A statistic
(STATS201.stats 201 20: Probability and probability distributions)

Anna Helga Jónsdóttir
Sigrún Helga Lund

December 15, 2012

Statistics

Statistic

A **statistic** is a number that is calculated by some method from our data.

- We look at our measurements as random variables because the outcome can change each time the experiment is repeated.
- Statistics are calculated from our measurements.
- If the outcomes change, the statistics can also change!
- That means that statistics are in fact random variables!

Sampling distribution

Sampling distribution

Every statistic is a random variable and has therefore some probability distribution. That distribution is called the **sampling distribution** of the statistic.

The sampling distribution depends on

- The **probability distribution of the measurements** that the statistic is calculated for.
- The **number of measurements**.

When certain criteria are fulfilled the sampling distribution of some statistics follow certain known types. Statistical inference normally relies on that fact.

Example

Let X_1 and X_2 be random variables that describe the outcome when a dice is thrown.

Below are shown all possible outcomes of the statistic $X_1 + X_2$.

					(6,1)					
				(5,1)	(5,2)	(6,2)				
			(4,1)	(4,2)	(4,3)	(5,3)	(6,3)			
		(3,1)	(3,2)	(3,3)	(3,4)	(4,4)	(5,4)	(6,4)		
	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)	
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
2	3	4	5	6	7	8	9	10	11	12

Example

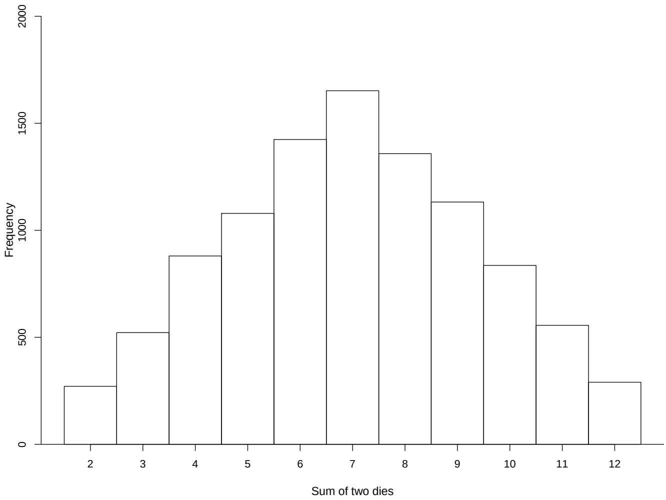

Figure: Simulation of 10000 throws of two dice.

Expected value of the sum of two random variables

Formulas for the expected value of random variables

If X and Y are two random variables, then

$$
\begin{aligned}
E[X+Y] &= E[X]+E[Y] \\
E[X-Y] &= E[X]-E[Y]
\end{aligned}
$$

Variance of the sum of two random variables

Formulas for the variance of random variables

If X and Y are two independent random variables then

$$Var[X+Y] = Var[X] + Var[Y]$$
$$Var[X-Y] = Var[X] + Var[Y]$$

Expected value and variance of the mean

Expected value and variance of the mean

If $X_1, \ldots, X_n$ are independent and identically distributed random variables with expected value $E[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$, then the following holds for the mean of them, denoted $\bar{X}$:

$$E[\bar{X}] = \mu$$

$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

Standard error

Standard error

If $\bar{X}$ is the mean of $X_1, \ldots, X_n$, independent and identically distributed random variables with variance $Var[X_i] = \sigma^2$, then their **standard error** is

$$\sigma/\sqrt{n}$$

It is the standard deviation of the mean of the measurements.

The probability distribution of the mean of normally distributed random variables

The probability distribution of the mean of normally distributed random variables

If $X_1, \ldots, X_n$ are normally distributed random variables with mean μ and variance σ^2, then $\bar{X}$ follows also a normal distribution, with mean μ and variance σ^2/n.

$$\text{That is if } \; X_i \sim N(\mu, \sigma^2) \; \text{ then } \; \bar{X} \sim N(\mu, \sigma^2/n).$$

Central limit theorem

Central limit theorem

If $X_1, \ldots, X_n$ are independent and identically distributed variables then $\bar{X}$ follows normal distribution with mean μ and variance σ^2/n

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

if n is large enaugh.

Notice that we do not need to know the probability distribution of the measurements!

Central limit theorem

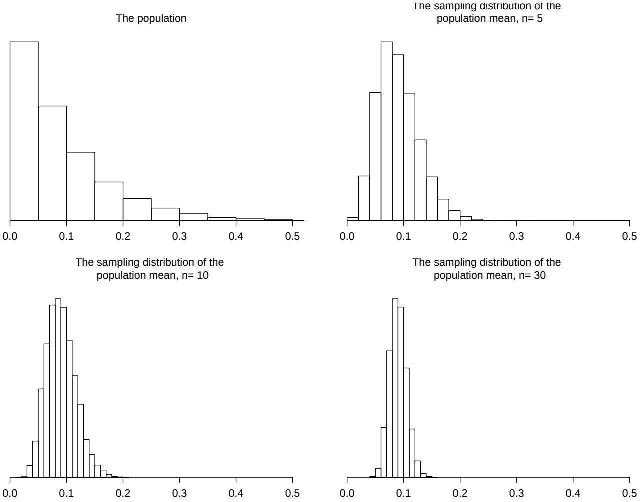

Figure: The sampling distribution of a mean when the random variables follow a very skewed distribution.

Central limit theorem

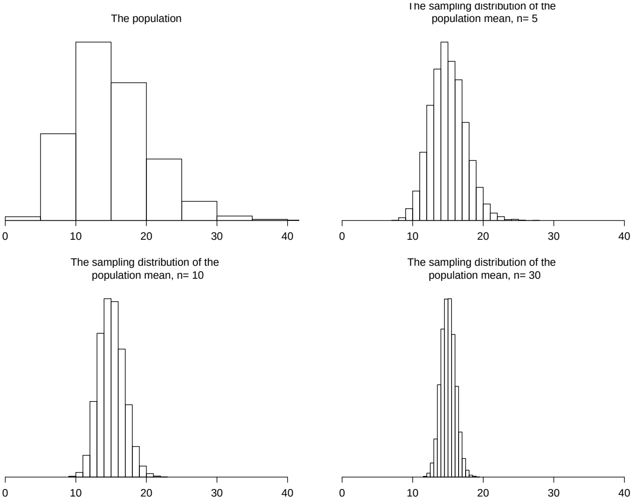

Figure: The sampling distribution of a mean when the random variables follow a slightly skewed distribution.

Estimators and test statistics

There are two groups of important statistics.

- **Estimators** estimate the parameters of the probability distribution that the random variables follow.
  Example: Estimator that estimatesμ when the measurements are normally distribution.
  Example: Estimator that estimates p when the measurements are binomially distributed.
- **Test statistics** allow us to make statistical inference.
  Example: Test statistic that allows us to infer whether the variance of two population is the same.
  Example: Test statistic that allows us to infer whether the mean of a population differs from 20.