Linear Models in R (M1–MIDO)
Lab Session 4 — Solutions

Henri PANJO

Table of contents

Dataset Overview: *data_pokemon.csv*

This dataset is adapted from a popular Kaggle Pokémon dataset.
Even if you are not familiar with Pokémon, the data is straightforward:
it combines numeric statistics with categorical attributes, making it well-suited for applying **Ordinary Least Squares (OLS)** in R.

What it contains

- Unique identifiers and names for each Pokémon
- Battle statistics (health, attack, defense, special attack, special defense, speed)
- Categorical features (primary/secondary type, generation, legendary flag)

Fields (Codebook)

- `id`: Unique Pokémon ID
- `name`: Pokémon name
- `type_1`: Primary type (e.g., Water, Fire)
- `type_2`: Secondary type (optional)
- `hp`: Hit points (overall health)
- `attack`: Physical attack strength (**we will use this as** y in most regressions)
- `defense`: Physical defense strength
- `sp_attack`: Special (non-physical) attack strength
- `sp_defense`: Special defense strength
- `speed`: Speed / turn order
- `generation`: Game generation label
- `legendary`: Indicator for legendary status (TRUE/FALSE)

Note on notation

- We treat `attack` as the outcome variable Y.
- Predictor variables (e.g., `defense`, `speed`) will be denoted as $x_1, x_2, \dots$.
- Factors like `type_1` or `legendary` will be included as categorical predictors.

Setup

To keep numbers readable and reproducible, we set display options:

```r
options(scipen = 999, digits = 5)
```

We also load the packages used during this session.

> **Warning**
>
> Don't worry if you don't know them all — we'll introduce functions as we need them. Some provide regression tools, others are for data visualization or diagnostics.

```r
library(broom)
library(performance)
library(parameters)
library(datawizard)
library(see)
library(effectsize)
library(insight)
library(correlation)
library(modelbased)
library(glue)
library(scales)
library(GGally)
library(ggpubr)
library(car)
library(lmtest)
library(multcomp)
library(rstatix)
library(matrixTests)
library(ggfortify)
library(qqplotr)
library(patchwork)
library(ggrepel)
library(gtsummary)
library(kableExtra)
library(openxlsx)
library(janitor)
library(marginaleffects)
library(collapse)
library(tidyverse)
```

```r
source("helper_functions4.R")
```

Data management

We first load the dataset and create the same variables as lab session 3: `typeg`, `second_type`
We also transform the variables `legendary` and `generation` into `factors`.

- Load the data

```r
pok <- read_csv("data_pokemon.csv", show_col_types = FALSE)
```

- Define the 3-level grouping map

```r
type_map3 <- list(
  elemental_env = c("Fire", "Water", "Grass", "Electric", "Ice", "Flying", "Poison"),
  physical_material = c("Bug", "Fighting", "Ground", "Rock", "Steel", "Normal"),
  mystical_supernatural = c("Psychic", "Ghost", "Dragon", "Fairy", "Dark")
)
```

- Variable creation with `mutate()` from `{dplyr}` and labelling with `relabel()`

```r
pok <- pok |>
  mutate(
    typeg = case_when(
      type_1 %in% type_map3$elemental_env ~ "Elemental", # "Elemental",
      type_1 %in% type_map3$physical_material ~ "Physical", # "Physical",
      type_1 %in% type_map3$mystical_supernatural ~ "Mystical", # "Mystical",
      .default = NA_character_
    )
  ) |>
  mutate(second_type = ifelse(type_2 == "None", 0, 1) |> factor(labels = c("No", "Yes"))) |>
  mutate(typeg = fct_infreq(typeg), legendary = as.factor(legendary)) |>
  mutate(generation = factor(generation, labels = paste0("Gen", 1:6))) |>
  relabel(
    typeg = "Primary Type", second_type = "Secondary type",
    legendary = "Legendary", generation = "Pokemon generation",
    attack = "Attack power", speed = "Speed power", defense = "Defense power",
    hp = "Hit points (health)", sp_attack = "Special attack power",
    sp_def = "Special defense power", id = "ID", name = "Pokemon name"
  )
```

- We use `tab_freq1()` from `helper_functions4.R` to get `factor` distributions

```
tab_freq1(pok, c("typeg", "second_type", "legendary", "generation"), digits = 1) |>
  kable(align = "l", padding = 2) |>
  row_spec(c(1, 5, 8, 11), bold = TRUE)
```

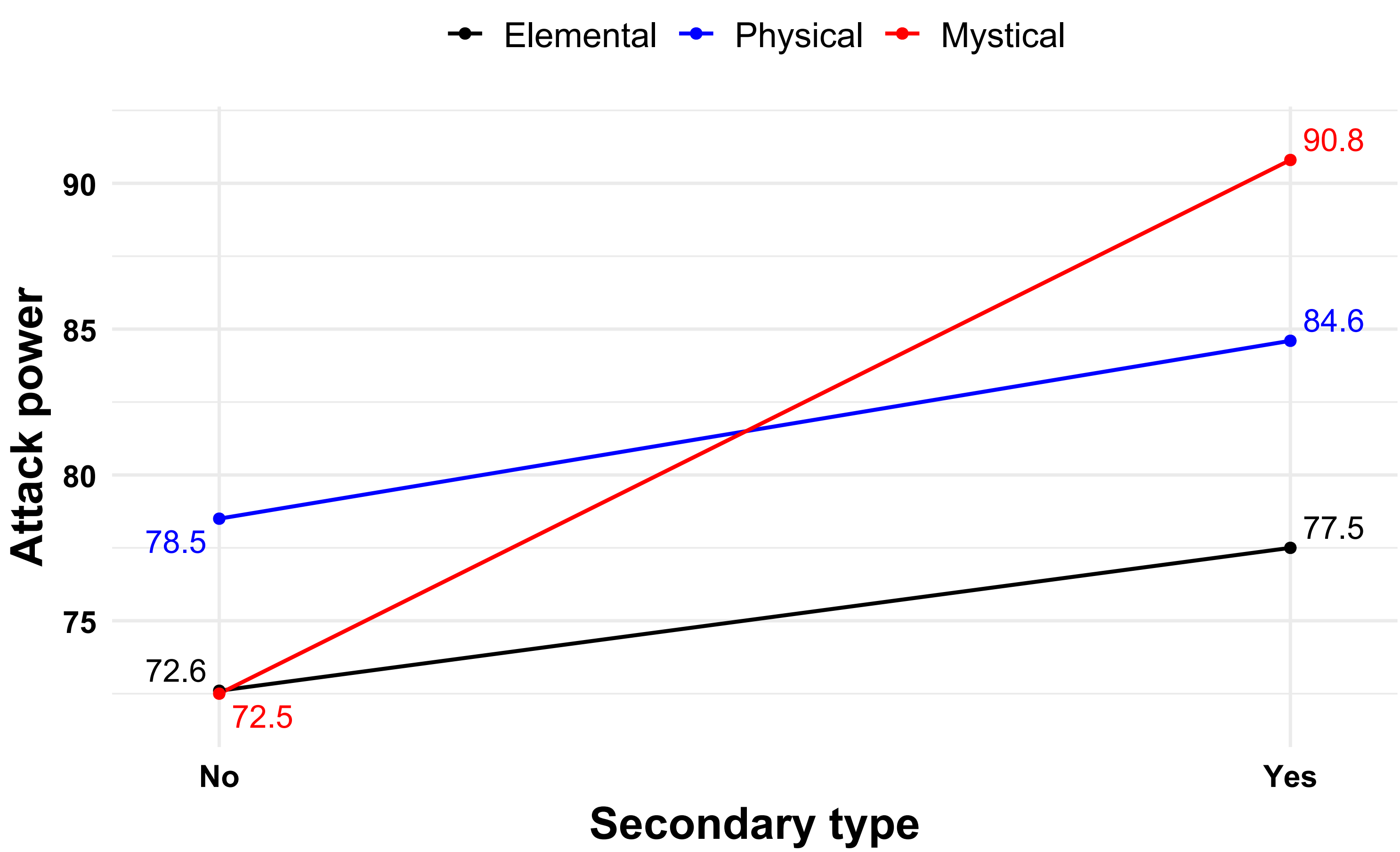
Variable	Count (n)	Percent (%)
Primary Type		
Elemental	334	41.8%
Physical	297	37.1%
Mystical	169	21.1%
Secondary type		
No	386	48.2%
Yes	414	51.7%
Legendary		
No	735	91.9%
Yes	65	8.1%
Pokemon generation		
Gen1	166	20.8%
Gen2	106	13.2%
Gen3	160	20.0%
Gen4	121	15.1%
Gen5	165	20.6%
Gen6	82	10.2%

Question 1: Attack mean with respect to `second_type` and `typeg`

Plot the mean value of `attack` for each combination of `second_type` and `typeg`.

Solution

```
group_by(pok, second_type, typeg) |>
  summarise(m = mean(attack) |> round(1)) |>
  ggplot(aes(x = second_type, y = m, color = typeg, group = typeg, label = m)) +
  geom_line(linewidth = 0.75) +
  geom_point(size = 1.5) +
  geom_text_repel(size = 4.5, show.legend = FALSE, seed = 123) +
  scale_color_manual(values = c("black", "blue", "red")) +
  scale_x_discrete(expand = expansion(0.10)) +
  scale_y_continuous(expand = expansion(0.1)) +
  labs(x = vlabels(pok$second_type), y = vlabels(pok$attack), color = NULL) +
  theme_minimal(base_size = 14) +
  labs_pubr(18) +
  theme(legend.text = element_text(size = 14), legend.position = "top")
```

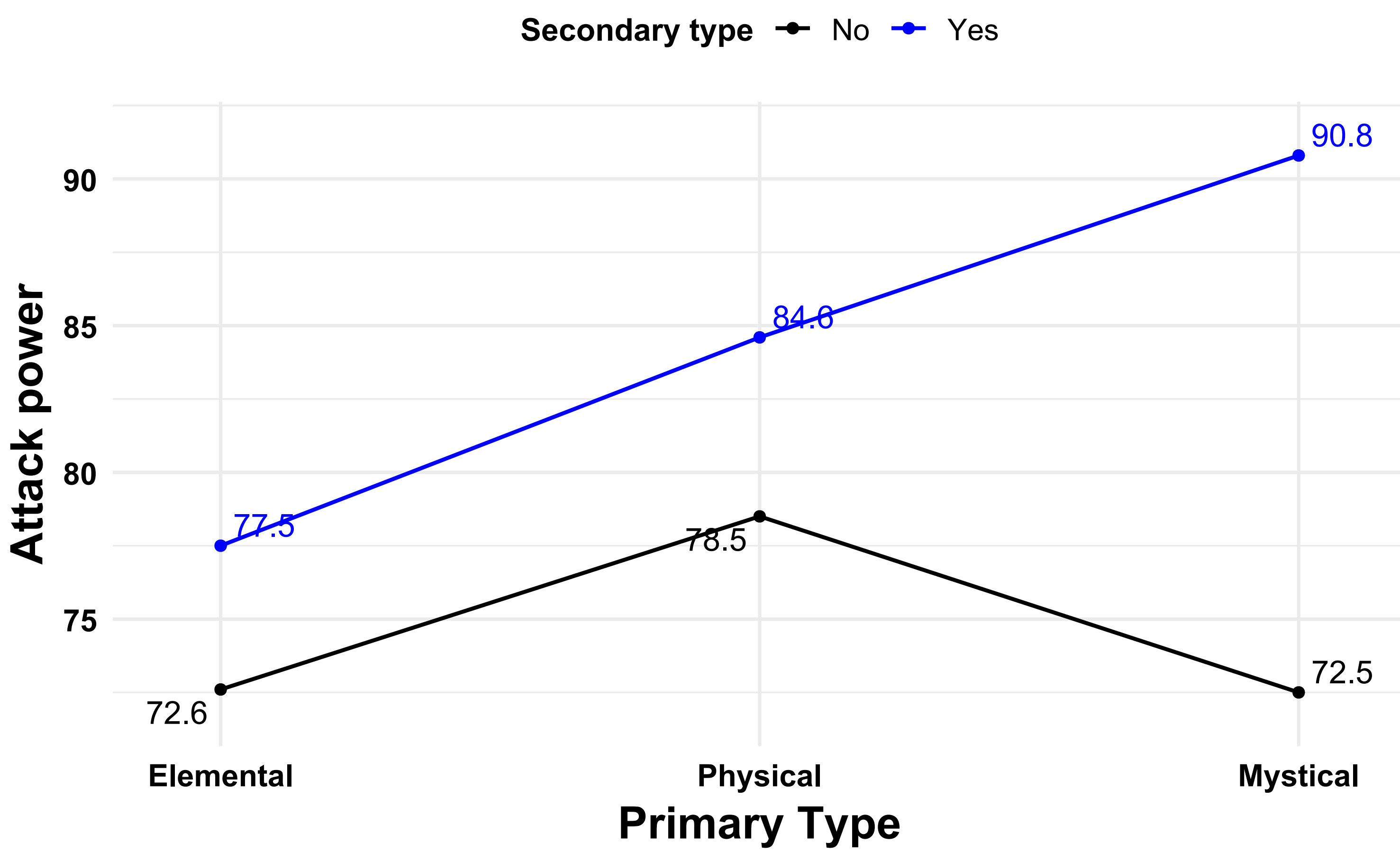


- Using `mean_by_group()` (`helper_functions4.R`) to get a table

```
group_by(pok, typeg) |>
  group_modify(~ mean_by_group(.x, "attack", "second_type")) |>
  group_by(typeg) |>
  mutate(typeg = if_else(is.na(N), typeg, NA)) |>
  frename(vlabels(pok$typeg), cols = 1) |>
  kable(align = "l") |>
  row_spec(c(1, 4, 7), bold = TRUE)
```

Primary Type	Variable	N	Mean (SD)
Elemental	**Secondary type**		
	No	177	72.6 (24.7)
	Yes	157	77.5 (28.9)
Physical	**Secondary type**		
	No	125	78.5 (33.4)
	Yes	172	84.6 (35.0)
Mystical	**Secondary type**		
	No	84	72.5 (36.3)
	Yes	85	90.8 (37.7)

```r
group_by(pok, second_type, typeg) |>
  summarise(m = mean(attack) |> round(1)) |>
  ggplot(aes(x = typeg , y = m, color = second_type, group = second_type, label = m)) +
  geom_line(linewidth = 0.75) +
  geom_point(size = 1.5) +
  geom_text_repel(size = 4.5, show.legend = FALSE, seed = 123) +
  scale_color_manual(values = c("black", "blue", "red")) +
  scale_x_discrete(expand = expansion(0.10)) +
  scale_y_continuous(expand = expansion(0.1)) +
  labs(x = vlabels(pok$typeg), y = vlabels(pok$attack), color = vlabels(pok$second_type)) +
  theme_minimal(base_size = 14) +
  labs_pubr(18) +
  theme(legend.text = element_text(size = 12), legend.position = "top")
```



Question 2: Effect of speed on attack for each primary type

Create separate scatter plots of `attack` versus `speed` for each Pokémon type (`typeg`).

Hint: You can use `scatter_plot()` (`helper_functions4.R`)

Solution

- Scatter plots for Elemental

```
scatter_speed1 <- filter(pok, typeg == levels(typeg)[1]) |>
  scatter_plot("speed", "attack", nbreaks = 10, color = "grey50", ols_line = "black") +
  labs(title = levels(pok$typeg)[1])
```

- Scatter plots for Physical

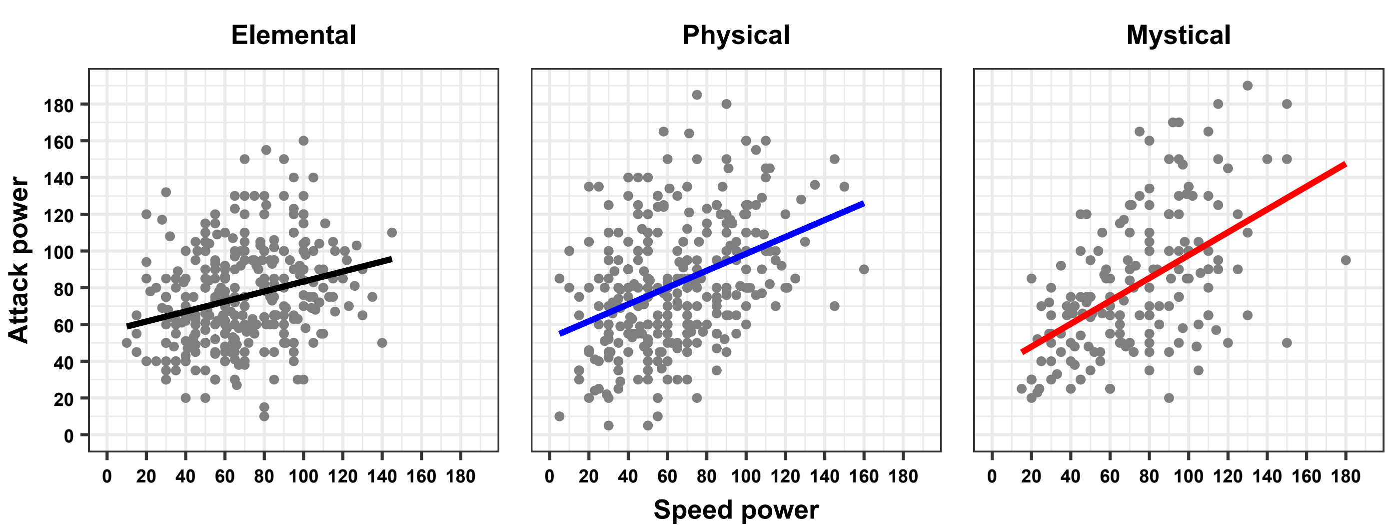
```
scatter_speed2 <- filter(pok, typeg == levels(typeg)[2]) |>
  scatter_plot("speed", "attack", nbreaks = 10, color = "grey50", ols_line = "blue") +
  labs(title = levels(pok$typeg)[2])
```

- Scatter plots for Mystical

```
scatter_speed3 <- filter(pok, typeg == levels(typeg)[3]) |>
  scatter_plot("speed", "attack", nbreaks = 10, color = "grey50", ols_line = "red") +
  labs(title = levels(pok$typeg)[3])
```

- Combine plots with `{patchwork}`

```
scatter_speed1 + scatter_speed2 + scatter_speed3 + plot_layout(axes = "collect") &
  coord_cartesian(xlim = c(0, 190), ylim = c(0, 190)) & labs_pubr(10) &
  theme(plot.title  = element_text(hjust = 0.5))
```



Question 3: Interaction between 2 categorical variables (1)

1. Fit the two models `mod_main1` and `mod_interact1` below, and interpret estimated coefficients.

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

$$\begin{aligned}\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \\ \beta_4 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon\end{aligned}$$

2. Determine whether `mod_interact1` provides a significantly better fit than `mod_main1`. (Equivalently: test whether the interaction terms are jointly equal to zero.)

Solution

Model `mod_main1`

```r
mod_main1 <- lm(attack ~ second_type + typeg, data = pok)

model_parameters(mod_main1, ci_method = "residual", digits = 1) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 1)
```

```
          Parameter  Coefficient [CI]      p
1       (Intercept) 71.1 [67.1, 75.1] <0.001
2 second type [Yes]  8.2 [ 3.7, 12.6] <0.001
3  typeg [Physical]  6.2 [ 1.2, 11.2]  0.016
4  typeg [Mystical]  6.5 [ 0.5, 12.4]  0.032
```

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

Profile	$\mathbb{E}(\texttt{attack} \mid \cdot)$
No secondary type, Elemental	β_0
No secondary type, Physical	$\beta_0 + \beta_2$
No secondary type, Mystical	$\beta_0 + \beta_3$
A Secondary, Elemental	$\beta_0 + \beta_1$
A Secondary, Physical	$\beta_0 + \beta_1 + \beta_2$
A Secondary, Mystical	$\beta_0 + \beta_1 + \beta_3$

Coefficient	Interpretation
$\hat{\beta}_0 = 71.1$ $(p < 0.001)$	Mean attack for Elemental Pokémon with no secondary type
$\hat{\beta}_1 = 8.2$ $(p < 0.001)$	Δ between secondary type, Yes vs No (for Elemental Pokémon)
$\hat{\beta}_2 = 6.2$ $(p = 0.016)$	Δ between Physical and Elemental types, when No secondary type
$\hat{\beta}_3 = 6.5$ $(p = 0.032)$	Δ between Mystical and Elemental types, when No secondary type

Model `mod_interact1`

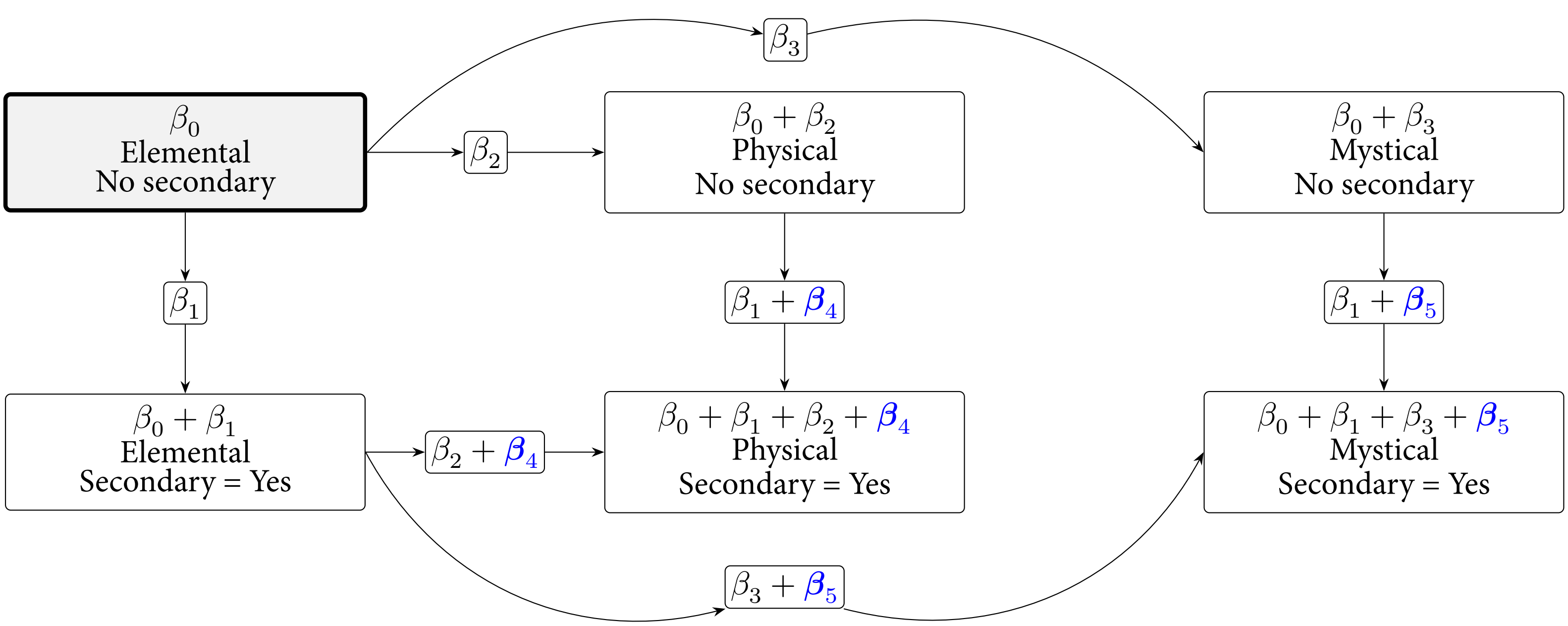
```
mod_interact1 <- lm(attack ~ second_type * typeg, data = pok)

model_parameters(mod_interact1, ci_method = "residual", digits = 1) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 1)
```

```
                             Parameter  Coefficient [CI]      p
1                          (Intercept) 72.6 [67.9, 77.4] <0.001
2                    second type [Yes]  4.9 [-2.0, 11.8]  0.165
3                     typeg [Physical]  5.9 [-1.5, 13.2]  0.116
4                     typeg [Mystical] -0.1 [-8.5,  8.2]  0.974
5 second type [Yes] × typeg [Physical]  1.2 [-8.9, 11.3]  0.821
6 second type [Yes] × typeg [Mystical] 13.4 [ 1.5, 25.3]  0.027
```

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \beta_4 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

Profile	$\mathbb{E}(\texttt{attack} \mid \cdot)$
No secondary type, Elemental	β_0
No secondary type, Physical	$\beta_0 + \beta_2$
No secondary type, Mystical	$\beta_0 + \beta_3$
Secondary type, Elemental	$\beta_0 + \beta_1$
Secondary type, Physical	$\beta_0 + \beta_1 + \beta_2 + \boldsymbol{\beta_4}$
Secondary type, Mystical	$\beta_0 + \beta_1 + \beta_3 + \boldsymbol{\beta_5}$



Interpretation of coefficients

```
                              Parameter  Coefficient [CI]      p
1                           (Intercept) 72.6 [67.9, 77.4] <0.001
2                     second type [Yes]  4.9 [-2.0, 11.8]  0.165
3                      typeg [Physical]  5.9 [-1.5, 13.2]  0.116
4                      typeg [Mystical] -0.1 [-8.5,  8.2]  0.974
5 second type [Yes] × typeg [Physical]  1.2 [-8.9, 11.3]  0.821
6 second type [Yes] × typeg [Mystical] 13.4 [ 1.5, 25.3]  0.027
```

- `(Intercept)`, $\hat{\beta}_0 = 72.6$ ($p < 0.001$): mean `attack` for Pokémon with **No secondary type** and **Elemental** (reference group).

- `second type [Yes]`, $\hat{\beta}_1 = 4.9$ ($p = 0.165$): effect of having a secondary type for **Elemental**. It is the Δ between **secondary type** *vs* **No secondary type**, for **Elemental**

- `typeg [Physical]`, $\hat{\beta}_2 = 5.9$ ($p = 0.116$): Δ between **Physical** and **Elemental** Pokémon, for **No secondary type**

- `typeg [Mystical]`, $\hat{\beta}_3 = -0.1$ ($p = 0.974$): Δ between **Mystical** and **Elemental** Pokémon, for **No secondary type**

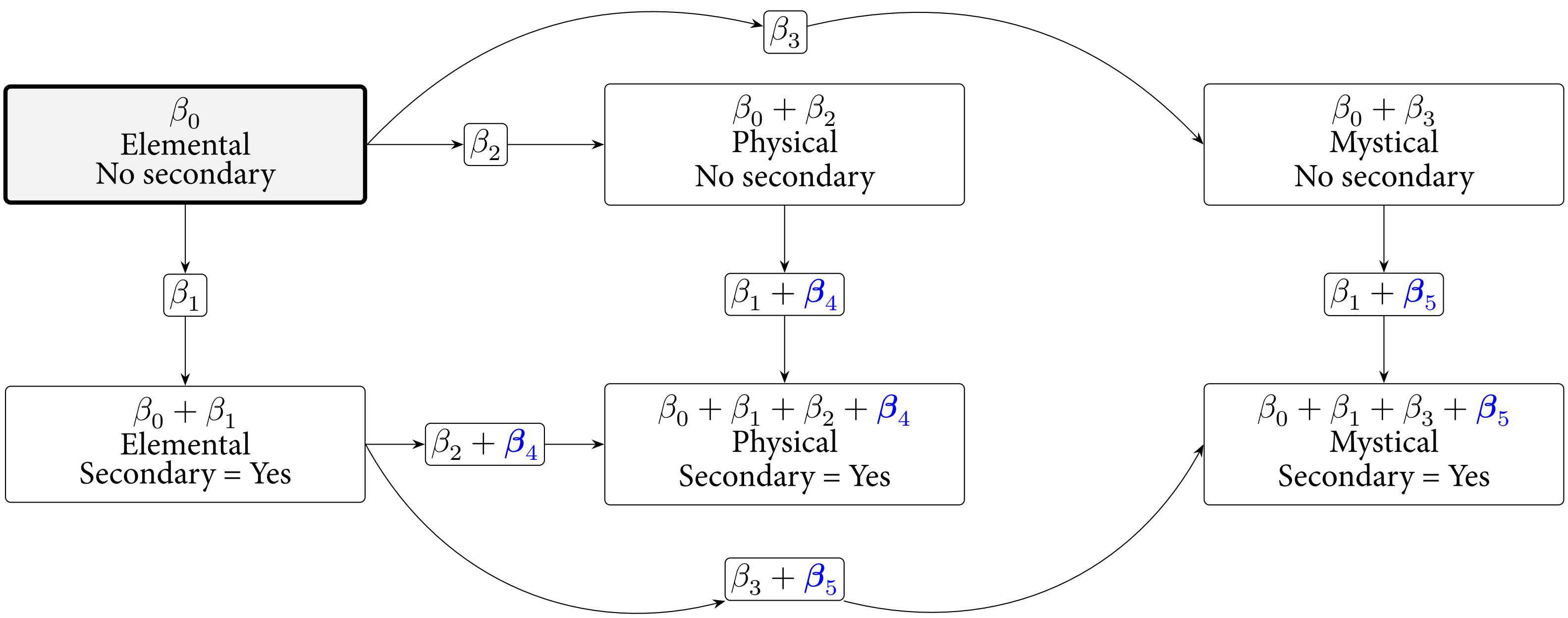
- `second type [Yes] × typeg [Physical]`, $\hat{\beta}_4 = 1.2$ ($p = 0.821$): measures how the effect of having a secondary type changes when moving from Elemental to Physical
  - ▹ β_1 : effect of secondary type for Elemental
  - ▹ $\beta_1 + \beta_4$: effect of secondary type for Physical
  - ▹ so β_4 is the **difference of these two effects**

- `second type [Yes] × typeg [Mystical]`, $\hat{\beta}_5 = 13.4$ ($p = 0.027$): measures how the effect of having a secondary type changes when moving from Elemental to Mystical
  - ▹ β_1 : effect of secondary type for Elemental
  - ▹ $\beta_1 + \beta_5$: effect of secondary type for Mystical
  - ▹ so β_5 is the **difference of these two effects**

Important Remarks: Symmetry of interaction

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \beta_4 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

Profile	$\mathbb{E}(\texttt{attack} \mid \cdot)$
No secondary type, Elemental	β_0
No secondary type, Physical	$\beta_0 + \beta_2$
No secondary type, Mystical	$\beta_0 + \beta_3$
Secondary type, Elemental	$\beta_0 + \beta_1$
Secondary type, Physical	$\beta_0 + \beta_1 + \beta_2 + \boldsymbol{\beta}_4$
Secondary type, Mystical	$\beta_0 + \beta_1 + \beta_3 + \boldsymbol{\beta}_5$



- β_2: Physical *vs* Elemental with **No secondary type**
- $\beta_2 + \beta_4$: Physical *vs* Elemental with **A secondary type**
- So β_4 is the **difference of these two effects**: it measures how the primary-type difference (Physical vs Elemental) changes when a Pokémon has a secondary type.
- If $\beta_4 = 0$, then having a secondary type **does not modify** the difference between Physical and Elemental Pokémon.
- You can now interpret the β_5 interaction the same way

Testing if the interaction terms are useful

$$\begin{aligned}\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \\ \beta_4 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon\end{aligned}$$

$$H_0 : \beta_4 = \beta_5 = 0 \qquad \text{vs} \qquad H_1 : \text{at least one } \beta_j \neq 0$$

- Nested F test between `mod_interact1` and `mod_main1`

```
anova(mod_main1, mod_interact1) |> as.data.frame()
```

```
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    796 819090 NA        NA     NA       NA
2    794 813540  2    5550.1 2.7084 0.067258
```

- `Anova()` on `mod_interact1`

```
Anova(mod_interact1, type = 3) |>
  as.data.frame() |>
  round(4)
```

```
                     Sum Sq  Df  F value Pr(>F)
(Intercept)        934202.7   1 911.7647 0.0000
second_type          1975.3   1   1.9279 0.1654
typeg                2973.9   2   1.4512 0.2349
second_type:typeg    5550.1   2   2.7084 0.0673
Residuals          813539.9 794       NA     NA
```

- `linearHypothesis()` from `{car}`

```
linearHypothesis(mod_interact1,
  c("second_typeYes:typegPhysical = 0", "second_typeYes:typegMystical = 0")
) |>
  as.data.frame()
```

```
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1    796 819090 NA        NA     NA       NA
2    794 813540  2    5550.1 2.7084 0.067258
```

- `waldtest()` from `{lmtest}`

```
waldtest(mod_interact1, "second_type:typeg") |> as.data.frame()
```

```
  Res.Df Df      F   Pr(>F)
1    794 NA     NA       NA
2    796 -2 2.7084 0.067258
```

Question 4: Interaction between 2 categorical variables (2)

Run the code below and interpret the coefficients

```
mod_interact1bis1 <- lm(attack ~ typeg + second_type:typeg, data = pok)
mod_interact1bis2 <- lm(attack ~ second_type + typeg:second_type, data = pok)
```

Solution

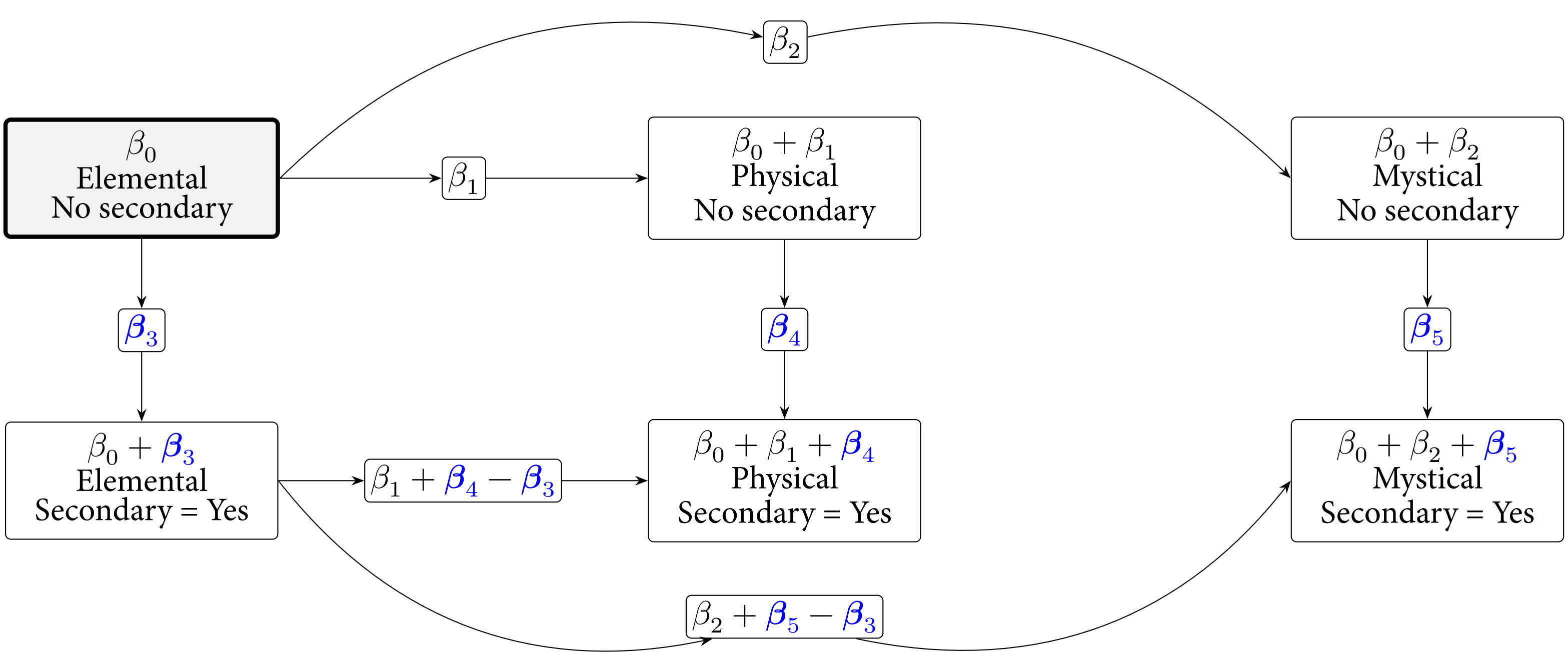
Model `mod_interact1bis1`

```
mod_interact1bis1 <- lm(attack ~ typeg + second_type:typeg, data = pok)

model_parameters(mod_interact1bis1, ci_method = "residual", digits = 1) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 1)
```

```
                          Parameter  Coefficient [CI]      p
1                       (Intercept) 72.6 [67.9, 77.4] <0.001
2                  typeg [Physical]  5.9 [-1.5, 13.2]  0.116
3                  typeg [Mystical] -0.1 [-8.5,  8.2]  0.974
4 typeg [Elemental] × second typeYes  4.9 [-2.0, 11.8]  0.165
5  typeg [Physical] × second typeYes  6.0 [-1.4, 13.4]  0.109
6  typeg [Mystical] × second typeYes 18.3 [ 8.6, 27.9] <0.001
```

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{typeg}_{\texttt{Physical}} + \beta_2 \texttt{typeg}_{\texttt{Mystical}} + \beta_3 \texttt{typeg}_{\texttt{Elemental}} \times \texttt{second\_type}_{\texttt{Yes}} + \beta_4 \texttt{typeg}_{\texttt{Physical}} \times \texttt{second\_type}_{\texttt{Yes}} + \beta_5 \texttt{typeg}_{\texttt{Mystical}} \times \texttt{second\_type}_{\texttt{Yes}} + \varepsilon$$



Interpretation of the Coefficients

- `(Intercept)`, $\hat{\beta}_0 = 72.6$ $(p < 0.001)$: mean `attack` for Pokémon in the **reference group**
- `typeg [Physical]`, $\hat{\beta}_1 = 5.9$ $(p = 0.116)$: Δ between **Physical** and **Elemental**, for Pokémon with **No secondary type**
- `typeg [Mystical]`, $\hat{\beta}_2 = -0.1$ $(p = 0.974)$: Δ between **Mystical** and **Elemental**, for Pokémon with **No secondary type**
- `typeg [Elemental] × second typeYes`, $\hat{\beta}_3 = 4.9$ $(p = 0.165)$: Δ between **A secondary type** and **No secondary type**, for Pokémon **in Elemental**
- `typeg [Physical] × second typeYes`, $\hat{\beta}_4 = 6.0$ $(p = 0.109)$: Δ between **A secondary type** and **No secondary type**, for Pokémon **in Physical**
- `typeg [Mystical] × second typeYes`, $\hat{\beta}_5 = 18.3$ $(p < 0.001)$: Δ between **A secondary type** and **No secondary type**, for Pokémon **in Mystical**

Test of interaction with this specification

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{typeg}_{\texttt{Physical}} + \beta_2 \texttt{typeg}_{\texttt{Mystical}} + \beta_3 \texttt{typeg}_{\texttt{Elemental}} \times \texttt{second\_type}_{\texttt{Yes}} + \\ \beta_4 \texttt{typeg}_{\texttt{Physical}} \times \texttt{second\_type}_{\texttt{Yes}} + \beta_5 \texttt{typeg}_{\texttt{Mystical}} \times \texttt{second\_type}_{\texttt{Yes}} + \varepsilon$$

$$H_0 : \beta_3 = \beta_4 = \beta_5 \qquad \text{vs} \qquad H_1 : \beta_j \neq \beta_k \text{ for at least one pair}$$

$$H_0 : \begin{cases} \beta_3 - \beta_4 = 0 \\ \beta_3 - \beta_5 = 0 \end{cases} \qquad \text{vs} \qquad H_1 : \text{ at least one of these equalities fails}$$

- With `linearHypothesis()` from `{car}`

```r
linearHypothesis(mod_interact1bis1,
  c("typegElemental:second_typeYes = typegPhysical:second_typeYes",
    "typegElemental:second_typeYes = typegMystical:second_typeYes")
) |>
  as_tibble()
```

```
# A tibble: 2 x 6
  Res.Df     RSS    Df `Sum of Sq`     F `Pr(>F)`
   <dbl>   <dbl> <dbl>       <dbl> <dbl>    <dbl>
1    796 819090.    NA          NA  NA      NA
2    794 813540.     2       5550.  2.71    0.0673
```

- We obtain the same p-value as earlier

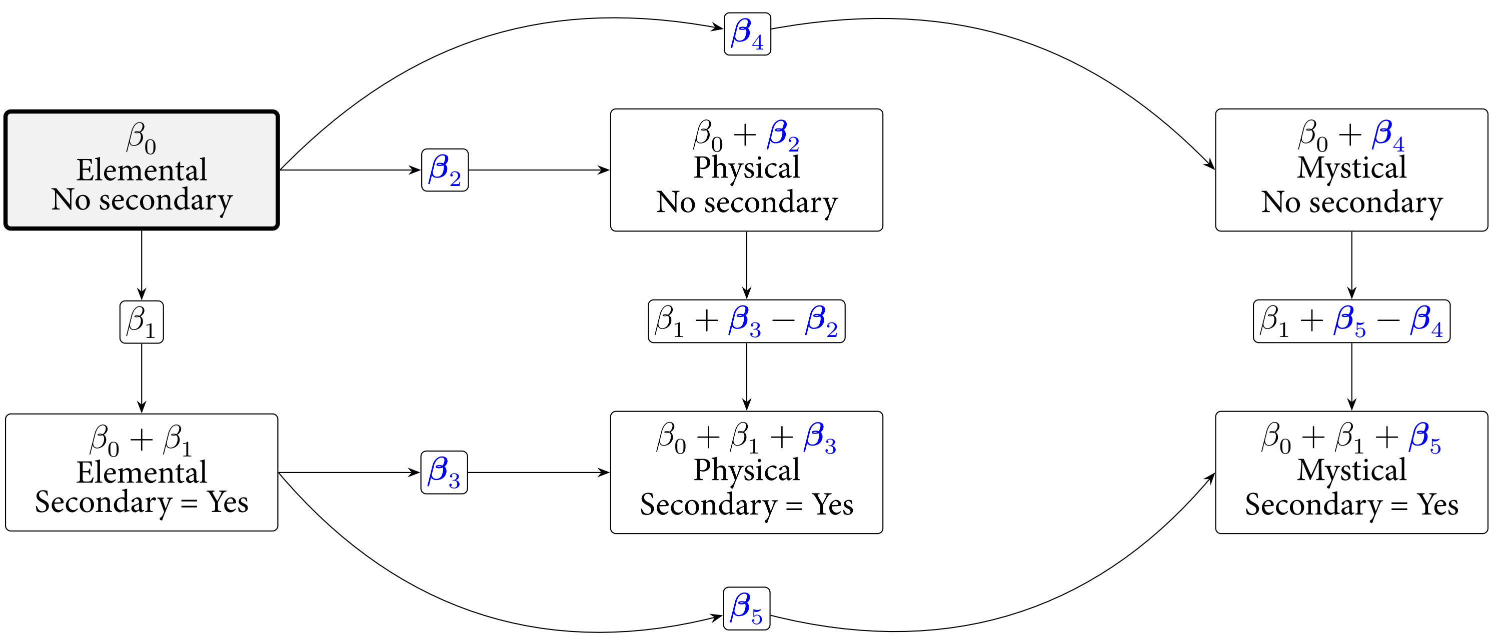
Model `mod_interact1bis2`

```r
mod_interact1bis2 <- lm(attack ~ second_type + typeg:second_type, data = pok)

model_parameters(mod_interact1bis2, ci_method = "residual", digits = 1) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 1)
```

```
                             Parameter  Coefficient [CI]      p
1                          (Intercept) 72.6 [67.9, 77.4] <0.001
2                    second type [Yes]  4.9 [-2.0, 11.8]  0.165
3  second type [No] × typegPhysical  5.9 [-1.5, 13.2]  0.116
4 second type [Yes] × typegPhysical  7.0 [ 0.1, 14.0]  0.046
5  second type [No] × typegMystical -0.1 [-8.5,  8.2]  0.974
6 second type [Yes] × typegMystical 13.3 [ 4.8, 21.7]  0.002
```

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{second\_type}_{\texttt{No}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_4 \texttt{second\_type}_{\texttt{No}} \times \texttt{typeg}_{\texttt{Mystical}} + \beta_5 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$



Interpretation of the Coefficients

- `(Intercept)`, $\hat{\beta}_0 = 72.6$ ($p < 0.001$): mean attack for Pokémon in the reference group (**Elemental**, **No secondary type**)

- `second type [Yes]`, $\hat{\beta}_1 = 4.9$ ($p = 0.165$): Δ between **secondary type** = **Yes** vs **No**, for Pokémon in **Elemental**

- `second type [No] × typegPhysical`, $\hat{\beta}_2 = 5.9$ ($p = 0.116$): Δ between **Physical** and **Elemental**, for Pokémon with **No secondary type**

- `second type [Yes] × typegPhysical`, $\hat{\beta}_3 = 7.0$ ($p = 0.046$): Δ between **Physical** and **Elemental**, for Pokémon with **A secondary type**

- `second type [No] × typegMystical`, $\hat{\beta}_4 = -0.1$ ($p = 0.974$): Δ between **Mystical** and **Elemental**, for Pokémon with **No secondary type**

- `second type [Yes] × typegMystical`, $\hat{\beta}_5 = 13.3$ ($p = 0.002$): Δ between **Mystical** and **Elemental**, for Pokémon with **A secondary type**

Test of interaction with this specification

$$\begin{aligned}\texttt{attack} = {} & \beta_0 + \beta_1 \texttt{second\_type}_{\texttt{Yes}} + \beta_2 \texttt{second\_type}_{\texttt{No}} \times \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Physical}} + \\ & \beta_4 \texttt{second\_type}_{\texttt{No}} \times \texttt{typeg}_{\texttt{Mystical}} + \beta_5 \texttt{second\_type}_{\texttt{Yes}} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon\end{aligned}$$

$$H_0 : \begin{cases} \beta_2 - \beta_3 = 0 \\ \beta_4 - \beta_5 = 0 \end{cases} \qquad \text{vs} \qquad H_1 : \text{ at least one of these equalities fails}$$

- With `linearHypothesis()` from `{car}`

```r
linearHypothesis(mod_interact1bis2,
  c("second_typeNo:typegPhysical = second_typeYes:typegPhysical",
    "second_typeNo:typegMystical = second_typeYes:typegMystical")
) |>
  as_tibble()
```

```
# A tibble: 2 x 6
  Res.Df     RSS    Df `Sum of Sq`     F `Pr(>F)`
   <dbl>   <dbl> <dbl>       <dbl> <dbl>    <dbl>
1    796 819090.    NA          NA  NA      NA
2    794 813540.     2       5550.  2.71   0.0673
```

- We obtain the same p-value as earlier

Question 5: Interaction between a categorical and a continuous

We now fit the 3 following models. Run the code below and interpret the estimated coeffcients.

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{speed} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{speed} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \beta_4 \texttt{speed} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5 \texttt{speed} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

$$\begin{aligned}\texttt{attack} = \beta_0 &+ \beta_1 \texttt{typeg}_{\texttt{Physical}} + \beta_2 \texttt{typeg}_{\texttt{Mystical}} + \beta_3 \texttt{typeg}_{\texttt{Elemental}} \times \texttt{speed} + \beta_4 \texttt{typeg}_{\texttt{Physical}} \times \texttt{speed} + \\ &\beta_5 \texttt{typeg}_{\texttt{Mystical}} \times \texttt{speed} + \varepsilon\end{aligned}$$

```r
mod_main2 <- lm(attack ~ speed + typeg, data = pok)
mod_interact2 <- lm(attack ~ speed * typeg, data = pok)
mod_interact2bis <- lm(attack ~ typeg + speed:typeg, data = pok)
```

Solution

Model `mod_main2`

```r
mod_main2 <- lm(attack ~ speed + typeg, data = pok)

model_parameters(mod_main2, ci_method = "residual", digits = 2) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 2)
```

```
        Parameter     Coefficient [CI]      p
1     (Intercept) 44.89 [39.00, 50.78] <0.001
2           speed  0.44 [ 0.36,  0.51] <0.001
3 typeg [Physical]  9.18 [ 4.51, 13.86] <0.001
4 typeg [Mystical]  4.45 [-1.08,  9.98]  0.115
```

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{speed} + \beta_2 \texttt{typeg}_{\texttt{Physical}} + \beta_3 \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

Type	$\mathbb{E}(\texttt{attack} \mid \texttt{speed}, \texttt{type})$
Elemental/Environmental	$\beta_0 \; + \; \beta_1 \, \texttt{speed}$
Physical/Material	$(\beta_0 + \beta_2) \; + \; \beta_1 \, \texttt{speed}$
Mystical/Supernatural	$(\beta_0 + \beta_3) \; + \; \beta_1 \, \texttt{speed}$

Coefficient	Interpretation
$\hat{\beta}_0 = 44.89 \; (p < 0.001)$	Mean attack for Elemental Pokémon at $\texttt{speed} = 0$
$\hat{\beta}_1 = 0.44 \; (p < 0.001)$	Effect of speed on attack (common slope)
$\hat{\beta}_2 = 9.18 \; (p < 0.001)$	Δ Physical vs Elemental at all speed
$\hat{\beta}_3 = 4.45 \; (p = 0.115)$	Δ Mystical vs Elemental at all speed

Model `mod_interact2`

```
mod_interact2 <- lm(attack ~ speed * typeg, data = pok)

model_parameters(mod_interact2, ci_method = "residual", digits = 2) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 2)
```

```
                 Parameter       Coefficient [CI]      p
1              (Intercept)  56.18 [ 47.28, 65.08] <0.001
2                    speed   0.27 [  0.15,  0.39] <0.001
3          typeg [Physical]  -3.58 [-15.57,  8.40]  0.557
4          typeg [Mystical] -20.70 [-35.11, -6.28]  0.005
5 speed × typeg [Physical]   0.19 [  0.02,  0.35]  0.027
6 speed × typeg [Mystical]   0.35 [  0.17,  0.54] <0.001
```

$$\texttt{attack} = \beta_0 + \beta_1\texttt{speed} + \beta_2\texttt{typeg}_{\texttt{Physical}} + \beta_3\texttt{typeg}_{\texttt{Mystical}} + \beta_4\texttt{speed} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5\texttt{speed} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

Type	$\mathbb{E}(\texttt{attack} \mid \texttt{speed}, \texttt{type})$
Elemental	$\beta_0 + \beta_1\ \texttt{speed}$
Physical	$(\beta_0 + \beta_2)\ +\ (\beta_1 + \boldsymbol{\beta_4})\ \texttt{speed}$
Mystical	$(\beta_0 + \beta_3)\ +\ (\beta_1 + \boldsymbol{\beta_5})\ \texttt{speed}$

Coefficient	Interpretation
$\hat{\beta}_0 = 56.18\ (p < 0.001)$	Mean attack for Elemental Pokémon when $\texttt{speed} = 0$
$\hat{\beta}_1 = 0.27\ (p < 0.001)$	Effect of speed (slope) for Elemental Pokémon.
$\hat{\beta}_2 = -3.58\ (p = 0.557)$	Δ between Physical and Elemental, at $\texttt{speed} = 0$
$\hat{\beta}_3 = -20.70\ (p = 0.005)$	Δ between Mystical and Elemental, at $\texttt{speed} = 0$
$\hat{\beta}_4 = 0.19\ (p = 0.027)$	Δ in speed effect (slope) between Physical and Elemental
$\hat{\beta}_5 = 0.35\ (p < 0.001)$	Δ in speed effect (slope) between Mystical and Elemental

Test of interaction

$$\texttt{attack} = \beta_0 + \beta_1\texttt{speed} + \beta_2\texttt{typeg}_{\texttt{Physical}} + \beta_3\texttt{typeg}_{\texttt{Mystical}} + \beta_4\texttt{speed} \times \texttt{typeg}_{\texttt{Physical}} + \beta_5\texttt{speed} \times \texttt{typeg}_{\texttt{Mystical}} + \varepsilon$$

$$H_0 : \beta_4 = \beta_5 = 0 \qquad \text{vs} \qquad H_1 : \text{at least one } \beta_j \neq 0$$

```
waldtest(mod_interact2, "speed:typeg") |> as.data.frame()
```

```
  Res.Df Df      F      Pr(>F)
1    794 NA     NA          NA
2    796 -2 7.0346 0.00093687
```

Model `mod_interact2bis`

```
mod_interact2bis <- lm(attack ~ typeg + speed:typeg, data = pok)

model_parameters(mod_interact2bis, ci_method = "residual", digits = 2) |>
  format_table(select = "{estimate} [{ci}]|{p}", digits = 2)
```

```
                   Parameter       Coefficient [CI]      p
1                (Intercept)  56.18 [ 47.28, 65.08] <0.001
2           typeg [Physical]  -3.58 [-15.57,  8.40]  0.557
3           typeg [Mystical] -20.70 [-35.11, -6.28]  0.005
4 typeg [Elemental] × speed   0.27 [  0.15,  0.39] <0.001
5  typeg [Physical] × speed   0.46 [  0.35,  0.57] <0.001
6  typeg [Mystical] × speed   0.62 [  0.48,  0.76] <0.001
```

$$\texttt{attack} = \beta_0 + \beta_1 \texttt{typeg}_{\texttt{Physical}} + \beta_2 \texttt{typeg}_{\texttt{Mystical}} + \beta_3 \texttt{typeg}_{\texttt{Elemental}} \times \texttt{speed} + \beta_4 \texttt{typeg}_{\texttt{Physical}} \times \texttt{speed} + \beta_5 \texttt{typeg}_{\texttt{Mystical}} \times \texttt{speed} + \varepsilon$$

Type	$\mathbb{E}(\texttt{attack} \mid \texttt{speed}, \texttt{typeg})$
Elemental	$\beta_0 + \boldsymbol{\beta_3}\,\texttt{speed}$
Physical	$(\beta_0 + \beta_1) + \boldsymbol{\beta_4}\,\texttt{speed}$
Mystical	$(\beta_0 + \beta_2) + \boldsymbol{\beta_5}\,\texttt{speed}$

Coefficient	Interpretation
$\hat{\beta}_0 = 56.18\ (p < 0.001)$	Mean attack for Elemental Pokémon at $\texttt{speed} = 0$
$\hat{\beta}_1 = -3.58\ (p = 0.557)$	Δ between Physical and Elemental Pokémon at $\texttt{speed} = 0$
$\hat{\beta}_2 = -20.70\ (p = 0.005)$	Δ between Mystical and Elemental Pokémon at $\texttt{speed} = 0$
$\hat{\beta}_3 = 0.27\ (p < 0.001)$	Speed effect (slope) for Elemental Pokémon
$\hat{\beta}_4 = 0.46\ (p < 0.001)$	Speed effect (slope) for Physical Pokémon
$\hat{\beta}_5 = 0.62\ (p < 0.001)$	Speed effect (slope) for Mystical Pokémon

Test of interaction with this specification

$$H_0 : \begin{cases} \beta_3 - \beta_4 = 0 \\ \beta_3 - \beta_5 = 0 \end{cases} \qquad \text{vs} \qquad H_1 : \text{ at least one of these equalities fails}$$

```
linearHypothesis(mod_interact2bis,
  c("typegElemental:speed = typegPhysical:speed",
    "typegElemental:speed = typegMystical:speed")
) |>  as.data.frame()
```

```
  Res.Df    RSS Df Sum of Sq      F      Pr(>F)
1    796 706203 NA        NA     NA          NA
2    794 693907  2     12296 7.0346 0.00093687
```

Session Info

Package	Version
broom	1.0.10
car	3.1-3
collapse	2.1.4
correlation	0.8.8
datawizard	1.3.0
effectsize	1.0.1
GGally	2.4.0
ggfortify	0.4.19
ggpubr	0.6.2
ggrepel	0.9.6
glue	1.8.0
gtsummary	2.4.0
insight	1.4.2
janitor	2.2.1
kableExtra	1.4.0
lmtest	0.9-40
marginaleffects	0.30.0
matrixTests	0.2.3.1
modelbased	0.13.0
multcomp	1.4-29
openxlsx	4.2.8
parameters	0.28.2
patchwork	1.3.2
performance	0.15.2
qqplotr	0.0.7
rstatix	0.7.3
scales	1.4.0
see	0.12.0
tidyverse	2.0.0