**Problem Set #4**

**Geog 2000: Introduction to Geographic Statistics**

**Instructor: Dr. Paul C. Sutton**

**#1) Simple Linear Regression (Ordinary Least Squares) by Hand**

|  |
| --- |
|  |
| **#2) Chi-Square (χ2) with Nominal and Spatial Data**  **For each problem:**   1. **State a null and an alternative hypothesis;** 2. **Define and calculate the value of a test statistic;** 3. **Find the 95% Significance value for that test statistic (aka p = .05);** 4. **Fail to reject or reject the Null Hypothesis.** 5. **Provide a decision for the problem.** |
| 1. A somewhat nutty person clams that blue-eyed people have a greater tendency to become avid Sailors (e.g. they own a sailboat). [They claim the Sea is reflected in their eyes]. Suppose you test this person’s ‘theory’ by randomly sampling 100 people.   Remember to:   1. Remember to State your null and alternative hypotheses 2. Define and calculate your test statistic 3. Look up an appropriate value in a table to evaluate your hypothesis with 95% confidence 4. Do you fail to reject or reject the Null Hypothesis 5. State your conclusion in no uncertain Terms |
|  |
| 1. Earthquakes in California: Are they randomly distributed?   What you need to determine is if earthquakes are randomly distributed throughout the space of California?  Remember to:   1. State your null and alternative hypotheses. 2. Define and calculate a test statistic for this data. 3. Look up the 95% critical value for your test statistic. 4. Accept or Fail to Accept your Null Hypothesis 5. Conclude something about the spatial distribution of earthquakes |
|  |
| 1. **Winners of the Lottery: Are they randomly distributed?** What you need to determine is if lottery winners are randomly distributed throughout the population of this metropolitan area. Test this hypothesis statistically.   Remember to:   1. State your null and alternative hypotheses. 2. Define and calculate a test statistic for this data. 3. Look up the 95% critical value for your test statistic. 4. Accept or Fail to Accept your Null Hypothesis 5. Conclude something about the spatial distribution of lottery winners |
|  |
| **#3) Match the scatterplot to a diagnostic and draw a residual diagram.** |
|  |
| **4) In Terms of Experimental Design….**  A) Explain the ideas of Replication, Local Control, and Randomization as they pertain to Experimental Design. |
|  |
| B) Describe a rough design for the following experiments or research questions: |
| 1. The effect of consuming caffeine on short term memory |
| 1. Pain reliever effectiveness: What’s a better headache reliever – Advil or Excedrin |
| 1. Gender and Income: Who makes more money – Men or Women? |
| 1. SAT scores as a predictor of success in college. |
| 1. For each of the above four experiments or research designs describe issues will potentially confound and complicate your conclusions? |
| 1) |
| 2) |
| 3) |
| 4) |
| **#5) How many statistical methods are there? More than we’ll ever know**  We only touched briefly upon several other statistical methods. Google the following methods and provide a 3-5 sentence description of these approaches as to what they are used for and how they work |
| 1) Logistic Regression |
| 2) Cluster Analysis |
| 3) Factor Analysis |
| 4) Principle Components Analysis |
| 5) Multivariate Regression |
| **#6) Straight Up Linear Regression Problem (Use JMP))Do a linear regression on the ‘faked’ data below and tell me he value of the intercept and sleop for the best Ordinary Least Squares fit line through these data points.**   1. Show how you would calculate the slope, intercept and R2 2. Calculate the predicted CSAP score of schools with 0%, 25%, 5-%, 75% and 100% of their students on free lunch programs 3. Calculate the 95% Confidence Interval for the estimated or predicted CSAP score of a school with 50% of its students on the free lunch program 4. Provide a one or two sentence interpretation of the results and comment on the potential ‘theoretical’ causality involved |
| The ‘a’ intercept: |
| The ‘b’ slope: |
| R2 |
|  |
|  |
| Interpretation: |
| **# 7) The Governor’s Interpretation….**  When the former republican Governor of Colorado was presented with the relationship between CSAP scores and Free Lunch program participation that you characterized in the previous question (#6) he apocryphally said: *“Oh, I get it. If we want to raise average CSAP scores we need to cancel the Free Lunch Program. Gee that’s great we actually spend less money to raise test scores.”*  **Comment on this statement with respect to your regression results. Is he right?** |
|  |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Metrics** | **Set I** | **Set II** | **Set III** | **Set IV** | | Mean X |  |  |  |  | | Mean Y |  |  |  |  | | Regression Line |  |  |  |  | | Standard error of slope |  |  |  |  | | t-value of slope |  |  |  |  | | Sum of squares |  |  |  |  | | Residual Sum of Squares |  |  |  |  | | Correlation coefficient |  |  |  |  | | R2 |  |  |  |  |   **Scatter Plots:** |
| **Comments:** |
| **#9) Playing around with JMP’s Visualization Capability**  Paste cool 3-d pictures here: |
| How can a three-dimensional visualization like this be akin to a remote sensing task of image classification? |
| Will pixels near each other in this visualization space necessarily be near each other in real space on the ground? Explain. |
| **#10) It’s Book Report Time…** |