1

$$H_0 : \text{gender and the dominant leadership style are independent}$$

$$\text{against}$$

$$H_a : \text{gender and the dominant leadership style are dependent}$$

Gender	Authoritarian	Democratic	Laissez-faire	total
Male	12	22	9	43
Female	20	13	3	36
total	32	35	12	79

Under H_0,

$$\nu = (2-1)(3-1) = (1)(2) = 2$$

$$O = \{12, 22, 9, 20, 13, 3\}$$

$$E = \left\{\frac{1376}{79}, \frac{1505}{79}, \frac{516}{79}, \frac{1152}{79}, \frac{1260}{79}, \frac{432}{79}\right\}$$

The test statistic $\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(2)$.

$$\chi^2_{\text{calc}} \approx 6.75$$

$$p\text{-value} = P\left(\sum \frac{(O_i - E_i)^2}{E_i} \geqslant \chi^2_{\text{calc}}\right) \approx 0.0343$$

Since p-value ≈ 0, there is some evidence for the rejection of H_0. Moreover, from the largest contributors to the test statistic, Males and Females who are identified as Authoritarian dominant, we can see that there are more females than males, contrary to the total number of males and females. Therefore, we conclude that there is sufficient evidence that gender and the dominant leadership style are dependent.

2 (a)

(i)

$$\bar{x} = \frac{\sum fx}{\sum f} = 1.33$$

$$H_0 : \text{the data follows a Poisson distribution}$$

$$\text{against}$$

$$H_a : \text{the data does not follow a Poisson distribution}$$

Under H_0,

$$\nu = (6-1) - 1 = 4$$

$$O = \{24, 36, 28, 8, 3, 1\}$$

$$E = \left\{e^{-1.33}, 1.33e^{-1.33}, \frac{1.33^2 e^{-1.33}}{2}, \frac{1.33^3 e^{-1.33}}{6}, \frac{1.33^4 e^{-1.33}}{24}, \frac{1.33^5 e^{-1.33}}{120}\right\}$$

The test statistic $\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(2)$.

$$\chi^2_{\text{calc}} \approx 1.76$$

$$p\text{-value} = P\left(\sum \frac{(O_i - E_i)^2}{E_i} \geqslant \chi^2_{\text{calc}}\right) \approx 0.780$$

Since p-value $> 5\%$, we do not reject H_0, and conclude that there is insufficient evidence at a 5% level that the data does not follow a Poisson distribution.

(ii)

$$H_0 : \text{the data follows a Poisson distribution with mean } 1.33$$

against

$$H_a : \text{the data does not follow a Poisson distribution with mean } 1.33$$

Under H_0,

$$\nu = 6 - 1 = 5$$

$$\vdots$$

(b)

$$H_0 : \text{there is no association between the channel watched most and the region}$$

against

$$H_a : \text{there is association between the channel watched most and the region}$$

Region	North	South	East	West	total
Channel 5	10	17	$10 + a$	$23 - a$	60
Channel 8	5	8	$20 - a$	$7 + a$	40
total	15	25	30	30	100

Under H_0,

$$\nu = (2 - 1)(4 - 1) = (1)(3) = 3$$

$$O = \{10, 17, 10 + a, 23 - a, 5, 8, 20 - a, 7 + a\}$$

$$E = \{9, 15, 18, 18, 6, 10, 12, 12\}$$

The test statistic $\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(3)$.

$$\begin{aligned}
\chi^2_{\text{calc}} &= \frac{1}{9} + \frac{4}{15} + \frac{(a-8)^2}{18} + \frac{(a-5)^2}{18} + \frac{1}{6} + \frac{2}{5} + \frac{(a-8)^2}{12} + \frac{(a-5)^2}{12} \\
&= \frac{34}{36} + \frac{5}{36}\left(a^2 - 16a + 64 + a^2 - 10a + 25\right) \\
&= \frac{10a^2 - 130a + 479}{36} \quad \square
\end{aligned}$$

$$
\begin{aligned}
p\text{-value} &> 5\% \\
P\left(\sum \frac{(O_i - E_i)^2}{E_i} \geqslant \chi^2_{\text{calc}}\right) &> 0.05 \\
\frac{10a^2 - 130a + 479}{36} &< c \\
10a^2 - 130a + 479 - 36c &< 0 \\
a &\in (1.76, 11.2)
\end{aligned}
$$

3

(i)

$$
\begin{aligned}
\chi^2_{\text{calc}} &= \frac{(a-np_1)^2}{np_1} + \frac{(b-np_2)^2}{np_2} + \frac{(c-np_3)^2}{np_3} + \frac{(d-np_4)^2}{np_4} \\
&= \frac{a^2 - 2anp_1 + n^2p_1^2}{np_1} + \frac{b^2 - 2bnp_2 + n^2p_2^2}{np_2} + \frac{c^2 - 2cnp_3 + n^2p_3^2}{np_3} + \frac{d^2 - 2dnp_4 + n^2p_4^2}{np_4} \\
&= \frac{a^2}{np_1} + \frac{b^2}{np_2} + \frac{c^2}{np_3} + \frac{d^2}{np_4} - 2(a+b+c+d) + n(p_1+p_2+p_3+p_4) \\
&= \frac{a^2}{np_1} + \frac{b^2}{np_2} + \frac{c^2}{np_3} + \frac{d^2}{np_4} - n \quad \square
\end{aligned}
$$

2.1 (ii)

$$
\begin{aligned}
n &= 26 + 19 + 10 + 45 = 100 \\
p_1 + p_2 + p_3 + p_4 &= 1 \\
p + 3p_3 + p_3 + p &= 1 \\
4p_3 &= 1 - 2p \\
p_3 &= \frac{1-2p}{4}
\end{aligned}
$$

$$
\begin{aligned}
\chi^2_{\text{calc}} &= \frac{26^2}{100p} + \frac{19^2}{100 \cdot 3(1-2p)/4} + \frac{10^2}{100(1-2p)/4} + \frac{45^2}{100p} - 100 \\
&= \frac{676}{100p} + \frac{361}{75(1-2p)} + \frac{300}{75(1-2p)} + \frac{2025}{100p} - 100 \\
&= \frac{2701}{100p} + \frac{661}{75(1-2p)} - 100
\end{aligned}
$$

$$
p_0 \approx \boxed{0.356}
$$

(iii)

H_0 : the population of the city in the country conforms to the figures

against

H_a : the population of the city in the country does not conform to the figures

Under H_0,

$$\nu = 4 - 1 = 3$$

$$O = \{26, 19, 10, 45\}$$

$$E = \{100p_0, 75 - 150p_0, 25 - 50p_0, 100p_0\}$$

The test statistic $\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(3)$.

$$\chi^2_{\text{calc}} \approx 6.47$$

$$p\text{-value} = P\left(\sum \frac{(O_i - E_i)^2}{E_i} \geqslant \chi^2_{\text{calc}}\right) \approx 0.0907$$

Since p-value $< 10\%$, we reject H_0, and conclude that there is sufficient evidence at a 10% level that the population of the city in the country does not conform to the figures.

(iv)

H_0 will be rejected because other values of p will lead to a greater test statistic, making p-value smaller.

(v)

$$(0.174, 0.346)$$

(I forget everything about CI...)

(vi)

Advantage: the range of the interval is smaller, giving more precision to the actual value of the proportion.

Disadvantage: the interval is less likely to contain the actual value.