BIOS 665: Problem Set 5

Assigned: November 9, 2017

Due: November 28, 2017

Reminder: For all hypothesis tests, please state the method, the null hypothesis, the test statistic, the degrees of freedom, the p-value, and the interpretation of the test using a two-sided significance level of 5%, unless otherwise stated.

Helpful hints: For estimates and tests, simply copying and pasting SAS output without any commentary will not earn full credit, especially on exams. Highlighting is not considered commentary. However, commentary can be as simple as: The 95% CI for the odds ratio is (\_, \_).

I have followed the Honor Code. Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. The contingency table shown below displays data from a randomized study to compare an existing dental treatment to an experimental dental treatment for the prevention of dental caries (cavities). An event was defined as the first occurrence of at least one cavity detected at a follow-up visit (such that a participant could only have experienced the event once during the study). Following treatment, there were biannual follow-ups lasting for a period of 2 years.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Treatment | No caries by 2 years | Follow-up period for dental caries (months) | | | |  | Follow-up period for withdrawal (months) | | | | Total |
| 0-6 | 6-12 | 12-18 | 18-24 |  | 0-6 | 6-12 | 12-18 | 18-24 |
| Existing | 39 | 2 | 5 | 14 | 35 |  | 2 | 2 | 7 | 10 | 116 |
| Experimental | 60 | 1 | 2 | 8 | 17 |  | 2 | 1 | 10 | 15 | 116 |

* 1. Present the data in life table format separately for each treatment group.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Treatment= Existing | | |  |
| Interval | No Caries | Dental Caries | Withdrawal | At Risk |
| 0-6 | 112 | 2 | 2 | 116 |
| 6-12 | 105 | 5 | 2 | 112 |
| 12-18 | 84 | 14 | 7 | 105 |
| 18-24 | 39 | 35 | 10 | 84 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Treatment= Experimental | | |  |
| Interval | No Caries | Dental Caries | Withdrawal | At Risk |
| 0-6 | 113 | 1 | 2 | 116 |
| 6-12 | 110 | 2 | 1 | 113 |
| 12-18 | 92 | 8 | 10 | 110 |
| 18-24 | 60 | 17 | 15 | 92 |

* 1. Provide life table estimates for the cumulative probabilities (and standard errors) of no occurrence of a cavity by the end of each of the four periods for each treatment group. State the assumptions for these estimates, and assume subjects who withdraw may be treated as not having caries at the time of withdrawal.

Assumptions:

* The first observation for each interval/stratification level includes subjects who withdrew and the second level will include subjects who has the event
* Patients who experienced the event (cavities) are not censored
* Withdrawal is independent of condition being studied
* Multiple withdrawals occur uniformly throughout the interval

|  |  |  |
| --- | --- | --- |
| Interval | Estimated Survival Rates | Standard Errors |
| Treatment= Existing | | |
| 0-6 | 0.9826 | 0.0122 |
| 6-12 | 0.9383 | 0.0226 |
| 12-18 | 0.8089 | 0.0376 |
| 18-24 | 0.4505 | 0.0498 |
| Treatment=Experimental | | |
| 0-6 | 0.9913 | 0.00866 |
| 6-12 | 0.9737 | 0.0150 |
| 12-18 | 0.8995 | 0.0288 |
| 18-24 | 0.7185 | 0.0455 |

* 1. Which treatment has a more favorable outcome? Provide a statistical test at the two-sided 0.05 significance level.

Ho: The distribution of time to cavity is the same for existing and experimental treatment.

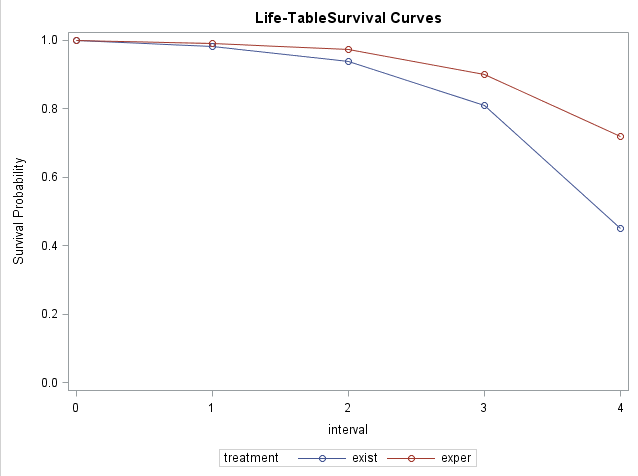
H1: Otherwise

Mantel-Cox Test

DF=1

= 13.4851

P-Value=0.0002



Conclusion:

The p-value is less than alpha of 0.05 so we reject Ho.

There is a difference in survival between existence and experimental treatment.

The experimental treatment has the more favorable outcome.

1. Fit a piecewise exponential model to the data presented in Problem 1 in order to describe the relationship of time to occurrence of dental caries with regard to the main effects for treatment and follow-up period, as well as considering their interaction. When necessary, use 0-6 months as the reference group for follow-up period, and use the Existing treatment as the reference group for treatment.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Treatment= Existing | | | |
| Interval | No Caries | Dental Caries | Withdrawal | Total person month exposure |
| 0-6 | 112 | 3 | 3 | 690 |
| 6-12 | 105 | 5 | 2 | 651 |
| 12-18 | 84 | 14 | 7 | 567 |
| 18-24 | 37 | 35 | 12 | 363 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Treatment= Experimental | | | |
| Interval | No Caries | Dental Caries | Withdrawal | Total person month exposure |
| 0-6 | 114 | 2 | 2 | 696 |
| 6-12 | 110 | 3 | 1 | 672 |
| 12-18 | 92 | 9 | 9 | 606 |
| 18-24 | 58 | 17 | 17 | 450 |

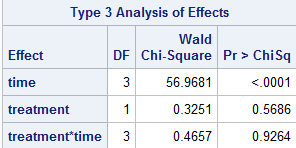
1. Mathematically specify the structure of the model, including the interaction, defining all variables used in the model. You do not need to interpret model parameters.

Logit()= α+++++++

|  |  |  |
| --- | --- | --- |
| Parameter | Value | Definition |
| α | -5.434 | Intercept: existing treatment and 0-6 month follow up |
|  | .572 | 6-12 month follow up |
|  | 1.757 | 12-18 month follow up |
|  | 3.196 | 18-24 month follow up |
|  | -.415 | Experimental treatment |
|  | -.130 | Interaction for time 6-12 and experimental treatment |
|  | -.103 | Interaction for time 12-18 and experimental treatment |
|  | -.584 | Interaction for time 18-24 and experimental treatment |

\*with is the number of caries and n is the person month exposure

H0: H1=otherwise



Wald Chi-Square ~

Wald Chi-Square=0.4657

p-value=0.9264

Since the p-value is greater than alpha=0.05, we fail to reject H0 and conclude that the time\*treatment interaction is not significant.

1. Regardless of your conclusion from (b), fit the model with main effects for treatment and follow-up period, but not including their interaction. Interpret the estimated model parameters. Provide a 95% confidence interval for the model parameter (or an appropriate transformation of this parameter) corresponding to the treatment variable.

Logit()= α++++

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter | Definition | Value | Interpretation |
| α | Intercept: existing treatment and 0-6 month follow up | -5.299 | Log incidence density for existent treatment and 0-6 month follow up |
|  | 6-12 month follow up | .524 | Increment for 6-12 month follow up |
|  | 12-18 month follow up | 1.722 | Increment for 12-18 month follow up |
|  | 18-24 month follow up | 2.987 | Increment for 18-24 month follow up |
|  | Experimental treatment | -.793 | Increment for treatment experimental |

|  |  |  |
| --- | --- | --- |
| Odds Ratio | Estimate | 95% CI |
| Treatment(experimental vs existing) | 0.413 | (0.260,0.657) |

1. Provide model-predicted values using the model in (c) for cumulative probabilities of no occurrence of a cavity by 6 months, 12 months, 18 months, and 24 months, respectively, for each treatment.

|  |  |  |
| --- | --- | --- |
| Cumulative probabilities of no occurrence of a cavity | | |
| Time | Estimated within failure | Estimated Survival |
| 6 months, Experimental treatment | =0.0013 | 1=0.992 |
| 12 months, Experimental treatment | =0.0031 | 0.992\*=0.9737 |
| 18 months, Experimental treatment | =0.01136 | 0.9737\*=0.9095 |
| 24 months, Experimental treatment | =0.0416 | 0.90995\*=0.7089 |
| 6 months, Existence treatment | =0.00311 | 1=0.9815 |
| 12 months, Existence treatment | =0.0076 | 0.9815=0.9377 |
| 18 months, Existence treatment | =0.0275 | 0.9377=0.7950 |
| 24 months ,Existence treatment | =0.101 | 0.7950=0.434 |

1. For this problem, you will be analyzing a dataset for a randomized, controlled trial among women of childbearing age to evaluate the longitudinal effects of an educational intervention. The primary response variable is the participants’ self-rating of health as either “good” or “poor”. The researchers would like to assess the effect of the intervention on self-rated health across the follow-up period.

REPEATED.sas7bdat contains data on women enrolled in this trial. These data were measured at 4 points in time: at the time of randomization, then 3, 6, and 12 months post-randomization.

Each observation in the dataset contains values for the following variables:

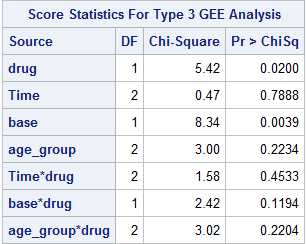
* ID: unique participant identification code
* TIME: the visit number for this observation of this participant
  + 2 corresponds to the 3 month post-randomization visit
  + 3 corresponds to the 6 month post-randomization visit
  + 4 corresponds to the 12 month post-randomization visit
* RX: the group to which the participant has been randomized
  + control
  + intervention
* HEALTH: participant’s self-rated level of health for this visit
  + Good
  + Poor
* AGE\_GROUP: participant’s age group at time of randomization
  + 15-24 (years old)
  + 25 to 34 (years old)
  + 35+ (years old)
* BASE: participant’s self-rated level of health at randomization
  + Good
  + Poor
    1. Fit a GEE repeated measures logistic regression model across all study follow-up visits (but not including time of randomization) to describe the marginal relationship of the log odds of participants’ self-rating of good health to the main effects of randomized group, visit (as a class variable), health self-rating at time of randomization, and age group as explanatory variables. Use 15-24 year old women in the control group with a good health assessment at randomization as your reference group, and use 3 months post-randomization as your reference visit. Assume an exchangeable working correlation structure. Present a table of all parameter estimates, their standard errors, the score statistics, and the corresponding p-values.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Estimate | Standard Error | ZScore Statistics | p-value |
| Intercept | .381 | .410 | .93 | .353 |
| Drug 0 | 1.949 | .466 | 4.18 | <.0001 |
| Time 3 | -.076 | .309 | 0.24 | .807 |
| Time 4 | -.223 | .320 | -.70 | .485 |
| Base Poor | -1.688 | .448 | -3.77 | .0002 |
| Age 25-34 | .963 | .440 | 2.19 | .029 |
| Age 35+ | 1.080 | .663 | 1.63 | .103 |

\*reference group drug=1,time=2,base=good, age\_group=15-24

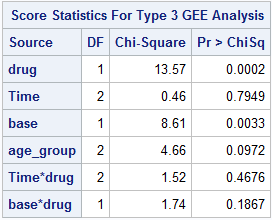
* + 1. Assess goodness of fit through consideration of all pairwise interactions with intervention (only), using manual backward selection to eliminate interactions (one-at-a-time) that are not significant at the 0.05 level. If any interactions are significant at the 0.05 level, include them and re-fit the model. Present and justify your choice for the final model.

Trial 1 Model: health =drug time base age\_group drug\*time drug\*base drug\*age\_group



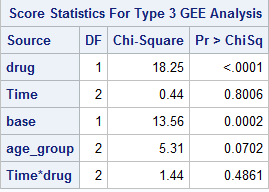
Trial 1 Conclusion: No significant interaction effects

Trial 2 Model: health =drug time base age\_group drug\*time drug\*base



Trial 2 Conclusion: No significant interaction effects

Trial 3 Model: health =drug time base age\_group drug\*time



Trial 3 Conclusion: No significant interaction effects

Final Model: health =drug time base age\_group

* + 1. Regardless of your final model in (b), use the main effects model from part (a), and:

1. provide the odds ratio and corresponding 95% confidence interval that pertain to the overall intervention effect. Interpret this result in one sentence.

Model: health =drug time base age\_group

Log OR: 1.9493 log OR 95% CI ={1.0356,2.8629}

OR=exp(log OR)=7.0235 95% CI OR={2.8167,17.5129}

The subjects on intervention had odds of good health that were 7.0235 times the odds of good health for those on control treatment. The 95% confidence interval does not include the null value so we can conclude that patients on intervention do significantly better than patients on control.

1. provide model-predicted probabilities of self-rated good health -- one for each follow-up visit -- for an individual who is 25-34 years of age, has poor health at randomization, and is randomized to the intervention arm.

|  |  |  |
| --- | --- | --- |
| Model | Equation | Value |
| Follow Up 2, Intervention, Age 25-34, Poor health baseline |  | ==0.833 |
| Follow Up 3, Intervention, Age 25-34, Poor health baseline |  | ==0.822 |
| Follow Up 4, Intervention, Age 25-34, Poor health baseline |  | ==0.799 |

1. For this question, consider the data from Question 1 of Problem Set 4 (restated below).

A social scientist is interested in how region and education level are associated with people’s feelings towards physical activity. She conducts a survey on randomly selected individuals, asking them their geographic region (West Coast, Midwest, and East Coast) and education level (college graduate, high school graduate, less than high school). Then she asked how strongly they agree with the following statement: “U.S. employers should offer greater incentives to have physically active employees.” Participants could choose one of these options: “Disagree”, “Neutral”, or “Agree”.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Education Level | Region | Level of agreement with statement regarding physical activity | | |
| Disagree | Neutral | Agree |
| College Graduate | West Coast | 18 | 15 | 48 |
|  | Midwest | 13 | 19 | 21 |
|  | East Coast | 28 | 28 | 52 |
| High School Graduate | West Coast | 46 | 23 | 24 |
|  | Midwest | 22 | 20 | 21 |
|  | East Coast | 48 | 18 | 23 |
| Less than high school | West Coast | 13 | 15 | 28 |
|  | Midwest | 15 | 16 | 17 |
|  | East Coast | 15 | 17 | 24 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Level of agreement with statement regarding physical activity (pooled across regions) | | |
| Education Level | Disagree ( ordinal 1) | Neutral (ordinal 2) | Agree (ordinal 3) |
| Less than high school  (ordinal 1) | 43 | 48 | 69 |
| High School (ordinal 2) | 116 | 61 | 68 |
| College (ordinal 3) | 59 | 62 | 121 |

* 1. Under minimal assumptions, and pooling across regions, conduct a statistical test to assess the association of education level (treated as a nominal variable) with the level of agreement (also treated as a nominal variable). Justify your method.

Since we are looking for general association and with no scale (nominal column and row variables) it implies we want to use the test.

=38.6521

=4

p-value=<0.0001

* 1. Under minimal assumptions, and pooling across regions, conduct a statistical test to assess the association of education level (treated as a nominal variable) with the level of agreement (treated as an ordinal variable), in terms of a location shift. Justify your method.

Since we are looking for the location shift with ordinal column variables it implies we want to use the test.

=36.8894

=2

p-value=<0.0001

* 1. Under minimal assumptions, and pooling across regions, conduct a statistical test to assess whether increasing level of education (treated as an ordinal variable) provides a progressive location shift in the level of agreement (also treated as an ordinal variable). Justify your method.

Since we are looking for the progressive location shift with row and column variable ordinal it implies we want to use the test.

=3.3328

=1

p-value=0.0679

* 1. For each of the tests conducted in (a), (b), and (c), compare and contrast your results in terms of the nature of the association between education level and the level of agreement. Please provide this summary in a short paragraph of 3-5 sentences.

For pooling across regions, we can conclude that when education level and level of agreement are nominal variables they are statistically associated. When education is treated as a nominal variable and level of agreement is treated as a ordinal variable, we can conclude there is a statistical location shift. When level of education and level of agreement are both ordinal, we can conclude that there is not a progressive location shift.