1) a) $\text{logit}\ \pi = \alpha + \gamma_1 C_1 + \gamma_2 C_2 + \gamma_3 C_3 + \gamma_4 C_4 + \gamma_5 C_5 + \gamma_6 C_6 + \gamma_7 C_7 + \gamma_8 S$

$\Rightarrow C_1 = \begin{cases} 1, & \text{Beijing} \\ 0, & \text{ow} \end{cases}$, $C_2 = \begin{cases} 1, & \text{Harbin} \\ 0, & \text{ow} \end{cases}$, $C_3 = \begin{cases} 1, & \text{Nanjing} \\ 0, & \text{ow} \end{cases}$, $C_4 = \begin{cases} 1, & \text{Nanchang} \\ 0, & \text{ow} \end{cases}$, $C_5 = \begin{cases} 1, & \text{Shanghai} \\ 0, & \text{ow} \end{cases}$, $C_6 = \begin{cases} 1, & \text{Shenyang} \\ 0, & \text{ow} \end{cases}$, $C_7 = \begin{cases} 1, & \text{Taiyuan} \\ 0, & \text{ow} \end{cases}$, if $C_1 = C_2 = \cdots = C_7 = 0$, then "Zhengzhou".

$S = \begin{cases} 1, & \text{yes} \\ 0, & \text{no} \end{cases}$

Fitted Model : $\text{logit}\ \pi = -0.52 - 0.03 C_1 - 0.01 C_2 - 0.08 C_3 - 0.02 C_4 - 0.03 C_5 - 0.06 C_6 - 0.77 C_7 + 0.77 S$

$\Rightarrow$ $H_0 : \gamma_8 = 0$ v.s. $H_1 : \gamma_8 \neq 0$

$\Rightarrow$ The p-value of the estimated coefficient is significantly less than 0.05 so we may reject the null hypothesis.

Interpretation about "smoking": Keeping other city variables fixed, the estimated odds of getting lung cancers for smokers was about 2.18 times odds than for non-smokers.

b) H_0 : model fits v.s. H_1 : model does not fit

$\Rightarrow$ We have $\chi^2 = 5.2$ with the degrees of freedom of 7, so we don't have enough evidence to reject the null hypothesis.

c) The standardized residuals seem to lie within the 95% interval. There is no indication of the lack of fit by the residuals.

2) a) $\text{logit}\ \pi = \alpha + \gamma_1 W_1 + \gamma_2 W_2$, W_1 : weight, W_2 : width

$\Rightarrow$ $H_0 : \gamma_1 \neq 0$ or $\gamma_2 \neq 0$ v.s. $H_1 : \gamma_1 = \gamma_2 = 0$

The deviance is 32.9 with the degrees of freedom of 2 ; we fail to reject the null hypothesis.

b) The Wald statistics of the two variables indicate that they are both not affecting.

There may be an interaction effect or multicollinearity.

The model with the interaction term fits poorly so the interaction effect is not the source of additional explanatory power.

The Pearson Correlation of 0.88 indicates that they are highly correlated ; one of the two variables is not necessary for the final model.