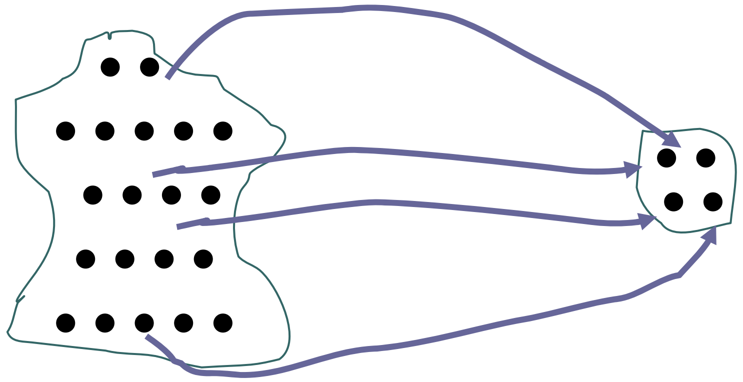
Statistics
Sample

Population:

Sample: independent identically distributed (iid)



Examples:
- production
- marketing research

Sample function: a function of the observed values in the sample used for making general conclusion about the entire population.

Sample
Mean, mode & median

Sample mean / average:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Mode:

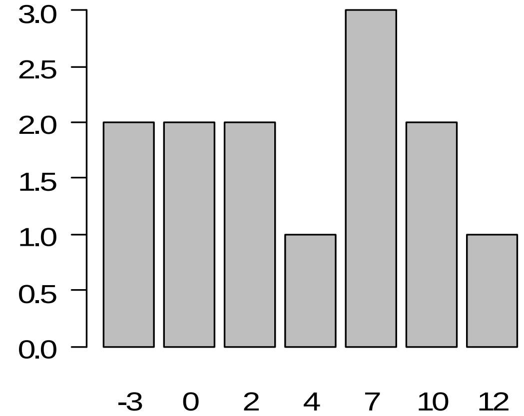
the value(s) with highest frequency

Median: $X_{(i)}$ increasing sequence

$$\widetilde{X} = \begin{cases} X_{(n+1)/2} & n \text{ odd} \\ \dfrac{X_{n/2} + X_{n/2+1}}{2} & n \text{ even} \end{cases}$$

Example: Grade sample:

4, 12, 7, 2, 10, 7, -3, 0, 0, -3, 2, 7, 10



Mean: $\overline{x} = 55/13 = 4.23$

Mode: $m = 7$

Median: $\widetilde{x} = 4$

Sample
Range & variance

Range:

$$R = X_{(n)} - X_{(1)}$$

Sample variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Sample standard deviation:

$$S = \sqrt{S^2}$$

Example cont.:

Range: $r = 15$

Variance: $s^2 = \frac{1}{13-1}\sum_{i=1}^{15}(X_i - 4.23)^2 = 25.03$

Standard div: $s = \sqrt{25.03} = 5.00$

Notice!! Lower case letters: Based on observations
Upper case letters: Based on random variables

Sample mean
Normal distribution

Theorem:
Let X_1, X_2, ... X_n be independent normal distributed random variables with same mean μ and same finite variance σ^2, that is

$$X_i \sim N(\mu, \sigma^2),\ i = 1, 2, \ldots, n \quad \text{iid.}$$

It follows that

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and then

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Sample mean

Distribution

The Central Limit Theorem (CLT):

Let $X_1, X_2, \ldots X_n$ be independent identically distributed random variables with same mean μ and same finite variance σ^2. Then the distribution of

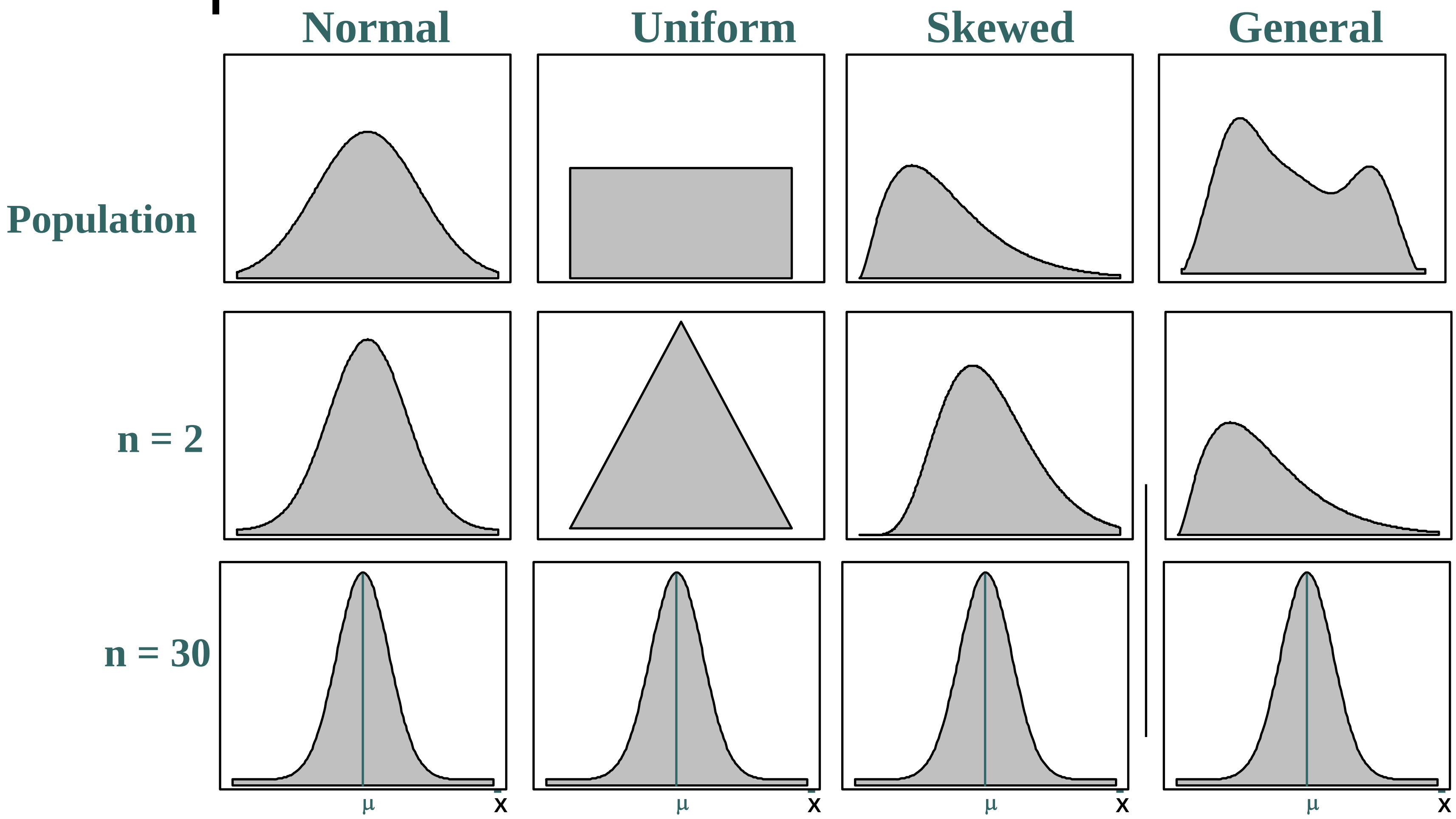
$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

will tend towards the standard normal distribution as $n \to \infty$.

How large should n be before the approximation is good?

- Most distributions: $n \geq 30$
- Normal distribution : for all n

Examples



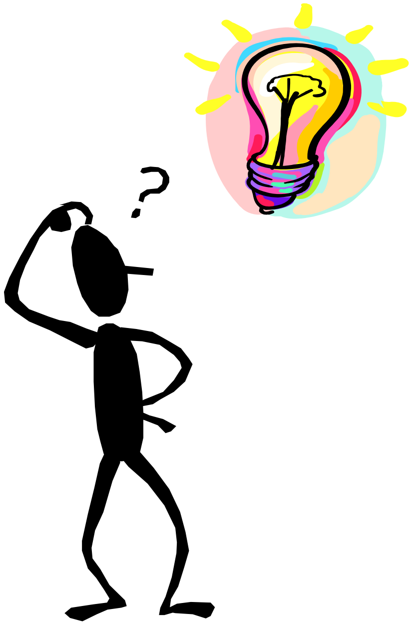
Sample mean
Example

Problem: Production of light bulbs

A company produces bulbs with a life time X, which is approximately normal distributed with

mean: $\mu = 800$ hours

standard deviation: $\sigma = 40$ hours



(a) Find the probability that a sample consisting of 16 bulbs has a mean life time less than 775 hours?

(b) If you observe a sample mean life time of 775 hours, would you believe that the population mean is in fact 800 hours?

Two sample means
Comparison

Theorem:
Assume two independent samples are taken from two populations with means μ_1 and μ_2, respectively, and finite variances σ_1^2 and σ_2^2, respectively.

Then for the difference between the two sample means $\bar{X}_1 - \bar{X}_2$, we have

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

and hence

$$\frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Sample variance
Distribution

Theorem:
Lad X_1, X_2, ... X_n be independent normal distributed random variables with mean μ and variance σ^2.

Then
$$\frac{(n-1)\,S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1)$$

When calculating S^2, it's usually more convenient to use
$$S^2 = \frac{1}{n\,(n-1)}\left[n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \right]$$

Sample variance
Example

Problem: Car batteries

A producer of car batteries claims that the life time of their batteries are normal distributed with

mean: $\mu = 3$ year
standard deviation: $\sigma = 1$ year

Sample of 5 batteries: 1.9 2.4 3.0 3.5 4.2

(a) Calculate sample standard deviation.

(b) Do you believe that the standard deviation is 1 year?

Sample mean
Distribution (unknown variance)

Typically the variance σ^2 is unknown.
If we replace the unknown variance by s^2 we obtain:

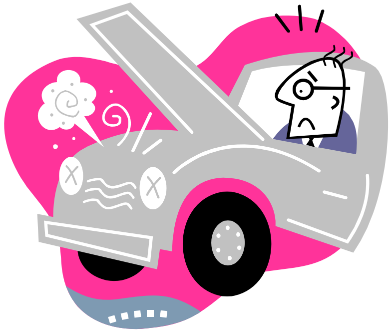
Theorem:
Lad $X_1, X_2, \ldots X_n$ be independent normal distributed random variables with mean μ and variance σ^2 (unknown).

Then $$\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$$

Sample mean
Example (unknown variance)

Problem cont.: Car batteries
Producer claims that the life time of their batteries are normal distributed with

mean: μ = 3 years
standard deviation: unknown



Sample of 5 batteries: 1.9 2.4 3.0 3.5 4.2

Do you believe that mean life time is 3 years?

Two sample variances
Comparison

Theorem:
If two independent samples are taken from two normal populations with variances σ_1^2 and σ_2^2, respectively, then

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Notice!! $f_{1-\alpha}(n_1, n_2) = \dfrac{1}{f_\alpha(n_2, n_1)}$

Eg. $f_{0.95}(6, 10) = \dfrac{1}{f_{0.05}(10, 6)}$