CONFIDENCE INTERVALS FOR POPULATION MEAN

- $x_1, \ldots, x_n$ is an i.i.d. sample selected from population random variable X with unknown mean μ and variance σ^2.

- $\bar{x} = \sum_{i=1}^{n} x_i/n$ = sample mean (an unbiased estimate of μ)

- $s^2 = \dfrac{\sum_i (x_i - \bar{x})^2}{n-1} = \dfrac{\sum_i x_i^2 - (\sum_i x_i)^2/n}{n-1}$
  = sample variance (an unbiased estimate of σ^2)

<u>100(1-α)% Large Sample Confidence Interval for μ</u>

$$\boxed{\bar{x} \pm \sqrt{s^2/n}\; z_{\alpha/2}}$$

Notes:
1. X can have any distribution.
2. n $\geq$ 30 required for Central Limit Theorem validity.

Special Case: X ~ Bernoulli(p)

$\bar{p} = \bar{x}$, $\bar{q} = 1-\bar{x}$, and $s^2 = \bar{p}\bar{q}\left(\dfrac{n}{n-1}\right)$

Now the confidence interval is

$$\boxed{\bar{p} \pm \sqrt{\bar{p}\bar{q}/(n-1)}\; z_{\alpha/2}}$$

Example: A single die is thrown 1000 times resulting in 165 "ones". Based on these data, the 95% confidence interval for the probability of receiving a "one" in a single toss is
$.165 \pm \sqrt{(.165*.835)/999}\;\; z_{.025}) = .165 \pm .0117*1.96 =$
$[.1420, .1880]$

100(1-α)% Small Sample Confidence Interval for μ

$$\boxed{\bar{x} \pm \sqrt{s^2/n}\ t_{n-1,\alpha/2}}$$

Notes:
1. $n \geq 2$.
2. X must have a Normal distribution. The mean and variance are unknown.

More examples:

1. 1000 replications of a Monte Carlo search simulation are performed yielding vector $x = x_1,...,x_{1000}$. Each x_i is either a 1 (detection occurred) or 0 (detection did not occur). Using MATLAB notation, the best estimate of Pd is mean(x). And the approximate 95% confidence interval is

mean(x) ± std(x)/sqrt(1000) * 1.96

= mean(x) ± sqrt(mean(x)*(1-mean(x))/999) * 1.96,

since each x_i is the realization of a Bernoulli random variable.

2. Now 50 simulation experiments like 1. above are conducted yielding a vector $P = (Pd_1, ..., Pd_{50})$. Now our best estimate of Pd is mean(P), and a good 95% confidence interval for this estimate is

mean(P) ± std(P)/sqrt(50) * 1.96

An alternate approximate 95% confidence interval is obtained by considering all 1000*50 replications as independent Bernoulli trials. This gives,

mean(z) ± std(z)/sqrt(50,000) * 1.96

= mean(z) ± sqrt(mean(z)*(1-mean(z))/49,999) * 1.96

where $z = x_1, ..., x_{50,000}$ is a (0,1) (i.e., Bernoulli) vector.