Project 2 Linear Regression – Instructions

MAT 1272 and 1372 – Statistics

**General Requirements**

1. Choose a research question:
   1. that can be addressed using linear regression
   2. for which you can collect data to analyze
2. Devise a plan to collect your data
3. Submit a project plan form to the instructor and obtain approval
4. Once your research question and data collection plan are approved, carry out your research:
   1. Collect data
   2. Conduct linear regression analysis, using guidelines below
   3. Write your results in a report, using the outline given below
5. Turn in your printed report (do not e-mail).
6. For ONE of your projects, you are required to give a class presentation of your results. If you choose to present this project, refer to the presentation instructions. The presentation is a separate grade from the project grade.

**Work with a team (3 to 4 members per team).** Your sample size must include at least 25 individuals ***per team member*** (teams of 3 must have n=75, etc.)

**Data Collection Options**

1. Reliable/reputable websites (e.g., sponsored by the census bureau, professional sports leagues, universities, real estate agencies, car manufacturers, consumer groups, financial institutions, well-known product manufacturers, restaurants/fast food companies, weather tracking agencies, county/city/state/federal government organizations, etc.)
2. Visit to one or more locations where item(s) you are researching can be found (e.g., stores to write down prices, rivers to count turtles, car dealers to write down data about cars, etc.)
3. Survey that you create and administer to people in your population of interest; this may include
   1. questions that you wrote yourself
   2. a quiz/survey you found online or in a book, magazine, etc. (you should be able to compute a numeric score from each participant’s answers by using scoring instructions given with the survey; however, these scoring instructions should NOT be placed on the survey that is given to your participants.) You may use a survey from the collection of surveys listed on the openlab.
4. Other resource by permission (if you have an idea, ask your instructor).

**Project Design**

To see the required components and examples for this type of project design, consult the Project Examples document on the openlab.

**Project Plan Form**

Download a copy of the project plan form on the openlab.

Complete this form and submit to your instructor. If the form is not completed satisfactorily, it will be returned to you for revision.

***Notes 1) Your project plan must be approved by the instructor before you may begin your project! 2) Attach your approved and signed project plan form to the front of your final project report.***

**Project Report**

Your final report should be *written in paragraph form* and should include the following sections. The sections are numbered here for your convenience, but they do not need to be numbered in your report.

1. **Introduction.** State the topic of your study as a research question AND as a specific hypothesis that you tested. Your hypothesis should indicate what type of correlation you expected to see (positive or negative) and how strong you expected the correlation to be (weak, moderate, or strong). Even if you expect a weak correlation, your hypothesis should describe a statistically significant result that you expected to find AND the ***practical reason*** that you expected this result (your rationale).

2. **Define Population(s).** Define clearly the population(s) that you intend for your study to represent. (Examples: all NFL football players, all cars manufactured this year, all students at City Tech, all small towns in Westchester County; all shoppers at the Target, all single family houses for sale Brooklyn, etc.)

3. **Define Variable(s).** Define clearly the variable(s) that you will obtain during your data collection (e.g., gender, age, salary, price, miles per gallon, score on a particular personality test, miles commuted one-way to school daily, major, etc.) This must be specific: “time spent watching TV” is too vague; “number of hours spent watching TV in the last 3 days” would be specific enough. If your variable is a measurement (e.g., height) give units (e.g., inches). If your variable is a score (e.g., on a personality quiz), give the range of possible scores (e.g., 0 to 15).

4. **Data Collection.** Describe your data collection process and sampling strategy. If you located data on a website, provide the URL and tell how you selected individuals from that website to include in your sample. If you obtained data from an agency, office, store, or other similar source, explain where you went and how you selected individuals to include in your sample. If you surveyed individuals directly or took measurements, describe how you selected individuals for the sample. If you used a survey, this section must include a copy of your survey. If you collected measurements, describe the device you used (e.g., tape measure, odometer, scale, stopwatch, etc.) If participants were required to do something to be measured, describe the measurement process (e.g., “each participant was asked to hold their breath as long as possible; the number of seconds they held their breath was timed with a stopwatch.”) *No matter what data collection process you used, address: a) what steps you took to avoid bias in your sample; and b) whether you believe the sample(s) you obtained were representative of the population. Tell why or why not.* **Include a table with ALL of your raw (not summarized) data as an appendix at the end of the report**.

5. **Study Design.** Identify the statistical analysis you conducted to analyze your data. Give relevant design details (e.g., which variable was selected as the explanatory (input) variable, and which the response variable? Why? What type of correlation did you expect? And so on.) State your null and alternative hypotheses about the correlation, both in words and in appropriate mathematical symbols.

6. **Results: Descriptive Statistics.** Give descriptive statistics ***for each of your two quantitative variables***. Note that you will be reporting summary statistics for both your explanatory variable and your response variable. Report each set of descriptive statistics using both a table and a chart as described below. All tables and charts should be placed directly in your report.

1. Table: Give sample size, mean, standard deviation, and 5-number summary.
2. Chart: Show a histogram that illustrates the distribution of the variable.

7. **Results: Statistical Analysis.** Report the results of your analysis; include all items below.

1. Scatterplot with a graph of the regression line
2. Value of the correlation coefficient *r* and interpretation of its meaning
3. Equation of regression line
4. An example of a prediction using the regression equation
5. Discussion of the slope of the regression line and its meaning
6. Value of R2 for the regression model and interpretation of its meaning
7. Indication of significance of correlation, with discussion of significance level

8. **Findings.** Interpret the results of your statistical analysis in the context of your original research question. Do your analyses support your expected findings? Explain.

9. **Discussion.** What conclusions, if any, do you believe you can draw as a result of your study? If the results were not what you expected, what factors might explain your results? What did you learn from the project about the population(s) you studied? What did you learn about the research variables? What did you learn about the specific statistical analysis you conducted?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **0 – 3 Pts** | **4 Pts** | **5 Pts** | **Possible** | **Earned** |
| **Project Plan** | Incomplete or missing project plan or approval not obtained. | Approval obtained on project plan, but approval sheet missing from submitted report. | Approved project plan form, signed by instructor, is attached to project report. | 5 |  |
| **Report Format** | Project report is not submitted as a formal paper OR  Final report is written in Excel or other program without proper formatting | Project report is submitted as a formal paper, but with minor issues (e.g., poor or inconsistent formatting, not typed) | Project report is submitted as a formal paper in paragraph form with full sentences, and is typed and well formatted. | 5 |  |
| **Writing and Readability** | Report is poorly organized and hard to follow; charts and tables not embedded in report; many writing errors, awkward sentences | Report is reasonably organized and readable with few writing errors; all charts and tables are embedded in report | Report is exceptionally well organized and well written, with all charts and tables embedded in report | 5 |  |
| **Originality and Initiative** | Topic selected is not original; it has been studied frequently by other students. | An original research topic was selected for this project. | An original research topic was selected AND report demonstrates initiative in carrying the project out. | 5 |  |
| **Overview of Research** | Research question and expected findings not stated clearly.  Rationale not explained. | Two of the three criteria at right are satisfied. | Stated clearly:  1) Research question  2) Expected findings  3) Rationale | 5 |  |
| **Research Population Specified** | Target population(s) of research project not defined, poorly defined, or incorrect | Reasonable target population(s) defined, but not fully appropriate | Target population(s) well defined and fully appropriate | 5 |  |
| **Definition of Variables** | Research variables not adequately defined. | Research variables named, but details are lacking on how variables are measured or quantified | Research variables named; measurement and possible values of each variable are clearly defined | 5 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **0 – 3 Pts** | **4 Pts** | **5 Pts** | **Possible** | **Earned** |
| **Data Collection: Sampling** | Ineffective sampling procedures OR poor description of sampling strategy | Two of the three criteria at right are satisfied. | 1) Sampling strategy explained fully  2) Sound measures taken to avoid bias explained  3) Representative sample addressed adequately | 5 |  |
| **Descriptive Statistics: Research Variables** | Descriptive statistics omitted or not given clearly for **each** research variable | A summary is given for **each** variable, but at least one summary is incomplete. | **Each** research variable is described fully, using mean, standard deviation, and 5-number summary. | 5 |  |
| **Data Representation: Charts/Graphs** | Report does not use sufficient charts or graphs to display data. | Most charts and tables are present, but some are missing OR  All charts are present but contain minor flaws. | **Each** research variable is represented with appropriate charts and tables. | 5 |  |
| **Data Representation: Raw Data** | Raw data used in the study are not provided. | Raw data are given, but presentation is incomplete or disorganized. | All raw data are included in table(s) in an appendix | 5 |  |
| **Statistical Analysis: Scatter Plot** | Scatter plot missing, incorrect, not adequately labeled, or poorly illustrated | Three of the four criteria at right are satisfied | 1) Scatter plot present and correct  2) Plot is well illustrated, scaled, and readable  3) Axes are well labeled  4) Shape of scatter plot is discussed | 10 |  |
| **Statistical Analysis: Correlation** | Correlation coefficient r not reported or incorrect  Correlation not explained adequately and significance not reported | Three of the four criteria at right are satisfied | Correctly reported:  1) Correlation coefficient r  2) Meaning of the correlation  3) Significance of correlation  4) Significance level | 5 |  |
| **Statistical Analysis: Significance** | Significance of correlation is not addressed correctly | Significance of correlation and significance level are both reported correctly | Criteria at left are satisfied AND null and alternative hypotheses for significance test are stated correctly in mathematical terms | 5 |  |
| **Statistical Analysis: Regression Line** | Regression line missing, incorrect, or poorly illustrated  Equation of regression line not reported  Regression equation not explained adequately | Regression line is present and correct; regression equation is reported and is correct | Regression line and equation are both correct and a prediction example is used to demonstrate the regression equation | 5 |  |
| **Statistical Analysis: Slope** | Slope of regression line is not reported correctly  No correct interpretation is given of slope | Slope is reported correctly and interpretation is partially correct | Slope is reported and interpreted accurately | 5 |  |
| **Statistical Analysis: *R2*** | *R2* is not reported correctly or no correct interpretation of *R2* | *R2* is reported correctly and interpretation is partially correct | *R2* is reported and interpreted accurately | 5 |  |
| **Conclusion and Discussion** | Implications of results not discussed; no attempt to explain findings. Report draws unwarranted conclusions or uses inappropriately certain language (e.g., “we proved”, “our hypothesis is true”) | Adequate discussion of results and their practical implications; reasonable explanation of findings offered; no unwarranted conclusions. | Discussion of results is insightful; adds meaning and significance to the report; no unwarranted conclusions. | 5 |  |