BUEC-333
Statistical analysis of economic data

Summer 2015

Table of contents

Normality, and the CLT

- I will link **random sampling** to **normality** using the **Central Limit Theorem (CLT)**
- How do you get an RV with a normal distribution?
  1. The RV simply **is** normal
  2. It is a **linear** combination of Normal RVs
     - Average
     - Sum
  3. It is a sum/average of **many** random variables

Central Limit Theorem

The **CLT** states:

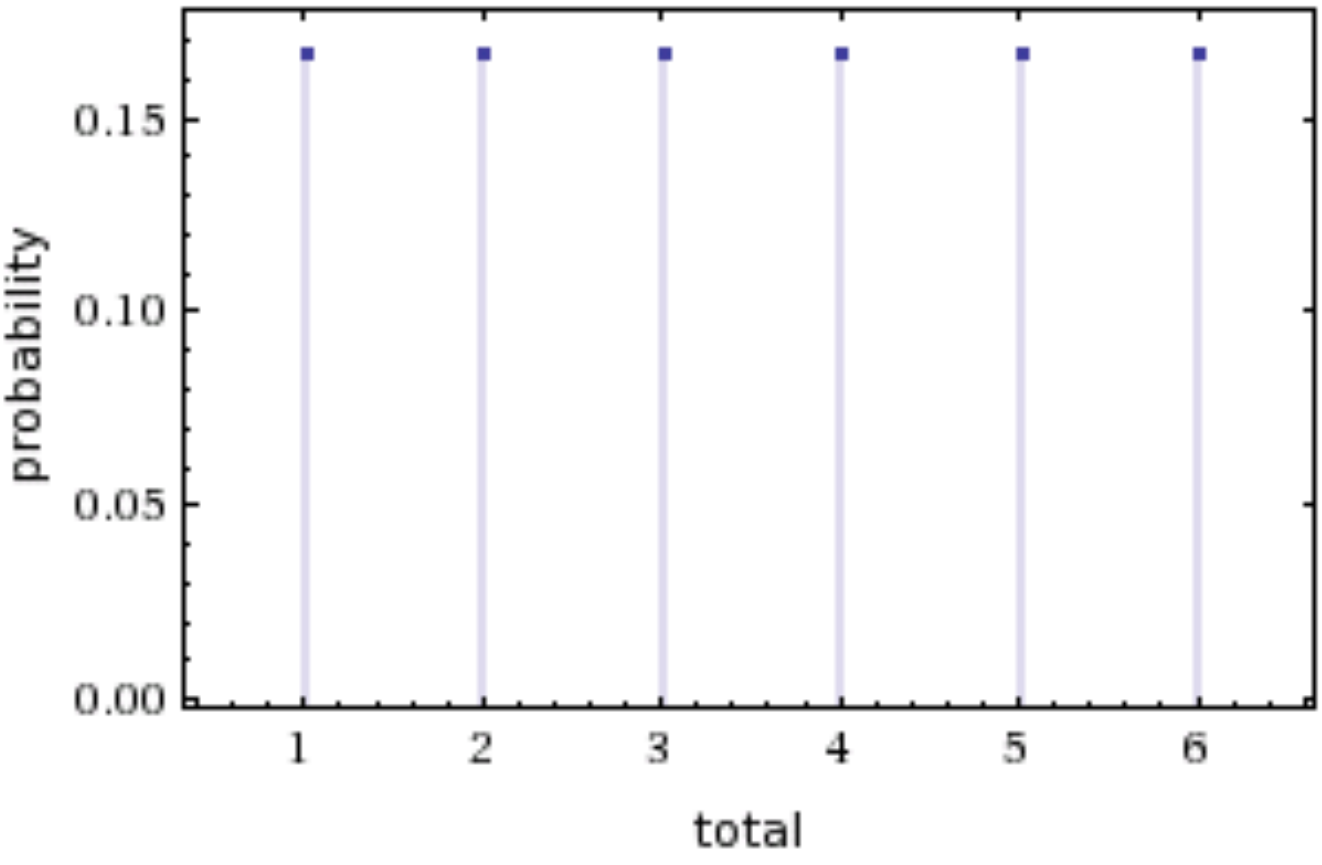
"The **mean** of a number of independent, identically distributed random variables will tend to be **normally** distributed, regardless of their distribution, if the number of different random variables is **large** enough."

Main example: sample mean $\bar{X}$

Example: die

- 6 possible **outcomes**: {1,2,3,4,5,6}
- Each happens with **probability** 1/6
- So, the **probability distribution** looks like:

Probability distribution for one throw



(assuming fair 6-sided dice)

Simulations

- Use the computer to “throw” many times (N)
- Look at the sampling distribution of the average
- Look at what happens as N increases

Changing N

- As we **increase** the number of die throws N ...
- Sampling distribution gets closer to a **normal**: CLT
- Variance decreases: if $\text{Var}(X) = \sigma^2$, then

$$\text{Var}(\bar{X}) = \sigma^2/N$$

Histogram

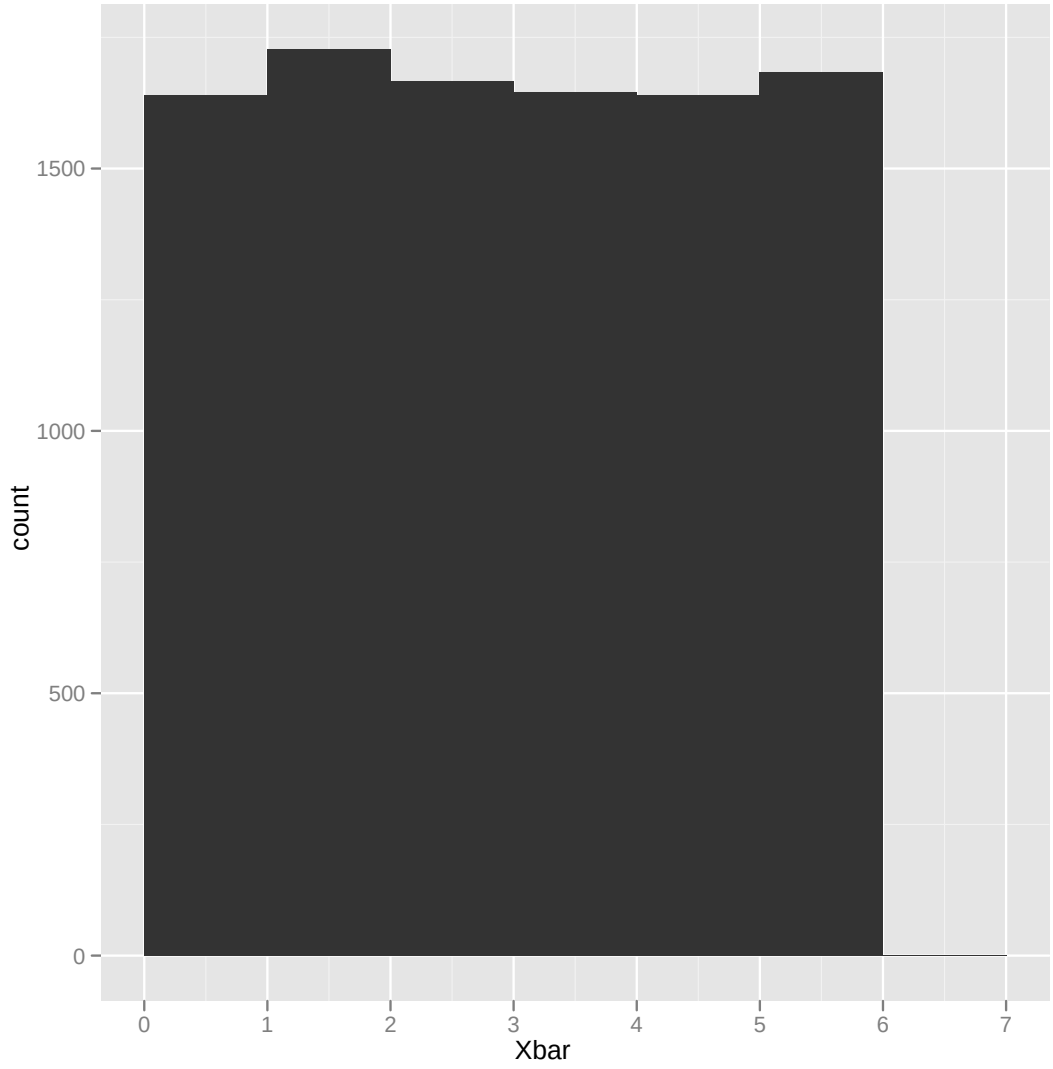


Figure: $N = 1$

Histogram

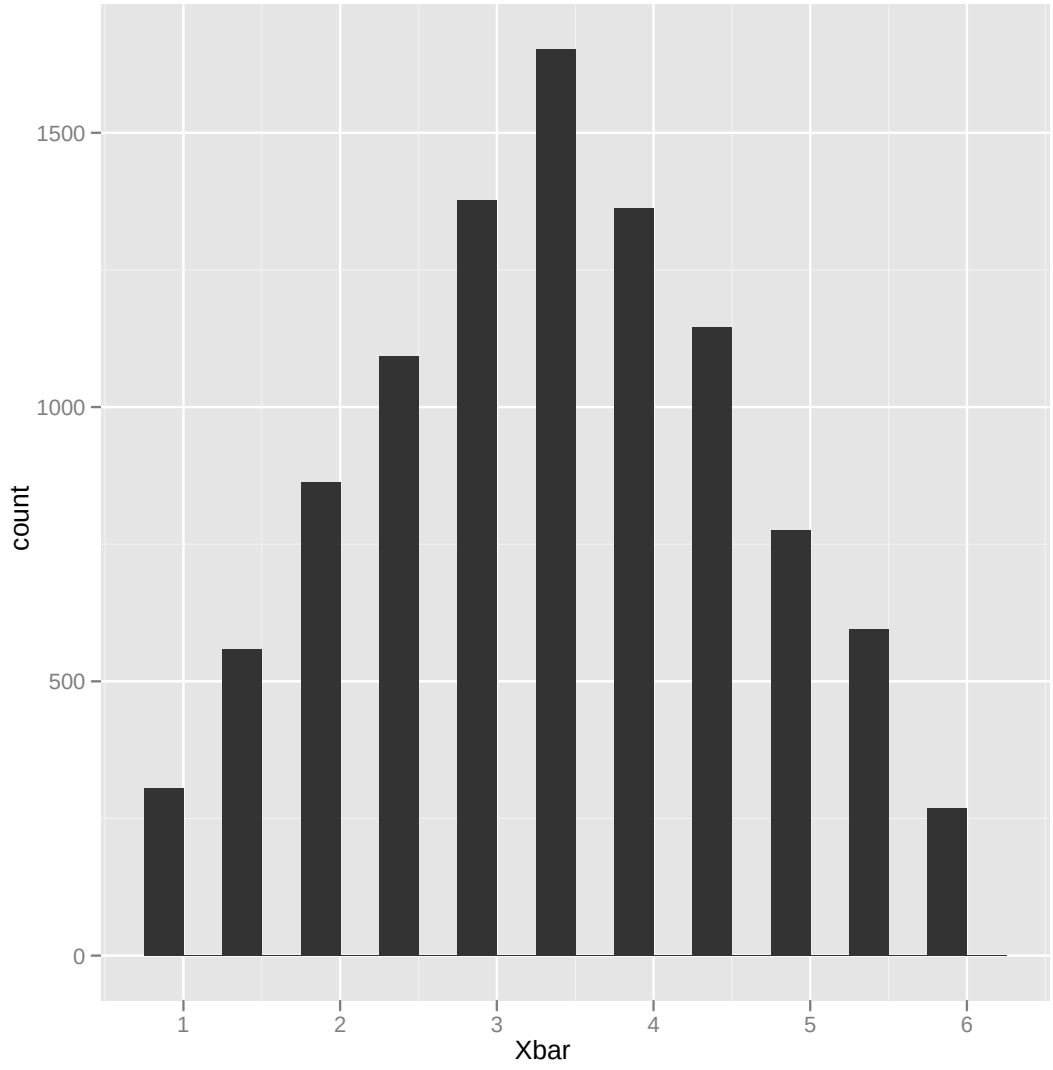


Figure: $N = 2$

Histogram

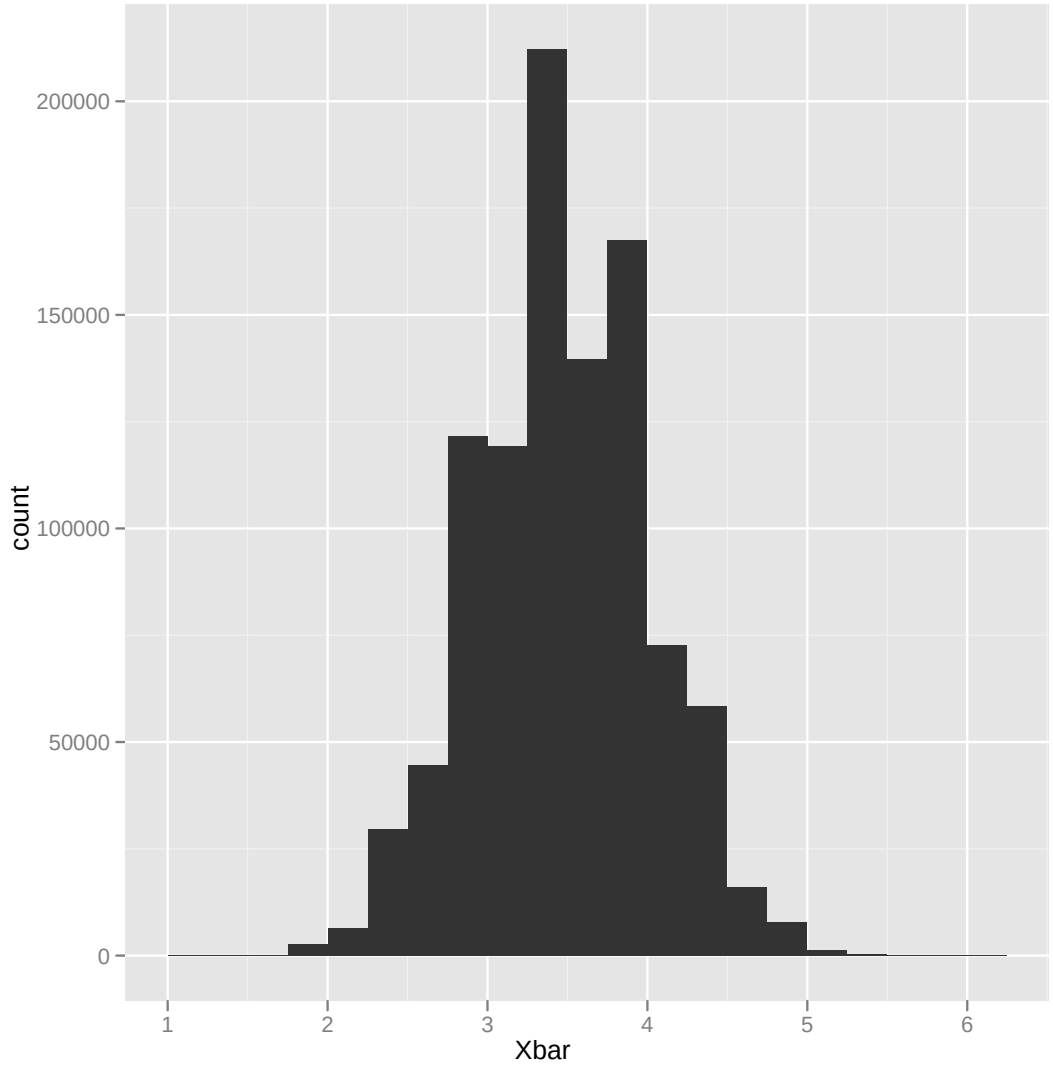


Figure: $N = 10$

Histogram

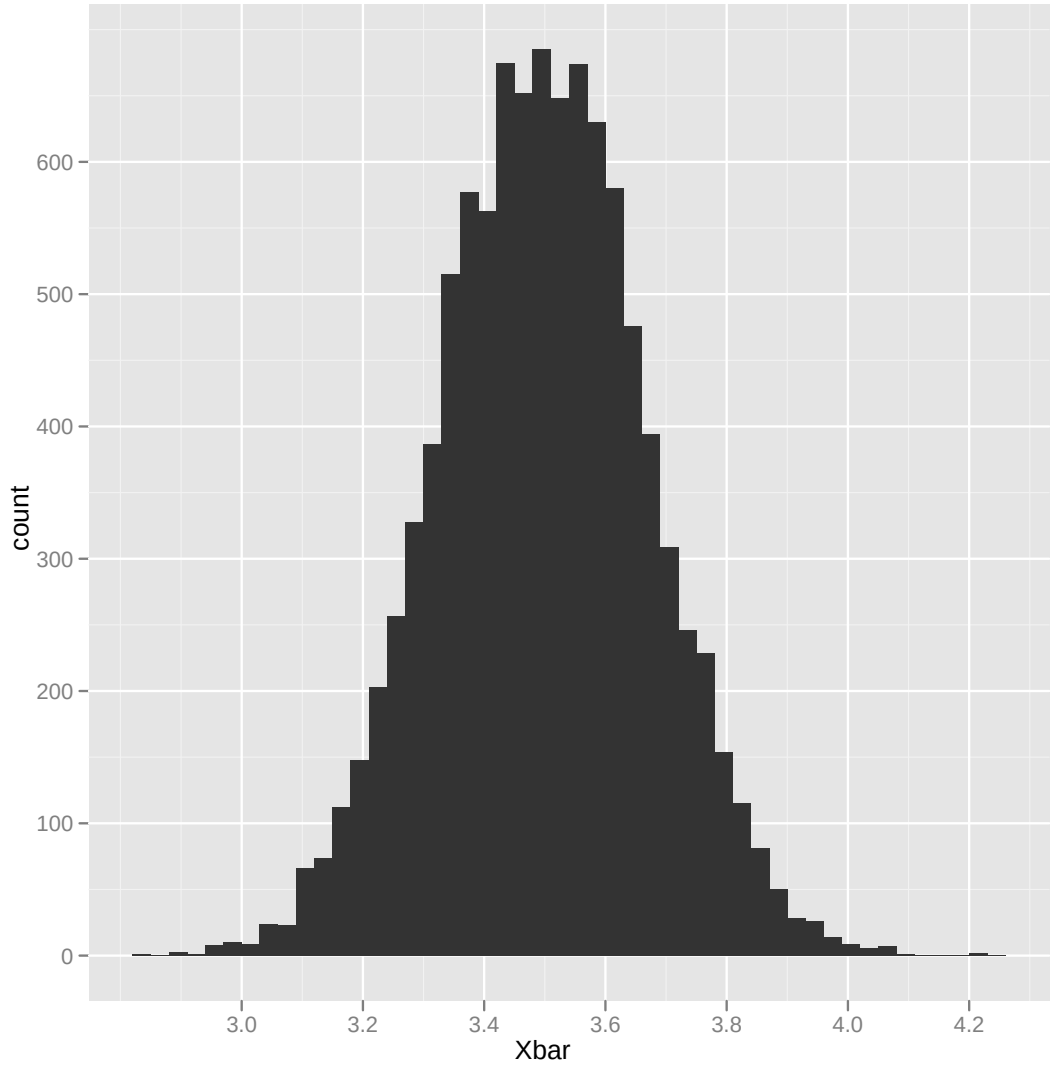


Figure: $N = 100$

Histogram

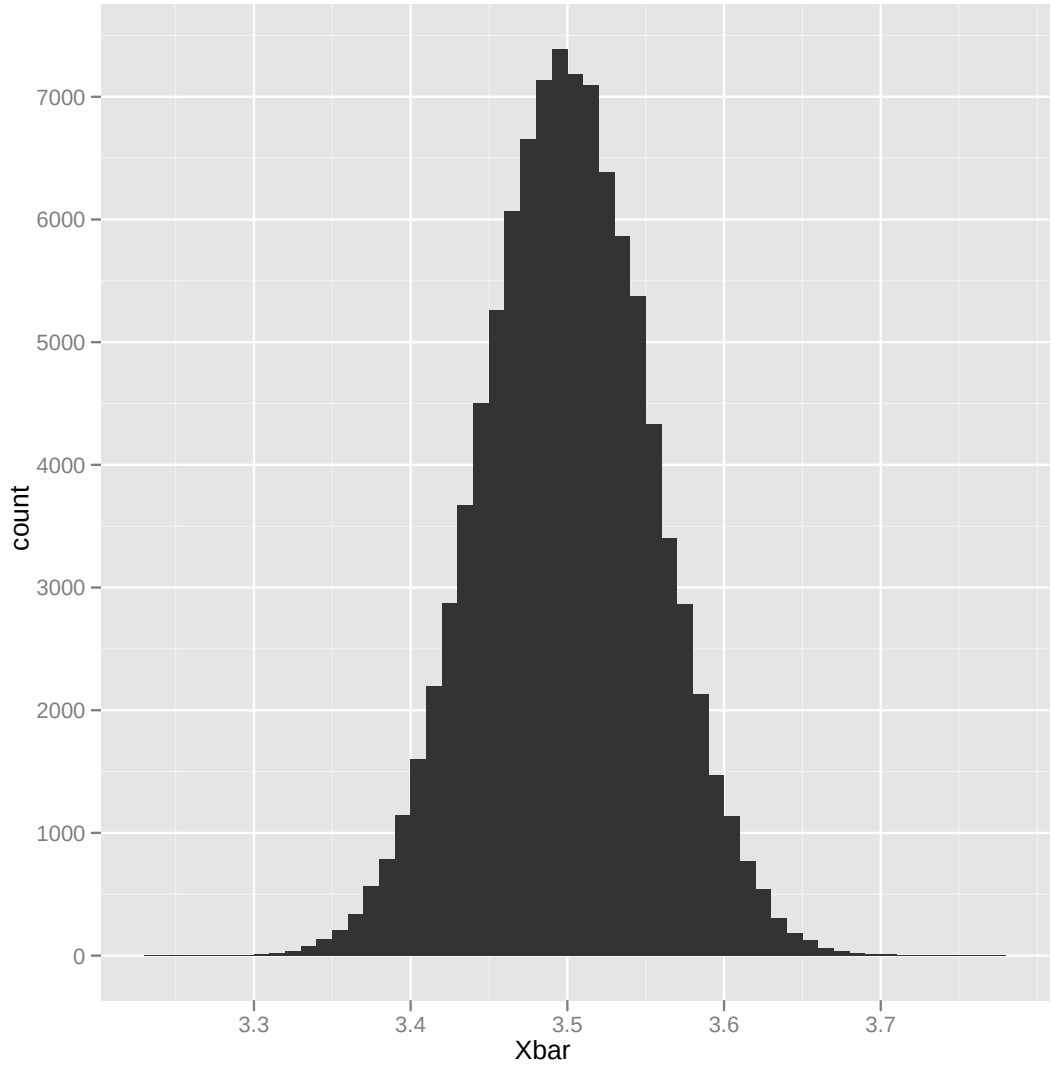


Figure: $N = 1000$