**DATA, MEASUREMENT & STATISTICS**

**analog** = **quantitative** **= continuous data:** can take on any value within a finite or infinite interval. e.g. mass, length, time, energy.

**auxiliary data:** data from an external source, such as census data, that is incorporated or linked in some way to the data collected by the study

**bias:** generally, a situation that contributes to making results deviate from the true state

**count data:** observations can take only the non-negative **integer** values (1,2,3 …)

**data distribution:** arrangement of data shown using, for example, a histogram (A histogram consists of tabular **frequencies**, shown as adjacent rectangles, built over discrete intervals (bins), with an area equal to the frequency of the observations in the interval)

**data exhaust**: data generated as trails or information byproducts resulting from all digital or online activities

**data mining**: the process of discovering new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems

**data pattern:** distribution and regularities in the data

**dependent variable**: the response that is measured, it depends on factors that are manipulated by the experimenter; e.g. the growth speed of plants

**digital data**: discrete, discontinuous binary digits

**field data**: raw unprocessed data, collected in an uncontrolled, actual environment

**independent variable**: is varied or manipulated by the researcher; e.g. the amount of sunlight

**interval data:** can be measured on a linear scale where the distance between any two adjacent units of is the same but the zero point is arbitrary; e.g. temperature in on the Celsius scale. Ratios between interval data are not meaningful, because we can’t say that 20oC is twice as hot as 10oC.

**level of measurement**: classification of scales into 4 basic types: "nominal", "ordinal", "interval" and "ratio".

**metadata**: a description of other data, i.e. description of how data are organized (structural metadata) and description of data itself (descriptive metadata); e.g. bibliographic classification used in libraries.

**nominal data**: observations can only be assigned to a given category

**order of magnitude**: an approximate measure of the size of a number, equal to the logarithm (base 10) rounded to a whole number. For example, the order of magnitude of 1500 is 3, because 1500 = 1.5 × 103.

**ordinal data**: observations can be ranked or rated but can’t be measured

**qualitative** **= categorical data:** can take on a finite number of values in terms in terms of some quality or categorization that in some cases may be 'informal' or may use relatively ill-defined, subjective characteristics such as warmth and flavor; e.g. sex, level of education, employment status

**random variable:** a variable whose value is subject to variations due to chance, it doesn’t have a fixed value but can **take on** a set of possible different values, each with an associated probability.; e.g. the outcome of a coin toss.

**ratio data:** are measured on a scale that has an absolute zero point; e.g. mass, length, time, energy. It makes sense to talk about ratios: the speed of 20mph is twice as big as that of 10mph.

**statistical inference** is the process of drawing conclusions from data subject to random variation, for example, observational errors or sampling variation

**time series:** a sequence of data taken at successive equally spaced points in time

|  |  |
| --- | --- |
| 1. a benchmark | punkt odniesienia |
| 1. a fall / rise **in** sales | spadek / wzrost (czegoś) np. sprzedaży |
| 1. a measure of length, time | miara długości, czasu, itp. |
| 1. ACCESS | mieć dostęp do |
| 1. ACCOUNT FOR | stanowić, np. *Kobiety* ***stanowią*** *70 % pracowników* |
| 1. ACQUIRE /əˈkwaɪə/ | zdobywać, wchodzić w posiadanie |
| 1. altitude | wysokość n.p.m. |
| 1. AMOUNT TO | wynosić np. suma wynosi 120 |
| 1. ANALYZE (US) / ANALYSE (UK) | analizować |
| 1. APPROXIMATE | przybliżać |
| 1. approximately = roughly | około, w przybliżeniu |
| 1. area | powierzchnia |
| 1. auxiliary data | dane pomocnicze |
| 1. BENCHMARK | wyznaczać standardy |
| 1. bias | obciążenie, błąd systematyczny |
| 1. breathalyser | alkomat |
| 1. bubble level | poziomnica |
| 1. CALCULATE = COMPUTE | obliczyć |
| 1. callipers (UK), calipers (US) | suwmiarka |
| 1. census | spis powszechny |
| 1. CHART, PLOT, GRAPH | przedstawić na wykresie |
| 1. chart, plot, graph | wykres |
| 1. COLLECT | zbierać, gromadzić |
| 1. confidential data | dane poufne |
| 1. CONTINUE to fall | nadal spadać |
| 1. continuous data | dane ciągłe |
| 1. CONVERT … into | przeformatować |
| 1. COPY from … to | kopiować |
| 1. count / discrete data | dane skokowe (dyskretne) np. liczba dzieci |
| 1. data breach | wyciek poufnych danych |
| 1. data breakdown, cross-classification | przekrój (danych) |
| 1. data broken down / cross-classified by sex | dane w przekroju na płeć |
| 1. data exhaust | ślad cyfrowy |
| 1. data set | zbiór danych |
| 1. data-driven decision making | podejmowanie decyzji w oparciu o dane |
| 1. dating methods | metody datowania |
| 1. degree | stopień (np. temperatury) |
| 1. demand for | popyt na, zapotrzebowanie na |
| 1. dependent, response, target variable | zmienna zależna |
| 1. DERIVE from | wywodzić, wyciągać z (np. danych) |
| 1. DETECT | mierzyć, wykrywać |
| 1. detector (metal ⌐, particle ⌐) | wykrywacz, (np. metalu, cząstek) |
| 1. DETERIORATE | pogarszać |
| 1. DETERMINE /dɪˈtɜːmɪn/ = ESTABLISH | określić, zmierzyć, ustalić |
| 1. DIGITISE /dydżytaiz/ | przekształcać na format cyfrowy |
| 1. DISPLAY data on a graph, chart, plot | przedstawiać dane na wykresie |
| 1. DISSEMINATE | rozpowszechniać |
| 1. distance | odległość |
| 1. DISTRIBUTE | rozpowszechniać |
| 1. drive space / capacity / volume | pojemność dysku |
| 1. downward / falling trend | tendencja spadkowa |
| 1. ESTIMATE /estimeit/ | szacować |
| 1. estimate /estimǝt/ | szacunek, oszacowanie, wartość oszacowana |
| 1. event | zdarzenie losowe |
| 1. facts and figures | fakty i dane liczbowe |
| 1. FALL, DECREASE, DECLINE, | spaść |
| 1. figure | rysunek w publikacji (zawierający wykres lub mapę) |
| 1. findings of a study | wyniki badania |
| 1. FLUCTUATE between ... and ... | wahać się pomiędzy … a … |
| 1. gauge /geidż/ fuel ⌐, pressure ⌐ | wskaźnik, miernik (np. paliwa, ciśnienia) |
| 1. height | wzrost, wysokość |
| 1. IMPROVE | polepszać |
| 1. **in**-accurate = **im**-precise = **in**-exact = | nie-dokładny, nie-precyzyjny |
| 1. independent, explanatory variable   predictor | zmienna niezależna, wyjaśniająca, predyktor |
| 1. INFER from | wnioskować na podstawie |
| 1. inference | wnioskowanie |
| 1. interval between 0 and 3 | przedział |
| 1. KEEP rising = CONTINUE to rise | ciągle rośnie |
| 1. location | położenie |
| 1. MAKE / TAKE measurement | robić pomiary |
| 1. mean = average | średnia, średni |
| 1. MEASURE | mierzyć |
| 1. measuring glass / cup | miarka kuchenna |
| 1. meter /miːtə/ | miernik |
| 1. missing data | braki danych |
| 1. nonresponse | braki odpowiedzi |
| 1. normal distribution | rozkład normalny (Gaussa) |
| 1. normally-distributed data | dane o rozkładzie normalnym |
| 1. OBTAIN | zdobywać, uzyskiwać |
| 1. ORDER | szeregować, porządkować |
| 1. order of magnitude | rząd wielkości (np. tysiące, miliony, miliardy) |
| 1. outcome (of an experiment / elections / the meeting, the process) | wynik |
| 1. outlier | obserwacja odstająca, o nietypowej wartości w stosunku do większości |
| 1. pattern | prawidłowość (w danych) |
| 1. People in the north of England are 20*%* **more likely to**die before they reach 75 | Jest 20% większe prawdopodobieństwo, że … w porównaniu do …  Prawdopodobieństwo, że … jest o 20% większe niż dla … |
| 1. physical quantity | wielkość fizyczna (np. masa) |
| 1. PRESENT = DISPLAY | prezentować, pokazywać |
| 1. pressure | ciśnienie |
| 1. PROCESS | przetwarzać |
| 1. PROTECT against / from | chronić |
| 1. protractor | kątomierz |
| 1. quantity = amount | ilość |
| 1. random | losowy |
| 1. RANGE from … to | być w zakresie od .. do |
| 1. RANK | ustawić w kolejności |
| 1. rate (birth rate, unemployment rate) | odsetek, stopa |
| 1. **ratio of** circumference **to** diameter | stosunek obwodu do średnicy |
| 1. ratio scale | skala ilorazowa, stosunkowa |
| 1. REACH a high / low of 19.50 | osiągnąć niski / wysoki poziom o wartości … |
| 1. RECOVER | odzyskiwać (zniszczone/uszkodzone) dane |
| 1. RELEASE | udostępniać |
| 1. REMAIN at the same level | pozostać na tym samym poziomie |
| 1. remote sensing | teledetekcja |
| 1. RETRIEVE | odnajdywać potrzebne informacje (w bazie danych lub na dysku) |
| 1. RISE, INCREASE, GROW, GO UP | wzrastać |
| 1. ROUND OFF to the nearest tenth, hundredth | zaokrąglić do części dziesiętnej, setnej |
| 1. ruler | linijka |
| 1. sample size | wielkość próby |
| 1. sampling | losowanie, pobieranie próby |
| 1. scales | waga (narzędzie) |
| 1. SCRAPE date, scraped data /skreipd/ | zbierać pozyskiwać dane z sieci, dane pozyskane ze stron www |
| 1. SECURE | zabezpieczać |
| 1. sensor | czujnik |
| 1. share / percentage | udział (procentowy) |
| 1. SHARE … with | udostępniać, dzielić się |
| 1. spatial analysis | analiza przestrzenna |
| 1. speed | prędkość |
| 1. speed camera | fotoradar |
| 1. speed(radar) gun | radar policyjny |
| 1. speedometer | prędkościomierz |
| 1. stopwatch | stoper |
| 1. STORE | przechowywać (np. na dyskach) |
| 1. study | badanie naukowe |
| 1. survey | badanie reprezentacyjne |
| 1. surveying | miernictwo, geodezja (szczegółowa) |
| 1. tachometer /tæˈkɒmɪtə/   = rev-counter, RPM gauge | obrotomierz |
| 1. tape measure | taśma miernicza |
| 1. TEND to occur | mieć tendencję do występowania |
| 1. the trend LEVELs OFF | trend wyrównuje się, osiąga stały poziom |
| 1. the trend RECOVERs | trend poprawia się (przestaje spadać i zaczyna rosnąć) |
| 1. time series | szereg czasowy |
| 1. **un**-certain | niepewny |
| 1. upward / growing / rising trend | tendencja wzrostowa |
| 1. volume | 1. objętość 2) głośność 3) ilość |
| 1. WEIGH /wei/ | ważyć |
| 1. weight / mass | waga / masa |
| 1. weighted average | średnia ważona |

**Be able to explain:**

* 4 scales of measurement
* quantitative vs qualitative data
* continuous vs discrete data
* analogue vs digital data