Problem 1a) - Hypothesis Testing and Confidence Intervals

Data: 9 successes in 10 trials, $H_0 : \pi = 0.5$

i) Tests and Confidence Intervals

1. **Wald Test**:
   - Statistic: $Z = \frac{\hat{\pi} - 0.5}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} = \frac{0.9 - 0.5}{\sqrt{0.9 \times 0.1/10}} = 4.216$
   - 95% CI: $\hat{\pi} \pm z_{0.975}\sqrt{\hat{\pi}(1-\hat{\pi})/n} = [0.714, 1.086]$
2. **Score Test**:
   - Statistic: $Z = \frac{\hat{\pi} - 0.5}{\sqrt{0.5 \times 0.5/10}} = 2.529$
   - 95% CI: Solve $\frac{|\hat{\pi} - \pi|}{\sqrt{\pi(1-\pi)/10}} \leq 1.96 \rightarrow [0.595, 0.982]$
3. **Likelihood Ratio Test**:
   - Statistic: $-2\ln(\Lambda) = 2\left[\ell(\hat{\pi}) - \ell(0.5)\right] = 5.796$
   - 95% CI: $\{\pi : -2\ln(\Lambda) \leq 3.841\} \approx [0.634, 0.985]$

ii) Exact p-values

- **One-sided p-value**: $P(X \geq 9) = \sum_{k=9}^{10} \binom{10}{k} 0.5^{10} = 0.0107$
- **Two-sided p-value**: $2 \times 0.0107 = 0.0214$

Problem 1b) - Relative Risk and CI

Contingency Table:

Race	Yes	No/Undecided	Total
White	31	12	43
Black	9	4	13

- **Relative Risk (RR)**:

$$RR = \frac{31/43}{9/13} = 1.03$$

- **95% CI for RR**:

$$\ln(RR) \pm 1.96\sqrt{\frac{12}{31 \times 43} + \frac{4}{9 \times 13}} \rightarrow [0.57, 1.86]$$

Problem 2a) - Contingency Table

HAART Data:

- Total patients: 1464
- Females: 63.52% (930)
- Defaulters: 22.61% (331), including 142 males

Contingency Table:

	Default	Non-Default	Total
Male	142	392	534
Female	189	741	930
Total	331	1133	1464

Joint/Marginal Distributions:

- $P(\text{Male}, \text{Default}) = 142/1464 = 0.097$
- $P(\text{Female}) = 930/1464 = 0.635$

Problem 2b) - Probabilities

1. $P(\text{Female} \cap \text{Default}) = 189/1464 = 0.129$
2. $P(\text{Male}) = 534/1464 = 0.365$
3. $P(\text{Default} \mid \text{Female}) = 189/930 = 0.203$

Problem 3) - Logistic Model for Case-Control Study

Data:

	Lung Cancer	Control
Smoking	452	215
Non-Smoking	548	785

(i) Logistic Model:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \text{Smoking}$$

- **Fit**: $\hat{\beta}_0 = -0.366, \hat{\beta}_1 = 0.785$

(ii) Probabilities:

- Smokers: $p = \frac{e^{-0.366+0.785}}{1+e^{-0.366+0.785}} = 0.678$
- Non-smokers: $p = \frac{e^{-0.366}}{1+e^{-0.366}} = 0.410$

(iii) Log-Odds:

- Smokers: $-0.366 + 0.785 = 0.419$
- Non-smokers: -0.366

(iv) Odds Ratio:

$$OR = e^{0.785} = 2.$$

Problem 4) - Logistic Regression for SBP

Data:

Age (x_i)	High SBP (y_i)	Total (n_i)
30	7	32
35	4	16
40	10	25
45	14	32

(i) Logistic Model:

$$\log\left(\frac{p_i}{1-p_i}\right) = -2.30 + 0.056\text{Age}$$

(ii) Predicted Proportions:

- Age 30: $\hat{p} = 0.219$
- Age 45: $\hat{p} = 0.437$

(iii) Odds Ratio (5-year increase):

$$OR = e^{5\times 0.056} = 1.32$$

Problem 5) - Multiple Logistic Regression

Data:

Gestational Age (x_1)	Weight Gain (x_2)	Normal Birth Weight (y)	Total (n)
30	14	15	25
32	8	18	35
34	12	6	15
36	11	4	16
38	10	1	8

(i) Model:

$$\log\left(\frac{p}{1-p}\right) = 1.92 - 0.12x_1 + 0.08x_2$$

(ii) Prediction:

For $x_1 = 31, x_2 = 9$:

$$\hat{p} = \frac{e^{1.92-0.12\times 31+0.08\times 9}}{1+e^{1.92-3.72+0.72}} = 0.324$$

Problem 6) - Three-Way Table Analysis

Death Penalty Data:

Victim	Defendant	Yes	No	Total
White	White	2	13	15
	Black	1	5	6
Black	White	0	5	5
	Black	1	15	16

(a) Conditional Odds Ratios (with +0.5 adjustment):

- White victims: $OR = \frac{2.5 \times 5.5}{13.5 \times 1.5} = 0.68$
- Black victims: $OR = \frac{0.5 \times 15.5}{5.5 \times 1.5} = 0.94$

(b) Marginal OR:

$$OR = \frac{3 \times 20}{1 \times 18} = 3.33$$

Conclusion: Marginal OR suggests bias, but conditional ORs (stratified by victim race) do not.

(c) Cochran-Mantel-Haenszel Test:

$$\chi^2_{CMH} = 0.24 \quad (p = 0.624)$$

→ Fail to reject independence.

Problem 7) - Loglinear Models

(a) Model (XY, XZ, YZ):

- Holds iff X and Y have homogeneous association across Z levels (i.e., $\theta_{XY(k)} = \theta_{XY}$ for all k).

(b) Model (XZ, YZ):

- Holds iff $X \perp\!\!\!\perp Y \mid Z$ (conditional independence).