Tutorial: Confidence and Predictive Intervals in Stata

1. Let X denote a random variable that represents BMI at baseline for the Framingham cohort. Assume that X is normally distributed. What is the mean of X? The standard deviation?

```
. summarize bmi1
```

2. Construct a 95% predictive interval for X. Pick a random observation from the dataset. Does your interval contain the BMI for the randomly selected observation?

   95% predictive interval for X is defined as $\mu \pm 1.96\sigma$.

3. Suppose we now draw repeated samples of size 100 from the Framingham cohort. What is a 95% predictive interval for $\bar{X}$?

   95% predictive interval for $\bar{X}$ is defined as $\mu \pm 1.96\sigma/\sqrt{n}$.

4. Take a sample of size 100 from the Framingham dataset. Does your predictive interval for $\bar{X}$ contain the mean from the 100 person subsample?

```
. sample 100, count
. sum bmi1
```

5. Construct a 95% confidence interval for the mean BMI in this sample. Does the 95% confidence interval contain the mean BMI for the entire cohort?

   A 95% CI for μ is defined as $\bar{X} \pm 1.96\sigma/\sqrt{n}$.