` **The University of Hong Kong**
School of Public Health

CMED 6040 Advanced Statistical Methods II
TUTORIAL 1

(Extension of Assignment 1)

Background For public health surveillance purposes, a study has been carried out to assess the prevalence of obesity (defined as BMI>25) among male HKU undergraduates. The study took a random sample of 185 undergraduate males in May 2023, and assessed each of their heights and weights. In total, 13 students were found to be obese.

The log-likelihood for the prevalence θ is given by $\log l(\theta) = 13log\theta + 172\log(1-\theta)$. The MLE estimates for θ is 13/185 = 0.07. By using a normal approximation, the 95% confidence interval is (0.03, 0.11).

(a) Using the likelihood ratio method, obtain a 95% confidence interval for θ.

$$2\left|\log l(\hat{\theta}) - \log l(\theta)\right| \sim \chi_1^2$$

(b) Using the bootstrap method, obtain a 95% confidence interval for θ.

(c) Suppose the study was also carried out in 7 other tertiary institutions. Their results are summarized below:

Institutions	#1	#2	#3	#4	#5	#6	#7
No. obese	18	21	10	11	10	17	12
No. male undergraduate	161	272	154	85	101	221	150

Estimate the overall prevalence using maximum likelihood method.

(d) Suppose it was hypothesized that institutions which were able to recruit more participants (e.g. n > 200) may have a different prevalence of obesity. Estimate the relative difference using the maximum likelihood method.

You may assume that the obesity prevalence is θ for schools with fewer participants, and $k\theta$ for schools with more participants.

(e) When the sample size is large, according to maximum likelihood theory

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta)),$$

where $I^{-1}(\theta)$ is the information matrix

$$I(\theta) = -E[\frac{\partial^2 logL(\theta)}{\partial\theta\partial\theta'}]$$

$\frac{\partial^2 logL(\theta)}{\partial\theta\partial\theta'}$ is the second derivative of the log-likelihood, also named Hessian, which can be obtained by setting "hessian=T" in the optim function in R.

Compute the standard error for the estimated prevalence of obesity <u>in the first tertiary institution</u> and calculate its 95% confidence interval.

(f) Referring to (d), compute the 95% confidence interval for k and test the hypothesis H_0: k=1.

[Hint: use solve() to compute the inverse of a matrix]

(g) Perform a likelihood ratio test for (f).