

Statistics

Name: _____

Date: _____

Concepts

- Correlation of a Distribution
- Mean Deviation
- Measures of Dispersion (theory)
- Measures of Position (theory)
- Percentile Rank
- Scatter Plots
- Standard Deviation
- Statistics
- The Double Entry Table
- The Linear Correlation Coefficient
- The Mayer Line
- The Median-Median Line
- The Quintile Rank
- The Regression Line
- The Standard Score (Z-score)
- Types of Graphs in Statistics - Secondary 4

Correlation of a Distribution

1. At a climbing gym, a coach records how many training sessions per week each member does and how long (in seconds) that member takes to top the same route.

Sessions per week	1	2	3	4	5	6
Time to top the route (s)	96	88	81	75	70	66

Describe the correlation between the two variables: state its *direction*, its *form* and its *strength*, and justify each choice in one sentence.

Mean Deviation

2. A ferry makes six crossings of a river in one afternoon. The crossing times, in minutes, are

18, 21, 24, 24, 27, 30.

Calculate the mean and the mean deviation of this distribution, and interpret the mean deviation in the context.

Measures of Dispersion

3. Two machines fill 6 bags of coffee each. A quality inspector rates every bag out of 60 according to how close its mass is to the target. The scores obtained are

Machine A: 10, 10, 10, 50, 50, 50 Machine B: 10, 30, 30, 30, 30, 50.

- a) Verify that the two distributions have the same mean and the same range.
- b) Calculate the mean deviation of each and state which machine is more regular.
- c) Explain why the range is a poor measure of dispersion here.

Measures of Position**4.**

- a) Sort the following into *measures of position* and *measures of dispersion*: the minimum, the mean deviation, $R_{100}(x)$, the third quartile, the standard deviation, $R_5(x)$, the range, the maximum.
- b) A measure of position tells you nothing about how spread out a distribution is. Illustrate this by giving two distributions of five values each in which the value 3 has the same percentile rank but the ranges are very different.

Percentile Rank

5. Twenty archers each shoot one end. Their scores, in ascending order, are

14, 16, 18, 19, 21, 21, 22, 23, 23, 23, 25, 26, 27, 28, 28, 30, 31, 33, 35, 38.

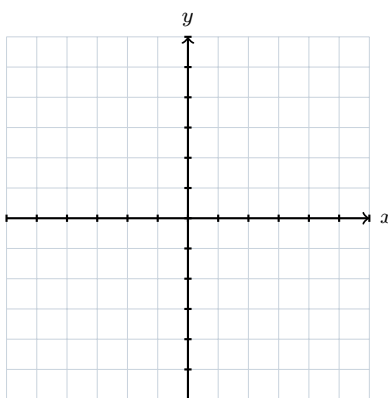
- a) Determine $R_{100}(23)$.
- b) Determine $R_{100}(31)$.

Scatter Plots

6. In a community garden, the height of a bean plant is measured once a week.

Time since planting (weeks)	1	2	3	4	5	6
Height (dm)	2	3	3	4	5	6

- Which variable is the independent one? Justify.
- Draw the scatter plot in the first quadrant of the grid below (1 square = 1 week horizontally, 1 square = 1 dm vertically).
- Describe the correlation shown.



Standard Deviation

7. An outfitter records the number of kayaks rented on five consecutive days:

14, 18, 20, 22, 26.

Treating these five days as the whole population, calculate the mean and the standard deviation.

Statistics

8. A transit agency wants to estimate how long the 48 000 residents of a town wait for a bus. It times 250 residents chosen at random.

- a) Identify the population and the sample.
- b) For each variable recorded, state whether it is qualitative, quantitative discrete, or quantitative continuous: (i) the waiting time in minutes; (ii) the bus line taken; (iii) the number of transfers made on the trip.

The Double Entry Table

9. The 50 members of a cycling club are classified by age and by the distance they ride each week.

	[0, 50[km	[50, 100[km	[100, 150[km	Total
[15, 25[years	3	8	5	16
[25, 35[years	4		6	20
[35, 45[years	7	5	2	14
Total	14	23	13	50

- a) Find the missing entry, and check it two different ways.
- b) What percentage of the members aged from 25 to 35 ride at least 100 km per week?

The Linear Correlation Coefficient

10. A scatter plot is drawn and the cloud of points is outlined by a rectangle of length $L = 12$ cm and width $w = 3$ cm. The points fall from left to right. Estimate the linear correlation coefficient and qualify the strength of the linear relationship.

The Mayer Line

11. A food truck owner compares the midday temperature with the number of smoothies sold.

Temperature ($^{\circ}\text{C}$)	2	4	6	8	10	12	14	16
Smoothies sold	14	19	21	26	32	35	38	39

Find the rule of the Mayer line, then predict the sales on a day at 20°C .

The Median-Median Line

12. A hydroponic grower records the daily hours of LED light given to nine lettuce trays and the mass harvested (in grams).

Light (h/day)	4	5	6	8	9	10	12	13	14
Mass (g)	19	25	22	30	38	34	44	49	46

Find the rule of the median-median line and use it to predict the mass harvested with 16 h of light per day.

The Quintile Rank

13. Fifteen robotics teams score the following points in a qualifying round, listed in descending order:

96, 93, 91, 88, 88, 85, 82, 80, 78, 75, 74, 70, 68, 65, 60.

- a) Find $R_5(85)$ using the formula.
- b) Find $R_5(74)$ by separating the distribution into five groups, and check it with the formula.

The Regression Line

14. A candle maker models the mass y (in grams) of a burning candle after x hours by the regression line $y = -7.5x + 240$.

- a) Interpret the values -7.5 and 240 in this context.
- b) Predict the mass after 12 h of burning.
- c) After how many hours is the candle completely consumed, according to the model?

The Standard Score (Z-score)

15. On a physics test, the group mean is 68 and the standard deviation is 8. Sofia scored 80.

- a) Calculate Sofia's standard score.
- b) Another student has a standard score of -0.75 . What was their mark?

Types of Graphs in Statistics — Secondary 4

16. For each situation, name the most appropriate type of graph and justify your choice in one sentence.

- a) The share of a town's recycling made up of glass, paper, metal and plastic.
- b) The monthly water level of a reservoir over three years.
- c) The distribution of the 180 finishing times of a road race, grouped in classes of 5 minutes.
- d) The link between the number of hours of sunshine and the mass of tomatoes harvested on 25 farms.

Synthesis — drawing on several sheets in this topic

17. A running club records, for ten of its members, the hours trained per week and the time taken to run 5 km.

Training (h/week)	1	2	3	4	5	6	7	8	9	10
Time over 5 km (min)	35	34	32	30	29	29	28	27	26	25

- a) Describe the correlation between the two variables.
- b) Determine the rule of the Mayer line for this distribution.
- c) A member plans to train 12 h per week. Predict her time and state one reason to treat the prediction with caution.
- d) In this club, the faster the runner, the better the performance. Determine the percentile rank of the member who runs 5 km in 27 min.

Want the harder problems and the worked solutions?

The *Challenge Problems* set and the *Answer Keys* for Statistics — with the reasoning written out in full — are available at tutorinmontreal.ca.