Alphabetic order →

1. Let $C_1 = \begin{cases} 1 & \text{Beijing} \\ 0 & \text{o.w.} \end{cases}$ $C_2 = \begin{cases} 1 & \text{Harbin} \\ 0 & \text{o.w} \end{cases}$ ... $C_7 = \begin{cases} 1 & \text{Taiyuan} \\ 0 & \text{o.w} \end{cases}$

$C_1 = C_2 = \cdots = C_7 = 0$ for Zhengzhou.

(a) The fitted model is

$$\text{logit}\,\hat{\pi} = -0.5199 - 0.0288\,C_1 - 0.0106\,C_2 - 0.0837\,C_3 - 0.0230\,C_4 + 0.0268\,C_5 - 0.0565\,C_6 - 0.7745\,C_7 + 0.7771\,(\text{smoking})$$

From Wald test $\left(\frac{\hat{\beta}_{\text{smoking}}}{SE} = \frac{0.7771}{0.0468},\ p\text{-value} < 0.0001\right)$, we reject $H_0: \beta_{\text{smoking}} = 0$.

Controlling for city, estimated odds of lung cancer for smokers was $e^{0.7771} = 2.18$ times odds for nonsmokers

(b) Pearson $X^2 = 5.1999$ $df = 7$ $p\text{-value} = 0.635583$

So, we cannot reject H_0: model holds

Therefore, there is no lack of fit.

(c) Absolute values of standardized Pearson residuals and standardized deviance residuals are less than 2.0.

So, there is no lack of fit.

```
data prob5_18;
 input city $ smoking cancer noncancer@;
 n=cancer+noncancer;
cards;
 Beijing 1 126 100
 Beijing 0  35  61
 Shanghai 1 908 688
 Shanghai 0 497 807
 Shenyang 1 913 747
 Shenyang 0 336 598
 Nanjing 1 235 172
 Nanjing 0 58 121
 Harbin 1 402 308
 Harbin 0 121 215
 Zhengzhou 1 182 156
 Zhengzhou 0 72 98
 Taiyuan 1 60 99
 Taiyuan 0 11 43
 Nanchang 1 104 89
 Nanchang 0 21 36
run;

proc genmod data=prob5_18;
 class city smoking(ref=first)/param=ref ;
 model cancer/n=city smoking / dist=bin link=logit r;
run;
```

2. $\text{logit}\,\hat{\pi} = -9.35 + .834(\text{weight}) + .307(\text{width})$

a. LR statistic $= 32.9$ ($df = 2$), P-value $< .0001$. There is extremely strong evidence that at least one variable affects the response.

b. Wald statistics are $(.834/.671)^2 = 1.55$ and $(.307/.182)^2 = 2.85$. These each have $df = 1$, and the P-values are .21 and .09. These predictors are highly correlated (Pearson Correlation = .887), so this is the problem of multicollinearity

```
data crab;
 infile 'C:\Users\Ondra\Desktop\LSU\CDA\crabs_SAS.dat';
 input color spine width satell weight;
 if satell>0 then y=1; if satell=0 then y=0; n=1;
 weight=weight/1000; color=color-1;
 if color=4 then dark=0; if color<4 then dark=1;
;

proc print data=crab noobs;
var color spine width satell weight y dark;
run;

proc corr data=crab;
var weight width;
run;

proc genmod data=crab;
model y/n = weight width /dist=bin link=logit lrci type3;
run;

proc genmod data=crab;
model y/n = /dist=bin link=logit lrci type3;
run;
```