University of California, Los Angeles
Department of Statistics

Statistics 13 **Instructor: Nicolas Christou**

Lab 6: Central limit theorem

Access the following data:

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics13/soil.txt",
header=TRUE)
```

Consider the variable `lead`. Answer the following questions:

a. Construct the histogram of `lead`.

b. Compute the mean and standard deviation of `lead` an assume that these are the population mean and standard deviation.

c. Let $X_1, X_2, \ldots, X_{200}$ be a random sample selected from `lead`. Based on the central limit theorem what is the distribution of the sum of these 200 random variables? How about the distribution of the sample mean of these 200 random variables?

d. Now let's verify the central limit theorem using `R`: Sample 100000 values of `lead` with replacement. Therefore you will get a vector of 100000 values. Collapse this vector into a 500×200 matrix. Each column of this matrix represents a random sample of size $n = 200$ from `lead`. Please attach the `R` code.

e. Compute the sample mean of each column of the matrix in part (d). You will get 200 sample means. Construct the histogram using the 200 sample means. What do you observe? Compare this histogram with your answer to part (c).

f. Compute the range of the distribution of the sample mean based on the central limit theorem using $\bar{x} \pm 2.5 \times \frac{\sigma}{\sqrt{200}}$. The histogram must approximately agree with this range.

g. Repeat part (d) as follows: Sample 2000 values of `lead` with replacement. Collapse this vector into a 5×400 matrix. Each column of this matrix represents a random sample of size $n = 5$ from `lead`. Compute the sample mean of each column of this matrix You will get 400 sample means. Construct the histogram using these 400 sample means. What do you observe?

h. Use similar `R` commands to verify the central limit theorem on another variable of your choice. Submit the `R` code and the histogram.