Simple linear regression

*with the **moderndive** package*

one explanatory variable

```
y ~ x
```

lm()
coef()
get_regression_points()
geom_smooth()

Fertility rate and life expectancy across countries

```
y ~ x
```

The **model formula**. Read the tilde as “is modeled by”. The response variable goes on the left, the explanatory variable on the right.

Response ~ Explanatory

Every modeling function in this chapter takes a formula and a data frame.

lm()

Fit a **linear regression** model: the single straight line that comes closest to every point in the cloud.

Fit a Linear Model



lm()

Example

```
demographics_model <- lm(fert_rate ~ life_exp, data = UN_data_ch5)
```

the verb (function)

the model formula response ~ explanatory

the data frame

coef()

Pull the **regression coefficients** out of a fitted model: the intercept and the slope you interpret.

Read the Coefficients

Two numbers: the intercept, then the slope for the explanatory variable.

coef()

Example

the fitted model

```
coef(demographics_model)
```

the verb (function)

```
(Intercept)    life_exp
 12.5992926  -0.1372879
```

get_regression_points()

Return the **fitted value** and the **residual** for every observation, so you can see how far each point sits from the line.

Fitted Values & Residuals

residual = observed y minus fitted y-hat

get_regression_points()

the fitted model

```
get_regression_points(demographics_model)
```

the verb (function)

Example

```
# A tibble: 181 × 5
     ID fert_rate life_exp fert_rate_hat residual
  <int>     <dbl>    <dbl>         <dbl>    <dbl>
1     1       4.3     53.6          5.23   -0.934
2     2       1.4     79.5          1.69   -0.289
3     3       2.7     78.0          1.89    0.813
4     4       5       62.1          4.07    0.928
# ℹ 177 more rows
```

geom_smooth()

Draw the **regression line** on top of a scatterplot.

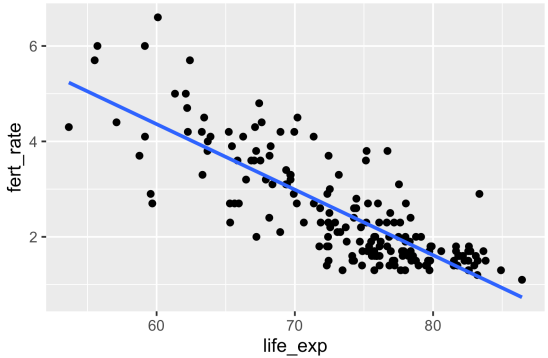
```
ggplot(UN_data_ch5, aes(x = life_exp, y = fert_rate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

what kind of layer to add

fit a straight line

no confidence band

Example



Interpreting

What the two regression coefficients actually mean.

Intercept & Slope

- The **intercept** is the fitted value when x is zero.
- The **slope** is the change in y for a one-unit increase in x.
- A **residual** is what the model missed for one observation.
- Correlation measures strength; the **slope** carries the units.