Multiple regression

two models, one question

*more than one explanatory variable, with **moderndive***

`+` vs `*`

```
lm(y ~ x1 + x2)
lm(y ~ x1 * x2)
geom_parallel_slopes()
coef()
```

Fertility rate, life expectancy, and income group

Why add a variable?

One explanatory variable can mislead you when a second one is doing part of the work.

Hold The Other One Fixed

- A slope in simple regression ignores everything else.
- In multiple regression each slope holds the **other** variables fixed.
- Adding a categorical variable gives each group its own intercept.
- Adding an interaction gives each group its own **slope** too.

$$y \sim x1 + x2$$

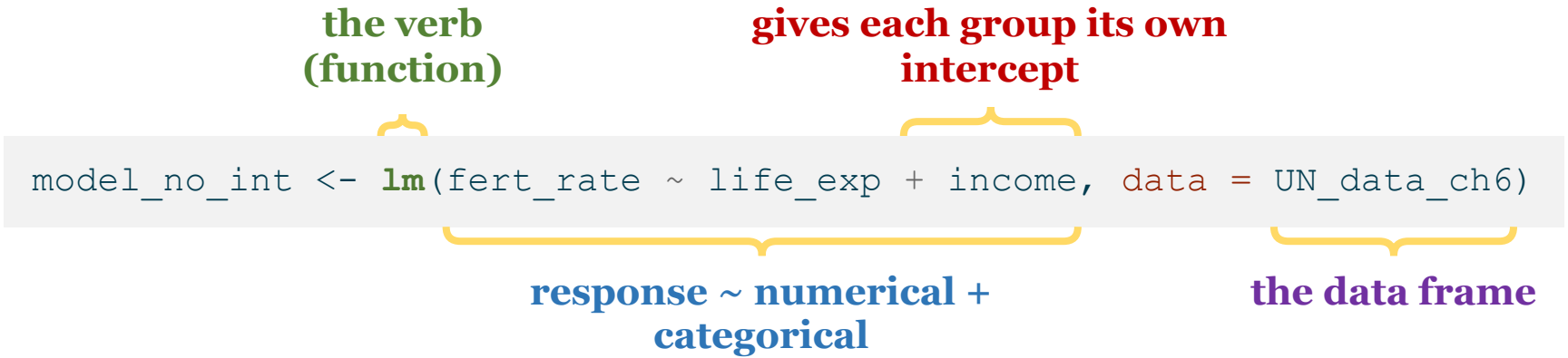
The **parallel slopes** model. Each group gets its own intercept, but every group shares one common slope.

Parallel Slopes

One slope for everybody, shifted up or down by group.

lm()

Example


the verb (function)
gives each group its own intercept
model_no_int <- lm(fert_rate ~ life_exp + income, data = UN_data_ch6)
response ~ numerical + categorical
the data frame

geom_parallel_slopes()

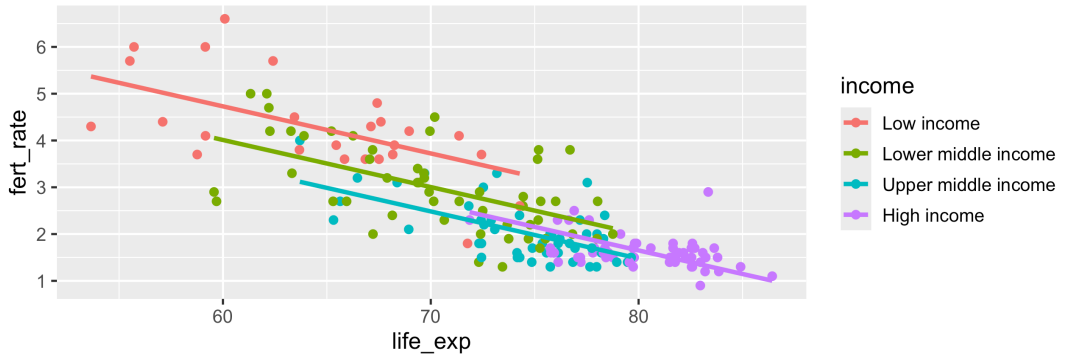
Draw the parallel slopes model. Same code as a scatterplot, with **geom_smooth()** swapped out.

one color per group

```
ggplot(UN_data_ch6, aes(x = life_exp, y = fert_rate, color = income)) +
  geom_point() +
  geom_parallel_slopes(se = FALSE)
```

one line per group, all the same slope

Example



$$y \sim x1 * x2$$

The **interaction** model. Each group gets its own intercept **and** its own slope, so the lines are free to fan out.

Interaction

Use it when the effect of x differs from group to group.

lm()

The star fits an **interaction**: each group gets its own slope as well as its own intercept.

Example

the verb (function)

interaction: slopes differ too

```
model_int <- lm(fert_rate ~ life_exp * income,
                data = UN_data_ch6)
```

the data frame

coef()

the fitted model

```
coef(model_int)
```

the verb (function)

Read the coefficients. With an interaction, the extra rows are **offsets** from the baseline group, not slopes on their own.

Example

```
                                (Intercept)                                    life_exp
                                11.91826460                                -0.11848075
                  incomeLower middle income                  incomeUpper middle income
                                -1.50420617                                -1.89291757
                     incomeHigh income life_exp:incomeLower middle income
                                -6.57953685                                 0.01265027
life_exp:incomeUpper middle income                  life_exp:incomeHigh income
                                 0.01117481                                 0.07221696
```

Reading the coefficients

The baseline group is the first level of the factor: here, Low income.

Baseline & Offsets

- The **(Intercept)** entry belongs to the baseline group.
- A group's entry is how far its intercept sits from the baseline.
- In an interaction model, an **x:group** entry shifts that group's slope.
- Add the offset to the baseline to get the group's own value.

Which model?

Fit both, look at both, then choose.

Parallel Or Interaction

- Plot the data colored by group before fitting anything.
- If the clouds trend at visibly different angles, try the interaction.
- If the lines would be nearly parallel anyway, the simpler model wins.
- Compare the two fitted models' coefficients side by side with **coef()**.