

## PRACTICAL Week 9 – Binary Logistic Regression

For this session, we will use a subset of the Health Survey for England.

We are interested in identifying the risk factors of hypertension. Hypertension is characterized by elevated values of systolic or diastolic blood pressure. The cut-off points are 140mm Hg and 90 mm Hg, respectively.

For this analysis we will focus only on the Systolic Blood Pressure. In our dataset it can be found as the variable “omsysval”.

1. Open the dataset Health\_Survey\_England.sav
2. Paste all calculations and answers to the syntax file and save it regularly.
3. **Create dichotomous variable.**
  - a. Recode omsysval into new binary variables indicating whether the participant has hypertension or not (hint: if omsysval>140 mm Hg). Before recoding, inspect the variable omsysval for outliers and decide what to do with them.
    - i. What is the prevalence of hypertension?
4. **Test whether age predicts hypertension.**
  - a. Model fit
    - i. Does the model fit the data better than an empty model?
    - ii. What are the *Cox and Snell R-square* and *Nagelkerke R-square*?
    - iii. What is the overall percentage of cases that were correctly predicted by our model?
  - b. Results
    - i. What is the effect of age on hypertension?
      1. What is the magnitude and significance of the slope?
      2. What is the magnitude and significance of the OR?
    - ii. Write-up the results.
5. **Test whether sex predicts hypertension, after adjusting for the effect of age.**
  - a. Model fit
    - i. Does the model fit the data better than an empty model?
    - ii. What are the *Cox and Snell R-square* and *Nagelkerke R-square*?
    - iii. How did these values change now that we added another predictor?
    - iv. What is the overall percentage of cases that were correctly predicted by our model?
    - v. Did the accuracy rate change after adding another predictor?
  - b. Results
    - i. What is the effect of sex on hypertension?

1. What is the magnitude and significance of the slope?
    2. What is the magnitude and significance of the OR?
  - ii. Write-up the results.
6. **Identify which of these common risk factors is associated with hypertension:** age, sex, bmi, cholesterol level, ethnicity, physical activity, smoking status and income.
- a. Build the model
    - i. What can you say about the model fit?
    - ii. What about its accuracy?
  - b. Results
    - i. Which of these predictors is associated with higher risk of elevated SBP?
    - ii. Assuming that I am only interested in assessing the effect of income on hypertension, how would you write up the result?

**Submit your practical:**

- Save your syntax file to Homework Vaults