

# APH103 Survey Sampling Coursework - Group T

Kunyang He, Jiayi Li, Baiding Chen, Yuang Li

April 2023

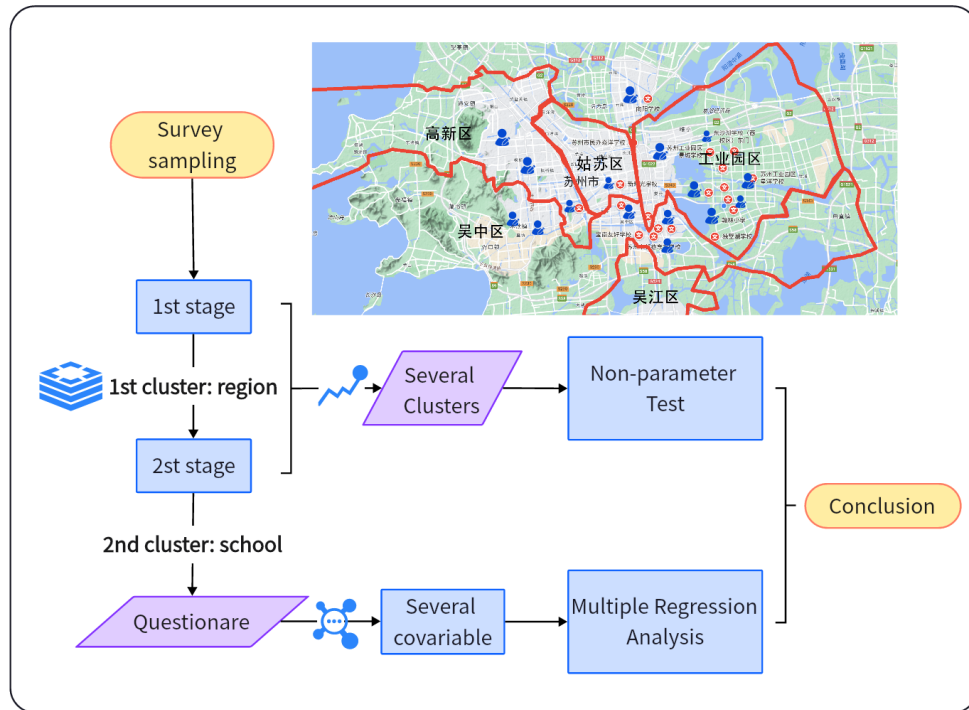


Figure 1: Schematics of the proposed survey sampling strategy

# Contents

<b>1</b>	<b>Background information</b>	<b>3</b>
1.1	HPV background . . . . .	3
1.2	Population Characteristics . . . . .	3
<b>2</b>	<b>Sampling Design [6]</b>	<b>4</b>
2.1	Sampling Methods . . . . .	4
2.1.1	Simple Random Sampling . . . . .	4
2.1.2	Stratified Sampling . . . . .	4
2.1.3	Systematic Sampling . . . . .	5
2.1.4	Cluster Sampling . . . . .	5
2.1.5	Discussion and Decision for this Case Study . . . . .	6
2.2	Two-stage Sampling with sampling choice specification . . . . .	6
<b>3</b>	<b>Questionnaire</b>	<b>6</b>
<b>4</b>	<b>SAP procedure</b>	<b>8</b>
4.1	Simulating the data . . . . .	8
4.1.1	Covariate indicators . . . . .	9
4.2	Kruskal-Wallis Test : test for cluster sampling . . . . .	10
4.3	Logistic regression analysis . . . . .	11
4.3.1	Data analysis . . . . .	11
<b>5</b>	<b>Summary</b>	<b>12</b>
<b>A</b>	<b>Kruskal-Wallis Test</b>	<b>13</b>
<b>B</b>	<b>Logistic regression analysis</b>	<b>14</b>
B.1	Likelihood function . . . . .	14
B.2	MLE method . . . . .	14

# 1 Background information

## 1.1 HPV background

According to the International Agency for Research on Cancer (IARC), there were approximately 604,000 new cases of cervical cancer and 342,000 deaths from cervical cancer worldwide in 2020, making cervical cancer the fourth leading cause of cancer death in women [5]. Infection with the Human Papillomavirus (HPV) can increase the incidence of cervical cancer. Studies have shown that HPV vaccines (types 16 and 18) have strong protective efficacy in preventing certain carcinogenic HPV-related cervical diseases, and can prevent more than 70% of cervical cancers. Awareness of the HPV vaccine is generally low, with 29.8% of women of childbearing age have heard of cervical cancer and 17.7% of women have heard of the HPV vaccine; Among them, 82.7% of college students have heard about cervical cancer and 31.1% about HPV vaccine [1]. Sexual transmission is the main route of HPV virus transmission. The study shows that 7.4% of college students have had sex, among which only 39.3% use condoms for the first sex, and only 40.9% use condoms every time. Therefore, college students are one of the high-risk groups for HPV infection.

To address this issue, understanding the factors that influence HPV vaccine uptake is critical. Previous studies have identified several societal and personal factors. By conducting a survey to assess the HPV vaccination rate and identify factors associated with personal (Age, Domain of expertise, Parental support, and Awareness) and societal (Finance, Vaccine storage, Promotion) can develop targeted interventions to improve HPV vaccine coverage in Suzhou and related regions.

## 1.2 Population Characteristics

Societal factors (Finance, Vaccine storage, Promotion) and personal factors (Age, Domain of expertise, Parental support, and Awareness) are collaboratively considered as covariates having an influence on the overall vaccination rate in the place of interest. More specifically, Societal factors can be directly reflected by the choice of subsections in each region, whereas personal factors can be stratified by various age groups respective to the survey setting.

Based on the population characteristics, we shall implement a two-stage sampling procedure with social factors (reflected by subsections of the region) considered in the first stage and the choice of age group considered in the second stage.

## 2 Sampling Design [6]

### 2.1 Sampling Methods

#### 2.1.1 Simple Random Sampling

**Methodology:** Simple Random Sampling is a common sampling method in statistics, where each member of the population has an equal chance of being selected for the sample. The procedure involves the following steps: Define the population of interest, which is the population of students in the Suzhou region; Determine the desired sample size, which is the number of students in different age groups and subregions. Assign a number to each member of the population (sampling frame); Use a random number generator or a table of random numbers to select the sample.

**Discussion:** It is easy to understand and produce samples that are likely to be representative of the population if the sample size is large enough, requiring minimal knowledge of the population, which guarantees that each member of the population has an equal chance of being selected to reduce the risk of bias. Nevertheless, it may not be appropriate for small populations or populations with a highly skewed distribution, leading to the exclusion of some population members. Additionally, it may be time-consuming and expensive if the population is large or geographically dispersed (if the regions and age groups of interest are distributed across Suzhou).

#### 2.1.2 Stratified Sampling

**Methodology:** Stratified Sampling is a sampling method that divides a population into homogeneous subgroups, called strata, based on some characteristic of interest, such as age groups or subsections of regions in our sampling context. Samples are then taken from each stratum using a random sampling technique. The procedure involves the following steps in our workflow: Define the population of interest and the characteristics that will be used to form the strata; Divide the population into mutually exclusive strata based on the selected characteristic; Determine the sample size for each stratum based on its proportion to the total population or the desired precision of the estimate; Select a random sample from each stratum.

**Discussion:** Stratified Sampling ensures that each stratum is represented in the sample, which allows for more precise estimates of population characteristics. Stratified sampling can be used to over-sample minority or rare groups, which may be of particular interest, and improve the accuracy and efficiency of the sampling process by reducing the variability of the estimate within each stratum. Stratified sampling requires prior knowledge of the population characteristics, which may not always be available or accurate. It can be time-consuming and expensive to implement, especially if the population is large and complex. Nevertheless, the sampling error can be high if the strata are not well

defined or the sample size for each stratum is too small (i.e. if samples available from specific age groups or sub-regions are rare compared to the remaining units in our sampling context)

### 2.1.3 Systematic Sampling

**Methodology:** Systematic sampling is a probability sampling method that involves selecting individuals from a population at regular intervals. The mathematical formulation for systematic sampling is as follows:

Let  $N$  be the size of the population of interest. Let  $n$  be the desired sample size. The sampling interval ( $k$ ) is calculated by dividing the population size by the sample size, such that  $k = N/n$ . The first individual to be selected is chosen at random from the first  $k$  individuals in the population. Subsequent individuals are selected at regular intervals of  $k$ , until the desired sample size is reached. The probability of selecting any particular individual in the population is not equal, but depends on the position of the first selected individual. However, the sample is still considered to be representative of the population, assuming that the first selected individual is chosen randomly and that there is no systematic pattern in the selection process.

**Discussion:** It is less subject to selection errors, especially when a good sampling frame is not available. Additionally, it provides greater information per unit, which generally spreads more uniformly over the entire population and thus may provide more information. Nevertheless, if there is an order in the population that is not known to us, it has limitation in estimating variance (biased), with larger errors than stratified random sampling.

### 2.1.4 Cluster Sampling

**Methodology:** Cluster sampling is a probability sampling method that involves dividing a population into clusters, randomly selecting clusters, and then selecting all individuals within the selected clusters as the sample. The mathematical formulation for Cluster sampling is as follows: Let  $N$  be the size of the population of interest, and let  $n$  be the desired sample size. The population is first divided into clusters, which are groups of individuals that are similar to each other in some way (e.g., subsections of regions, age groups in our sampling context). A random sample of clusters is then selected, and all individuals within the selected clusters are included in the sample.

**Discussion:** It is cost-effective with the ability to provide a representative sample of large, geographically dispersed populations (diverse subsections of Suzhou); it is also useful when it is not possible or practical to sample individuals directly from the population. Nevertheless, there is potential for bias if the selected clusters are not representative of the population or if there is variability within clusters. Additionally, cluster sampling may require a larger sample size than other sampling methods to achieve the same level of precision.

### 2.1.5 Discussion and Decision for this Case Study

After meticulous investigations of the four classical sampling methods, namely Simple Random Sampling, Stratified Sampling, Systematic Sampling, and Cluster Sampling, we need to choose one from the four and perform a Two-stage sampling procedure. In our survey sampling context, different subsections (e.g. Suzhou Industrial Park, Wuzhong District, Huqiu District, Kunshan District, etc) of Suzhou can be seen as primary sampling units (PSUs), while diverse schools (Clusters of junior high schools, senior high schools, and universities) in each subsection can be seen as individuals. It is decent for us to conclude that the Cluster Sampling method is useful in both stages since age groups and subregions naturally contain clusters rather than strata. Conversely, Simple Random Sampling does not take clusters of age groups and subregions into consideration; Stratified Sampling is not useful because age groups and regions are not homogenous subgroups; Systematic Sampling is not applicable because age groups and subsections do not form regular intervals. In conclusion, We shall choose the Two-stage sampling method using Cluster Sampling in each stage as specifications.

## 2.2 Two-stage Sampling with sampling choice specification

Two-stage sampling is a probability sampling method that involves sampling in two stages or phases. The first stage involves selecting a subset of primary sampling units (PSUs) from the population of interest, and the second stage involves selecting a sample of individuals from within each selected PSU.

Intrinsically, subsections of regions and different age groups contain various clusters in the sample space. In our survey sampling context, different subsections (e.g. Suzhou Industrial Park, Wuzhong District, Huqiu District, Kunshan District, etc) of Suzhou can be seen as primary sampling units (PSUs), while diverse schools (Clusters of junior high schools, senior high schools, universities) in each subsection can be seen as individuals within PSUs. For the first stage, we selected a subset of districts which naturally reflects clusters of regions with differences in societal factors. Subsequently, we select a sample of schools from diverse types within the selected PSU.

## 3 Questionnaire

Question 1-5 collect the basic information on every coefficient. Each option (A to D) corresponds to a value (1 to 4) one by one. Scores on the multiple-choice questions indicate knowledge about the HPV vaccine. The high grade is, the better knowledge have. Question 6-8 collect some subject information which can help to analyze the possible reasons for high or low vaccination rate.

- 1 What's your age?

- A 12-15
- B 15-18
- C 18-22
- D 22+

2 What do you know about the HPV vaccine?

- A Not at all
- B Nearly don't know, only have heard about the name
- C A bit, I know it is good for my health
- D Very well, the HPV vaccine protects against viral infections and cervical cancer

3 How important do you think HPV is to one's physical health?

- A Not important at all, I don't need to vaccinate
- B Not very important, vaccinations cause me a lot of trouble
- C Somewhat important, I will go there if there is a convenient place for vaccination nearby
- D Very important, I'm even willing to pay more for vaccinations

4 Your parents' education level?

- A Junior high school
- B Technical secondary school or high school
- C Junior College or bachelor degree
- D Postgraduate or above

5 Estimated annual income of your family?

- A Below 10w
- B 10w-40w
- C 40w-90w
- D Above 90w

6 Have you made an appointment for the HPV vaccine?

- A Yes
- B No

**If you answered "no" to question 3,**

7\* [multiple choice]What are the reasons for not getting the HPV vaccine?

- A Social reasons (lack of medical resources: trouble with vaccination, lack of vaccines)

- B Subjective reasons (doubts about vaccines, parents don't understand)
- C Economic reasons (high price)
- D Postgraduate or above

**If you answered "yes" to question 3,**

7\* How many doses of the HPV vaccine have you received?

- A 0 stitches (in line)
- B 1 stitch
- C 2 stitches
- D 3 stitches

8\* Where did you get your knowledge of the HPV vaccine [multiple choice]?

- A Medical institution or school
- B Guardians
- C Friends
- D Television or Internet

## 4 SAP procedure

### 4.1 Simulating the data

```
#Creating data, for HPV categories, (1 for vaccination and 0 for
                                non-vaccination)

#Binomial
n=1200 #for simulating the data.
a = 1
b = 1
z = rnorm(n)
eta = a + b*z
formula <- y ~ 1 + z
prob = exp(eta)/(1 + exp(eta))
Ntrials = sample(1:10, size=n, replace=TRUE)
HPV = rbinom(n, size = Ntrials, prob = prob)

#For age struture, with the assumption, we construct the left -
                                skewness distribution.
nnorm <- function(mu = 0, sigma = 3, lambda = 0){
  function(x){
    x <- (x - mu)/sigma
    f <- 1/(sqrt(2*pi))*exp(-x^2/2)*pnorm(x*lambda)
    return(f)
  }
}

x <- seq(12,28,0.01)
n = length(x)
```



```

Lambda <- 3
Age <- data.frame(
  x = x,
  y = nnorm(mu = 19, lambda = 3)(x),
  z = rep(Lambda, each = n),
  z1 = as.factor(rep(Lambda, each = n))
)

Age <- ceiling(Age)

```

#### 4.1.1 Covariate indicators

We use a map  $\mathbf{X}_{ij} : \{A, B, C, D\} \rightarrow \{1, 2, 3, 4\}$  to build up a connection between indicators' value and Questionnaire MCQ.

$$X(\omega) := \begin{cases} 1 & \text{if } \omega = A \\ 2 & \text{if } \omega = B \\ 3 & \text{if } \omega = C \\ 4 & \text{if } \omega = D \end{cases}$$

Here, A,B,C,D is the questionnaire answer. During the procedures. We use a map to let these categories' data become the numeric data that can be used in logistic regression analysis.

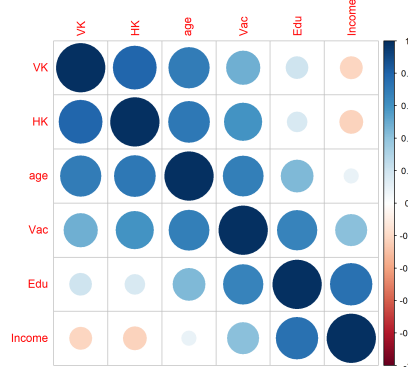


Figure 2: Correlation matrix of variables. When the absolute value of the correlation coefficient is close to 1. it means that the two corresponding variables are highly correlated with each other: then the absolute value of the correlation coefficient is close to 0, it means that the two corresponding variables appear no or unobservable correlation. (HK : HPV knowledge, VK : Vaccine knowledge, Edu : education, Vac : Vaccine (0 or 1))

Here is the dataset we create and use.

```
> head(APH103)
# A tibble: 6 × 7
  student Vacciane age Education Income `HPV knowledge` `Vaccine knowledge`
  <dbl>   <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1     1     1    19         3         2         1         3
2     2     1    20         2         3         2         2
3     3     1    22         4         4         4         3
4     4     0    18         1         1         3         4
5     5     1    19         2         4         2         1
6     6     0    25         4         2         1         2
```

Figure 3: The head of the data we collected

## 4.2 Kruskal-Wallis Test : test for cluster sampling

Here, KW test [2] is used to complete the nonparametric test of multi-group comparison. This method is chosen to test the equality of more than two (p) population probability distributions. The null hypothesis of our hypothesis test here is that all population medians are all equal. Each dot here represents the overall vaccination rate calculated by the school. Thirty schools in each district were sampled. The results show that there are significant differences between regions. We calculated the difference between the two regions using the pairtest and found that the difference between the two regions is also very significant.

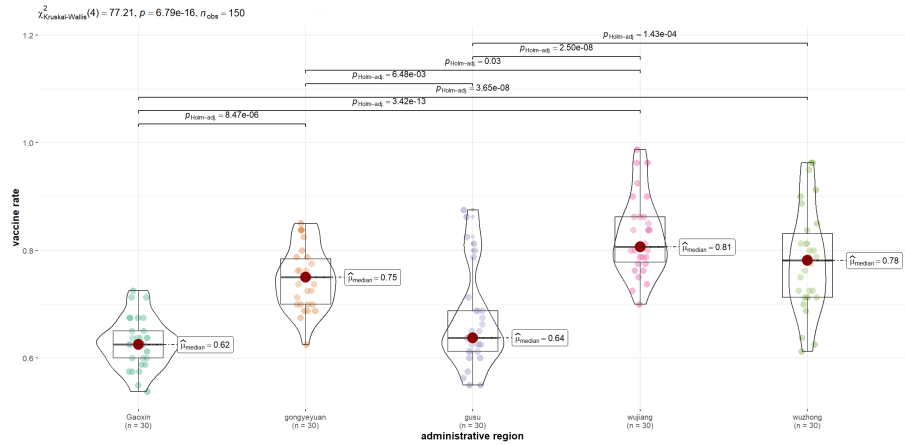


Figure 4: KW test statistical plot and analysis

- State Hypothesis
  - $H_0$  All of the population medians are all equal.
  - $H_1$  Not all of the population medians are equal.
- Rank the data values from smallest to largest. Resolve ties using the mean rank.

- Calculate the sum of the ranks for each group,  $R_1$ ,  $R_2$ , and  $R_3$ .
- Calculate H (the test statistic).

### 4.3 Logistic regression analysis

#### 4.3.1 Data analysis

Let  $\mathbf{Y} = \{Y_i, \dots, Y_n\}$  to be the observations we collected for vaccinated or not vaccinated HPV [3].  $\{\theta_1, \dots, \theta_5\}$  represents the covariates which the questionnaire collected.

Based on the following formula:

$$\begin{aligned}
 Y_i &\sim \text{Binom}(p_i) \\
 \text{logit}(p) &= \log\left(\frac{p}{1-p}\right), 0 \leq p \leq 1 \\
 \eta_i &= \log\left(\frac{p_i}{1-p_i}\right) = \theta_0 + \theta_1 X_{i1} + \dots + \theta_k X_{ik} \\
 OR &= e^\theta
 \end{aligned}$$

Our statistical model will essentially look something like the following (k=6 as the number of covariates is 5) :

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{15} \\ 1 & X_{21} & X_{22} & \dots & X_{25} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n5} \end{bmatrix}_{n \times 6} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_5 \end{bmatrix}_{6 \times 1}$$

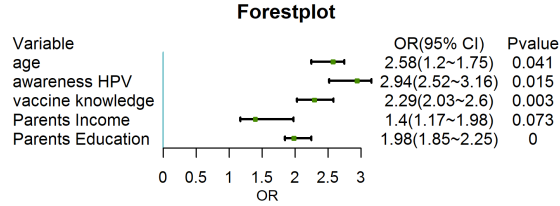


Figure 5: The HPV dataset result [4]

After the completion of logistic regression analysis, through the analysis of forest plot and OR, it can be observed that the OR values of the five groups of covariables are all significantly greater than one, indicating that with the increase of the assigned value of covariable, it has a significant positive impact on the improvement of vaccine coverage rate. In particular, it can be seen that hpv awareness has the highest OR value, indicating that vaccination awareness

is particularly important in the process of vaccination. In addition, the P value of the Parent income; The effect of 0.05 was not significant, and the variance was large, indicating that family income could not significantly affect the successful vaccination rate.

## 5 Summary

Throughout this survey sampling research report, intrinsic features of the population of interest are considered meticulously and a decent strategy is designed respectively. After meticulous investigations and discussions of the four classical sampling methods, namely Simple Random Sampling, Stratified Sampling, Systematic Sampling, and Cluster Sampling, We implement the two-stage sampling method and apply cluster sampling in each stage based on the subsections of the Suzhou region and diverse age groups accordingly.

Intrinsically, subsections of regions and different age groups contain various clusters in the sample space. In our survey sampling context, different subsections (e.g. Suzhou Industrial Park, Wuzhong District, Huqiu District, Kunshan District, etc) of Suzhou can be seen as primary sampling units (PSUs), while diverse schools (Clusters of junior high schools, senior high schools, and universities) in each subsection can be seen as individuals within PSUs. For the first stage, we selected a subset of districts which naturally reflects clusters of regions with differences in societal factors. Subsequently, a sample of schools and corresponding individuals within specific PSUs are selected. A questionnaire is then designed based on the sampling strategy that utilizes subsections and age groups as two clusters in the two-stage sampling.

Based on a preliminary analysis of population characteristics and the selection of a decent sampling strategy, we implement the proposed SAP procedure to derive insightful discoveries on the pattern of covariates and their impacts on the vaccination rate. In the SAP section, we first use a map from choices to the indicator's value numerically for subsequent statistical computation. To quantify the significance of differences between various clusters such as subsections of regions and age groups, the Kruskal-Wallis Test is utilized to complete the non-parametric test for multi-group comparison. For further investigations on the impacts of diverse covariates on vaccination rates, we interpret logistic regression results with a detailed analysis of the forest plot and OR. Specifically, the odds ratio (OR) values for the five groups of covariates exhibit a statistically significant deviation from unity, implying that an increase in the value of the assigned covariate has a positive effect on the improvement of the vaccination rate. Specifically, the covariate pertaining to awareness of HPV exhibits the highest OR value, thus highlighting the vital role of vaccination awareness in the vaccination process. However, the parent income covariate has a P value of 0.05 and does not show statistical significance. Furthermore, the covariate's large variance indicates that family income has no substantial impact on the

vaccination rate.

## References

- [1] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray. Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods. *International journal of cancer*, 144(8):1941–1953, 2019.
- [2] P. E. McKight and J. Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.
- [3] S. Menard. *Applied logistic regression analysis*. Number 106. Sage, 2002.
- [4] N. Muñoz, X. Castellsagué, A. B. de González, and L. Gissmann. Hpv in the etiology of human cancer. *Vaccine*, 24:S1–S10, 2006.
- [5] W. H. Organization. Human papillomavirus (hpv) and cervical cancer. Retrieved from. [https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer/](https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer/), 2021.
- [6] C. Soloff, D. Lawrence, and R. Johnstone. *Sample design*. Australian Institute of Family Studies Melbourne, 2005.

## Appendix

### A Kruskal-Wallis Test

Total variability of ranks (TV), Variability between group (BV), Variability within group (WV).

average rank =  $\frac{1+2+\dots+n}{n} = \frac{n+1}{2}$  The average of the total sum of squares of

$$\begin{aligned}
 \text{rank(aTV)} &= \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} \left( R_{ij} - \frac{n+1}{2} \right)^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^{n_i} \left( i - \frac{n+1}{2} \right)^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n_i} i^2 - \frac{n(n+1)^2}{4} \right) \\
 &= \frac{1}{n-1} \left( \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \right) = \frac{n(n+1)}{12} \\
 BV &= \sum_{i=1}^k n_i \left( \frac{R_{i.}}{n_i} - \frac{n+1}{2} \right)^2
 \end{aligned}$$

$$KW(H) = \frac{BV}{aTV} = \frac{12}{n(n+1)} \sum_{i=1}^k n_i \left( \frac{R_i}{n_i} - \frac{n+1}{2} \right)^2 = \frac{12}{n(n+1)} \sum_{i=1}^k R_i \frac{R_i}{n_i} - 3(n+1)$$

$R_i$  and  $n_i$  is the rank sum and sample size of the  $i$ th group, respectively

## B Logistic regression analysis

$$\theta^T \mathbf{X} = \sum_{i=1}^n \theta_i X_i = \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Logistic Regression is a classification algorithm that works by trying to learn a function that approximates  $P(Y | X)$ . It makes the central assumption that  $P(Y | X)$  can be approximated as a **sigmoid function** applied to a linear combination of input features. Mathematically, for a single training datapoint  $(\mathbf{x}, y)$  Logistic Regression assumes:

$$P(Y = 1 | \mathbf{X} = \mathbf{x}) = \sigma(z)$$

$$z = \theta_0 + \sum_{i=1}^m \theta_i x_i$$

This assumption is often written in the equivalent forms:

$$P(Y = 1 | \mathbf{X} = \mathbf{x}) = \sigma(\theta^T \mathbf{x})$$

$$P(Y = 0 | \mathbf{X} = \mathbf{x}) = 1 - \sigma(\theta^T \mathbf{x})$$

Using these equations for probability of  $Y | X$ , we can create an algorithm that select values of  $\theta$  that maximize that probability for all data. I am first going to state the log probability function and partial derivatives with respect to  $\theta$ .

### B.1 Likelihood function

$$L(\theta) = \prod_{i=1}^n P(Y = y^{(i)} | X = \mathbf{x}^{(i)})$$

$$= \prod_{i=1}^n \sigma(\theta^T \mathbf{x}^{(i)})^{y^{(i)}} \cdot [1 - \sigma(\theta^T \mathbf{x}^{(i)})]^{(1-y^{(i)})}$$

### B.2 MLE method

$$LL(\theta) = \sum_{i=1}^n \log(f(\mathbf{x}^{(i)}, y^{(i)} | \theta)) = \sum_{i=1}^n \log(f(\mathbf{x}^{(i)} | \theta) P(y^{(i)} | \mathbf{x}^{(i)}, \theta))$$

$$= \sum_{i=1}^n \log(f(\mathbf{x}^{(i)}) f(y^{(i)} | \mathbf{x}^{(i)}, \theta))$$

$$\theta_{MLE} = \operatorname{argmax} LL(\theta) = \operatorname{argmax} (\sum_{i=1}^n \log f(\mathbf{x}^{(i)}) + \log f(y^{(i)} | \mathbf{x}^{(i)}, \theta))$$

$$= \operatorname{argmax} (\sum_{i=1}^n \log f(y^{(i)} | \mathbf{x}^{(i)}, \theta))$$

```
#In R we fit a GLM in the same way as a linear model except using
#glm instead of lm and we must
#also specify the type of GLM to
#fit using the family argument.

df <- HPV_collected
fit.full <- glm(Vac ~ VK + HK + Edu + Income + age,
               data=df, family = binomial())

fit.result <- summary(fit.full)
forestplot(labeltext=as.matrix(fp[,1:3]),
           mean=fp$OR_mean,
           lower=fp$OR_1,
           upper=fp$OR_2,
           zero=0,
           boxsize=0.2,
           lineheight = unit(7, 'mm'),
           colgap=unit(2, 'mm'))
```

```

lwd.zero=1.5,
lwd.ci=2,
col=fpColors(box='#458B00',
              summary='#8B008B',
              lines = 'black',
              zero = '#7AC5CD'),
xlab="OR",
lwd.xaxis =1,
txt_gp = fpTxtGp(ticks = gpar(cex = 0.85),
                  xlab = gpar(cex = 0.8),
                  cex = 0.9),
lty.ci = "solid",
title = "Forestplot",
line.margin = 0.08)

```