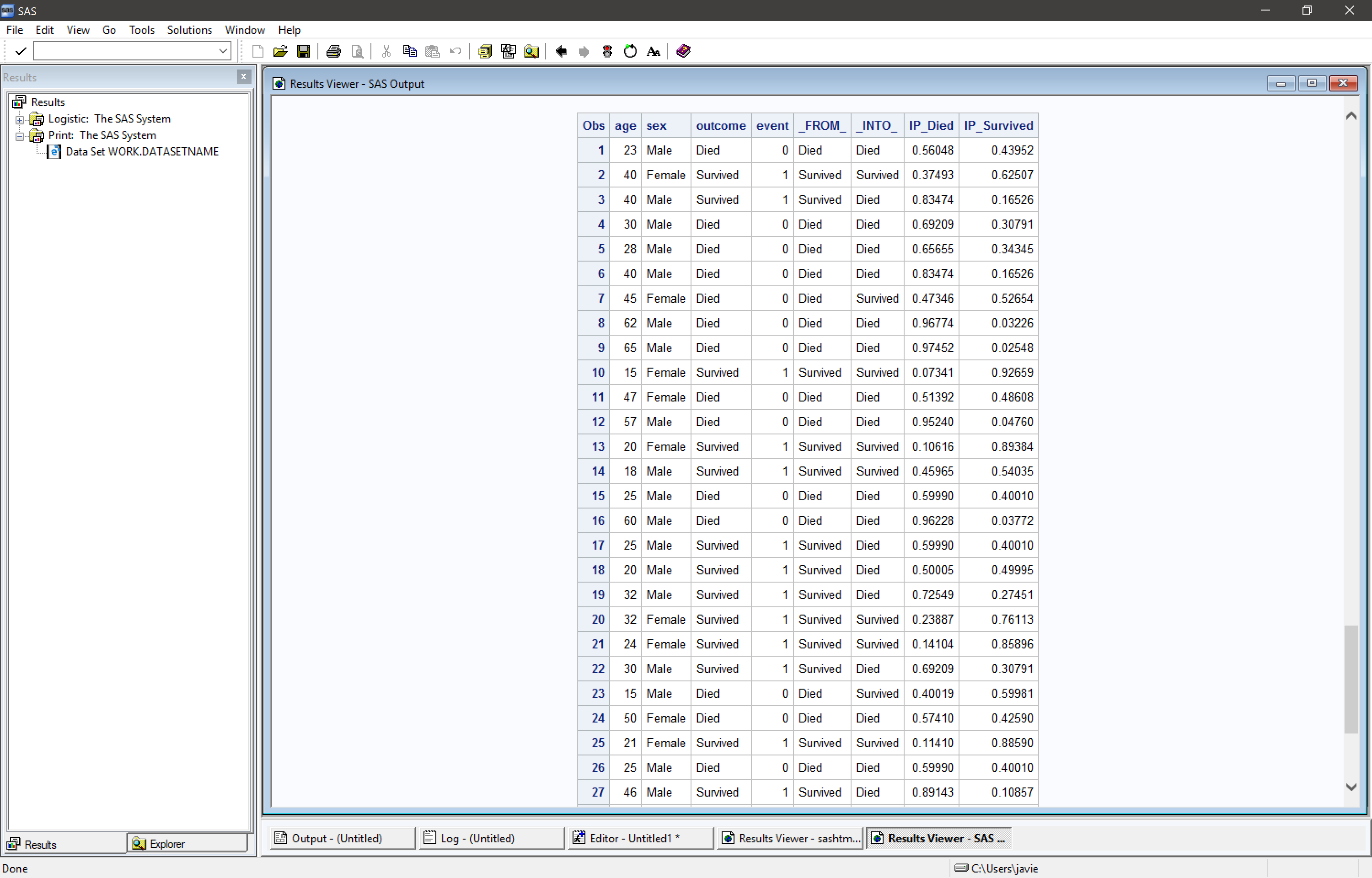
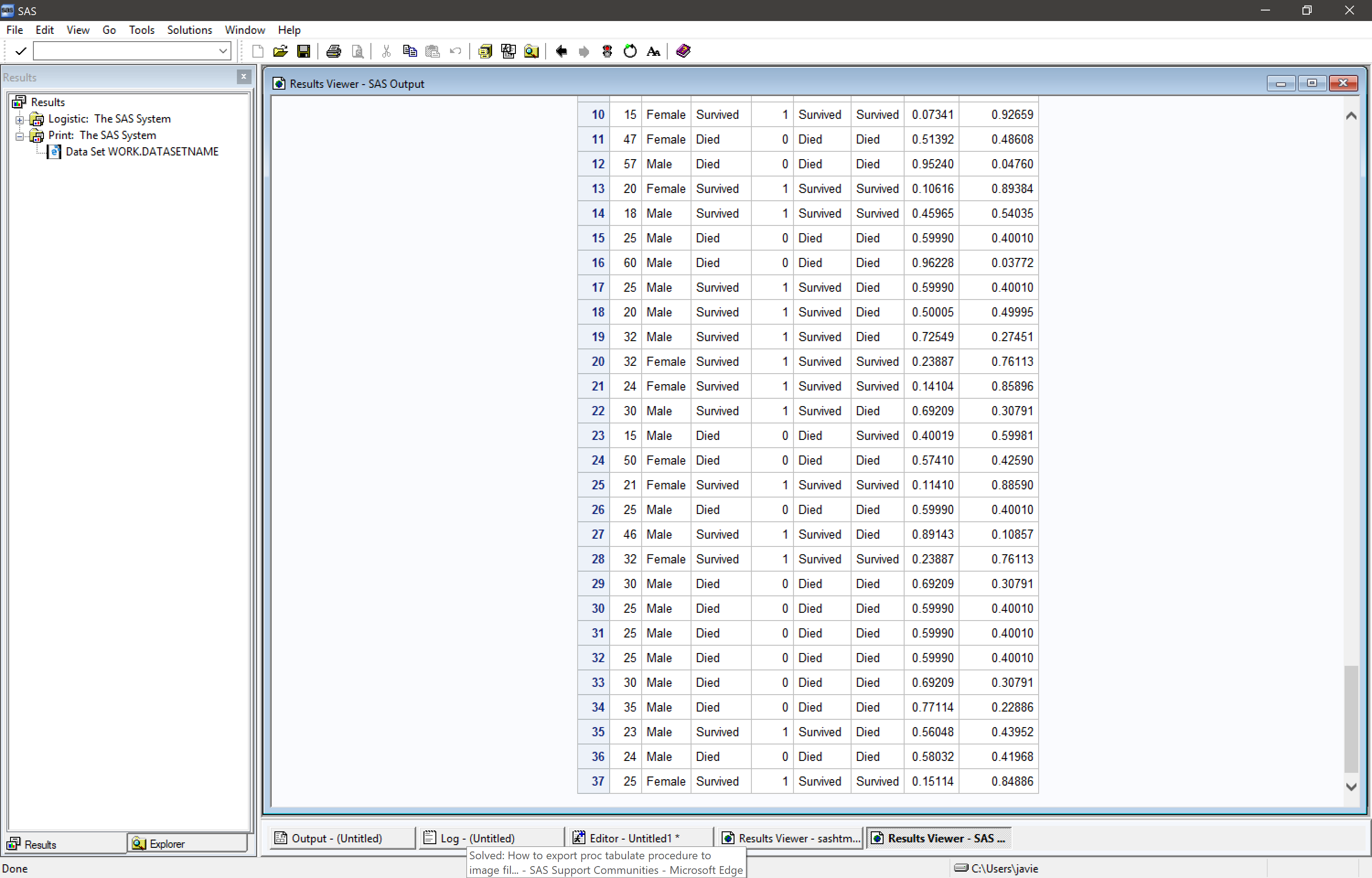
UNIT 12 HW

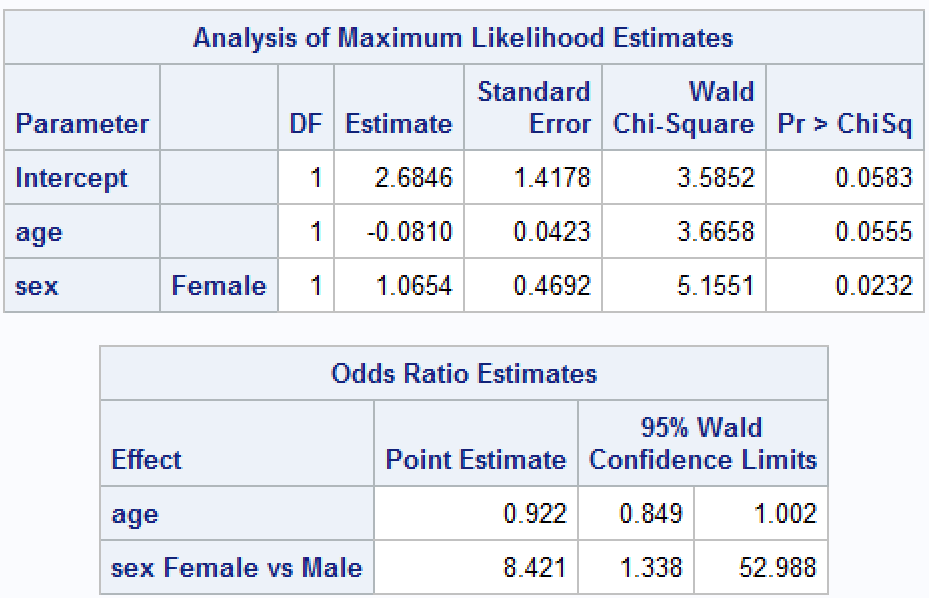
1. The **Donner Party** was a group of [American pioneers](https://en.wikipedia.org/wiki/American_pioneers) led by [George Donner](https://en.wikipedia.org/wiki/George_Donner) and [James F. Reed](https://en.wikipedia.org/wiki/James_F._Reed) who set out for California in a [wagon train](https://en.wikipedia.org/wiki/Wagon_train) 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](https://en.wikipedia.org/wiki/Sierra_Nevada_(U.S.)). Some of the pioneers resorted to [cannibalism](https://en.wikipedia.org/wiki/Cannibalism) to survive. The Donner.txt data set contains the ages, genders and status (died or survived) of each person in the party.
   1. 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.





The IP\_Survived column dictates the probability of survival for each member.

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

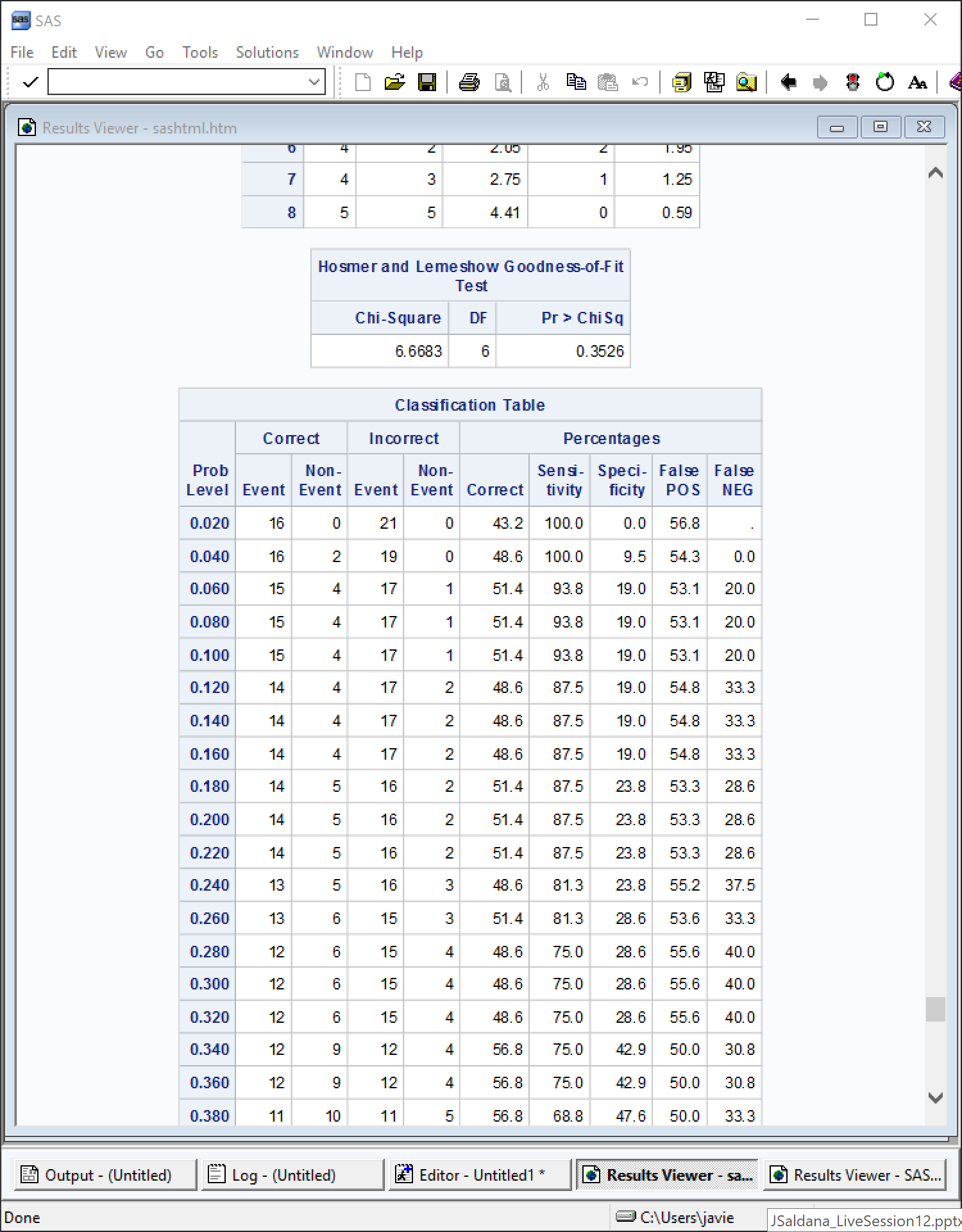


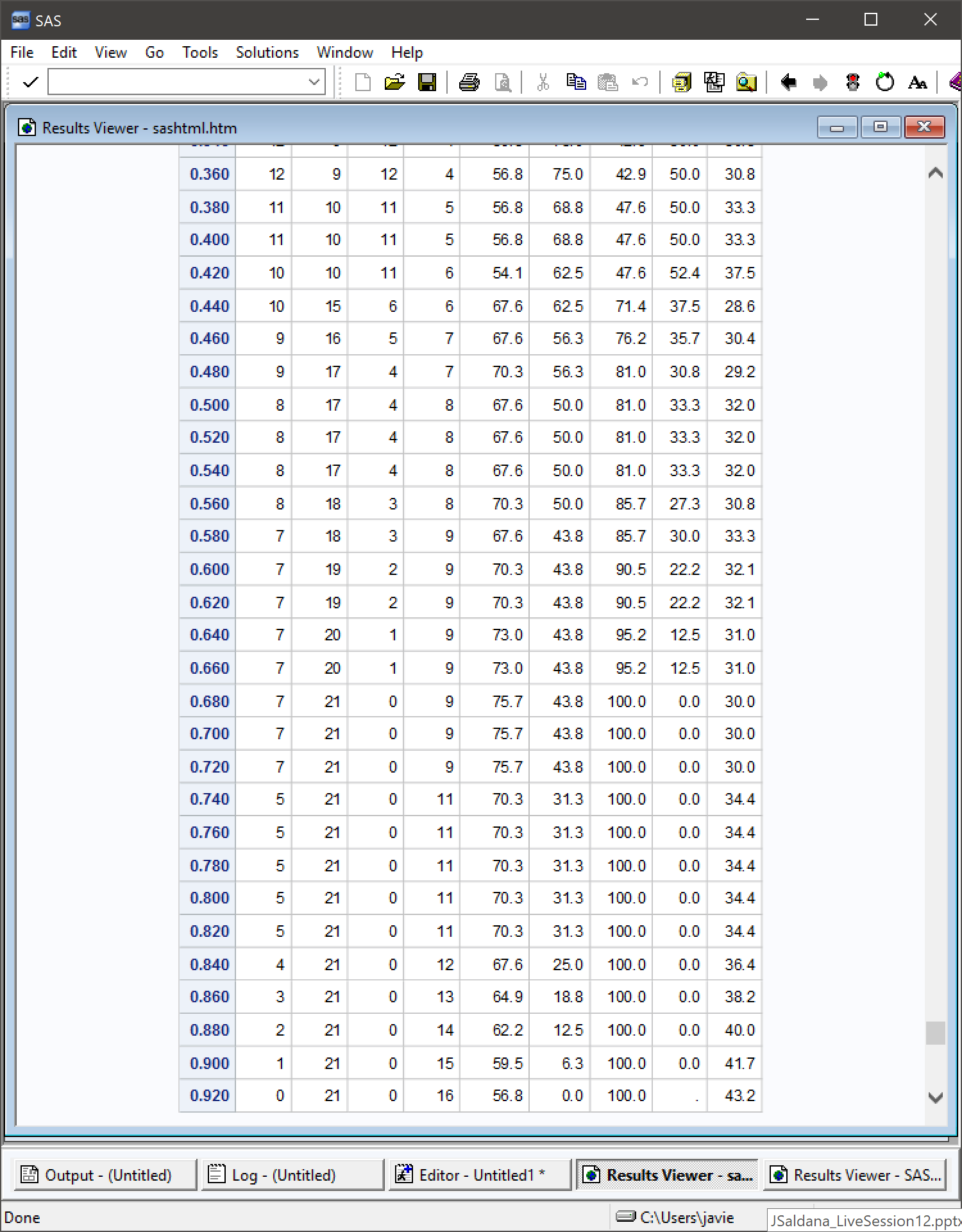
The odds ratio increases by a mean of .992 with a 95% confidence level between 0.849 and 1.002 for every unit change in age.

* 1. 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.

1. 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.)

The misclassification rate is 29.7%, which means that 29.7% of the time, a person who died was classified as survived. Sensitivity pertains to the true positive rate, which is the ability to correctly identify those likely to survive. Specificity is the true negative rate, which is the ability to correctly identify those not likely to survive.





1. 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.)

Bird keeping isn’t the cause of lung cancer. Instead, they found that there are potential confounding variables that increase the likelihood of dysfunction of lung macrophages, which can diminish the immune system response. Thus, increasing the odds of getting lung cancer (or any other disease for that matter).