Submitting your assignment via Canvas (Applied Categorical Data Analysis):

- Please include **your full name and the assignment number (e.g., HW#4) at the top of the first page**, as well as **a page number** at the top-right (or top-left) corner of each page.
- You must show your work to obtain full credit. Just a number will not be sufficient as a solution to a problem where a numerical answer is expected.
- Please ensure that your uploaded document is:
  - a single file, with the pages in the *correct* order;
  - in portrait mode with the pages correctly rotated;
  - clearly legible.

Read: Lecture Notes 6 - 8, and Textbook Sections 3.1 - 3.2, 4.1 - 4.2.

1. (*Modified* from Textbook problems 4.1, 4.2, pages 121-122) A study used logistic regression to determine characteristics associated with Y = whether a cancer patient achieved remission (1 = yes). The most important explanatory variable was a labeling index (LI) that measures proliferative activity of cells after a patient receives an injection of tritiated thymidine. It represents the percentage of cells that are "labeled." Table 4.8 shows the grouped data. Software (using R) reports **output I** for a logistic regression model using LI to predict $\pi = P(Y = 1)$. Answer questions according to **R output I**.

Table 4.8. Data for Exercise 4.1 on Cancer Remission

LI	Number of Cases	Number of Remissions	LI	Number of Cases	Number of Remissions	LI	Number of Cases	Number of Remissions
8	2	0	18	1	1	28	1	1
10	2	0	20	3	2	32	1	0
12	3	0	22	2	1	34	1	1
14	3	0	24	1	0	38	3	2
16	3	0	26	1	1			

Source: Reprinted with permission from E. T. Lee, *Computer Prog. Biomed.*, **4**: 80–92, 1974.

(a) Write down the equation for the logistic regression model of LI on remission in cancer patients (using the parameters β_0 and β_1).

(b) Find the estimated probabilities of remission $\hat{\pi}$ when $LI = 8,\ 26$, and 34.

(c) Show that the rate of change in $\hat{\pi}$ is 0.036 when $LI = 26$.

(d) The lower quartile and upper quartile for LI are 14 and 28. Show that $\hat{\pi}$ increases by 0.42, from 0.15 to 0.57, between those values.

(e) When LI increases by 1, show the estimated odds of remission multiply by 1.16.

(f) Construct a 95% confidence interval for the true probability π of remission when $LI = 26$.

(g) Conduct a Wald test for the LI effect at $\alpha = 0.05$. Interpret.

(h) Construct a 95% Wald confidence interval for the odds ratio corresponding to a 1-unit increase in LI. Interpret.

(i) Conduct a likelihood-ratio test for the LI effect at $\alpha = 0.05$. Interpret.

(j) Construct the 95% likelihood-ratio confidence interval for the odds ratio. Interpret.

2. For the horseshoe crab data ("Crabs.txt"), fit the logistic regression model for π = probability of a satellite, using **weight** ($= X$) as the predictor. Use R to complete Parts (a) through (h). **Also provide your R codes and outputs.**

   (a) Report the ML prediction equation (using the estimates of β_0 and β_1).

   (b) Find $\hat{\pi}$ at the weight values 1.20 kg, 2.44 kg, and 5.20 kg, which are the sample minimum, mean, and maximum.

   (c) Find the rate of change in $\hat{\pi}$ when $x = 3.0$ kg.

   (d) Find the weight at which $\hat{\pi} = 0.50$.

   (e) Construct a 95% Wald confidence interval to describe the effect of weight on the odds of a satellite. Interpret the interval.

   (f) Conduct the Wald test of the hypothesis that weight has no effect. Report the value of the test statistic and state your conclusion using $\alpha = 0.05$.

   (g) Conduct the likelihood-ratio test of the hypothesis that weight has no effect. Report the value of the test statistic and state your conclusion using $\alpha = 0.05$.

   (h) Construct the 95% likelihood-ratio confidence interval for the odds ratio. Interpret.

```
# R output I (for Problem #1):

# Data: LI = Explanatory variable, Y = Response variable

> LI <- c(8,8,10,10,12,12,12,14,14,14,16,16,16,18,
+ 20,20,20,22,22,24,26,28,32,34,38,38,38)

> y <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,1,0,0,1,1,0,1,1,1,0)

> length(y)
[1] 27          # 27 patients

> length(LI)
[1] 27


# Fitting using the glm( ) function: default link function, (link = "logit")

> fit1 <- glm(y ~ LI, family=binomial)

> summary(fit1)

 Coefficients:
             Estimate Std. Error z value  Pr(>|z|)
(Intercept) -3.77714    1.37862  -2.740   0.00615
LI           0.14486    0.05934   2.441   0.01464
---
    Null deviance: 34.372  on 26  degrees of freedom
Residual deviance: 26.073  on 25  degrees of freedom


> vcov(fit1)
             (Intercept)           LI
(Intercept)   1.90060430 -0.076525261
LI           -0.07652526  0.003521352


> confint(fit1)
                  2.5 %     97.5 %
(Intercept) -6.9951909 -1.4098443
LI           0.0425232  0.2846668
```