

(a) We have a binary outcome, where 0 and 1 represent absence and presence, resp., of byssinosis. For this situation, it is appropriate to use a model that relates the prob. that a worker suffers from byssinosis (denoted by p) to the covariates, as in the log. reg. model. Note that the data is given in an aggregated form.

Using only DUST, the model takes the form

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)},$$

where

$$x_1 = \begin{cases} 1, & \text{if medium dustiness,} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$x_2 = \begin{cases} 1, & \text{if low dustiness,} \\ 0, & \text{otherwise.} \end{cases}$$

(b) The estimated probability $\hat{p}$, that a worker suffers from byssinosis is associated with dustiness as follows:

$$\text{If DUST = high : } \hat{p} = \frac{\exp(\hat{\beta}_0)}{1 + \exp(\hat{\beta}_0)} \approx 0.157$$

$$\text{If DUST = medium : } \hat{p} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1)} \approx 0.014$$

$$\text{If DUST = high : } \hat{p} = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_2)} \approx 0.012$$

Thus, high dustiness is associated with a considerable increase in risk.

(c)

$$p = \frac{\exp(\quad)}{1 + \exp(\quad)}$$

where x_1 and x_2 are defined as in (a) and

$$x_3 = \begin{cases} 1, & \text{if employed 10-20 years} \\ 0, & \text{otherwise,} \end{cases}$$

and

$$x_4 = \begin{cases} 1, & \text{if employed } > 20 \text{ years,} \\ 0, & \text{otherwise.} \end{cases}$$

Let p_1 and p_2 be the risk probabilities of two workers, whose workplace has the same level of dustiness. Worker 1 has been employed 10-20 years and worker 2 has been employed < 10 years.

The OR is given by:

$$OR = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}} = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3)}{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)} = e^{\beta_3}$$

Using the output of Model 2, the estimated OR is:

$$\hat{OR} = e^{\hat{\beta}_3} = e^{0.5643} \approx 1.76$$

Thus, the odds for a worker employed 10-20 years is 76% higher than that of worker employed < 10 years.

A 95% CI for the OR is given by

$$e^{\hat{\beta}_3 \pm 1.96 \text{ se}(\hat{\beta}_3)} = e^{0.564 \pm 0.486}$$

Thus, we are 95% confident that the OR is between

$$e^{0.564 - 0.486} = 1.08 \text{ and } e^{0.564 + 0.486} = 2.86$$

(d) let p_1 be defined as in (c). Let p_3 be the risk probability of a worker with a workplace having the same dustiness and who has been employed >20 years.

The OR is given by:

$$OR = \frac{\frac{p_3}{1-p_3}}{\frac{p_1}{1-p_1}} = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4)}{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3)} = e^{\beta_4 - \beta_3}.$$

Using the output of Model 2, the estimated OR is:

$$\hat{OR} = e^{\hat{\beta}_4 - \hat{\beta}_3} = e^{0.673 - 0.564} = 1.12.$$

Thus, the odds for worker 3 is 12% higher than the odds for worker 1.

(e) To test if smoking has a significant effect on the risk of suffering from byssinosis, we use the G test statistic:

$$G = D_2 - D_3,$$

where D_2 is the residual deviance for Model 2 (smoking excluded) and D_3 is the deviance for Model 3 (smoking included). Under H_0 (no effect of smoking), G will be approx. χ^2 distributed with 1 d.o.f. Using the outputs of Models 2 and 3, we find that

$$G = 23.53 - 12.09 = 11.44,$$

which gives a p-value of less than 0.5% (see table), and we conclude that smoking has a significant effect.