

**(3.3) Line of Best Fit**

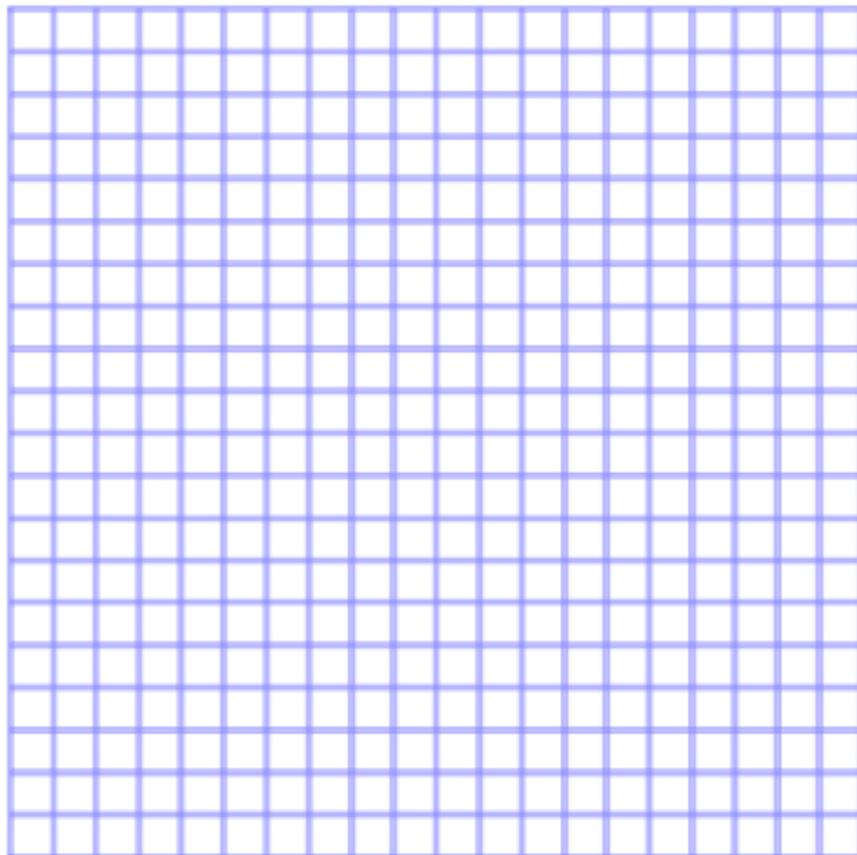
V Farms wants to find an alternative method to pricing their pumpkins, (rather than weighing them). They decide to investigate the relationship between weight and circumference. You can help V Farms by measuring the weight and circumference of the pumpkins sampled from their pumpkin patch.

Using the grid below, plot the collected data as a scatter plot.

Which variable would be the independent variable? \_\_\_\_\_.

Why? \_\_\_\_\_.

Circumference (cm)	Weight (kg)



### Lines of Best Fit

A **Line of Best Fit** (aka Trend Line, or Regression Line) is a line drawn through a set of data that best represents the **relationship** between two variables.

A good line of best fit:

- Is as close as possible to all data points.
- Follows the trend of the data points.

1. Draw a line of best fit that best describes the trend in the data on your graph.

### Extrapolation and Interpolation

One of the most important reasons to make a scatter plot is to find the trend in the data so it can be used to make predictions.

**Interpolation:** predictions are made from the **existing** points on a scatter plot.

**Extrapolation:** predictions are made about points **beyond** those on the scatter plot.

2. Find the equation of that line of best fit.

**Interpolate the data:**

3. Use the graph and your line of best fit to estimate the weight of a pumpkin that has a circumference of 85cm.

The weight of a pumpkin with a circumference of 85 cm would be \_\_\_\_\_.

**Extrapolate the data:**

4. Estimate the weight of a pumpkin that has a circumference of 200 cm.  
(*HINT: Use the equation of the line of best fit*)

### Outliers

An **outlier** is a point or points in a data set that lie(s) **outside** the trend of the other points.

Outliers are the result of **measurement errors** or **special cases**. For example, an antique car’s price would be an outlier in a set of data comparing price and age of cars.

Since outliers are so unusual or inaccurate, we **ignore** these points when we create a line of best fit. This prevents the outliers from significantly changing the trend line and the predictions made from it.

(NOTE: we usually circle the outliers and label them as such on a scatter plot)

The weight of a pumpkin with a circumference of 200 cm would be \_\_\_\_\_.

5. Suppose V Farms brings you a pumpkin that is very tall and skinny. It has a circumference of 90 cm but weighs 11 kg. Add this pumpkin to your scatter plot. Would your line of best fit be the same? Explain.

### Correlation Coefficient

- The strength of the linear relationship is measured with the **correlation coefficient** ( $r$ ).
- The correlation coefficient is a number from **-1** to **+1**.
- If the number is positive, then this indicates **positive** correlation. If the number is negative, this indicates **negative** correlation.

Negative				Positive				
Strong	Moderately Strong	Moderately Weak	Weak	Weak	Moderately Weak	Moderately Strong	Strong	
-1	-0.75	-0.50	-0.25	0	+0.25	+0.50	+0.75	+1

- The closer the number is to +1 or -1, the **stronger** the relationship. Conversely, the closer the number is to 0, the **weaker** the relationship.

6. Make a prediction about the  $r$  value for your pumpkin line of best fit. Explain your prediction.

**Coefficient of Determination**

- The **coefficient of determination** ( $r^2$ ) is an alternate measure of the strength of the linear relationship.
- Since it is squared, there are no negative values.
- If  $r^2=0.67$ , then 67% of the variance in the dependent variable is due to a change in the independent variable.

7. Use a graphing calculator to graph this data, and come up with the  $r^2$  value for the line of best fit.

Coefficient of Determination ( $r^2$ ) is \_\_\_\_\_.

This means

---

---