|  |  |
| --- | --- |
| Description | Use lifestyle.csv and lifestyle\_variables.xlsx (this contains variable descriptions) to answer questions 1-17. This is a dataset from a questionnaire collected in 1998 by Prof. Putnam at Harvard. It contains responses on internet usage, political sensitive topics, and demographics (http://bowlingalone.com/). Assume that you are a marketing manager of your favorite political party and campaign management asks you to investigate the relationship between their opinion patterns and life style variables.  Use experiment\_price\_ad.csv to answer questions 18 – 20. It contains the results of a test market study for a new instant breakfast product. The goal of this test was to determine an optimal introductory marketing plan for use when the product is introduced on a national scale and to obtain improved sales prediction. This dataset is an experimental outcome in which the values of the independent variables were chosen by the analyst. The first column gives the recorded sales from the experiment. There were 3 levels of prices: 24, 29, 34 cents per oz. and 2 levels of advertising: High and Low (coded as 1/0). In addition, the data includes the size of the stores where these experiments were conducted. |
| Instructions | 1. While you can study and work together with your teammates, you are expected to submit your own R code/output (or worksheet if the assignment involves excel analysis). Please treat this like an exam as your assignments are graded even though you’ll have week to complete summary each.  2. The R code/output or Excel worksheet will not be graded. However, they will be examined for uniqueness (i.e. not copied from your teammates), and availability of content indicating you worked on the problems. So, please do not copy the R code or Excel worksheet used for classwork/practice problems or from your teammates. You will get 10 points for submitting your individual R code and Excel worksheet.  3. Team-level interaction is encouraged so that you can verify your answers or your approach to solving the problem. But please do your assignments individually so you are adequately prepared for the exam. It is OK to disagree with your teammates and submit different answers.  4. Numerical answers are available as closed intervals such as [2,3] indicating the answer say x lies within the bounds i.e. 2 <= x <=3. |
| Multiple Attempts | Not allowed. This test can only be taken once. |
| Force Completion | This test can be saved and resumed later. |

### Expand Question Completion Status:



### QUESTION 1

1. For the test of independence between ‘buyamer’ (“Americans should always try to buy American products”) and ‘emplmerg’ (“Employment Status of Respondent”), what is the observed count of “Moderately Agree” and “Retired”, and the expected count of “Moderately Disagree” and “Not Employed”?

|  |  |  |  |
| --- | --- | --- | --- |
|  | A. | 113 and 58 |  |
|  | B. | 58 and 113 |  |
|  | C. | 171 and 58 |  |
|  | D. | None of the above |  |

**2 points**

### QUESTION 2

1. What is the null hypothesis for the test of independence between ‘buyamer’ and ‘emplmerg’?

|  |  |  |
| --- | --- | --- |
|  | A. | The level of agreement with the statement “Americans should always try to buy American products” is dependent on the “Employment status of the respondent”. |
|  | B. | The level of agreement with the statement “Americans should always try to buy American products” is independent of the “Employment status of the respondent”. |
|  | C. | The level of agreement with the statement “Americans should always try to buy American products” is related to the “Employment status of the respondent”. |
|  | D. | None of the above |

**2 points**

### QUESTION 3

1. What is the chi-squared test statistic for the test of independence between ‘buyamer’ and ‘emplmerg’?

|  |  |  |
| --- | --- | --- |
|  | A. | [33, 34] |
|  | B. | [32, 33] |
|  | C. | [35, 36] |
|  | D. | None of the above |

**2 points**

### QUESTION 4

1. What can you say about the p-value of the test of independence between ‘buyamer’ and ‘emplmerg’?

|  |  |  |
| --- | --- | --- |
|  | A. | Indicates no significance at the 90% confidence level |
|  | B. | Indicates significance at the 95% confidence level |
|  | C. | The p-value indicates non-significance |
|  | D. | None of the above |

**2 points**

### QUESTION 5

1. What is your conclusion regarding the test of independence between ‘buyamer’ and ‘emplmerg’?

|  |  |  |
| --- | --- | --- |
|  | A. | The level of agreement with the statement “Americans should always try to buy American products” is not independent of the “Employment status of the respondent” |
|  | B. | The level of agreement with the statement “Americans should always try to buy American products” is independent of the “Employment status of the respondent” |
|  | C. | The level of agreement with the statement “Americans should always try to buy American products” is not dependent of the “Employment status of the respondent” |
|  | D. | None of the above |

**2 points**

### QUESTION 6

1. What is the correlation between “I’m in favor of the death penalty” and “respondent’s age”?

|  |  |  |
| --- | --- | --- |
|  | A. | [-0.9, -0.8] |
|  | B. | [0, +0.1] |
|  | C. | [-0.1, 0] |
|  | D. | None of the above |

**2 points**

### QUESTION 7

1. What is the null hypothesis for the test of correlation between and “I’m in favor of the death penalty” and “respondent’s age” and what is your conclusion?

|  |  |  |
| --- | --- | --- |
|  | A. | Hypothesis: Correlation is zero;Conclusion: There is a significant positive albeit small correlation. |
|  | B. | Hypothesis: There is no correlation;Conclusion: There is a significant negative albeit small correlation. |
|  | C. | Hypothesis: There is no correlation;Conclusion: There is a high degree of negative correlation. |
|  | D. | None of the above |

**2 points**

### QUESTION 8

1. In a simple bivariate regression of netdome (“I don't have a clue what the internet is and what it can do for me”) on age (i.e. respondent age as independent variable), what is R-squared and what is its purpose?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.115, 0.125]; It explains the percentage variation in X accounted for by Y |
|  | B. | [0.126, 0.136]; It explains percentage variation in Y accounted for by X |
|  | C. | [0.115, 0.136]; It explains percentage variation in Y accounted for by X |
|  | D. | None of the above |

**2 points**

### QUESTION 9

1. What can you say about the overall significance of bivariate regression model between ‘netdome’ and ‘age’?

|  |  |  |
| --- | --- | --- |
|  | A. | The model is significant overall based on the significance of the beta coefficient |
|  | B. | The model is significant based on the p-value of the F-statistic |
|  | C. | The model is not significant based on the p-value of the F-statistic |
|  | D. | None of the above |

**2 points**

### QUESTION 10

1. In the bivariate regression of ‘netdome’ on ‘age’, what is the unstandardized coefficient of the independent variable and is it significant?

|  |  |  |
| --- | --- | --- |
|  | A. | [0.038, 0.045] and significant |
|  | B. | [0.038, 0.045] and not significant |
|  | C. | [0.38, 0.45] and highly significant |
|  | D. | None of the above |

**2 points**

### QUESTION 11

1. What is estimated prediction equation for the bivariate regression of ‘netdome’(use Y) on ‘age’ (use X)?

|  |  |  |
| --- | --- | --- |
|  | A. | Y = 1.184 + 0.42 X |
|  | B. | Y = 1.184 + 0.042 X |
|  | C. | Y = 0.042 + 1.184 X |
|  | D. | None of the above |

**2 points**

### QUESTION 12

1. What will be predicted value of ‘netdome’ if the age is known as 70 years using the bivariate regression model?

|  |  |  |
| --- | --- | --- |
|  | A. | [2, 3] |
|  | B. | [3, 4] |
|  | C. | [4, 5] |
|  | D. | None of the above |

**2 points**

### QUESTION 13

1. If you were to build bivariate regression models to predict ‘netdome’ using first the independent variable ‘age’ and then the independent variable ‘sex’, which of the two models would explain the variation of dependent variable (netdome) better? (hint: see the R-square value)

|  |  |  |
| --- | --- | --- |
|  | A. | Sex |
|  | B. | Age |
|  | C. | Their predictive performance will be equivalent |
|  | D. | None of the above |

**2 points**

### QUESTION 14

1. If you run a multiple regression of ‘liberal’ (“Generally speaking, would you consider yourself to be …”) on ‘age’ and ‘politics’ (“I’m interested in politics”, how would you interpret the coefficient of age based on your estimates?

|  |  |  |
| --- | --- | --- |
|  | A. | As people get older, they tend to be less liberal |
|  | B. | As people get older, they tend to be more liberal |
|  | C. | There is no relationship between age on how liberal people are |
|  | D. | None of the above |

**2 points   reg**

### QUESTION 15

1. What is the coefficient of politics in a multiple regression of ‘liberal’ on ‘age’ and ‘politics’?

|  |  |  |
| --- | --- | --- |
|  | A. | [-0.18, -0.16] |
|  | B. | [-0.19, -0.18.5] |
|  | C. | [0.16, 0.18] |
|  | D. | None of the above |

**2 points**

### QUESTION 16

1. What is the estimated multiple regression equation (best approximate) for predicting ‘liberal’(Y) using ‘age’(X1) and ‘politics’(X2)?

|  |  |  |
| --- | --- | --- |
|  | A. | Y = 1.581 + 0.044X2 – 0.17X1 |
|  | B. | Y = 1.581 + 0.044X1 – 0.17X2 |
|  | C. | Y = 0.44X1 – 0.17X2 |
|  | D. | None of the above |

**2 points**

### QUESTION 17

1. What is the predicted value of ‘netdome’ using a multiple regression model when the value age is set as 50 and the value for politics is set as 1?

|  |  |  |
| --- | --- | --- |
|  | A. | [3.5, 4.5] |
|  | B. | [3.1, 3.4] |
|  | C. | [4.1, 4.4] |
|  | D. | None of the above |

**2 points**

### QUESTION 18

1. What can you say about the impact of ad levels on sales?

|  |  |  |
| --- | --- | --- |
|  | A. | Ad levels significantly impact sales at the 95% confidence level |
|  | B. | Ad levels significantly impact sales at the 90% confidence level |
|  | C. | There is no evidence to indicate that Ad levels impact sales |
|  | D. | None of the above |

**2 points**

### QUESTION 19

1. Using a multiple linear regression model to predict sales with both ad level and price, what can you say about the impact of independent variables?

|  |  |  |
| --- | --- | --- |
|  | A. | Both ad levels and price have significant impact on sales |
|  | B. | Both ad levels and price do not impact sales |
|  | C. | Ad level has a significant impact whereas price has no significant impact |
|  | D. | None of the above |

**2 points**

### QUESTION 20

1. Using a multiple linear regression model to predict sales with ad level, price, and store size, what can you say about the impact of ad level on sales?

|  |  |  |
| --- | --- | --- |
|  | A. | Ad level has a significant impact on sales after controlling for store size |
|  | B. | Ad level has no significant impact on sales |
|  | C. | Impact of ad level on sales depends on price |
|  | D. | None of the above |

**2 points**

### QUESTION 21

1. Please upload your R code and output here (in pdf format)
   1. Attach File

Attach Local File

**10 points**

*Click Save and Submit to save and submit. Click Save All Answers to save all answers.*