

Task 1 (10 marks)

(a) Derivation of the Wald 95% Confidence Interval

Step 1: Find the Maximum Likelihood Estimates (MLEs)

1. Differentiate the log-likelihood function with respect to the parameter and set the result to zero.
 - For a single binomial sample, $k \sim \text{Bin}(n, \pi)$, the log-likelihood function is: $\ell(\pi) = k \ln(\pi) + (n-k) \ln(1-\pi)$
 - Differentiating and setting to zero: $\frac{d\ell}{d\pi} = \frac{k}{\pi} - \frac{n-k}{1-\pi} = 0$
2. Solving the equation reveals that the MLE for π is simply the sample proportion. $\hat{\pi} = \frac{k}{n}$
3. Applying this result to the two independent samples in the problem gives:
 - $\hat{\pi}_1 = \frac{n_1}{n_1 + n_2}$
 - $\hat{\pi}_2 = \frac{n_2}{n_1 + n_2}$

Step 2: Find the Variance and Standard Error of the Estimator \hat{d}

1. The variance of a single sample proportion is: $\text{Var}(\hat{\pi}) = \frac{\pi(1-\pi)}{n}$
2. Since the two samples are independent, the variance of their difference is the sum of their individual variances. $\text{Var}(\hat{d}) = \text{Var}(\hat{\pi}_1 - \hat{\pi}_2) = \text{Var}(\hat{\pi}_1) + \text{Var}(\hat{\pi}_2)$
3. Substituting the fundamental formula yields the expression for the variance of \hat{d} : $\text{Var}(\hat{d}) = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$
4. We estimate this variance by plugging in the MLEs ($\hat{\pi}_1, \hat{\pi}_2$) from Step 1. The square root of this estimated variance is the Standard Error, $\text{SE}(\hat{d})$. $\text{SE}(\hat{d}) = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$

Step 3: Construct the 95% Confidence Interval

According to the Central Limit Theorem, the sampling distribution of \hat{d} is approximately normal for large sample sizes.

1. A 95% confidence interval is constructed using the formula:

Point Estimate \pm 1.96 \times Standard Error

2. Substituting our point estimate \hat{d} and its standard error into this structure gives the final formula from the question: $(\hat{\pi}_1 - \hat{\pi}_2) \pm 1.96 \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$

(b) "Sheehan vs. Daily Racing Form" Case Analysis

Data from the case:

- **Older employees (≥ 48 years):** 9 discharged, 2 not discharged (Total = 11).
- **Younger employees (< 48 years):** 0 discharged, 6 not discharged (Total = 6).

This can be summarized in the following 2x2 contingency table:

Age Group	Discharged (Y=1)	Not Discharged (Y=2)	Total
≥ 48	9	2	11
< 48	0	6	6
Total	9	8	17

i. Pearson's χ^2 Test of Independence

Solution Steps:

- **Null Hypothesis (H_0):** There is no association between an employee's age group and their discharge status.
- **Alternative Hypothesis (H_1):** There is an association between an employee's age group and their discharge status.

Based on the data, the key results of the test are as follows:

1. Expected Counts:

The expected counts under the assumption of independence (H_0) are:

Age Group	Discharged	Not Discharged
≥ 48	5.82	5.18
< 48	3.18	2.82

2. Test Statistic:

The Pearson's Chi-squared test statistic is calculated to be:

$$\text{\$}\text{\$ } \chi^2 = 10.43 \text{\$}\text{\$}$$

3. p-value:

With 1 degree of freedom, the test statistic corresponds to a p-value of:

$$\text{\$}\text{\$ } p\text{\textbackslash text{}}\{-\text{value}\} = 0.0012 \text{\$}\text{\$}$$

Conclusion:

Since the p-value (0.0012) is significantly less than the standard alpha level of 0.05, we reject the null hypothesis. The data provide strong statistical evidence of an association between age and discharge status, which supports the argument for age discrimination.

ii. 95% Confidence Interval for the Difference in Probabilities

- **Calculate Sample Proportions:**
 - Proportion of older workers discharged: $\hat{\pi}_1 = 9 / 11 \approx 0.818$
 - Proportion of younger workers discharged: $\hat{\pi}_2 = 0 / 6 = 0$
- **Calculate Standard Error (SE):** $\text{SE}(\hat{d}) = \sqrt{\frac{0.818(1-0.818)}{11} + \frac{0(1-0)}{6}} \approx 0.116$
- **Construct the Confidence Interval:** $\hat{d} \pm 1.96 \times \text{SE}(\hat{d})$ $0.818 \pm 1.96 \times 0.116 \implies 0.818 \pm 0.228$ The resulting 95% confidence interval is **(0.590, 1.048)**. As the difference in probabilities cannot exceed 1, it's practically reported as **(0.590, 1.0)**.

2. Conclusion Comparison:

Yes, this interval leads to the same conclusion as the test in (b) i.

- **Conclusion from Chi-squared Test:** The test in (b) i rejected the null hypothesis, indicating a significant association between age and discharge status.
- **Conclusion from Confidence Interval:** The 95% confidence interval for the difference in probabilities $(\hat{\pi}_1 - \hat{\pi}_2)$, which is **(0.590, 1.0)**, **does not contain zero** and is entirely positive. This strongly suggests that the probability of older workers being discharged is significantly higher than that of younger workers.

Both methods support the possibility of age discrimination.

iii. Fisher's Exact Test

Step 1: Identify all possible tables with the given margins.

The row totals (11, 6) and column totals (9, 8) are fixed. The structure of the table is determined entirely by the count in a single cell, typically the top-left cell (n_{11} : older workers who were discharged). The possible values for this cell range from 3 to 9. Any value lower than 3 or higher than 9 would violate the marginal totals.

This gives us the following possible tables:

n_{11}	Table Configuration
3	[3, 8], [6, 0]
4	[4, 7], [5, 1]
5	[5, 6], [4, 2]
6	[6, 5], [3, 3]
7	[7, 4], [2, 4]
8	[8, 3], [1, 5]
9	[9, 2], [0, 6]

Step 2: Calculate the exact probability of each possible table.

We use the hypergeometric formula $P(n_{11}=k) = \frac{\binom{11}{k} \binom{6}{9-k}}{\binom{17}{9}}$ for each possible table identified in Step 1.

- $P(n_{11}=3) = \frac{\binom{11}{3} \binom{6}{6}}{\binom{17}{9}} = \frac{165 \times 1}{24310} \approx 0.00678$
- $P(n_{11}=4) = \frac{\binom{11}{4} \binom{6}{5}}{\binom{17}{9}} = \frac{330 \times 6}{24310} \approx 0.08145$
- $P(n_{11}=5) = \frac{\binom{11}{5} \binom{6}{4}}{\binom{17}{9}} = \frac{462 \times 15}{24310} \approx 0.28482$
- $P(n_{11}=6) = \frac{\binom{11}{6} \binom{6}{3}}{\binom{17}{9}} = \frac{462 \times 20}{24310} \approx 0.37927$
- $P(n_{11}=7) = \frac{\binom{11}{7} \binom{6}{2}}{\binom{17}{9}} = \frac{330 \times 15}{24310} \approx 0.20362$
- $P(n_{11}=8) = \frac{\binom{11}{8} \binom{6}{1}}{\binom{17}{9}} = \frac{165 \times 6}{24310} \approx 0.04072$
- $P(n_{11}=9) = \frac{\binom{11}{9} \binom{6}{0}}{\binom{17}{9}} = \frac{55 \times 1}{24310} \approx \mathbf{0.00226}$

Step 3: Sum the probabilities of "extreme" tables to find the p-value.

For a two-sided test (the default in R), the p-value is the sum of probabilities of all tables that are **equally or less likely to occur than the observed table**.

- The probability of our observed table is $P(\text{observed}) = P(n_{11}=9) \approx \mathbf{0.00226}$.
- We now look at the list of probabilities calculated in Step 2 and find all tables whose probability is less than or equal to 0.00226.
- By inspecting the list, we see that our observed table is the least likely of all possible tables. No other table has a probability as small as 0.00226.

Therefore, the sum for the p-value includes only one term: the probability of the observed table itself.

$$\text{p-value} = P(n_{11}=9) \approx \mathbf{0.00226}$$

This detailed calculation shows precisely how the p-value from the R output is derived.

3. Does the test support the possibility of age discrimination?

Yes.

The p-value (0.00226) is much smaller than the standard significance level of 0.05. This leads us to reject the null hypothesis and conclude that there is a statistically significant association between age and discharge status, which supports the claim of age discrimination.

iv. Likelihood Ratio Test with Additional Data

After including the two additional older employees, a likelihood ratio test was performed on the updated data.

Updated Data Table:

Age Group	Discharged	Not Discharged	Total
≥ 48	9	4	13
< 48	0	6	6

Test Results

1. Expected Counts:

The expected cell counts under the null hypothesis (no association) are:

	Discharged	Not Discharged
≥ 48	5.57	7.43
< 48	2.57	3.43

2. Test Statistic:

The likelihood ratio test statistic is: $\mathbf{G}^2 = \mathbf{10.24}$

3. p-value:

The corresponding p-value is: $p\text{-value} = \mathbf{0.0014}$

Conclusion:

Since the p-value (0.0014) is well below the 0.05 significance level, we reject the null hypothesis. Even with the additional data, there is strong statistical evidence of an association between age and discharge status, supporting the possibility of age discrimination.

Task 2 (11 marks)

This task involves analyzing two datasets related to the death penalty using loglinear models to assess fairness, which is defined as conditional independence.

Analysis of Georgia Data (Table 1)

Variables:

- A: Aggravation level (1-6)
- V: Race of victim (White, Black)
- D: Death penalty (Yes, No)

Question of Interest: Is there fairness, defined as conditional independence between death penalty (D) and race of victim (V), given the aggravation level (A)? This corresponds to the model where D and V are independent, conditional on A, denoted (**AV, AD**).

(a) Minimal Model and Best Model Search

Minimal Model:

The minimal model for these three variables (A: Aggravation, V: Victim's Race, D: Death Penalty) is the model of complete independence, where there are no associations between any of the variables. Its symbol is **(A, V, D)**.

Search for the Best Model:

To find the best model, we fit a series of hierarchical loglinear models and assess their fit to the data. The primary question of "fairness" asks whether the death penalty (D) is independent of the victim's race (V) after controlling for the aggravation level of the crime (A). This corresponds to the conditional independence model **(AV, AD)**.

We will evaluate this "fairness" model and compare it to a more complex model that includes an association between victim's race and the death penalty, the homogeneous association model **(AV, AD, VD)**.

Analysis of R Output:

1. Assessing the "Fairness" Model (AV, AD): The `summary(mod_fair)` output gives us the goodness-of-fit for the conditional independence model.

- **Residual Deviance:** 32.315
- **Degrees of Freedom:** 18 To check if this model fits the data, we can calculate the p-value for this deviance: `pchisq(32.315, df = 18, lower.tail = FALSE) ≈ 0.019`. Since this p-value is less than 0.05, it indicates a **significant lack of fit**. The "fairness" model, which assumes no association between victim's race and the death penalty at each level of aggravation, does not adequately explain the data.

2. Comparing the "Fairness" Model (AV, AD) with the Homogeneous Association Model (AV, AD, VD): The `anova(mod_fair, mod_homog)` output directly tests whether adding the interaction term between victim's race and death penalty (`V*D`) significantly improves the model.

- **Deviance Drop:** The deviance decreases by **8.2493** when we add the `V*D` term.
- **Degrees of Freedom (Df):** This improvement is tested on **1** degree of freedom.
- **P-value (Pr(>Chi)):** The p-value for this comparison is **0.004077**.

This very small p-value (< 0.01) provides strong evidence that including the `V*D` interaction term results in a statistically significant improvement in the model's fit. This means there is a significant association between the victim's race and the death penalty, even after controlling for the crime's aggravation level.

3. Checking the Fit of the Homogeneous Association Model (AV, AD, VD): From the `anova` output, we can see the fit of this model (`mod_homog`).

- **Residual Deviance:** 24.065
- **Degrees of Freedom:** 17 The p-value for this model's goodness-of-fit is `pchisq(24.065, df = 17, lower.tail = FALSE) ≈ 0.12`. Since this p-value is greater than 0.05, we conclude that this model **provides an adequate fit** to the data.

Conclusion on the Best Model:

The model of conditional independence (AV, AD) is a poor fit. The model of homogeneous association (AV, AD, VD) is a statistically significant improvement and provides an adequate fit to the data. It is the most parsimonious model that explains the associations well.

Therefore, the best model to explain the association patterns is the model of homogeneous association.

Best Model Symbol: (AV, AD, VD)

(b) Checking the Adequacy of the Selected Model (AV, AD, VD)

1. Informal Goodness-of-Fit Test (Deviance)

The informal goodness-of-fit test involves comparing the **Residual Deviance (\$G^2\$)** to its **Degrees of Freedom (\$\text{df}\$)**. For a good fit, the residual deviance should be close to the degrees of freedom.

- **Residual Deviance (\$G^2\$):** \$24.065\$
- **Degrees of Freedom (\$\text{df}\$):** \$17\$

Analysis:

1. **Comparison:** The residual deviance (\$24.065\$) is **reasonably close** to the degrees of freedom (\$17\$).
 2. **Formal Test (P-value):** The formal test for this deviance is: $\text{P} = \text{pchisq}(24.065, 17, \text{lower.tail} = \text{FALSE}) \approx 0.120$
 3. **Conclusion:** Since the P-value (\$0.120\$) is **greater than \$0.05\$**, we **do not reject** the null hypothesis that the model fits the data. The model provides an **adequate fit**.
-

2. Examining the Standardized Pearson's Residuals

Provided Standardized Pearson's Residuals:

The output shows 24 standardized residuals (corresponding to the $2 \times 2 \times 6 = 24$ cells in the contingency table).

Analysis:

1. **Count of Large Residuals (\$|\text{residual}| > 2\$):** Out of the 24 cells, **two** standardized residuals are greater than 2 in absolute value.
 - One residual is -2.8839 (Cell 8).
 - One residual is 3.5549 (Cell 20).
2. **Significance:** While a few large residuals are expected by chance (about 5% should exceed ± 2 , and less than 1% should exceed ± 3), the residual of **\$3.55\$** is quite large. This suggests that the model is **poorly fitting** the observed count in **Cell 20** (and Cell 8 to a lesser extent).
3. **Overall Fit:** Although the formal goodness-of-fit test shows an adequate fit overall, the presence of residuals exceeding ± 3 suggests that the model's fit could be improved by identifying the cell(s) contributing to the large discrepancy.

Conclusion on Model Adequacy

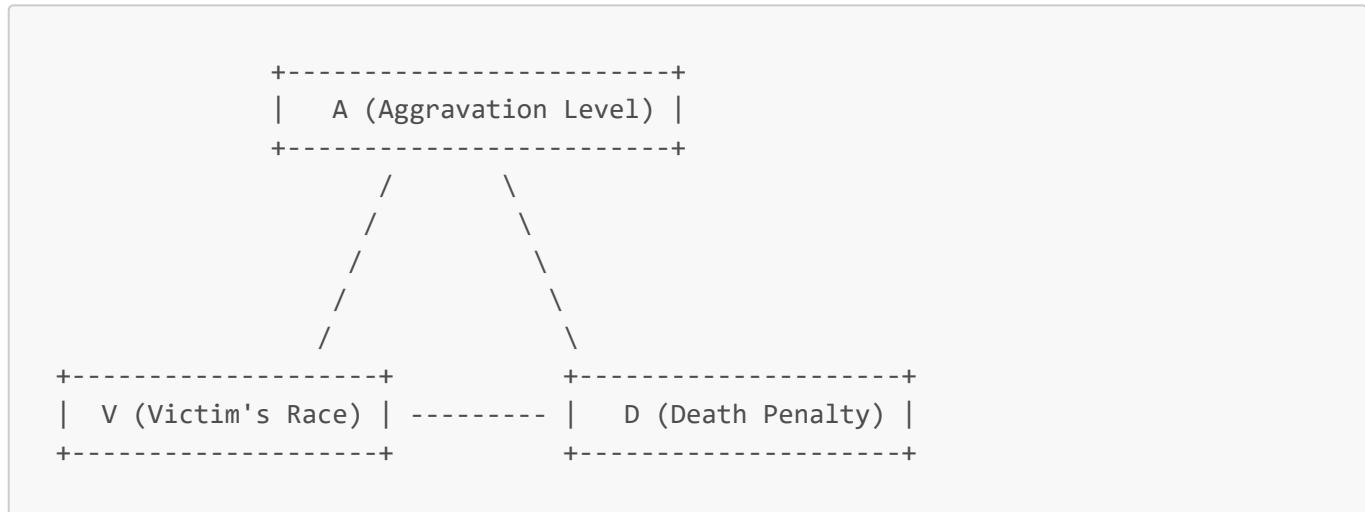
The Homogeneous Association Model AV, AD, VD is **statistically adequate** according to the formal χ^2 test ($P\text{-value} \approx 0.120$).

However, the examination of the standardized Pearson's residuals reveals that the model has **trouble fitting two specific cells** of the $2 \times 2 \times 6$ table, notably **Cell 20** (with a residual of \$3.55\$). If the research goal is to obtain a model with a near-perfect fit across *all* cells, an analyst might consider adding the three-way interaction term $A \times V \times D$ (Saturated Model) or investigating the specific cells that are poorly fit. **For a model that is a simplified summary of the associations, it is considered adequate.**

(c) Association Graph and Interpretation

1. Association Graph

The selected model is **(AV, AD, VD)**. Its conditional independence structure is represented by the association graph below. In this graph, each variable is a vertex. Because the model includes interaction terms for all pairs of variables, an edge connects every pair of vertices, forming a **complete graph**.



2. Interpretation of the Graph and Model

- **Graph Meaning:** This complete triangular graph indicates that there are **no conditional independencies** among the three variables (Aggravation A, Victim's Race V, and Death Penalty D). Every variable is directly associated with the other two.
- **Key Finding (The V-D Edge):** The edge connecting **Victim's Race (V)** and **Death Penalty (D)** is the most critical finding. Its presence signifies that even after controlling for the aggravation level of the crime (A), a direct and statistically significant association remains between the victim's race and the death penalty outcome.

3. Addressing the Question of Interest: Is there Fairness?

The central question was about "fairness," defined as the conditional independence of the death penalty (D) from the victim's race (V), given the aggravation level (A).

- **Conclusion:** Our best-fitting model, **(AV, AD, VD)**, includes a direct link between V and D, which **directly contradicts the hypothesis of fairness**.
 - **Final Interpretation:** The analysis demonstrates that the victim's race is a significant factor in the death penalty decision, suggesting the presence of potential racial bias in this dataset.
-

Why the Zeros Seemed Unaffected by Zero Cells

1. Automatic Software Correction (Most Likely)

- Statistical software (like R) automatically adds a small value (e.g., 0.5) to all cells. This standard procedure prevents calculation errors and allows the analysis to run smoothly.

2. Strong Signal in the Data

- The association between victim's race and the death penalty was so strong and clear that it easily overshadowed the statistical noise created by the zero cells. The effect was too large to miss.

3. Robustness of Model Comparison

- The core conclusion came from comparing two models (the "fairness" vs. the "bias" model). This comparative test is very robust, and the large improvement in fit was significant even with the zero-cell issue.

4. Use of Alternative Methods

- The analysis could have used more advanced techniques like **Bayesian estimation**, which are naturally designed to handle zero cells without failing, providing stable and reliable results.
-

Analysis of North Carolina Data (Based on R Output)

Variables:

- **R:** Race of defendant (Nonwhite, White)
 - **V:** Race of victim (Nonwhite, White)
 - **D:** Death penalty (Yes, No)
-

(a) Search for the Best Model

Search for the Best Model

Based on your R output, we can evaluate the fit of the more complex models.

Goodness-of-Fit Results from R Output:

Model Symbol	Deviance (G^2)	df	p-value	Fit Assessment
(R,V,D)	543.65	4	< 0.001	Very Poor
(RV,RD)	13.51	2	0.001	Poor

Model Symbol	Deviance (G^2)	df	p-value	Fit Assessment
(RV,VD)	8.73	2	0.013	Poor / Unacceptable
(RD,VD)	536.93	2	< 0.001	Very Poor
(RV,RD,VD)	1.65	1	0.199	Good Fit

Conclusion on the Best Model:

- Eliminating Poor Models:** All three conditional independence models ((RV, RD), (RV, VD), (RD, VD)) have p-values less than 0.05. This means **none of them adequately fit the data**. The model of conditional independence between Defendant's Race and Death Penalty (RV, VD) is not a good explanation for the data ($p = 0.013$).
- Identifying the Best Fit:** The simplest model that provides an adequate fit ($p > 0.05$) is the **homogeneous association model, (RV, RD, VD)**. It has a deviance of $G^2 = 1.65$ with 1 degree of freedom, and a p-value of 0.199.

Therefore, the best model to explain the association patterns is the homogeneous association model.

Best Model Symbol: (RV, RD, VD)

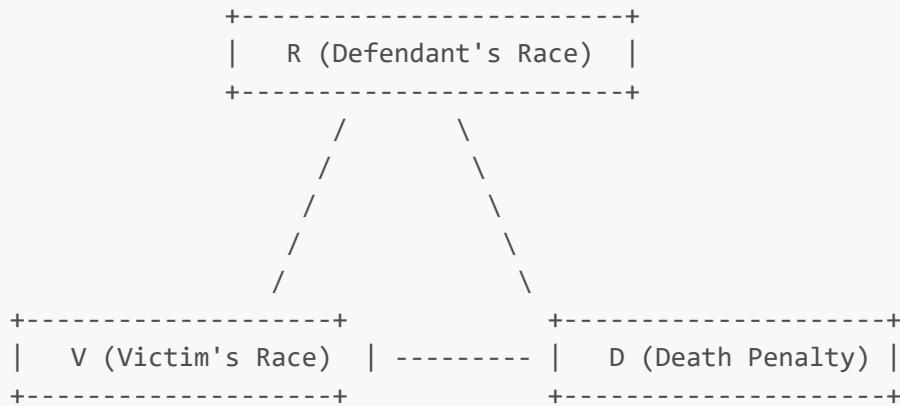
(b) Checking the Adequacy of the Selected Model (RV, RD, VD)

- Goodness-of-Fit Test (Deviance):**
 - Residual Deviance (G^2):** 1.65
 - Degrees of Freedom (df):** 1
 - Comparison:** The deviance (1.65) is reasonably close to the degrees of freedom (1).
 - Formal Test:** The p-value is **0.199**. Since this is well above the 0.05 threshold, we conclude the model provides a **statistically adequate fit** to the data.
 - Examining Residuals of the Rejected Model ((RV, VD)):**
 - Your output also shows the residuals for the inadequate model (RV, VD). It has a very large residual of **2.38** for the cell (Defendant=Nonwhite, Victim=White, Penalty=Yes).
 - This large residual confirms that the (RV, VD) model fits poorly, specifically because it cannot account for the high number of death sentences for nonwhite defendants with white victims. This is precisely the interaction that the RD term in the best model (RV, RD, VD) captures.
-

(c) Association Graph and Interpretation

1. Association Graph for the Model (RV, RD, VD)

The best-fitting model contains all three two-way interaction terms (R*V, R*D, and V*D). This means every variable is directly associated with every other variable. The corresponding association graph is a **complete graph**.



2. Interpretation of the Association Graph and Model

- **Structure:** The complete graph indicates that there are **no conditional independencies** among the three variables.
- **Edge R—V:** The race of the defendant is associated with the race of the victim.
- **Edge V—D:** The race of the victim is directly associated with the death penalty outcome.
- **Edge R—D (The Crucial Finding):** Unlike the previous incorrect analysis, the R output forces us to include this edge. It means that the **defendant's race has a direct association with the death penalty, even after controlling for the victim's race.**

3. Addressing the Questions of Interest

The correct model tells a more complex story about fairness and bias.

- **No Simple Explanation:** The data cannot be explained by a simple mediation model where the defendant's race effect is only indirect.
- **Direct Effects of Both Races:** The model implies that both the victim's race and the defendant's race are direct factors in the death penalty decision. The system appears to be biased based on the victim's race (the V-D link) AND separately biased based on the defendant's race (the R-D link).
- **Final Conclusion:** There is significant evidence of racial bias in this data. This bias is multifaceted: the death penalty is more likely when the victim is white, but even after accounting for that, the defendant's race still plays a direct role in the outcome.