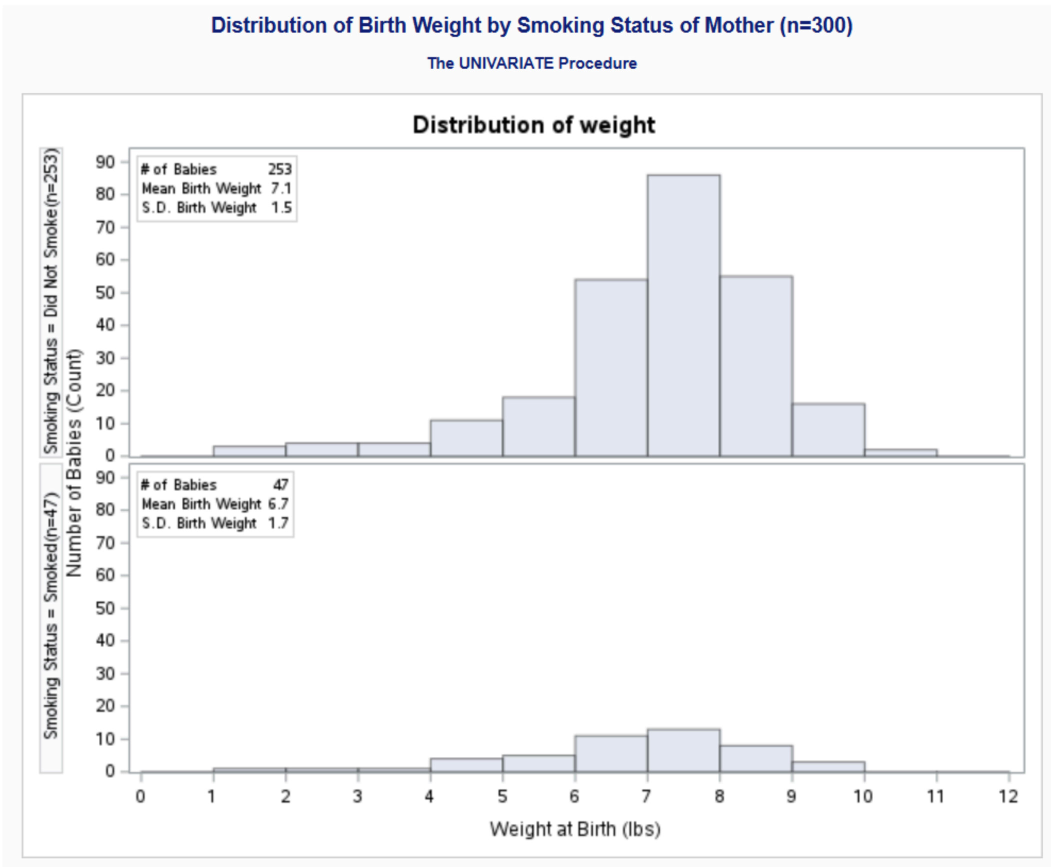
I have completed the Course Evaluation.

"I have neither given nor received unauthorized help on this assignment" - Audrija Bhattacharya

Question 1

a. Stacked Histograms



b. Risk Difference Test
   i. H_0: $p_{smoking} = p_{nonsmoking}$ The proportion of babies born at a low birth weight is the same between mothers who do and do not smoke in the population of babies born in NC.
      H_a: H_0: $p_{smoking} \neq p_{nonsmoking}$ The proportion of babies born at a low birth weight is not the same between mothers who do and do not smoke in the population of babies born in NC.
   ii. Results: Risk Difference = 0.08, ASE= 0.06, Z = 1.35, p = 0.18

The FREQ Procedure

Frequency Percent Row Pct Col Pct

Table of smoke by lowbirthweight	lowbirthweight		
smoke	NO	YES	Total
NS	214 71.33 84.58 85.60	39 13.00 15.42 78.00	253 84.33
S	36 12.00 76.60 14.40	11 3.67 23.40 22.00	47 15.67
Total	250 83.33	50 16.67	300 100.00

Risk Difference Test	
H0: P1 - P2 = 0 Wald Method	
Risk Difference	0.0799
ASE (H0)	0.0592
Z	1.3496
One-sided Pr > Z	0.0886
Two-sided Pr > |Z|	0.1771
Column 1 (lowbirthweight = NO)	

iii. Interpretation: Under the assumption that the proportion of babies born at a low birth weight is the same between mothers who do and do not smoke in the population of babies born in NC, obtaining a Z statistic extreme as 1.35 or more is 0.18. At a 0.05 significance level, this is not a statistically significant result, and our result is consistent with the assumption of the proportion of babies born at a low birth weight being the same between mothers who do and do not smoke in the population of babies born in NC.

c. 95% Confidence Interval: (-0.21, 0.05), where the unit is the difference in proportion between (lowbirthweight =yes)/smoking and (lowbirthweight =yes)/non smoking.
   i. This means that if we were to take many repeated samples from the population and calculate a 95% confidence interval of difference in proportion between babies with low birth weight that have smoking and non smoking mothers for each sample, approximately 95% of samples would contain the true population difference.
   ii. This confidence interval contains the value 0, which is consistent with our failure to reject the null assumption of no difference. The interval containing 0 means that having a difference of 0 in the population is among the plausible values. However, although both of these results are consistent with our null assumption of no difference, they do not provide evidence for it nor do they prove it.

d. Both the Chi Squared test and the Fisher's exact test can be used in this scenario. We are able to use the Chi Squared test for a 2x2 table, as $Z^2 = X^2$, and we know the sample fulfills the risk difference assumptions and additionally has fixed rows and columns with sufficient cell counts, meaning it fulfills all necessary assumptions. The scenario in which we must use the Fisher's exact test is when the cell counts are insufficient, as the Chi Squared approximation requires a minimum sample size in each cell, while Fisher's does not. Since the counts in each cell are all greater than or equal to 5, and conservatively are also all greater than or equal to 10, both tests work in this scenario.

Question 2

a. $Risk_{Ivermectin}$:52/241 = 21.6%, $Risk_{Control}$: 43/249 = 17.3%
   i. Risk Ratio: 21.6/17.3 = 1.25. This risk ratio means that in the sample, patients given Ivermectin were at a 25% higher risk of experiencing a progression to severe disease than the patients given the control medication.

b. SAS Calculations
   i. Odds Ratio: 1.32

Odds Ratio and Relative Risks			
Statistic	**Value**	**95% Confidence Limits**	
Odds Ratio	1.3181	0.8407	2.0664
Relative Risk (Column 1)	1.2494	0.8690	1.7964
Relative Risk (Column 2)	0.9479	0.8688	1.0343

Sample Size = 490

   ii. This odds ratio means that in the sample, the group treated with Ivermectin had 1.32 times the odds of experiencing a progression to severe disease compared to the control group.
   iii. The 95% Confidence Interval is (0.8407, 2.0664)

c. Since there are 5 secondary tests, the Bonferroni correction of the p-value for *all cause in hospital mortality* would be $0.09 * 5 = 0.45$. This value would then be interpreted against the threshold of $\alpha = 0.05$, and we would get a nonsignificant result as 0.45 is greater than 0.05. Conversely, we could have fixed p value at 0.09, and instead made our threshold $0.05/5 = 0.01$, but we were told to hold our α at 0.05 in this scenario. Either way, 0.09 is greater than 0.01 and our result is still nonsignificant.