The University of Hong Kong
School of Public Health

CMED 6020 Advanced Statistical Methods I
TUTORIAL 2 – Applied regression

Influenza vaccine effectiveness

An observational study investigated the effect of influenza vaccination on influenza infection in the general population. To estimate the vaccine effect, potential confounding effects and effect modification will be considered. Data are available from 2000 adults aged 20 to 60y (saved in 'fluvaccine.csv'), with the following variables:

Variable	Data label	Remark
flu	Influenza infection during the flu peak season	0: No; 1: Yes
age	Age in years	
male	Male sex	0: Female; 1: Male
shealth	Self-rated health rating	0-10 (good health)
smoking	Current smoker	0: No; 1: Yes
bmi	Body mass index (kg/m^2)	
vac	Pre-seasonal influenza vaccination	0: No; 1: Yes
abT	Pre-seasonal antibody titer (but after vaccination)	by 2-fold serial dilution

1. Read the dataset into R. Perform some exploratory analyses of the data (especially related to the main study objective).

2. Fit a regression model to estimate the crude effect of flu vaccination. Please also label the table.

Variable	***Odds ratio***	***95% CI****
Flu vaccination	0.68	0.47 – 1.00

*CI: confidence interval

3. Based on the literature, it's well known that self-reported health status is a confounder for the effect of flu vaccine on infection. Is the data consistent with this finding?

4. Since self-reported health status is a confounder, we could perform stratified analysis to estimate the vaccine effect at each self-reported health level. Stratify the variable *shealth* into 2 or 3 levels and estimate the strata-specific vaccine effect.

5. Assess if there are also other potential confounders for estimating the vaccination effect.

6. It was recognized that smoking and age may modify the vaccine effect. Assess if this is the case in the study.

7. Based on the model with an interaction effect, calculate the vaccine effect and the corresponding 95% CI for each subgroup of the effect modifier.

   The variance of $\beta_1 + \beta_2$ can be obtained by $var(\beta_1) + var(\beta_2) + 2cov(\beta_1, \beta_2)$.

8. When testing the interaction effect between vaccine and age, also assess if there is collinearity problem. If so, is there a way to handle it?

9. Subjects who received flu vaccine will usually have an increase in antibody titer. It is hypothesized that the protection of flu largely depends on antibody production. Assess if this is the case for the study. [Note: first transform the antibody titer using a log function with base 2]

10. Suppose we purely rely on statistical methods to select models. Using AIC to compare between models, what would be the selected model? [Hint: use stepAIC function in package "MASS"]

Suppose we select the final model based on both previous knowledge and statistical significance. We are interested to assess the total effect of vaccination on flu infection.

11. Produce residual plots for the final logistic regression model. Do you identify any violation of assumptions?

12. [for reference only] Calculate the area under the receiver operating curve (AUROC, or c-statistics) for the final model.

13. Determine your final model, summarize your results in a table and draw your conclusion on the study.

Variable	***Adjusted odds ratio***	***95% CI****
Flu vaccination	0.36	0.23-0.58
Self-reported health rating	0.71	0.66-0.76
Current smoker	1.66	1.02-2.70
Flu vaccination × current smoker	2.64	1.02-6.83

*CI: confidence interval

Based on a logistic regression accounting for confounders self-reported health rating, flu vaccination was associated with a 64% (95% CI 42 - 77%) reduction in risk of flu infection for non-smokers. The estimated reduction of risk for current smokers is 4% (95% CI -122 - 58%). We found that current smoker is an important effect modifier for vaccine, while age may not modify the vaccine effect. The final model has a satisfactory fit to the data and a discriminative power of 72%.