MA4413 Statistics for Computing

Lecture 6B : Normal Distribution

Kevin O'Brien

kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

Autumn 2012

Today's Class

- More on the Normal Distribution
  - Last class: definitions related to statistical inference.
  - Fundamental part of statistical inference is the Central Limit Theorem.
  - CLT based on Normal Distribution.
- Introduction to Quantile functions
  - The `qnorm` function
- A revision of the course material relevant to the mid-term.

Probability Density Function

Recall: As the Normal distribution is a continuous distribution, the PDF for a particular observed value will not give us an intuitive result (as far as this module is concerned). It is, in fact, the height of the density curve at a particular point.
Nonetheless, the relevant `R` code may be included in exam questions, so as to add complexity to questions.

```
> dnorm(0.7)
[1] 0.3122539
>
> dnorm(1.7)
[1] 0.09404908
```

Sample Question

Suppose X is a normally distributed random variable with mean $\mu = 2000$ and standard deviation $\sigma = 200$. Compute the probability of X being less than (or equal to) 2340.

$$P(X \leq 2340)$$

As always, we compute the z-score that corresponds to 2340.

$$z_o = \frac{x_o - \mu}{\sigma} = \frac{2340 - 2000}{200} = 1.7$$

R Implementation

Using the following R code, we can determine $P(Z \leq 1.7)$.

```
> pnorm(1.7)
[1] 0.9554345
```

Direct R Implementation

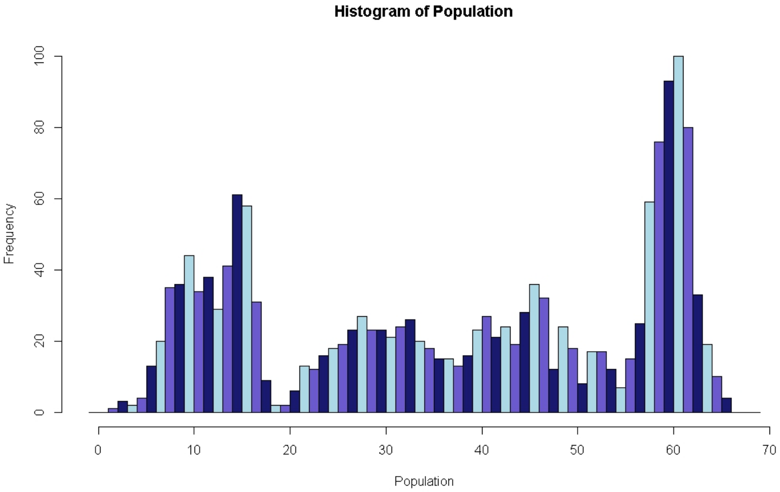
This can easily be implemented directly - without using the standardization formula, by specifying the normal mean and normal standard deviation directly. However, we will not be using this approach in this module.

```
> pnorm(2340,mean=2000,sd=200)
[1] 0.9554345
```

Central Limit Theorem

- The main aspect of the CLT that we shall consider is that many statistics (e.g sample mean and other related statistics, such as the sample variance) are normally distributed.
- Consider the population, characterized by the histogram on the next slide.
- While the population is not normally distributed, the population of sample statistic will be normally distributed.

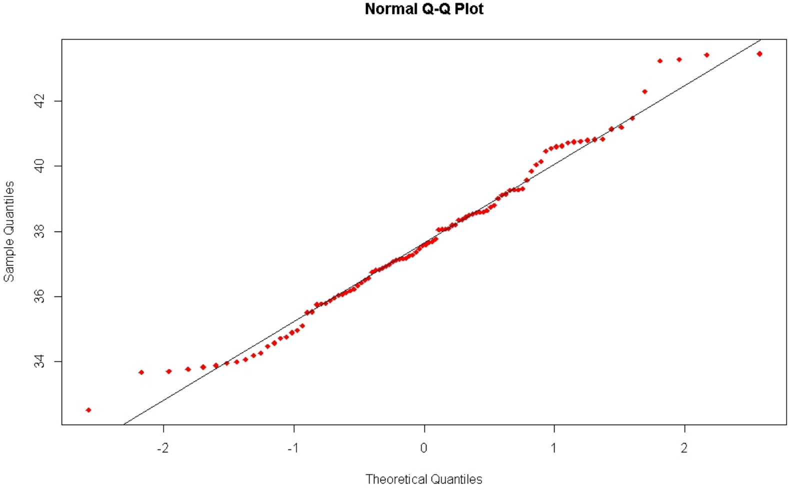
Central Limit Theorem



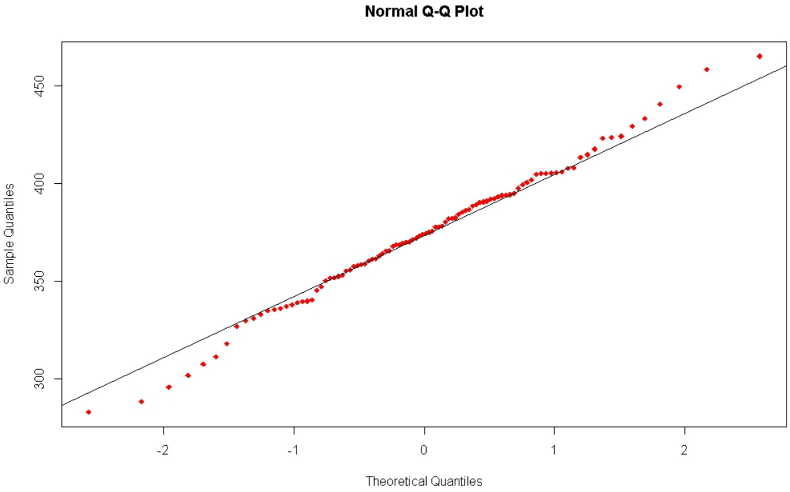
Central Limit Theorem

- Consider an experiment whereby a sample of 60 members of this population was taken, and the following sample statistics were computed
    - Sample mean $\bar{X}$
    - Sample variance s^2
    - Sample median $\tilde{X}$
- This experiment was performed 100 times (i.e. 100 independent samples were taken).
- The sample statistics were collated by type and examined to determine normality.
- A data set can be tested for normality using a very simple graphical procedure known as the 'Normal Probability Plot', or Q-Q plot.
- A data set can be assumed to be normally distributed if the points on the Q-Q plot follow the trendline.

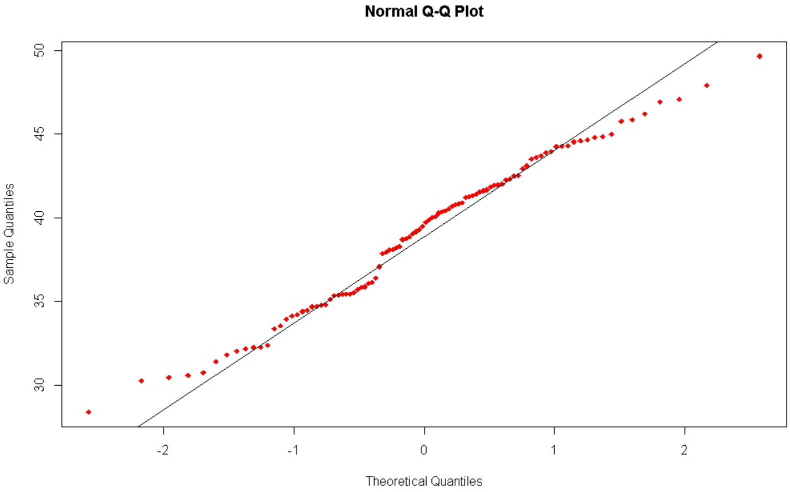
CLT: Sample Mean Q-Q plot



CLT: Sample Variance Q-Q plot



CLT: Sample Median Q-Q plot



Central Limit Theorem: Sampling Distributions

- In each of the three plots, the points follow the trend-line quite closely in each case.
- As we can see, the population of these statistics are normally distributed.
- We refer to these distributions as ‘Sampling Distributions’.
- While the statistic that we will be dealing with in this module do have normal sampling distributions, it must be noted that many statistics have sampling distributions other than the normal distribution.

Quantile Functions

- The Cumulative Distribution Function is used to identify the probability of a random variable being below a threshold value k.

$$P(X \leq k)$$

- In short, we compute a probability values associated with a specified value.
- (In `R`, this is carried out using the `p-` family of functions.)
- With Quantile Functions, we are performing the opposite operation, i.e. for a specified probability, we determine the threshold value k.
- For some value p, we computed k such that

$$P(X \leq k) = p$$

- (In `R`, this is performed using the `q-` family of functions.)

Quantile Functions

- Recall that, from the Murdoch Barnes Tables, $P(Z \leq 1.96) = 0.975$ and $P(Z \leq 1.28) = 0.8997$

```
> qnorm(0.975)
[1] 1.959964
>
> qnorm(0.8997)
[1] 1.279844
```