Meaningful Interpretation

Applied Regression in R

Aleš Vomáčka

22. 02. 2026

Faculty of Arts, Charles University

The Problem with Regression Coefficients

The Problem with Regression Coefficients

Making complex models is easy. Interpreting them is not.

```
lm(sex_partners ~ poly(sexlib_index, 2) * gender
```

The Problem with Regression Coefficients

Parameter	Coefficient
(Intercept)	7.493
sexlib_index	104.866
sexlib_index^2	15.837
genderWomen	-0.403
sexlib_index:genderWomen	-33.203
sexlib_index^2:genderWomen	-3.346

The Problem with Regression Coefficients

Interpreting more complex models using just regression coefficients is (borderline) impossible.

But we have other tools to do it - **predictions, plots, contrasts and marginal effects**.

Predictions

Predictions

Instead of interpreting everything at once, we can look at expected (predicted) values while controlling for other variables.

For example: *What is the expected number of sexual partners for men and women, after controlling for sexual liberalism?*

Predictions

For example: *What is the expected number of sexual partners for men and women, after controlling for sexual liberalism?*

But what does **„controlling for“** even mean?

Predictions

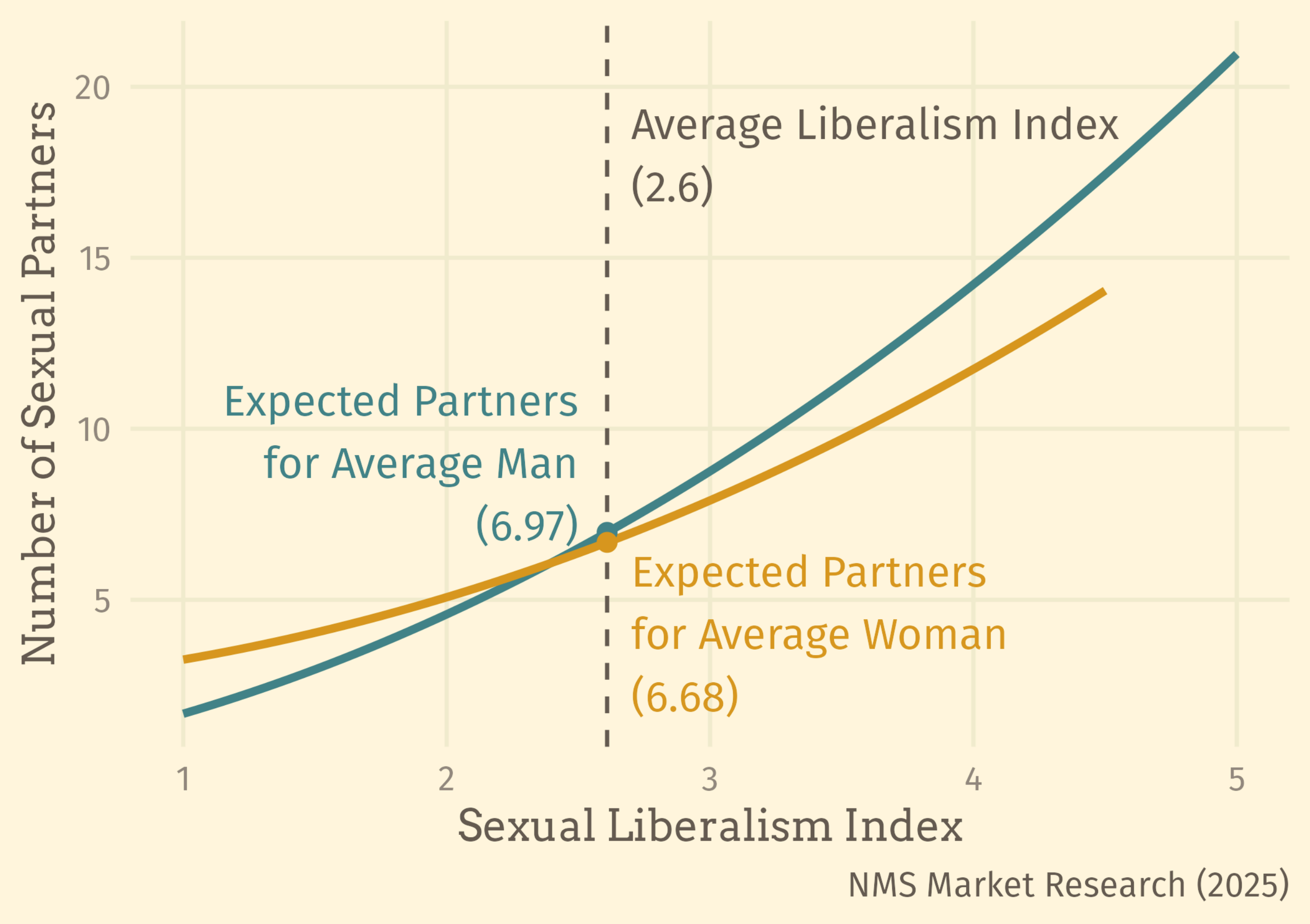
Option 1: We can compute the expected number of partners of men and women for **people with average level of sexual liberalism**.

Using the `modelbased` package:

```
estimate_means(model, by = "gender", estimate = "typical")
```

Expected Number of Partners for an Average Man and Woman

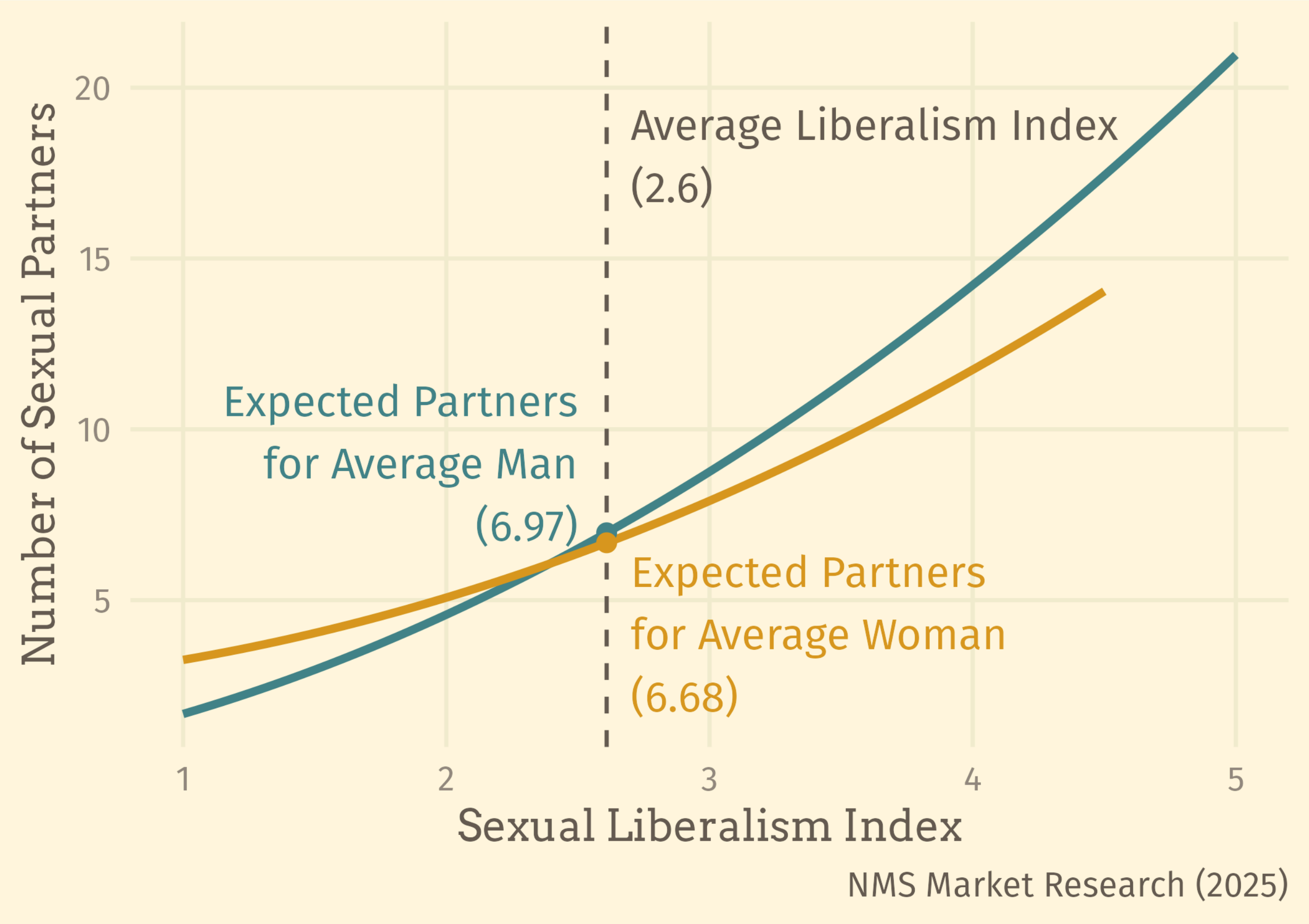
gender	Mean
Men	6.97
Women	6.68



Expected Number of Partners for an Average Man and Woman

What's happening:

1. Plug average value of sex. liberalism (2.6) into the regression formula.
2. Compute expected value for Men and Women.



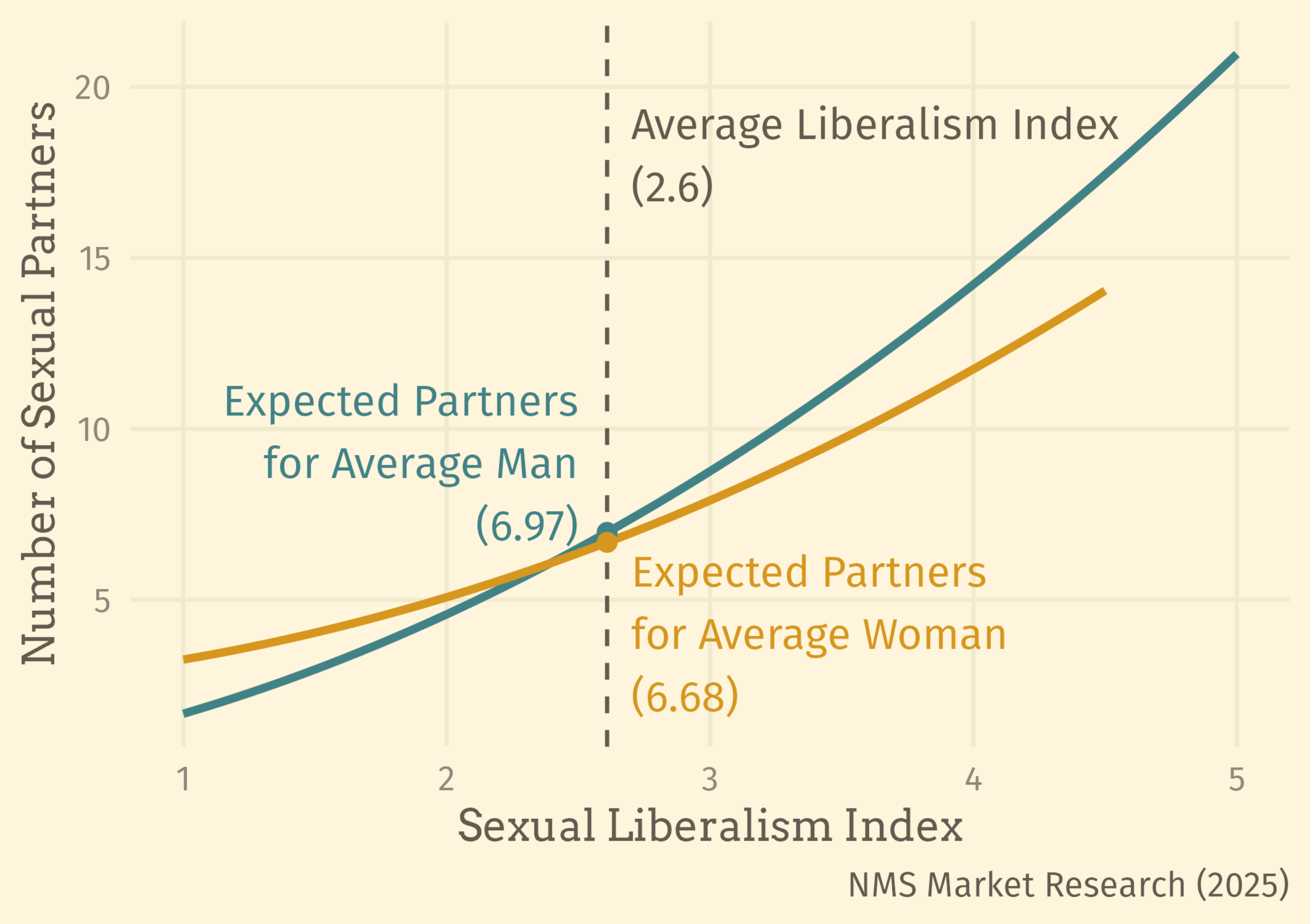
Expected Number of Partners for an Average Man and Woman

Advantage:

- Simple to compute
- Close to regression coefficients.

Disadvantage:

- ???



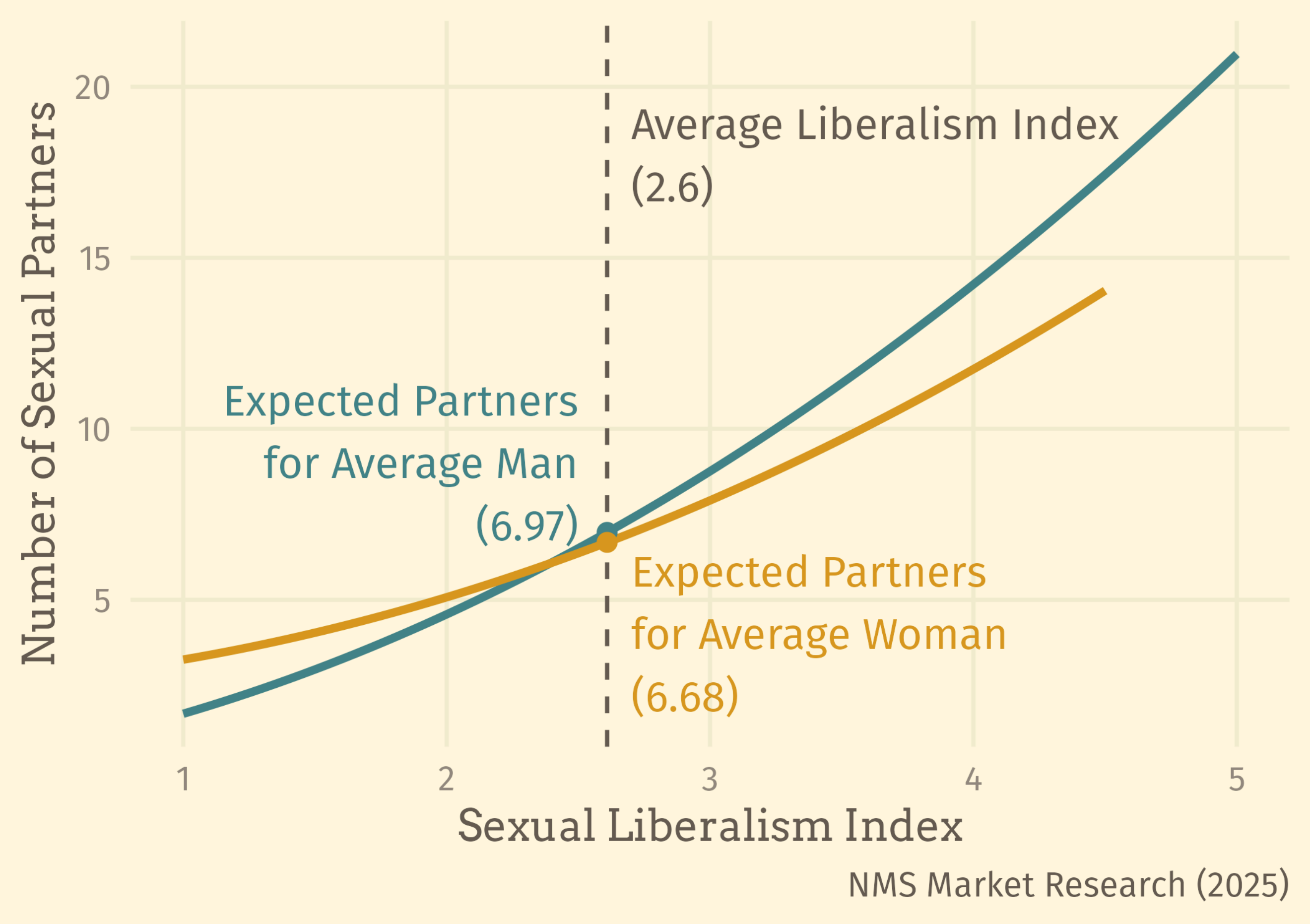
Expected Number of Partners for an Average Man and Woman

Advantage:

- Simple to compute
- Close to regression coefficients.

Disadvantage:

- Only really works for simple models



Questions?

Predictions

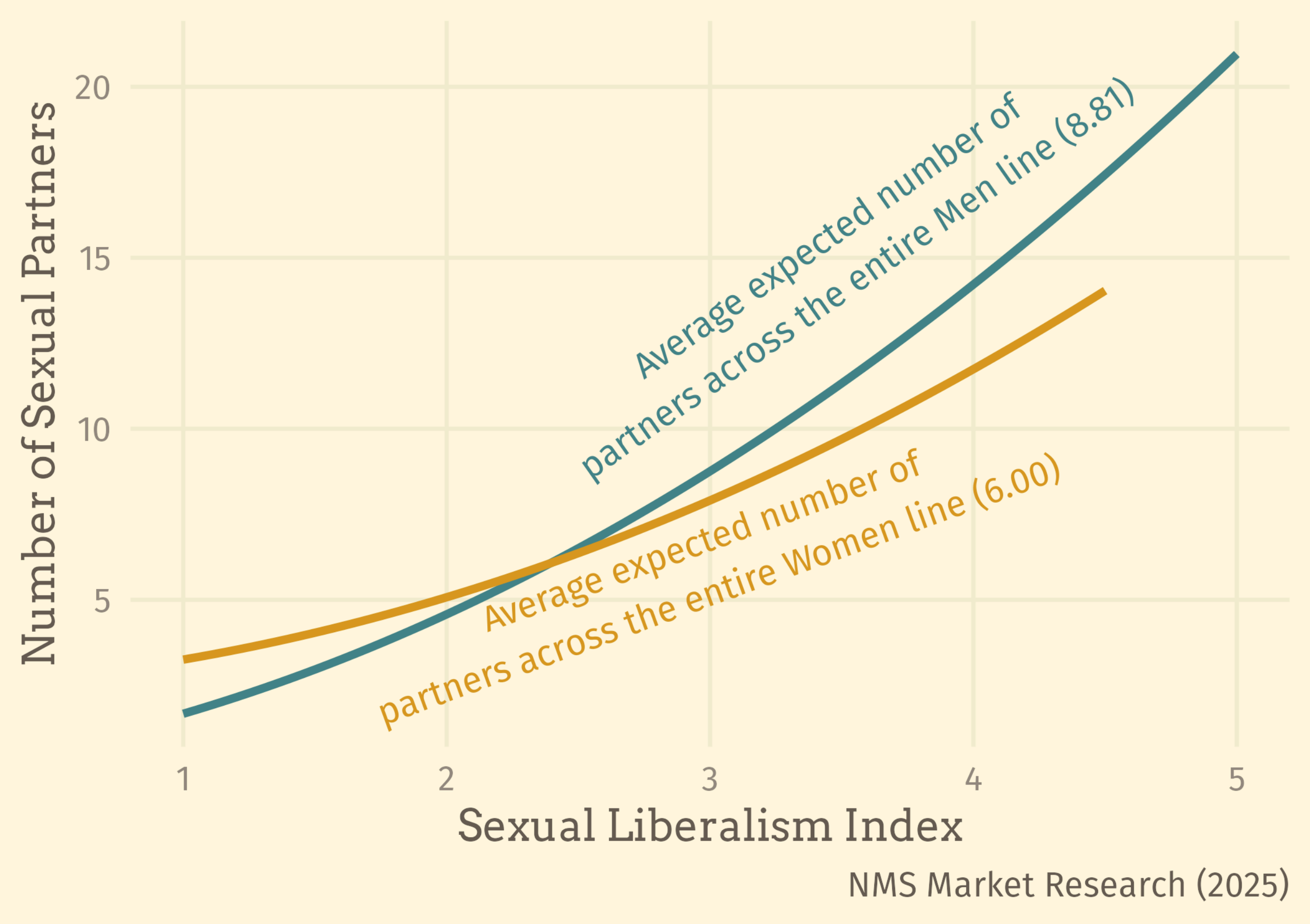
Option 2: We can compute **the average expected number of partners** of men and women across different levels of sexual liberalism.

Using the `modelbased` package:

```
estimate_means(model, by = "gender", estimate = "average")
```

Average Expected Number of Partners for Men and Women

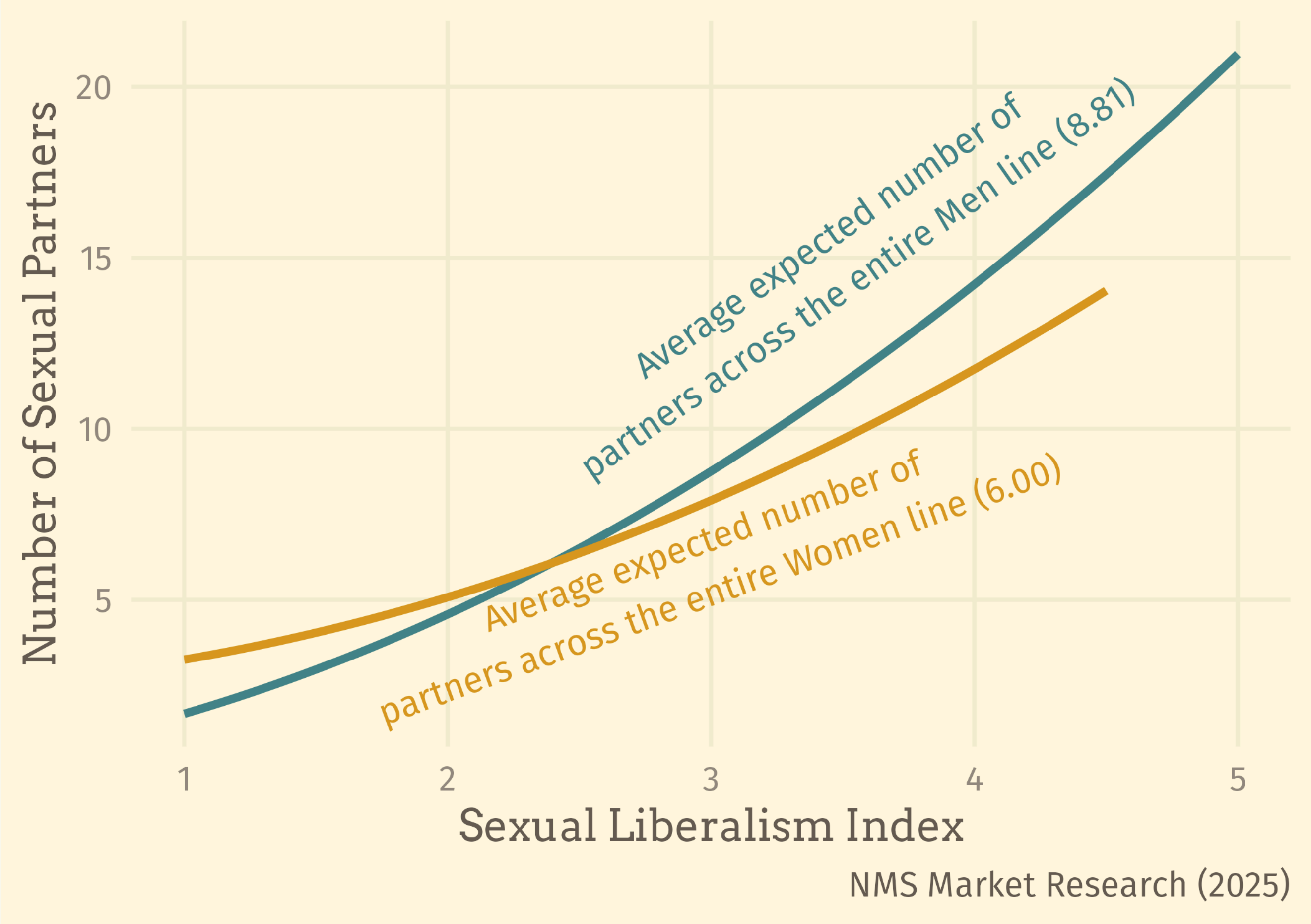
gender	Mean
Men	8.81
Women	6



Average Expected Number of Partners for Men and Women

What's happening:

1. Compute predicted value for every respondent.
2. Average predicted values separately for Men and Women.



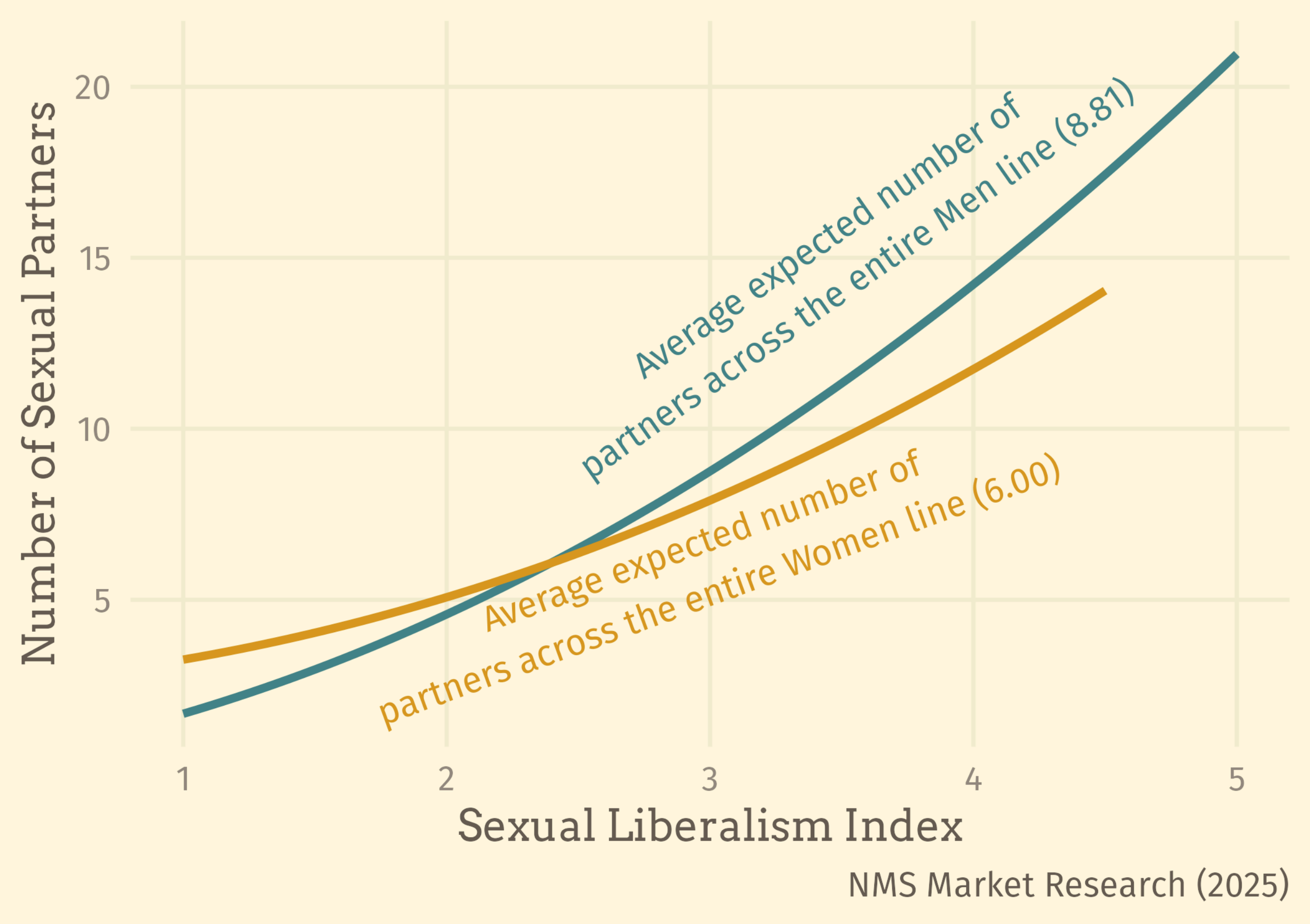
Average Expected Number of Partners for Men and Women

Advantage:

- Takes into account nonlinear relationships and interactions.

Disadvantage:

- Harder to compute.



Questions?

Personal recommendations

By default, use the average expected values (`estimate = "average"`). This is what most people intuitivelly mean by „*controlling for variables*“. This is sometimes called the **average marginal effect**.

If you have computational problems, fix the controlling variables at the average/typical value (`estimate = "typical"`). This is sometimes called **marginal effect at the mean**.

Predictions

Option 3: We can compute the expected number of partners, if everyone was a man and everyone was a women, while sexual liberalism remains unchanged.

This way, we can make predictions about **broader (hypothetical) population**.

```
estimate_means(model, by = "gender", estimate = "population")
```

Predictions

What's happening:

1. Compute the model
2. Set everyone's gender in your data into „Man".
3. Compute expected number of partners.
4. Repeat for category „Woman"

gender	Mean
Men	7.49
Women	7.09

Predictions

Not very common in typical (haha) sociology research.

But very comon in causal inference/policy evaluations:

- *What would the school performance be if all children got publicly funded school meals versus if none did?*

- *What would the death rate be if everyone was vaccinated vs if no one was?*

Questions?

InteRmezzo!

Contrasts

Contrasts

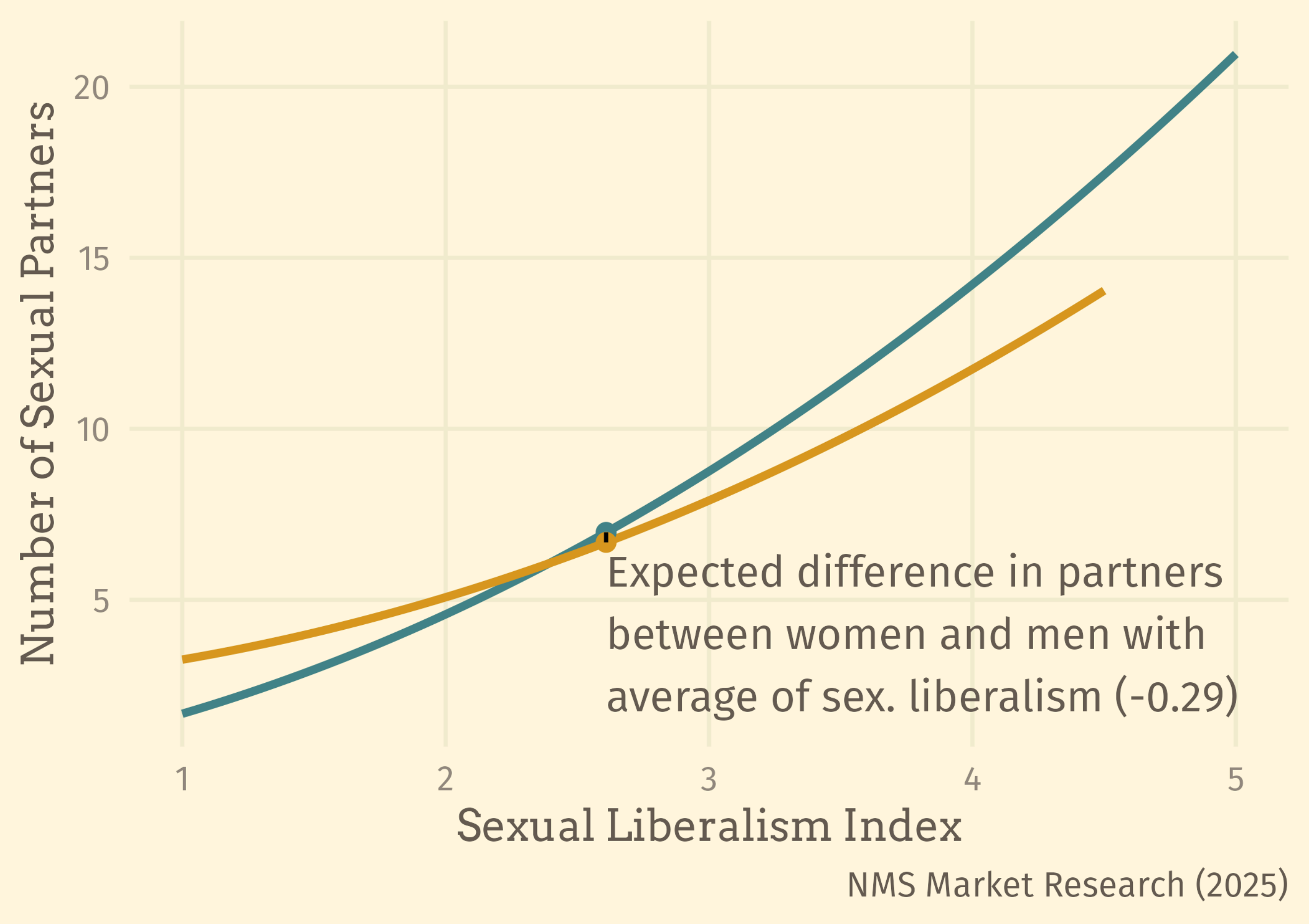
We know how to compute predictions, e.g. expected number of partners for men and women.

Based on these predictions, we can also compute differences between groups. These are called **contrasts**.

The whole `typical` vs `average` vs `population` things repeats again.

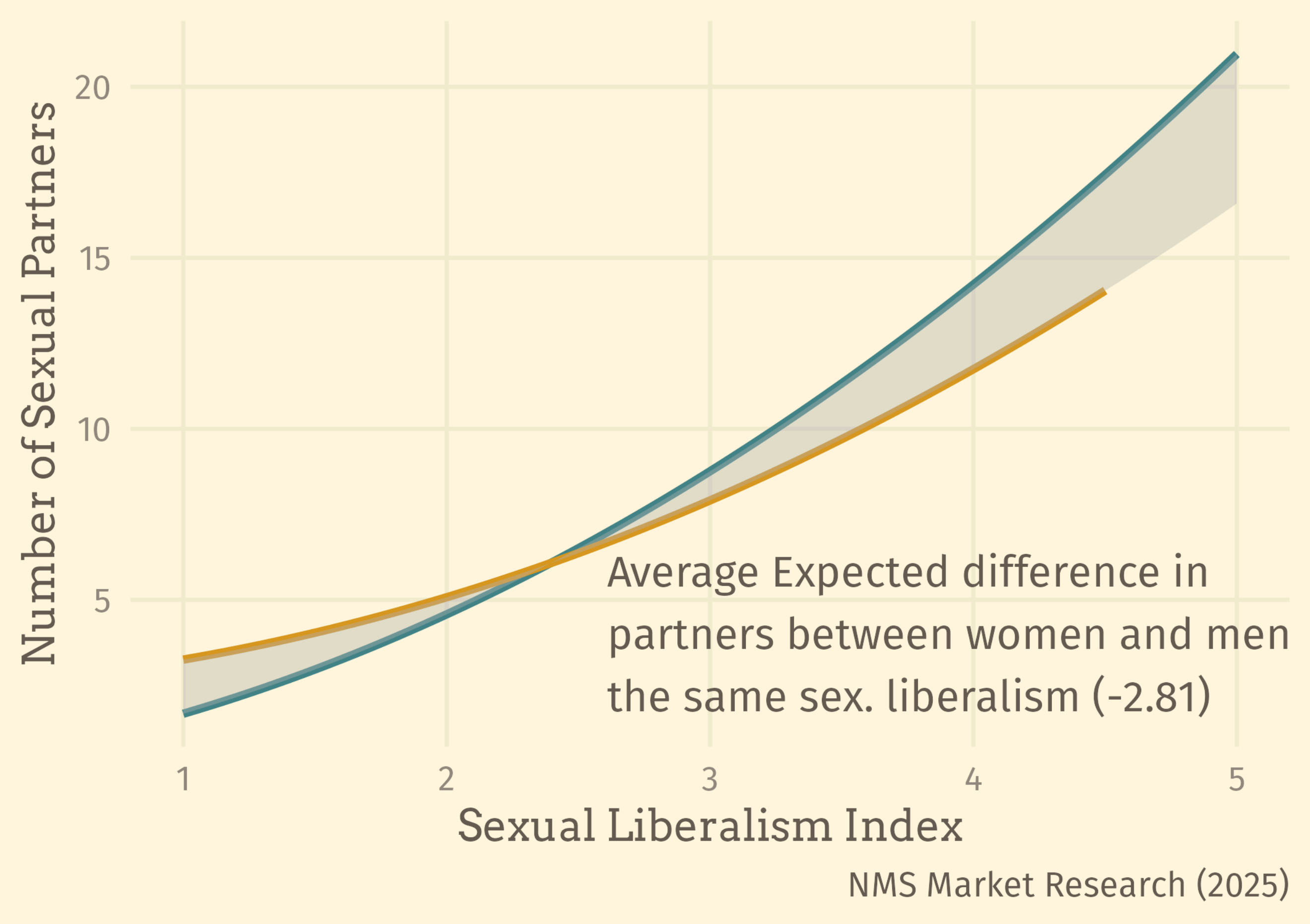
Contrasts - Typical

G1	G2	Diff
Women	Men	-0.29



Contrasts - Average

G1	G2	Diff
Women	Men	-2.8



Contrasts - Population

G1	G2	Diff
Women	Men	-0.4

The expected difference betwee population where everyone is a men and where one is a women.

(Assuming the distribution of sexual liberalism is same for both.)

Questions?

InteRmezzo!

Slopes

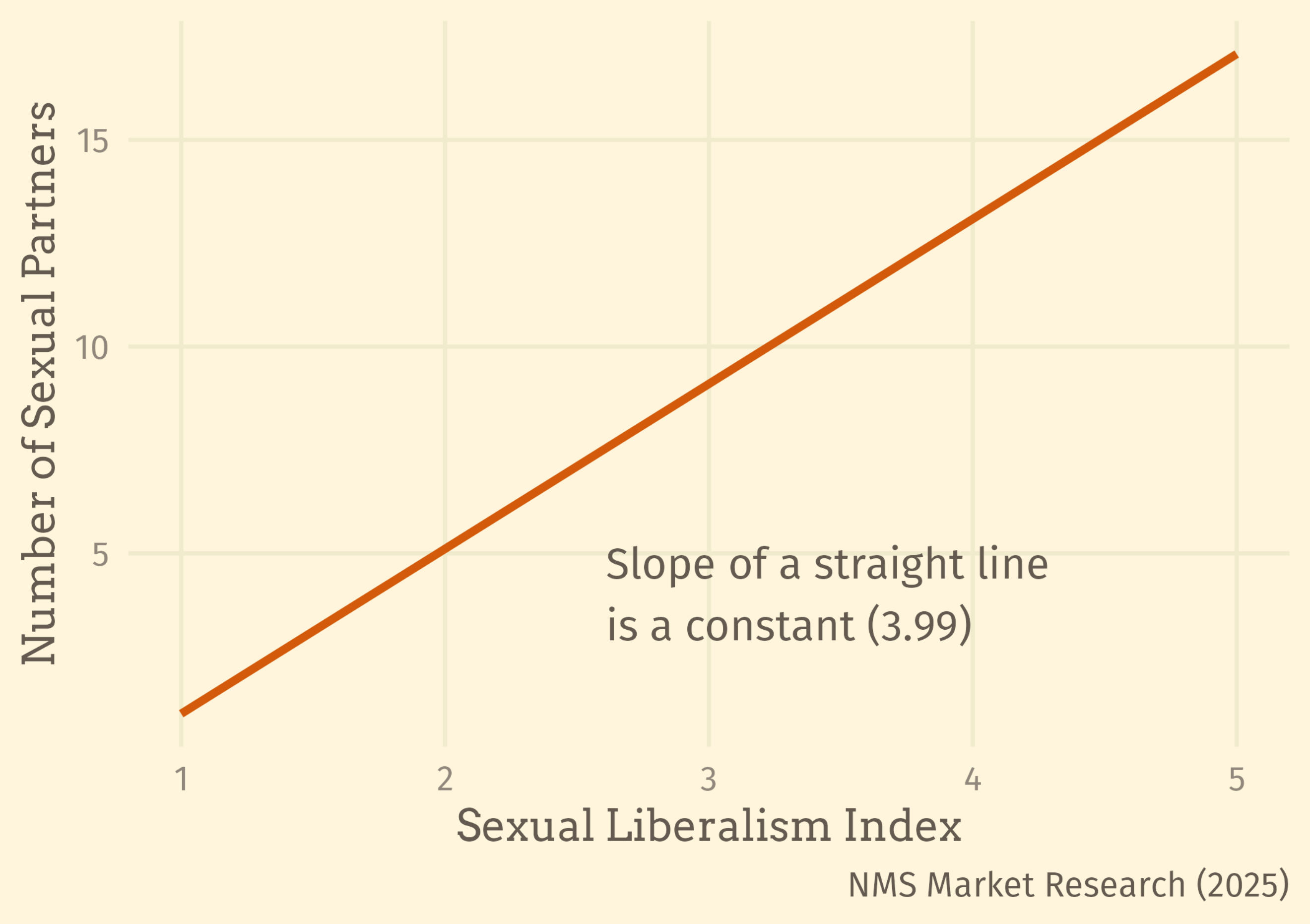
Slopes

We know how to compare groups, but how to quantify relationships between numerical variables?

We look ath its slope!

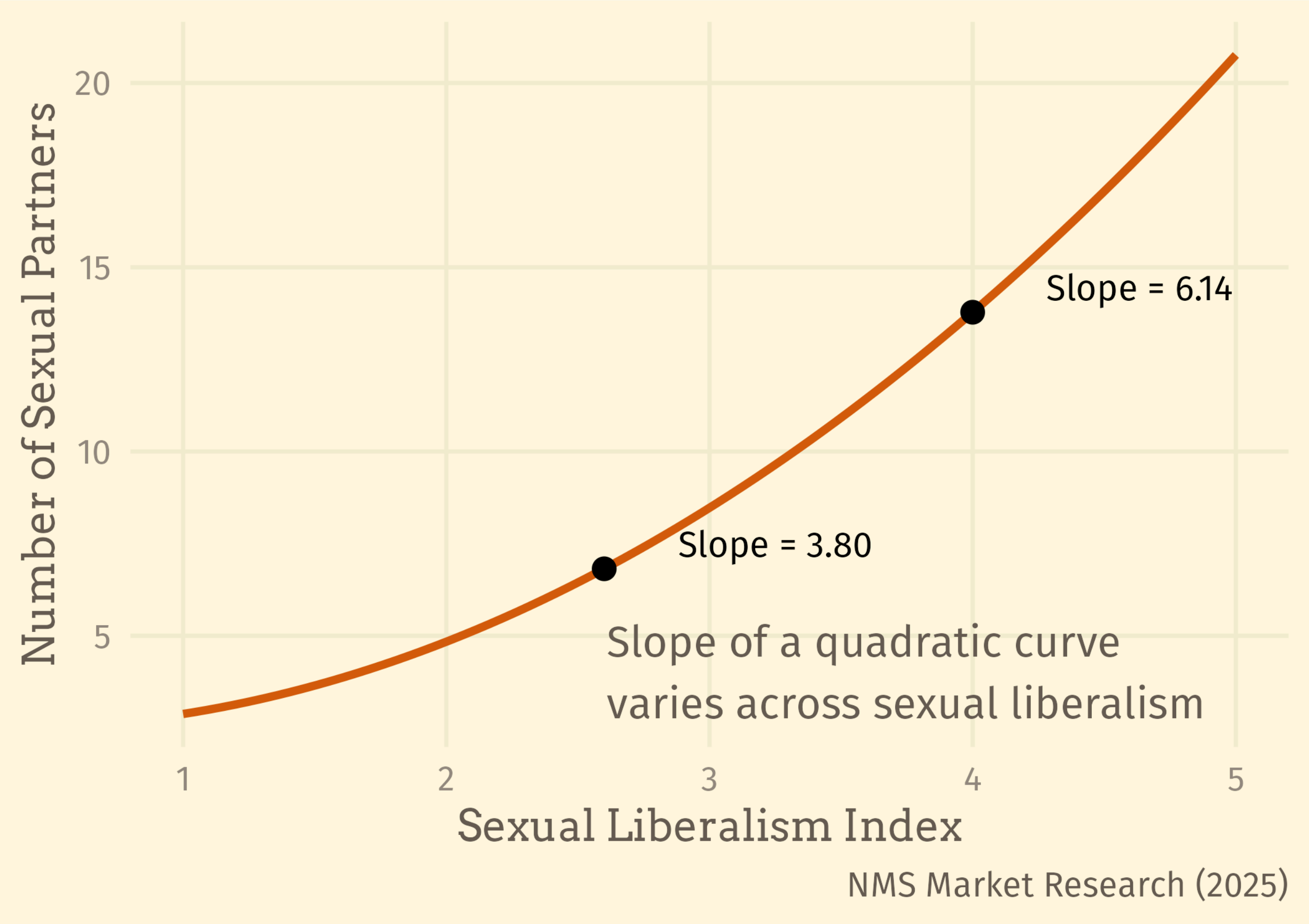
Slopes

Quantifying linear relationships is easy - the slope is the same everywhere.



Slopes

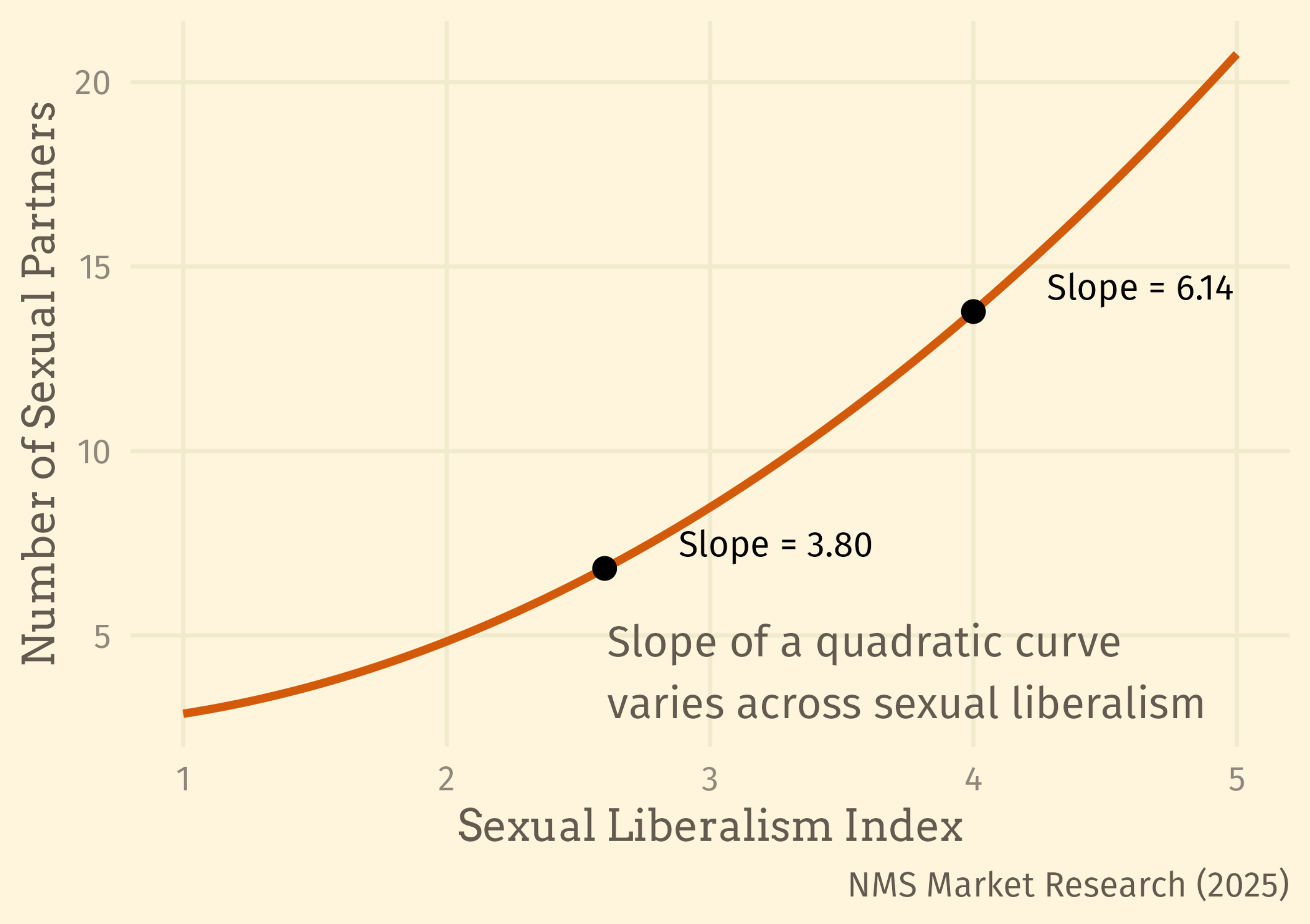
Quantifying non-
-linear relationships
is harder - the
strenght of
relationship
changes!



Slopes

The solution is to compute the **average slope**!

(usually for observed values of the numerical predictor)



Slopes

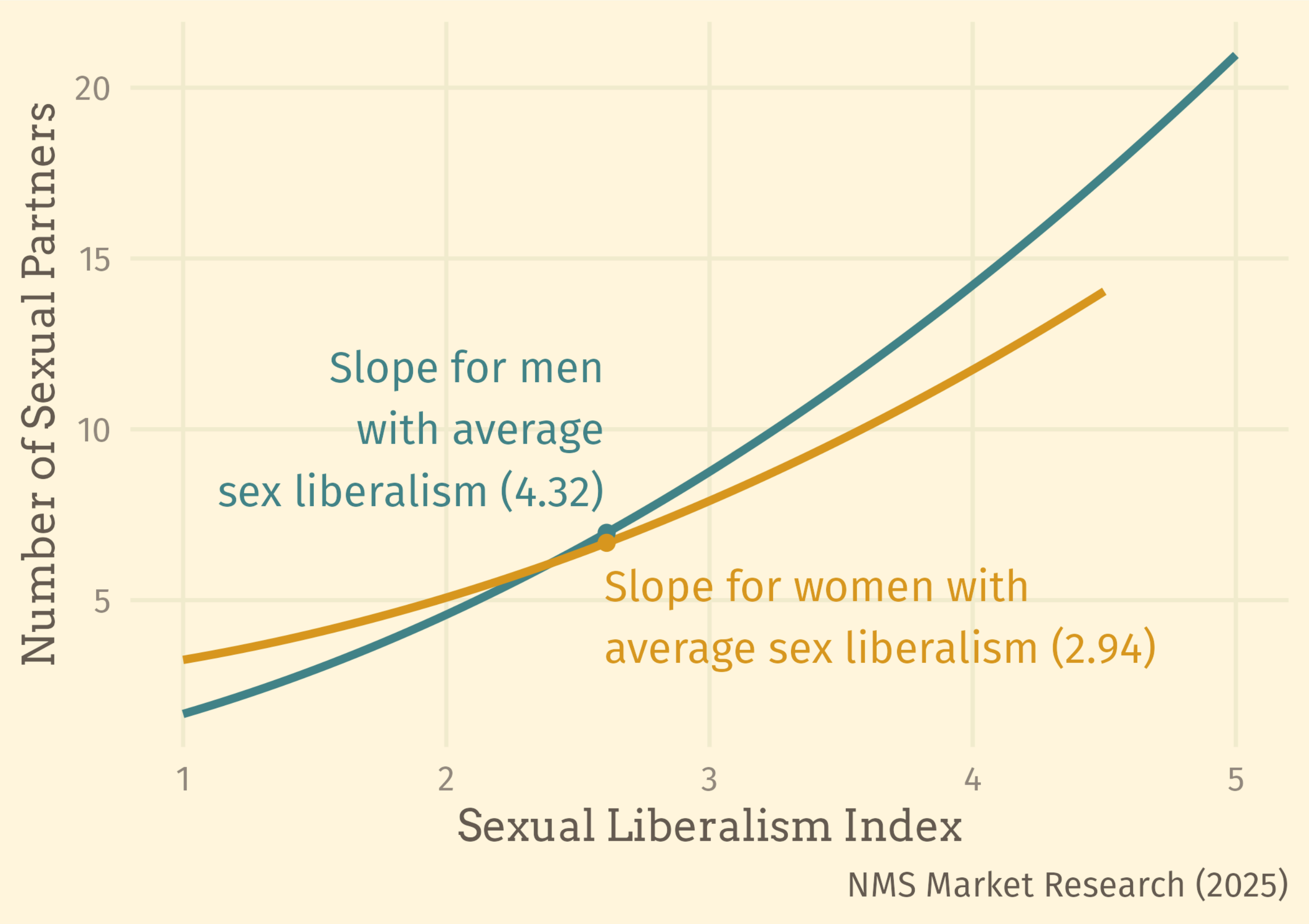
In R, we can do it using the `modelbased` package as usual. E.g.:

```
estimate_slope(model,
               slope = sex_libindex, estimate = "average")
```

The fuction doesn't allow for `estimate = "population"`, but `"typical"` and `"average"` work.

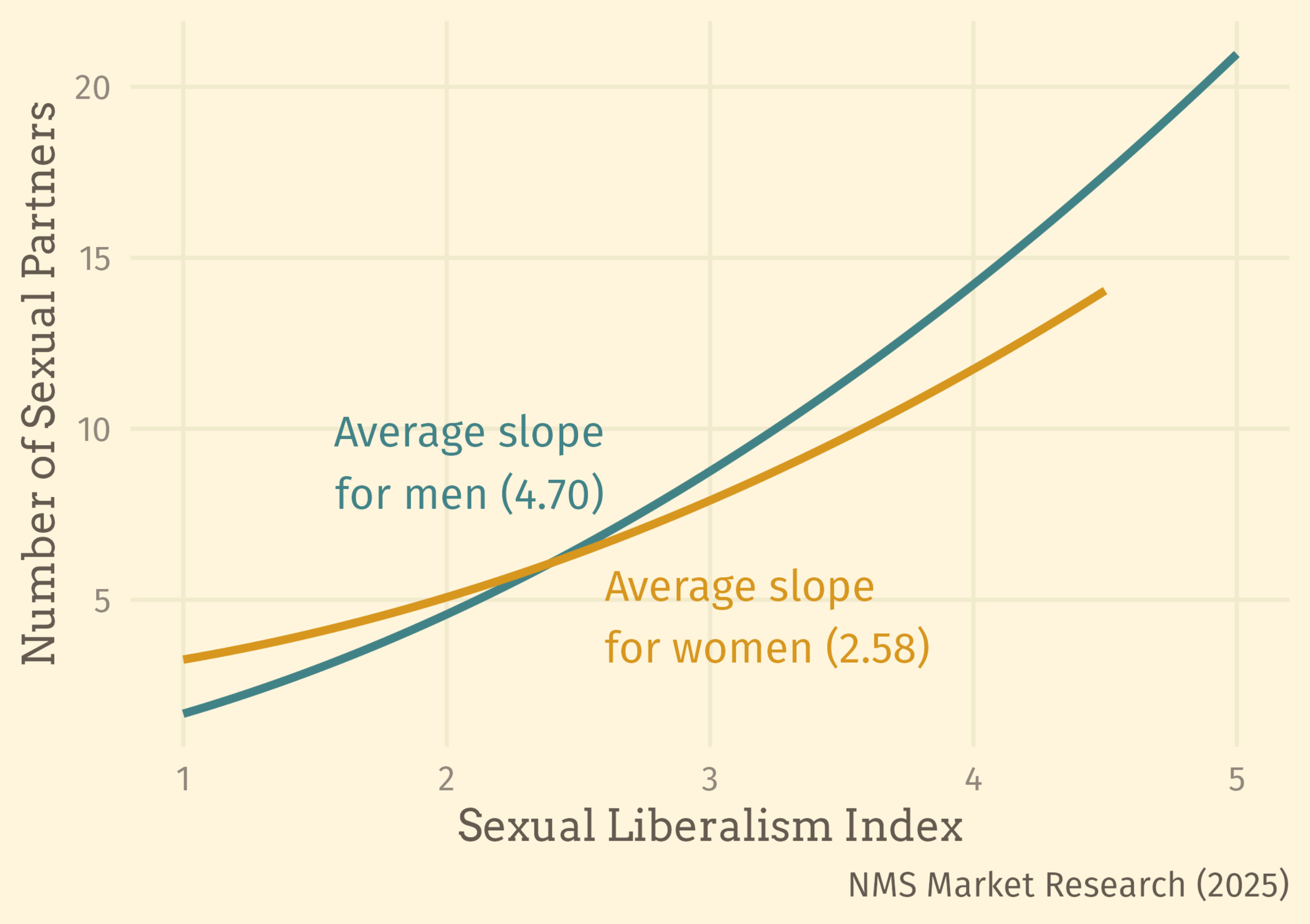
Slopes - Typical

gender	Slope
Men	4.3
Women	2.9



Slopes - Average

gender	Slope
Men	4.7
Women	2.6



Slopes

We can also compute the „total“ average slope using:

```
estimate_slopes(model, slope = "sexlib_index")
# Average Slope: 3.74
```

On average, one unit increase in sexual liberalism is associated with 3.74 higher number of sexual partners.

Notice ho 3.74 is the average of the average slopes for men and women (4.7 and 2.6)

Slopes - Bonus

Protip: We can compute contrasts for slopes!

This allows you to compare if a relationship between two variables has different strength in different subpopulations.

```
estimate_contrasts(model, contrast = "sexlib_index",
                   by = "gender", estimate = "average")
```

Slopes - Bonus

```
estimate_contrasts(model, contrast = "sexlib_index",
                   by = "gender", estimate = "average")
```

```
# Level1 | Level2 | Difference | 95% CI         | p
# ------------------------------------------------------------
# Women  | Men    |      -2.12 | [-3.51, -0.73] | 0.003
```

This is very underused research angle in sociology!

Questions?

InteRmezzo!