

(a) If we assume that each driver in a group has the same risk of being involved in an accident each day, the counting process of accidents is a Poisson process, and have ~~not~~ the number of accidents in a three-month period is Poisson distributed.

(b) The estimated RRs are:

- Level 2 : $RR(2) = e^{0.149} \approx 1.16$
- Level 3 : $RR(3) = e^{0.389} \approx 1.48$
- Level 4 : $RR(4) = e^{0.553} \approx 1.74$

Thus, level 2, 3 and 4 have an accident rate that is 16%, 48% and 74% higher, resp., than level 1.

(c) The target RR is $RR(2) = e^{\beta_1}$, where β_1 is the regression coefficient associated with level 2, and level 1 is the reference.

We first compute the 95% CI for β_1 :

$$\hat{\beta}_1 \pm 1.96 \text{ se}(\hat{\beta}_1) = 0.1493 \pm 1.96 \cdot 0.0505 = (0.0503, 0.2483)$$

By exponentiation, we obtain the 95% CI for $RR(2)$:

$$(e^{0.0503}, e^{0.2483}) = (1.05, 1.28)$$

(d) We compute RRs for different age groups (comparing ^{age} against the youngest age group)

• 25-29 : $RR = \dots = 0.83$

• 30-35 : $RR = \dots = 0.71$

• >35 : $RR = \dots = 0.58$

Thus, drivers that are 25-29, 30-35, >35 years old have an accident rate that is 83%, 71%, 58%, resp., of the rate of a driver that is <25 years old.

Let β_2 be regression coefficient for drivers in ^{age} category 25-29. To test $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$, we compute the Wald test statistic:

$$Z = \frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)} = -\frac{0.191}{0.083} = -2.30$$

This gives a p-value that is < 0.05 and we can reject H_0 at sig. level $\alpha = 0.05$, implying a sig. difference in accident rates.

(e) For the model with VOLUME as the only covariate, the intercept is the log of the rate for cars with motor volume < 1 litre.

For the model with VOLUME, AGE, AREA as covariates, the intercept is the log of the rate for cars with < 1 litre motor, with drivers that are < 25 years and do not live in big city.

As the intercept in the second model represents the youngest age group, with the highest accident rate, it is reasonable that it is higher than the intercept of the first model.

(4) An interaction between AGE and AREA means that the effect of age is not the same across areas.

To test the hypothesis of no interaction, we compute the G-statistic, which is the difference in deviance:

$$G = 51.42 - 40.91 = 10.51$$

Under H_0 (no interaction), the G-statistic follows a χ^2 distribution with $3 \cdot 3 = 9$ d.o.f. (i.e., the number of parameters used to model the interaction). This gives a p-value that is > 0.10 , meaning there is no significant support for an interaction effect.