**MATH F432: APPLIED STATISTICAL METHODS**

**Project 3**

**Cirrhosis Analysis**

*By*

*Akhil Agnihotri (2016A4PS0353H)*

*Deeksha Kartik (2016A7PS0809H)*

*G. Jaya Aadityaa (2016A4PS0213H)*

*Skanda Vaidyanath (2016A7PS0236H)*

**Legend**

age --- Age in year (continuous)

gender --- Gender of patients (male or female) (String categorical)

bilirubin --- Serum Bilirubin (mg/dl) (continuous)

albumin --- Serum Albumin (g/dl) (continuous)

stage --- Histolic stage of disease (numeric categorical)

clotting\_time --- Standardised blood clotting time (continuous)

time\_rem --- Time to death/Liver Transplantation (continuous)

BV\_mal --- Blood vessel malformation in the skin (numeric

categorical)

ascites --- Presence of ascites (numeric categorical)

liver --- Enlarged liver/Presence of Hepatomegaly (numeric

categorical)

alkaline\_phosphate --- Alkaline Phosphate (continuous)

GOT --- Serum Glutamic Oxaloacetic Transaminase (continuous)

cholesterol --- Serum Cholesterol (continuous)

triglycerides --- Triglycerides (mg/dl) (continuous)

platelet --- Platelet count (continuous)

drug --- Drug used (String categorical)

status --- Presence or Absence of Cirrhosis (numeric categorical)

edema --- Presence or Absence of Edema (String categorical)

copper --- Copper content (continuous)

gender\_num --- Gender of patients (male or female) (numeric categorical)

drug\_num --- Drug used (numeric categorical)

edema\_num --- Presence or Absence of Edema (numeric categorical)

**Two way MANOVA**

GLM age bilirubin albumin clotting\_time time\_rem alkaline\_phosphate GOT cholesterol Triglycerides

platelet copper BY status gender

/METHOD=SSTYPE(3)

/INTERCEPT=INCLUDE

/PLOT=PROFILE(status\*gender) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO

/EMMEANS=TABLES(status)

/EMMEANS=TABLES(gender)

/EMMEANS=TABLES(status\*gender)

/PRINT=DESCRIPTIVE

/CRITERIA=ALPHA(.05)

/DESIGN= status gender status\*gender.

**General Linear Model**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 18-NOV-2018 14:17:05 |
| Comments | |  |
| Input | Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 272 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics are based on all cases with valid data for all variables in the model. |
| Syntax | | GLM age bilirubin albumin clotting\_time time\_rem alkaline\_phosphate GOT cholesterol Triglycerides  platelet copper BY status gender  /METHOD=SSTYPE(3)  /INTERCEPT=INCLUDE  /PLOT=PROFILE(status\*gender) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO  /EMMEANS=TABLES(status)  /EMMEANS=TABLES(gender)  /EMMEANS=TABLES(status\*gender)  /PRINT=DESCRIPTIVE  /CRITERIA=ALPHA(.05)  /DESIGN= status gender status\*gender. |
| Resources | Processor Time | 00:00:01.67 |
| Elapsed Time | 00:00:01.47 |

|  |  |  |
| --- | --- | --- |
| **Between-Subjects Factors** | | |
|  | | N |
| status | 0 | 163 |
| 1 | 109 |
| gender | female | 238 |
| male | 34 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | status | gender | Mean | Std. Deviation | N |
| age | 0 | female | 46.77 | 9.801 | 150 |
| male | 53.69 | 13.181 | 13 |
| Total | 47.33 | 10.235 | 163 |
| 1 | female | 52.11 | 9.679 | 88 |
| male | 58.29 | 10.673 | 21 |
| Total | 53.30 | 10.127 | 109 |
| Total | female | 48.75 | 10.072 | 238 |
| male | 56.53 | 11.720 | 34 |
| Total | 49.72 | 10.588 | 272 |
| bilirubin | 0 | female | 1.642666666666667 | 1.801727234928410 | 150 |
| male | 3.184615384615385 | 2.823073429747018 | 13 |
| Total | 1.765644171779141 | 1.936918144190939 | 163 |
| 1 | female | 6.277272727272726 | 6.644066184573631 | 88 |
| male | 2.880952380952381 | 1.855160113741950 | 21 |
| Total | 5.622935779816514 | 6.165084376994950 | 109 |
| Total | female | 3.356302521008404 | 4.824115557520808 | 238 |
| male | 2.997058823529412 | 2.237488485715103 | 34 |
| Total | 3.311397058823530 | 4.579971909353926 | 272 |
| albumin | 0 | female | 3.605733333333332 | .329912224785078 | 150 |
| male | 3.596153846153846 | .320508410226067 | 13 |
| Total | 3.604969325153375 | .328213221559297 | 163 |
| 1 | female | 3.338181818181819 | .455713525295098 | 88 |
| male | 3.607142857142857 | .445265570835461 | 21 |
| Total | 3.390000000000001 | .464074147970790 | 109 |
| Total | female | 3.506806722689075 | .401764951694783 | 238 |
| male | 3.602941176470588 | .396916367129832 | 34 |
| Total | 3.518823529411762 | .401699450395459 | 272 |
| clotting\_time | 0 | female | 10.369333333333326 | .852233281440900 | 150 |
| male | 10.984615384615383 | .745929122913334 | 13 |
| Total | 10.418404907975463 | .858597712883254 | 163 |
| 1 | female | 11.234090909090915 | 1.022492600500036 | 88 |
| male | 11.033333333333331 | 1.124425779379561 | 21 |
| Total | 11.195412844036696 | 1.040511387901171 | 109 |
| Total | female | 10.689075630252100 | 1.007674342609665 | 238 |
| male | 11.014705882352940 | .984464891702417 | 34 |
| Total | 10.729779411764710 | 1.008796708891223 | 272 |
| time\_rem | 0 | female | 2285.55 | 979.266 | 150 |
| male | 2364.31 | 1302.983 | 13 |
| Total | 2291.83 | 1004.105 | 163 |
| 1 | female | 1507.99 | 1098.283 | 88 |
| male | 1599.81 | 1197.863 | 21 |
| Total | 1525.68 | 1112.979 | 109 |
| Total | female | 1998.05 | 1089.573 | 238 |
| male | 1892.12 | 1276.401 | 34 |
| Total | 1984.81 | 1112.586 | 272 |
| alkaline\_phosphate | 0 | female | 1482.513333333333300 | 1164.308855543148900 | 150 |
| male | 1859.830769230769000 | 2644.882903326287800 | 13 |
| Total | 1512.606134969325200 | 1332.487461395770700 | 163 |
| 1 | female | 2873.552272727272000 | 2892.579042362820500 | 88 |
| male | 2258.590476190476700 | 2274.127103287053000 | 21 |
| Total | 2755.073394495413600 | 2785.170611392303000 | 109 |
| Total | female | 1996.847058823531000 | 2092.013710486796000 | 238 |
| male | 2106.123529411764000 | 2390.984960481301600 | 34 |
| Total | 2010.506617647058200 | 2127.179167624052300 | 272 |
| GOT | 0 | female | 112.573599999999970 | 53.892090556378600 | 150 |
| male | 116.191538461538470 | 45.274106073662360 | 13 |
| Total | 112.862147239263820 | 53.142179682529810 | 163 |
| 1 | female | 143.834886363636370 | 60.590085168091600 | 88 |
| male | 131.114761904761930 | 46.755024181262880 | 21 |
| Total | 141.384220183486200 | 58.202602148357240 | 109 |
| Total | female | 124.132394957983240 | 58.329106250757214 | 238 |
| male | 125.408823529411760 | 46.091395618804240 | 34 |
| Total | 124.291948529411800 | 56.870918847688486 | 272 |
| cholesterol | 0 | female | 335.86 | 193.536 | 150 |
| male | 388.54 | 227.047 | 13 |
| Total | 340.06 | 196.148 | 163 |
| 1 | female | 439.11 | 300.351 | 88 |
| male | 352.05 | 149.178 | 21 |
| Total | 422.34 | 279.250 | 109 |
| Total | female | 374.04 | 243.226 | 238 |
| male | 366.00 | 180.435 | 34 |
| Total | 373.03 | 236.026 | 272 |
| Triglycerides | 0 | female | 110.81 | 51.245 | 150 |
| male | 140.54 | 42.663 | 13 |
| Total | 113.18 | 51.140 | 163 |
| 1 | female | 145.20 | 83.894 | 88 |
| male | 129.76 | 58.955 | 21 |
| Total | 142.23 | 79.691 | 109 |
| Total | female | 123.53 | 67.167 | 238 |
| male | 133.88 | 52.883 | 34 |
| Total | 124.82 | 65.557 | 272 |
| platelet | 0 | female | 271.47 | 87.243 | 150 |
| male | 247.31 | 72.851 | 13 |
| Total | 269.55 | 86.236 | 163 |
| 1 | female | 254.93 | 104.295 | 88 |
| male | 232.57 | 96.221 | 21 |
| Total | 250.62 | 102.739 | 109 |
| Total | female | 265.36 | 94.033 | 238 |
| male | 238.21 | 87.143 | 34 |
| Total | 261.96 | 93.480 | 272 |
| copper | 0 | female | 66.95 | 53.053 | 150 |
| male | 153.08 | 134.777 | 13 |
| Total | 73.82 | 66.948 | 163 |
| 1 | female | 134.05 | 106.475 | 88 |
| male | 164.19 | 76.035 | 21 |
| Total | 139.85 | 101.715 | 109 |
| Total | female | 91.76 | 83.574 | 238 |
| male | 159.94 | 100.694 | 34 |
| Total | 100.28 | 88.619 | 272 |

The mean age of the participants in the sample is 50 years. A healthy human has a low level of Bilirubin and its usually slightly higher in men. However, we see that women with Cirrhosis have abnormally high levels of it, which is not so in the case of men. The Albumin levels are similar for diseased and healthy humans. Clotting time is slightly higher in the case of diseased humans, and this can be attributed to their lower Platelet count. Time remaining depends on the status and not on gender. There is a significant increase in levels of Alkaline Phosphate and Glutamic Oxaloacetic Transaminase for those afflicted by the disease due to the presence of a damaged liver. The copper and Triglycerides levels are significantly higher for women with the disease but the changes in the levels for men is much lower.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Multivariate Tests** | | | | | | |
| Effect | | Value | F | Hypothesis df | Error df | Sig. |
| Intercept | Pillai's Trace | .993 | 3484.950b | 11.000 | 258.000 | .000 |
| Wilks' Lambda | .007 | 3484.950b | 11.000 | 258.000 | .000 |
| Hotelling's Trace | 148.583 | 3484.950b | 11.000 | 258.000 | .000 |
| Roy's Largest Root | 148.583 | 3484.950b | 11.000 | 258.000 | .000 |
| status | Pillai's Trace | .143 | 3.913b | 11.000 | 258.000 | .000 |
| Wilks' Lambda | .857 | 3.913b | 11.000 | 258.000 | .000 |
| Hotelling's Trace | .167 | 3.913b | 11.000 | 258.000 | .000 |
| Roy's Largest Root | .167 | 3.913b | 11.000 | 258.000 | .000 |
| gender | Pillai's Trace | .167 | 4.693b | 11.000 | 258.000 | .000 |
| Wilks' Lambda | .833 | 4.693b | 11.000 | 258.000 | .000 |
| Hotelling's Trace | .200 | 4.693b | 11.000 | 258.000 | .000 |
| Roy's Largest Root | .200 | 4.693b | 11.000 | 258.000 | .000 |
| status \* gender | Pillai's Trace | .069 | 1.734b | 11.000 | 258.000 | .066 |
| Wilks' Lambda | .931 | 1.734b | 11.000 | 258.000 | .066 |
| Hotelling's Trace | .074 | 1.734b | 11.000 | 258.000 | .066 |
| Roy's Largest Root | .074 | 1.734b | 11.000 | 258.000 | .066 |

|  |
| --- |
| a. Design: Intercept + status + gender + status \* gender |
| b. Exact statistic |

The interaction effect determines whether the effect of gender is consistent across the different statuses (Presence or absence of Liver Cirrhosis disease). Alternatively, but equivalently, the interaction effect determines whether the presence/absence of the disease is similar for males and females.

In this case, p = **.066** (i.e., the Wilks' Lambda row highlighted in yellow), which means that there is no statistically significant interaction effect. This means that the effect of gender (male or female) on the dependent variables is the same for those who have Liver Cirrhosis and those who don’t.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Between-Subjects Effects** | | | | | | |
| Source | Dependent Variable | Type III Sum of Squares | df | Mean Square | F | Sig. |
| Corrected Model | age | 3552.553a | 3 | 1184.184 | 11.830 | .000 |
| bilirubin | 1195.884b | 3 | 398.628 | 23.801 | .000 |
| albumin | 4.246c | 3 | 1.415 | 9.607 | .000 |
| clotting\_time | 44.649d | 3 | 14.883 | 17.256 | .000 |
| time\_rem | 38558912.503e | 3 | 12852970.834 | 11.602 | .000 |
| alkaline\_phosphate | 108950742.249f | 3 | 36316914.083 | 8.711 | .000 |
| GOT | 56038.062g | 3 | 18679.354 | 6.102 | .000 |
| cholesterol | 603913.595h | 3 | 201304.532 | 3.722 | .012 |
| Triglycerides | 69719.398i | 3 | 23239.799 | 5.688 | .001 |
| platelet | 38850.736j | 3 | 12950.245 | 1.490 | .218 |
| copper | 388942.550k | 3 | 129647.517 | 19.977 | .000 |
| Intercept | age | 311869.872 | 1 | 311869.872 | 3115.651 | .000 |
| bilirubin | 1371.895 | 1 | 1371.895 | 81.911 | .000 |
| albumin | 1403.803 | 1 | 1403.803 | 9528.610 | .000 |
| clotting\_time | 13346.369 | 1 | 13346.369 | 15474.703 | .000 |
| time\_rem | 422109475.391 | 1 | 422109475.391 | 381.025 | .000 |
| alkaline\_phosphate | 503722414.065 | 1 | 503722414.065 | 120.825 | .000 |
| GOT | 1779646.953 | 1 | 1779646.953 | 581.316 | .000 |
| cholesterol | 16110557.243 | 1 | 16110557.243 | 297.911 | .000 |
| Triglycerides | 1942948.576 | 1 | 1942948.576 | 475.557 | .000 |
| platelet | 7102411.414 | 1 | 7102411.414 | 817.188 | .000 |
| copper | 1883953.618 | 1 | 1883953.618 | 290.289 | .000 |
| status | age | 692.130 | 1 | 692.130 | 6.915 | .009 |
| bilirubin | 131.562 | 1 | 131.562 | 7.855 | .005 |
| albumin | .462 | 1 | .462 | 3.134 | .078 |
| clotting\_time | 5.853 | 1 | 5.853 | 6.786 | .010 |
| time\_rem | 16678802.211 | 1 | 16678802.211 | 15.055 | .000 |
| alkaline\_phosphate | 22468434.108 | 1 | 22468434.108 | 5.389 | .021 |
| GOT | 14960.880 | 1 | 14960.880 | 4.887 | .028 |
| cholesterol | 31263.202 | 1 | 31263.202 | .578 | .448 |
| Triglycerides | 3911.359 | 1 | 3911.359 | .957 | .329 |
| platelet | 6861.772 | 1 | 6861.772 | .790 | .375 |
| copper | 42898.389 | 1 | 42898.389 | 6.610 | .011 |
| gender | age | 1202.025 | 1 | 1202.025 | 12.009 | .001 |
| bilirubin | 24.119 | 1 | 24.119 | 1.440 | .231 |
| albumin | .472 | 1 | .472 | 3.203 | .075 |
| clotting\_time | 1.205 | 1 | 1.205 | 1.397 | .238 |
| time\_rem | 204094.084 | 1 | 204094.084 | .184 | .668 |
| alkaline\_phosphate | 396113.352 | 1 | 396113.352 | .095 | .758 |
| GOT | 581.107 | 1 | 581.107 | .190 | .663 |
| cholesterol | 8294.058 | 1 | 8294.058 | .153 | .696 |
| Triglycerides | 1430.778 | 1 | 1430.778 | .350 | .555 |
| platelet | 15182.960 | 1 | 15182.960 | 1.747 | .187 |
| copper | 94817.702 | 1 | 94817.702 | 14.610 | .000 |
| status \* gender | age | 3.913 | 1 | 3.913 | .039 | .843 |
| bilirubin | 171.046 | 1 | 171.046 | 10.213 | .002 |
| albumin | .544 | 1 | .544 | 3.694 | .056 |
| clotting\_time | 4.671 | 1 | 4.671 | 5.416 | .021 |
| time\_rem | 1196.304 | 1 | 1196.304 | .001 | .974 |
| alkaline\_phosphate | 6906090.412 | 1 | 6906090.412 | 1.657 | .199 |
| GOT | 1872.258 | 1 | 1872.258 | .612 | .435 |
| cholesterol | 136972.623 | 1 | 136972.623 | 2.533 | .113 |
| Triglycerides | 14309.410 | 1 | 14309.410 | 3.502 | .062 |
| platelet | 22.858 | 1 | 22.858 | .003 | .959 |
| copper | 21979.005 | 1 | 21979.005 | 3.387 | .067 |
| Error | age | 26826.212 | 268 | 100.098 |  |  |
| bilirubin | 4488.651 | 268 | 16.749 |  |  |
| albumin | 39.483 | 268 | .147 |  |  |
| clotting\_time | 231.140 | 268 | .862 |  |  |
| time\_rem | 296897550.169 | 268 | 1107826.680 |  |  |
| alkaline\_phosphate | 1117294775.979 | 268 | 4169010.358 |  |  |
| GOT | 820457.621 | 268 | 3061.409 |  |  |
| cholesterol | 14493001.107 | 268 | 54078.362 |  |  |
| Triglycerides | 1094948.132 | 268 | 4085.627 |  |  |
| platelet | 2329264.896 | 268 | 8691.287 |  |  |
| copper | 1739300.653 | 268 | 6489.928 |  |  |
| Total | age | 702800.000 | 272 |  |  |  |
| bilirubin | 8667.110 | 272 |  |  |  |
| albumin | 3411.666 | 272 |  |  |  |
| clotting\_time | 31590.650 | 272 |  |  |  |
| time\_rem | 1406987263.000 | 272 |  |  |  |
| alkaline\_phosphate | 2325706744.040 | 272 |  |  |  |
| GOT | 5078484.546 | 272 |  |  |  |
| cholesterol | 52946717.000 | 272 |  |  |  |
| Triglycerides | 5402676.000 | 272 |  |  |  |
| platelet | 21034044.000 | 272 |  |  |  |
| copper | 4863665.000 | 272 |  |  |  |
| Corrected Total | age | 30378.765 | 271 |  |  |  |
| bilirubin | 5684.535 | 271 |  |  |  |
| albumin | 43.729 | 271 |  |  |  |
| clotting\_time | 275.789 | 271 |  |  |  |
| time\_rem | 335456462.673 | 271 |  |  |  |
| alkaline\_phosphate | 1226245518.228 | 271 |  |  |  |
| GOT | 876495.682 | 271 |  |  |  |
| cholesterol | 15096914.702 | 271 |  |  |  |
| Triglycerides | 1164667.529 | 271 |  |  |  |
| platelet | 2368115.632 | 271 |  |  |  |
| copper | 2128243.202 | 271 |  |  |  |

|  |
| --- |
| a. R Squared = .117 (Adjusted R Squared = .107) |
| b. R Squared = .210 (Adjusted R Squared = .202) |
| c. R Squared = .097 (Adjusted R Squared = .087) |
| d. R Squared = .162 (Adjusted R Squared = .153) |
| e. R Squared = .115 (Adjusted R Squared = .105) |
| f. R Squared = .089 (Adjusted R Squared = .079) |
| g. R Squared = .064 (Adjusted R Squared = .053) |
| h. R Squared = .040 (Adjusted R Squared = .029) |
| i. R Squared = .060 (Adjusted R Squared = .049) |
| j. R Squared = .016 (Adjusted R Squared = .005) |
| k. R Squared = .183 (Adjusted R Squared = .174) |

Note: We make an alpha correction to account for multiple ANOVAs being run, such as a Bonferroni correction. As such, in this case, we accept statistical significance at *p* < .025.

We can see from this table that Status (presence of disease) has a statistically significant effect on Serum Bilirubin (p= 0.005<0.025). Similarly, we can infer that Status has a statistically significant effect on Age. Clotting time, Time remaining, Alkaline Phosphate, Copper levels.

Equivalently, Gender has a Statistical significance on Age and Copper levels. Clearly, Gender has a significant effect on fewer variables.

The combined effect of Status and Gender, have a significant effect on Age and Time remaining. The combined effect has a statistically low significance on the variables due to low interaction effect.

Certain variables such as Albumin, Alkaline Phosphate, Glutamic oxaloacetic transaminase, Cholesterol and Triglycerides are not significantly varied by the variation in factors.

.

**Estimated Marginal Means**

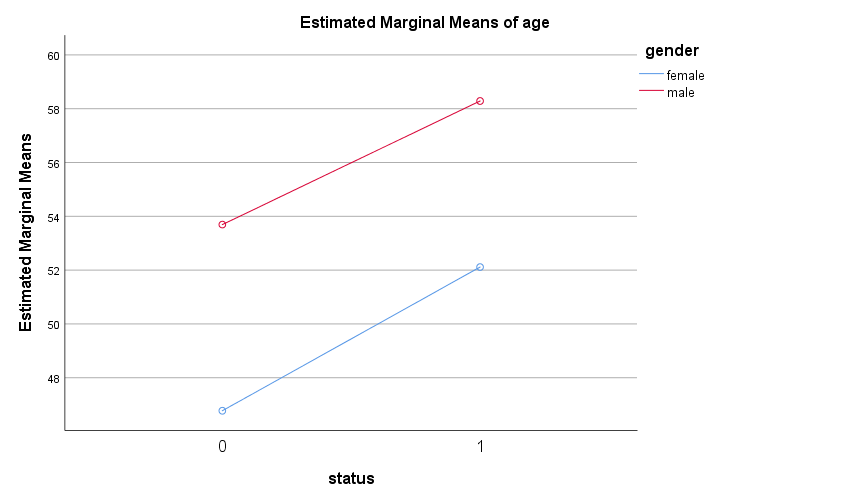
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1. Status** | | | | | |
| Dependent Variable | status | Mean | Std. Error | 95% Confidence Interval | |
| Lower Bound | Upper Bound |
| age | 0 | 50.233 | 1.446 | 47.385 | 53.080 |
| 1 | 55.200 | 1.215 | 52.808 | 57.592 |
| bilirubin | 0 | 2.414 | .592 | 1.249 | 3.578 |
| 1 | 4.579 | .497 | 3.601 | 5.558 |
| albumin | 0 | 3.601 | .055 | 3.492 | 3.710 |
| 1 | 3.473 | .047 | 3.381 | 3.564 |
| clotting\_time | 0 | 10.677 | .134 | 10.413 | 10.941 |
| 1 | 11.134 | .113 | 10.912 | 11.356 |
| time\_rem | 0 | 2324.927 | 152.154 | 2025.359 | 2624.496 |
| 1 | 1553.899 | 127.811 | 1302.258 | 1805.540 |
| alkaline\_phosphate | 0 | 1671.172 | 295.164 | 1090.037 | 2252.307 |
| 1 | 2566.071 | 247.941 | 2077.911 | 3054.232 |
| GOT | 0 | 114.383 | 7.998 | 98.635 | 130.130 |
| 1 | 137.475 | 6.719 | 124.246 | 150.703 |
| cholesterol | 0 | 362.199 | 33.617 | 296.012 | 428.386 |
| 1 | 395.581 | 28.239 | 339.983 | 451.178 |
| Triglycerides | 0 | 125.676 | 9.240 | 107.484 | 143.868 |
| 1 | 137.483 | 7.762 | 122.201 | 152.765 |
| platelet | 0 | 259.391 | 13.477 | 232.857 | 285.924 |
| 1 | 243.752 | 11.321 | 221.463 | 266.040 |
| copper | 0 | 110.015 | 11.646 | 87.086 | 132.944 |
| 1 | 149.118 | 9.783 | 129.858 | 168.378 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2. Gender** | | | | | |
| Dependent Variable | gender | Mean | Std. Error | 95% Confidence Interval | |
| Lower Bound | Upper Bound |
| age | female | 49.443 | .672 | 48.121 | 50.766 |
| male | 55.989 | 1.765 | 52.513 | 59.465 |
| bilirubin | female | 3.960 | .275 | 3.419 | 4.501 |
| male | 3.033 | .722 | 1.611 | 4.455 |
| albumin | female | 3.472 | .026 | 3.421 | 3.523 |
| male | 3.602 | .068 | 3.468 | 3.735 |
| clotting\_time | female | 10.802 | .062 | 10.679 | 10.924 |
| male | 11.009 | .164 | 10.686 | 11.332 |
| time\_rem | female | 1896.768 | 70.665 | 1757.638 | 2035.898 |
| male | 1982.059 | 185.722 | 1616.398 | 2347.719 |
| alkaline\_phosphate | female | 2178.033 | 137.084 | 1908.133 | 2447.932 |
| male | 2059.211 | 360.284 | 1349.864 | 2768.557 |
| GOT | female | 128.204 | 3.715 | 120.890 | 135.518 |
| male | 123.653 | 9.763 | 104.431 | 142.875 |
| cholesterol | female | 387.487 | 15.613 | 356.747 | 418.226 |
| male | 370.293 | 41.034 | 289.504 | 451.082 |
| Triglycerides | female | 128.009 | 4.291 | 119.560 | 136.458 |
| male | 135.150 | 11.279 | 112.944 | 157.356 |
| platelet | female | 263.203 | 6.259 | 250.879 | 275.526 |
| male | 239.940 | 16.450 | 207.552 | 272.328 |
| copper | female | 100.499 | 5.409 | 89.850 | 111.148 |
| male | 158.634 | 14.215 | 130.646 | 186.621 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **3. Status \* Gender** | | | | | | |
| Dependent Variable | status | gender | Mean | Std. Error | 95% Confidence Interval | |
| Lower Bound | Upper Bound |
| age | 0 | female | 46.773 | .817 | 45.165 | 48.382 |
| male | 53.692 | 2.775 | 48.229 | 59.156 |
| 1 | female | 52.114 | 1.067 | 50.014 | 54.213 |
| male | 58.286 | 2.183 | 53.987 | 62.584 |
| bilirubin | 0 | female | 1.643 | .334 | .985 | 2.301 |
| male | 3.185 | 1.135 | .950 | 5.419 |
| 1 | female | 6.277 | .436 | 5.418 | 7.136 |
| male | 2.881 | .893 | 1.123 | 4.639 |
| albumin | 0 | female | 3.606 | .031 | 3.544 | 3.667 |
| male | 3.596 | .106 | 3.387 | 3.806 |
| 1 | female | 3.338 | .041 | 3.258 | 3.419 |
| male | 3.607 | .084 | 3.442 | 3.772 |
| clotting\_time | 0 | female | 10.369 | .076 | 10.220 | 10.519 |
| male | 10.985 | .258 | 10.477 | 11.492 |
| 1 | female | 11.234 | .099 | 11.039 | 11.429 |
| male | 11.033 | .203 | 10.634 | 11.432 |
| time\_rem | 0 | female | 2285.547 | 85.939 | 2116.345 | 2454.748 |
| male | 2364.308 | 291.920 | 1789.559 | 2939.056 |
| 1 | female | 1507.989 | 112.200 | 1287.082 | 1728.895 |
| male | 1599.810 | 229.682 | 1147.600 | 2052.019 |
| alkaline\_phosphate | 0 | female | 1482.513 | 166.714 | 1154.279 | 1810.748 |
| male | 1859.831 | 566.298 | 744.873 | 2974.789 |
| 1 | female | 2873.552 | 217.658 | 2445.015 | 3302.090 |
| male | 2258.590 | 445.561 | 1381.346 | 3135.835 |
| GOT | 0 | female | 112.574 | 4.518 | 103.679 | 121.468 |
| male | 116.192 | 15.346 | 85.978 | 146.405 |
| 1 | female | 143.835 | 5.898 | 132.222 | 155.448 |
| male | 131.115 | 12.074 | 107.343 | 154.887 |
| cholesterol | 0 | female | 335.860 | 18.987 | 298.477 | 373.243 |
| male | 388.538 | 64.497 | 261.553 | 515.524 |
| 1 | female | 439.114 | 24.790 | 390.306 | 487.921 |
| male | 352.048 | 50.746 | 252.136 | 451.959 |
| Triglycerides | 0 | female | 110.813 | 5.219 | 100.538 | 121.089 |
| male | 140.538 | 17.728 | 105.635 | 175.442 |
| 1 | female | 145.205 | 6.814 | 131.789 | 158.620 |
| male | 129.762 | 13.948 | 102.300 | 157.224 |
| platelet | 0 | female | 271.473 | 7.612 | 256.486 | 286.460 |
| male | 247.308 | 25.857 | 196.400 | 298.215 |
| 1 | female | 254.932 | 9.938 | 235.365 | 274.498 |
| male | 232.571 | 20.344 | 192.517 | 272.625 |
| copper | 0 | female | 66.953 | 6.578 | 54.003 | 79.904 |
| male | 153.077 | 22.343 | 109.086 | 197.068 |
| 1 | female | 134.045 | 8.588 | 117.137 | 150.953 |
| male | 164.190 | 17.580 | 129.579 | 198.802 |

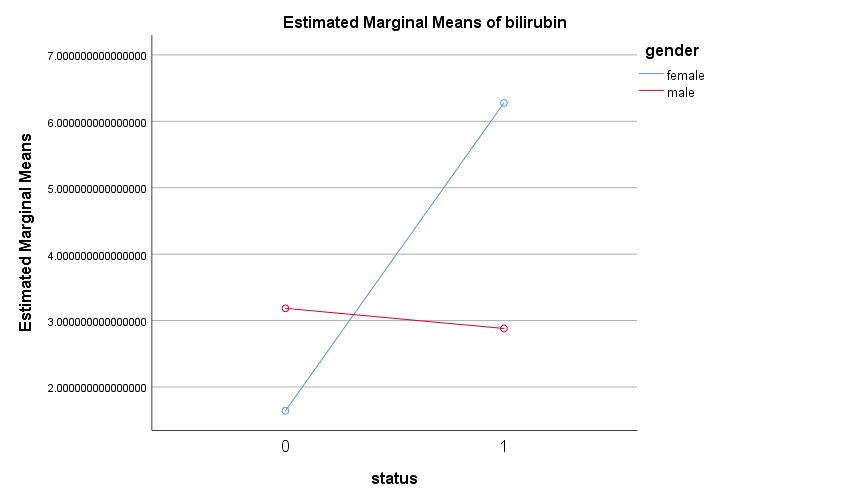
**Profile Plots**

**Age**



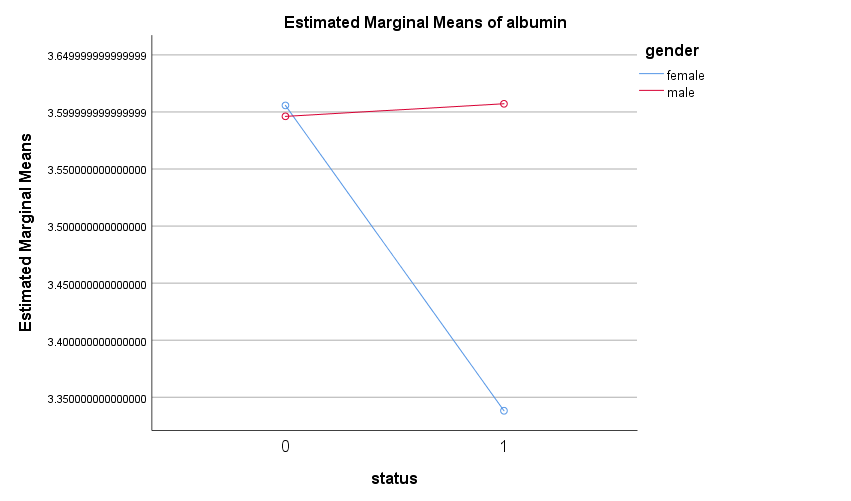
No statistically significant interaction as lines are parallel.

**Bilirubin (mg/dl )**

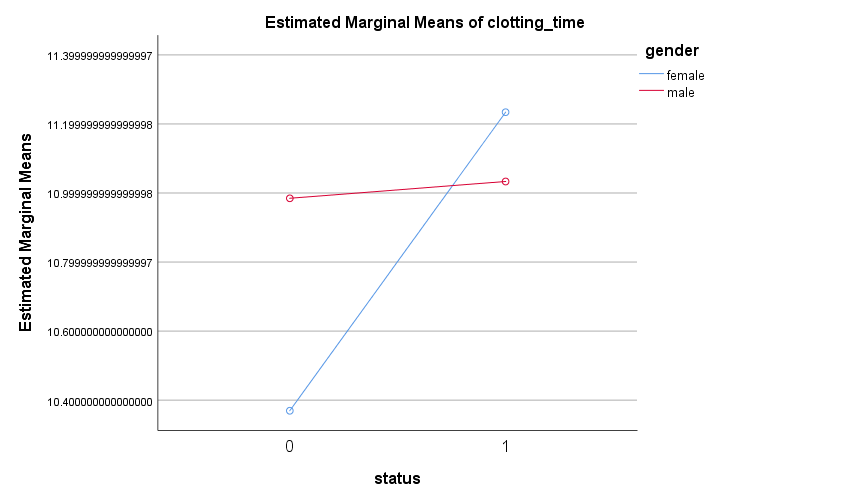


Statistically significant interaction as lines are crossing.

**Albumin**

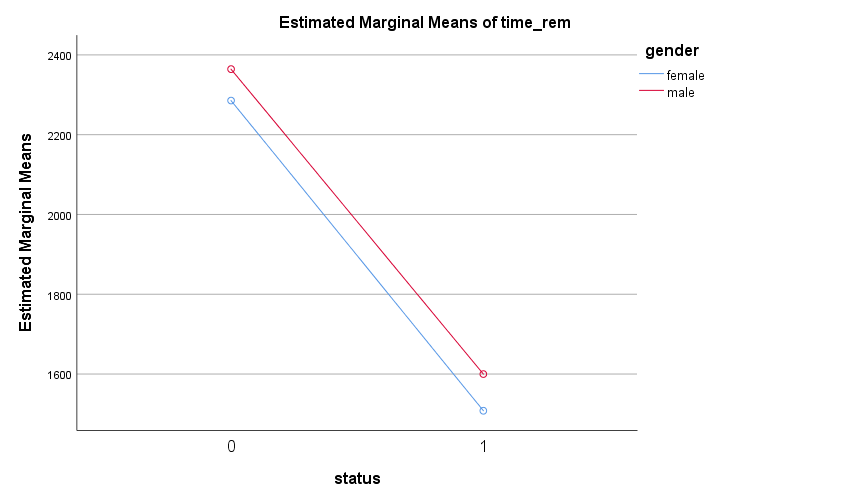
Statistically significant interaction as lines are crossing

**clotting\_time**



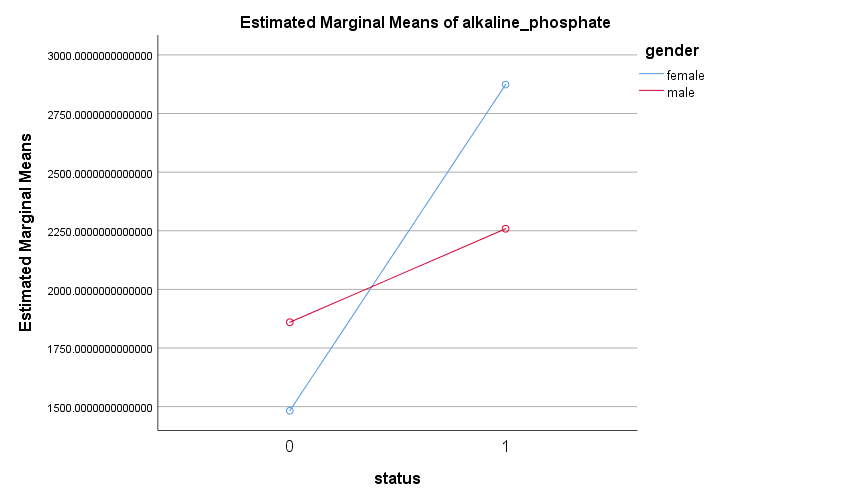
Statistically significant interaction as lines are crossing.

**time\_rem**



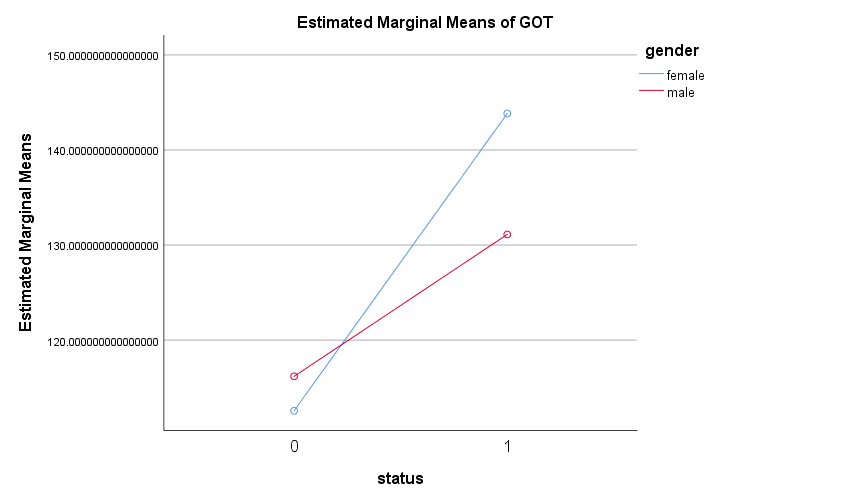
No statistically significant interaction as lines are parallel

**alkaline\_phosphate**



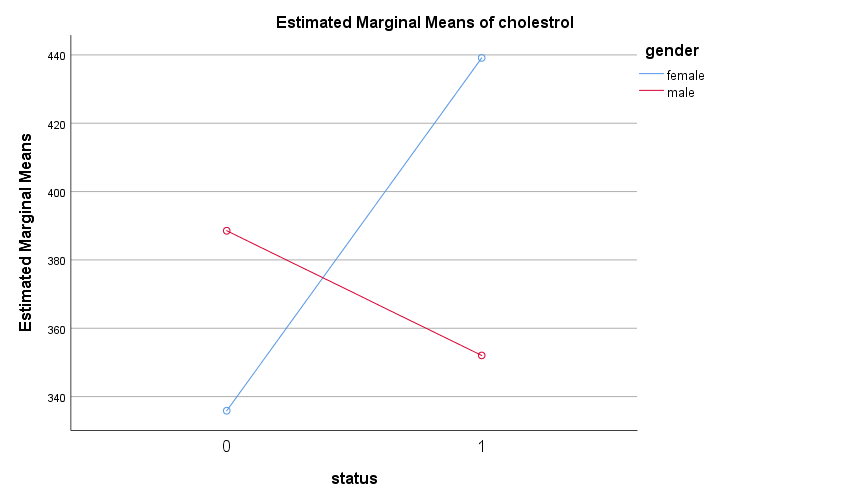
Statistically significant interaction as lines are crossing.

**GOT**



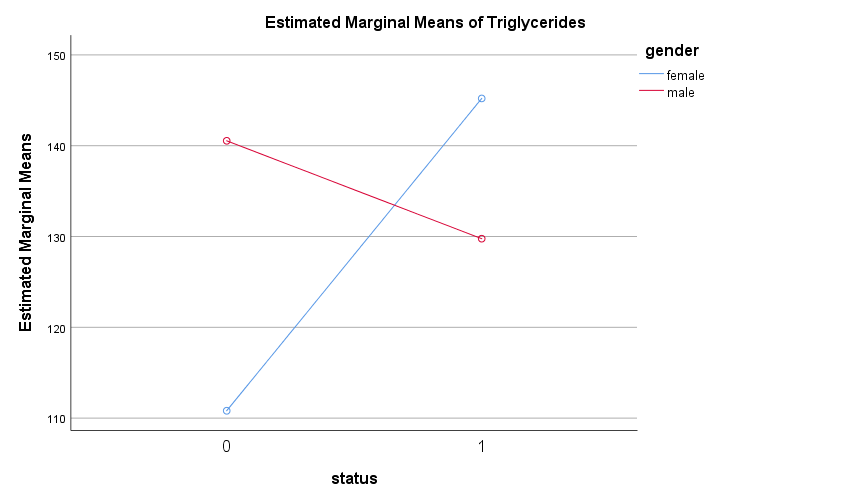
Statistically significant interaction as lines are crossing.

**cholesterol**



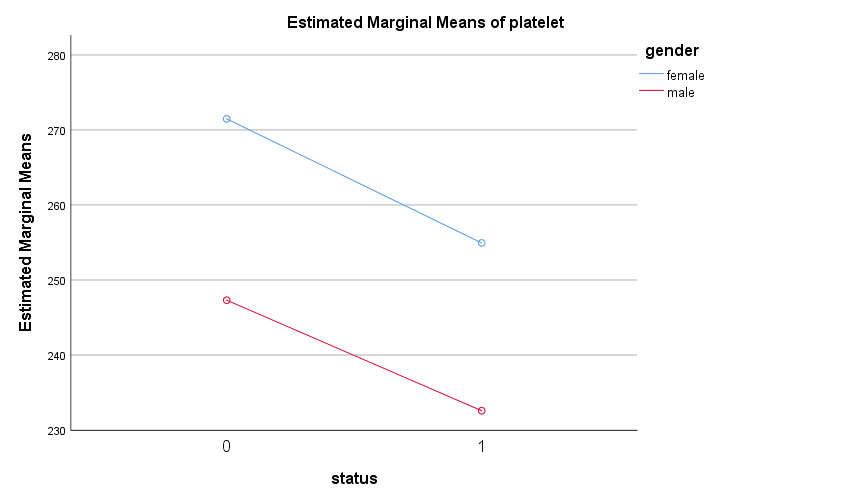
Statistically significant interaction as lines are crossing.

**Triglycerides**



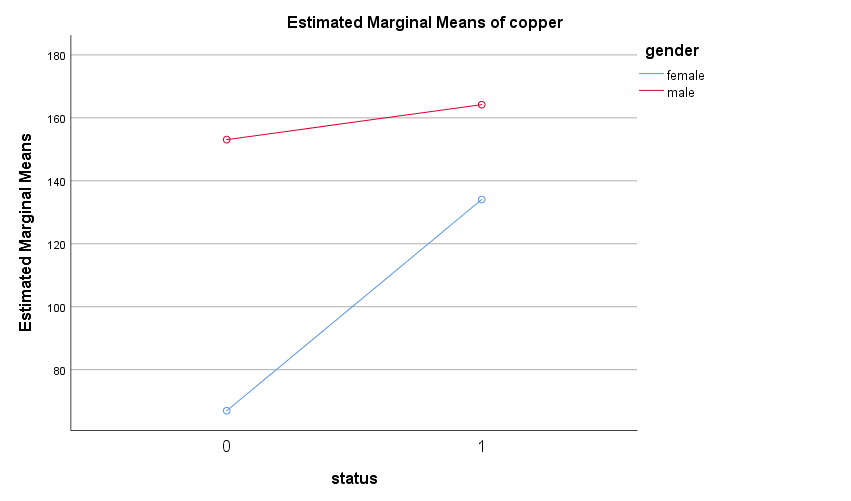
Statistically significant interaction as lines are crossing.

**platelet**



No statistically significant interaction as lines are parallel.

**copper**

Statistically significant interaction as lines are non-parallel.

AUTORECODE VARIABLES=gender

/INTO gender\_num

/PRINT.

gender into gender\_num

Old Value New Value Value Label

female 1 female

male 2 male

AUTORECODE VARIABLES=drug edema

/INTO drug\_num edema\_num

/PRINT.

drug into drug\_num

Old Value New Value Value Label

D-penicillamine 1 D-penicillamine

placebo 2 placebo

edema into edema\_num

Old Value New Value Value Label

edema despite diuretic therapy 1 edema despite diuretic therapy

edema, no diuretic therapy 2 edema, no diuretic therapy

no edema 3 no edema

NPAR TESTS

/K-W=stage BV\_mal liver ascites gender\_num drug\_num edema\_num BY status(0 1)

/STATISTICS DESCRIPTIVES

/MISSING ANALYSIS.

**Non-Parametric Tests**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 18-NOV-2018 20:28:15 |
| Comments | |  |
| Input | Active Dataset | DataSet1 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 272 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing. |
| Cases Used | Statistics for each test are based on all cases with valid data for the variable(s) used in that test. |
| Syntax | | NPAR TESTS  /K-W=stage BV\_mal liver ascites gender\_num drug\_num edema\_num BY status(0 1)  /STATISTICS DESCRIPTIVES  /MISSING ANALYSIS. |
| Resources | Processor Time | 00:00:00.03 |
| Elapsed Time | 00:00:00.04 |
| Number of Cases Alloweda | 241979 |
| a. Based on availability of workspace memory. | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | |
|  | N | Mean | Std. Deviation | Minimum | Maximum |
| stage | 272 | 3.04 | .854 | 1 | 4 |
| BV\_mal | 272 | .29 | .453 | 0 | 1 |
| liver | 272 | .53 | .500 | 0 | 1 |
| ascites | 272 | .08 | .267 | 0 | 1 |
| gender\_num | 272 | 1.13 | .331 | 1 | 2 |
| drug\_num | 272 | 1.49 | .501 | 1 | 2 |
| edema\_num | 272 | 2.79 | .533 | 1 | 3 |
| status | 272 | .40 | .491 | 0 | 1 |

**Kruskal-Wallis Test**

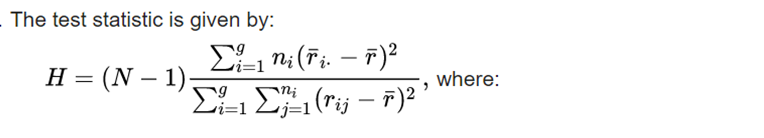
Grouping by Edema, grouping by gender, grouping by drug.

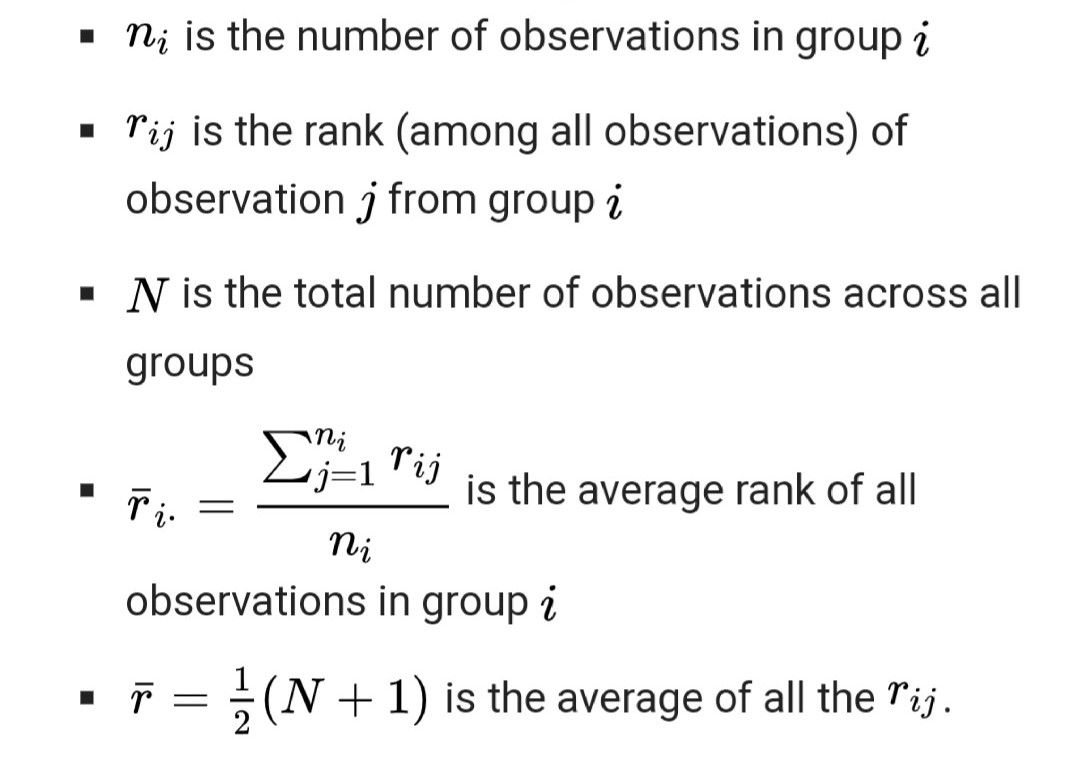
Three separate Kruskal Wallis Tests

|  |  |  |  |
| --- | --- | --- | --- |
| **Ranks** | | | |
|  | ed\_num | N | Mean Rank |
| @5 | edema despite diuretic therapy | 16 | 207.56 |
| edema, no diuretic therapy | 25 | 163.90 |
| no edema | 231 | 128.61 |
| Total | 272 |  |

|  |  |
| --- | --- |
| **Test Statisticsa,b** | |
|  | @5 |
| Kruskal-Wallis H | 20.796 |
| df | 2 |
| Asymp. Sig. | .000 |
| a. Kruskal Wallis Test | |
| b. Grouping Variable: ed\_num | |
| |  |  |  |  | | --- | --- | --- | --- | | **Ranks** | | | | |  | gender | N | Mean Rank | | @5 | female | 238 | 135.25 | | male | 34 | 145.26 | | Total | 272 |  |  |  |  | | --- | --- | | **Test Statisticsa,b** | | |  | @5 | | Kruskal-Wallis H | .545 | | df | 1 | | Asymp. Sig. | .460 | | a. Kruskal Wallis Test | | | b. Grouping Variable: gender | |  |  |  |  |  | | --- | --- | --- | --- | | **Ranks** | | | | |  | drug | N | Mean Rank | | @5 | D-penicillamine | 138 | 142.34 | | placebo | 134 | 130.49 | | Total | 272 |  |  |  |  | | --- | --- | | **Test Statisticsa,b** | | |  | @5 | | Kruskal-Wallis H | 1.744 | | df | 1 | | Asymp. Sig. | .187 | | a. Kruskal Wallis Test | | | b. Grouping Variable: drug | | | |

The Kruskal-Wallis H test is used to check if the factors have a statistically significant effect on the stage variable (test variable). The results from the three tests show that the test is significant only for grouping by edema but not for gender and drug. This indicates that the histolic stages are more or less same for both male and female and for the actual drug and penicillin. But in the case of edema, we find that the histolic stage is significantly higher for edema despite diuretic therapy, when compared to edema, no diuretic therapy and no edema. This is a very meaningful result.





**Chi-Square Goodness of Fit Test**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **gender \* status Crosstabulation** | | | | |
| Count | | | | |
|  | | status | | Total |
| 0 | 1 |
| gender | female | 150 | 88 | 238 |
| male | 13 | 21 | 34 |
| Total | | 163 | 109 | 272 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Chi-Square Tests** | | | | | |
|  | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | 7.613a | 1 | .006 |  |  |
| Continuity Correctionb | 6.616 | 1 | .010 |  |  |
| Likelihood Ratio | 7.447 | 1 | .006 |  |  |
| Fisher's Exact Test |  |  |  | .008 | .005 |
| Linear-by-Linear Association | 7.585 | 1 | .006 |  |  |
| N of Valid Cases | 272 |  |  |  |  |
| a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.63. | | | | | |
| b. Computed only for a 2x2 table | | | | | |

This test shows that (considering only Pearson Chi-Square), the test is statistically significant at 5% los. The 2-sided p-value is 0.006 as given in the table and this indicates that there is a difference in the number of liver cirrhosis affected males and females. The number for female dominates males.

**Discriminant Analysis**

DISCRIMINANT

/GROUPS=status(0 1)

/VARIABLES=@1 @3 @4 @5 @6 @7 @8 @9 @10 @11 @12 @13 @14 @15 drug edema\_num copper gender

/ANALYSIS ALL

/PRIORS EQUAL

/HISTORY

/STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR TABLE

/PLOT=COMBINED SEPARATE MAP

/CLASSIFY=NONMISSING POOLED.

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 11-NOV-2018 21:53:39 |
| Comments | |  |
| Input | Data | C:\Deeksha\3-1\Applied Statistical Methods\Project\_Dataset.sav |
| Active Dataset | DataSet4 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 272 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing in the analysis phase. |
| Cases Used | In the analysis phase, cases with no user- or system-missing values for any predictor variable are used. Cases with user-, system-missing, or out-of-range values for the grouping variable are always excluded. |
| Syntax | | DISCRIMINANT  /GROUPS=status(0 1)  /VARIABLES=@1 @3 @4 @5 @6 @7 @8 @9 @10 @11 @12 @13 @14 @15 drug edema\_num copper gender  /ANALYSIS ALL  /PRIORS EQUAL  /HISTORY  /STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR TABLE  /PLOT=COMBINED SEPARATE MAP  /CLASSIFY=NONMISSING POOLED. |
| Resources | Processor Time | 00:00:00.44 |
| Elapsed Time | 00:00:00.32 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Analysis Case Processing Summary** | | | |
| Unweighted Cases | | N | Percent |
| Valid | | 272 | 100.0 |
| Excluded | Missing or out-of-range group codes | 0 | .0 |
| At least one missing discriminating variable | 0 | .0 |
| Both missing or out-of-range group codes and at least one missing discriminating variable | 0 | .0 |
| Total | 0 | .0 |
| Total | | 272 | 100.0 |

The above table tells us if there’s any missing data. As we can see, no data is missing.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group Statistics** | | | | | |
| status | | Mean | Std. Deviation | Valid N (listwise) | |
| Unweighted | Weighted |
| 0 | @1 | 47.325153374233130 | 10.234825223367249 | 163 | 163.000 |
| @3 | 1.765644171779141 | 1.936918144190941 | 163 | 163.000 |
| @4 | 3.604969325153375 | .328213221559285 | 163 | 163.000 |
| @5 | 2.834355828220859 | .869679972933010 | 163 | 163.000 |
| @6 | 10.418404907975457 | .858597712883336 | 163 | 163.000 |
| @7 | 2291.828220858895700 | 1004.104684394689000 | 163 | 163.000 |
| @8 | .208588957055215 | .407551976993579 | 163 | 163.000 |
| @9 | .435582822085890 | .497361058406413 | 163 | 163.000 |
| @10 | .018404907975460 | .134824652379835 | 163 | 163.000 |
| @11 | 1512.606134969325000 | 1332.487461395770400 | 163 | 163.000 |
| @12 | 112.862147239263850 | 53.142179682529780 | 163 | 163.000 |
| @13 | 340.061349693251540 | 196.148019606872250 | 163 | 163.000 |
| @14 | 113.184049079754600 | 51.140479396330120 | 163 | 163.000 |
| @15 | 269.546012269938700 | 86.236482387059800 | 163 | 163.000 |
| drug | 1.484662576687117 | .501304818693349 | 163 | 163.000 |
| edema\_num | 2.932515337423313 | .275072029979224 | 163 | 163.000 |
| copper | 73.822085889570560 | 66.948010070593810 | 163 | 163.000 |
| gender | 1.079754601226994 | .271747774577746 | 163 | 163.000 |
| 1 | @1 | 53.302752293577980 | 10.126518712945188 | 109 | 109.000 |
| @3 | 5.622935779816512 | 6.165084376994954 | 109 | 109.000 |
| @4 | 3.390000000000002 | .464074147970783 | 109 | 109.000 |
| @5 | 3.339449541284404 | .735604859285484 | 109 | 109.000 |
| @6 | 11.195412844036703 | 1.040511387901140 | 109 | 109.000 |
| @7 | 1525.678899082568700 | 1112.979217475380400 | 109 | 109.000 |
| @8 | .403669724770642 | .492898946427905 | 109 | 109.000 |
| @9 | .669724770642202 | .472484486744345 | 109 | 109.000 |
| @10 | .165137614678899 | .373019745974459 | 109 | 109.000 |
| @11 | 2755.073394495412700 | 2785.170611392302000 | 109 | 109.000 |
| @12 | 141.384220183486200 | 58.202602148357340 | 109 | 109.000 |
| @13 | 422.339449541284400 | 279.250392757556800 | 109 | 109.000 |
| @14 | 142.229357798165130 | 79.691259921757350 | 109 | 109.000 |
| @15 | 250.623853211009160 | 102.739223007936720 | 109 | 109.000 |
| drug | 1.504587155963303 | .502288341474946 | 109 | 109.000 |
| edema\_num | 2.577981651376147 | .723847461699189 | 109 | 109.000 |
| copper | 139.853211009174320 | 101.714544799911540 | 109 | 109.000 |
| gender | 1.192660550458716 | .396210386971542 | 109 | 109.000 |
| Total | @1 | 49.720588235294116 | 10.587670317512423 | 272 | 272.000 |
| @3 | 3.311397058823528 | 4.579971909353926 | 272 | 272.000 |
| @4 | 3.518823529411766 | .401699450395450 | 272 | 272.000 |
| @5 | 3.036764705882353 | .853973462737392 | 272 | 272.000 |
| @6 | 10.729779411764707 | 1.008796708891267 | 272 | 272.000 |
| @7 | 1984.805147058823400 | 1112.585603776117300 | 272 | 272.000 |
| @8 | .286764705882353 | .453084357716412 | 272 | 272.000 |
| @9 | .529411764705882 | .500054262304137 | 272 | 272.000 |
| @10 | .077205882352941 | .267409856319439 | 272 | 272.000 |
| @11 | 2010.506617647059000 | 2127.179167624052800 | 272 | 272.000 |
| @12 | 124.291948529411780 | 56.870918847688465 | 272 | 272.000 |
| @13 | 373.033088235294140 | 236.025787431930900 | 272 | 272.000 |
| @14 | 124.823529411764710 | 65.556587466028260 | 272 | 272.000 |
| @15 | 261.963235294117600 | 93.479591717780990 | 272 | 272.000 |
| drug | 1.492647058823529 | .500867491435167 | 272 | 272.000 |
| edema\_num | 2.790441176470588 | .533231972044455 | 272 | 272.000 |
| copper | 100.283088235294120 | 88.618823898194440 | 272 | 272.000 |
| gender | 1.125000000000000 | .331328534518200 | 272 | 272.000 |

This table shows the means on each independent variable for people without Cirrhosis (status = 0), with Cirrhosis (status = 1) and overall means on each variable.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tests of Equality of Group Means** | | | | | |
|  | Wilks' Lambda | F | df1 | df2 | Sig. |
| @1 | .923 | 22.470 | 1 | 270 | .000 |
| @3 | .829 | 55.681 | 1 | 270 | .000 |
| @4 | .931 | 20.020 | 1 | 270 | .000 |
| @5 | .916 | 24.863 | 1 | 270 | .000 |
| @6 | .857 | 45.051 | 1 | 270 | .000 |
| @7 | .886 | 34.843 | 1 | 270 | .000 |
| @8 | .955 | 12.629 | 1 | 270 | .000 |
| @9 | .947 | 15.064 | 1 | 270 | .000 |
| @10 | .927 | 21.128 | 1 | 270 | .000 |
| @11 | .918 | 24.192 | 1 | 270 | .000 |
| @12 | .939 | 17.425 | 1 | 270 | .000 |
| @13 | .971 | 8.147 | 1 | 270 | .005 |
| @14 | .953 | 13.409 | 1 | 270 | .000 |
| @15 | .990 | 2.693 | 1 | 270 | .102 |
| drug | 1.000 | .103 | 1 | 270 | .748 |
| edema\_num | .893 | 32.200 | 1 | 270 | .000 |
| copper | .866 | 41.714 | 1 | 270 | .000 |
| gender | .972 | 7.775 | 1 | 270 | .006 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pooled Within-Groups Matrices** | | | | | | | | | | | | | | | | | | | |
|  | | @1 | @3 | @4 | @5 | @6 | @7 |  |  |  |  |  |  |  |  |  |  |  |  |
| Correlation | @1 | 1.000 | -.048 | -.178 | .107 | .156 | -.049 |  |  |  |  |  |  |  |  |  |  |  |  |
| @3 | -.048 | 1.000 | -.214 | .138 | .197 | -.331 |  |  |  |  |  |  |  |  |  |  |  |  |
| @4 | -.178 | -.214 | 1.000 | -.240 | -.095 | .332 |  |  |  |  |  |  |  |  |  |  |  |  |
| @5 | .107 | .138 | -.240 | 1.000 | .130 | -.286 |  |  |  |  |  |  |  |  |  |  |  |  |
| @6 | .156 | .197 | -.095 | .130 | 1.000 | .013 |  |  |  |  |  |  |  |  |  |  |  |  |
| @7 | -.049 | -.331 | .332 | -.286 | .013 | 1.000 |  |  |  |  |  |  |  |  |  |  |  |  |
| @8 | -.072 | .184 | -.161 | .262 | .151 | -.192 |  |  |  |  |  |  |  |  |  |  |  |  |
| @9 | -.003 | .170 | -.223 | .362 | .036 | -.188 |  |  |  |  |  |  |  |  |  |  |  |  |
| @10 | .171 | .265 | -.276 | .200 | .254 | -.174 |  |  |  |  |  |  |  |  |  |  |  |  |
| @11 | -.101 | .024 | -.039 | -.073 | -.022 | .225 |  |  |  |  |  |  |  |  |  |  |  |  |
| @12 | -.201 | .368 | -.135 | .067 | -.032 | -.113 |  |  |  |  |  |  |  |  |  |  |  |  |
| @13 | -.212 | .370 | -.028 | -.045 | -.104 | -.090 |  |  |  |  |  |  |  |  |  |  |  |  |
| @14 | -.039 | .395 | -.057 | .070 | -.062 | -.102 |  |  |  |  |  |  |  |  |  |  |  |  |
| @15 | -.126 | -.028 | .171 | -.197 | -.161 | .126 |  |  |  |  |  |  |  |  |  |  |  |  |
| drug | .127 | -.114 | -.032 | -.104 | -.099 | .001 |  |  |  |  |  |  |  |  |  |  |  |  |
| edema\_num | -.180 | -.304 | .319 | -.190 | -.305 | .258 |  |  |  |  |  |  |  |  |  |  |  |  |
| copper | -.023 | .367 | -.155 | .189 | .083 | -.268 |  |  |  |  |  |  |  |  |  |  |  |  |
| gender | .208 | -.106 | .130 | -.013 | .048 | .027 |  |  |  |  |  |  |  |  |  |  |  |  |

**Analysis 1**

**Box's Test of Equality of Covariance Matrices**

|  |  |  |
| --- | --- | --- |
| **Log Determinants** | | |
| status | Rank | Log Determinant |
| 0 | 18 | 58.173 |
| 1 | 18 | 68.868 |
| Pooled within-groups | 18 | 65.411 |

|  |
| --- |
| The ranks and natural logarithms of determinants printed are those of the group covariance matrices. |

|  |  |  |
| --- | --- | --- |
| **Test Results** | | |
| Box's M | | 799.163 |
| F | Approx. | 4.332 |
| df1 | 171 |
| df2 | 166513.139 |
| Sig. | .000 |

|  |
| --- |
| Tests null hypothesis of equal population covariance matrices. |

**Summary of Canonical Discriminant Functions**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Eigenvalues** | | | | |
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | .681a | 100.0 | 100.0 | .636 |

|  |
| --- |
| a. First 1 canonical discriminant functions were used in the analysis. |

This table tells us about the “latent variable” that we have constructed (i.e. the discriminant

function), which helps in differentiating between the groups.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wilks' Lambda** | | | | |
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 | .595 | 135.520 | 18 | .000 |

This is the multivariate test – Wilk’s lambda, just like in MANOVA. Because p < 0.05, we can say that the model is a good fit for the data. This multivariate test is a goodness of fit statistic, just like the F-test is for regression.

|  |  |
| --- | --- |
| **Standardized Canonical Discriminant Function Coefficients** | |
|  | Function |
| 1 |
| @1 | .339 |
| @3 | .145 |
| @4 | .057 |
| @5 | .107 |
| @6 | .380 |
| @7 | -.343 |
| @8 | .092 |
| @9 | .091 |
| @10 | .065 |
| @11 | .453 |
| @12 | .190 |
| @13 | .126 |
| @14 | .052 |
| @15 | -.002 |
| drug | .056 |
| edema\_num | -.059 |
| copper | .104 |
| gender | .157 |

Based on these, we can write out the equation for the discriminant function:

DF = 0.339\*age + 0.145\*bilirubin + 0.057\*albumin + 0.107\*stage + 0.380\*clot – 0.343\*time + 0.092\*skin + 0.091\*liver + 0.065\*ascites + 0.453\*phosphate + 0.190\*GOT + 0.126\*cholesterol + 0.052\*triglycerides – 0.002\*platelet + 0.056\*drug – 0.059\*edema + 0.104\*copper + 0.157\*gender

Using this equation, given someone’s scores on age, bilirubin, albumin, stage, clot, time, skin, liver, ascites, phosphate, GOT, cholesterol, triglycerides, platelet, drug, edema, cooper and gender, we can calculate their score on the discriminant function.

|  |  |
| --- | --- |
| **Structure Matrix** | |
|  | Function |
| 1 |
| @3 | .550 |
| @6 | .495 |
| copper | .476 |
| @7 | -.435 |
| edema\_num | -.419 |
| @5 | .368 |
| @11 | .363 |
| @1 | .350 |
| @10 | .339 |
| @4 | -.330 |
| @12 | .308 |
| @9 | .286 |
| @14 | .270 |
| @8 | .262 |
| @13 | .211 |
| gender | .206 |
| @15 | -.121 |
| drug | .024 |

|  |
| --- |
| Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  Variables ordered by absolute size of correlation within function. |

This table tells us the correlation between each feature and the discriminant function.

|  |  |
| --- | --- |
| **Canonical Discriminant Function Coefficients** | |
|  | Function |
| 1 |
| @1 | .033 |
| @3 | .035 |
| @4 | .146 |
| @5 | .130 |
| @6 | .406 |
| @7 | .000 |
| @8 | .207 |
| @9 | .186 |
| @10 | .253 |
| @11 | .000 |
| @12 | .003 |
| @13 | .001 |
| @14 | .001 |
| @15 | .000 |
| drug | .112 |
| edema\_num | -.117 |
| copper | .001 |
| gender | .479 |
| (Constant) | -8.241 |

|  |
| --- |
| Unstandardized coefficients |

|  |  |
| --- | --- |
| **Functions at Group Centroids** | |
| status | Function |
| 1 |
| 0 | -.672 |
| 1 | 1.005 |

|  |
| --- |
| Unstandardized canonical discriminant functions evaluated at group means |

These are the group centroids. If someone’s score on the discriminant function is closer to -0.672, then the person is likely not to have Liver Cirrhosis. If the person’s score on the DF is closer to 1.005, then the person is likely to have Liver Cirrhosis. We can figure out which group a person is in, by calculating a cut score halfway between the two centroids:

Cut score = ( -0.672 + 1.005 ) / 2 = 0.1665

If an individual person’s score on the DF is below 0.1665, then the person is likely not to have Liver Cirrhosis. If their DF score is above 0.1665, then they are likely to have Liver Cirrhosis.

**Classification Statistics**

|  |  |  |
| --- | --- | --- |
| **Classification Processing Summary** | | |
| Processed | | 272 |
| Excluded | Missing or out-of-range group codes | 0 |
| At least one missing discriminating variable | 0 |
| Used in Output | | 272 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Prior Probabilities for Groups** | | | |
| status | Prior | Cases Used in Analysis | |
| Unweighted | Weighted |
| 0 | .500 | 163 | 163.000 |
| 1 | .500 | 109 | 109.000 |
| Total | 1.000 | 272 | 272.000 |

|  |  |  |
| --- | --- | --- |
| **Classification Function Coefficients** | | |
|  | status | |
| 0 | 1 |
| @1 | .650 | .706 |
| @3 | -.442 | -.383 |
| @4 | 31.551 | 31.795 |
| @5 | 5.230 | 5.448 |
| @6 | 15.603 | 16.284 |
| @7 | -.002 | -.003 |
| @8 | 2.889 | 3.237 |
| @9 | 4.827 | 5.138 |
| @10 | 13.420 | 13.845 |
| @11 | .001 | .001 |
| @12 | .112 | .118 |
| @13 | .002 | .002 |
| @14 | .029 | .031 |
| @15 | .039 | .039 |
| drug | 10.516 | 10.703 |
| edema\_num | 20.414 | 20.217 |
| copper | .001 | .003 |
| gender | -.788 | .016 |
| (Constant) | -211.554 | -225.657 |

|  |
| --- |
| Fisher's linear discriminant functions |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Resultsa** | | | | | |
|  |  | status | Predicted Group Membership | | Total |
| 0 | 1 |
| Original | Count | 0 | 143 | 20 | 163 |
| 1 | 27 | 82 | 109 |
| % | 0 | 87.7 | 12.3 | 100.0 |
| 1 | 24.8 | 75.2 | 100.0 |

|  |
| --- |
| 1. 82.7% of original grouped cases correctly classified. |

This table gives information about actual group membership vs predicted group membership.

Overall % correctly classified = 82.7%

Sensitivity = 143/163 = 87.7%

Specificity = 82/109 = 75.2%

Positive Predictive Values (PPV) = 143/(143+27) = 84.1%

Negative Predictive Values (NPV) = 82/(82 + 20) = 80.4%

DISCRIMINANT

/GROUPS=status(0 1)

/VARIABLES=@1 @3 @4 @5 @6 @7 @8 @9 @10 @11 @12 @13 @14 @15 drug edema\_num copper gender

/ANALYSIS ALL

/METHOD=WILKS

/FIN=3.84

/FOUT=2.71

/PRIORS EQUAL

/HISTORY

/STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR TABLE

/PLOT=COMBINED SEPARATE MAP

/CLASSIFY=NONMISSING POOLED.

**Discriminant**

|  |  |  |
| --- | --- | --- |
| **Notes** | | |
| Output Created | | 11-NOV-2018 21:54:44 |
| Comments | |  |
| Input | Data | C:\Deeksha\3-1\Applied Statistical Methods\Project\_Dataset.sav |
| Active Dataset | DataSet4 |
| Filter | <none> |
| Weight | <none> |
| Split File | <none> |
| N of Rows in Working Data File | 272 |
| Missing Value Handling | Definition of Missing | User-defined missing values are treated as missing in the analysis phase. |
| Cases Used | In the analysis phase, cases with no user- or system-missing values for any predictor variable are used. Cases with user-, system-missing, or out-of-range values for the grouping variable are always excluded. |
| Syntax | | DISCRIMINANT  /GROUPS=status(0 1)  /VARIABLES=@1 @3 @4 @5 @6 @7 @8 @9 @10 @11 @12 @13 @14 @15 drug edema\_num copper gender  /ANALYSIS ALL  /METHOD=WILKS  /FIN=3.84  /FOUT=2.71  /PRIORS EQUAL  /HISTORY  /STATISTICS=MEAN STDDEV UNIVF BOXM COEFF RAW CORR TABLE  /PLOT=COMBINED SEPARATE MAP  /CLASSIFY=NONMISSING POOLED. |
| Resources | Processor Time | 00:00:00.39 |
| Elapsed Time | 00:00:00.33 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Analysis Case Processing Summary** | | | |
| Unweighted Cases | | N | Percent |
| Valid | | 272 | 100.0 |
| Excluded | Missing or out-of-range group codes | 0 | .0 |
| At least one missing discriminating variable | 0 | .0 |
| Both missing or out-of-range group codes and at least one missing discriminating variable | 0 | .0 |
| Total | 0 | .0 |
| Total | | 272 | 100.0 |

**Analysis 1**

**Box's Test of Equality of Covariance Matrices**

|  |  |  |
| --- | --- | --- |
| **Log Determinants** | | |
| status | Rank | Log Determinant |
| 0 | 6 | 41.153 |
| 1 | 6 | 45.633 |
| Pooled within-groups | 6 | 44.077 |

|  |
| --- |
| The ranks and natural logarithms of determinants printed are those of the group covariance matrices. |

|  |  |  |
| --- | --- | --- |
| **Test Results** | | |
| Box's M | | 305.727 |
| F | Approx. | 14.195 |
| df1 | 21 |
| df2 | 197487.488 |
| Sig. | .000 |

|  |
| --- |
| Tests null hypothesis of equal population covariance matrices. |

**Stepwise Statistics**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables Entered/Removeda,b,c,d** | | | | | | | | | |
| Step | Entered | Wilks' Lambda | | | | | | |  |
| Statistic | df1 | df2 | df3 | Exact F | | |  |
| Statistic | df1 | df2 |  |
| 1 | @3 | .829 | 1 | 1 | 270.000 | 55.681 | 1 | 270.000 |  |
| 2 | @6 | .762 | 2 | 1 | 270.000 | 41.973 | 2 | 269.000 |  |
| 3 | @11 | .714 | 3 | 1 | 270.000 | 35.797 | 3 | 268.000 |  |
| 4 | @7 | .661 | 4 | 1 | 270.000 | 34.240 | 4 | 267.000 |  |
| 5 | @1 | .631 | 5 | 1 | 270.000 | 31.139 | 5 | 266.000 |  |
| 6 | @12 | .621 | 6 | 1 | 270.000 | 27.003 | 6 | 265.000 |  |

Based on this table, bilirubin is the best single predictor, clot is the next best one and so on. We would include 6 variables in a model namely – bilirubin, clot, phosphate, time, age and GOT to get the best possible prediction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables in the Analysis** | | | | |
| Step | | Tolerance | F to Remove | Wilks' Lambda |
| 1 | @3 | 1.000 | 55.681 |  |
| 2 | @3 | .961 | 33.477 | .857 |
| @6 | .961 | 23.603 | .829 |
| 3 | @3 | .960 | 29.438 | .792 |
| @6 | .960 | 23.110 | .775 |
| @11 | .999 | 18.108 | .762 |
| 4 | @3 | .838 | 9.546 | .685 |
| @6 | .952 | 25.583 | .724 |
| @11 | .936 | 26.894 | .728 |
| @7 | .829 | 21.394 | .714 |
| 5 | @3 | .831 | 11.214 | .657 |
| @6 | .924 | 17.410 | .672 |
| @11 | .930 | 28.538 | .698 |
| @7 | .825 | 18.062 | .674 |
| @1 | .957 | 12.722 | .661 |
| 6 | @3 | .729 | 5.486 | .633 |
| @6 | .917 | 18.499 | .664 |
| @11 | .928 | 26.433 | .682 |
| @7 | .825 | 17.517 | .662 |
| @1 | .927 | 14.949 | .656 |
| @12 | .822 | 4.359 | .631 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables Not in the Analysis** | | | | | |
| Step | | Tolerance | Min. Tolerance | F to Enter | Wilks' Lambda |
| 0 | @1 | 1.000 | 1.000 | 22.470 | .923 |
| @3 | 1.000 | 1.000 | 55.681 | .829 |
| @4 | 1.000 | 1.000 | 20.020 | .931 |
| @5 | 1.000 | 1.000 | 24.863 | .916 |
| @6 | 1.000 | 1.000 | 45.051 | .857 |
| @7 | 1.000 | 1.000 | 34.843 | .886 |
| @8 | 1.000 | 1.000 | 12.629 | .955 |
| @9 | 1.000 | 1.000 | 15.064 | .947 |
| @10 | 1.000 | 1.000 | 21.128 | .927 |
| @11 | 1.000 | 1.000 | 24.192 | .918 |
| @12 | 1.000 | 1.000 | 17.425 | .939 |
| @13 | 1.000 | 1.000 | 8.147 | .971 |
| @14 | 1.000 | 1.000 | 13.409 | .953 |
| @15 | 1.000 | 1.000 | 2.693 | .990 |
| drug | 1.000 | 1.000 | .103 | 1.000 |
| edema\_num | 1.000 | 1.000 | 32.200 | .893 |
| copper | 1.000 | 1.000 | 41.714 | .866 |
| gender | 1.000 | 1.000 | 7.775 | .972 |
| 1 | @1 | .998 | .998 | 21.507 | .768 |
| @4 | .954 | .954 | 7.182 | .807 |
| @5 | .981 | .981 | 13.187 | .790 |
| @6 | .961 | .961 | 23.603 | .762 |
| @7 | .890 | .890 | 10.925 | .797 |
| @8 | .966 | .966 | 4.067 | .817 |
| @9 | .971 | .971 | 5.808 | .812 |
| @10 | .930 | .930 | 6.086 | .811 |
| @11 | .999 | .999 | 18.575 | .775 |
| @12 | .865 | .865 | 1.947 | .823 |
| @13 | .863 | .863 | .008 | .829 |
| @14 | .844 | .844 | .495 | .828 |
| @15 | .999 | .999 | 1.691 | .824 |
| drug | .987 | .987 | 1.144 | .826 |
| edema\_num | .908 | .908 | 10.567 | .798 |
| copper | .865 | .865 | 13.216 | .790 |
| gender | .989 | .989 | 10.702 | .797 |
| 2 | @1 | .969 | .934 | 13.749 | .725 |
| @4 | .951 | .923 | 5.332 | .747 |
| @5 | .970 | .948 | 8.989 | .737 |
| @7 | .884 | .850 | 12.754 | .728 |
| @8 | .953 | .937 | 1.927 | .757 |
| @9 | .971 | .935 | 5.265 | .747 |
| @10 | .887 | .887 | 1.969 | .757 |
| @11 | .999 | .960 | 18.108 | .714 |
| @12 | .853 | .821 | 3.534 | .752 |
| @13 | .830 | .807 | 1.012 | .759 |
| @14 | .823 | .794 | 1.993 | .757 |
| @15 | .974 | .937 | .266 | .761 |
| drug | .981 | .952 | 1.948 | .757 |
| edema\_num | .845 | .845 | 3.840 | .751 |
| copper | .865 | .837 | 11.731 | .730 |
| gender | .984 | .948 | 7.896 | .740 |
| 3 | @1 | .960 | .934 | 15.973 | .674 |
| @4 | .950 | .922 | 4.340 | .703 |
| @5 | .964 | .947 | 10.327 | .687 |
| @7 | .829 | .829 | 21.394 | .661 |
| @8 | .952 | .936 | 2.129 | .708 |
| @9 | .970 | .934 | 4.452 | .702 |
| @10 | .887 | .887 | 2.199 | .708 |
| @12 | .849 | .821 | 2.325 | .708 |
| @13 | .822 | .807 | .309 | .713 |
| @14 | .810 | .794 | .705 | .712 |
| @15 | .941 | .937 | 1.638 | .710 |
| drug | .981 | .951 | 1.799 | .709 |
| edema\_num | .844 | .844 | 4.236 | .703 |
| copper | .857 | .837 | 8.537 | .692 |
| gender | .983 | .947 | 8.026 | .693 |
| 4 | @1 | .957 | .825 | 12.722 | .631 |
| @4 | .859 | .749 | .435 | .660 |
| @5 | .900 | .773 | 4.012 | .651 |
| @8 | .930 | .810 | .549 | .660 |
| @9 | .948 | .809 | 1.850 | .656 |
| @10 | .876 | .812 | .871 | .659 |
| @12 | .849 | .730 | 2.179 | .656 |
| @13 | .821 | .713 | .447 | .660 |
| @14 | .809 | .706 | .771 | .659 |
| @15 | .932 | .821 | .638 | .659 |
| drug | .980 | .828 | 1.289 | .658 |
| edema\_num | .810 | .795 | 1.221 | .658 |
| copper | .823 | .772 | 3.828 | .652 |
| gender | .983 | .828 | 7.214 | .644 |
| 5 | @4 | .831 | .749 | .000 | .631 |
| @5 | .894 | .772 | 2.833 | .624 |
| @8 | .921 | .805 | 1.155 | .628 |
| @9 | .948 | .806 | 1.781 | .627 |
| @10 | .855 | .800 | .146 | .630 |
| @12 | .822 | .729 | 4.359 | .621 |
| @13 | .796 | .713 | 1.624 | .627 |
| @14 | .809 | .699 | .626 | .629 |
| @15 | .926 | .818 | .244 | .630 |
| drug | .961 | .824 | .403 | .630 |
| edema\_num | .790 | .790 | .294 | .630 |
| copper | .823 | .766 | 3.793 | .622 |
| gender | .946 | .823 | 3.928 | .622 |
| 6 | @4 | .822 | .727 | .041 | .620 |
| @5 | .891 | .729 | 2.428 | .615 |
| @8 | .919 | .722 | 1.320 | .617 |
| @9 | .947 | .721 | 1.909 | .616 |
| @10 | .850 | .698 | .292 | .620 |
| @13 | .771 | .660 | .825 | .619 |
| @14 | .802 | .611 | .960 | .618 |
| @15 | .921 | .727 | .115 | .620 |
| drug | .961 | .724 | .400 | .620 |
| edema\_num | .790 | .704 | .273 | .620 |
| copper | .813 | .691 | 2.926 | .614 |
| gender | .944 | .721 | 3.461 | .613 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Wilks' Lambda** | | | | | | | | | |
| Step | Number of Variables | Lambda | df1 | df2 | df3 | Exact F | | |  |
| Statistic | df1 | df2 |  |
| 1 | 1 | .829 | 1 | 1 | 270 | 55.681 | 1 | 270.000 |  |
| 2 | 2 | .762 | 2 | 1 | 270 | 41.973 | 2 | 269.000 |  |
| 3 | 3 | .714 | 3 | 1 | 270 | 35.797 | 3 | 268.000 |  |
| 4 | 4 | .661 | 4 | 1 | 270 | 34.240 | 4 | 267.000 |  |
| 5 | 5 | .631 | 5 | 1 | 270 | 31.139 | 5 | 266.000 |  |
| 6 | 6 | .621 | 6 | 1 | 270 | 27.003 | 6 | 265.000 |  |

This table shows the Wilk’s Lambda for each step. The model is a good fit for the data with just one predictor, or two predictors and so on up to 6 predictors.

**Summary of Canonical Discriminant Functions**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Eigenvalues** | | | | |
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | .611a | 100.0 | 100.0 | .616 |

|  |
| --- |
| a. First 1 canonical discriminant functions were used in the analysis. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wilks' Lambda** | | | | |
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 | .621 | 127.386 | 6 | .000 |

|  |  |
| --- | --- |
| **Standardized Canonical Discriminant Function Coefficients** | |
|  | Function |
| 1 |
| @1 | .390 |
| @3 | .271 |
| @6 | .433 |
| @7 | -.445 |
| @11 | .508 |
| @12 | .228 |

If we wanted to construct a predictive equation using just the best predictors, it would be:

DF = 0.390\*age + 0.271\*bilirubin + 0.433\*clot - 0.445\*time + 0.508\*phosphate + 0.228\*GOT

|  |  |
| --- | --- |
| **Structure Matrix** | |
|  | Function |
| 1 |
| @3 | .581 |
| @6 | .522 |
| @7 | -.459 |
| edema\_numa | -.395 |
| @11 | .383 |
| @1 | .369 |
| @4a | -.367 |
| coppera | .349 |
| @12 | .325 |
| @10a | .308 |
| @5a | .241 |
| @14a | .197 |
| @9a | .168 |
| @8a | .167 |
| @13a | .142 |
| @15a | -.095 |
| druga | -.037 |
| gendera | .036 |

|  |
| --- |
| Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  Variables ordered by absolute size of correlation within function. |
| a. This variable not used in the analysis. |

|  |  |
| --- | --- |
| **Canonical Discriminant Function Coefficients** | |
|  | Function |
| 1 |
| @1 | .038 |
| @3 | .065 |
| @6 | .463 |
| @7 | .000 |
| @11 | .000 |
| @12 | .004 |
| (Constant) | -7.252 |

|  |
| --- |
| Unstandardized coefficients |

|  |  |
| --- | --- |
| **Functions at Group Centroids** | |
| status | Function |
| 1 |
| 0 | -.637 |
| 1 | .953 |

|  |
| --- |
| Unstandardized canonical discriminant functions evaluated at group means |

If we wanted to know whether an individual’s score on this new, simpler DF suggested that they had Liver Cirrhosis or not, we must compare their score on the DF to these centroids. If their score were closer to -0.637, they are likely not to have Liver Cirrhosis and if their score was closer to 0.953, they are likely to e have Liver Cirrhosis. These centroids are different from the ones obtained when all 18 variables were in the model, just as the discriminant coefficients are different.

**Classification Statistics**

|  |  |  |
| --- | --- | --- |
| **Classification Processing Summary** | | |
| Processed | | 272 |
| Excluded | Missing or out-of-range group codes | 0 |
| At least one missing discriminating variable | 0 |
| Used in Output | | 272 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Prior Probabilities for Groups** | | | |
| status | Prior | Cases Used in Analysis | |
| Unweighted | Weighted |
| 0 | .500 | 163 | 163.000 |
| 1 | .500 | 109 | 109.000 |
| Total | 1.000 | 272 | 272.000 |

|  |  |  |
| --- | --- | --- |
| **Classification Function Coefficients** | | |
|  | status | |
| 0 | 1 |
| @1 | .372 | .433 |
| @3 | -.632 | -.529 |
| @6 | 11.966 | 12.702 |
| @7 | .002 | .001 |
| @11 | .000 | .001 |
| @12 | .077 | .084 |
| (Constant) | -77.713 | -89.492 |

|  |
| --- |
| Fisher's linear discriminant functions |

**Conclusion**

We applied four different tests to the data given after conducting some initial analysis on the descriptive statistics. First, a two-way MANOVA was carried to do find the variables which had continuous data. Graphs were plotted for the estimated marginal means for all continuous variables with usage of both factors and compared. Next, we conducted a Kruskal-Wallis H Test on the test variable histolic stage, grouped separately by gender, edema and drug. The test was significant for grouping by edema indicating that the histolic stages are different for each category of edema. Next, we applied a chi-square goodness of fit test to conclude that females have a higher tendency to have a status of liver cirrhosis than males. Finally, we did a discriminant analysis as well to predict the status column from the other features. We performed discriminant analysis at once and step by step to remove some variables from the analysis and obtained a simpler DF.