HW 6

1] A study used logistic regression to determine characteristics associated with $Y =$ if 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 recieves an injection of tritiated thymidine. It represents the percentage of cells that are "labeled." The table shows the grouped data. Software reports the 2nd table for a logistic regression model using LI to predict $\pi = P(Y=1)$.

(Tables can be viewed in textbook)

a) Show how software obtained $\hat{\pi} = 0.068$ when $LI = 8$

The model is that of a linear equation that outputs a log odds ratio.
So by using the model to find an appropriate log odds ratio, then solving the log odds ratio, a probability of success can be found.

model: $\text{logit}(\pi) = -3.7771 + 0.1449\, LI \rightarrow \text{logit}(\pi(8)) = -3.7771 + 0.1449(8) = -2.6179$

$$\pi(8) = \frac{\exp(-2.6179)}{1 + \exp(-2.6179)} \boxed{= 0.068}$$

b) Show that $\hat{\pi}(26.0) = 0.50$

$\text{logit}(\pi(26.0)) = -3.7771 + 0.1449(26.0) = -0.0097$

$$\pi(26.0) = \frac{\exp(-0.0097)}{1 + \exp(-0.0097)} \boxed{= 0.49758 \approx 0.50}$$

c) Show that the rate of change in $\hat{\pi}$ is 0.009 when $LI = 8$ and 0.036 when $LI = 26$.

rate of change $= \beta \cdot \pi(x) \cdot [1 - \pi(x)]$

rate of change $(LI = 8) = (0.1449)[(0.068)(1 - (0.068))] \boxed{= 0.00918 = 0.009}$

rate of change $(LI = 26) = (0.1449)[(0.50)(1 - (0.50))] \boxed{= 0.03623 \approx 0.036}$

d) The lower quartile and the 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.

$\text{logit}(\pi(14)) = -3.7771 + .1449(14) = -1.7485 \rightarrow \pi(14) = \frac{\exp(-1.7485)}{1 + (-1.748)} = 0.15$

$\pi(26) = 0.50$

$\text{logit}(\pi(28)) = -3.7771 + 1.449(28) = 0.2801 \longrightarrow \pi(28) = \frac{\exp(0.2801)}{1 + \exp(0.2801)} = 0.57$

increasing

$\pi(28) - \pi(14) = 0.57 - 0.15 \boxed{= 0.42}$

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

odds $= \exp(\alpha + \beta X) = e^{\alpha}(e^{\beta})^{X}$

$\hookrightarrow (e^{\beta})^{X} = e^{\beta X} \rightarrow e^{\beta(X+1)} = e^{\beta X} + e^{\beta}$

$e^{\beta} = e^{(0.1449)} = 1.1559 \approx 1.16$

$\therefore$ Every increase in x of 1 results in a 1.16 multiplicative increase in odds of remission.

3] In the first nine decades of the 20^{th} Century in Baseball's National League, the percentage of times the starting pitcher pitched a complete game were: 72.7, 63.4, 50.0, 44.3, 41.6, 32.8, 27.2, 22.5, 13.3 (by decade).

a) Treating the number of games the same in each decade, the linear probability model has ML fit $\hat{\pi} = 0.7578 - 0.0694x$, where x is the decade ($x = 1, 2, \ldots 9$) Interpret -0.0694.

-0.0694 is the β of this model. It shows that the probability of a starting pitcher pitching an entire game decreases with each following decade. Each 1 decade increase in X decreases the estimation of the odds of complete games being pitched by a multiplicative factor of $e^{-0.0694} = 0.93295$, or said differently, a decrease of roughly 0.07.

b) Substituting $x = 12$, predict the percentage of complete games for 2010-2019 (decade 12), Is this prediction plausible? Why?

$\hat{\pi}(12) = 0.7578 - 0.0694(12) \boxed{= -0.075}$

This is not a plausible prediction, it is not possible to have a negative percentage of complete games.

c) The logistic regression ML fit is $\hat{\pi} = \exp(1.148 - 0.315x) / [1 + \exp(1.148 - 0.315x)]$. Obtain $\hat{\pi}$ for $x = 12$. Is this more plausible than the prediction in (b)?

$$\hat{\pi}(12) = \frac{\exp(1.148 - 0.315(12))}{1 + \exp(1.148 - 0.315(12))} \quad \boxed{= 0.067 \rightarrow 6.7\%}$$

This is more plausible because it fits better with the decreasing trend observed in the data, and it is not a negative result like in (b).

4| Consider the snoring and heart disease data:

Snoring		Heart Disease Yes	No
Never	(0)	24	1355
Occasional	(2)	35	603
Nearly Every Night	(4)	21	192
Every Night	(5)	30	224

the logistic regression ML fit is $logit(\hat{\pi}) = -3.866 + 0.397x$.

a) Interpret the estimated effect of x.

For every unit increase in snoring level the odds increase by $e^{(0.397)} = 1.487 \approx 1.5$

b) Estimate the probabilities of having heart disease at snoring levels 0 and 5.

$logit(\pi(0)) = -3.866 + 0.397(0) = -3.866 \rightarrow \pi(0) = \frac{exp(-3.866)}{1 + exp(-3.866)} = \boxed{0.02051}$

$logit(\pi(5)) = -3.866 + 0.397(5) = -1.881 \rightarrow \pi(5) = \frac{exp(-1.881)}{1 + exp(-1.881)} = \boxed{.13227}$

c) Describe the estimated effect of snoring on the odds of heart disease.

Increases in snoring are estimated to increase the odds of contracting heart disease.