Performance Based Learning and Assessment Task

Regression Analysis Activity

I. ASSESSSMENT TASK OVERVIEW & PURPOSE:

In this activity, students will collect and analyze data for answering one of two questions using regression techniques. Students will then use this information to make prediction and evaluate the accuracy of their predictions.

II. UNIT AUTHOR:

Eric Tatum, Galax High School, Galax City Public Schools

III. COURSE:

Algebra II

IV. CONTENT STRAND:

Data Analysis

V. OBJECTIVES:

The student will be able to:

- Gather data using physical and/or electronic measurements
- Create appropriate regression curves and lines based on an analysis of data
- Use variable relationship data to make predictions about other data points
- Evaluate the accuracy of predictions by researching and/or collecting more data
- Assess data collection techniques to improve the accuracy of the data

VI. REFERENCE/RESOURCE MATERIALS:

Graphing calculators, spreadsheet software, graph paper, measuring tape, metersticks, yardsticks, rulers, stopwatches, CBLs with timed light sensors

VII. PRIMARY ASSESSMENT STRATEGIES:

The task includes an assessment component that performs two functions: (1) for the student it will be a checklist and provide a self-assessment and (2) for the teacher it will be used as a rubric. This assesses each group's data collection, organization, and representation. This also assesses prediction accuracy and comparison. Finally, it assesses how well each group works together.

VIII. EVALUATION CRITERIA:

The assessment list for the activity will be used for evaluation.

IX. INSTRUCTIONAL TIME:

The class introductory activity is estimated to take one 45-50 minute period, and the group data collection and analysis activity is estimated to take three 45-50 minute periods.

Regression Analysis Activity

Strand

Data Analysis

Mathematical Objective(s)

The student will be able to:

- Gather data using physical and/or electronic measurements
- Create appropriate regression curves and lines based on an analysis of data
- Use variable relationship data to make predictions about other data points
- Evaluate the accuracy of predictions by researching and/or collecting more data
- Assess data collection techniques to improve the accuracy of the data

Related SOL

- AII.9 – The student will collect and analyze data, determine the equation of the curve of best fit, make predictions, and solve real-world problems, using mathematical models. Mathematical models will include polynomial, exponential, and logarithmic functions.

NCTM Standards

- Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them
- Select and use appropriate statistical methods to analyze data
- Develop and evaluate inferences and predictions that are based on data
- For bivariate measurement data, be able to display a scatterplot, describe its shape, and determine regression coefficients, regression equations, and correlation coefficients using technological tools;
- Identify trends in bivariate data and find functions that model the data or transform the data so that they can be modeled.

Materials/Resources

All of the materials below are provided to students to they will have options on how to collect their data and answer questions. Not all materials are necessarily required.

- Graphing calculators
- Spreadsheet software
- Graph paper
- Measuring tape
- Metersticks
- Yardsticks
- Rulers
- Stopwatches
- CBLs with timed light sensors

Assumption of Prior Knowledge

- Students should be comfortable working with bivariate data, using regression aspects of graphing calculators and/or spreadsheet software, physically measuring length and/or time with good levels of precision and accuracy.
- Students may struggle to decide a good way to collect data or pick an appropriate regression model for their data. The teacher should provide support as needed.
- Students will benefit from some prior experience with basic curve fitting for polynomial and exponential relationships. This activity would also serve as a good extension from modeling linear relationships to polynomial and exponential relationships.
- Relevant contexts for this activity are as follows: human biological norms and trends, US population growth, the physics associated with freefall.

Introduction: Setting Up the Mathematical Task

- In this activity, you will work to answer one of the following questions.
    - How long does it take a basketball to hit the ground when dropped from the roof of the school?
    - What is the relationship between the heights of people and their arm spans?

  You and your group will collect data pertaining to one of the questions asked above. You will decide how you will collect your data and how you will analyze it. You will then use your data to make predictions, and you will compare your predictions with all other same-question groups in the class. Finally you will evaluate your predictions when compared to the other same-question groups.

- These questions will take three to four days in class to answer. You should allot one day for data collection, one day for data analysis, and one day to make predictions and evaluate the accuracy of those predictions.
- Students will complete a teacher-led introductory activity that answers the question, "How many people will be living in the U.S. in the year 2100?
- Intentional questions for students:
    - How will you collect your data?
    - Is your data accurate and precise?
    - What affects the accuracy and precision of your data?
    - What influenced your regression model choice?
    - How accurate are your predictions?
    - What affects the accuracy of your predictions?
- Students will work individually for the introductory teacher-led activity. Students will form groups of three to four students to answer one of the two questions above.
- Students will organize their data and compare their results with other groups to make their understanding public.

Student/Teacher Actions:

- Students should work in their groups to determine a plan and timeline for how to answer their chosen question.
- The teacher should offer guidance where appropriate and help groups determine if their plan and timeline is reasonable.
- Common misconceptions or errors that students may need help with are as follows:
    - Poor (inaccurate or imprecise) data collection techniques

    - Collecting too few data points
    - Collecting data points that are not varied enough
    - Choosing poor regression models
- If available, technology can be integrated into this activity. Some possibilities are as follows:
    - Using graphing calculators or a spreadsheet program to find regression models
    - Using stopwatches to time the length of a sample drop
    - Using CBLs to time the length of a sample drop

Monitoring Student Responses

- Students will be assessed on their data collection, results, and predictions.
- Students will be assessed on their abilities to work in groups. Each group member should be included.
- Students will be expected to justify their predictions using regression analysis.
- The teacher will monitor student progress in each group and provide support where needed. The intentional questions above can help to confirm student understanding or expose student misconceptions.
- Students who are ready to move on can work to research the physical relationship posed in the questions to gain a deeper understanding of why each regression model is appropriate for them.

- Students will summarize the activity by comparing their group results to the other groups answering the same question.

- Evidence of student learning can be obtained through their displays of data collection and their justifications of their answers to one of the two questions.

Introduction

In this activity, you will work to answer one of the following questions.

- How long does it take a basketball to hit the ground when dropped from the roof of the school?

- What is the relationship between the heights of people and their arm spans?

You and your group will collect data pertaining to one of the questions asked above. You will decide how you will collect your data and how you will analyze it. You will then use your data to make predictions, and you will compare your predictions with all other same-question groups in the class. Finally you will evaluate your predictions when compared to the other same-question groups.

All of the materials below are provided to students to they will have options on how to collect their data and answer questions. Not all materials are necessarily required.

- Graphing calculators
- Spreadsheet software
- Graph paper
- Measuring tape
- Metersticks
- Yardsticks
- Rulers
- Stopwatches
- CBLs with timed light sensors

You are free to use any of the above materials (or other approved materials if you have other ideas) to collect data to answer your question.

We will first go through this process together by answering the following question.

- How many people will be living in the U.S. in the year 2100?

Population Data Sheet Name_____

Use the following table to organize your data collection. You may not need the entire space. For the U.S. population question, we will use census data from page 143 of https://www.census.gov/prod/2002pubs/pol02marv.pdf, but your group will have to gather your own data to answer your question. You should give each column a title (header).

Population Regression Analysis Name_____

Decide how you want to visually display your data and create an approximation below. You may use computer software to display your data, or you may draw it (use a straightedge or graph paper to increase your accuracy).

Now provide as much information about your data as you can.

Type of Model: Linear Quadratic Exponential Other:_____

How well does your model fit your data? _____

Explain your reasoning. _____

Regression Equation: _____

Population Prediction Sheet Name_____

Predict the time the number of people living in the USA or the year when given the values in the table below. Show your work in the space below.

US Population	Year
500,000,000	
The number of people currently living in China today.	
	50 years from now
	2100

How accurate do you think your predictions are?

_____
_____
_____

How does your data affect the accuracy of your predictions?

_____
_____
_____

Why would it be useful to be able to predict the population of the United States in the future?

_____
_____
_____

Arm Span Data Sheet Name_____

Use the following table to organize your data collection. You may not need the entire space. In the U.S. population question, the data was provided. Now your group will have to gather your own data to answer your question. You should give each column a title (header).

Arm Span Regression Analysis Name_____

Decide how you want to visually display your data and create an approximation below. You may use computer software to display your data, or you may draw it (use a straightedge or graph paper to increase your accuracy).

Now provide as much information about your data as you can.

Type of Model: Linear Quadratic Exponential Other:_____

How well does your model fit your data? _____

Explain your reasoning. _____

Regression Equation: _____

Arm Span Prediction Sheet Name_____

Predict the arm span or height of a person when given the values in the table below. Show your work in the space below.

Arm Span	Height
64 cm	
	5.75 feet
	Michael Jordan's
6' 2"	

How accurate do you think your predictions are?

_____
_____
_____

How does your data affect the accuracy of your predictions?

_____
_____
_____

Why would it be important to know how the height of a person relates to his or her arm span?

_____
_____
_____

Drop Length Data Sheet Name_____

Use the following table to organize your data collection. You may not need the entire space. For the U.S. population question, we utilized census data, but now your group will have to gather your own data to answer your question. You should give each column a title (header).

Drop Length Regression Analysis Name_____

Decide how you want to visually display your data and create an approximation below. You may use computer software to display your data, or you may draw it (use a straightedge or graph paper to increase your accuracy).

Now provide as much information about your data as you can.

Type of Model: Linear Quadratic Exponential Other:_____

How well does your model fit your data? _____

Explain your reasoning. _____

Regression Equation: _____

Drop Length Prediction Sheet Name_____

Predict the time it takes for the basketball to hit the ground or the height from which the basketball was dropped when given the values in the table below. Show your work in the space below.

Height of Drop	Time
10 feet	
	3.7 seconds
	0.1 minutes
The roof of the school	

How accurate do you think your predictions are?

_____
_____
_____

How does your data affect the accuracy of your predictions?

_____
_____
_____

Why would it be important to know how long it will take something to hit the ground when dropped from a particular height?

_____
_____
_____

Student Comparison Sheet Name_____

Compare your results with the groups who studied the same question.

Explain how each group collected their data, noting the similarities and differences for each group.

_____
_____
_____
_____
_____

How did the amount of data each group collected differ? What influence do you think this had on your comparisons? Which one do you think may have provided a more accurate equation and why?

_____
_____
_____
_____
_____

Look at the prediction table you completed, and compare your findings. Why are your predictions not exactly the same? Are one group's predictions better than another's? Why or why not?

_____
_____
_____
_____
_____

Explain how you could change what you did in this experiment to make your predictions more accurate.

_____
_____
_____
_____
_____

Student Check Sheet and Rubric Name_____

Number	Element	Point Value	Self-Assessment	Teacher Assessment
1	I collected sufficient data for my question.	2		
2	I organized the data in a useful way.	2		
3	I chose an appropriate model for my data.	2		
4	I analyzed the data in a mathematical way.	2		
5	I based my predictions on the data I collected.	2		
6	I compared and contrasted my predictions with other groups.	2		
7	I evaluated the reasonableness of my predictions.	2		
8	I included everyone in my group.	2		
9	I used my time in a productive manner.	2		
10	My work is neat, legible, and grammatically correct.	2		

#	Element	0	1	2
1	I collected sufficient data (at least ten points) for my question.	No data collected.	Insufficient data collected.	Sufficient data collected.
2	I organized the data in a useful way.	No data collected.	Data is unorganized or not usefully organized.	Data is organized in a useful way.
3	I chose an appropriate model for my data.	No model chosen.	Inappropriate model	Appropriate model
4	I analyzed the data in a mathematical way.	No data analyzed.	Data is not analyzed using regression techniques.	Data is analyzed using regression techniques.
5	I based my predictions on the data I collected.	No data collected or no predictions are made.	Predictions are not based on the data collected	Predictions are based on the data collected
6	I compared and contrasted my predictions with other groups.	No group comparisons are made.	Predictions are compared to at least one but not all other same-question groups.	Predictions are compared to all other same-question groups.
7	I evaluated the reasonableness of my predictions.	No predictions are made or no evaluations are made.	N/A	I evaluated the reasonableness of my predictions.
8	I included everyone in my group.	I did not include everyone in my group.	N/A	I included everyone in my group.
9	I used my time in a productive manner.	I did not complete the task.	I spent class time off task, doing other things.	I stayed on task and used my time wisely.
10	My work is neat, legible, and grammatically correct.	My work meets none of the three qualifications.	My work meets one or two qualifications.	My work meets all three qualifications.

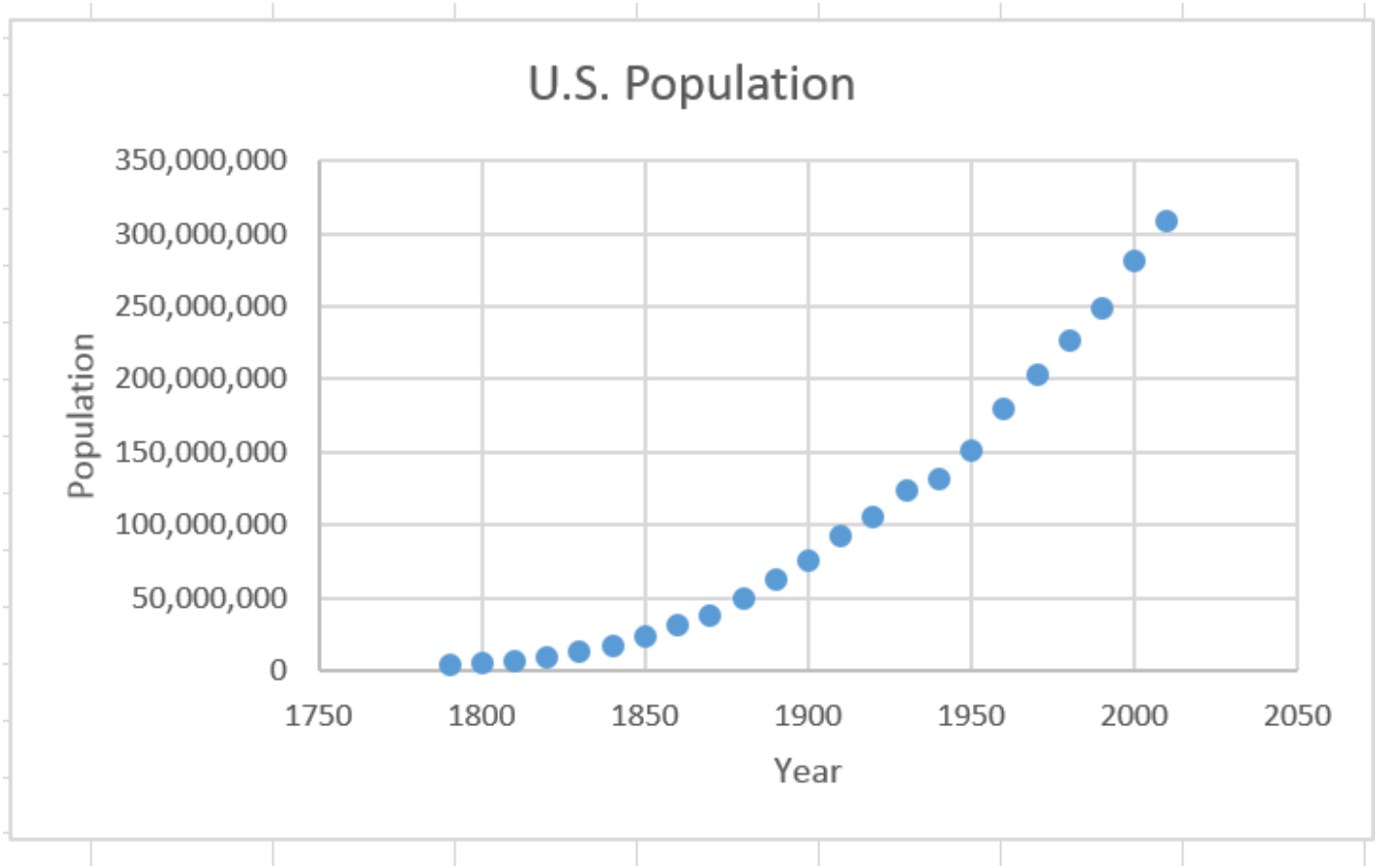
Population Data Sheet Name_____

Use the following table to organize your data collection. You may not need the entire space. For the U.S. population question, we will use census data from page 143 of https://www.census.gov/prod/2002pubs/pol02marv.pdf, but your group will have to gather your own data to answer your question. You should give each column a title (header).

Year	U.S. Population
1790	3,929,214
1800	5,236,631
1810	7,239,881
1820	9,638,453
1830	12,866,020
1840	17,069,453
1850	23,191,876
1860	31,443,321
1870	38,558,371
1880	49,371,340
1890	62,979,766
1900	76,212,168
1910	92,228,531
1920	106,021,568
1930	123,202,660
1940	132,165,129
1950	151,325,798
1960	179,323,175
1970	203,211,926
1980	226,545,805
1990	248,709,873
2000	281,421,906
2010	308,745,538

Population Regression Analysis Name_____

Decide how you want to visually display your data and create an approximation below.



Now provide as much information about your data as you can.

Type of Model: Linear (Quadratic) Exponential Other:_____

How well does your model fit your data? The model fits the data really well.

Explain your reasoning. For the model, $r^2 = 0.991$, so the model is almost perfect.

Regression Equation: $y = 6784.41326x^2 - 24420167.26x + 21980675310$ (from Excel)

Population Prediction Sheet Name_____

Predict the time the number of people living in the USA or the year when given the values in the table below. Use the space below to show your work.

US Population	Year
500,000,000	2070
617,586,541	2100
483,294,552	50 years from now (2065)
The number of people currently living in China today. (1,393,783,836)	2252

How accurate do you think your predictions are?

The data may not be completely accurate because we are extrapolating and many other factors may affect the U.S. population before those dates occur. The data would be more accurate if it were finding populations between 1790 and 2010.

How does your data affect the accuracy of your predictions?

Our predictions are based on our data, so the more accurate our data is, the more accurate our predictions will be.

Why would it be useful to be able to predict the population of the United States in the future?

Urban development, food production, and resource availability are all important reasons for predicting the future U.S. population.

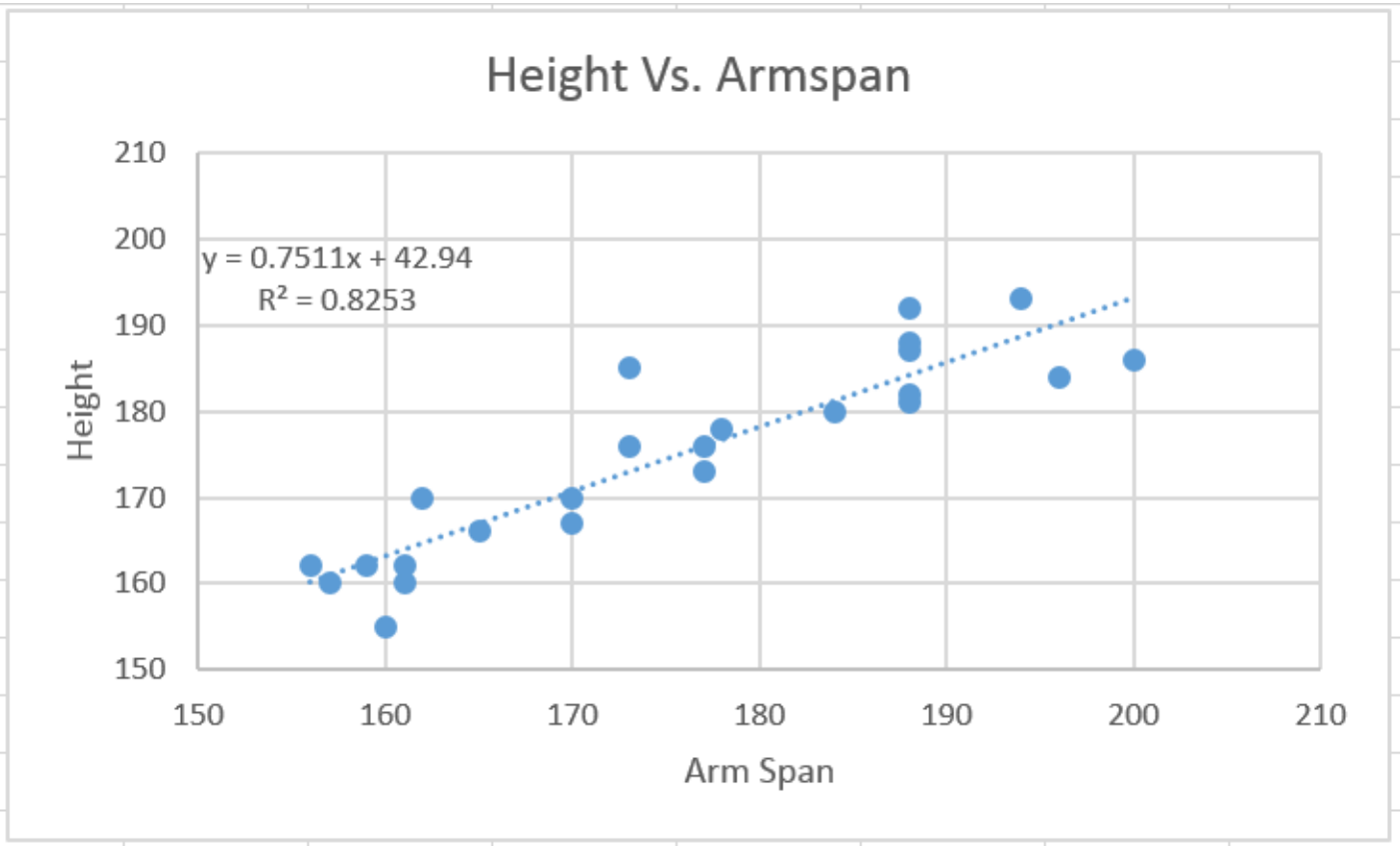
Arm Span Data Sheet Name_____

Use the following table to organize your data collection. You may not need the entire space. For the U.S. population question, we will use census data from page 143 of https://www.census.gov/prod/2002pubs/pol02marv.pdf, but your group will have to gather your own data to answer your question. You should give each column a title (header).

Arm Span (cm)	Height (cm)
156	162
157	160
159	162
160	155
161	160
161	162
162	170
165	166
170	170
170	167
173	185
173	176
177	173
177	176
178	178
184	180
188	188
188	187
188	182
188	181
188	192
194	193
196	184
200	186

Arm Span Regression Analysis Name_____

Decide how you want to visually display your data and create an approximation below.



Now provide as much information about your data as you can.

Type of Model: (Linear) Quadratic Exponential Other:_____

How well does your model fit your data? The model fits the data moderately well.

Explain your reasoning. For the model, $r^2 = 0.8253$. Other models had a slightly higher value for r^2, but the increase in the complexity of the equations is not really worth the negligible increase in accuracy.

Regression Equation: $height = 0.7511(armspan) + 42.94$

Arm Span Prediction Sheet Name_____

Predict the arm span or height of a person when given the values in the table below. Use the space below to show your work.

Arm Span	Height
64 cm	91 cm
176 cm	5.75 feet (175.26 cm)
207 cm	Michael Jordan's (198 cm)
6' 2" (188 cm)	184 cm

How accurate do you think your predictions are?

My data contains arm spans between 156 cm and 200 cm. The predictions that fall between these two numbers are probably more accurate than other predictions because I am interpolating instead of extrapolating.

How does your data affect the accuracy of your predictions?

My data points that are interpolated are probably more accurate than the other data predictions. I could get more accurate data by sampling a wider range of arm spans and heights.

Why would it be important to know how the height of a person relates to his or her arm span?

Athletes could be considered more desirable for a team if they have a long arm span for their height. As well, heights that are very small or large for a given arm span could be predictors of physical problems for people.

Sample Benchmark for Drop Length Question

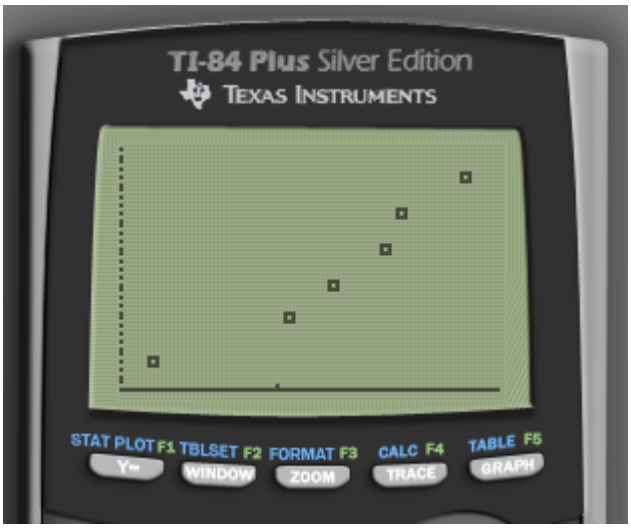
Drop Length Data Sheet Name_____

Use the following table to organize your data collection. You may not need the entire space. For the U.S. population question, we will use census data from page 143 of https://www.census.gov/prod/2002pubs/pol02marv.pdf, but your group will have to gather your own data to answer your question. You should give each column a title (header).

Height of drop.	Time it takes to hit the ground.
4 ft	0.755 sec
10 ft	1.026 sec
15 ft	1.113 sec
20 ft	1.217 sec
25 ft	1.25 sec
30 ft	1.375 sec

Drop Length Regression Analysis Name_____

Decide how you want to visually display your data and create an approximation below.



Now provide as much information about your data as you can.

Type of Model: Linear (Quadratic) Exponential Other:_____

How well does your model fit your data? The model fits the data moderately well.

Explain your reasoning. The value of r^2 was closest to 1 for a quadratic model. Also, free fall motion has a quadratic relationship.

Regression Equation: $height = 47.56(time)^2 - 57.25(time) + 19.81$

Drop Length Prediction Sheet Name_____

Predict the time it takes for the basketball to hit the ground or the height from which the basketball was dropped when given the values in the table below. Use the space below to show your work.

Height of Drop	Time
10 feet	1.026 seconds
459 feet	3.7 seconds
1389 feet	0.1 minutes
The roof of Galax High School	1.88 seconds

How accurate do you think your predictions are?

I think the first and last predictions are the most accurate because they are closest to the heights of our sample data. The second and third predictions are probably less accurate because we had to extrapolate instead of interpolate.

How does your data affect the accuracy of your predictions?

Our predictions are based on our data, so the more accurate our data is, the more accurate our predictions will be.

Why would it be important to know how long it will take something to hit the ground when dropped from a particular height?

Skydivers would want to know how long they have until they reach the ground so they can open their parachutes in time.

Student Comparison Sheet Name_____

Compare your results with the groups who studied the same question.

Explain how each group collected their data, noting the similarities and differences for each group.

Our group collected drop length data using a stopwatch, while the other group used a CBL with a light sensor. We used a stairwell to measure greater heights, while the other group made sample drops from as high as their tallest group member could reach. We measured the heights of the drops in feet, while the other group measured the heights in centimeters. We both dropped a basketball every time.

How did the amount of data each group collected differ? What influence do you think this had on your comparisons? Which one do you think may have provided a more accurate equation and why?

Our group had 6 sample drops that ranged in height from 4 feet to 30 feet. The other group had 15 sample drops that ranged in height from 5 centimeters to 213 centimeters. Our group had fewer, but more diverse data points. The other group had more, but less diverse data points. I think our group had more accurate predictions because our sample drops covered a larger range of heights.

Look at the prediction table you completed, and compare your findings. Why are your predictions not exactly the same? Are one group's predictions better than another's? Why or why not?

We converted everything to centimeters so we could compare our findings easier. For similar drop heights our times were pretty close to each other, but they were not exactly the same because of human error. It is difficult to stop a stopwatch at the same time, and it is difficult to start the CBL light sensor at the same time.

Explain how you could change what you did in this experiment to make your predictions more accurate.

We could do five drops from the same height instead of one to get a better reading of the actual drop time. For the CBL light sensor, the person dropping the basketball could drop the basketball with one hand, and start the light sensor with the other hand to increase the accuracy and precision of the timing.

Student Check Sheet and Rubric Name_____

Number	Element	Point Value	Self-Assessment	Teacher Assessment
1	I collected sufficient data for my question.	2	2	
2	I organized the data in a useful way.	2	2	
3	I chose an appropriate model for my data.	2	2	
4	I analyzed the data in a mathematical way.	2	2	
5	I based my predictions on the data I collected.	2	2	
6	I compared and contrasted my predictions with other groups.	2	2	
7	I evaluated the reasonableness of my predictions.	2	2	
8	I included everyone in my group.	2	2	
9	I used my time in a productive manner.	2	2	
10	My work is neat, legible, and grammatically correct.	2	2	

#	Element	0	1	2
1	I collected sufficient data (at least ten data points) for my question.	No data collected.	Insufficient data collected.	Sufficient data collected.
2	I organized the data in a useful way.	No data collected.	Data is unorganized or not usefully organized.	Data is organized in a useful way.
3	I chose an appropriate model for my data.	No model chosen.	Inappropriate model	Appropriate model
4	I analyzed the data in a mathematical way.	No data analyzed.	Data is not analyzed using regression techniques.	Data is analyzed using regression techniques.
5	I based my predictions on the data I collected.	No data collected or no predictions are made.	Predictions are not based on the data collected	Predictions are based on the data collected
6	I compared and contrasted my predictions with other groups.	No group comparisons are made.	Predictions are compared to at least one but not all other same-question groups.	Predictions are compared to all other same-question groups.
7	I evaluated the reasonableness of my predictions.	No predictions are made or no evaluations are made.	N/A	I evaluated the reasonableness of my predictions.
8	I included everyone in my group.	I did not include everyone in my group.	N/A	I included everyone in my group.
9	I used my time in a productive manner.	I did not complete the task.	I spent class time off task, doing other things.	I stayed on task and used my time wisely.
10	My work is neat, legible, and grammatically correct.	My work meets none of the three qualifications.	My work meets one or two qualifications.	My work meets all three qualifications.