(1, 1pt) 7 (77 subjects but null has 69 dfs - allowing one df for the intercept leaves 7 unused subjects)

(2, 2pts) This data has many predictors (a total of $2^8 = 256$ level combinations), but only 77 subjects, so we would expect that the covariate classes based on the predictors will be small in size. Alternatively, since each subject has different food preferences, we would also expect that with so many options, the number of people who ate the exact same food items will not be large. In such a sparse situation, the observed responses and the fitted values might not be properly compared, and the chi-squared asymptotic distribution cannot be used to check the fit of the model because the class sizes are small (asymptotic condition doesn't hold). So we can't tell from the deviance whether this model fits or not.

(3, 1pt) logistic$(-2.02177) = \frac{exp(-2.02177)}{1+exp(-2.02177)} = 0.1169361$

(4, 1pt) $exp(3.196) = 24.435$. So eating the lettuce increased the odds of getting sick by a factor of about 24.

(5, 2pts) Usually we could compare the deviances and make a test using the chi-square distribution. However, note that the null df in the second (75) is larger than that in the first (69) due to the missing values. This makes a direct comparison of the deviances problematic because the data used in the two models is different.

(6, 2pts) If we assume that the food poisoning is caused by a single type of bacteria then there would be no reason to expect a synergistic combination of effects (a synergistic effect refers to a situation where the combined effect of multiple factors is greater than the sum of their individual effects). Main effects alone would be sufficient.

(7, 1pt) Greater than 5ft perch, more than 2in branch, sunny at midday. In the summary, they have one less count than the other categories in the same predictors.

(8, 2pts) The residual deviance (14.205) is a bit less than the degrees of freedom (17) so the model fits. Some might worry about sparseness but this problem was specifically excluded by the question.

(9, 2pts) They are approximate because the Wald stats are only asymptotically normal, and the null distribution is normal(0, 1).

(10, 2pts) Fit a model without time of day, i.e. y $\sim$ height+diameter+light, and compute the difference in deviance and compare to a chi-square with 2 dfs.

(11, 1.5pts) Outliers, incorrect (i.e., too simple) predictor form (e.g., might need interactions), response not Poisson (e.g., overdispersion).

(12, 2pts) Judging from the signs of the coefficient estimates gives the answer: midday in the shade on a low (<5ft) and narrow (<=2in) perch would maximize the number of lizards.

(13, 2pts) They would not change. This is because the over-dispersion model only changes the variance structure by multiplying the dispersion parameter, which does not alter the relative proportion of variances. Therefore, in estimating the coefficients using IRWLS, the weights remain unchanged. However, their standard errors will change accordingly.

(14, 2pts) Based on the information available, a reasonable estimate is $\hat{\sigma}^2 = 66.927/17 = 3.936882$. We usually estimate σ^2 by X^2/df, where X^2 is the Pearson's X^2 statistic. Based on the assumtion given in question (8), X^2 should be close to deviance (66.927) in size (because they are asymptotically equivalent).

(15, 2pts) exp(2.5977)$\times$logistic(1.945)=exp(2.5977)$\times \frac{exp(1.945)}{1+exp(1.945)} = 13.43281 \times 0.8749004 = 11.75237$

(16, 1pt) No. It makes no sense to compare these models since one is for the proportion and the other is for the total. The former deviance is based on binomial distribution, while the latter one is based on Poisson distribution.

(17, 1.5pts) Gender: nominal, Income: interval; Satisfaction: ordinal.

(18, 2pts) There will be 3 predictors: Gender, Income, and Satisfaction, each with 2, 4, and 4 levels, respectively. The response will be the number of individuals out of 104 in each of the $2 \times 4 \times 4 = 32$ different level combinations.