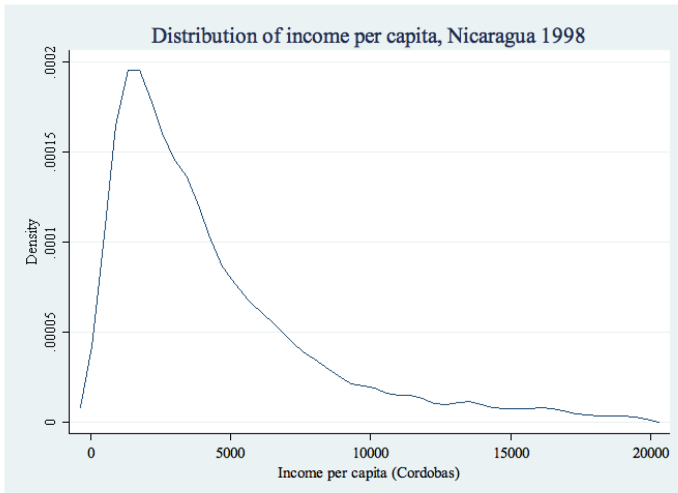
Estimation of the mean of a population and construction of a confidence interval

Population = 22339 persons from Nicaragua with average income/capita μ ("true mean"), and true standard deviation σ. [True means and standard errors are usually unknown, as they could only be known if we had a census of the whole population of households, with their income/capita. In our case, $\mu = 5507$ córdobas, and $\sigma = 10771$ córdobas; in 1998 US$1~ 10 córdobas].

```
label var i_income "Income per capita (Cordobas)"
kdensity i_income if i_income<20000, title ("Distribution of income per capita,
Nicaragua 1998")
```



1. Define an estimator $\bar{x}$ for the population mean μ and an estimator s^2 for the population variance σ^2

One sample of 2,234 persons:

Compute sample mean and sample variance of income/capita

```
. sum income
  Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    income |      2234     5396.51    8011.15          0   168552.9
```

$$\bar{x} = \frac{\sum x_i}{n} = 5397 \text{ and } s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = 8011^2 * \frac{2234}{2233}$$

2. Property of the estimator $\bar{x}$

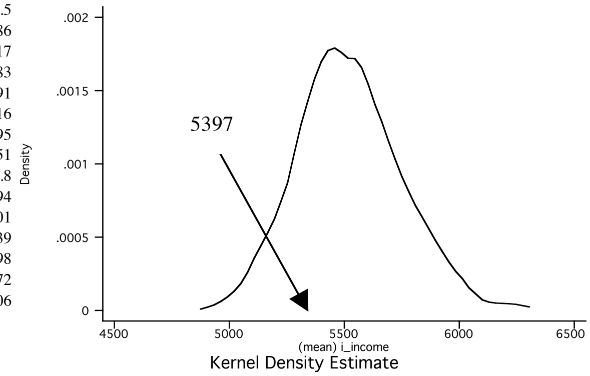
It is unbiased, i.e., $E(\bar{x}) = \mu$, and its variance is $\text{var}(\bar{x}) = \frac{1}{n}\text{var}(x) = \frac{1}{n}\sigma^2$ or standard deviation:

$$\text{sd}(\bar{x}) = \frac{1}{\sqrt{n}}\sigma$$

From the sample, we can compute one value, $\bar{x} = 5397$, and one estimation for its variance: $\widehat{\text{var}(\bar{x})} = \frac{1}{n}s^2 = 28727$ or $\widehat{\text{sd}(\bar{x})} = 169$

Check numerically:
Repeat 512 times: Here are some values for $\bar{x}$:

5546.19	5524.84	5608.63	5656.64	5009.88	5723.81	5281.5
5471.63	5598.41	5349.31	5462.35	5515.2	5677.64	5995.86
5503.55	5160.92	5976.73	5491.89	5641.75	5484.19	5460.17
5703.7	5384.34	5384.08	5976.77	5611.49	5272.66	5178.83
5280.9	5324.98	5394.74	5759.59	5558.9	5622.06	5548.91
5707.15	5679.66	5133.75	5717.11	5584.75	5509.09	5559.16
5157.59	5846.83	5118.19	5820.1	5653.99	5130.67	5377.95
5686.97	5493.53	5810.17	5265.17	5495.76	5442.15	5665.51
5490.38	5498.91	5997.88	5150.2	5578.23	5426.48	5669.8
5061.16	5862.29	5446.75	5484.8	5971.43	5173.89	5364.94
5396.54	5963.28	5508.36	5901.36	5383.45	5299.95	5555.01
5851.09	5914.9	5308.24	5759.63	5571.92	5160.21	5602.39
5749.47	5387.94	5474.79	5524.78	5553.74	5377.61	5295.98
5514.25	5402.17	5501.91	5646	5377.47	5533.67	5512.72
5433.35	5507.5	5398.02	5117.08	5917.58	5493.15	5497.06



It looks like a Normal distribution.

Distribution of the 512 computed $\bar{x}$:

```
. sum minc
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        minc|       512     5520.92      217.61     4992.86    6259.81
```

If we had an infinite number of repetition, rather than 512, we would have found the true values:

$$E(\bar{x}) = \mu = (5507) \text{ Cordobas and } \operatorname{var}(\bar{x}) = \frac{\operatorname{var}(x)}{2234} = \left(\frac{10771^2}{2234} = 51931\right) \text{ or } \operatorname{sd}(\bar{x}) = (228)$$

3. Reminder on the Normal distribution: