**COGS 536 FINAL, Leman Nur Erkan**

**1.a.** repeated measures ANOVA. In this study, the participants are the same but the tests are changing, therefore it is repeated measures.

**b.** dependent var.s= ponzo, contours and solid dosts, independent=control. All are continuous variables, scale is ratio variables, because the results are measurements of the lengths, which can be zero.

**c.** since a variable is not normally distributed if sig. < 0.05. in both Kolmogorov & Shapiro tests, in table 1.2 suggests that the variables are normally distributed. In table 1.3, sphericity assumption is not rejected since sig > 0.05, and lastly the three mean is different p < 0.05, in table 1.4 and also all pairwise comparisons are significant in table 1.5. Therefore independence is occuring.

**d.** in table 1.5, there are p < 0.05 by pairwise comparisons, therefore there are significant differences among the conditions. Also in table 1.4, we see the overall difference btw. the means,

For the trial, the condition is affected by sphericity F(2, 8)=111,52, p=.000.

Post hoc analysis with a Bonferroni adjustment revealed significant changes that from condition 1 to condition 2(1.20 (95% CI, 0.61 to 1.44), p < .002), and from condition 1 to condition 3 (1.9 (95% CI, 1.28 to 2.53), p = .001), and from condition 1 to condition 4 (2.8 (95% CI, 1.86 to 3.78) p = .001), and from condition 2 to condition 3 (0.88 (95% CI, 0.12 to 1.74) p = .048), and from condition 2 to condition 4 (1.80 (95% CI, .94 to 2.67) p = .003), and from condition 3 to condition 4 (0.92 (95% CI, .11 to 1.73) p = .032).

**e.** 0.05/4=0.0125 this the adjusted alpha level, therefore sig<0.0125, only the conditions 2 & 3 and 3 &4 are significantly differ from each other.

**f. The** planned contrasts compare only significant differences rather than comparing all the means of the pairs. The idea is based on weighting the means by categorizing them and then comparing the differences. By this way, more appropriate comparisons can be done by weighting the significant differences between the groups, and by this way some variables can be eliminated before the pairwise comparison, they are getting zero weights and others take -1 and +1.

**2.a.** Likert-scale variables=independent var.s, purchasing situations=dependent var.s

**b.** This test checks the equality of multiple variance-covariance matrices.(null H=covariance matrices are equal) sig.=.000 we reject the null H, no equality of matrices.

**c.** There was a statistically significant difference in purchasing situations based on likert-scales of the customers, F (14, 182) = 15.37, p < .0005; Wilk's Λ = 0.210, partial $\eta2$ = .54.

**d.** Eigenvalues shows the total amount of variance. If this value is greater than 1, it means higher than average. In the Eigenvalues table, it is seen that these two factors explain the outcome one hundred percent, in the cumulative % part. The squared version of the canonical correlation gives you the effect size. Wilks' lambda (λ), how well the prediction model fits: a small value indicates a large distribution between groups. and significance is =.000, statistically significant.

**e.** The standardized canonical discriminant function, on the other hand, allows all variables to be compared on a single scale. the variables in the standardized canonical discriminant function of the individual variable gives the highest predictor capability. Since a model with 2 different functions is formed, there are different values for the first and second functions. While the price level gave the highest positive effect in the second function, the service gave the highest effect in the first function.

**f.**The structure matrix displays the most correlated variables within each function. For example, delivery speed is the variable with the highest correlation for the first function, and in the second function the highest correlated variable is price level.
**g.** canonical discriminant function coefficients and structural matrices are consistent for price level, price flexibility, manufacturer image in both functions and service is consistent in function 1. That is, these calculated values gave the same results in both tables. Therefore, we can say that the price level has a better correlation with the second function two than the first function. In Function 1, however, price flexibility showed a better correlation. The manufacturer image contains correlations for both functions, but the second is a better predictor than the first. service is a good predictor for function 1 only. That is, these calculated values gave the same results in both tables. Therefore, we can say that the price level has a better correlation than the function-two-to the first function. In Function 1, however, price flexibility showed a better correlation. The manufacturer image contains correlations for both functions, but the second is a better predictor than the first. service is a good predictor for function 1 only.
**h.**While MANOVA can only tell the significant differences between groups, discriminant analysis can also tell which significant variable caused the difference between groups. Post hoc tests are done to show which group the difference between groups originates from. This analysis is done if the F value is significant in the analysis of variance. However, since group differences in discriminant analyses are already based on predictor variables, it does what these two tests -manova and post hoc- do at once.
**3.a.**predictors=tar, outcome=CO
**b.**r square=.911
**c.** y =mx+c; → tar x.901+1.799=CO
**d.** in anova table, regression means the variance that is explained by regression, in residuals that are error-not explained by the regression. Since the regression is more than residuals, the regression model has a better fit to the baseline model
**e.** the predictor is significant since the R squared in model summary is very high= .911, which means the predictor variable can explain the outcome variable more than 90 percent.It is same for the beta value in coefficients table, higher beta means stronger relation between the variables. Standardized coefficients are the standard versions of the coefficients, putting variables in the same scale, therefore if there is more than one variable, effects of the var.s can be compared based on their beta values.
**f.** outcome=CO, predictors= nicotine, tar
**g.** even there is r squared value in model summary, since it is multilinear regression, variance is represented y adjusted r sq. which is =.902
**h.** tar x .915 - nicotine x .225 +1.831= CO
**i.** In anova table, regression means the variance that is explained by regression, in residuals that are error-not explained by the regression. Since the regression is more than residuals, the regression model has a better fit to the baseline model
**j.** Tar is the significant predictor for the CO, beta value of the is higher than nicotine, and it is .969 and nicotine beta value -.016, that is, .016 change in nicotine corresponds to .969 change in tar.
**k.** In the coefficient table, since VIF values are greater than 1, it means that there is multicollinearity between the variables, i.e. predictors may be correlated. It means there is unnecessary information in the regression equation, and removing some variables would not significantly affect the r squared value. Other methods for regression can be used for overcoming this problem. For example, the stepwise regression method tries to derive the

highest possible regression equation for all selected independent variables, and does not include variables that have little or no effect - making multicollinearity into the model. Similarly, the best subsets regression method can be used. Or, principal component analysis can be done before regression and it can be tried to determine in advance which variables may have more impact.

**4.a.** independent/predictor=N of months in training, Does x-ray show fracture
dependent/outcome= decision is correct or not

**b.** prediction accuracy is 57.1 for baseline model and 91.4 for the new model. The percentage is increased by the new model and (chi-square=36.49, df=2, p<.000) so the new model is significantly better in the omnibus test table.

**c. & d.** odds ratio represent how much the unit chance occurs with associated explanatory variables. OR 1= 1 unit change after a unit change in the predictor, no association between the exposure and outcome == odds ratio 1 in %95 CI. The OR of the N of months is higher than 1, therefore it has a greater likelihood of getting the correct decision - for N months of training, for x-ray; %23 chance of the event occurring, less likelihood for correct decision occuring. Exp(B)=13 for N of months, therefore a change in N of months affects the output as 13 times and, Exp(B)=5.9 for x-ray, means a change in this variable affects the output as 5.9 times.

**e.** linear separability means to diagnose whether two sets of points are separated by a line. a hyperplane can be formed to separate the dots in the figure, line y=-x lines can be drawn in the figures and separate two sets of points in the spatial plane. logistic regression is basically a linear separator. After a certain point, the decision boundary is formed and this line gives two different decision outputs. Two different sets of points can be defined in the given figures.