**Introduction**

This report is aimed at providing the board of Deakin University's hospital renal unit with a clear and simple understanding of the key relationships among various health indicators affecting the CKD risk score. It provides an analysis of a sample of 250 patients suffering from CKD.

**Section 1 – Answer to Questions**

**Q1:** The renal unit assumed that the patients were evenly spread across various CKD risk categories, but based on the analysis of CKD risk categories, it turns out that the given data has an uneven spread. The high-risk category has the highest frequency of 48% while the low-risk category has the lowest frequency of 6%, representing the uneven spread of patients. The CKD risk score has a central tendency of 58.14 and shows a slight negative skewness of 0.24. One low-end outlier can be found, but no high-end outlier is present. The data set has a score range between 4.32 and 96.13.

**Q2:** This question is about understanding the relationships between various health factors and CKD risk score. It is evident that alcohol has a limited association with CKD, which suggests that there might be certain underlying factors responsible for the high-risk condition apart from alcohol intake. The mean and median of blood urea has been seen to increase across the CKD risk categories, suggesting that the higher the CKD risk, elevated is the blood urea level. The correlation coefficient of haemoglobin being 0.06, suggests a weak relationship with CKD risk score, whereas, on the contrary, Serum creatinine and packed cell volume indicate a slight negative relationship with a coefficient of -0.07 and

-0.02 respectively.

**Q3:** Assuming that the data is normally distributed, probabilities are calculated for patients having a CKD risk score between 50 to 80 across different physical activities. It is evident that the active physical activity profile has the highest probability, while the typical physical activity profile has the lowest probability of patients with scores ranging between 50 to 80.

Furthermore, assuming that the blood pressure was normally distributed, the cut-off of the top one-third of patients in each CKD risk category was calculated, and it turns out that patients in the high-risk category have the highest cut-off off while the moderate-risk category has the lowest.

**Q4:** Based on the analysis of the relationship between blood glucose and CKD risk score, it is evident that the moderate risk category has the highest average blood glucose, while the urgent-risk category has the lowest average blood glucose. The z-value of blood glucose across various CKD-risk categories turned out to be 1.96 with \_\_ showing the lowest interval lower limit while \_\_\_ showing the highest interval upper limit.

**Q5:** The Renal unit assumed that the CKD risk score of each resident was more than 50. The hypothesis test was used to analyse the residence type whose mean CKD score is more than 50, and it turned out that both rural and urban areas had a mean CKD score more than 50. Urban residence showed a higher average CKD score than rural residence.

# Section 2 – Analysis of Key metric

**Chronic Kidney Disease (CKD)**

Based on the analysis of relationships between various health factors and CKD risk categories, it is clear that blood urea and serum creatinine have a strong influence on the CKD risk score, as it is directly proportional to the risk score. Patients with higher Serum creatinine values had higher CKD risk scores. Due to this consistency across the data and the fact that the coefficients of various factors like haemoglobin, serum creatinine, and packed cell volume are so close to zero that it suggests no linear association with the CKD risk score, serum creatinine and blood urea turned out to be the strongest biochemical influences on the CKD risk score. Blood urea levels show a steady increase across various CKD risk categories. Blood glucose was also varied across the four CKD risk categories, and upon analysis, it was clear that it was not by chance, as supported by the 95% confidence interval. The higher blood glucose levels were concentrated in more severe categories; it can be said that the blood glucose is strongly associated with an elevated CKD risk score. The blood pressure is also directly proportional to the CKD risk score, as the urgent-risk category showed the highest blood pressure level, while the low-risk category showed the least. Likewise, moderate alcohol intake shows the lowest value of the urgent CKD risk category, while no alcohol intake shows the highest of the same, suggesting the presence of some underlying factor affecting the score. In the case of residents, the test statistics far exceeded the baseline, rejecting the null hypothesis. This indicates that the residence does not have much influence on the risk score, and it does not matter whether the patient is from rural or urban areas; both groups are equally prone to the disease. Moreover, the variability across the groups is the same, suggesting that the data is widespread.

The renal unit made a few assumptions about the data set. The first assumption was that the patients are evenly spread across various CKD risk categories. The analysis showed that this was not the case, and the patients are unevenly spread. The high-risk group was the most frequent, while the low-risk group was the least frequent. The next assumption was that the average blood glucose level of patients of both rural and urban residence exceeded 50. The hypothesis test confirmed this assumption for both groups. Urban patients recorded slightly higher averages than the rural patients. The third assumption was that alcohol intake had a major impact on the CKD risk score, but that was not the case, as proved earlier in the report.

**Conclusion:**

The analysis of the dataset showed that the populations of the patients are not widespread. Health indicators such as serum creatinine, blood urea, blood glucose and blood pressure have a strong association with the CKD risk score, while factors such as alcohol intake and residence are not the dominant drivers of the risk scores. Likewise, lifestyle factors like physical activity have a mild influence on CKD risk score.