Additional note for Logistic Regression.

1. Comments for Problem 2 (Disease outbreak) in Lab 8.
   a. The (MLE) estimated logistic regression model is:

   $$log\left(\frac{\pi}{1-\pi}\right) = -2.29 + 0.027(\text{age}) + 0.045(\text{ses2}) + 0.253(\text{ses3}) + 1.24(\text{sec2})$$

   You can also express the results in the form of the estimated response function.

   $$\hat{\pi} = \frac{\exp(-2.29 + 0.027(\text{age}) + 0.045(\text{ses2}) + 0.253(\text{ses3}) + 1.24(\text{sec2}))}{1 + \exp(-2.29 + 0.027(\text{age}) + 0.045(\text{ses2}) + 0.253(\text{ses3}) + 1.24(\text{sec2}))}$$

   b. Interpreting the estimated slope for Age.

   After controlling the effects of other predictors, if age increases by 1,

   - the log-odds of getting the disease will increase 0.027.
   - odds of getting the disease will change by a multiplier (or factor) of e^0.027 = 1.03.
   - **Odds ratio is 1.03 (most common)**

   c. Interpreting the estimated slope for city sector 2 (sec2, dummy variable)

   - After controlling the effects of other predictors (age and social economic status), the odds ratio between city sector 2 over city sector 1 (baseline) is 3.46 (= e^1.24).
   - The odds of residents in sector 2 getting disease is 3.46 times as large as the odds of residents in sector 1.

2. For "Wald's test", we can the Z notation and the standard Normal distribution. When used to test one slope (Ho: $\beta_k = 0$ vs Ha: $\beta_k \neq 0$), it's the same test that can have 2 names. We call it "Wald's" to recognize the statistician who developed it. We can also it call it "Z-test" or "Standard Normal test" to recognize the sampling distribution used in the test. The "z value" and p-values in the logistic regression summary output is the result from Wald's test.

3. About Likelihood Ratio Test (LRT, aka, Deviance test) in logistic regression. (Problem 14.19, c, d.)
   a. LRT in logistic regression (and GLM) is similar to the partial F-test in linear regression in terms of:
   - They test the same Ho/Ha about the slopes (often more than 1 slope).

- They both use the idea of Full model vs. Reduced model.

b. LRT is different from the partial F-test in terms of
   - LRT applies to logistic regression, non-Normal responses. (Partial F-test applies to linear regression with assumptions such as constant variance and Normal errors.)
   - LRT compares the Deviance (aka. Residual deviance) from the full and the reduced models. (Partial F-test compares the SSEs from both models.)
   - LRT uses Chi-square distribution to compute the p-value for the test. (Partial F-test uses the F-distribution.)

c. The degrees of freedom used in LRT is determined by the difference of the number of parameters between the full and the reduced models. That is:
   (number of parameters in the Full model) – (number of parameters in the Reduced model)

d. To compute the upper-tail (right-hand side) probability for Chi-square distribution in R, use function pchisq( ). For example, if the difference of the Deviance between 2 models is 12.3, and the difference of the number of parameters between 2 models is 3, then, the resulting *p-value* is

$$p\text{-}value = P\left(\chi^2_{(df=3)} > 12.3\right) = 1 - P\left(\chi^2_{(df=3)} \leq 12.3\right)$$

   In R, the above probability can be calculated using code.

   1 – pchisq(12.3, df = 3)