

Python Fast Track Quiz Day 2

* Erforderlich

* Dieses Formular wird Ihren Namen aufzeichnen. Bitte tragen Sie Ihren Namen ein.

1

What are you supposed to do when getting a new dataset? *

- ☐ Check the correctness of the variables' data types (e.g. numeric, binary, string).
- ☐ Split your dataset into training and testing data.
- ☐ Look through the variables and describe the meaning and the type.
- ☐ Create a correlatin matrix.

2

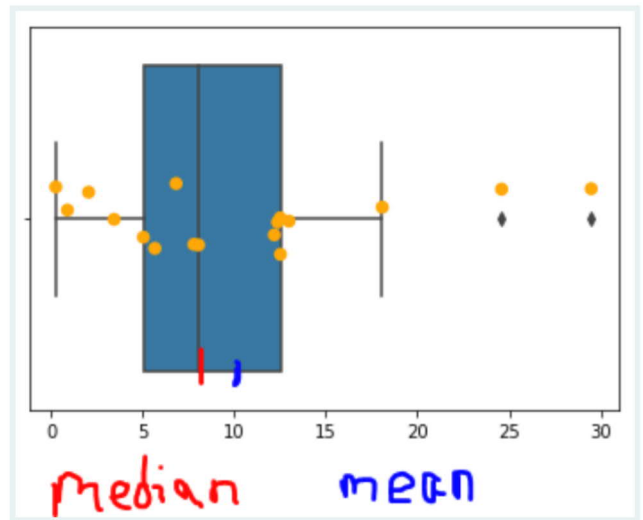
What are best practices for programming in python? *

- ☐ Use separate virtual environments for all your projects.
- ☐ Only import specific parts of a package (e.g. "from sklearn.impute import KNNImputer") instead of the whole thing ("import sklearn").
- ☐ Write all the code you need in one python file.
- ☐ Avoid using nested loops (loop in a loop).
- ☐ In functions, use "return" rather than "print".
- ☐ Write all the code you need in one python function.
- ☐ Insert comments (e.g. "# data import") in your code.

	Actual positive (1)	Actual negative (0)
Predicted positive (1)	5	50
Predicted negative (0)	10	10000

Why can imbalanced data (unequal group sizes) be a problem in machine learning? *

- ☐ There is no problem. The accuracy ($(5+10000)/(5+50+10+10000)$) of 99.4% is great and we can deploy the model.
- ☐ The precision ($(5) / (5+50)$) indicates that only 9.1% of our positively predicted cases are actually positive. Thus our model is actually not really good at not identifying actual negative cases.
- ☐ The recall ($(5) / (5+10)$) indicates that only 33.3% of all of our actually positive cases were identified correctly. Thus our model is not really at identifying actual positive cases.
- ☐ The influence of a majority group on a statistical model is often too large.



Why is it always useful to also provide the median in addition to the mean as a summary statistic? *

- ☐ The median can also be used for categorical variables.
- ☐ The median is generally the more accurate summary statistic compared to the mean.
- ☐ The median is more robust to outliers than the mean.
- ☐ The median provides information about the skewness of your data distribution.

Dieser Inhalt wurde von Microsoft weder erstellt noch gebilligt. Die von Ihnen übermittelten Daten werden an den Formulareigentümer gesendet.

 Microsoft Forms