**Executive Summary**

Insurances are one of the major instruments of the modern-day financial markets. Understanding the complex dynamics that drive insurance costs has become more important than ever in current times. This executive summary sums up the findings of various experimental designs, deep diving into the complex relationship between demographics such as geographic regions, family size, and body composition of an individual on insurance charges. In our experimental study, we considered two research questions. In the first research question, we looked at the effect of ‘Region’ on insurance charges. In the second research question, we looked at the impact of the ‘Number of Children’ an individual has on insurance charges. We tried to answer our research questions using the following experimental design models: two Completely Randomized Designs (CRDs); two Randomized Block Designs (RBDs); one Completely Randomized Factorial Design (CRF); one Randomized Block Factorial Design (RBF); and two Completely Randomized Analysis of Covariance (CRAC).

Investigating the impact of ‘Region’ on insurance charges using Completely Randomized Design (CRD), we found that insurance charges were not equal across all levels of region (northeast, northwest, southeast, and southwest). In other words, there are variations in insurance charges among the different regions. However, conducting the post-hoc tests such as REGWQ and Scheffe’s revealed that there were no pairwise differences in insurance charges across regions. The planned comparison using orthogonal contrasts revealed that the ‘Northwest’ region was different from the combined group of other regions. We investigated the same research question using Randomized Block Design (RBD) using ‘Age Group’ as a blocking factor to control for the variability associated with age group. We concluded that region remains a significant factor even accounting for the age group. In other words, the geographic region plays a role in determining insurance charges, and this effect remains significant even when considering potential age-related variations. The post-hoc tests such as REGWQ and Tukey’s HSD determined that ‘Southeast’ region was different to ‘Northwest and Southwest’ regions. Tukey’s test for non-additivity indicated that there is significant non-additive interaction between ‘Region’ and ‘Age Group’. In other words, interaction between the main factor and blocking factor existed. Hence, the effect of main factor was not uniform across all levels of the age group. However, the specific combinations of age group and region through post-hoc tests did not show where the interactions were significant.

Investigating the impact of ‘Number of Children’ on insurance charges using (CRD), we found that insurance charges are not equal across all levels of number of children (0,1,2,3,4,5). In other words, there are variations in insurance charges among individuals having different numbers of children. However, conducting the post-hoc tests such as REGWQ and Scheffe’s revealed that there were no pairwise differences in insurance charges across all levels of number of children. The planned comparison using orthogonal contrasts revealed that individuals with 2 and 3 children were significantly different to the combined group of individuals with other numbers of children. We investigated the same research question using (RBD) using ‘Age group’ as a blocking factor to control for the variability associated with age group. We concluded that ‘Number of Children’ remains a significant factor even accounting for the age group. In other words, the number of children plays a role in determining insurance charges, and this effect remains significant even when considering potential age-related variations. The post-hoc tests such as REGWQ and Tukey’s HSD determined that insurance charges of people having 3 children were different from people having 5 children. While there were no differences in the other groups. Tukey’s test for non-additivity indicated that there was significant non-additive interaction between the number of children and the age group. In other words, interaction between the main factor and blocking factor existed. Hence, the effect of main factor is not uniform across all levels of the age group. However, the specific combinations of age group and number of children using post-hoc tests fail to identify the significance of interactions.

Investigating the impact of ‘Region’ and ‘Number of Children’ on insurance charges using Completely Randomized Factorial Design (CRF), we found that the main effect of ‘Region’ was insignificant while the main effect of ‘number of children’ was significant. While the interaction between the two factors was insignificant as well, indicating that the combined effect of the two factors was not significantly different from what would be expected based on their individual effects. Post-hoc tests of REGWQ and Bonferroni on ‘number of children’ failed to identify where exactly the difference were since all levels of ‘number of children’ fell into the same group. We investigated the same research question using Randomized Block Factorial Design (RBF) while using ‘Age group’ as a blocking factor to control for the variability associated with age group. Our findings were the same as CRF i.e. the insignificance of ‘Region’ and the significance of “Number of children’. Post-hoc test of REGWQ revealed that insurance charges of people having 3 and 2 children were different from people having 5 children, while all other pairwise differences were insignificant. While Bonferroni test concluded that insurance charge of people having no children was different from people having 2 children. Unlike in the CRF, the interaction between ‘Region' and ‘number of children’ was significant in RBF. While conducting the post-hoc tests to test the significance of interaction effect of the 24 possible combinations of ‘Region’ and ‘Number of Children’, it was revealed that not a single combination of interaction was significant.

Investigating the impact of ‘Region’ on insurance charges including ‘BMI’ as a covariate using Completely Randomized Analysis of Covariance (CRAC), we found that insurance charges were equal across all levels of region (northeast, northwest, southeast, and southwest). In other words, the ‘Region’ became insignificant from significant as indicated by the CRD. The inclusion of BMI as covariate in the model indicated that the differences among regions become less important and the variation in insurance charges could be explained by differences in BMI. This model also revealed that the assumptions such as the linear relationship between the insurance charges (dependent variable) and the BMI (covariate) in general as well as across all levels of ‘Region’ and the homogeneity of regression slope were not violated. Hence, this strengthens the reliability of the results of this experimental design.

Investigating the impact of ‘Number of Children’ on insurance charges including ‘BMI’ as a covariate using Completely Randomized Analysis of Covariance (CRAC), we found that insurance charges were different in at least one level of ‘Number of Children’ controlling for BMI. In other words, ‘Number of Children’ remained significant despite including BMI as a covariate in the model. However, the post-hoc test of Tukey’s adjusted LSMEANS failed to identify the exact nature of those differences. This model revealed that Adjusting for BMI does not make the ‘Number of Children’ less important (as was the case with region). In other words, the ‘Number of Children’ itself is a strong factor affecting insurance charges. This model also revealed that the assumption of linear relationship between the ‘Number of Children’ (dependent variable) and the BMI (covariate) in general as well as across all levels of ‘Number of Children’ was not violated. However, the assumption of homogeneity of regression slope was violated. Hence, giving rise to the biasedness in the results of this model.

In conclusion, based on the results of all the experimental design models, it can be concluded that ‘Number of Children’ is the most important factor which drives the insurance charges as compared to the ‘Region’. However, the data of the study has a few limitations which might be a source of biasedness and unreliability of results. These limitations are the violation of normality and homogeneity of variance assumptions of the dependent variable across each level of the factors. The homogeneity of regression slope in the ANCOVA is also violated which raises questions on the validity of the results. All in all, ‘Number of Children’ can be concluded as the most important factor that drives the insurance costs.