24400 HW8

Bin Yu

December 5, 2024

Question 1

(a)

To find the likelihood function $f_{X|\Theta}(x_1, \ldots, x_n|\theta)$, note that $X_1, \ldots, X_n$ are independent given Θ. Therefore, the joint probability mass function is the product of the individual PMFs:

$$\begin{aligned} f_{X|\Theta}(x_1, \ldots, x_n|\theta) &= \prod_{i=1}^{n} f_X(x_i|\theta) \\ &= \prod_{i=1}^{n} \left((1-\theta)^{x_i-1}\theta\right) \\ &= \theta^n (1-\theta)^{\sum_{i=1}^{n}(x_i-1)} \\ &= \theta^n (1-\theta)^{S-n}, \end{aligned}$$

where $S = \sum_{i=1}^{n} x_i$.

(b)

The joint distribution $f(x_1, \ldots, x_n, \theta)$ is given by multiplying the likelihood function with the prior distribution:

$$\begin{aligned} f(x_1, \ldots, x_n, \theta) &= f_{X|\Theta}(x_1, \ldots, x_n|\theta) f_\Theta(\theta) \\ &= \theta^n (1-\theta)^{S-n} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{n+\alpha-1}(1-\theta)^{S-n+\beta-1}. \end{aligned}$$

(c)

Based on Baye's theorem:

$$f_{\Theta|X}(\theta|x_1, \ldots, x_n) = \frac{f_{X|\Theta}(x_1, \ldots, x_n|\theta) f_\Theta(\theta)}{f_X(x_1, \ldots, x_n)}$$

Therefore, the posterior distribution of Θ given the data can be seen as proportional to the product of the likelihood function and the prior distribution:

$$\begin{aligned} f_{\Theta|X}(\theta|x_1, \ldots, x_n) &\propto f_{X|\Theta}(x_1, \ldots, x_n|\theta) f_\Theta(\theta) \\ &\propto \theta^{n+\alpha-1}(1-\theta)^{S-n+\beta-1}. \end{aligned}$$

Therefore, the posterior distribution is Beta(α', β') with parameters:

$$\alpha' = n + \alpha, \quad \beta' = S - n + \beta.$$

Therefore, the posterior density function is:

$$f_{\Theta|X}(\theta|x_1, \ldots, x_n) = \frac{\Gamma(\alpha' + \beta')}{\Gamma(\alpha')\Gamma(\beta')} \theta^{\alpha'-1}(1-\theta)^{\beta'-1}, \quad 0 \leq \theta \leq 1.$$

where,

$$\alpha' = n + \alpha, \quad \beta' = S - n + \beta.$$

(d)

The posterior mean of Θ given the data is the mean of the Beta distribution Beta(α', β'):

$$E[\Theta|X_1, \ldots, X_n] = \frac{\alpha'}{\alpha' + \beta'} = \frac{n + \alpha}{S + \alpha + \beta}.$$

Question 2

(a)

The probability density function (pdf) $f(x|\theta)$ is:

$$f(x|\theta) = \begin{cases} \frac{1}{2\theta}, & \text{if } x \in [-\theta, \theta], \\ 0, & \text{otherwise.} \end{cases}$$

Since $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.), the likelihood function of the data is:

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta) = \left(\frac{1}{2\theta}\right)^n \prod_{i=1}^{n} \mathbf{1}_{\{X_i \in [-\theta, \theta]\}}.$$

The indicator function will be 1 if all X_i are in $[-\theta, \theta]$, that is, $\theta \geq \max_{1 \leq i \leq n} |X_i|$, thus,

$$L(\theta) = \left(\frac{1}{2\theta}\right)^n \mathbf{1}_{\{\theta \geq \max_{1 \leq i \leq n} |X_i|\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

(b)

The likelihood function will be positive only if all observations X_i lie in $[-\theta, \theta]$, and since for $\theta \geq \max_i |X_i|$, $L(\theta)$ decreases as θ increases, the maximum occurs at:

$$\hat{\theta} = \max_{1 \leq i \leq n} |X_i|.$$

(c)

the maximum likelihood estimator is:

$$\hat{\theta} = \max_{1 \le i \le n} |X_i|.$$

$$P\left(\hat{\theta} < (1-\varepsilon)\theta\right) = P\left(\max_{1 \le i \le n} |X_i| < (1-\varepsilon)\theta\right).$$

$$P\left(\max_{1 \le i \le n} |X_i| < (1-\varepsilon)\theta\right) = P\left(|X_1| < (1-\varepsilon)\theta,\, |X_2| < (1-\varepsilon)\theta, \ldots, |X_n| < (1-\varepsilon)\theta\right).$$

Since the X_i are independent and identically distributed (i.i.d.), the joint probability is the product of the individual probabilities:

$$P\left(|X_1| < (1-\varepsilon)\theta, \ldots, |X_n| < (1-\varepsilon)\theta\right) = \prod_{i=1}^{n} P\left(|X_i| < (1-\varepsilon)\theta\right).$$

The event $|X_i| < (1-\varepsilon)\theta$ corresponds to $X_i \in (-(1-\varepsilon)\theta,\, (1-\varepsilon)\theta)$. The probability for a single X_i:

$$\begin{aligned}
P\left(|X_i| < (1-\varepsilon)\theta\right) &= P\left(X_i \in (-(1-\varepsilon)\theta,\, (1-\varepsilon)\theta)\right) \\
&= \int_{-(1-\varepsilon)\theta}^{(1-\varepsilon)\theta} f(x|\theta)\, dx \\
&= \int_{-(1-\varepsilon)\theta}^{(1-\varepsilon)\theta} \frac{1}{2\theta}\, dx \\
&= \frac{1}{2\theta}\, [x]_{-(1-\varepsilon)\theta}^{(1-\varepsilon)\theta} \\
&= \frac{1}{2\theta}\left((1-\varepsilon)\theta - (-(1-\varepsilon)\theta)\right) \\
&= \frac{1}{2\theta}\left(2(1-\varepsilon)\theta\right) \\
&= (1-\varepsilon).
\end{aligned}$$

Therefore,

$$\prod_{i=1}^{n} P\left(|X_i| < (1-\varepsilon)\theta\right) = \left((1-\varepsilon)\right)^n.$$

$$P\left(\hat{\theta} < (1-\varepsilon)\theta\right) = (1-\varepsilon)^n.$$

(d)

the maximum likelihood estimator is:

$$\hat{\theta} = \max_{1 \le i \le n} |X_i|.$$

$$P\left(\hat{\theta} > (1+\varepsilon)\theta\right) = P\left(\max_{1 \le i \le n} |X_i| > (1+\varepsilon)\theta\right).$$

Since each X_i is drawn from the uniform distribution Uniform$[-\theta, \theta]$, so the possible values of X_i are in the interval $[-\theta, \theta]$:

$$|X_i| \le \theta \quad \text{for all } i = 1, 2, \ldots, n.$$

Therefore:

$$(1+\varepsilon)\theta > \theta \geq |X_i| \quad \text{for all } i.$$

The probability that any single X_i satisfies $|X_i| > (1+\varepsilon)\theta$ is zero:

$$P\left(|X_i| > (1+\varepsilon)\theta\right) = 0.$$

The probability that the maximum of $|X_i|$ exceeds $(1+\varepsilon)\theta$ is also zero, because if none of the $|X_i|$ can exceed $(1+\varepsilon)\theta$, their maximum either:

$$P\left(\max_{1\leq i\leq n} |X_i| > (1+\varepsilon)\theta\right) = 0.$$

Therefore:

$$P\left(\hat{\theta} > (1+\varepsilon)\theta\right) = 0.$$

Question 3

(a)

Since $X_1, \ldots, X_n \sim N(0, \theta)$, the likelihood function for a given θ is:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{X_i^2}{2\theta}\right).$$

The likelihood ratio is:

$$LR = \frac{L(\theta=1)}{L(\theta=2)} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(1)}} \exp\left(-\frac{X_i^2}{2(1)}\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(2)}} \exp\left(-\frac{X_i^2}{2(2)}\right)}.$$

Simplifies to:

$$LR = 2^{n/2} \exp\left(-\frac{1}{4}\sum_{i=1}^{n} X_i^2\right).$$

(b)

Under H_0 where $\theta = 1$, the test statistic LR becomes:

$$LR = 2^{n/2} \exp\left(-\frac{1}{4}S\right),$$

where $S = \sum_{i=1}^{n} X_i^2$.

Reject H_0 if $LR \leq c = 1$. Therefore, the Type I error (significance level α) is:

$$\alpha = P_{H_0}(LR \leq 1) = P\left(2^{n/2} \exp\left(-\frac{S}{4}\right) \leq 1\right).$$

Taking natural logarithms:

$$
\begin{aligned}
&P\left(\ln\left(2^{n/2}\exp\left(-\frac{S}{4}\right)\right)\leq \ln 1\right)\\
=&P\left(\frac{n}{2}\ln 2-\frac{S}{4}\leq 0\right)\\
=&P\left(-\frac{S}{4}\leq -\frac{n}{2}\ln 2\right)\\
=&P\left(\frac{S}{4}\geq \frac{n}{2}\ln 2\right)\\
=&P\left(S\geq 2n\ln 2\right).
\end{aligned}
$$

Under H_0, Since $X_i \sim N(0,1)$, we have $X_i^2 \sim \chi^2(1)$. Therefore, $S=\sum_{i=1}^n X_i^2 \sim \chi^2(n)$.

For $n=40$:

$$2n\ln 2 = 2\times 40\times 0.6931 \approx 55.45.$$

Thus, the Type I error is:

$$\alpha = P\left(\chi^2(40)\geq 55.45\right).$$

From the chi-squared distribution table in R:

$$P\left(\chi^2(40)\leq 55.45\right)\approx 0.9471.$$

Therefore:

$$\alpha = 1-0.9471 = 0.0529.$$

The Type I error (significance level) is approximately $\alpha \approx 0.0529$.

(c)

Under H_1 where $\theta = 2$, the test statistic LR remains:

$$LR = 2^{n/2}\exp\left(-\frac{S}{4}\right),$$

where $S=\sum_{i=1}^n X_i^2$.

Each X_i is distributed as $X_i \sim N(0,2)$ under H_1, therefore, $X_i^2 = 2Z_i^2$.

Since $Z_i^2 \sim \chi^2(1)$, it follows that $X_i^2 \sim 2\chi^2(1)$.

$$S=\sum_{i=1}^n X_i^2 = 2\sum_{i=1}^n Z_i^2 \sim 2\chi^2(n).$$

Calculate the probability of failing to reject H_0 when H_1 is true:

$$\beta = P_{H_1}(LR>1).$$

$$\beta = P_{H_1}\left(2^{n/2}\exp\left(-\frac{S}{4}\right) > 1\right).$$

Take natural logarithms:

$$\begin{aligned}
&P_{H_1}\left(\ln\left(2^{n/2}\exp\left(-\frac{S}{4}\right)\right) > 0\right) \\
&= P_{H_1}\left(\frac{n}{2}\ln 2 - \frac{S}{4} > 0\right) \\
&= P_{H_1}\left(-\frac{S}{4} > -\frac{n}{2}\ln 2\right) \\
&= P_{H_1}\left(S < 2n\ln 2\right).
\end{aligned}$$

Since $S \sim 2\chi^2(n)$, we can write:

$$\beta = P\left(2\chi^2(n) < 2n\ln 2\right) = P\left(\chi^2(n) < n\ln 2\right).$$

For $n = 40$:

$$n\ln 2 = 40 \times 0.6931 \approx 27.72.$$

$$\beta = P\left(\chi^2(40) < 27.72\right).$$

Using the chi-squared distribution table in R:

$$P\left(\chi^2(40) < 27.72\right) \approx 0.0709.$$

So:

$$\beta \approx 0.0709.$$

(d)

Given: Sample size: $n = 40$

Sample mean: $\bar{X} = 0.8$

Sample variance: $S^2 = 0.9$

To calculate the LR statistic, we need $S = \sum_{i=1}^{n} X_i^2$.

The sample variance S^2 is defined as:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n\bar{X}^2$$

Therefore,

$$\sum_{i=1}^{n} X_i^2 = (n-1)S^2 + n\bar{X}^2$$

$$(n-1)S^2 = (40-1) \times 0.9 = 39 \times 0.9 = 35.1$$
$$n\bar{X}^2 = 40 \times (0.8)^2 = 40 \times 0.64 = 25.6$$
$$S = 35.1 + 25.6 = 60.7$$

Therefore,

$$LR = 2^{20} \exp\left(-15.175\right)$$
$$LR \approx 0.269$$

Thus,

$$LR \approx 0.269$$

Question 4

(a)

Under H_0, the density function of X_i is:

$$f_0(x) = \begin{cases} 1, & \text{if } x \in [0,1], \\ 0, & \text{otherwise.} \end{cases}$$

$$f_0(x) = 1 \cdot \mathbf{1}_{[0,1]}(x).$$

Under H_1:

The density function of X_i is:

$$f_1(x) = \begin{cases} 6x(1-x), & \text{if } x \in [0,1], \\ 0, & \text{otherwise.} \end{cases}$$
$$f_1(x) = 6x(1-x) \cdot \mathbf{1}_{[0,1]}(x).$$

Since $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.), the likelihood functions under H_0 and H_1 are:

Under H_0:

$$L_0 = \prod_{i=1}^{n} f_0(X_i) = \prod_{i=1}^{n} \left[1 \cdot \mathbf{1}_{[0,1]}(X_i)\right] = \mathbf{1}_{[0,1]^n}(X_1, \dots, X_n).$$

Under H_1:

$$L_1 = \prod_{i=1}^{n} f_1(X_i) = \prod_{i=1}^{n} \left[6X_i(1-X_i) \cdot \mathbf{1}_{[0,1]}(X_i)\right] = \left(\prod_{i=1}^{n} 6X_i(1-X_i)\right) \cdot \mathbf{1}_{[0,1]^n}(X_1, \ldots, X_n).$$

The likelihood ratio statistic is:

$$\Lambda(X_1, \ldots, X_n) = \frac{L_0}{L_1} = \frac{\mathbf{1}_{[0,1]^n}(X_1, \ldots, X_n)}{\left(\prod_{i=1}^{n} 6X_i(1-X_i)\right) \cdot \mathbf{1}_{[0,1]^n}(X_1, \ldots, X_n)} = \frac{1}{\prod_{i=1}^{n} 6X_i(1-X_i)}.$$

(b)

For $n = 1$, the likelihood ratio simplifies to:

$$\Lambda(X) = \frac{1}{6X(1-X)}.$$

We reject H_0 if $\Lambda(X) \leq c = 1$, and accept H_0 if $\Lambda(X) > 1$.

$$\frac{1}{6X(1-X)} \leq 1.$$

$$\begin{aligned} 6X(1-X) - 1 &\geq 0. \\ -6X^2 + 6X - 1 &\geq 0. \\ 6X^2 - 6X + 1 &\leq 0. \end{aligned}$$

$$X = \frac{6 \pm \sqrt{(-6)^2 - 4 \times 6 \times 1}}{2 \times 6} = \frac{6 \pm \sqrt{36 - 24}}{12} = \frac{6 \pm \sqrt{12}}{12}.$$

$$\begin{aligned} X_1 &= \frac{6 - 3.4641}{12} = \frac{2.5359}{12} \approx 0.2113, \\ X_2 &= \frac{6 + 3.4641}{12} = \frac{9.4641}{12} \approx 0.7887. \end{aligned}$$

The quadratic $6X^2 - 6X + 1$ opens upwards, so the expression is less than or equal to zero between the roots X_1 and X_2.

Therefore, we reject H_0 when:

$$X \in [0.2113, 0.7887].$$

We accept H_0 when:

$$X \in [0, 0.2113) \cup (0.7887, 1].$$

(c)

The Type I error (α) is the probability of rejecting H_0 when H_0 is true.

Under H_0, $X \sim \text{Uniform}[0, 1]$:

$$\alpha = P_{H_0}(\text{Reject } H_0) = P_{H_0}(X \in [0.2113, 0.7887]) = 0.7887 - 0.2113 = 0.5774.$$

Therefore, the Type I error is approximately $\alpha = 0.5774$

(d)

The power of the test is the probability of correctly rejecting H_0 when H_1 is true.

Under H_1, X has density:

$$f_1(x) = 6x(1 - x), \quad x \in [0, 1].$$

$$\text{Power} = P_{H_1}(\text{Reject } H_0) = \int_{0.2113}^{0.7887} 6x(1 - x)\, dx.$$

$$\int_{0.2113}^{0.7887} (6x - 6x^2)\, dx = \left[3x^2 - 2x^3\right]_{0.2113}^{0.7887}.$$

$$\text{Power} = 0.8858 - 0.1152 = 0.7706.$$

Therefore, the power of the test is 0.7706.

Question 5

We are given the counts of students registering for a course:

- Number of Undergraduates: U
- Number of Master's students: M
- Number of PhD students: P
- Total number of students: $n = U + M + P$

Each student registers independently with probabilities p_u, p_m, and p_p respectively, where $p_u + p_m + p_p = 1$.

The likelihood function for observing the counts $U = u$, $M = m$, and $P = p$ is given by the multinomial distribution:

$$L(p_u, p_m, p_p) = \frac{n!}{U!\, M!\, P!}\, p_u^U\, p_m^M\, p_p^P$$

Given the hypothesis that there should be twice as many Master's students as PhD students. Therefore, we can express the probabilities in terms of a single parameter p:

$$\begin{cases} p_p = p \\ p_m = 2p \\ p_u = 1 - 3p \end{cases}$$

all probabilities must be between 0 and 1:

$$p_p = p \geq 0 \implies p \geq 0$$
$$p_m = 2p \geq 0 \implies p \geq 0$$
$$p_u = 1 - 3p \geq 0 \implies p \leq \frac{1}{3}$$

Thus, $p \in [0, \frac{1}{3}]$.

Substitute p_u, p_m, and p_p into the likelihood function, and we get the likelihood function:

$$L(p) = \frac{n!}{U!\,M!\,P!}\,(1-3p)^U\,(2p)^M\,p^P$$

The log-likelihood function:

$$\ell(p) = \ln L(p) = \ln\left(\frac{n!}{U!\,M!\,P!}\right) + U\ln(1-3p) + M\ln(2p) + P\ln p$$
$$\frac{d\ell}{dp} = U\left(\frac{-3}{1-3p}\right) + M\left(\frac{2}{2p}\right) + P\left(\frac{1}{p}\right)$$
$$\frac{d\ell}{dp} = -\frac{3U}{1-3p} + \frac{M}{p} + \frac{P}{p}$$
$$\frac{d\ell}{dp} = -\frac{3U}{1-3p} + \frac{M+P}{p}$$

Set the derivative equal to zero to find the maximum:

$$-\frac{3U}{1-3p} + \frac{M+P}{p} = 0$$
$$-3Up + (M+P)(1-3p) = 0$$
$$-3(U+M+P)p + (M+P) = 0$$

Since $U + M + P = n$:

$$-3np + (M+P) = 0$$

$$p = \frac{M+P}{3n}$$

Therefore, the MLE of p is:

$$\hat{p} = \frac{M+P}{3n}$$

Question 6

(a)

Let X_i be the number of successes for the i-th student, where $i = 1, 2, \ldots, 50$. Each X_i follows a binomial distribution with parameters $n = 4$ (number of trials) and p (probability of success):

$$X_i \sim \text{Binomial}(4, p).$$

The probability mass function (PMF) of X_i is:

$$P(X_i = k) = \binom{4}{k} p^k (1-p)^{4-k}, \quad \text{for } k = 0, 1, 2, 3, 4.$$

The likelihood function for the observed data $X_1, \ldots, X_{50}$ is:

$$L(p) = \prod_{i=1}^{50} P(X_i) = \prod_{i=1}^{50} \left[\binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} \right].$$

(b)

The likelihood function for the observed data $X_1, X_2, \ldots, X_{50}$ is:

$$L(p) = \prod_{i=1}^{50} P(X_i) = \prod_{i=1}^{50} \left[\binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} \right].$$

log-likelihood:

$$\ell(p) = \ln L(p) = \sum_{i=1}^{50} \left[\ln \binom{4}{X_i} + X_i \ln p + (4 - X_i) \ln(1-p) \right].$$

$$\ell(p) = \sum_{i=1}^{50} \ln \binom{4}{X_i} + \sum_{i=1}^{50} X_i \ln p + \sum_{i=1}^{50} (4 - X_i) \ln(1-p).$$

Denote:

Total successes: $S = \sum_{i=1}^{50} X_i$.

Total failures: $F = \sum_{i=1}^{50} (4 - X_i) = 4 \times 50 - S = 200 - S$.

Taking the Derivative of the Log-Likelihood with Respect to p:

$$\frac{d\ell}{dp} = \frac{S}{p} - \frac{200 - S}{1-p}.$$

Setting the Derivative Equal to Zero to Find the Maximum:

$$\frac{S}{p} - \frac{200 - S}{1 - p} = 0.$$

$$S - Sp = 200p - Sp.$$

$$S = 200p.$$

$$p = \frac{S}{200}.$$

The Maximum Likelihood Estimator (MLE) of p is:

$$\hat{p} = \frac{1}{200} \sum_{i=1}^{50} X_i.$$

(c)

Calculate S using the observed data:

$$\begin{aligned} S &= (0)(9) + (1)(12) + (2)(11) + (3)(14) + (4)(4) \\ &= 0 + 12 + 22 + 42 + 16 = 92. \end{aligned}$$

Then the MLE of p is:

$$\hat{p} = \frac{92}{200} = 0.46.$$

(d)

For a binomial distribution with $n = 4$ and $p = 0.46$, we can calculate $P(X = k)$ for $k = 0, 1, 2, 3, 4$:

For $k = 0$:

$$P(X = 0) = \binom{4}{0}(0.46)^0(0.54)^4 = 1 \times 1 \times (0.54)^4 \approx 0.085.$$

For $k = 1$:

$$P(X = 1) = \binom{4}{1}(0.46)^1(0.54)^3 = 4 \times 0.46 \times (0.54)^3 \approx 0.290.$$

For $k = 2$:

$$P(X = 2) = \binom{4}{2}(0.46)^2(0.54)^2 = 6 \times (0.46)^2 \times (0.54)^2 \approx 0.370.$$

For $k = 3$:

$$P(X = 3) = \binom{4}{3}(0.46)^3(0.54)^1 = 4 \times (0.46)^3 \times 0.54 \approx 0.210.$$

For $k = 4$:

$$P(X = 4) = \binom{4}{4}(0.46)^4(0.54)^0 = 1 \times (0.46)^4 \times 1 \approx 0.045.$$

$$E_k = n \times P(X = k).$$

$$E_0 = 50 \times 0.085 = 4.25.$$

$$E_1 = 50 \times 0.290 = 14.5.$$

$$E_2 = 50 \times 0.370 = 18.5.$$

$$E_3 = 50 \times 0.210 = 10.5.$$

$$E_4 = 50 \times 0.045 = 2.25.$$

Expected Counts Table:

# of successes (k)	0	1	2	3	4
Observed (O_k)	9	12	11	14	4
Expected (E_k)	4.25	14.5	18.5	10.5	2.25

(e)

The test statistic is:

$$\chi^2 = \sum_{k=0}^{4} \frac{(O_k - E_k)^2}{E_k}.$$

For $k = 0$:

$$\frac{(9 - 4.25)^2}{4.25} = \frac{(4.75)^2}{4.25} = \frac{22.5625}{4.25} \approx 5.307.$$

For $k = 1$:

$$\frac{(12 - 14.5)^2}{14.5} = \frac{(-2.5)^2}{14.5} = \frac{6.25}{14.5} \approx 0.431.$$

For $k = 2$:

$$\frac{(11 - 18.5)^2}{18.5} = \frac{(-7.5)^2}{18.5} = \frac{56.25}{18.5} \approx 3.041.$$

For $k = 3$:

$$\frac{(14 - 10.5)^2}{10.5} = \frac{(3.5)^2}{10.5} = \frac{12.25}{10.5} \approx 1.167.$$

For $k = 4$:

$$\frac{(4 - 2.25)^2}{2.25} = \frac{(1.75)^2}{2.25} = \frac{3.0625}{2.25} \approx 1.361.$$

$$\chi^2 = 5.307 + 0.431 + 3.041 + 1.167 + 1.361 = 11.307.$$

(f)

$$\text{d.f.} = (\text{number of categories}) - (\text{number of parameters estimated}) = 5 - 1 = 4.$$

Using the chi-squared distribution table in R for $\chi^2 = 11.307$ with 4 degrees of freedom:

$$p\text{-value} = P(\chi^2(4) \geq 11.307) \approx 0.023.$$