z

REVISION

1)(a) A clinical trial is conducted to evaluate a new treatment. This experiment has nine successes in the first 10 trials.

i) Test the hypothesis of $H_0: \pi = 0.5$ using the three methods (Wald, Score and Likelihood ratio) and construct the corresponding confidence intervals.

ii) Find the exact one sided and two-sided p-values.

b) The following table was taken from the General Social Survey:

Belief in Afterlife			
Race	Yes	No or Undecided	Total
White	31	12	43
Black	9	4	13

Find the relative risk and the 95% approximate confidence interval for the true value of the relative risk.
($z_{0.95} = 1.645$; $z_{0.975} = 1.96$)

2) a) In the HAART Data used by Seid et al. (2014), there are 1464 patients. Of these 63.52 % are females. 22.61% of these patients were defaulted including 142 males.

(i). Construct the contingency table. (ii). Find the joint and marginal distributions.

b) If a patient is selected at random, what is the probability that the patient is

(i) a female and defaulter? (ii) a male? (iii) defaulter if the member is female?

c) Suppose we have a 2 × 2 contingency table where the notation is defined as follows:

	Column 1	Column 2	Total
Row 1	n_{11}	n_{12}	n_{1+}
Row 2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	N

Show that the odds ratio may be expressed only in terms of the (1,1)th cell frequency n_{11} and the marginal frequencies by $\theta = \frac{n_{11}(n1+n_{11}-(n_{1+}+n_{+1}))}{(n_{1+}-n_{11})(n_{+1}-n_{11})}$

3) We consider a case-control study, in which 1000 males are in case group having lung cancer and 1000 males are in control group who do not have lung cancer. They were asked about whether they were exposed to smoking or not. The data are given as follows:

	Lung Cancer (cases)	No Lung Cancer (Control)
Smoking	452	215
Non-Smoking	548	785
Total	1000	1000

(i) Fit a logistic model

(ii) Determine the probability of having lung cancer in both groups

(iii) Compute log-odds for smokers and non-smokers

z

(iv) Obtain the odds ratio for lung cancer

4) A hypothetical study of 105 patients was conducted to investigate the effect of age on systolic blood pressure (1 for high SBP and 0 for normal SBP). Data on the age and systolic blood pressure are recorded in Table below. solve the following

(i) Fit a logistic model between SBP and age

(ii) Obtain the predicted values of proportion for the given values of age in the data

(iii) Calculate the odds ratio for 5 years increase in age

S/N	Age (x_i)	Number if Patient having high SBP (y_i)	Total Number of Patients (n_i)
1	30	7	32
2	35	4	16
3	40	10	25
4	45	14	32

5) The following table provides the information of birth weight of neonate (1 for normal birth weight and 0 low birth weight), gestational age of fetus and increase in mother's weight during pregnancy:

S/N	Gestational Age (x_1)	Increase in Mother's weight (x_2)	Number of Neonates having Normal birth weight (y)	Total Number of Neonates (n)
1	30	14	15	25
2	32	8	18	35
3	34	12	6	15
4	36	11	4	16
5	38	10	1	8
Total	170	55	44	100

(i) Fit the multiple regression model to check the relationship of birth weight with the other variables given in above data.

(ii) By method of matrices, estimate the proportion of neonates having normal birth weight if gestational age is 31 years and increase in mother's weight is 9 kg.

6) 4) Suppose we have the following three-way contingency table:

Victim's	Defendant's	Death Penalty		
Race	Race	Yes	No	Total
White	White	2	13	15
	Black	1	5	6
	Total	3	18	21
Black	White	0	5	5

z

	Black	1	15	16
	Total	1	20	21

a) Compute and interpret the sample conditional odds ratios, adding 0.5 to each cell to reduce the impact of the 0 cell count.

b) Compute and interpret the sample marginal odds ratio between defendant's race and the death penalty verdict. Compare the conclusion made here with those made in (a).

c) Test the hypothesis that death penalty is independent of defendant's race, controlling for victim's race using the Cochran-Mantel-Haenszel test with α = 5%. ($\chi_1^2(0.95) = 3.8415$, $\chi_2^2(0.95) = 5.9915$; $\chi_3^2(0.95) = 7.8147$)

7) 5) Consider a three-way contingency table with categorical variables *X* having 2 categories, *Y* having 2 categories and *Z* having $K \geq 2$ categories. (Hint: To show "A if and only if B", you need show both "A implies B" and "B implies A")

a) Show that the loglinear model (*XY, XZ, Y Z*) holds if and only if *X* and *Y* have homogeneous association controlling for *Z*.

b) Show that the loglinear model (*XZ, Y Z*) holds if and only if *X* and *Y* are conditionally independent controlling for *Z*.