STATISTICAL METHODS IN BIOLOGY

Malaika Fatima (SP21.BSI.030)

DECEMBER 26, 2024

DESCRIPTIVE TESTS:

We compare diabetic vs non-diabetic datasets means heathy vs diseases .

Table 1: Illustrate the descriptive statistics tests result from spss.

Descriptive Statistics				
	Minimum	Maximum	Mean	Std. Deviation
Age	22.00	52.00	34.2000	9.37859
Blood Pressure	106.00	155.00	128.6500	15.16323
Cholesterol	170.00	235.00	203.1000	20.85766
Glucose	78.00	110.00	93.4500	9.54477
BMI	56.27	70.99	65.6978	4.61471

Here's a brief interpretation of the descriptive statistics provided:

1. **Age**:
    - The age of individuals in the sample ranges from 22 to 52 years.
    - The average age is 34.2 years, with a standard deviation of 9.38 years, indicating a moderate spread of ages around the mean.
2. **Blood Pressure**:
    - The blood pressure values range from 106 to 155 mmHg.
    - The mean blood pressure is 128.65 mmHg, with a standard deviation of 15.16 mmHg, suggesting moderate variability in blood pressure among the individuals.
3. **Cholesterol**:
    - Cholesterol levels range from 170 to 235 mg/dL.
    - The mean cholesterol level is 203.1 mg/dL, with a standard deviation of 20.86 mg/dL, indicating a moderate spread in cholesterol values.
4. **Glucose**:
    - Glucose levels range from 78 to 110 mg/dL.
    - The mean glucose level is 93.45 mg/dL, with a standard deviation of 9.54 mg/dL, showing a moderate variability in glucose levels.
5. **BMI (Body Mass Index)**:
    - BMI ranges from 56.27 to 70.99.

- The average BMI is 65.7, with a standard deviation of 4.61, suggesting that most individuals in the sample have a BMI closer to the higher end of the scale, but there's still some variability.

Table 2: Illustrate the results of tests of normality.

Tests of Normality

	Gender	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Blood Pressure	0	0.182	8	.200*	0.896	8	0.268
	1	0.174	12	.200*	0.887	12	0.108
Cholesterol	0	0.284	8	0.056	0.770	8	0.013
	1	0.104	12	.200*	0.950	12	0.643
Glucose	0	0.181	8	.200*	0.966	8	0.862
	1	0.185	12	.200*	0.927	12	0.347
BMI	0	0.369	8	0.002	0.670	8	0.001
	1	0.139	12	.200*	0.957	12	0.744
Physical Activity	0	0.235	8	.200*	0.802	8	0.030
	1	0.270	12	0.016	0.746	12	0.002

Here is the interpretation with specific values included:

- **Blood Pressure**:
  - Females (0): KS (p = .200*), SW (p = 0.268) → Normally distributed.
  - Males (1): KS (p = .200*), SW (p = 0.108) → Normally distributed.
- **Cholesterol**:
  - Females (0): KS (p = 0.056), SW (p = 0.013) → Not normally distributed
  - (SW $p \leq 0.05$).
  - Males (1): KS (p = .200*), SW (p = 0.643) → Normally distributed.
- **Glucose**:
  - Females (0): KS (p = .200*), SW (p = 0.862) → Normally distributed.
  - Males (1): KS (p = .200*), SW (p = 0.347) → Normally distributed.
- **BMI**:
  - Females (0): KS (p = 0.002), SW (p = 0.001) → Not normally distributed.
  - Males (1): KS (p = .200*), SW (p = 0.744) → Normally distributed.
- **Physical Activity**:
  - Females (0): KS (p = .200*), SW (p = 0.030) → Not normally distributed (SW $p \leq 0.05$).
  - Males (1): KS (p = 0.016), SW (p = 0.002) → Not normally distributed.

Summary:

- Normally distributed: Blood Pressure (both genders), Cholesterol (males), Glucose (both genders), BMI (males).
- Not normally distributed: Cholesterol (females), BMI (females), Physical Activity (both genders).

Pie charts:

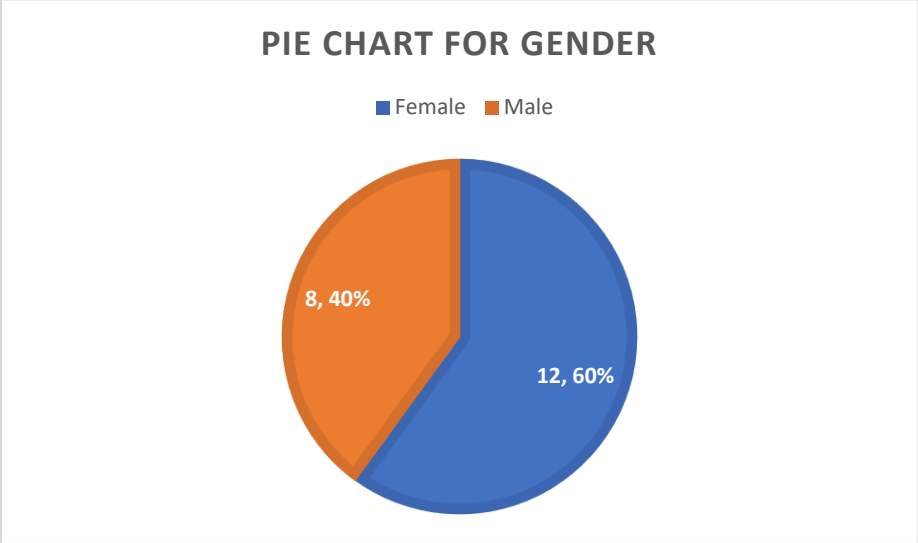

FIGURE: Illustrate the gender distribution pie charts

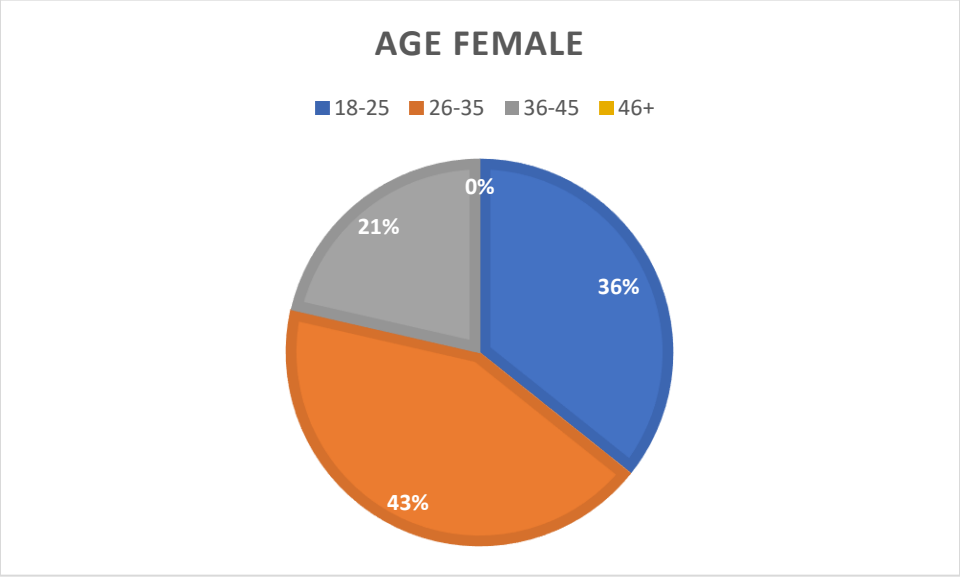

FIGURE: Illustrate the AGE distribution of Female pie charts

Histograms:

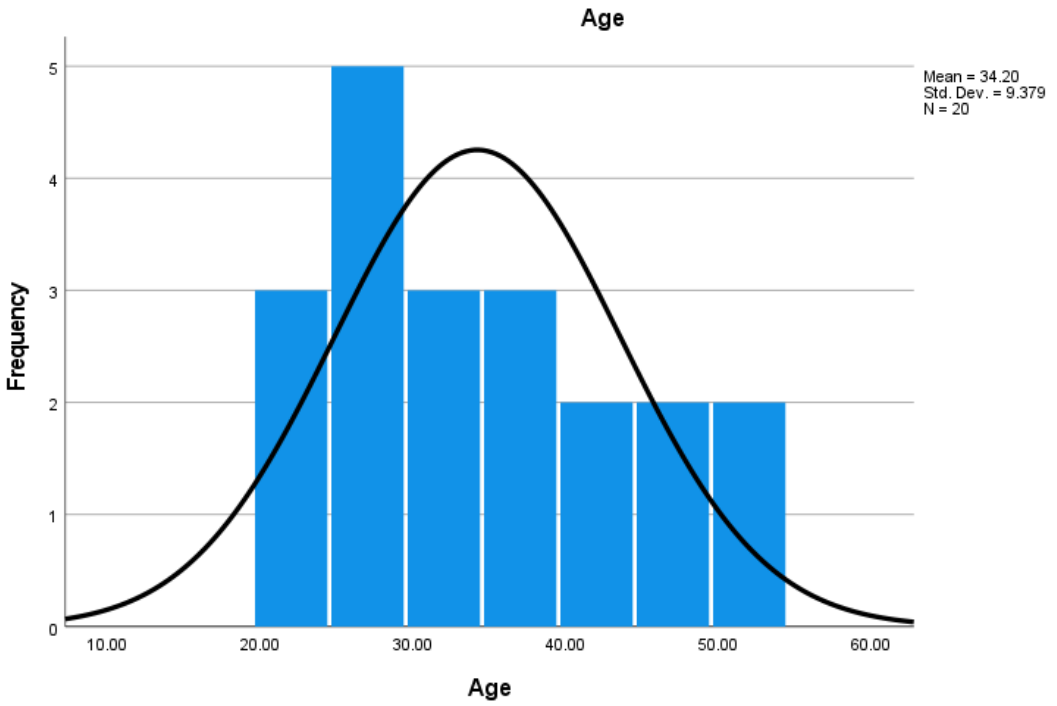

Figure: Illustrate histogram of age variable

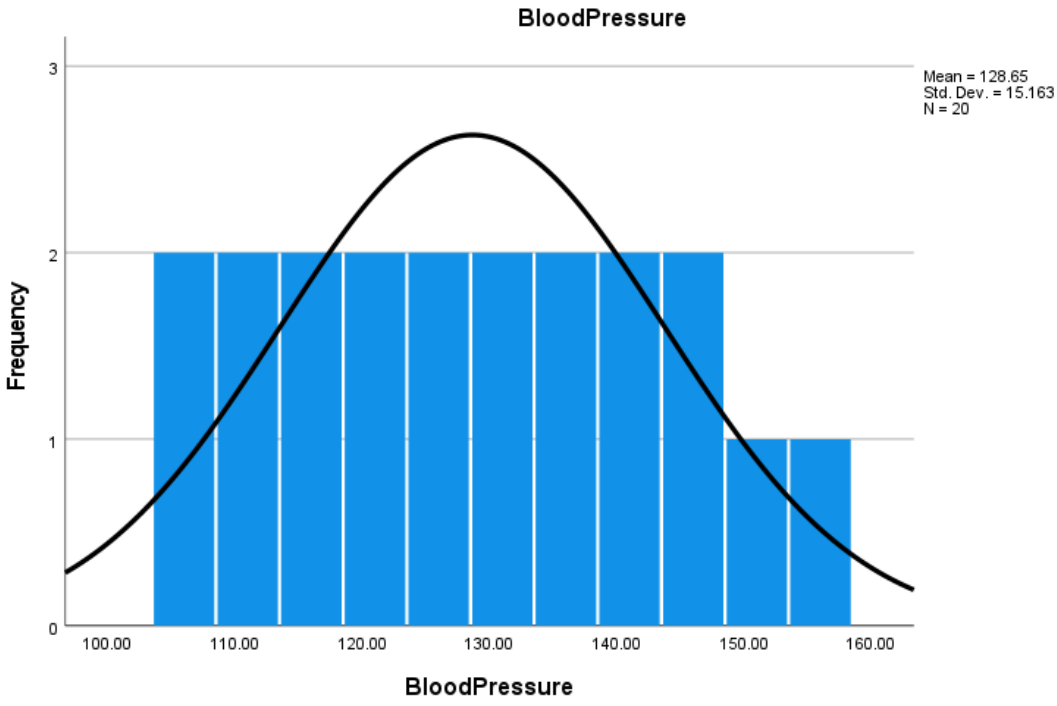

Figure: Illustrate histogram of Blood pressure variable

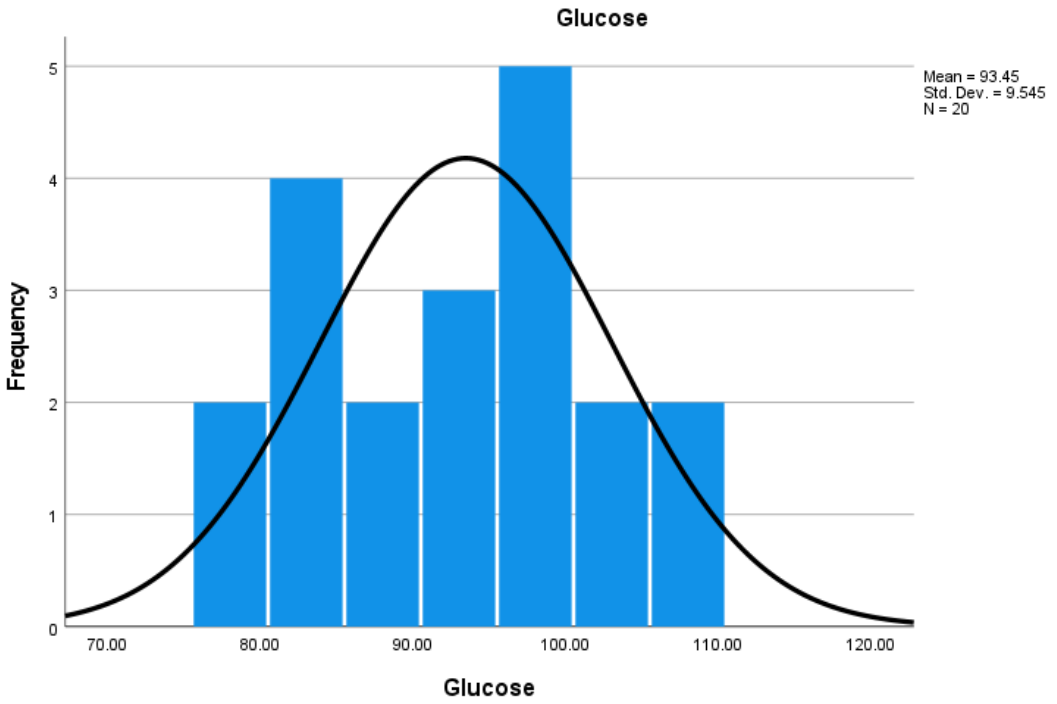

Figure: Illustrate histogram of Glucose variable

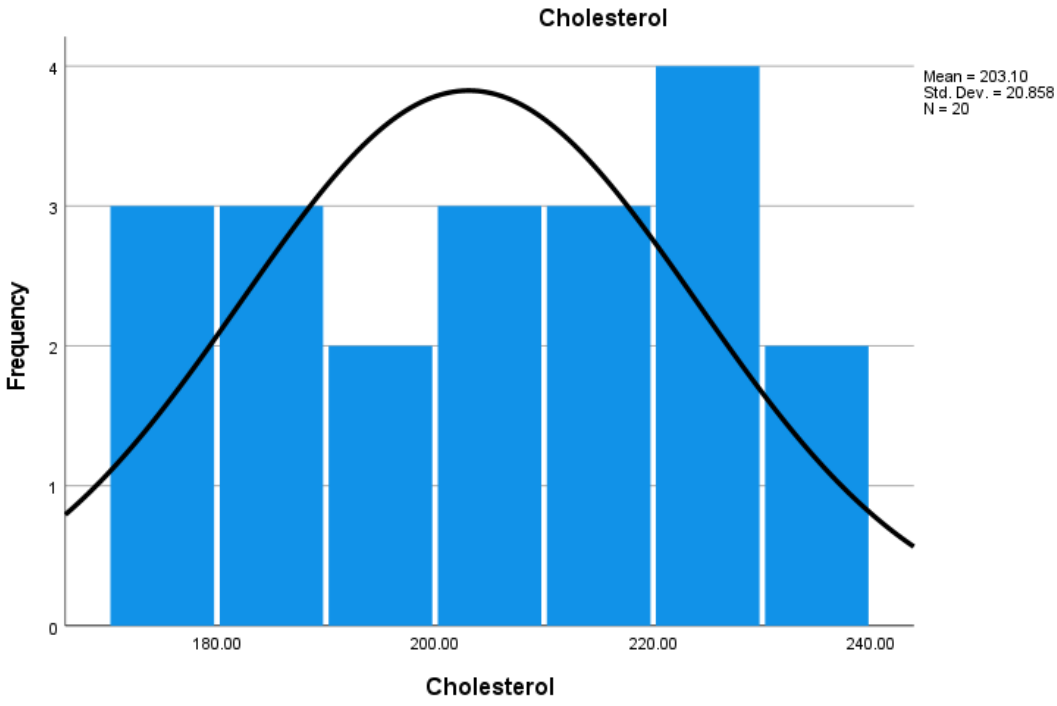

Figure: Illustrate histogram of cholesterol variable

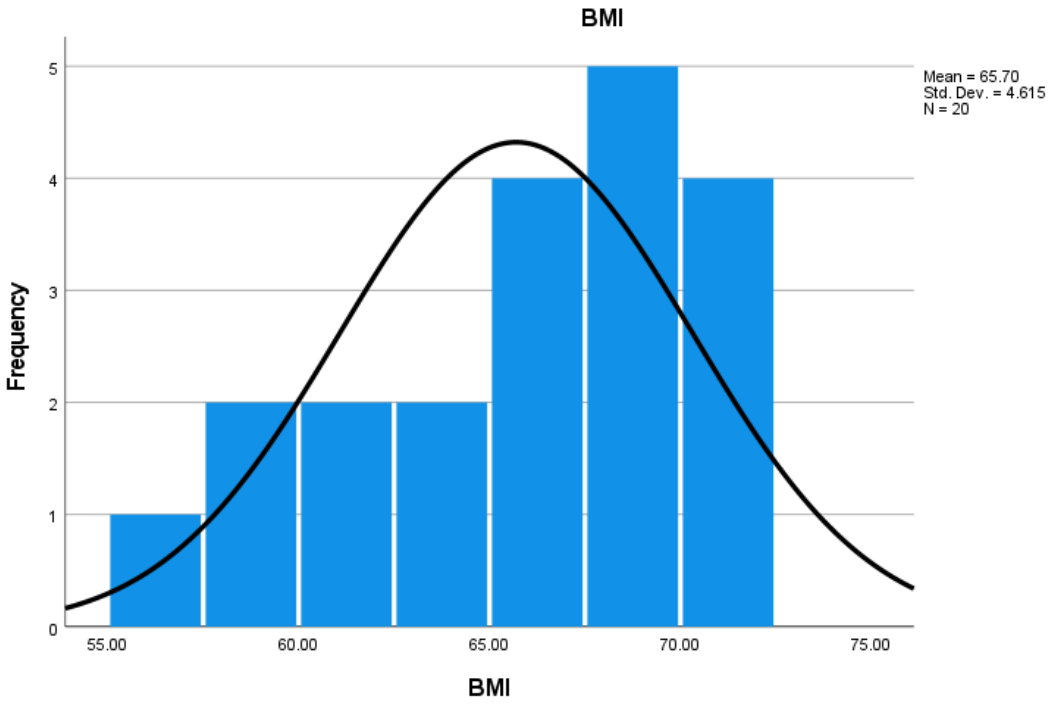

Figure: Illustrate histogram of BMI variable

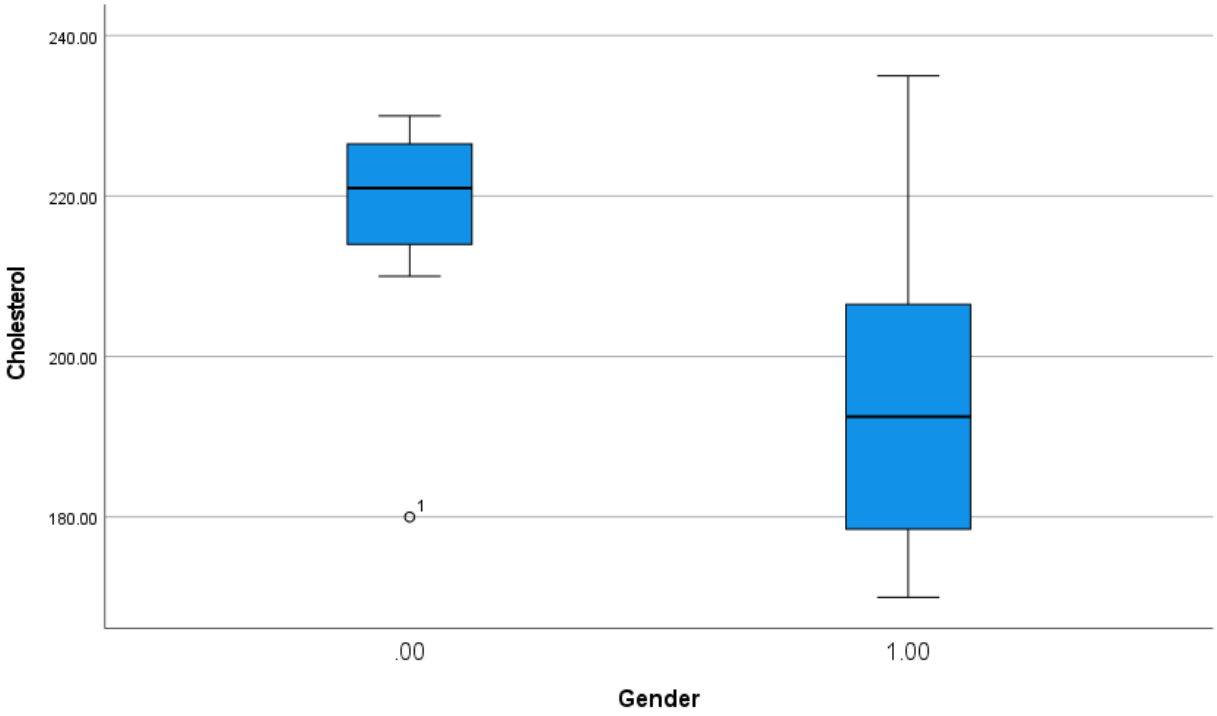

Figure: Illustrate box plots of gender vs cholesterol variable

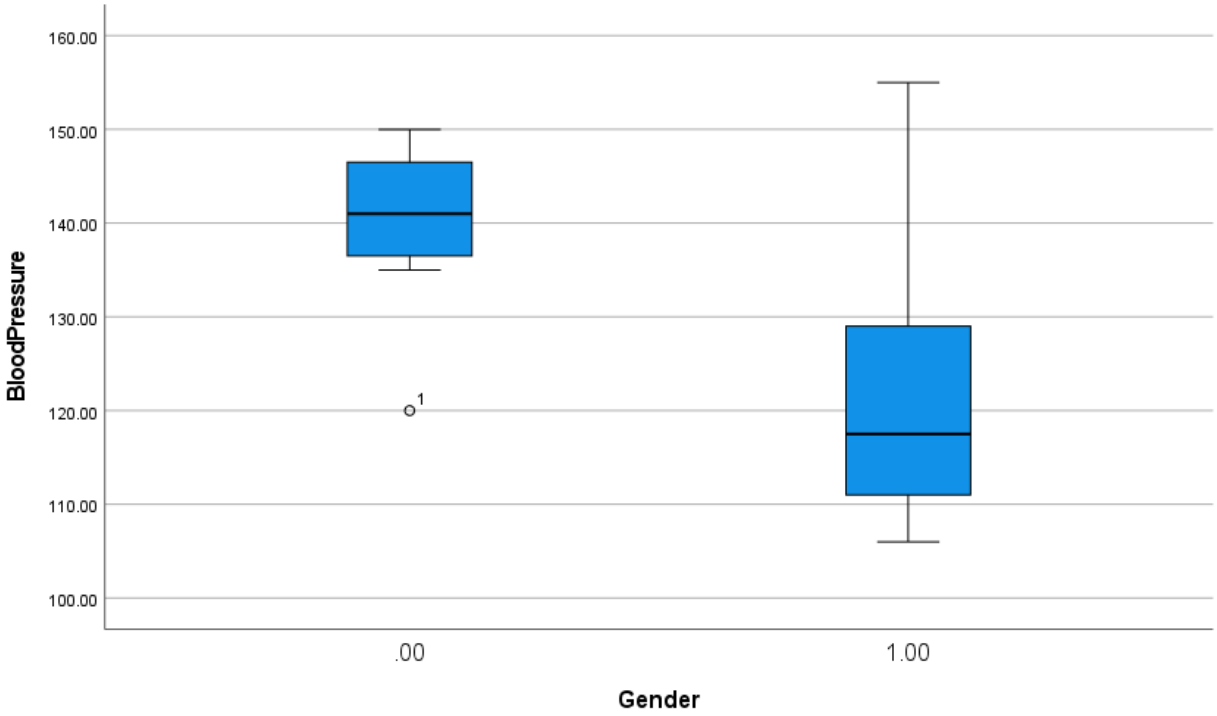

Figure: Illustrate box gender vs blood pressure

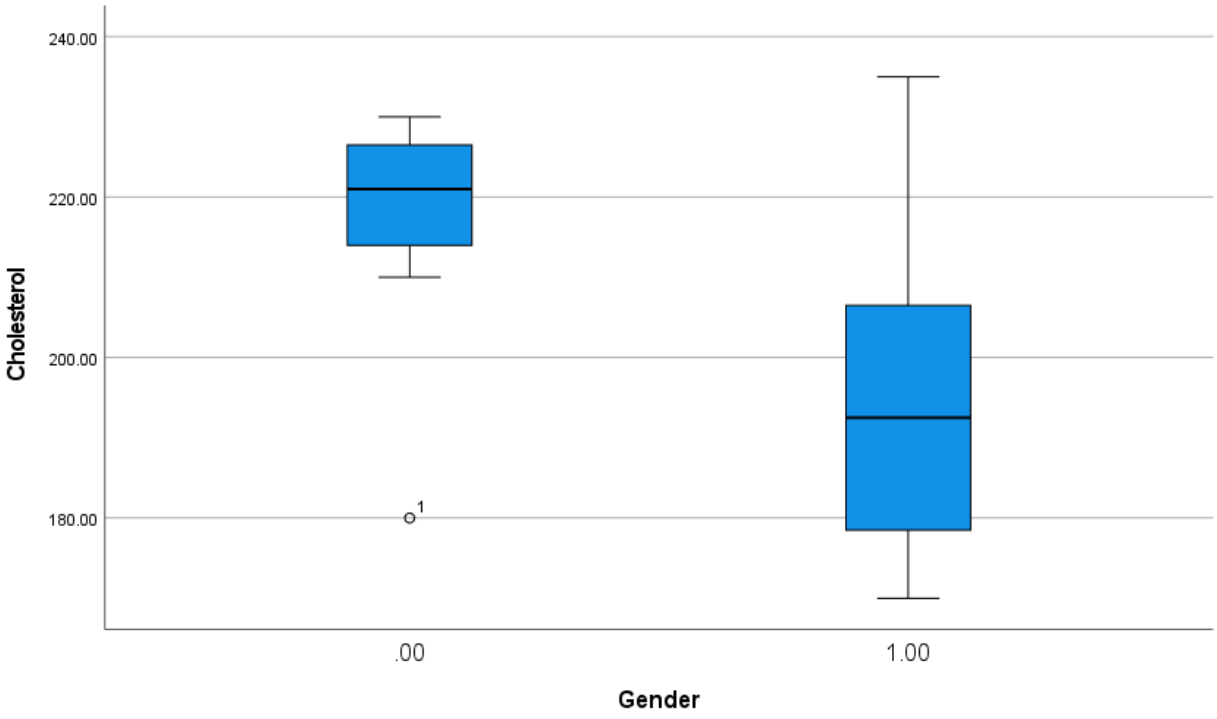

Figure: Illustrate box gender vs cholesterol

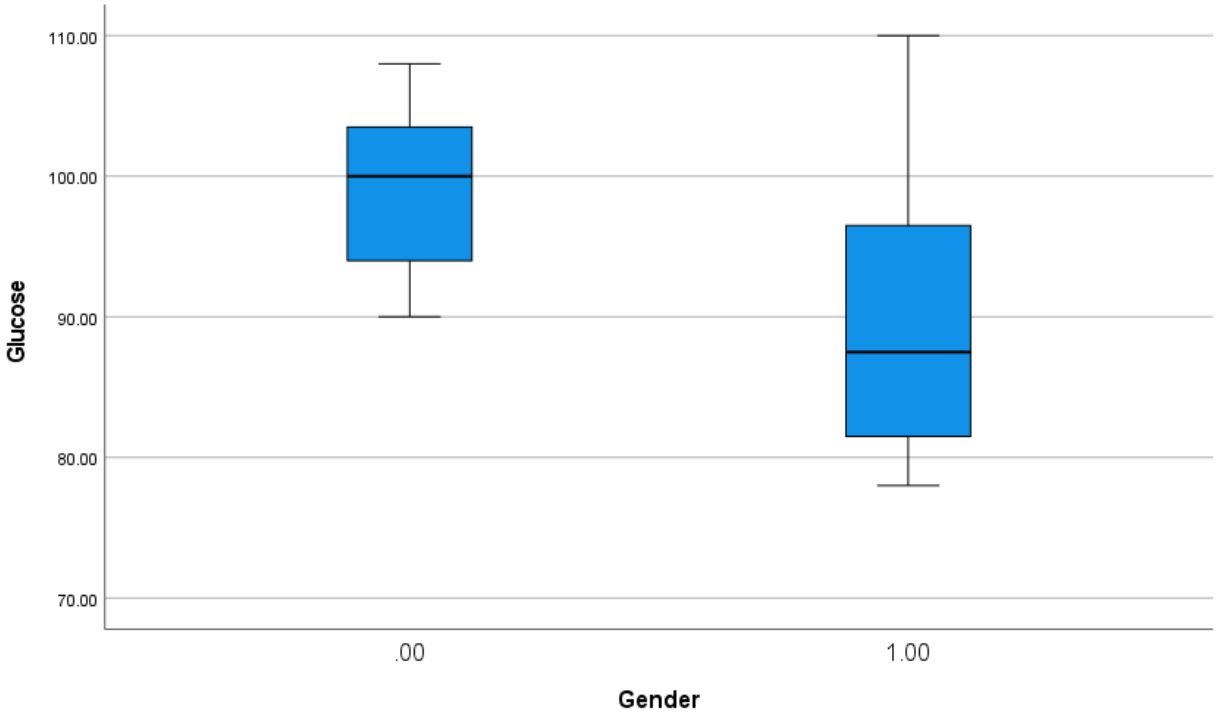

Figure: Illustrate box gender vs glucose

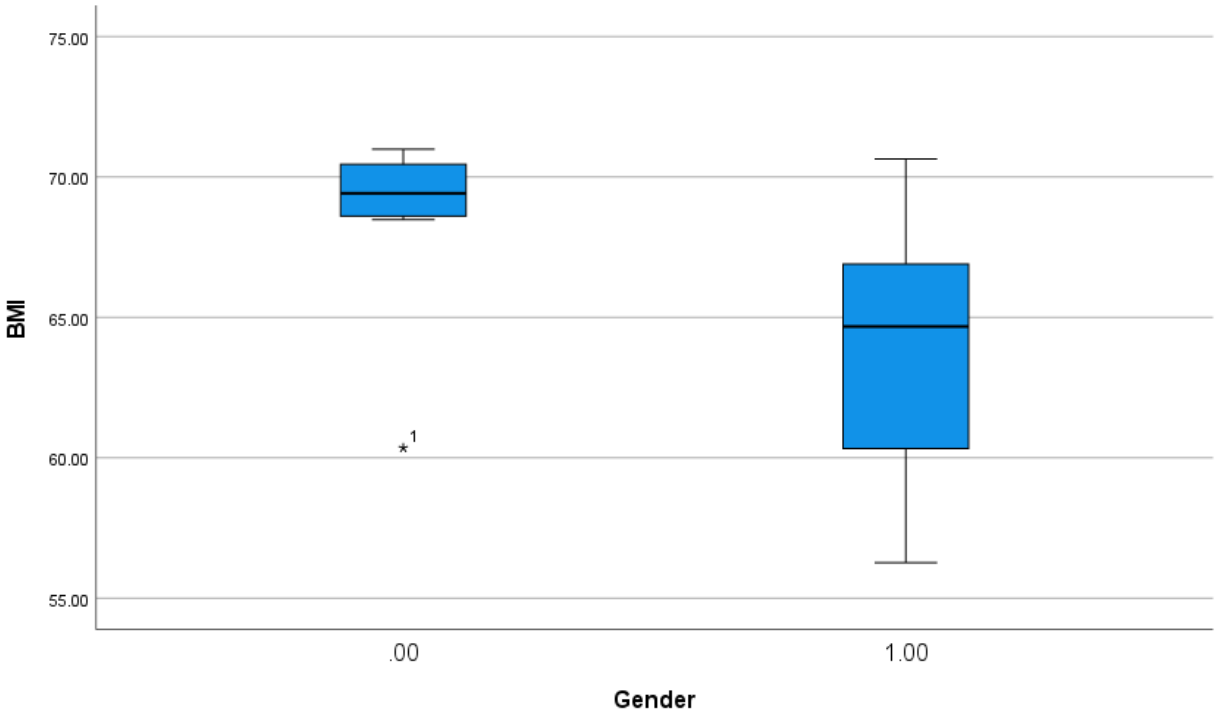

Figure: Illustrate box gender vs cholesterol

MEASURE OF CENTRAL TENDENCY:

Table 3: Illustrate the results of central tendency by means, S.D, min, and max values.

MEASURE OF CENTRAL TENDENCIES						
		Age	Blood Pressure	Cholesterol	Glucose	BMI
Diabetic	Mean ± Std	34.28±8.25	128.28±13.45	203.55±17.76	92.91±8	66.34±3.78
	Median	32.00	130.00	208.00	94.00	67.07
	Min ± Max	24±50	108±150	175±230	81±105	59.04±70.99
Non-Diabetic	Mean ± Std	34.12±11.13	129.12±17.88	202.56±25.27	94.12±11.65	64.92±5.62
	Median	30.00	125.00	205.00	96.00	66.41
	Min ± Max	22±52	106±155	170±235	78±110	56.27±70.7

The interpretation of the above test results is as follows:

Key Findings:

1. **Blood Pressure**:
   - Both females and males have p-values greater than 0.05 for both KS and SW tests.
   - **Conclusion**: Blood pressure is normally distributed for both genders.
2. **Cholesterol**:
   - Females have an SW p-value of 0.013 (< 0.05), indicating non-normality.
   - Males have p-values > 0.05 for both KS and SW tests.
   - **Conclusion**: Cholesterol is not normally distributed for females but is normally distributed for males.
3. **Glucose**:
   - Both genders have p-values > 0.05 for KS and SW tests.
   - **Conclusion**: Glucose is normally distributed for both genders.
4. **BMI**:
   - Females have p-values ≤ 0.05 for both KS and SW tests, indicating non-normality.
   - Males have p-values > 0.05 for both KS and SW tests.
   - **Conclusion**: BMI is not normally distributed for females but is normally distributed for males.
5. **Physical Activity**:
   - Both genders have SW p-values ≤ 0.05, indicating non-normality.
   - **Conclusion**: Physical activity is not normally distributed for either gender.

This indicates that parametric tests (e.g., t-tests or ANOVA) can be used for normally distributed variables.

GRAPHPAD:

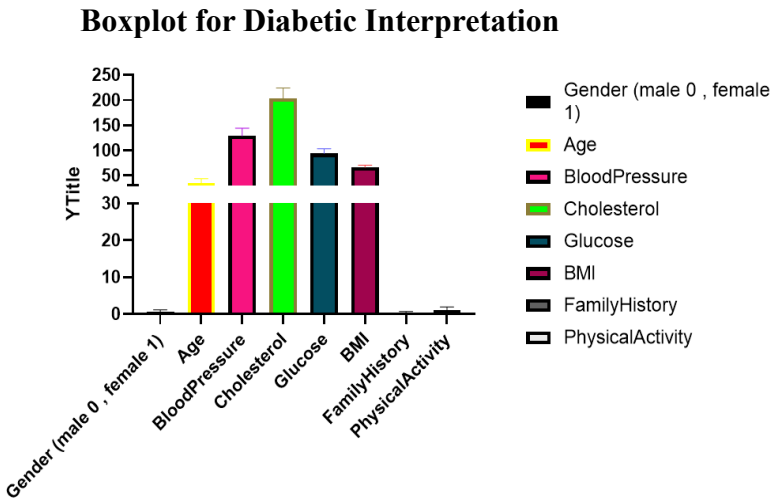

Figure: Illustrate two segmented box plots in graph pad

Correlation Analysis:

Table 4: Illustrate the correlation analysis, the association between variables.

Correlations									
	Gender	Age	Blood Pressure	Cholesterol	Glucose	BMI	Family History	Physical Activity	Outcome
Gender	1								
Age	-.574**	1							
Blood Pressure	-.613**	.966**	1						
Cholesterol	-.543*	.951**	.975**	1					
Glucose	-.498*	.906**	.955**	.948**	1				
BMI	-.516*	.900**	.923**	.969**	.902**	1			
Family History	-0.236	-0.101	0.061	0.003	0.059	0.074	1		
Physical Activity	0.000	-0.061	0.045	0.080	0.054	0.025	0.258	1	
Outcome	0.082	0.009	-0.028	0.024	-0.064	0.157	0.058	-.562**	1
**. Correlation is significant at the 0.01 level (2-tailed).									
*. Correlation is significant at the 0.05 level (2-tailed).									

Correlation Analysis from r value:

- **Strong Positive Relationships**: Age, Blood Pressure, Cholesterol, Glucose, and BMI are highly correlated with each other ($p < 0.01$).
- **Negative Relationships**: Gender negatively correlates with Age, Blood Pressure, Cholesterol, Glucose, and BMI ($p < 0.05$).
- **Physical Activity**: Negatively correlates with Outcome ($p < 0.01$).
- **Family History**: Does not show significant correlations with other variables.

Heat Map:

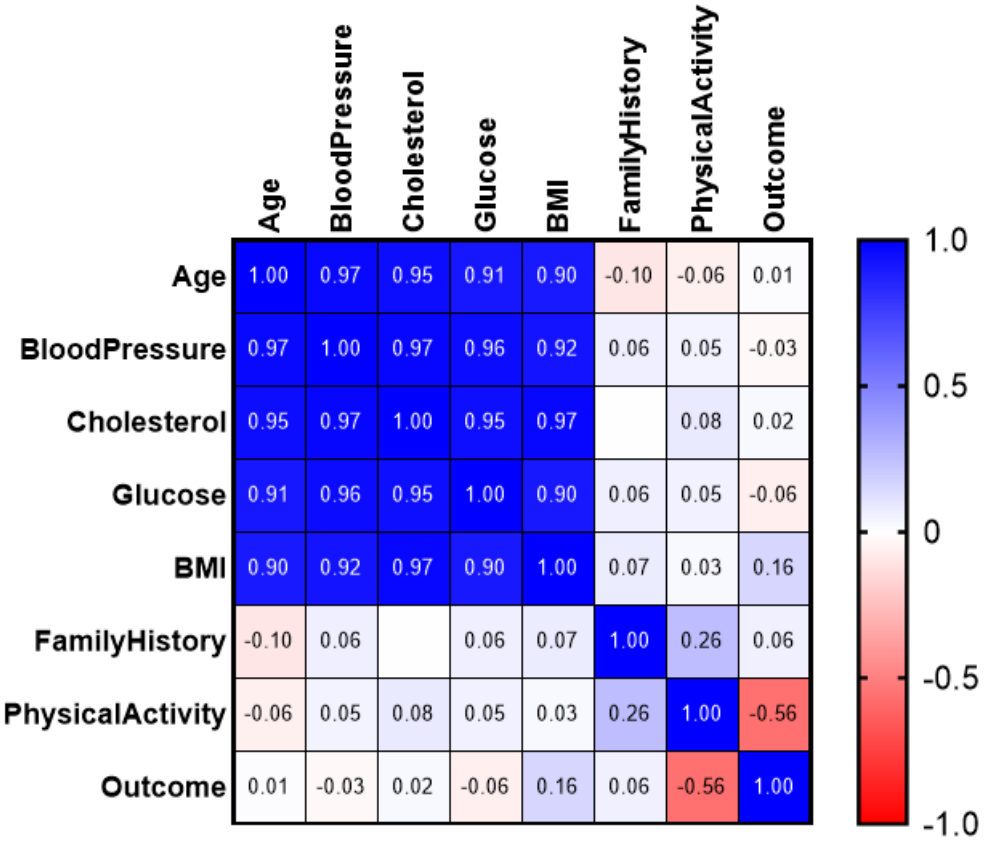

Figure: Illustrate the heat map with respective groups with correlation r

T test:

Independent test:

Table 5: Illustrate the results of independent sample t test

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means

		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Gender	Equal variances assumed	0.395	0.537	-0.349	18	0.731	-0.08081	0.23132	-0.56679	0.40518
	Equal variances not assumed			-0.348	16.895	0.732	-0.08081	0.23239	-0.57134	0.40972
Age	Equal variances assumed	1.842	0.192	-0.037	18	0.971	-0.16162	4.33070	-9.26009	8.93686
	Equal variances not assumed			-0.036	14.467	0.972	-0.16162	4.46607	-9.71144	9.38821
Blood Pressure	Equal variances assumed	2.076	0.167	0.120	18	0.906	0.83838	6.99934	-13.86667	15.54344
	Equal variances not assumed			0.116	14.615	0.909	0.83838	7.20766	-14.55970	16.23647
Cholesterol	Equal variances assumed	2.841	0.109	-0.103	18	0.919	-0.98990	9.62889	-21.21944	19.23964
	Equal variances not assumed			-0.099	13.948	0.922	-0.98990	9.98042	-22.40323	20.42343
Glucose	Equal variances assumed	2.133	0.161	0.273	18	0.788	1.20202	4.39850	-8.03888	10.44292
	Equal variances not assumed			0.263	13.725	0.796	1.20202	4.56904	-8.61607	11.02011
BMI	Equal variances assumed	2.819	0.110	-0.675	18	0.508	-1.42033	2.10453	-5.84179	3.00113
	Equal variances not assumed			-0.648	13.527	0.528	-1.42033	2.19038	-6.13370	3.29305
Family History	Equal variances assumed	0.249	0.624	-0.247	18	0.808	-0.05051	0.20481	-0.48079	0.37978
	Equal variances not assumed			-0.248	17.579	0.807	-0.05051	0.20357	-0.47892	0.37791

Physical Activity	Equal variances assumed	4.368	0.051	2.881	18	0.010	1.01010	0.35056	0.27361	1.74659
	Equal variances not assumed			3.043	16.225	0.008	1.01010	0.33197	0.30715	1.71305

ANOVA: MULTIPLE GROUP COMPSRIOSN TESTS

Dependent variables: age, blood pressure, glucose, BMI, cholesterol

Grouping variable: physical activity (3 groups 0,1,2)

Table 6: Illustrate the anova tests as physical activity contain 3 classes

ANOVA						
		Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**
Age	**Between Groups**	**45.450**	**2**	**22.725**	**.238**	**.791**
	Within Groups	**1625.750**	**17**	**95.632**		
	Total	**1671.200**	**19**			
Blood Pressure	**Between Groups**	**73.800**	**2**	**36.900**	**.146**	**.865**
	Within Groups	**4294.750**	**17**	**252.632**		
	Total	**4368.550**	**19**			
Glucose	**Between Groups**	**51.575**	**2**	**25.787**	**.261**	**.773**
	Within Groups	**1679.375**	**17**	**98.787**		
	Total	**1730.950**	**19**			

BMI	**Between Groups**	**3.926**	**2**	**1.963**	**.083**	**.920**
	Within Groups	**400.689**	**17**	**23.570**		
	Total	**404.615**	**19**			
Cholesterol	**Between Groups**	**54.675**	**2**	**27.338**	**.057**	**.945**
	Within Groups	**8211.125**	**17**	**483.007**		
	Total	**8265.800**	**19**			

Table 7:Illustrate the Multiple comparison group test as there are multiple groups in one variable

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) Physical Activity	(J) Physical Activity	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Age	**.00**	**1.00**	**-2.87500**	**5.98850**	**.882**	**-18.2376**	**12.4876**
		2.00	**1.25000**	**4.88959**	**.965**	**-11.2935**	**13.7935**
	1.00	**.00**	**2.87500**	**5.98850**	**.882**	**-12.4876**	**18.2376**
		2.00	**4.12500**	**5.98850**	**.773**	**-11.2376**	**19.4876**
	2.00	**.00**	**-1.25000**	**4.88959**	**.965**	**-13.7935**	**11.2935**
		1.00	**-4.12500**	**5.98850**	**.773**	**-19.4876**	**11.2376**

Blood Pressure	**.00**	**1.00**	**-5.25000**	**9.73330**	**.853**	**-30.2194**	**19.7194**
		2.00	**-1.50000**	**7.94721**	**.981**	**-21.8874**	**18.8874**
	1.00	**.00**	**5.25000**	**9.73330**	**.853**	**-19.7194**	**30.2194**
		2.00	**3.75000**	**9.73330**	**.922**	**-21.2194**	**28.7194**
	2.00	**.00**	**1.50000**	**7.94721**	**.981**	**-18.8874**	**21.8874**
		1.00	**-3.75000**	**9.73330**	**.922**	**-28.7194**	**21.2194**
Glucose	**.00**	**1.00**	**-4.37500**	**6.08646**	**.756**	**-19.9890**	**11.2390**
		2.00	**-1.12500**	**4.96958**	**.972**	**-13.8737**	**11.6237**
	1.00	**.00**	**4.37500**	**6.08646**	**.756**	**-11.2390**	**19.9890**
		2.00	**3.25000**	**6.08646**	**.856**	**-12.3640**	**18.8640**
	2.00	**.00**	**1.12500**	**4.96958**	**.972**	**-11.6237**	**13.8737**
		1.00	**-3.25000**	**6.08646**	**.856**	**-18.8640**	**12.3640**
BMI	**.00**	**1.00**	**.94321**	**2.97300**	**.946**	**-6.6836**	**8.5700**
		2.00	**-.25452**	**2.42744**	**.994**	**-6.4818**	**5.9727**
	1.00	**.00**	**-.94321**	**2.97300**	**.946**	**-8.5700**	**6.6836**
		2.00	**-1.19772**	**2.97300**	**.915**	**-8.8245**	**6.4291**
	2.00	**.00**	**.25452**	**2.42744**	**.994**	**-5.9727**	**6.4818**

		1.00	1.19772	2.97300	.915	-6.4291	8.8245
Cholesterol	.00	1.00	-2.62500	13.45837	.979	-37.1505	31.9005
		2.00	-3.62500	10.98871	.942	-31.8150	24.5650
	1.00	.00	2.62500	13.45837	.979	-31.9005	37.1505
		2.00	-1.00000	13.45837	.997	-35.5255	33.5255
	2.00	.00	3.62500	10.98871	.942	-24.5650	31.8150
		1.00	1.00000	13.45837	.997	-33.5255	35.5255