W5 Solutions

. (a) For this data, $p_1 = 17/147 = 0.1156$ and $p_2 = 218/646 = 0.3375$. Thus, the sample difference in proportions, $p_1 - p_2 =$ -0.2218. The estimated standard error,

$$\hat{\sigma}(p1 - p2) = \sqrt{\frac{0.1156(1-0.1156)}{147} + \frac{0.3375(1-0.3375)}{646}} = 0.0323.$$

Therefore, the 90% C.I. for difference of proportions is

$$-0.2218 \pm 1.645 \times 0.0323 = (-0.2749, -0.1687).$$

The confidence interval does not contain 0 and entirely on the negative side, from which we can conclude that head injuries are not independent of wearing helmets and the probability of having head injuries without helmets is higher than that with helmets.
(b) The sample odds ratio,

$$\hat{\theta} = \frac{17 \times 428}{218 \times 130} = 0.2567$$

and sample log-odds ratio is, $\log \hat{\theta} =$ -1.3597. The estimated standard error for $\log\hat{\theta}$ is

$$\hat{\sigma}(\log \hat{\theta}) = \sqrt{\frac{1}{17} + \frac{1}{218} + \frac{1}{130} + \frac{1}{428}} = 0.2710.$$

Therefore, the 90% C.I. for log-odds ratio is

$$-1.3597 \pm 1.645 \times 0.2710 = (-1.8055, -0.9139).$$

Therefore, the 90% C.I. for odds ratio is,

$$(e^{-1.8055}, e^{-0.9139}) = (0.1644, 0.4010).$$

The confidence interval is smaller than 1, which also supports the fact observed in part (a).
(c) The following table shows the expected frequencies:

(c) The following table shows the expected frequencies:

	Wearing Helmet	
Head Injury	Yes	No
Yes	43.56	191.44
No	103.44	454.56

Therefore, to test for independence, the Pearson's chi-square test statistic is, $\chi^2 = 28.256$ and likelihood ratio test statistic is $G^2 = 32.5432$ with 1 degrees of freedom. The p-values for these tests are, 1.063×10^{-7} and 1.165×10^{-8} respectively. Both of these tests suggest that head injuries are dependent on wearing helmets.

(a) Calculate the crude (Marginal) odds ratio $\hat{\theta}_{XY}$. [4 Marks]

Solution: The marginal table,

(c) Calculate the Cochran-Mantel-Haenszel odds ratio $\hat{\theta}_{CMH}$ for the table above [4 Marks]

Solution: The $\hat{\theta}_{CMH} = 0.3434343$

Law	Accepted	Rejected	Total
Males	490	210	700
Females	280	220	500
Total	770	430	1200

The marginal odds ratio, $\hat{\theta} = \dfrac{490 \times 220}{210 \times 280} = 1.833$

(b) Let the conditional odds ratios for the kth table be $\hat{\theta}_k = \frac{a_k d_k}{b_k c_k}$. The Cochran-Mantel-Haenszel odds ratio is calculated as $\hat{\theta}_{CMH} = \frac{\sum_k w_k \hat{\theta}_k}{\sum_k w_k} = \frac{\sum_k \frac{a_k d_k}{n_k}}{\sum_k \frac{b_k c_k}{n_k}}$.

Show that, $w_k = \dfrac{b_k c_k}{n_k}$ [3 Marks]

Solution: The numerator of $\hat{\theta}_{CMH} = \sum_k wk\hat{\theta}_k$. We know, $\hat{\theta}_k = \dfrac{a_k d_k}{b_k c_k}$. Thus,

$$
\begin{aligned}
w_k\hat{\theta}_k &= \frac{a_k d_k}{n_k} \\
\Rightarrow w_k &= \frac{a_k d_k}{n_k \hat{\theta}_k} \\
\Rightarrow w_k &= \frac{a_k d_k b_k c_k}{n_k a_k d_k} \\
\Rightarrow w_k &= \frac{b_k c_k}{n_k} \qquad Q.E.D
\end{aligned}
$$

. If Y_1 and Y_2 are independent Poisson random variable with parameters μ_1 and μ_2 respectively. Then find the conditional distribution of Y_1 given $Y_1 + Y_2 = n$. That is calculate $P(Y_1 = k | Y_1 + Y_2 = n)$. [5 Marks]

Solution:

$$
\begin{aligned}
\mathbb{P}(Y_1 = k | Y_1 + Y_2 = n) &= \frac{\mathbb{P}(Y_1 = k, Y_1 + Y_2 = n)}{\mathbb{P}(Y_1 + Y_2 = n)} \\
&= \frac{\mathbb{P}(Y_1 = k, Y_2 = n - k)}{\mathbb{P}(Y_1 + Y_2 = n)} \\
&= \frac{\mathbb{P}(Y_1 = k)\mathbb{P}(Y_2 = n - k)}{\mathbb{P}(Y_1 + Y_2 = n)} \\
&= \frac{\frac{e^{-\mu_1}\mu_1^k}{k!}\frac{e^{-\mu_2}\mu_2^{n-k}}{(n-k)!}}{\frac{e^{-\mu_1+\mu_2}(\mu_1+\mu_2)^n}{n!}} \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^k \left(\frac{\mu_2}{\mu_1+\mu_2}\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^k \left(1 - \frac{\mu_1}{\mu_1+\mu_2}\right)^{n-k}
\end{aligned}
$$

Thus the conditional distribution is Binomial with parameters $n = n$ and $\pi = \dfrac{\mu_1}{\mu_1 + \mu_2}$

If $Y_1 \sim \text{Bin}(n_1, \pi)$ and $Y_2 \sim \text{Bin}(n_2, \pi)$ are independent. Then find the conditional distribution of Y_1 given $Y_1 + Y_2 = m$. That is calculate $P(Y_1 = k | Y_1 + Y_2 = m)$. [5 Marks]

Solution:

$$
\begin{aligned}
\mathbb{P}(Y_1 = k | Y_1 + Y_2 = m) &= \frac{\mathbb{P}(Y_1 = k, Y_1 + Y_2 = m)}{\mathbb{P}(Y_1 + Y_2 = m)} \\
&= \frac{\mathbb{P}(Y_1 = k, Y_2 = m - k)}{\mathbb{P}(Y_1 + Y_2 = m)} \\
&= \frac{\mathbb{P}(Y_1 = k)\mathbb{P}(Y_2 = m - k)}{\mathbb{P}(Y_1 + Y_2 = m)} \\
&= \frac{\binom{n_1}{k}\pi^k(1-\pi)^{n_1-k}\binom{n_2}{(m-k)}\pi^{m-k}(1-\pi)^{n_2-m+k}}{\binom{(n_1+n_2)}{m}\pi^m(1-\pi)^{n_1+n_2-m}} \\
&= \frac{\binom{n_1}{k}\binom{n_2}{(m-k)}}{\binom{(n_1+n_2)}{m}}
\end{aligned}
$$

Let, $n_1 + n_2 = N$ and $n_1 = n$, then the probability bcomes,

$$
\mathbb{P}(Y_1 = k | Y_1 + Y_2 = m) = \frac{\binom{n}{k}\binom{N-n}{(m-k)}}{\binom{N}{m}}
$$

This is the PMF of a Hypergeometric(N, m, n) distribution