

LING115 Homework #6
Due October 6, 2010

Instructions:

Create a `hw6` directory under your home directory and make sure you save all your work in that directory.

1. [3 points] Copy `hw6_stat.py` under `/home/ling115/hw6_out/` to your `hw6` directory. The file `hw6_stat.py` contains the definition of a function that calculates the average of a list of values. In the same file, define a function called `sd` which takes a list of numbers as its argument and returns the standard deviation. The standard deviation of a list of numbers can be calculated as follows:

- Calculate the arithmetic mean (average) of the list.
- Initialize a variable called `sum_squared_diff` to zero.
- For each number in the list, do the following:
 - Compute the difference between the number and the arithmetic mean.
 - Square the difference.
 - Increase `sum_squared_diff` by the squared difference.
- Divide `sum_squared_diff` by the length of the list. This is the variance.
- Standard deviation is the square root of the variance.

For example, the arithmetic mean of `[1,2,3,4,5]` is 3 and its standard deviation would be

$$\sqrt{\frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5}} = 1.414 \dots$$

To calculate the square root of a number, say `x`, you must first `import math` and do `math.sqrt(x)`.

2. [3 points] Write a Python program named `hw6.py` that prints out the average number of words and its standard deviation in all the files listed under a directory that the user specifies. For example,

```
$ python hw6.py /data/TREEBANK/RAW/WSJ/00/
```

should print out something like the following

```
Average = 403.262626263  
Standard Deviation = 456.11573903
```

Feel free to use the codes specified in the lecture note. I strongly encourage you to use the functions in `hw6_stat.py` – the `avg` function that is already there and the `sd` function you defined in the previous question.