

Practical 5

Jumping Rivers

Z-score

1. Write a function that processes data by applying the slide and squish rule to implement the z-score transformation
2. Load the random data from the first practical using `jrpyml.get_numeric_list()` and apply your transformation. Ensure that the mean and variance are indeed as expected.

Means of distributions

The `numpy.random.choice` function can be used to generate samples from an array e.g

1. Write a function that takes 3 arguments
 - Some array of data, `x`
 - A sample size, `n`
 - A replicate count, `r`

This function should find `r` means of sample size `n` from the array `x` and return them

2. Generate a histogram of the number of votes on the movies data
3. Does this distribution look normal?
4. Using your newly written function for calculating means, calculate 1000 averages on samples of size 5000 from the collection of all votes
5. Draw a histogram of these means
6. Does the distribution now look normal?
7. Experiment with different sample sizes, what do you observe

What we are doing here is exploring a piece of mathematics theory known as the central limit theorem which we will explore in the next chapter.