Multiple (hierarchical) regression

E0420

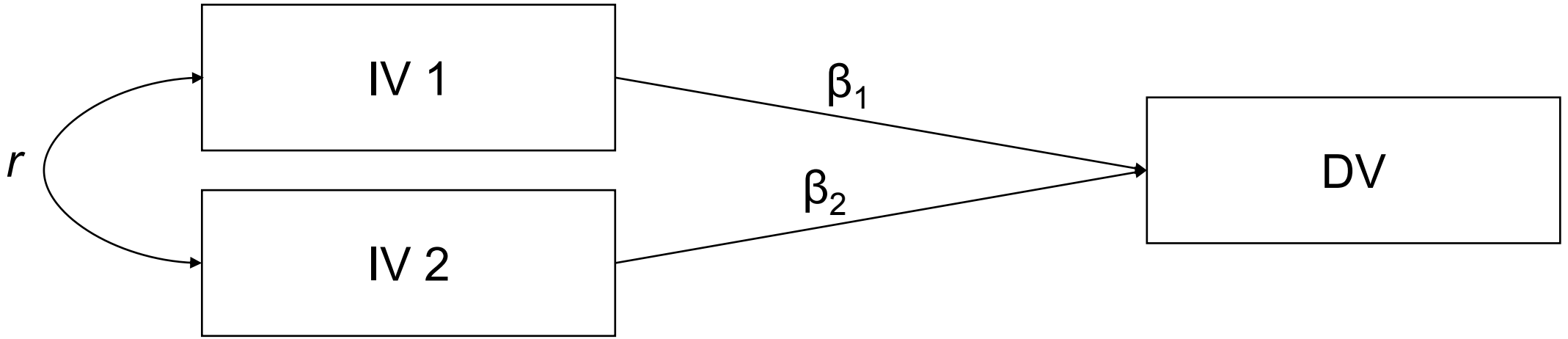
Week 8

Multiple (linear) regression

- Uses more than one explanatory/independent variable

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots + \beta_n X_n + \varepsilon$$

- The estimates now reflect the relationship accounting for the effect of other variables (conditional effect – partial regression coefficient)
- Multiple regression also accounts for the correlation between the IVs


IV 1
IV 2
r
β1
β2
DV

Example

$y = B_0 + 983x_1 + 122x_2 + e$

For any given level of x_2, when there is a one unit increase in x_1 then y will increase by 983

Assumptions

- Just like simple regression
  - Linear relationship
  - Normally distributed residuals
  - Independent residuals
  - Homoscedasticity

  AND
  - Absence of multicollinearity – independent variables should not be highly correlated with each other

Multicollinearity

- When IVs are too intercorrelated
  - *r* ~ 0.70 and more
- The model as a whole still “works”, but the estimates for individual predictors cannot be trusted
  - Large SEs (non-significant estimates), improbable Bs (too high, different sign)
- Diagnostics
  - Variance inflation factor (VIF) – how much the variance is inflated due to multicollinearity
    - Tolerance (1/VIF)

Hierarchical linear regression

- Enter the IVs in steps (models)
- Steps can include 1 or more IVs
- Order of entry – variable is entered after the variables that might be source of spurious relation have been entered
- Typically: control variables first, the variable of interest last

Example

H1: happiness will be related to number of pets

- Model 1:
  - Happiness = intercept + gender + age
- Model 2:
  - Happiness = interecept + gender + age + number of friends
- Model 3:
  - Happiness = interecept + gender + age + number of friends + number of pets

https://data.library.virginia.edu/hierarchical-linear-regression

R^2 change

- The total % of variability in DV explained by IVs
- We can assess the incremental R^2 of each Model (step)
  - The F-test is used to statistically assess the change in model R^2 after adding more variables
- Sometimes the IV estimate (B) is significant but the R^2 is not
  - The effect of the IV might be spurious

Interactions

- Used for testing moderation
- When two variables affect the DV beyond their additive effect

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

- The interaction term should be entered in a separate step

Adjusted R^2

- R^2 either increases or remains the same (if it decreases, you have a problem!)
- Selecting a model based on R^2 would then always prefer the model with more predictors
- Adjusted R^2 takes into account the number of predictors entered and increases only when the variables explain meaningful amount of variance

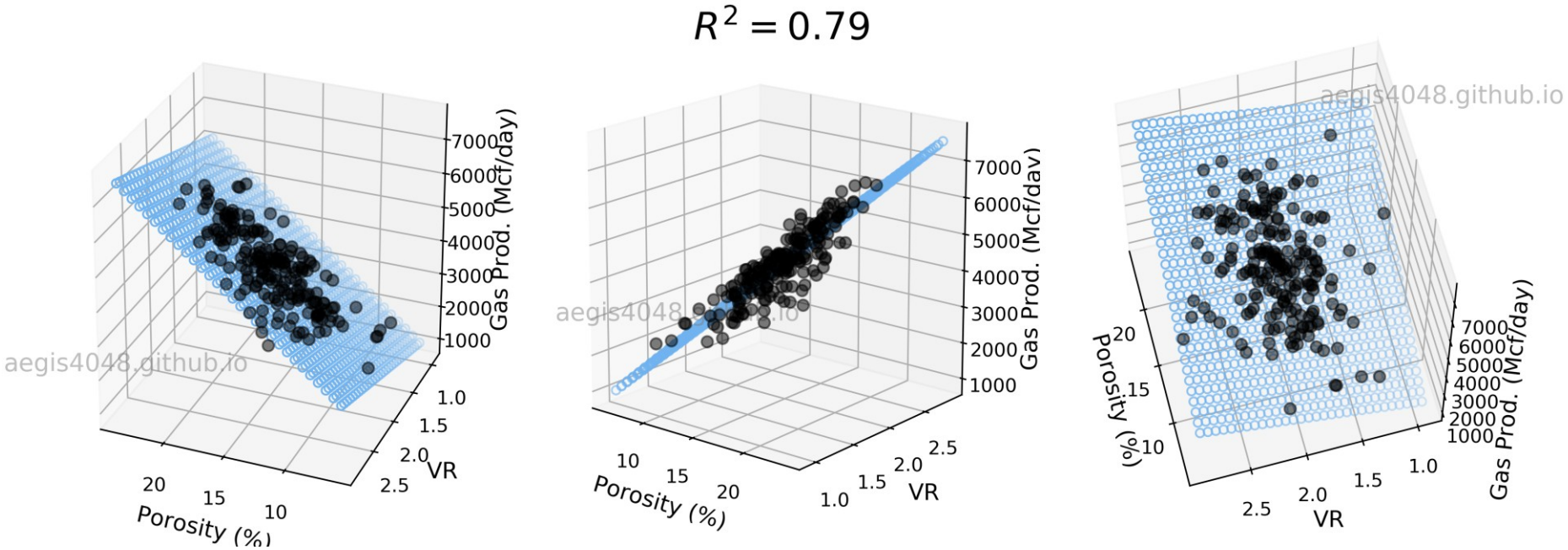
Stepwise regression

- Not to be confused with hierarchical linear regression
- The order of variables entered is selected by the computer to maximize the R^2 at each step
  - Forward selection, backwards elimination, bidirectional
- Atheoretical, might be too dependent on the specific sample (cannot be reproduced), capitalizing on chance

Plotting

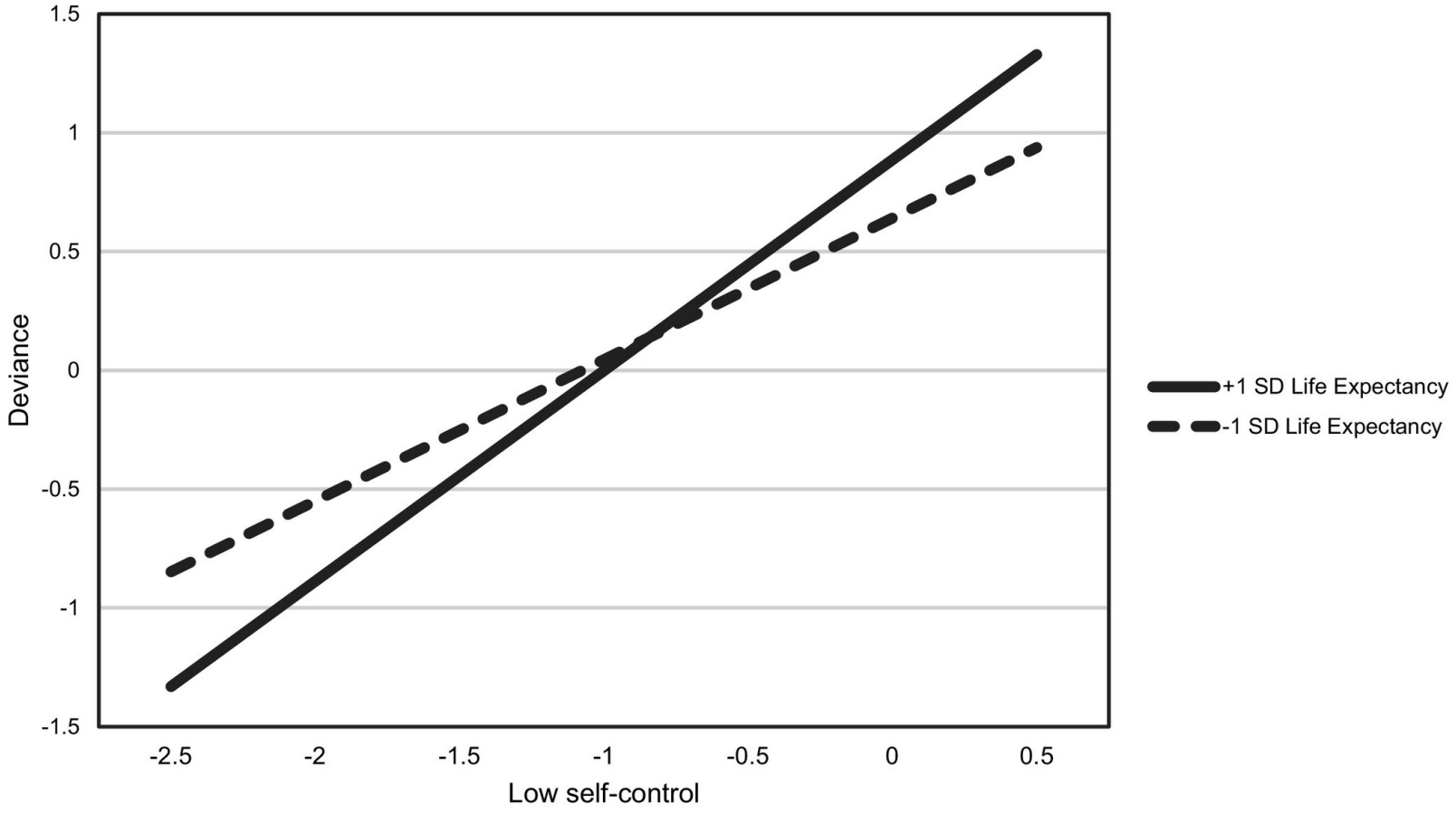
- Cannot be easily plotted due to the fact that there can be only one X variable
- Plotting can use grouping variable (e.g., sex)
  - For interactions
- Plot the focal variable (with partial coefficient)

Two IVs



https://aegis4048.github.io/mutiple_linear_regression_and_visualization_in_python

Interaction



Vazsonyi, A. T., Ksinan, A. J., & Javakhishvili, M. (2021). Problems of cross-cultural criminology no more! Testing two central tenets of Self-Control Theory across 28 nations. *Journal of Criminal Justice*, 101827.

Write-up

- The results of multiple regression analysis showed that both age (β = .12, $p<.001$) and extraversion (β = .56, $p<.001$) were significant positive predictors of aggressive tendencies. The full model explained 35.8% of the variance ($F(2,55)=5.56$, $p<.01$) .