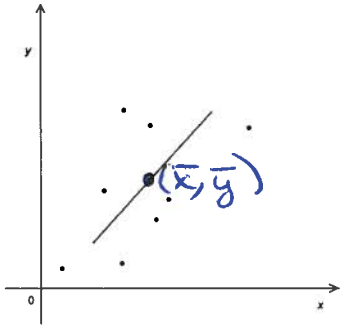
10.2 The line of best fit.

A line of best fit or trend line is drawn on a scatter diagram to find the

- direction of an association between 2 variables and
- to show the trend
- The line of best fit can then be used to make predictions.

To draw a line of best fit

- By eye draw a line that will balance the number of points above the line with the number of points below the line.
- An improvement is to have a reference point.
  - One point for the line to pass through.
  - This is the mean point
  - Calculated by finding the mean of the x-values and the mean of the y-values
  - The mean point is written as $(\bar{x}, \bar{y})$.



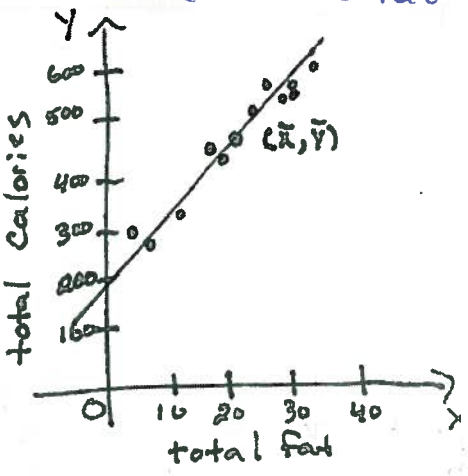
- A line of best fit is also called a regression line.

Example:

Is there a relationship between the grams of fat and the total calories in fast food?

Meal	Total fat (g)	Total Calories
Hamburger	9	260
Cheeseburger	13	320
Quarter Pounder	21	420
Quarter Pounder with cheese.	30	530
Big Burger	31	560
Toasted Sandwich	31	550
Chicken Wings	34	590
Crispy Chicken	25	500
Fish Fillet	28	560
Grilled Chicken	20	440
Grilled Chicken Light	5	300

Fast Food Calories vs Fat



$\bar{x} = 22.45$ $\bar{y} = 457.27$

Find the mean number of grams of fat and the mean number of calories.
Construct a scatter diagram for this data. (Be sure to give title + label)
Plot the mean point on your scatter diagram and use it to draw a line of best fit.

note: $22.4\dot{5}$ is the same as 22.45 (repeating decimal)

Hmwk 10B

Raw data rarely fit a straight line exactly. Usually you must be satisfied with rough predictions.

- The equation of the line of best fit, also called the regression line, can be used for prediction purposes.

Mr. Kaiser's 10 students' scores, out of 100, for their classwork and final exam are shown below.

Student	Ryan	Julia	Megan	Olive	Lilly	Keith	Elana	Gylf	Lexi	Hannah
Classwork	95	66	88	75	90	82	50	45	84	80
Final	95	59	85	77	92	70	40	50	80	absent

Hannah was absent for the final. Do not include her grades in finding the mean point.

a. Find the mean classwork score. $675/9 = 75$ $(\bar{x}, \bar{y}) = (75, 72)$

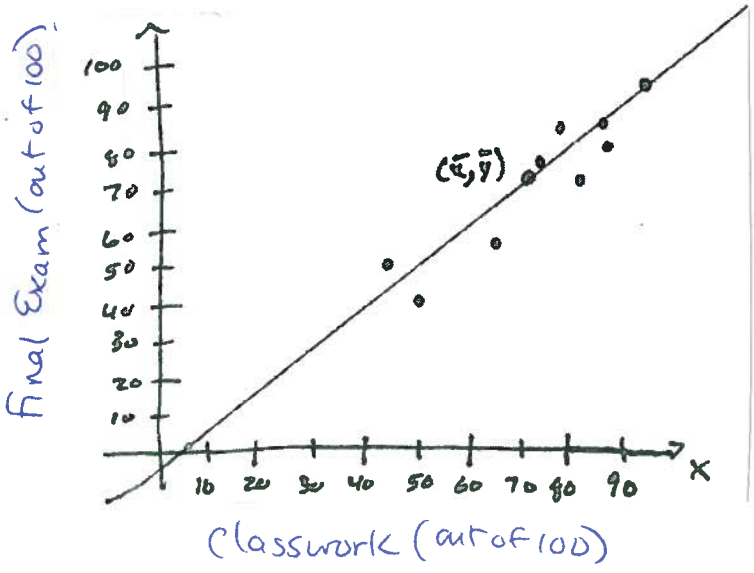
b. Find the mean final exam score. $648/9 = 72$

c. Construct a scatter diagram and draw a line of best fit through you mean point.

d. Find the equation of the regression line.

e. Use the equation of the regression line to estimate Hannah's score for the final exam.

Classwork vs Final Exam



d. Using the mean pt + Ryan's scores

slope or gradient → $m = \frac{95-72}{95-75} = \frac{23}{20}$

$y - y_1 = m(x - x_1)$ use mean pt

$y - 72 = \frac{23}{20}(x - 75)$

$y - 72 = \frac{23}{20}x - 86\frac{1}{4}$

$y = \frac{23}{20}x - 14\frac{1}{4}$ (or $y = 1.15x - 14.25$)

e. plug in $x = 84$

$y = \frac{23}{20}(84) - 14\frac{1}{4} = 82.35$ Final

Exercise 10C

* using the line of best fit to predict a data value within the range of given data is called interpolation — more reliable than extrapolation

Understanding the regression line:

Example: A study was done to investigate the relationship between the age x in years of a younger person and the time y in minutes in which the child can run one kilometer. The equation of the regression line was found to be $y = 20 - \frac{1}{2}x$.

← In stats they use $y = b + mx$ (it's the same as $y = mx + b$)

Interpret the slope and y-intercept.

The slope as it relates to this problem is as a child gets one year older their time to run one km decreases by ½ a minute or 30 seconds.

The y-int is not relevant. It means that a person 0 yrs old can run a km in 20 minutes. That is impossible!

> The y-intercept is the height of the line when $x = 0$, and might not always have a meaning. Be careful with you interpretation of the intercept. Sometime the value $x = 0$ is impossible or represents a dangerous extrapolation outside the range of the data.

Example: A biologist wants to study the relationship between the number of trees x per hectare and the number of birds y per hectare. She calculates the equation of the regression line to be $y = 8 + 5.4x$. State the gradient and the y-intercept them.

The gradient says that for every 1 tree per hectare the population of birds increases by 5.4

The y-int says if there are no trees then there are 8 birds per hectare.

> NOTE that all these interpretations follow a pattern:
>
> The **gradient** of the line is the amount by which y increases when x increases by 1 unit. (slope! but always over 1 unit)

Exercise 10d

10.3 Least squares regression

The term regression is used in statistics quite differently to other contexts.

- Regression is used for many sorts of curve fitting.
- We can construct a scatter diagram to illustrate the data, find a mean point and draw a line of best fit (regression line) through the mean point.
- Inaccuracies occur because we only have one point to draw the line through and the line of best it drawn "by eye".