Assignment 3 - Data Analysis

Author : Gayi Komi Selassi
ID : RA2422021010003
Program : MSc Epidemiology & Biostatistics
Portfolio: Click Here
Access the GitHub Repository: Click Here

1. Frequency tables for Age_Group and Heart_Disease :

The FREQ Procedure

Age Group

Age_Group	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Middle-Aged	696	67.90	696	67.90
Senior	272	26.54	968	94.44
Young	57	5.56	1025	100.00

Heart Disease

Heart_Disease	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	499	48.68	499	48.68
Yes	526	51.32	1025	100.00

2. Average cholesterol (chol) and resting blood pressure (trestbps) grouped by Heart_Disease :

The MEANS Procedure

Heart_Disease	*N Obs*	*Variable*	*Label*	*Mean*
No	499	chol	Serum Cholesterol (mg/dL)	251.29
		trestbps	Resting Blood Pressure (mm Hg)	134.11
Yes	526	chol	Serum Cholesterol (mg/dL)	240.98
		trestbps	Resting Blood Pressure (mm Hg)	129.25

COMMENT:

The average cholesterol (chol) was slightly higher in patients without heart disease (251.29 mg/dL) compared to those with heart disease (240.98 mg/dL).

The average resting blood pressure (trestbps) was also higher in patients without heart disease (134.11 mm Hg) than in those with heart disease (129.25 mm Hg).

This result may appear counterintuitive, but it could be explained by treatment effects or other confounding factors (e.g., age, sex, lifestyle, or medication use).

Therefore, cholesterol and blood pressure alone are not sufficient predictors of heart disease in this dataset, and further statistical testing or multivariate analysis is recommended.

Exploratory Analysis of the HEART Dataset: Assignment 3

3. Relationship between sex and Heart_Disease:

The FREQ Procedure

Frequency	Table of sex by Heart_Disease		
sex(0 = Female, 1 = Male)	Heart_Disease		
	No	Yes	Total
0	86	226	312
1	413	300	713
Total	499	526	1025

Statistics for Table of sex by Heart_Disease

Statistic	DF	Value	Prob
Chi-Square	1	80.0737	<.0001
Likelihood Ratio Chi-Square	1	82.3927	<.0001
Continuity Adj. Chi-Square	1	78.8631	<.0001
Mantel-Haenszel Chi-Square	1	79.9956	<.0001
Phi Coefficient		-0.2795	
Contingency Coefficient		0.2692	
Cramer's V		-0.2795	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	86
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1.0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Sample Size = 1025

COMMENT:

Among females (sex=0), 86 had no heart disease while 226 had heart disease.

Among males (sex=1), 413 had no heart disease while 300 had heart disease.

The Chi-Square test shows a highly significant association between sex and heart disease (Chi-Square=80.07, p<0.0001).

This suggests that sex is strongly related to the prevalence of heart disease in this dataset.

Interpretation: Females have proportionally more cases of heart disease compared to males, even though the absolute number of male patients is higher.

Effect size measures (Phi = -0.28, Cramer's V = -0.28) indicate a moderate strength of association.

4. Top 5 oldest patients with heart disease with some variables:

Line_Number	*Age (years)*	*Sex (0 = Female, 1 = Male)*	*Age_Group*	*Heart_Disease*	*Resting Blood Pressure (mm Hg)*	*Serum Cholesterol (mg/dL)*	*Resting Electrocardiographic Results*	*Thalassemia (3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect)*
536	76	0	Senior	Yes	140	197	2	2
100	76	0	Senior	Yes	140	197	2	2
966	76	0	Senior	Yes	140	197	2	2
591	74	0	Senior	Yes	120	269	0	2
725	74	0	Senior	Yes	120	269	0	2

GitHub Repository: Click Here
Portfolio: Click Here