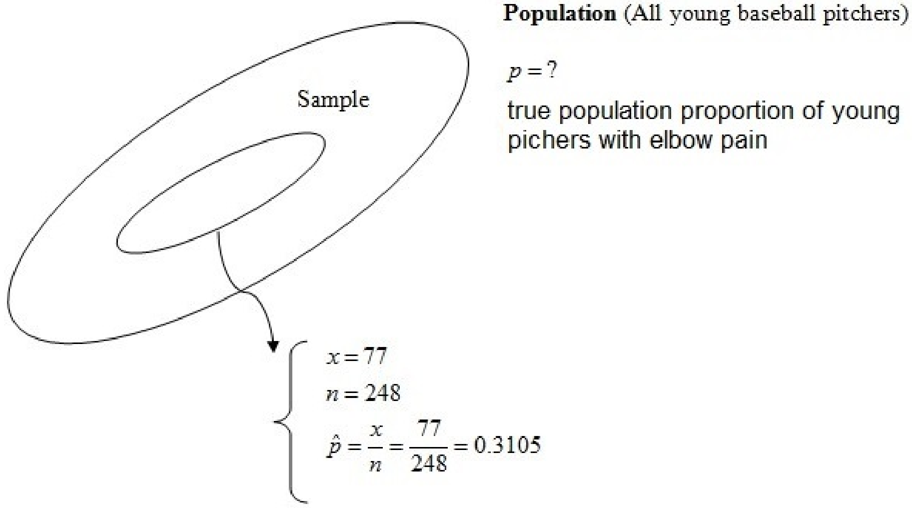
University of New Mexico

CONFIDENCE INTERVALS (Fall 2016)

PH 538: Public Health Biostatistical Methods I [Instructor: Dr. Fares Qeadan]

1.1 C.I. AROUND ONE POPULATION PROPORTION (p) [using sample statistics]:

Example: In a 1999-2000 longitudinal study of youth baseball, researchers found that 77 of 248 young pitchers complained of elbow pain after pitching. What is the 90% C.I. for the true population proportion of young pitchers with elbow pain?



Solution by hand: Since the CL is 90% then $Z_{\alpha/2} = di \quad invnormal((1-0.90)/2) \approx -1.645$. Thus,

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.3105 \pm 1.645\sqrt{\frac{0.3105(1-0.3105)}{248}} = (0.2621676, 0.3588324) \approx (0.2622, 0.3588).$$

We are 90% confident that the true population proportion of young pitchers with elbow pain is between 26.22% and 35.88%.

Solution by STATA:

```
. cii proportions 248 77, level(90) wald

                                                   — Binomial Wald —
    Variable |        Obs  Proportion    Std. Err.     [90% Conf. Interval]
-------------+-------------------------------------------------------------
             |        248    .3104839     .029381       .2621565    .3588112
```

1.2 C.I. AROUND ONE POPULATION PROPORTION (p) [using sample data]:

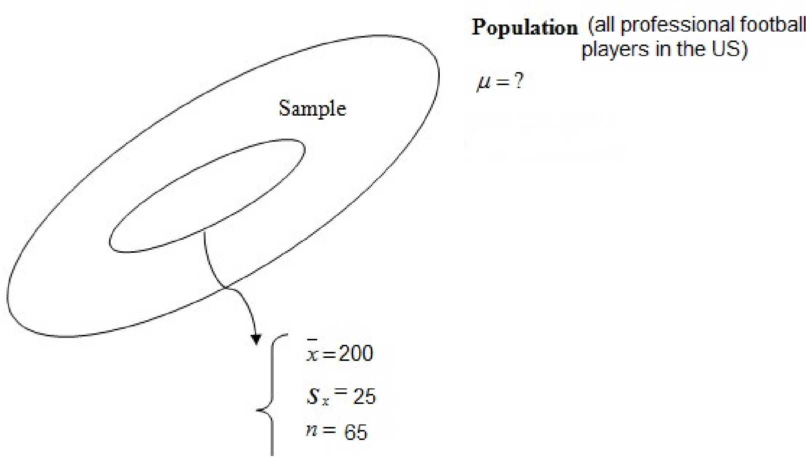
Example: Using the Diabetes dataset from Lab1, what is the 95% confidence interval around the true population proportion of diabetes?

Solution by STATA:

```
. use "C:\Users\Fares\Documents\PH538\Fall2016\handouts\diabetesfall16.dta"

. ci proportions diab, level(95) wald

                                                       — Binomial Wald —
    Variable |        Obs  Proportion    Std. Err.       [95% Conf. Interval]
-------------+-----------------------------------------------------------------
        diab |        390    .1538462     .0182699        .1180378    .1896545
```

We are 90% confident that the true population proportion of diabetes is between 11.80% and 18.96%.

2.1 C.I. AROUND ONE POPULATION MEAN (μ) when σ is unknown [using sample statistics]:
Example: A study from the US of 65 professional football players found that the weight of football players is roughly normally distributed with a sample mean of 200 pounds and a standard deviation of 25 pounds. What is the 99% C.I. for the true population mean weight of US football players?



Solution by hand: Since the CL is 99% then $t_{n-1,\alpha/2} = di \quad invt(64,(1-0.99)/2) \approx -2.655$. Thus,

$\bar{x} \pm t_{n-1,\alpha/2}\left(\frac{s_x}{\sqrt{n}}\right) = 200 \pm 2.655\left(\frac{25}{\sqrt{65}}\right) = (191.7672, 208.2328) \approx (191.77, 208.23)$.
We are 99% confident that the true population mean weight of US professional football players is between 191.77 and 208.23 pounds.

Solution by STATA:

```
. cii means 65 200 25, level(99)

    Variable |        Obs        Mean    Std. Err.       [99% Conf. Interval]
-------------+---------------------------------------------------------------
             |         65         200    3.100868        191.7676    208.2324
```

2.2 C.I. AROUND ONE POPULATION MEAN (μ) when σ is unknown [using sample data]:

Example: Using the Diabetes dataset from Lab1, what is the 95% confidence interval around the true mean age of subjects in the population of African Americans in Virginia?
Solution by STATA:

```
. use "C:\Users\Fares\Documents\PH538\Fall2016\handouts\diabetesfall16.dta"

. ci means age, level(95)

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
         age |        403    46.85112    .8125752        45.25369    48.44854
```

We are 95% confident that the true mean age of subjects in the population of African Americans in Virginia is between 45.25 and 48.45 pounds.

2.3 C.I. AROUND ONE POPULATION MEAN (μ) when σ is unknown by stratum (group) [using sample data]:

Example: Using the Diabetes dataset from Lab1, what is the 95% confidence interval around the true mean age of Female subjects in the population of African Americans in Virginia?
Solution by STATA:

Method 1: we could use the ”if” statement:

```
. use "C:\Users\Fares\Documents\PH538\Fall2016\handouts\diabetesfall16.dta"

. ci means age if gender=="female", level(95)

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
         age |        234    45.83333    1.085489        43.69471    47.97196
```

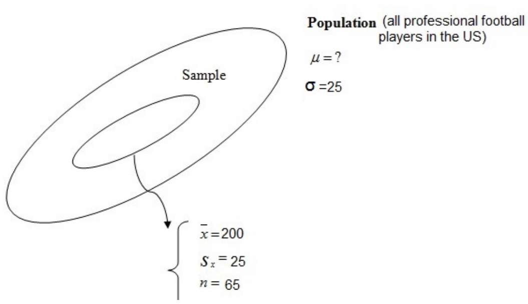
We are 95% confident that the true mean age of Females in the population of African Americans in Virginia is between 43.69 and 47.97 pounds.

Method 2: we could use the ”by” statement after sorting the data by the stratum/group:

```
. use "C:\Users\Fares\Documents\PH538\Fall2016\handouts\diabetesfall16.dta"

. sort gender

. by gender: ci means age, level(95)

--------------------------------------------------------------------------
-> gender = female

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
         age |        234    45.83333    1.085489        43.69471    47.97196

--------------------------------------------------------------------------
-> gender = male

    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
         age |        169    48.26036     1.21841        45.85499    50.66572
```

3.1 C.I. AROUND ONE POPULATION MEAN (μ) when σ is known [using sample statistics]:

Example: A study from the US of 65 professional football players found that the weight of football players is roughly normally distributed with a sample mean of 200 pounds. Assume that the population standard deviation for the weights is known and equals to 25 pounds. What is the 99% C.I. for the true population mean weight of US football players?



Solution by hand: Since the CL is 99% then $Z_{\alpha/2} = di \quad invnormal((1-0.99)/2) \approx 2.576$. Thus,

$$\bar{x} \pm Z_{\alpha/2}\left(\frac{\sigma_x}{\sqrt{n}}\right) = 200 \pm 2.576\left(\frac{25}{\sqrt{65}}\right) = (192.0122, 207.9878) \approx (192.01, 207.99).$$

Solution by STATA:

```
. ztesti 65 200 25 0, level(99)

One-sample z test
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [99% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      65         200    3.100868          25    192.0127    207.9873
------------------------------------------------------------------------------
    mean = mean(x)                                                z =  64.4981
Ho: mean = 0

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(Z < z) = 1.0000         Pr(|Z| > |z|) = 0.0000          Pr(Z > z) = 0.0000
```

We are 99% confident that the true population mean weight of US professional football players is between 192.01 and 207.99 pounds.