ECON 2250: Statistics for Economists

Jim Crozier

Fall 2022

Mid term practice 1

Suppose X is a column in a panda's dataframe and $x_1, x_2, ..., x_n$ are the rows of X where n is the number of observations. For reference, the python code for the math we are about to type is.

```
import pandas as pd
import numpy as np
df = pd.DataFrame({'X':[5, 7, 10]})
n_obs = len(df)
mean_x = df["X"].mean()
np.sqrt(sum((df["X"] - mean_x)**2)/(n_obs-1))
```

0.1 jargony jargon

Define the following words:

- scalar
- vector
- dataframe
- sum
- quantile
- median
-

0.2 Mean to Standard Deviation

Use the summation operator ($\sum$) to express the sum of x from 1 to n. I just want the mathematical expression here.

What does the scalar from above represent?

Use the summation operator ($\sum$) to express the average of X. I just want the mathematical expression here.

What does the scalar from above represent?

Refer to the above scalar as *mean_x* and write an equation for subtracting off *mean_x* from element of X. Refer to the elements as x_i, and I only want the equation.

What does the list (vector) from above represent? How many rows will it have?

Write an equation for squaring the above equation.

What does the list (vector) from above represent? How many rows will it have?

Write an equation for summing the squared elements above and dividing by n. This should look kind of similar to taking the average above.

Now take the square of the above equation. What do we call this observation, and what does it represent.

0.3 Quantiles : the sorting hat

Describe how you would find the min and max, and what they represent.

Describe how you find the median? What the core idea behind median and quantiles? What quanitle does the median represent?

What is the difference in finding the median of an even or odd number of observations.

Why do we care about the difference between the mean or median or other quantiles?