**Submitted By: Amna Shamshad, Asad Sajid, Kanika Yadav, *Zhaoyi Xu, Shenglin Wang***

# ANALYSIS FOR GLUCOSE LEVEL OF INDIVIDUALS

## Introduction

This report covers the effects of 2 treatments/factors namely smoking type and work type on the dependent variable glucose level. The covariate variable age and the blocking variable BMI is also taken into account to find the effect of the two factors on the glucose level using five experimental design techniques that include CRD, CRAC, RBD, RBF and CRF.. For smoking type, we used 3 types; formerly smokers, never smokers, smokers. For work type, we used 4 types; children, private job, government job, self employed. For ***Completely Randomized Analysis***, the results show that treatment type smoking type does not account for significant difference in the average glucose level of each treatment. The treatment type work status has a significant effect on the average glucose level for the individuals.. For completely randomized analysis of covariance, the age as the covariate, the results indicate that after accounting for the covariate, the smoking type becomes even more insignificant. The treatment work type becomes slightly more significant after accounting for the covariate indicating that the covariate age is an important variable. For ***Randomized Block Design*** we have chosen BMI Levels as the blocking factor, and verify the effect of two treatment factors smoking status and work type on Glucose level. It is observed that the effect of smoke type is insignificant also the work type has less effect on glucose level after evaluation with block design. It created distinctive independent results of treatment factors on the dependent variable glucose which insignificant indicating BMI levels has significant impact on the glucose levels. ***For Completely Randomized Factorial Design*** we have chosen Work_Type and Smoking_Status as the categorical variables, and verify the effect of two categorical variables on Avg_Glucose_Level. We tested the main effects and the interaction of two categorical factors over a nominal variable and randomized through the testing of each type of Work against each status of Smoker. It is observed that work type is insignificant but the smoking status has more effect on glucose level after completely randomized factorial design. The interaction of two categorical factors also have a significant effect on Avg_Glucose_Level. For ***Randomized Block Factorial Design*** we have chosen Work_Type and Smoking_Status as the categorical variables and BMI 1: < 25, BMI 2: > 25 and < 30, BMI 3: > 30 as blocking variable, and verify the effect of two categorical variables on Avg_Glucose_Level with a block. We tested the main effects and the interaction of two categorical factors over a nominal variable with a block and randomized through the testing of each type of Work against each status of Smoker. It is observed that block bmi is insignificant but the Work_Type and Smoking_Status has more effect on glucose level after randomized block factorial design. The interaction of two categorical factors also have a significant effect on Avg_Glucose_Level.

## Analysis

For the effect of smoking on the glucose level of individuals,there were **3 treatment groups**. We checked for the normality tests, both tests have a p-value < 0.05, they are not normally distributed. According to the Levene's Test and Brown and Forsythe's Test, we both get p-value > 0.05, which means the condition for **homogeneity of variances ( homoscedasticity) is met**, it's equal variances across groups in statistical analysis.

The null hypothesis is H0: $\mu1 = \mu2 = \mu3$. It means the mean glucose level of individuals for all three treatment groups is equal. The alternate hypothesis is that at least one pair is significantly different i.e. $\mu i \neq \mu j$ for some i,j where i and j are greater than equal to 1 and less than 4 (integer). At least one contrast among the smoking habits is significant.

For each post hoc test, there are 3 possible pairwise tests,

Null Hypothesis: $\mu1 = \mu2$ , Alternate Hypothesis: $\mu1 \neq \mu2$

Null Hypothesis: $\mu1 = \mu3$ , Alternate Hypothesis: $\mu1 \neq \mu3$

Null Hypothesis: $\mu2 = \mu3$ , Alternate Hypothesis: $\mu2 \neq \mu3$

From CRD, the p value is 0.2052 and null hypothesis is accepted. After accounting for the covariate age, the type III analysis shows a p value of 0.9594. This shows that the ***treatment factor smoking type becomes even more insignificant after accounting for covariate age*** since the p value is much greater than 0.05. The age is an important covariate as indicated by the p value. Even after accounting for covariate age, the null hypothesis for smoking type is accepted. For treatment group 1 i.e, formerly smokers, the equation for predicting glucose level Y from covariate X is **y= 71.79 + 0.75x**. For treatment group 2, people who never smoked, the equation for predicting glucose level Y from covariate X is **y= 79.98 + 0.63x.** For treatment group 3, people who smoke, the equation for predicting glucose level Y from covariate X is **y= 80.696 + 0.58x.** One of the assumptions of analysis of covariance is homogeneity of the within-groups population regression coefficients: H0: $\beta w1 = \beta w2 = \beta w3$. The three sample regression coefficients are similar and consistent with the homogeneity assumption; $\beta w1=0.75$, $\beta w2=0.63$ and $\beta w3=0.58$. The model also assumes that the slopes relating glucose levels to age are parallel for all treatment groups i.e smoking types. We checked this assumption by including the class-by-covariate interaction, SmokingType*Age, in the model and examining the ANOVA test for the significance of this effect. The X × A interaction is not significant: $F = 0.82$, $p = 0.4400$ Hence, the homogeneity assumption is tenable.

For treatment factor work type, we used **4 treatment groups** here: children, private job, government job, self employed. For their normality tests, since both tests have a p-value < 0.05, they are **not normally distributed**. According to the Levene's Test and and Brown and Forsythe's Test, we both get p-value < 0.05, Which means the condition for **homogeneity of variances ( homoscedasticity) is not met**, it can bias estimates and affect results.

For CRDt, the null hypothesis is H0: $\mu1 = \mu2 = \mu3 = \mu4$. It means the mean glucose level of individuals for all 4 treatment groups i.e job types (children, private job, government job, self employed) is equal. The alternate hypothesis is that at least one pair is significantly different.

Null Hypothesis: $\mu1 = \mu2$ , Alternate Hypothesis: $\mu1 \neq \mu2$

Null Hypothesis: $\mu1 = \mu3$ , Alternate Hypothesis: $\mu1 \neq \mu3$

Null Hypothesis: $\mu1 = \mu4$ , Alternate Hypothesis: $\mu1 \neq \mu4$

Null Hypothesis: $\mu2 = \mu3$ , Alternate Hypothesis: $\mu2 \neq \mu3$

Null Hypothesis: $\mu2 = \mu4$ , Alternate Hypothesis: $\mu2 \neq \mu4$

Null Hypothesis: $\mu1 = \mu3$ , Alternate Hypothesis: $\mu1 \neq \mu3$

Null Hypothesis: μ2= μ4 , Alternate Hypothesis: μ2 ≠ μ4

From the CRD, we get p-value < 0.001 and we have to reject the null hypothesis, and it shows at least one pair of the group is significantly different from others.The type III analysis shows a p value of <0.0001. ***The treatment factor work type is significant after accounting for the covariate age. The p value before and after adjusting for the covariate is the same i.e <0.0001.*** For treatment group 1 i.e, children, the equation for predicting glucose level Y from covariate X age is **y= 93.01 + 0.226x**. For treatment group 2, people who have a government job,, the equation for predicting glucose level Y from covariate X is **y= 71.19 + 0.718x.** For treatment group 3, people who have a private job, the equation for predicting glucose level Y from covariate X is **y= 71.87 + 0.827x.** For treatment group 4, people who are self employed, the equation for predicting glucose level Y from covariate X is **y= 84.52 + 0.478x.** The four sample regression coefficients are somewhat similar and consistent with the homogeneity assumption; βw1=0.226, βw2=0.718, βw3=0.827, βw4=0.478. The class-by-covariate interaction, JobType*Age in the model is significant but very close to 0.05, so the equal-slopes assumption can be justified to some extent. The multiple comparison tests show that there is ***a significant difference between glucose levels of individuals with government job and children (gov job vs children), children and self employed individuals (children vs self employed individuals).*** When the means for treatment groups are compared, they indicate that ***glucose level for children is lowest and self employed individuals is highest before accounting for the covariate age***. ***After including the covariate age, self employed individuals have lowest and children have the highest glucose level.***

On conducting RBD with treatment factor : Smoke Type and BMI level as blocking variable, we see that the individual treatment factor of bmi level results in p-value of 0.0027<0.05 which is significant and smoke status results in p-value 0.7721 > 0.05 alpha indicating smoke status does not have significant effect and we do not reject null hypothesis, smoking status has no significant impact on glucose level. This shows that the ***treatment factor smoking type becomes even more insignificant after accounting for blocking variable BMI levels.***
For Work type, the individual treatment factor of bmi level results in p-value of 0.1281>0.05 which is insignificant and work type results in p-value 0.9301 > 0.05 alpha indicating work type does not have significant effect and we do not reject null hypothesis. ***The treatment factor work type becomes insignificant after including the blocking variable BMI level***. On the CRF test, the work_type factor has the p value 0.01, which is less than alpha=0.05. So it can be considered as significant and we can reject the null hypothesis. But the factor of smoking status in p value is 0.612, which is bigger than alpha=0.05 indicating smoking status does not have significant effect and we do not reject null hypothesis, smoking status is not significant with the glucose levels. The interaction of work type and smoking status has the p value of 0.720 which is greater than alpha=0.05, so the interaction of two factors is not significant with the glucose levels and we can accept the null hypothesis. Removing the interaction from the model and just looking at work_type and smoke_status without the interaction still gives us the same results; work_type is significant, smoking_status is insignificant.

On RBF, by introducing the block of BMI levels, it becomes clear that most of the variation in glucose levels is due to the block, and not because of any of the treatments. The working type factor has the p value of 0.57 greater than the alpha=0.05. So it does not significantly affect the avg_glucose_level and we can not reject the null hypothesis. For smoking_status, it has the p value of 0.804, greater than the alpha=0.05, which is also not signficant to the glucose_level, so the null hypothesis can not be rejected. The block variable block bmi has the p value of 0.0002 less than the alpha=0.05, so it significantly affects the glucose_level and we can conclude that the blocking is significant and useful. The interaction between work type and smoking status has the p value of 0.172, greater than alpha=0.05, the interaction of two factors is not significant with the glucose_level and we can conclude that the null hypothesis cannot be rejected.

Removing the interaction from the model and just looking at work_type and smoke_status, in the presence of blocking variables, without the interaction still gives us the same results; work_type is not significant, smoking_status is insignificant, blocking is significant.

**Summary**

For treatment factor work type, work type becomes significant even after accounting for the covariate age, hence age is a useful covariate.When the means are adjusted for the covariate age, the difference among the mean glucose level slightly increases. The p value before and after accounting for covariate is <0.0001. The results are significant with the significant F test of the omnibus null hypothesis. There is a significant difference between glucose levels of individuals with government jobs and children (gov job vs children), children and self employed individuals (children vs self employed individuals). The glucose level for children is lowest and self employed individuals are highest before accounting for the covariate age. After including the covariate age, self employed individuals have lowest and children have the highest glucose level. The treatment factor smoking type becomes more insignificant after including the covariate age. The difference between the means of three smoking types decreases and almost becomes negligible after adjusting for the covariate. The results are consistent with the omnibus F test as the p value becomes greater after adjusting for covariate .

Results for Randomized block design indicate, BMI Levels are impacting the most in the result of treatments with smoke status or a work type which are comparatively insignificant. The individual Type III analysis is also narrowed down to the smoke status or work type as insignificant. Smoking type becomes more insignificant after including the blocking variable and treatment factor work type becomes insignificant after including blocking variable BMI levels. Hence, we can conclude Glucose levels i.e. dependent variable is heavily impacted by BMI levels.

Results of Completely Randomized Factorial Design indicate that work type is impacting the most in the treatments, but the smoking status is not significant to the glucose levels. And the interaction of these two factors is not significant with the glucose levels. So we can conclude that Glucose levels are heavily impacted by work type only.

Results of Randomized Block Factorial Design indicate that work type and smoking status and their interaction are not significant to the glucose levels in the presence of a blocking variable of BMI. So we can conclude that Glucose levels are heavily impacted by blocking only and blocking on BMI is useful.