Name_____

MTH1W Statistics Project- Linear Regression

Note: Linear Regression is the process of finding a line that best fits the data points available on the plot, so that we can use it to predict output values for inputs that are not present in the data set we have, with the belief that those outputs would fall on the line.

Learning Goals:

- To improve our data collection techniques, considering bias
- To apply our understanding of linear regression and statistical modelling to real data
- To be able to use a spreadsheet effectively to produce graphs and do statistical analyses
- To learn to work effectively with a partner on a project

Project Overview:

In this partner project, you will choose a topic, collect raw data on the topic, and perform various statistical analyses in Excel.

Need to do:

1. Choose a single partner (if you wish)
2. Choose a topic (various groups can do the same topic)
3. Make a hypothesis/prediction using full sentences.
4. Collect the data in a spreadsheet. Have the columns beside each other. You need between 50-100 data points, so that means about 25-50 movies, players, foods, etc….
5. Complete a summary chart for mean, median, mode, min, max, range, standard deviation, low normal, high normal, and analysing the normal range.
6. Create a Scatter Plot with a line of best fit, titles, and labels axes.
7. Calculate the correlation. What does this mean?
8. Reflect! Was your hypothesis/prediction correct? Did you find anything surprising? How might bias have influenced your data collection? Were there any attempts made by you to reduce bias? What would have made your research better? Etc….

What do I hand in? One Excel spreadsheet containing all of your raw data and analysis, clearly laid out in an easy-to-understand manner. Submit in Edsby

Due: End of Class on Wednesday, June 4

Presentation: You will need to present your findings to the class. Make a slideshow with the following slides:

1. Your name(s) and the topic
2. Your hypothesis/prediction.
3. A screenshot of your data and where you gathered it from.
4. Your summary chart.
5. Your scatter plot with line of best fit.
6. Your correlation and meaning of it.
7. Your reflection.

Presentations will begin on THURSDAY, JUNE 5.

Some Possible Topic Ideas:

🌍 World Issues

1. **Carbon Emissions per Capita vs. GDP per Capita (Countries)**
   – Do wealthier countries emit more CO_2?
2. **Literacy Rate vs. Female Employment Rate (Countries)**
   – Is literacy linked to women's economic participation?
3. **Renewable Energy Usage % vs. Air Quality Index (Countries)**
   – See if using renewables leads to better air quality.
4. **Access to Clean Water vs. Child Mortality Rate**
   – Investigate how water access affects children's health.
5. **Education Level vs. Average Salary in Canada**
   – Investigate if more education leads to higher income.
6. **A Country's GDP per capita vs. Life Expectancy**
   – See if wealthier nations have a higher life expectancy

🏀 Sports

1. **Average Player Height vs. Points Per Game in the NBA**
   – Explore if taller players tend to score more.
2. **Team Payroll vs. Number of Wins in MLB**
   – Analyze whether teams that spend more win more games.
3. **Olympic Gold Medals vs. Country GDP**
   – See if wealthier nations win more gold medals.
4. **Average Age of Soccer Teams vs. FIFA World Ranking**
   – Investigate whether age correlates with performance.
5. **NHL Goalie Save Percentage vs. Team Wins in a Season**
   – Explore if better goalies lead to more team wins.
6. **Average Yards per Game vs. Player Salary (NFL Running Backs)**
   – Do more productive players earn more?

🎤 🍿 Pop Culture/Movies

1. **Social Media Followers vs. Billboard Chart Rankings (Top Artists)**
   – Do more followers help artists reach #1?
2. **Number of Grammy Awards vs. Number of Albums Released**
   – Explore if releasing more albums leads to more awards.
3. **Movie Budget vs. Box Office Revenue**
   – Analyze whether spending more leads to higher profits.
4. **Opening Weekend Revenue vs. Total Revenue (Popular Films)**
   – Does a strong opening predict total success?
5. **Number of Awards Won vs. Metacritic Score**
   – Do more critically acclaimed movies win more awards?

Statistics Project Rubric

Presenter(s) _____.

Topic _____.

R	Level 1	Level 2	Level 3	Level 4
Thinking & Inquiry ((4+1+6) marks)				
No **bias** mentioned	Explanation of bias in data collection is limited	Explanation of bias in data collection is somewhat thorough	Explanation of bias in data collection is considerably thorough	Explanation of bias in data collection is highly thorough and accurate
No **Hypothesis/Prediction** mentioned	___/1T			
Reflection ____*was hypothesis correct?/found anything surprising?* ____*how must bias have influenced data collection?/any attempts to reduce bias?* ____*what could have made research better?*	___/6T			
Knowledge & Understanding ((5+10) marks)				
Scatterplot is missing	Scatterplot has major errors	Scatterplot is somewhat easy to read, correct, and thorough	Scatterplot is mostly easy to read, correct, and thorough	Scatterplot looks great and is easy to read with clear labelled axes (units where relevant), a descriptive and clear title, and a best fit line with its equation
Summary Chart is missing -Mean__, Median__, Mode__, min.__,	Summary Chart is mostly incomplete	Summary Chart is somewhat complete	Summary Chart is mostly complete	Summary Chart is complete

max.__, range__, standard deviation__, low normal__, high normal__, and analysis of normal range__	*(6 missing pieces)*	*(4 missing pieces)*	*(2 missing pieces)*	
Application (4 marks)				
Provides no explanation for **correlation coefficient** ____*what is the correlation coefficient* ____*positive/negative* ____*strong/weak* ____*the why behind your analysis*	Provides a weak explanation *(three pieces missing)*	Provides an okay explanation *(two pieces missing)*	Provides a good explanation *(one piece missing)*	Provides a complete and well-researched explanation
Communication (5 marks)				
N/A	Presenter speaks with limited effectiveness	Presenter speaks with some effectiveness	Presenter speaks with considerable effectiveness	**Presenter speaks clearly, confidently, and at a good pace, minimizing filler words like "um"**
Other				
Data Collection – 50 or more data points collected _____/3 PowerPoint – Text/images are well presented and work shows professionalism with appropriate spacing, graphics are attractive and support the theme/content of the presentation, all 7 slides expected are present _____/2				

Comments:

_____

_____

_____