UNIT 12 HW

1. The **Donner Party** was a group of American pioneers led by George Donner and James F. Reed who set out for California in a wagon train in May 1846. They were delayed by a series of mishaps and mistakes, and spent the winter of 1846–47 snowbound in the Sierra Nevada. Some of the pioneers resorted to cannibalism to survive. The Donner.txt data set contains the ages, genders and status (died or survived) of each person in the party.
   A. Use proc logistic to estimate the probability of survival for all the members of the party. Hint: use "output out = *datasetname* predpobs = I" in your SAS code.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Age	0.925	0.860	0.995
Sex Female vs Male	4.940	1.124	21.716

Women survived around 4.94% higher than men

   B. Interpret the age parameter estimate from a model that accounts for the gender as well and provide a 95% confidence interval.

The 95% confidence interval is (1.124,21.716)

   C. Finally, test the claim that there is a higher incidence of survival for males than for females after accounting for age. To provide evidence provide an estimate and interpretation (and confidence interval) of the odds ratio (odds or survival of males / odds of survival for females) again, after accounting for age.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	3.2304	1.3870	5.4248	0.0199
Age		1	-0.0782	0.0373	4.3988	0.0360
Sex	Male	1	-1.5973	0.7555	4.4699	0.0345

Parameter Estimates and Wald Confidence Intervals				
Parameter		Estimate	95% Confidence Limits	
Intercept		3.2304	0.5120	5.9488
Age		-0.0782	-0.1513	-0.00512
Sex	Male	-1.5973	-3.0780	-0.1165

Men had a lower chance of surviving than women by 1.5973% when accounting for age and with a 95% confidence interval of (-3.0780,-0.1165) from the Wald Test.

2. Using the Donner Party data again and the model that models the survival of the subject against the age and gender, provide a confusion table using the probability level threshold of .6 (classify a subject as survived if the estimated probability of survival is greater than .6.) Include In your answer the estimated mis-classification rate as well as the sensitivity and specificity. Finally, interpret the sensitivity and specificity statistics (one sentence each is fine.)

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.600	9	23	2	11	71.1	45.0	92.0	18.2	32.4

3. Consider again the bird keeping and lung cancer study. Remember that logistic regression can provide estimates of the probability. Why, however, would an estimate of the probability of getting lung cancer for a 40 year old female of high socio economic status who has smoked for 10 years and kept a bird not be appropriate here? Hint: This is the second case study in the book. You will need to read about the case study to get this answer. (One or two sentences is fine.)
   I really don't know what the answer would be for this and I read the case study in book also.