

**Question 1** (1 point) ✓ *Saved*

What is an example of a ratio measure?

- ☐ Attractiveness
- ☐ Weight
- ☐ Nationality
- ☐ Temperature

**Question 2** (1 point) ✓ *Saved*

What is NOT an example of data born digital?

- ☐ Digital photos from your last vacation
- ☐ Twitter post
- ☐ E-book version of a novel from 19th century
- ☐ Instagram post

**Question 3** (1 point) ✓ *Saved*

What are the examples of cost associated with data collection? (choose all that apply)

- ☐ Time
- ☐ Processing
- ☐ Storage
- ☐ Utility

**Question 4** (1 point) ✓ *Saved*

What is NOT a common pre-processing step in data analysis?

- ☐ Imputation
- ☐ Storage
- ☐ Conversion

**Question 5** (1 point) ✓ *Saved*

What is data imputation?

- ☐ Converting different units of measurement
- ☐ Mapping different objects across datasets
- ☐ Removing erroneous (impossible) data
- ☐ Interpolating/hypothesizing to fill in missing values

**Question 6** (1 point) ✓ *Saved*

What is NOT an informative feature for classifying risky vs. non-risky loans in a loan risk assessment system for individuals?

- ☐ Credit rating
- ☐ Zip code
- ☐ Birth date
- ☐ Income

**Question 7** (1 point) ✓ *Saved*

What are the examples of sensitive data? (choose all that applies)?

Select 3 correct answer(s)

- ☐ Vaccination record
- ☐ Religion
- ☐ Ethnic origin
- ☐ First name

**Question 8** (1 point) ✓ *Saved*

Which one is a good example of a low-level programming language?

- ☐ KNIME
- ☐ Assembly
- ☐ Python
- ☐ Java

**Question 9** (1 point) ✓ *Saved*

Which statement is true about interpreted languages?

- ☐ The source code is executed step by step
- ☐ Interpreted languages are faster to run compared to the compiled languages
- ☐ Programs are compiled to native code during compile time

**Question 10** (1 point) ✓ *Saved*

Which statements are TRUE regarding algorithms?

- ☐ A program can only contain a single algorithms
- ☐ An algorithm describes how to carry out a computational procedure
- ☐ An algorithm needs to be implemented into programs to be executed
- ☐ Algorithms includes rules to execute a process with machines

**Question 11** (1 point) ✓ *Saved*

Computational workflows ...

- ☐ do not include any cycles in their graph
- ☐ require user interaction during execution
- ☐ include links that represents data flow

**Question 12** (1 point) ✓ *Saved*

What features are not present in electronic notebooks?

- ☐ Data preparation steps
- ☐ Large scale processing
- ☐ Easy documentation and sharing
- ☐ Data visualization steps

**Question 13** (1 point) ✓ *Saved*

What are the benefits of using computational workflows?

- ☐ Documenting provenance
- ☐ Computational efficiency
- ☐ Simple programming paradigm
- ☐ Modular assembly

**Question 14** (1 point) ✓ *Saved*

Provenance in data analysis includes ... (choose all that applies)

- ☐ Documenting the input data that was used
- ☐ Documenting what components are used
- ☐ Documenting what values are assigned to parameters
- ☐ Documenting who ran the data analysis last

**Question 15** (1 point) ✓ *Saved*

"Researchers in Frankfurt used the same method as Smith et al (2000) to gather data and were able to confirm Smith's findings that dolphins understand numbers"

The description is an example of:

- ☐ Replication
- ☐ Repeatability
- ☐ Reproduction

**Question 16** (1 point) ✓ *Saved*

"Researchers in Alaska independently confirm that data from Canadian scientists shows that people discriminate against black shelter pets"

is a describing an example of:

- ☐ Replication
- ☐ In-lab replication
- ☐ Repeatability
- ☐ Reproduction

**Question 17** (1 point) ✓ *Saved*

An online vendor divides its users into two groups to evaluate the effect of a new recommender engine compared to an existing recommender in sales. What kind of study is this?

- ☐ Experimental study
- ☐ Observational study
- ☐ Usability study
- ☐ Market research

**Question 18** (1 point) ✓ *Saved*

Which statement is FALSE about observational studies

- ☐ They are performed when an intervention might cause harm
- ☐ They establish causal connections

**Question 19** (1 point) ✓ *Saved*

We are running an observational study to measure the possible effect of parents' level of education, income and body mass index (weight divided by square height) on children's obesity. What is the dependent variable:

- ☐ Parents' body mass index
- ☐ Children's body mass index
- ☐ Parents' education

**Question 20** (1 point) ✓ *Saved*

What are the other names for the dependent variable? (select all that applies)

- ☐ Response variable
- ☐ Outcome variable
- ☐ Explained variable
- ☐ Global variable

**Question 21** (1 point) ✓ *Saved*

What are the other names for the independent variable (select all that applies)?

- ☐ Manipulated variable
- ☐ Exposure variable
- ☐ Control variable
- ☐ Local variable

**Question 22** (1 point) ✓ *Saved*

In an experimental study, the independent variables are:

- ☐ Predicted
- ☐ Manipulated
- ☐ Estimated
- ☐ Measured

**Question 23** (1 point) ✓ *Saved*

Which kind of study can be used to establish causal relationship?

- ☐ Observational study
- ☐ Experimental study
- ☐ Usability study

**Question 24** (1 point) ✓ *Saved*

Mean, median and mode are descriptive statistics for

- ☐ Describing the shape of a distribution
- ☐ Measuring the center of data
- ☐ Measuring the reliability of data
- ☐ Measuring the spread of data

**Question 25** (1 point) ✓ *Saved*



Variables that affect both independent and dependent variables are:

- ☐ Control variables
- ☐ Latent variables
- ☐ Controlled variables
- ☐ Confounding variables

**Question 26** (1 point) ✓ *Saved*

What is NOT a measure of the spread of data?

- ☐ Range
- ☐ Standard deviation
- ☐ Inter-quartile range
- ☐ Skewness

**Question 27** (1 point) ✓ *Saved*

If you have a large number of samples the average converges to:

- ☐ Expected value
- ☐ Range
- ☐ Median
- ☐ Standard deviation

**Question 28** (1 point) ✓ *Saved*

Why is median reported for a measure of typical income in households?

- ☐ Median is less sensitive to asymmetry in distribution
- ☐ Median income is a representative income as it is the income of the household ranked in the middle
- ☐ Median is easier to calculate
- ☐ Average can be misleading if there are a few very wealthy people pushing its value higher

**Question 29** (1 point) ✓ *Saved*

Two random variable A and B are independent if (choose all that applies):

- ☐ Knowing the outcome of one does have any impact on the outcome of the other one
- ☐ Probability of A given B is equal to the probability of B, i.e.,  $P(A|B) = P(B)$
- ☐ The joint probability can be calculated by multiplying the probability of two variables, i.e.,  $P(A,B) = P(A)P(B)$
- ☐ Probability of A given B is equal to the probability of A, i.e.,  $P(A|B) = P(A)$

**Question 30** (1 point) ✓ *Saved*

Probability of drawing a second ace (in two consecutive draws with no replacement) from a stack of cards is:

- ☐ Not independent of the probability of the first card being an ace
- ☐ Does not change depending on the first card being an ace or not

**Question 31** (1 point) ✓ *Saved*

Prior in Bayesian inference ...

- ☐ is the same as posterior.
- ☐ is the probability before any evidence is present

**Question 32** (1 point) ✓ *Saved*

If  $x$  is normally distributed, approximately, what are the odds that  $x$  falls within two standard deviations of its mean? (pick the closest value)

- ☐ 90%
- ☐ 68%
- ☐ 95%
- ☐ 99%

**Question 33** (1 point) ✓ *Saved*

In descriptive statistics (choose all that applies):

- ☐ We discover disjoint clusters in the data
- ☐ We estimate parameters that describe the distribution of the sample
- ☐ We sample from a population to make an inference
- ☐ We summarize the data

**Question 34** (1 point) ✓ *Saved*

What is NOT the p-value in a t-test?

- ☐ Probability of type II error
- ☐ Probability of type I error

**Question 35** (1 point) ✓ *Saved*

Having a small p-value imply a strong effect

- ☐ True
- ☐ False

**Question 36** (1 point) ✓ *Saved*

What is a commonly reported effect size and at what level it is considered a strong effect?

- ☐ Cohen's  $d > 0.8$
- ☐ Cohen's  $\kappa > 0.5$
- ☐ Cohen's  $d > 0.5$
- ☐ Cohen's  $\kappa > 0.8$

**Question 37** (1 point) ✓ *Saved*

We like to see whether a continuous measure is significantly different across three groups. Which statistical test is more suitable?

- ☐ Regression
- ☐ ANOVA
- ☐ Chi-square test
- ☐ t-test

**Question 38** (1 point) ✓ *Saved*

We like to see whether a categorical outcome is significantly different across three groups. Which statistical test is more suitable?

- ☐ ANOVA
- ☐ t-test
- ☐ Chi-square test
- ☐ Regression

**Question 39** (1 point) ✓ *Saved*

We ran an experiment measuring a continuous outcome in two conditions, what statistical test is the most suitable to compare the dependent variables under two conditions?

- ☐ t-test
- ☐ ANOVA
- ☐ Chi-square test
- ☐ Regression

**Question 40** (1 point) ✓ *Saved*

What is type I error in statistical analysis?

- ☐  $\alpha$  error
- ☐ Concluding there is a difference where there is none

**Question 41** (1 point) ✓ *Saved*

What are type II errors in statistical tests?

- ☐ False negative
- ☒ False positive
- ☐ Rejecting a false null hypothesis
- ☐ Failing to reject a false null hypothesis

**Question 42** (1 point) ✓ *Saved*

How can we reduce type I error in statistical analysis? (choose all that applies)

- ☐ Increasing  $\alpha$
  - ☐ Increasing sample size
  - ☐ Reducing the cutoff  $\alpha$
  - ☐ Choosing two tailed t-test rather than one-tailed t-test
-

Submit Quiz

*42 of 42 questions saved*