

Glossary

A

accidental relationship A relationship that is based purely on coincidence.

aggregate data Data that are combined or summarized in such a way that the individual microdata can no longer be determined.

arrangement An ordered list of items.

attribute A quality or characteristic given to a person, group, or object.

B

bias Occurs when there is a prejudice for or against an idea or response. Biased samples can result from problems with either the sampling technique or the data collection method.

binomial probability distribution

A distribution with independent trials whose outcomes are either success or failure.

binomial theorem A formula used to expand $(a + b)^n$.

$$(x + y)^n = {}_n C_0 x^n y^0 + {}_n C_1 x^{n-1} y^1 + {}_n C_2 x^{n-2} y^2 + \dots + {}_n C_{n-1} x^1 y^{n-1} + {}_n C_n x^0 y^n$$

The general term is ${}_n C_r x^{n-r} y^r$.

C

categorical (qualitative) data Data that can be sorted into distinct groups or categories.

cause and effect relationship The correlation between two variables in which a change in one directly causes a change in the other.

combination A selection from a group of objects without regard to order. The number of r objects chosen from a set of n items is

$${}_n C_r = \frac{n!}{(n - r)! r!}$$

common cause relationship The correlation between two variables in which both variables change as a result of a third common variable.

complement A set of possible outcomes not included in an event.

compound events Multiple events in a probability experiment which may or may not affect each other.

conditional probability The probability of a second event occurring, given that a first event occurred. The sample space for the second event is reduced from the first event.

confidence interval The range of possible values of a measured statistic at a particular confidence level.

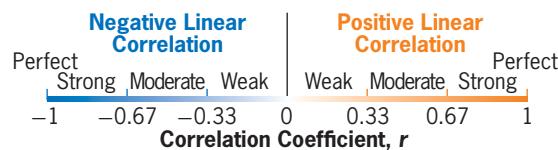
confidence level The probability that a particular statistic is within the range indicated by the margin of error.

continuity correction A correction applied when using the normal approximation to correct for the difference between a discrete and continuous distribution.

continuous random variable A variable that can have an infinite number of possible values in a given range.

control group The participants in an experiment who do not receive the specific treatment being measured.

correlation coefficient A measure of how well a linear model fits a two-variable set of data.



D

dependent events The occurrence or non-occurrence of one event influences the probability of the other event occurring.

discrete random variable A variable that can have only certain values within a given range.

E

event A set of outcomes that have a common characteristic.

expectation (expected value) The predicted average of all possible outcomes in a probability distribution.

$$\begin{aligned}E(X) &= x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \dots x_n \cdot P(x_n) \\&= \sum_{i=1}^n x_i \cdot P(x_i)\end{aligned}$$

experimental probability Probability based on experimental trials.

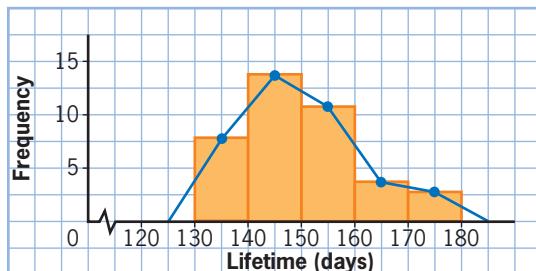
$P(A) = \frac{n(A)}{n(T)}$, where $P(A)$ is the probability that outcome A occurs, $n(A)$ is the number of times that outcome A occurred, and $n(T)$ is the total number of trials.

F

factorial A product of sequential natural numbers with the form
 $n! = n(n - 1)(n - 2) \dots \times 2 \times 1$.

frequency histogram A graph with intervals on the horizontal axis and frequencies on the vertical axis.

frequency polygon A segmented line that joins the midpoints of the top of each column in a frequency histogram.



fundamental counting principle If one event can occur in m ways and a second event can occur in n ways, then together they can occur in $m \times n$ ways.

H

hidden variable A variable that affects or obscures the relationship between two other variables.

hypergeometric probability distribution A distribution with dependent trials whose outcomes are either success or failure.

hypothesis A prediction about the relationship between variables or about the outcome of a research question.

I

independent events Situations in which the occurrence or non-occurrence of one event has no influence on the probability of the other event occurring.

indirect method A method of solving a permutation where you subtract the number of unwanted outcomes from the total number of outcomes without restrictions.

interquartile range (IQR) The difference between the first and third quartiles.

L

line of best fit A straight line that represents a trend in the scatter plot as long as the pattern is more or less linear.

linear correlation A relationship in which a change in one variable tends to correspond to a proportional change in another variable.

linear regression The formal process by which a line of best fit is mathematically determined.

M

margin of error The range of values that a particular measurement is said to be within.

mean The sum of the data entries divided by the number of entries.

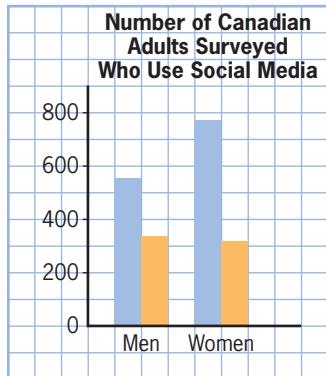
measurement bias When the survey collection method is such that the characteristics are consistently over- or under-represented.

median The middle value of all the data points when the data values are listed in order from least to greatest.

microdata An individual set of data about a single survey respondent.

mode The data value that occurs most often in a data set.

multiple bar graph A graph in which different quantities are represented by different colours and lengths of bars that are placed side-by-side.



mutually exclusive events Events that have different attributes and cannot occur simultaneously.

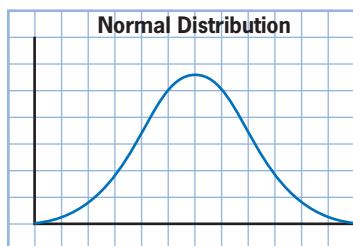
N

nominal data Qualitative data that cannot be ranked.

non-mutually exclusive events Different events that can happen at the same time.

non-response bias When the opinions of survey respondents differ in meaningful ways from those of non-respondents.

normal distribution A probability distribution around a central value, dropping off symmetrically to the right and left, forming a bell-like shape.



null set A set with no elements.

numerical (quantitative) data Data in the form of any number.

O

odds against The ratio of the probability that an event will not happen to the probability that it will.

$$A = P(A) : P(A')$$

odds in favour The ratio of the probability that an event will happen to the probability that it will not.

$$A = P(A') : P(A)$$

ordinal data Qualitative data that can be ranked.

outcome A possible result of an experiment.

outlier An element of the data set that is significantly different from the rest of the data points.

P

Pascal's triangle A triangular array of numbers in which each term is the sum of the two terms above it.

			1		
		1	1	1	
	1	2	1		
1	3	3	1		
1	4	6	4	1	

percentile The percent of all the data that are less than or equal to a specific data value.

$$p = \frac{(L + 0.5E)}{n}, \text{ where } p \text{ is the percentile, } L \text{ is the number of data less than the data point, } E \text{ is the number of data equal to the data point, and } n \text{ is the size of the population.}$$

permutation An arrangement of n distinct items in a definite order. The total number of these permutations is written nPn or $P(n, n)$. The number of permutations of n items is ${}_nP_n = n!$.

population All the individuals in a group that is being studied.

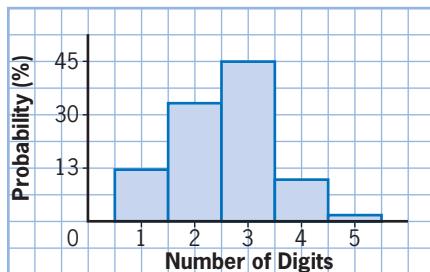
presumed relationship A relationship that makes sense but does not seem to have a causation factor.

primary source data Data that have been collected directly by the researcher and have not been manipulated or summarized.

probability The likelihood of something occurring.

probability distribution The probabilities for all possible outcomes of an experiment or sample space.

probability histogram A graph of a probability distribution in which equal intervals are marked on the horizontal axis and the probabilities associated with these intervals are indicated by the areas of the bars.



Q

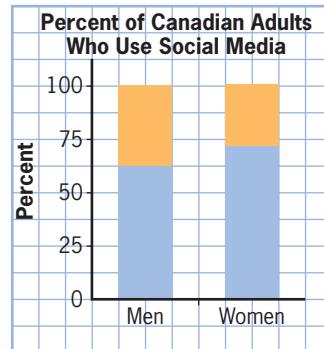
quartiles Three points that divide the data set into four equal groups. The first quartile (Q_1) is the middle number between the smallest number and the median. The second quartile (Q_2) is the median of the data set. The third quartile (Q_3) is the middle number between the median and the largest number in a data set.

R

random variable A quantity that can have a range of values. Designated by X , with individual values designated by x .

range The difference between the highest value and the lowest value of a data set.

relative split bar graph A graph in which different percents, totalling 100, are represented by different colours and lengths of bars that are placed one above the other.



reliable data Results of a study that can be duplicated in another study.

research question A question about a topic or problem that can be investigated or solved.

residual The difference between a data point's actual dependent value and the dependent value predicted by the line of best fit.

residual plot A graph which shows the value of each residual graphically as the vertical distance from a horizontal axis.

response bias When survey respondents change their answers to influence the results, to avoid embarrassment, or to give the answer they think the questioner wants.

reverse cause and effect relationship

A relationship in which the independent and dependent variable are reversed.

S

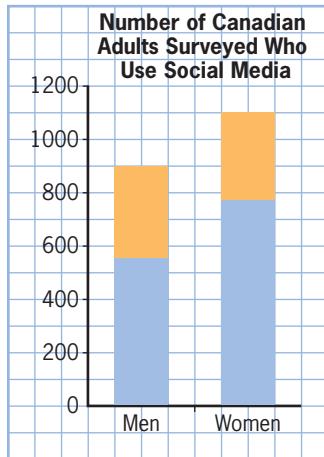
sample A group of items or people selected from the population.

sample space A collection of all possible outcomes. Sometimes called a sample set.

sampling bias When the sample does not closely represent the population.

secondary source data Data used by someone other than those who actually collected them.

split bar graph A graph in which different quantities are represented by different colours and lengths of bars that are placed one above the other.



standard deviation The average distance of the scores from the mean.

population standard deviation:

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

sample standard deviation:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}, \text{ where the population}$$

deviation is represented by $(x - \mu)$ and the sample deviation is represented by $(x - \bar{x})$.

subjective probability A probability estimate based on intuition.

T

theoretical probability A probability based on analysis of all possible outcomes.

$$P(A) = \frac{n(A)}{n(S)}, \text{ where } P(A) \text{ is the probability}$$

that event A can occur, $n(A)$ is the number of ways it can occur, and $n(S)$ is the total number of possible outcomes in the sample space.

treatment group The participants in an experiment who receive the specific treatment being measured.

U

uniform distribution Occurs when, in a single trial, all outcomes are equally likely.

$P(x) = \frac{1}{n}$, where n is the number of possible outcomes in the experiment.

V

valid data Results that accurately represent the entire population.

variability (in samples) Shows how samples are different from each other. The more similar the samples are, the lower the variability and the more accurately they represent the population.

variance The average squared difference of the scores from the mean.

population variance: sample variance:

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N} \quad s^2 = \frac{\sum(x - \bar{x})^2}{n - 1},$$

where the population deviation is represented by $(x - \mu)$ and the sample deviation is represented by $(x - \bar{x})$.

W

weighted mean The mean of a set of numbers that are given weightings based on their frequency.

Z

z-score The number of standard deviations an observation is from the mean.

population z-score: sample z-score:

$$z = \frac{x - \mu}{\sigma} \quad z = \frac{x - \bar{x}}{s}$$

Index

A

accidental relationship, 397
additive principle for mutually exclusive events, 37
see also rule of sum
adjust for inflation, 243
aggregate data, 224
airline flight delays, 277
APA style, 461
arrangement, 76–77, 119
attribute, 320
average net worth, 276

B

bar graph, 31, 195, 250
cluster bar graph, 195
multiple bar graph, 290, 291, 294, 295
relative split bar graph, 290, 291
split bar graph, 290, 291, 295
stacked bar graph, 294
bell curve, 332
bias, 216, 381
big numbers, 237
identification of, 234
infographics, interpretation of, 238–239
measurement bias, 233
misleading statistics, 235–236
non-response bias, 233
response bias, 233
sampling bias, 233
in two-variable data, 416–417
type of bias, 233
big numbers, 237
binomial probability distribution, 160, 163–165, 184
application, 166
counting successes, 162
and hypergeometric distributions, comparison of, 180
independent trials, 170
investigation of, 160–161
normal approximation, 365, 366–367
and normal distribution, comparison of, 362–363
outcomes, 366
probability in a binomial distribution, 162
binomial theorem, 124–125, 143
the birthday problem, 92, 95
box and whisker plot, 269, 270–273
see also side-by-side box plots
box office hits, 242

box plot. *See* box and whisker plot; side-by-side box plots
bubble plot, 428–429

C

Canadian census, 211, 302–303, 307
CANSIM, 222, 302, 303–304
capital sigma, 149, 279
careers
computer programmer, 60
data miner, 192
health care statistician, 378
insurance underwriter, 2
management science, 100
mechanical engineer, 316
meteorologist, 140
risk assessment manager, 248
statistician, 440
categorical (qualitative) data, 198, 225, 250
categorical nominal data, 198
categorical ordinal data, 198
cause and effect, 392–399
cause and effect relationship, 394–395
cell phones, and driving, 440, 455
census, 211, 222, 302–303, 307
census metropolitan area, 303
census profiles, 302–303
central tendency, measures of. *See* measures of central tendency
chances of winning, 129
Chebyshev, Pafnuty, 351
Chebyshev's Theorem, 351
checklists
probability project self-assessment checklist, 472
statistics project self-assessment checklist, 476
choose, then arrange, 119
choose from groups, 130
circle graph, 251
clinical trials, 221
cluster bar graph, 195
cluster sample, 207
clustered column graph, 294
coefficient of determination, 387, 391
combination, 111
choose, then arrange, 119
counting cases, 118–119
diagram, interpretation of, 112
items from a set, 111
more than one group, 112
non-ordered selections, 110–111
number of sums of money, 116
and Pascal's triangle, 122–125
probabilities using combinations, 128–132
problem solving with, 116–120
solving counting problems with, 100
total number of subsets, 117
combination of real trials, 31
comma-separated values, 303
common cause relationship, 395–397
comparison
discrete distributions, 180–184
experimental and theoretical probabilities, 26–32
filter, 424–427
graphical comparison of groups, 294–295
normal distribution and binomial distribution, 362–363
normal distribution and hypergeometric distribution, 364–365
two similar distributions, 181–183
types of data, 199
complement, 20–21
compound events, 44
compound independent events, 47
different compound events, 47–49
simple compound events, 44–45
computer programmer, 60
concave, 115
concept circles, 378
conditional probability, 51–52
confidence intervals, 352–358
confidence level, 352, 353, 357–358, 359
Consumer Price Index (CPI), 223, 307
contingency table, 411
continuity correction, 365
continuous data, 198, 322
continuous numerical data, 198
continuous probability distributions
calculation of probabilities, 332
confidence intervals, 352–358
connections to discrete random variables, 362–369
normal distribution, 344, 346–349, 362–369
probability density function, 322
uniform distributions, 322–323
continuous random variable, 146, 320–327
continuous sample space, 146
control, 213
control group, 213
convenience sample, 207
convex, 115
convex octagon, 115
correlation coefficient, 384, 386–388
correlation strength, 384–385

counting

- cards, 34–35
- cases, 118–119
- outcomes, 64–66
- permutations, 78
- repeated independent trials, 72
- repeated trials without replacement, 72
- successes, 162

counting story

101, 137

critiquing other projects, 470–471, 475, 479

cross-sectional study, 220

CSV (comma-separated values), 303

cube, 158

cubit, 326

culminating probability project

analysis of, 447–454

assessment of plan, 446

assessment of project, 454, 470–471

data collection, 446

finalization of project, 453–454

planning and carrying out, 442–446

planning the report, 452

prepping, 445

probability project content rubric, 473

probability project peer critique form, 475

probability project presentation rubric, 474

probability project self-assessment checklist, 472

writing the report, 452

writing your report, 452

culminating statistics project

analysis of, 466–469

assessment of data analysis, 468

assessment of project, 465, 471

assessment of results, 469

assessment of validity of process and results, 469

defining and posing the problem, 456–461

digging in to the centre of the data, 467

graphing the data, 465

planning and carrying out, 455–465

primary data, 464–465

research process, 455

searching for data, 457–461

searching for patterns, 467–468

secondary data, 461–464

statistics project content rubric, 477

statistics project peer critique form, 479

statistics project presentation rubric, 478

statistics project self-assessment checklist, 476

D

data

aggregate data, 224

analysis of database, 226–227

bias, 216, 233–239

categorical (qualitative) data, 198, 225, 250

collection of. *See* data collection

comparison of types of data, 199

concepts and graphical summaries, 196–203

continuous data, 198, 322

discrete data, 198, 321

distorted data, 418–419

graphing data, 465

grouped data, 257–260

interpretation of data, 196, 224–225, 228–229

microdata, 224

with more than one variable, 200–202

nominal data, 198

numerical (quantitative) data, 198, 250

ordinal data, 198, 250

outliers, 254, 256

primary source data, 224, 464–465

reliable data, 292

searching for data, 457–461

seasonal adjustment, 205

secondary data, 224, 234

secondary source data, 461–464

sensational use of data, 420–421

skewed data, 256

Statistics Canada data, 222–223, 224–225, 302–305

surveying for data, 462–464

trend on time graph, 228–229

uses and misuses of data, 416–421

valid data, 292

variability in, 197

data analysis

advanced techniques, 424–431

assessment of, 468

cause and effect, 392–399

database, analysis of, 226–227

dynamic analysis of two-variable data, 402–412

hidden variable, 411–412

interpretation of data, 196, 224–225, 228–229

line of best fit, 382–389

measures of central tendency, 252–262

measures of spread, 266–275

one-variable data analysis. *See* one-variable data analysis

outliers, accounting for, 406–410

residual plot, 402–405

standard deviation, 278–283

statistical summaries, 290–298

Statistics Canada data, 302–305

two-variable data analysis. *See* two-variable data analysis

uses and misuses of data, 416–421

z-score, 283–285

data collection

culminating probability project, 446

experiment, 212–214

principles of, 206–210

survey, 215–216, 217–218

data miner

data mining, 192

database

database analysis, 226–227

decimals, 4

degree, 125

demographic information, 215

dependent events, 49–51

dependent trials, 90, 170

dependent variable, 382, 401

derangement, 87

diagonal patterns in Pascal's triangle, 123

dice

octahedral dice, 96

Sicherman dice, 144

simulation, 29–31

tetrahedral dice, 167

Triple Your Chances, Double Your Counters, 442–443, 447–449

different compound events

47–49

digital files and collections

226

discrete data, 198, 321

discrete numerical data

198

discrete probability distributions

321

binomial distributions, 160–167

calculation of probabilities, 332

comparison and selection, 180–184

hypergeometric distributions,

170–178

uniform distributions, 154–159

discrete random variable

146, 320–321, 362–369

see also discrete probability distributions

discrete sample space

146

distance

428

distinct objects in fixed order

107

distorted data

418–419

dodecahedron

158

double blind study

243

double factorial

81

E

economics

401

encryption

60, 61, 99

event

16, 17

compound events, 44–45

compound independent events, 47

dependent events, 49–51

independent events, 45–49
multiple events, 44
mutually exclusive events, 35–38
non-mutually exclusive events, 38–41
odds of an event, 22–23
expectation (expected value), 148, 150
for a binomial distribution, 162
for a hypergeometric distribution, 172, 177
for whole number answer, 177
expected value. *See* expectation (expected value)
experiment, 212–214
experimental probability, 6, 450–451
application, 10
calculation, 7, 8–9
investigation, 7
predictions, 9
and theoretical probability, comparison of, 26–32
experimental studies, 213
exponential regression, 401
extrapolation, 383

F

factorial, 77
binomial probability, 365
double factorial, 81
fair game, 156
Fathom™
coefficient of determination, 387
comparison of binomial distribution and normal distribution, 362–363
comparison of hypergeometric distribution and normal distribution, 364–365
correlation coefficient, 387
dice simulation, 30–31
hidden variable, accounting for, 411–412
interquartile range and box and whisker plots, 271
legend attribute, 429–430
normal distribution, 346–347
outliers, accounting for, 406–407
probability distributions, 146
side-by-side box plots, 425–426
survey, 217
filter, 424–427
fixed order, 107
flowchart, 100
fractions, 4
Frayer model, 60
frequency distribution, 334–336
frequency histogram, 143, 144, 324–326
frequency polygon, 324–326, 333
frequency table, 316, 324–326
fundamental counting principle, 47, 70–72, 79, 82

G

game analysis, 3, 59
game construction, 443–446
games of chance, 442–446
geometric mean, 265
geometric probability distributions, 187
Google Docs, 218
graph, 165
bar graph, 31, 195, 250
box and whisker plot, 269, 270–273
bubble plot, 428–429
circle graph, 251
cluster bar graph, 195
clustered column graph, 294
comparison of groups, 294–295
data, graphing, 465
frequency histogram, 143, 144, 324–326
hidden variable, 428–430
histogram, 31, 251, 318
multiple bar graph, 290, 291, 294, 295
pictograph, 250
relative split bar graph, 290, 291
residual plot, 402–405, 406–410
scatter plot, 380, 406–410
side-by-side box plots, 299, 425–426
split bar graph, 290, 291, 295
stacked bar graph, 294
stacked column graph, 294
stem and leaf plot, 31
time graph, 228–229
time series graph, 307
graphic organizer, 140
graphical summaries, 250
graphing calculator
binomial distribution, 163–165, 166
coefficient of determination, 387
comparison of two similar distributions, 181–182
correlation coefficient, 386–387
dice simulation, 30
frequency polygon, 325–326
hypergeometric distribution, 173–175, 176
infinity, lack of symbol for, 340
interquartile range and box and whisker plots, 272–273
measures of central tendency, evaluation of, 255
normal distribution, 346
outliers, accounting for, 408–410
probability distributions, 145
random number generation, 15, 70
scatter plot, 380
spot landing contest, 340
three-coin-flip simulation, 27
 z -score, 284
grouped data, 257–260

H

health care statistician, 378
hidden variable, 411–412, 428–430
histogram, 31, 251, 318
hypergeometric probability distribution, 170, 172, 173–175, 184
application to selections, 176–177
and binomial distributions, comparison of, 180
dependent trials, 170
expectation for a hypergeometric distribution, 172
expectation formula, 177
investigation of, 170–171
normal approximation, 365, 367–368 and normal distribution, comparison of, 364–365
probability in a hypergeometric distribution, 172
hypothesis, 455

I

icosahedron, 158
identical elements, 104–105, 106
imperial units, 328, 340, 361
independent events, 45–49
independent trials, 89, 170
independent variable, 382, 401
indirect method, 85
counting cases, 118–119
total number of subsets, 117
inference, 213
infographics, 238–239
insurance underwriter, 2
interpolation, 383
interquartile range (IQR), 269, 270–273, 466
intersection set, 39
interval width, 333
IQR. *See* interquartile range (IQR)

L

leading question, 215
legend attribute, 429–430
like elements, 105–106
line of best fit, 382–389
linear correlation, 384
linear model, 387
linear regression, 388, 401
literacy strategy
concept circles, 378
flowchart, 100
Frayer model, 60
frequency table, 316
graphic organizer, 140
mathematics word wall, 248
mind map, 192
summary, 100
Venn diagram, 2
longitudinal study, 219

M

management science, 100
margin of error, 352, 356, 357, 361
mathematics word wall, 248
mean, 250, 252, 466
 geometric mean, 265
 for grouped data, 257
 harmonic mean, 265
 population mean, 252
 sample mean, 252, 318, 357–358
 trimmed mean, 265
 values within one standard deviation of mean, 348
 weighted mean, 148, 149, 261–262
measurement bias, 233
measures of central tendency, 250, 252–262
 evaluation of, 254–257
 interpretation of, 292–293
 mean, 250, 252
 see also mean
 median, 250, 252, 466
 mode, 250, 252, 466
measures of spread, 266–275
 interquartile range (IQR), 269, 270–273
 percentile rank, 267
 percentiles, 266–267, 268–269
 quartiles, 269, 273–274
 range, 269
 standard deviation, 278–283
 variance, 279, 281–282
mechanical engineer, 316
the media
 critical analysis of media claims, 296–297
 and statistical claims, 290–291
median, 250, 252, 466
meteorologist, 140
meteorology, 2
metric unit, 361
microdata, 224
mind map, 192, 445
misleading statistics, 235–236
MLA style, 462
mode, 250, 252, 466
money, number of sums of, 116
multiple bar graph, 290, 291, 294, 295
multiple events, 44
multiplicative counting rule. *See*
 fundamental counting principle
multiplicative principle for dependent events, 51
multiplicative principle for independent events, 47
 see also fundamental counting principle
multistage sample, 207
mutually exclusive events, 35–38

N

national census. *See* census
National Household Survey, 211
negative linear correlation, 384
negative *z*-score, 318
nominal data, 198
non-linear model, 387
non-mutually exclusive events, 38–41
non-ordered selections, 110–111
non-replacement of items, 66
non-response bias, 233
normal distribution, 332, 344
 applications of, 346–349
 areas under normal distribution curve, 480–481
 and binomial distribution, comparison of, 362–363
 and hypergeometric distribution, comparison of, 364–365
 normal distribution curve, 340
null set, 117
number of sums of money, 116
numerical (quantitative) data, 198, 250, 269

O

observational studies, 213, 219
octahedral dice, 96
octahedron, 158
odds, 21, 22–23
odds against, 21
odds in favour, 21
one-variable data analysis
 measures of central tendency, 252–262
 measures of spread, 266–275
 standard deviation, 278–283
 statistical summaries, 290–298
 Statistics Canada data, 302–305
 z-score, 283–285
opinion polls, 169
ordered selections, 91
ordinal data, 198, 250
organized counting, 64–67
outcome, 6, 17
 counting outcomes, 64–66
 determination of number of outcomes, 70–71
 illustration of outcomes, 64–66
outliers, 254, 256, 406–410

P

parameter, 181
Pascal, Blaise, 122
Pascal's triangle, 102, 122–125, 131
password encryption, 61, 99
patterns, 467–468
Pearson, Karl E., 344
percent, 4, 62, 236
percentage point increase, 236
percentile rank, 267
percentiles, 266–267, 268–269
permutations, 60, 77
 arrangement, 76–77
 counting permutations, 78
 derangement, 87
 factorials, 77
 fundamental counting principle, 70–72
 with identical elements, 104–105
 with like elements, 105–106
 of *n* items, 77
 with non-ordered elements, 104–107
organized counting, 64–67
probability problems, 88–93
of *r* items out of *n* items, 79
with restrictions, 79
rule of sum, 82–86
with several identical elements, 106
solving counting problems with, 100
of some items in a set, 78
pictograph, 250
pivot table, 426–427
placebo, 219, 221
platonic solids, 158
Poisson, Siméon-Denis, 187
Poisson distribution, 187
population, 206, 276
population formulas, 282
population mean, 252
population standard deviation, 279, 283
population variance, 279
population *z*-score, 283
positive linear correlation, 384
positive *z*-score, 318
powers, 100
predictions, 9
presentation, 470–471, 474, 478
presumed relationship, 397
primary source data, 224, 464–465
principle of inclusion and exclusion, 39, 84
probability, 6
 in a binomial distribution, 162
 combinations, use of, 128–132
 comparison of experimental and theoretical probabilities, 26–32
complement, 20–21
compound events, 44–45
conditional probability, 51–52
dependent events, 49–51
estimate of, 334–336
experimental probability, 6–11, 26–32, 450–451
 in a hypergeometric distribution, 172
independent events, 45–49
mutually exclusive events, 35–38
non-mutually exclusive events, 38–41

odds against, 21
odds in favour, 21
odds of an event, 22–23
outcomes, 6
permutations, use of, 88–93
for range of values, 319
simple probabilities, 6–12
subjective probability, 11–12, 22
sum of probabilities, 9
theoretical probability, 16–23, 26–32
probability density distribution, 323
probability density function, 322
probability distributions, 145
binomial distributions, 160–167,
 180, 184
comparison of distributions, 180–184
continuous probability distributions.
 See continuous probability
 distributions
discrete probability distributions.
 See discrete probability
 distributions
expectation of, 148–149
geometric probability distributions,
 187
hypergeometric distributions, 170–
 178, 180, 184
investigation of, 144–146
normal distribution, 344, 346–349,
 362–369
probability histogram, 147
two similar distributions,
 comparison of, 181–183
uniform distributions, 154–157, 184,
 322–323
probability histogram, 147
probability project. *See* culminating
 probability project
probability project peer critique form,
 475
probability project presentation rubric,
 474
proportion, 236

Q

qualitative data, 198
quantitative data, 198, 269
quartiles, 269, 273–274
questionable statistics, 235–236

R

random act, 5
random device, 153
random number generation, 15, 70
random variable, 145, 184
 continuous random variable, 146
 discrete random variable, 146
randomization, 60, 213
range, 269, 466
relationships, 394–398

relative frequency histogram, 144
relative split bar graph, 290, 291
reliable data, 292
repeated independent trials, 72
repeated sampling, 357
repeated trials without replacement, 72
replacement of items, 66
replication, 213
reports, 452, 470–471
research process, 455
research question, 455
residual, 403
residual plot, 402–405, 406–410
response bias, 233
reverse cause and effect relationship, 397
risk assessment manager, 248
routes on a grid, 124
rubrics
 probability project content rubric,
 473
 probability project presentation
 rubric, 474
 statistics project content rubric, 477
 statistics project presentation rubric,
 478
rule of sum, 37–38, 82–86, 117, 130

S

sample, 206, 207–209
sample mean, 252, 318, 357–358
sample size, 357
sample space, 16, 18
 continuous sample space, 146
 discrete sample space, 146
sample standard deviation, 279, 283,
 318, 357
sample variance, 279
sample *z*-score, 283
sampling bias, 233
sampling method, 208
scatter plot, 380, 406–410
scientific calculator
 binomial distribution, 163, 166
 hypergeometric distributions, 176
 z -score, 283–284
secondary source data, 224, 234, 461–464
selections
 hypergeometric probability
 distribution, 176–177
 non-ordered selections, 110–111
 ordered selections, 91
sensational use of data, 420–421
sensationalism, 421
set
 combinations of items from a set,
 111
 intersection set, 39
 null set, 117
 permutations of some items in a
 set, 78
subset, 117
union set, 39
set notation, 18
set parentheses, 18
SI units, 328, 361
Sicherman, George, 144
Sicherman dice, 144
side-by-side box plots, 299, 425–426
 see also box and whisker plot
sigma, 279
simple compound events, 44–45
simple probabilities, 6–12
 experimental probability, 6–11
 subjective probability, 11–12, 22
simple random sample, 207
simulation
 dice simulation, 29–31
 fish populations using game chips,
 353–354
 normal distribution, investigation of,
 346–347
 real-life events, 442
 technology, use of, 26
 three-coin-flip simulation, 26–29
simulator, 26
skewed data, 256
skewed graph, 165
socioeconomic database, 222, 302
split bar graph, 290, 291, 295
spread, measures of. *See* measures of
 spread
spreadsheet
 bubble plot, 428–429
 clustered column graph, 294
 comparison of two similar
 distributions, 183
 correlation coefficient, 388
 CSV (comma-separated values), 303
 grouped data, 259–260
 measures of central tendency,
 evaluation of, 256
 multiple bar graph, 294
 pivot table, 426–427
 stacked bar graph, 294
 stacked column graph, 294
 three-coin-flip simulation, 27–28
stacked bar graph, 294
stacked column graph, 294
standard deviation, 279, 466
 calculation of, 281–282
 investigation of, 278–279
 population standard deviation, 279,
 283
 sample standard deviation, 279, 283,
 318, 357
values within one standard deviation
 of mean, 348
visualization of spread, 280–281

- statistical summaries
- critical analysis of media claims, 296–297
 - graphical comparison of groups, 294–295
 - interpretation of, 290–298
 - measures of central tendency, interpretation of, 292–293
 - statistical claims in the media, 290–291
- statistician, 440
- Statistics Canada, 222–223, 224–225, 292, 302–305
- Statistics Canada report, 304
- statistics project. *See* culminating statistics project
- statistics project content rubric, 477
- statistics project peer critique form, 479
- statistics project presentation rubric, 478
- statistics project self-assessment
- checklist, 476
 - stem and leaf plot, 31
 - stratified sample, 207
 - studies, 212–214, 219, 220, 243
 - subjective probability, 11–12, 22
 - subset, 117
 - sum of probabilities, 9
 - summary, 100, 250
 - see also* statistical summaries
- survey, 212, 215–216, 217–218, 291, 356–357
- systematic sample, 207
- T**
- table, 338–339
- technology
- see also* Fathom™; graphing calculator; scientific calculator; spreadsheet
 - simulation of probability experiment, 26
 - survey, creation of, 217–218
- tetrahedral dice, 167
- tetrahedron, 158
- theoretical probability, 16
- see also* probability
 - calculation of, 18–20
 - complement, 20–21
 - and experimental probability, comparison of, 26–32
 - odds against, 21
 - odds in favour, 21
 - odds of an event, 22–23
 - outcomes and events, 17
 - probability, use of term, 34, 35
- time graph, 228–229
- time series, 382
- time series graph, 307
- total number of subsets, 117
- treatment group, 213
- tree diagram, 18, 45, 46, 48, 103
- trend on a time graph, 228–229
- trials
- clinical trials, 221
 - combination of real trials, 31
 - dependent trials, 90, 170
 - independent trials, 89, 170
 - repeated independent trials, 72
 - repeated trials without replacement, 72
- trimmed mean, 265
- Triple Your Chances, Double Your Counters, 442–443, 447–449
- two-variable data analysis
- advanced techniques, 424–431
 - bias, 416–417
 - cause and effect, 392–399
 - dynamic analysis, 402–412
 - hidden variable, 411–412
 - line of best fit, 382–389
 - outliers, accounting for, 406–410
 - residual plot, 402–405
 - uses and misuses of data, 416–421
- U**
- uniform distributions, 154–157, 184, 322–323
- union set, 39
- US Census Bureau, 292
- V**
- valid data, 292
- variability (in data), 197
- variability (in samples), 206
- variables
- accidental relationship, 397
 - cause and effect relationship, 394–395
 - common cause relationship, 395–397
 - continuous random variable, 146, 320–327
 - data with more than one variable, 200–202
 - dependent variable, 382, 401
 - discrete random variable, 146, 320–321, 362–369
 - hidden variable, 411–412, 428–430
 - independent variable, 382, 401
 - presumed relationship, 397
 - random variable, 145, 184
 - relationships, 394–398
 - reverse cause and effect relationship, 397
- variance, 279, 281–282, 466
- calculation of, 281–282
 - population variance, 279
 - sample variance, 279
- Venn diagram, 2, 18, 21, 35, 181
- visual representation, 21
- volatility, 289
- voluntary sample, 207
- W**
- weighted mean, 148, 149, 261–262
- Z**
- z*-score, 283–285, 318, 338–339