Question 1:

(a) More people would use the cycle paths on regular weekdays to travel to work or to travel to school or university.

(b) **Methods and Assumptions Checks:**

A Poisson log-linear model was fitted to the number of cyclists using the Hopetoun Street cycle path. The residual plot shows that we seem to have mostly captured the trend. The residual deviance is much bigger than the expected residual deviance given this degrees of freedom. The P-value for the model adequacy test is very small. Therefore we have very strong evidence that the data is over-dispersed, and there is more variation than we would expect from the Poisson model. Therefore, the model was refitted as a quasipoisson model. After fitting this, there was no evidence of interaction between Highbrook and Weekend so the interaction term was dropped.

The final model is approximately: $\log(\mu_i) = \beta_0 + \beta_1 Highbrook_i + \beta_2 Weekend1_i$

Where Weekend1_i is 1 if the ith day was a weekend or public holiday and 0 if it was a weekday, HopetounStreet$_i$, the number of cyclists using the Hopetoun Street cycle path on the ith day, has an overdispersed distribution with underlying mean μ_i. (Note: this is not a Poisson distribution in this case as we used a QuasiPoisson model.)

(c) **Executive Summary:**

We wish to the see if the number of cyclists using the Hopetown Street path was related to the numbers using the High Brook cycle path.

We found evidence that the number of cyclists using the Hopetown Street path was related to the numbers using the High Brook cycle path. However there was no evidence that this relationship depended on the type of day (regular weekday verses weekends and public holiday days.)

We estimate that for every 10 additional cyclists using the High Brook cycle path, the expected number of cyclists using the Hopetown Street path increased by between 5.8% and 14.9%.

We also estimate that on weekends and public holidays, the expected number of cyclists using the Hopetown Street path was between 61.3% and 71.4% lower than on regular week days.

(d) Medians.

Question 2:

(a) Logistic regression is a sensible approach here as we have a binary response variable: the subject either has significant coronary heart disease or they don't. Alternatively we can use logistic regression because we have a group size and the proportion with the required response (has significant coronary heart disease) for each group.

(b) The final model is approximately:

$log(odds_i) = \beta_0 + \beta_1 \times age_i$

Where the $odds_i$ are the odds of someone in the at-risk population having significant coronary heart disease.

(c) **Executive Summary:**

We were interested to see how age affected the risk of significant coronary heart disease for an at-risk population.

We have strong evidence that the odds of someone in the at-risk population having significant coronary heart disease increase as they get older.

We estimate that for each additional year of age, the odds of someone in the at-risk population having significant coronary heart disease increases by between 6.9 and 17.5%.