by dvě přímky nezachytily.

```
# Spočteme rozdíl mezi měřením a předpovědí pro každého kraba
data_krabi$odhad_sirky_mm <-
  fitted(object = mod_interakce)
data_krabi$residuum_mm <-
  data_krabi$zadni_sirka_mm - data_krabi$odhad_sirky_mm

# Nula znamená dokonalou shodu měření a předpovědi
plot(
  x = data_krabi$odhad_sirky_mm,
  y = data_krabi$residuum_mm,
  col = vec_barvy_pohlavi[as.character(data_krabi$pohlavi)],
  pch = vec_symboly_pohlavi[as.character(data_krabi$pohlavi)],
  xlab = "Odhadnutá zadní šířka (mm)",
  ylab = "Residuum (mm)",
  cex.axis = 1.2,
  cex.lab = 1.2
)
abline(h = 0, lty = 2)
```



Hledáme soustavné zakřivení nebo nápadně větší rozptyl na jedné straně: tím bychom poznali, že dvě přímky některou část vztahu nevystihují. V tomto grafu nic takového nevidíme. Residua leží nad nulou i pod ní v celém rozsahu odhadů, netvoří oblouk a jejich rozptyl je u malých i velkých odhadů podobný. Vraťte se také k bodovému grafu a prohlédněte kraby na koncích rozsahu

délek. V těchto místech je méně sousedních měření, a jednotlivý krajní bod může sklon přímky více ovlivnit. Takový bod je důvod k ověření měření a citlivosti závěru, ne k automatickému odstranění.

Diagnostika modelu sama nezaručí, že je vztah v populaci přesně lineární, že výběr krabů byl reprezentativní, ani že pohlaví něco způsobilo. Předpovědi mimo pozorované délky jsme nekontrolovali.

Doplňující: interakci můžeme číst jen tam, kde máme data

Rozdíl sklonů je odhadnut z oblastí, kde máme měření obou pohlaví. Kdyby samice byly pozorovány jen mezi kratšími délkami a samci jen mezi delšími délkami bez společného rozsahu, model by sice dokázal nakreslit dvě přímky, ale jejich srovnání při stejné délce by bylo založeno hlavně na extrapolaci.

Před interpretací interakce proto kontrolujeme nejen počet bodů, ale také **společný rozsah hodnot** a zastoupení všech důležitých kombinací. Řídká nebo prázdná část grafu znamená, že modelová předpověď je méně podepřená skutečnými srovnáními. Více parametrů nemůže vytvořit informaci v kombinaci, kterou jsme nepozorovali.

Co ukázal první příklad

U těchto změřených krabů modré formy roste odhadnutá zadní šířka s délkou u obou pohlaví. Sklon je u samic přibližně 0,407 mm šířky na 1 mm délky a u samců 0,282 mm/mm. Při 30 mm je modelový rozdíl **samec minus samice** -1,76 mm. Samčí sklon je o 0,124 mm/mm menší, takže u krabů s větší délkou model předpovídá ještě větší náskok samic. To odpovídá biologickému očekávání z úvodu, že samice s rostoucí velikostí mají relativně širší zadní část krunýře. Jde však o souvislost v těchto měřeních, nikoli o důkaz, že ji způsobuje pohlaví nebo nošení snůšky.

13 Liší se rozdíl v délce krunýře mezi pohlavími u obou barevných forem?

Vztah zadní šířky k délce jsme sledovali pouze u modré formy. Soubor obsahuje také oranžovou barevnou formu. U mnoha živočichů se samice a samci liší velikostí a tento rozdíl nemusí být stejný u příbuzných druhů nebo forem. Můžeme se proto zeptat na jiný rozměr týchž krabů: **je rozdíl v délce krunýře mezi samci a samicemi stejný u obou forem?** Odezvou je tentokrát délka krunýře v milimetrech. Barevná forma i pohlaví jsou kategoriální proměnné neboli skupinové znaky; pro tento příklad nepotřebujeme číselný prediktor.

Co tato data o velikosti krabů neřeknou

Autoři vybrali do každé ze čtyř skupin právě 50 krabů. Průměrné délky proto závisejí také na tom, které kraby sesbírali a změřili, nejen na tom, jak velcí krabi dané formy a pohlaví v přírodě obvykle jsou. Příklad nám ukáže, jak čist model se dvěma skupinovými znaky; tvrzení o velikosti krabů v přírodě by vyžadovalo náhodný výběr.

```
# Ze stejné tabulky vybereme novou odezvu i oba skupinové znaky
data_formy <-
  data_krabi_vse[, c("sp", "sex", "CL")]
names(data_formy) <-
  c("forma", "pohlavi", "delka_krunyre_mm")

# Obě skupinové proměnné převedeme na faktory s referenční úrovní
data_formy$forma <-
  factor(
    x = data_formy$forma,
    levels = c("B", "0"),
    labels = c("modrá", "oranžová")
  )
data_formy$pohlavi <-
  factor(
    x = data_formy$pohlavi,
    levels = c("F", "M"),
    labels = c("samice", "samec")
  )

# Počty ve čtyřech skupinách a chybějící délky
table(data_formy$forma, data_formy$pohlavi)
```

	samice	samec
modrá	50	50
oranžová	50	50

```
colSums(x = is.na(data_formy))
```

forma	pohlavi	delka_krunyre_mm
0	0	0

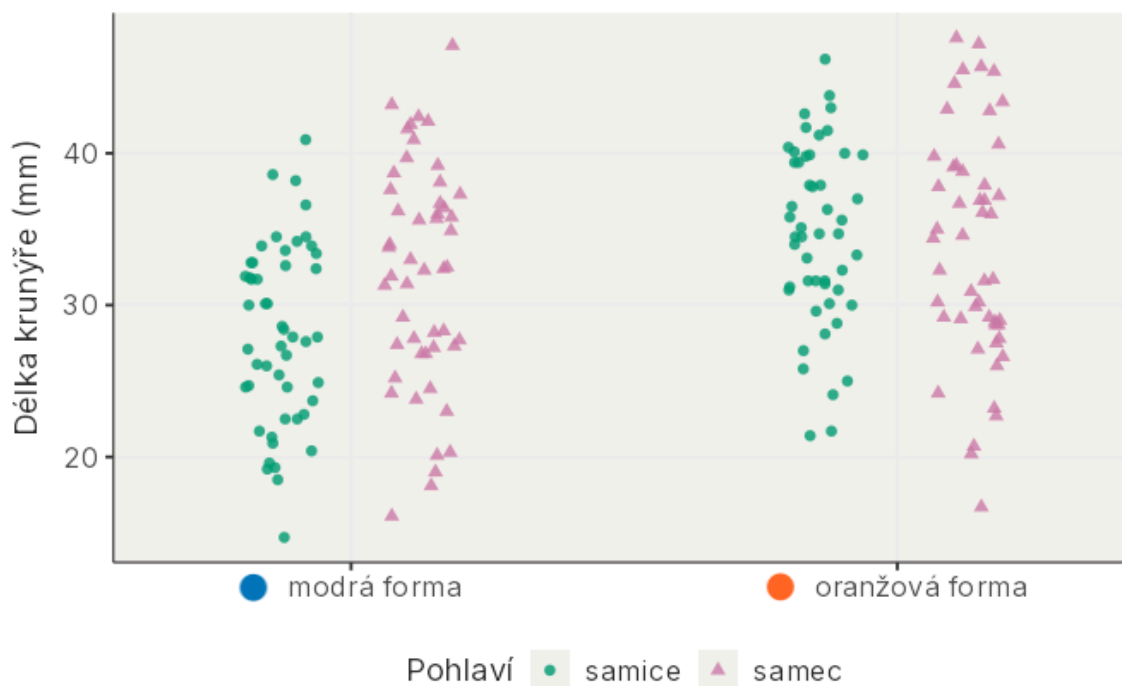
```
summary(object = data_formy$delka_krunyre_mm)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
14.70	27.27	32.10	32.11	37.23	47.60

V souboru je 200 krabů. V každé ze čtyř kombinací je 50 měření; v délce krunýře chybí 0 hodnot. Pozorované délky v celém souboru sahají od 14,7 do 47,6 mm. Jeden řádek stále představuje jednoho kraba. Názvy forem označují dvě skupiny v datovém souboru, nikoli zásah, který bychom krabům přidělili.

14 Jaké délky krunýře vidíme ve čtyřech skupinách?

Podívejte se na jednotlivé naměřené délky; každý bod představuje jednoho kraba. Body samic a samců jsou uvnitř každé barevné formy mírně posunuté do stran, aby se méně překrývaly; barvy a tvary bodů pro pohlaví zůstávají stejné jako v předchozích grafech. Modrý a oranžový kroužek u popisku připomíná barevnou formu.



Barevná forma	Samice: průměr (mm)	Samec: průměr (mm)
modrá	28,10	32,01
oranžová	34,62	33,69

Porovnejte obě formy

Ve které formě mají samci větší průměrnou délku než samice? Platí to i ve druhé formě?

Možná odpověď

U modré formy činí rozdíl **samec minus samice** +3,91 mm. U oranžové formy je to -0,93 mm, takže rozdíl má opačný směr.

Průměry popisují skupiny, zatímco jednotlivé body v obrázku připomínají velkou proměnlivost délek mezi kraby uvnitř každé skupiny.

15 Jak model zachytí změnu rozdílu mezi pohlavími podle formy?

Vrátíme se ke **stejnému obrázku se všemi jednotlivými kraby**. Body ani osy neměníme. Přidáme pouze velké kosočtverce pro čtyři průměry a v každé barevné formě spojíme průměr samic s průměrem samců. Sklon spojnice zde nevyjadřuje spojitou předpověď; pouze zvýrazňuje velikost a směr rozdílu mezi dvěma kategoriemi.



Na modré formě leží průměr samců nad průměrem samic; na oranžové pod ním. Zajímá nás, **o kolik se rozdíl samec minus samice změní mezi formami**. Tento rozdíl dvou rozdílů je druhým typem interakce:

změna rozdílu mezi pohlavími = průměr délky oranžových samců
 – průměr délky oranžových samic
 – průměr délky modrých samců
 + průměr délky modrých samic

Vezmeme čtyři průměry z tabulky a odečteme vždy dvě hodnoty ze stejného řádku:

$$\begin{aligned} \text{změna rozdílu mezi pohlavími} &= (33,688 - 34,618) - (32,014 - 28,102) \\ &= -0,930 - 3,912 \\ &= -4,842 \text{ mm} \end{aligned}$$

Stručně lze stejné čtyři průměry označit $\bar{y}_{O,M}$, $\bar{y}_{O,F}$, $\bar{y}_{B,M}$ a $\bar{y}_{B,F}$; O a B označují oranžovou a modrou formu, M a F samce a samici a pruh nad y průměr délky. Pak platí $\Delta = (\bar{y}_{O,M} - \bar{y}_{O,F}) - (\bar{y}_{B,M} - \bar{y}_{B,F})$, kde Δ označuje změnu rozdílu mezi pohlavími při přechodu od modré k oranžové formě.

Jediným modelem můžeme tyto čtyři skupinové průměry odhadnout společně. Hvězdička zahrne obě skupiny i jejich interakci:

```
# Referenční kombinací je modrá forma a samice
mod_formy <-
  lm(
    formula = delka_krunyre_mm ~ forma * pohlavi,
    data = data_formy
  )
summary(object = mod_formy)
```

```
Call:
lm(formula = delka_krunyre_mm ~ forma * pohlavi, data = data_formy)

Residuals:
    Min       1Q   Median       3Q      Max
-16.988  -4.636   0.184   5.130  15.086

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      28.1020     0.9499  29.584 < 2e-16 ***
formaoranžová      6.5160     1.3434   4.851 2.5e-06 ***
pohlavisamec       3.9120     1.3434   2.912 0.00401 **
formaoranžová:pohlavisamec -4.8420     1.8998  -2.549 0.01158 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6.717 on 196 degrees of freedom
Multiple R-squared: 0.1232, Adjusted R-squared: 0.1098
F-statistic: 9.181 on 3 and 196 DF, p-value: 1.031e-05
```

V tabulce **Coefficients** se opět soustředíte na sloupec **Estimate**. Řádky nejsou čtyři nezávislé průměry: jeden je referenční průměr a zbylé tři popisují, jak se od něj dostaneme k ostatním skupinám.

16 Co znamenají čtyři koeficienty modelu se dvěma skupinovými znaky?

Protože referenčními kategoriemi jsou **modrá forma** a **samice**, každý koeficient má přesně určené srovnání. Odhady vytáhneme ze stejného modelu, který vytvořil výpis `summary()`.

- **(Intercept)** je odhad průměrné délky **modré samice**: 28,102 mm. Najdete jej v tabulce průměrů v řádku modré formy a sloupci samic.
- **formaoranžová** je rozdíl **oranžová minus modrá mezi samicemi**: +6,516 mm. Po přičtení k interceptu dostaneme průměr oranžových samic; tento koeficient nepopisuje rozdíl forem u obou pohlaví stejně.
- **pohlavisamec** je rozdíl **samec minus samice v modré formě**: +3,912 mm. Souhlasí se svislou vzdáleností obou průměrů u modré formy.
- **formaoranžová:pohlavisamec** je změna rozdílu samec minus samice při přechodu **od modré k oranžové formě**: -4,842 mm. Jde přesně o výše vypočtený rozdíl dvou rozdílů, nikoli o další samostatnou skupinu.

```
# Interval patří k rozdílu dvou rozdílů, ne k jedné skupině
confint(
  object = mod_formy,
  parm = "formaoranžová:pohlavisamec"
)
```

	2.5 %	97.5 %
formaoranžová:pohlavisamec	-8.588674	-1.095326

Modelový 95% interval spolehlivosti pro změnu rozdílu sahá od -8,59 do -1,10 mm. Celý leží pod nulou: při předpokladech modelu data podporují menší rozdíl samec minus samice u oranžové než u modré formy. Protože autoři vybrali do každé skupiny pevný počet krabů, nelze tento interval automaticky zobecnit na všechny kraby.

Jak model sestaví předpověď pro oranžového samce? K odhadu modré samice přidá rozdíl forem mezi samicemi, rozdíl pohlaví u modré formy a změnu tohoto rozdílu u oranžové formy:

odhad délky oranžového samce = odhad délky modré samice
 + rozdíl forem mezi samicemi
 + rozdíl pohlaví u modré formy
 + změna rozdílu pohlaví u oranžové formy

Dosazené hodnoty jsou právě čtyři řádky výpisu modelu:

$$\begin{aligned}\text{odhad délky oranžového samce} &= 28,102 + 6,516 + 3,912 + (-4,842) \\ &= 38,530 + (-4,842) \\ &= 33,688 \text{ mm}\end{aligned}$$

V obecném zkráceném zápisu je to $\hat{y}_{O,M} = \beta_0 + \beta_O + \beta_M + \beta_{OM}$. Symbol $\hat{y}_{O,M}$ značí modelový odhad délky oranžového samce, β_0 referenční průměr modré samice, β_O rozdíl forem u samic, β_M rozdíl pohlaví u modré formy a β_{OM} změnu tohoto rozdílu u oranžové formy.

```
# Vytvoříme všechny čtyři kombinace a pro každou spočteme odhad
data_kombinace <-
  expand.grid(
    forma = levels(data_formy$forma),
    pohlavi = levels(data_formy$poohlavi)
  )
data_kombinace$delka_odhad_mm <-
  predict(
    object = mod_formy,
    newdata = data_kombinace
  )
data_kombinace
```

```
      forma pohlavi delka_odhad_mm
1   modrá   samice      28.102
2 oranžová samice      34.618
3   modrá   samec      32.014
4 oranžová samec      33.688
```

Barevná forma	Samice: odhad (mm)	Samec: odhad (mm)
modrá	28,10	32,01
oranžová	34,62	33,69

V řádku oranžové formy je rozdíl samec minus samice $-0,93$ mm. Získáme jej také sečtením koeficientů `pohlavisamec` a `formaoranžová:pohlavisamec`: $3,912 + (-4,842) = -0,930$ mm. **Samotný koeficient pohlaví tedy platí pro modrou formu; u oranžové se k němu přidá interakce.**

Doplňující: čtyři skupinové předpovědi pomocí emmeans

Stejně jako v předchozím příkladu můžeme z již odhadnutého modelu získat předpovědi pomocí `emmeans()`. Tentokrát žádnou číselnou hodnotu nenastavujeme: žádáme o **obě pohlaví uvnitř každé barevné formy**, tedy o čtyři skutečně zastoupené kombinace.

```
# Ze stejného modelu získáme odhad pro každou kombinaci formy a pohlaví
prumery_formy <-
  emmeans::emmeans(
    object = mod_formy,
    specs = ~ pohlavi | forma
  )
prumery_formy
```

```
forma = modrá:
  pohlavi 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:
  pohlavi 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

Sloupec `emmean` ukazuje stejné délky jako tabulka předpovědí z `predict()`. Protože model obsahuje obě skupiny i jejich interakci, tyto čtyři odhady jsou zde také rovny pozorovaným průměrům čtyř skupin. Intervaly `lower.CL` a `upper.CL` se vztahují ke **skupinovým průměrům**. Rozdíl **samec minus samice** pro každou formu vypočteme přímo:

```
# Opačné pořadí dvojice dává rozdíl samec minus samice
rozdily_formy <-
  emmeans::contrast(
    object = prumery_formy,
    method = "revpairwise"
  )
confint(object = rozdily_formy)
```

```
forma = modrá:
  contrast      estimate   SE  df lower.CL upper.CL
```

```
samec - samice      3.91 1.34 196      1.26      6.56
```

forma = oranžová:

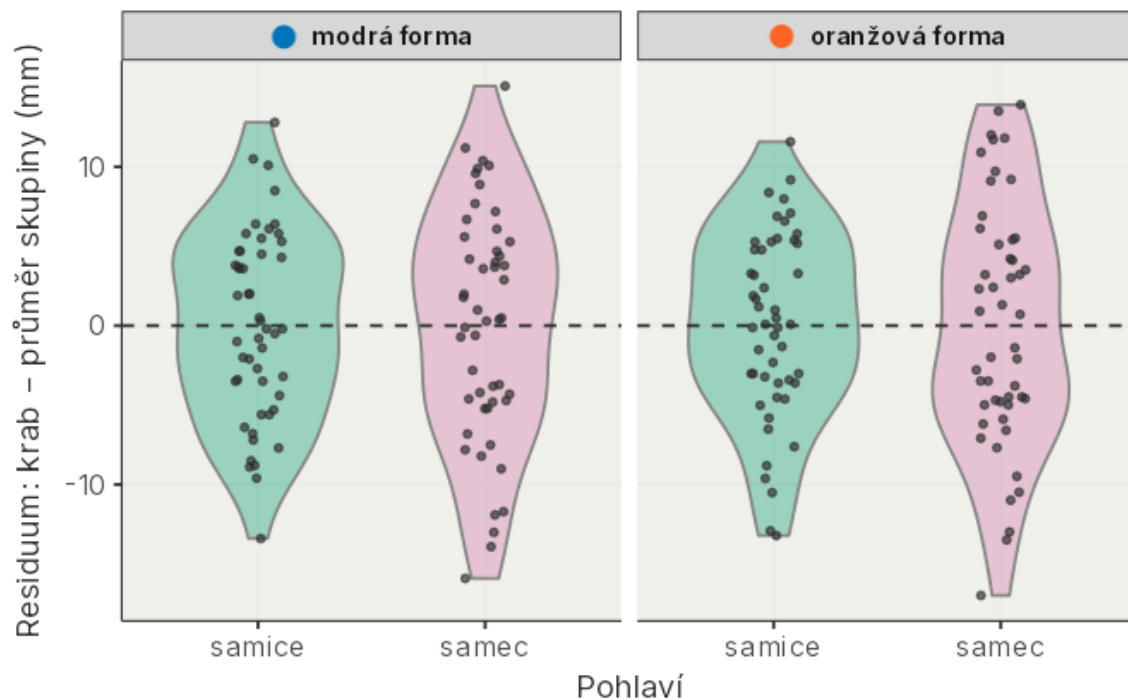
```
contrast      estimate    SE  df lower.CL upper.CL
samec - samice    -0.93 1.34 196    -3.58     1.72
```

Confidence level used: 0.95

Řádek modré formy odpovídá rozdílu +3,91 mm a řádek oranžové -0,93 mm. Interval u každého řádku se týká **rozdílu pohlaví v dané formě**; interval pro změnu tohoto rozdílu mezi formami jsme získali výše z interakčního koeficientu. Tyto tři otázky mají odlišné intervaly, i když vycházejí z téhož modelu.

17 Co o jednotlivých krabech čtyři průměry neříkají?

Model čtyř skupin přesně vystihuje jejich průměry. To však neznamená, že stejně dobře vystihuje každého kraba: jednotlivé délky se od skupinových průměrů výrazně liší. Zkontrolujeme proto residua uvnitř skupin; jejich nulový průměr je zde vlastností odhadu skupinových průměrů, nikoli důkazem bezchybného modelu. V houslovém grafu (anglicky *violin plot*) širší část naznačuje, kde se odchylky častěji vyskytují, a každý bod ukazuje skutečné měření jednoho kraba.



Ve všech skupinách jsou délky na obou stranách odhadu. Rozpětí a krajní hodnoty se mezi skupinami mohou lišit; graf proto nečtěte jako důkaz, že čtyři průměry popisují jednotlivce přesně. Bez popisu způsobu výběru krabů také nelze rozdíly automaticky zobecnit na všechny kraby ani z nich vyvozovat účinek pohlaví nebo barevné formy.

18 Přibývá objem dřeva s průměrem kmene stejně rychle u nízkých i vysokých stromů?

V obou příkladech s kraby byl aspoň jeden prediktor skupinou. Interakce však může spojovat i dva číselné prediktory. Lesníci odhadují, kolik dřeva je v kmeni, z jeho průměru a z výšky stromu. Podíváme se na měření pokácených stromů střemchy pozdní (*Prunus serotina*), severoamerického stromu, který se v českých lesích šíří jako invazní druh.

Co čekáte?

Představte si dvojici nízkých stromů stejné výšky, z nichž jeden má kmen o 1 cm tlustší, a stejnou dvojici vysokých stromů. O kolik více dřeva má tlustší strom v každé dvojici? Bude tento rozdíl u nízkých i vysokých stromů stejný?

Možná odpověď

Pravděpodobně ne. Centimetr průměru navíc znamená silnější vrstvu dřeva po celé délce kmene, takže delší kmen jí obsahuje více. Toto očekávání vychází z tvaru kmene, ne z fyziologie stromu. Pokud toto očekávání platí, sklon vztahu objemu a průměru bude záviset na výšce.

Soubor [stromy.csv](#) je součástí materiálů lekce; uložte jej do stejné složky `data` jako soubor s kraby. Hodnoty jsou v původních amerických jednotkách: průměr kmene v palcích, výška ve stopách a objem dřeva v krychlových stopách. Proto je hned převedeme:

```
# Načteme tabulku pokácených stromů
data_stromy <-
  read.csv(file = "data/stromy.csv")

# Převedeme jednotky na cm, m a m³
data_stromy$prumer_cm <-
  data_stromy$Girth * 2.54 # 1 palec = 2,54 cm
data_stromy$vyška_m <-
  data_stromy$Height * 0.3048 # 1 stopa = 0,3048 m
data_stromy$objem_m3 <-
  data_stromy$Volume * 0.0283168 # 1 krychlová stopa = 0,0283168 m³
```

```
# Zkontrolujeme typy sloupců, chybějící hodnoty a rozsahy
str(object = data_stromy)
```

```
'data.frame': 31 obs. of 6 variables:
 $ Girth      : num  8.3 8.6 8.8 10.5 10.7 10.8 11 11 11.1 11.2 ...
 $ Height     : int   70 65 63 72 81 83 66 75 80 75 ...
 $ Volume     : num   10.3 10.3 10.2 16.4 18.8 19.7 15.6 18.2 22.6 19.9 ...
 $ prumer_cm  : num   21.1 21.8 22.4 26.7 27.2 ...
 $ vyska_m    : num   21.3 19.8 19.2 21.9 24.7 ...
 $ objem_m3   : num    0.292 0.292 0.289 0.464 0.532 ...
```

```
colSums(x = is.na(data_stromy))
```

Girth	Height	Volume	prumer_cm	vyska_m	objem_m3
0	0	0	0	0	0

```
summary(object = data_stromy[, c("prumer_cm", "vyska_m", "objem_m3")])
```

prumer_cm	vyska_m	objem_m3
Min. :21.08	Min. :19.20	Min. :0.2888
1st Qu.:28.07	1st Qu.:21.95	1st Qu.:0.5493
Median :32.77	Median :23.16	Median :0.6853
Mean :33.65	Mean :23.16	Mean :0.8543
3rd Qu.:38.73	3rd Qu.:24.38	3rd Qu.:1.0562
Max. :52.32	Max. :26.52	Max. :2.1804

Jeden řádek představuje **jeden pokácený strom**. Sloupec *Girth* (anglicky obvod) je podle dokumentace ve skutečnosti průměr kmene měřený asi 1,37 m nad zemí. V souboru je 31 stromů a chybí 0 hodnot. Průměr kmene sahá od 21,1 do 52,3 cm, výška od 19,2 do 26,5 m a objem dřeva od 0,29 do 2,18 m³. Vyšší stromy mívají i tlustší kmen (korelace průměru a výšky je 0,52); oba prediktory tedy nenesou úplně nezávislou informaci, podobně jako dvě teploty v minulé lekci.

Stromů je málo a dokumentace neuvádí, kde a jak byly vybrány. Závěry proto vztahujeme k těmto 31 stromům a k pozorovaným rozsahům průměru a výšky.

Doplňující: odkud pocházejí data o stromech

Data pocházejí z knihy Meyer, H. A. (1953): *Forest Mensuration*, Penns Valley Publishers, převzaté do Ryan, T. A., Joiner, B. L. a Ryan, B. F. (1976): *The Minitab Student Handbook*,

Duxbury Press. Tabulku šíří základní balíček R `datasets` pod licencí GPL-2 nebo GPL-3; soubor `stromy.csv` je její nezměněná kopie. Stejnou tabulku si můžete vytvořit sami:

```
# Tabulka je součástí základní instalace R
data_stromy_zdroj <-
  as.data.frame(x = datasets::trees)

# Kontrola: 31 stromů, 3 sloupce a žádné chybějící hodnoty
dim(x = data_stromy_zdroj)
sum(is.na(x = data_stromy_zdroj))

# Uložíme tabulku do složky data jako soubor CSV
write.csv(
  x = data_stromy_zdroj,
  file = "data/stromy.csv",
  row.names = FALSE
)
```

19 Vystihne společný sklon nízké i vysoké stromy?

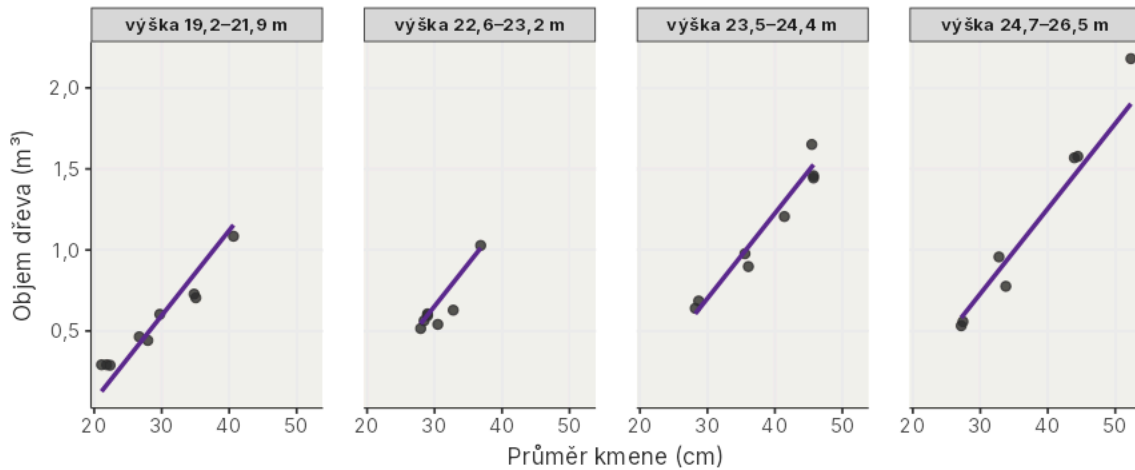
U krabů jsme kreslili jednu přímku pro každé pohlaví. Výška je však číslo a žádné skupiny nemá. Abychom viděli, jak vypadá vztah objemu a průměru u různě vysokých stromů, rozdělíme stromy **jen pro kreslení** do čtyř výškových pásem se zhruba stejným počtem stromů. Každý panel ukazuje stromy jednoho pásma. Do modelu pásma nevstupují: model dál pracuje s výškou jako s číslem.

Začneme známým aditivním modelem z minulé lekce, který má dva číselné prediktory spojené znakem +:

```
# Aditivní model: jeden společný sklon průměru pro všechny výšky
mod_aditivni <-
  lm(
    formula = objem_m3 ~ prumer_cm + vyska_m,
    data = data_stromy
  )
coef(object = mod_aditivni)
```

```
(Intercept)    prumer_cm    vyska_m
-1.64202494    0.05248820    0.03151742
```

Přímka v každém panelu je předpověď modelu pro **střední výšku daného pásma** (medián: 21,0; 22,9; 24,4; 25,3 m) a vede jen přes průměry kmenů pozorované v tomto pásmu. V pásmech je 9, 7, 8, 7 stromů.



Co v panelech vidíte?

Mají přímky ve všech panelech stejný sklon? Leží body nejnižších a nejvyšších stromů kolem svých průměků stejně?

Možná odpověď

Přímky jsou rovnoběžné: aditivní model dává všem výškám společný sklon 0,052 m³ na 1 cm průměru. V nejnižším pásmu však tenké kmeny leží nad přímkou a tlusté pod ní, takže body stoupají pozvolněji než příмка. V nejvyšším pásmu je to naopak a body stoupají strměji. Společný sklon tedy zřejmě nevystihuje nízké ani vysoké stromy.

20 Model s interakcí dovolí, aby se sklon průměru měnil s výškou

Stejně jako u krabů zapíšeme interakci hvězdičkou. Rozdíl je jen v tom, že druhý prediktor je číslo: interakční člen `prumer_cm:vyska_m` dovolí, aby se **sklon průměru plynule měnil s výškou stromu**.

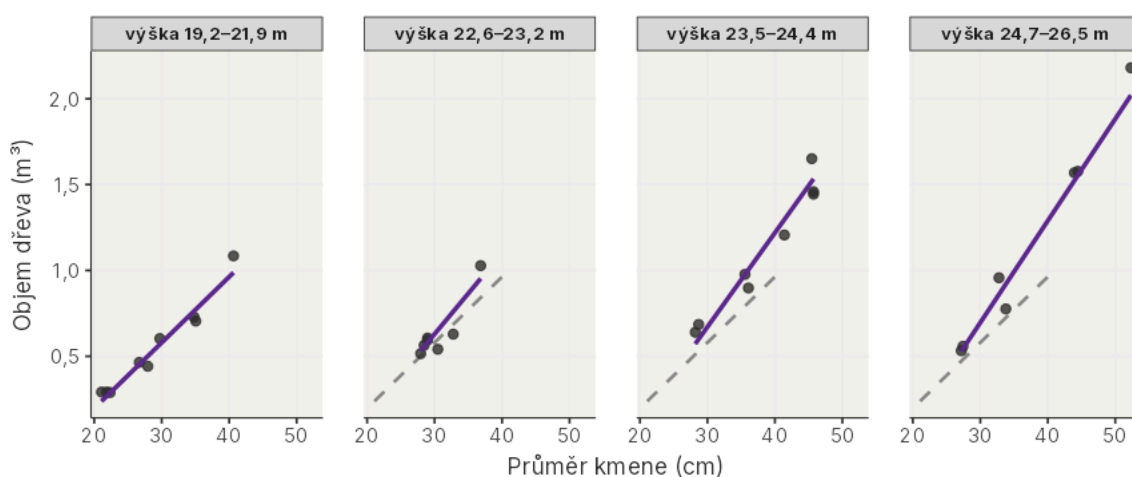
```
# Hvězdička dovolí, aby sklon průměru závisel na výšce
mod_stromy <-
  lm(
    formula = objem_m3 ~ prumer_cm * vyska_m,
    data = data_stromy
```

```
)  
coef(object = mod_stromy)
```

(Intercept)	prumer_cm	vyska_m	prumer_cm:vyska_m
1.965081588	-0.065283022	-0.120502790	0.004925111

Model má opět čtyři koeficienty. Jejich význam rozebereme v další části; nejprve se podíváme, jaké přímky z nich vzniknou.

Na stejné panely nyní nakreslíme přímky modelu s interakcí. Šedá přerušovaná čára opakuje v každém panelu přímku nejnižšího pásma, aby bylo vidět, jak se sklon s výškou mění.



Jak strmá je přímka při určité výšce? Sklon průměru složíme ze dvou koeficientů z výpisu `coef()`, z řádku `prumer_cm` a z interakčního řádku `prumer_cm:vyska_m`:

$$\begin{aligned} \text{sklon průměru při zvolené výšce} &= \text{koeficient průměru} \\ &+ \text{interakční koeficient} \times \text{zvolená výška} \end{aligned}$$

Při výšce 21 m dosadíme:

$$\text{sklon průměru} = -0,0653 + 0,00493 \cdot 21 \approx 0,038 \text{ m}^3 \text{ na 1 cm}$$

Stručně $\beta_1 + \beta_3 v$, kde v je zvolená výška v metrech. Koeficienty číslujeme v pořadí výpisu: β_0 je intercept, β_1 řádek `prumer_cm`, β_2 řádek `vyska_m` a β_3 řádek `prumer_cm:vyska_m`. Stejně dostaneme při výškách 21, 23, 25 m sklony 0,038; 0,048; 0,058 m^3 na 1 cm. S každým metrem výšky navíc přibude ke sklonu 0,0049 m^3 na 1 cm; to je hodnota interakčního koeficientu.

Stejně sklony ukážou i přímé předpovědi. Zvolíme dva průměry a dvě výšky, které leží uvnitř pozorovaných dat, a pro každou kombinaci spočteme odhad objemu:

```
# Zvolené průměry a výšky leží uvnitř pozorovaných rozsahů
prumery_porovnani_cm <- c(28, 36)
vysky_porovnani_m <- c(21, 25)

# Každá kombinace průměru a výšky dostane jeden řádek
data_porovnani_stromy <-
  expand.grid(
    prumer_cm = prumery_porovnani_cm,
    vyska_m = vysky_porovnani_m
  )
data_porovnani_stromy$objem_odhad_m3 <-
  predict(
    object = mod_stromy,
    newdata = data_porovnani_stromy
  )
data_porovnani_stromy
```

	prumer_cm	vyska_m	objem_odhad_m3
1	28	21	0.5025637
2	36	21	0.8077182
3	28	25	0.5721650
4	36	25	1.0349230

Stejný rozdíl v průměru, různý rozdíl v objemu

Kmen o průměru 36 cm místo 28 cm znamená podle modelu u stromu vysokého 21 m o 0,31 m³ dřeva více, u stromu vysokého 25 m o 0,46 m³ více. Po vydělení 8 cm dostaneme právě sklony průměru při těchto výškách.

21 Co znamená koeficient průměru, když se sklon mění s výškou?

Vraťme se ke čtyřem koeficientům z výpisu `coef(object = mod_stromy)`:

- **(Intercept)** je odhad objemu stromu s průměrem 0 cm a výškou 0 m. Takový strom neexistuje; číslo samo o sobě nevykládáme.
- **prumer_cm** je sklon průměru **při výšce 0 m**: -0,065 m³ na 1 cm. Záporná hodnota neznámá, že by tlustší kmen měl méně dřeva; platí pro výšku, kterou žádný změřený strom nemá.
- **vyska_m** je obdobně sklon výšky **při průměru 0 cm**: -0,121 m³ na 1 m. Ani tento koeficient sám nevykládáme.
- **prumer_cm:vyska_m** je interakční člen: o 0,0049 m³ na 1 cm se zvětší sklon průměru, když je strom o 1 m vyšší.

Koeficient průměru neplatí při každé výšce

U dvou číselných prediktorů leží hodnota 0 často daleko mimo data. Hlavní koeficient pak popisuje situaci, kterou jsme nepozorovali. Sklon jednoho prediktoru proto vždy uvádíme **při zvolené hodnotě druhého** uvnitř pozorovaného rozsahu.

Interakci lze číst i obráceně. Stejný koeficient říká, o kolik se zvětší sklon výšky (m^3 dřeva na 1 m výšky), když je kmen o 1 cm tlustší. U stromu s průměrem 28 cm přibude podle modelu na každý metr výšky $0,017 \text{ m}^3$ dřeva, u stromu s průměrem 36 cm $0,057 \text{ m}^3$. Obě věty popisují tutéž vlastnost modelu.

Jak přesně interakční koeficient známe, ukáže 95% interval spolehlivosti:

```
# Interval patří změně sklonu průměru na 1 m výšky
confint(
  object = mod_stromy,
  parm = "prumer_cm:vyska_m"
)
```

```
                2.5 %      97.5 %
prumer_cm:vyska_m 0.003095651 0.006754571
```

Interval pro změnu sklonu průměru (m^3 na 1 cm průměru) při zvýšení stromu o 1 m sahá od 0,0031 do 0,0068. Celý leží nad nulou: při předpokladech modelu tedy data podporují strmější vztah objemu a průměru u vyšších stromů. Platí to pro 31 změřených stromů a v pozorovaném rozsahu průměrů a výšek.

V tomto krátkém příkladu jsme nekontrolovali residua ani vliv jednotlivých stromů, jako jsme to udělali u krabů; 31 stromů je pro takovou kontrolu malý soubor. Objem kmene navíc roste zhruba s druhou mocninou průměru, takže část rozdílu mezi panely by mohl popsat i zakřivený vztah. Přímkou zde proto čteme jako jednoduchý popis v pozorovaném rozsahu, ne jako jediný možný model.

22 Kdy musíme při výkladu jednoho prediktoru uvést hodnotu druhého?

První příklad vyjádřil interakci jako **změnu rozdílu v šířce mezi pohlavími při různých délkách**: dva sklony nemusí být stejné. Druhý příklad ji vyjádřil jako **změnu rozdílu v délce mezi pohlavími u dvou barevných forem**: dvě srovnání samce se samicí nemusí dát stejný výsledek. Třetí příklad ji vyjádřil jako **plynulou změnu sklonu průměru s výškou stromu**: s každým metrem výšky je přímka o něco strmější. Ve všech třech případech musíme při výkladu jednoho prediktoru uvést hodnotu nebo úroveň druhého: ptáme se „**při jaké délce?**“, „**v jaké barevné formě?**“ nebo „**při jaké výšce?**“.

Doplňující: interakce závisí na měřítku odezvy

V příkladech s kraby popisujeme odezvu v milimetrech, takže interakce znamená změnu **rozdílu** v milimetrech. Jiná biologická otázka se může ptát na poměr: například zda jsou samci v obou formách o stejných 10 % delší než samice. Stejný procentní poměr může při různých průměrech vytvářet různé rozdíly v milimetrech.

Poměry a součiny se na logaritmickém měřítku mění v rozdíly a součty. Příklad se stromy to ukazuje na skutečných datech. Objem kmene je zhruba úměrný druhé mocnině průměru krát výšce, tedy *součinu*. Když objem, průměr i výšku nahradíme jejich logaritmy (transformace, ke které se vrátíme v lekci o tvaru rozdělení), změní se tento součin v součet a model bez interakce by mohl stačit. Interakční koeficient takového modelu má 95% interval spolehlivosti od -1,19 do 1,74. Interval je široký a zahrnuje nulu, takže na logaritmickém měřítku data interakci nedokládají, ačkoli v krychlových metrech ji ukazují zřetelně.

Tvrzení „interakce existuje“ proto není úplné bez uvedení měřítka modelu. V lekci o binárních a početních odezvách se setkáme s modely, které používají jiné měřítko než přímé milimetry; interakční koeficient se pak nejprve interpretuje na tomto modelovém měřítku a potom pomocí předpovědí v biologicky srozumitelných jednotkách.

23 Co tedy můžeme říct o krabech a stromech?

Vraťme se k hlavní otázce: **Je rozdíl v zadní šířce krunýře mezi stejně dlouhou samicí a samcem stejný u menších i větších krabů?**

U změřených krabů modré formy není. Zadní šířka roste s délkou u obou pohlaví, ale u samic rychleji: o 0,407 mm na 1 mm délky, u samců o 0,282 mm. Rozdíl samec minus samice se proto mění o -0,124 mm na každý milimetr délky (95% interval spolehlivosti -0,151 až -0,098 mm/mm). Při délce 30 mm má samec podle modelu zadní šířku o 1,76 mm menší než stejně dlouhá samice. Graf residuí neukázal zakřivení, které by dvě přímky přehlédly.

Výsledek odpovídá biologickému očekávání z úvodu, že samice mají s rostoucí velikostí relativně širší zadní část krunýře. Data však jsou pozorovací a není doloženo, že krabi byli vybráni náhodně: popisují souvislost u těchto 100 krabů a v pozorovaném rozsahu délek, nikoli příčinu. Druhý příklad ukázal, že stejná logika platí pro dvě skupinové proměnné: rozdíl v délce mezi pohlavími byl u oranžové formy o 4,84 mm menší než u modré, ale průměrné délky závisejí i na tom, které kraby autoři vybrali.

Třetí příklad přenesl stejnou myšlenku na dva číselné prediktory. U 31 změřených stromů byl vztah objemu a průměru tím strmější, čím byl strom vyšší: při 21 m přidal kmen o 8 cm tlustší podle modelu 0,31 m³ dřeva, při 25 m 0,46 m³. Na původním měřítku objemu to odpovídá očekávání z tvaru kmene; platí to jen v pozorovaném rozsahu průměrů a výšek.

24 Co bude následovat dál

Aditivní model a model s interakcí jsou dva kandidátní popisy téhož vztahu. Model s interakcí je pružnější, ale má více parametrů a jeho koeficienty se vykládají složitěji. V následující lekci se budeme ptát, jak mezi několika předem biologicky zdůvodněnými modely vybrat ten, který data popisuje dostatečně dobře bez zbytečné složitosti.

25 Shrnutí

- Interakce prediktorů znamená, že vztah jednoho prediktoru k odezvě závisí na hodnotě nebo úrovni druhého; potřebu interakce má podpořit biologická otázka i graf.
- $x * y$ je zkratka pro $x + y + x:y$; interakční člen u číselného a skupinového prediktoru popisuje rozdíl sklonů.
- Hlavní koeficienty v modelu s interakcí jsou podmíněné: platí při nulové hodnotě druhého prediktoru nebo v referenční skupině. Skupiny proto porovnáváme pomocí předpovědi při zvolené hodnotě.
- U dvou skupinových prediktorů je interakce rozdíl dvou rozdílů.
- U dvou číselných prediktorů interakce znamená, že se sklon jednoho plynule mění s hodnotou druhého; sklon proto uvádíme při zvolených hodnotách uvnitř dat.
- Závěr omezujeme na pozorovaný rozsah dat, na způsob jejich získání a na souvislost, nikoli příčinu.

25.1 Závěrečná otázka

Uveďte biologickou otázku, u které byste čekali, že vztah dvou znaků závisí na skupině nebo na hodnotě jiného znaku. Jak by vypadal graf?

Možné odpovědi

Například vztah hmotnosti a délky těla u ryb z teplé a studené vody: v teplejší vodě mohou ryby s rostoucí délkou přibírat rychleji, takže přímkami by se rozbíhaly. Nebo vztah dávky hnojiva a výnosu u dvou odrůd plodiny, z nichž jedna na hnojivo reaguje silněji. V obou případech by bodový graf ukázal dvě skupiny bodů s různým sklonem a aditivní model s rovnoběžnými přímkami by takový rozdíl nezachytil.

Místo skupiny může vztah měnit i další číselný znak: například růst semenáčků s teplotou může být strmější ve vlhčí půdě. Graf by pak měl několik panelů pro různou vlhkost půdy a přímkami růstu proti teplotě by byla s rostoucí vlhkostí strmější.

Slovníček pojmů

Pojmy použité v těchto materiálech a jejich definice. Úplný průběžně aktualizovaný slovníček je dostupný také [online](#).

Interaktivní slovníček je dostupný online: <https://cuni-natur-biostatistics.github.io/slovník/>