

#### 4.2.5 Προσομοιώσεις

Στη Στατιστική είναι αρκετά διαδεδομένη η άποψη (Chance & Rossman 2006 και Wood 2005) ότι η εκμάθηση των στατιστικών εννοιών διευκολύνεται με τη βοήθεια στατιστικών προσομοιώσεων (statistical simulations). Οι στατιστικές προσομοιώσεις χρησιμοποιούν πολλαπλά δείγματα για να προσεγγίσουν σύνθετες μαθηματικές διαδικασίες και υπολογισμούς (King et al. 2000, σελ. 349). Οι πολλαπλές αναπαραστάσεις της ίδιας στατιστικής έννοιας ενεργοποιούν το χρήστη, τον υποβοηθούν στη διερεύνηση και τελικά τον οδηγούν στην αποκαλυπτική μάθηση της υπό επεξεργασία έννοιας.

Ως πρώτο παράδειγμα θα μελετήσουμε την έννοια της καμπύλης συχνοτήτων μιας κατανομής. Επιπλέον, απαραίτητη προϋπόθεση σε πολλούς στατιστικούς ελέγχους είναι η κανονικότητα των πληθυσμών από τους οποίους προέρχονται τα δείγματα (Ενότητα 3.4). Η προσομοίωση και των δύο εννοιών προσεγγίζονται με τη βοήθεια του ιστογράμματος δειγμάτων που προέρχονται από πληθυσμούς με προκαθορισμένη κατανομή. Ειδικότερα, το αρχείο `simulation_normal.R` εξετάζει εποπτικά αν ικανοποιείται η προϋπόθεση της κανονικότητας, δηλαδή αν τα δεδομένα μας προέρχονται από έναν πληθυσμό που ακολουθεί την κανονική κατανομή με μέση τιμή  $\mu$  και διακύμανση  $\sigma^2$ . Για να επιτευχθεί αυτό σχηματίζεται το ιστόγραμμα συχνοτήτων για διαφορετικό αριθμό κλάσεων, το αντίστοιχο πολύγωνο συχνοτήτων, το πλήθος των στοιχείων κάθε κλάσης και την καμπύλη της κανονικής κατανομής που μπορεί να προσαρμοστεί στα δεδομένα μας. Η συνάρτηση `sim.norm()` δέχεται ως ορίσματα:

- το διάνυσμα δεδομένων `data` με προκαθορισμένη τιμή `rnorm(500, mean=8, sd=1)`. Δηλαδή, τα δεδομένα μας αποτελούνται από 500 τυχαίους αριθμούς που προέρχονται από την κανονική κατανομή με μέση τιμή  $\mu = 8$  και τυπική απόκλιση  $\sigma = 1$ .
- το `n` που αντιστοιχεί στο πλήθος των κλάσεων, με προκαθορισμένη τιμή `5`.
- την ετικέτα του οριζόντιου άξονα `x.lab`, με προκαθορισμένη τιμή `"ώρες εξωσχολικών δραστηριοτήτων"`.

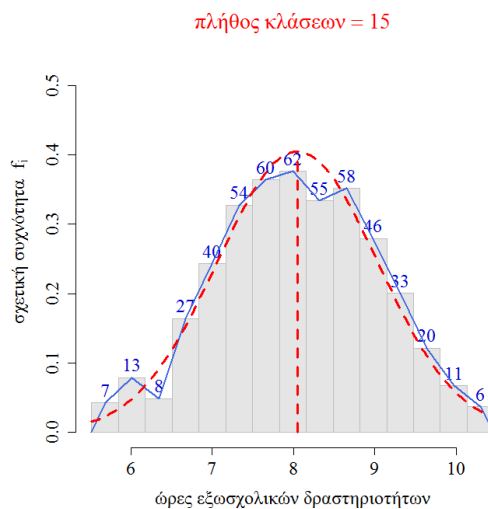
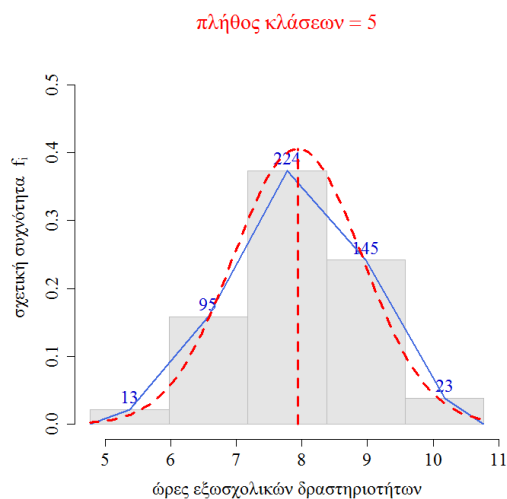
Οι προκαθορισμένες τιμές προέρχονται από το Παράδειγμα 9, που αφορούσε τις εξωσχολικές δραστηριότητες των μαθητών. Εννοείται ότι μπορούμε να τροποποιήσουμε οποιοδήποτε από αυτές. Η συνάρτηση καλείται με την εντολή:

```
> sim.norm (data=rnorm(500, mean=8, sd=1), n=5,
            x.lab="ώρες εξωσχολικών δραστηριοτήτων")
```

ή

```
> sim.norm ()
```

Στη δεύτερη περίπτωση η R δέχεται για τα ορίσματα της τις προκαθορισμένες τιμές, επομένως οι δύο παραπάνω εντολές παράγουν το ίδιο αποτέλεσμα, που απεικονίζεται στο παρακάτω σχήμα (στο αριστερό μέρος).



Με τον τρόπο αυτό μπορούμε να κάνουμε μια προσομοίωση των δεδομένων για τα διάφορα ορίσματα της συνάρτησης. Έτσι, μπορούμε να έχουμε το παραπάνω σχήμα για τις διάφορες τιμές του  $n$ . Μπορούμε να πληκτρολογήσουμε

```
> sim.norm (n=10)
```

για να έχουμε 10 κλάσεις. Ο πιο εύκολος τρόπος για να αποφύγουμε την πληκτρολόγηση είναι να πατήσουμε το πάνω βελάκι από τα τέσσερα βελάκια κατεύθυνσης που βρίσκονται συνήθως στο κάτω δεξιά μέρος του πληκτρολογίου.

Τότε εμφανίζεται η τελευταία εντολή στην οποία μπορούμε να κινηθούμε δεξιά και αριστερά με τα αντίστοιχα βελάκια κατεύθυνσης, να διαγράψουμε το **10** και να το αντικαταστήσουμε με οποιονδήποτε άλλο αριθμό επιθυμούμε. Π.χ.

```
> sim.norm (n=15)
```

Πρέπει να σημειωθεί ότι με τις εντολές

```
> sim.norm (n=10)
```

```
> sim.norm (n=15)
```

έχουμε διαφορετικές κλάσεις αλλά σε ελαφρώς διαφορετικά δεδομένα, διότι κάθε φορά που εκτελείται η συνάρτηση και δε δίνουμε δεδομένα καλείται η προκαθορισμένη τιμή της εντολής `rnorm`, η οποία βέβαια παρέχει διαφορετικά δεδομένα από την ίδια κανονική κατανομή. Επομένως, για να έχουμε διαφορετικό πλήθος κλάσεων για τα ίδια δεδομένα, πρέπει να εισάγουμε στην R ένα διάνυσμα δεδομένων, π.χ.

```
> x <- <data=rnorm(500, mean=8, sd=1)
```

και στην κλήση της συνάρτησης να γίνεται ως εξής:

```
> sim.norm (data=x, n=5, x.lab="")
```

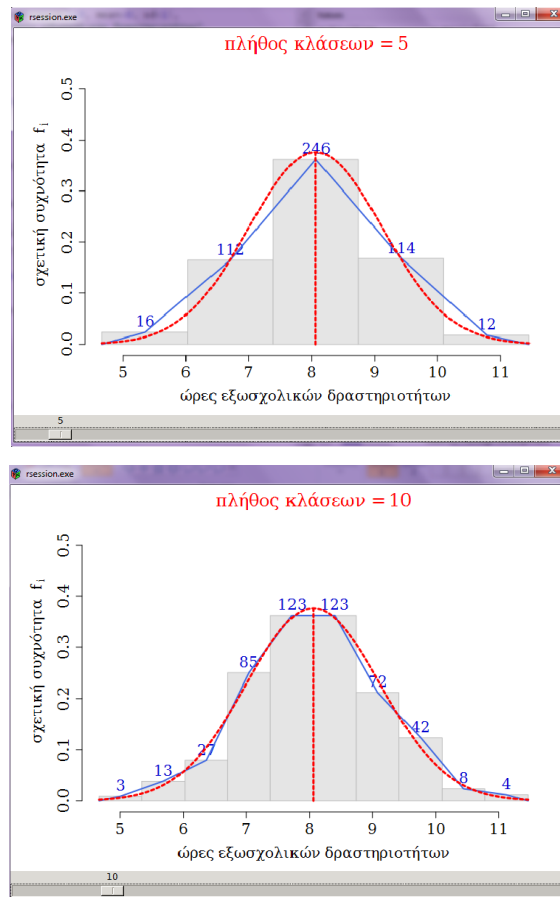
```
> sim.norm (data=x, n=10, x.lab="")
```

```
> sim.norm (data=x, n=20, x.lab="")
```

για να έχουμε αριθμό κλάσεων 5, 10 και 20 στα δεδομένα που ορίσαμε με το `x`. Επίσης μπορούμε να έχουμε δεδομένα από διαφορετικές κατανομές. Η R διαθέτει έναν ευρύ κατάλογο θεωρητικών κατανομών πιθανότητας, που μπορούν να δούμε με την εντολή:

```
> ?Distributions
```

Με το αρχείο `simulation_normal_var.R` έχουμε το ίδιο παραγόμενο αποτέλεσμα με το αρχείο `simulation_normal.R`, με την προσθήκη όμως ενός δρομέα, δηλαδή μιας οριζόντιας γραμμής κύλισης η τιμή της οποίας αντιστοιχεί στο πλήθος των κλάσεων. Καθώς ο δρομέας παίρνει τιμές από 2 έως 50, ταυτόχρονα μεταβάλλονται όλα τα στοιχεία του γραφήματος, δηλαδή το ιστόγραμμα και το αντίστοιχο πολύγωνο συχνοτήτων καθώς και καμπύλη της κανονικής κατανομής. Η αρχική τιμή του δρομέα είναι το 5. Το Σχήμα ... απεικονίζει δύο στιγμιότυπα του γραφήματος για πλήθος κλάσεων 5 και 10.



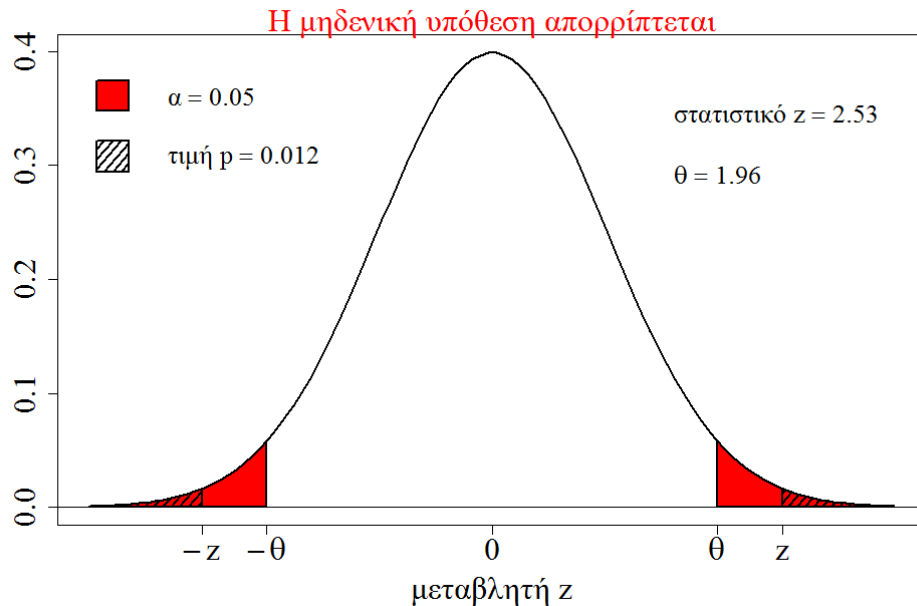
Μία άλλη προσομοίωση αφορά τον έλεγχο υποθέσεων του  $z$  test, όπως στο Παράδειγμα 9. Τα αρχεία [simulation\\_ztest.R](#) και [simulation\\_ztest\\_var.R](#) απεικονίζουν όλα τα σημαντικά χαρακτηριστικά του  $z$  test. Για να λειτουργήσουν αυτά τα αρχεία χρειάζεται να μεταφορτωθεί το πακέτο TeachingDemos για το πρώτο αρχείο και τα πακέτα TeachingDemos, cairoDevice και RGtk2 για το δεύτερο. Υπενθυμίζουμε, ότι για να γίνει αυτό τσεκάρουμε το άδειο κουτάκι που βρίσκεται μπροστά από το όνομα του κάθε απαιτούμενου πακέτου. Τότε στη κονσόλα του RStudio θα εμφανιστεί η εντολή `library` και το μήνυμα επιβεβαίωσης της ενεργοποίησης του πακέτου που θέλουμε να χρησιμοποιήσουμε.

Τα αρχεία [simulation\\_ztest.R](#) και [simulation\\_ztest\\_var.R](#) απεικονίζουν με την κωδωνοειδή γραφική παράσταση, την κατανομή της στατιστικής συνάρτησης  $z$ , που είναι κανονική με μέση τιμή μηδέν και διακύμανση 1. Βλέπουμε επίσης την τιμή του  $z$  και την τιμή του  $\theta$ , που καθορίζει την κρίσιμη περιοχή από το  $-\theta$  έως το  $\theta$ . Το εμβαδόν της κόκκινης σκιασμένης περιοχής παριστάνει το επίπεδο σημαντικότητας  $\alpha$ , ενώ το εμβαδόν της γραμμοσκιασμένης περιοχής την τιμή  $p$  του ελέγχου. Η συνάρτηση που ορίζει το πρώτο αρχείο ενεργοποιείται με την εντολή

```
> sim.ztest()
```

ή

```
> sim.ztest <- function( data=rnorm(500, mean=15, sd=1),  
                          mean.data=15, sd.data=1, alpha=0.05 )
```

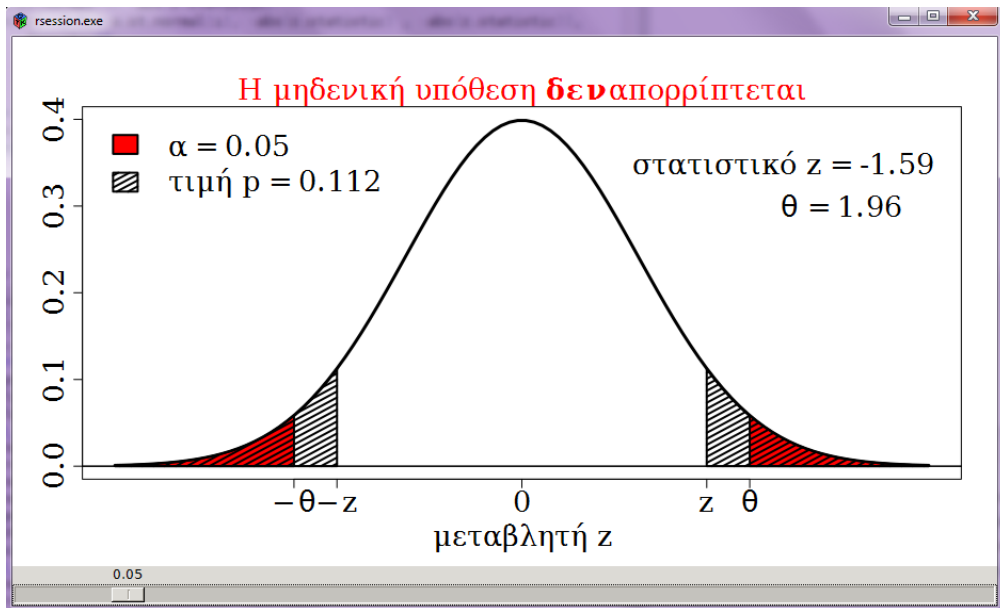


Στο παραπάνω σχήμα βλέπουμε ότι η τιμή  $p$  είναι μικρότερη από το επίπεδο σημαντικότητας  $\alpha$ , οπότε η μηδενική υπόθεση απορρίπτεται. Η συνάρτηση που ορίζει το δεύτερο αρχείο ενεργοποιείται με την εντολή

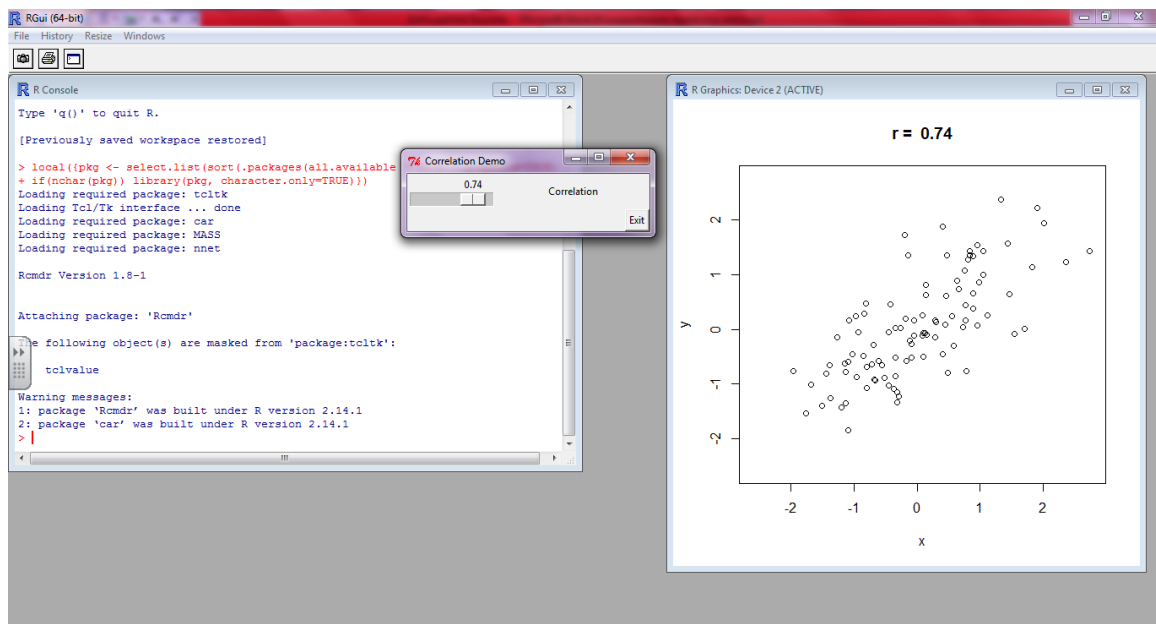
```
> sim.ztest.var()
```

και έχουμε το ίδιο αποτέλεσμα με τη συνάρτηση `sim.ztest()`, με διαφορά ότι η τιμή του επιπέδου σημαντικότητας  $\alpha$ , είναι μεταβαλλόμενη από τη γραμμή κύλισης στο κάτω μέρος του παραθύρου. Στο στιγμιότυπο που ακολουθεί η τιμή  $p$  είναι μεγαλύτερη από το επίπεδο σημαντικότητας, επομένως δεν μπορούμε να απορρίψουμε τη μηδενική υπόθεση.

Η προσομοίωση είναι επίσης πολύ χρήσιμη στη συσχέτιση και γραμμική παλινδρόμηση. Η δυνατότητα αυτή υλοποιείται υλοποίηση μέσω του **R Commander** και προϋποθέτει την εγκατάσταση και ενεργοποίηση του πακέτου TeachingDemos από το **R Commander**, επιλέγοντας το μενού **Tools** και μετά **Load package(s)**... . Στο επόμενο παράθυρο επιλέγουμε το TeachingDemos και μετά **OK**. Κατόπιν, χρειάζεται η εγκατάσταση και ενεργοποίηση του πακέτου RcmdrPlugin.TeachingDemos από το μενού



**Tools** και μετά **Load Rcmdr plug-in(s)...** . Αν η διαδικασία έγινε σωστά τότε στα μενού του **R Commander** έχει προστεθεί το μενού **Demos**, από το οποίο επιλέγουμε **Simple correlation**. Εμφανίζεται ένα διάγραμμα σκεδασμού ενός δείγματος με συντελεστή γραμμικής συσχέτισης  $r=0$ . Ταυτόχρονα, σε ένα μικρότερο παράθυρο έχουμε τη δυνατότητα μεταβολή του  $r$ , σε όλο το εύρος δυνατών τιμών, από το  $-1$  έως και το  $1$ . Ένα στιγμιότυπο της παραπάνω προσομοίωσης παρατίθεται στο επόμενο σχήμα.



Από το μενού **Demos** και πάλι επιλέγουμε **Simple linear regression** για να προσομοιώσουμε την γραμμική παλινδρόμηση. Όταν το **Add Point** είναι ενεργοποιημένο (μαύρο χρώμα), κάθε φορά που πατάμε με το ποντίκι στο εσωτερικό

του σχήματος προστίθεται ένα σημείο. Με αντίστοιχους χειρισμούς μπορούμε να διαγράψουμε ένα σημείο (Delete Point) ή να μετακινήσουμε ένα σημείο (Move Point). Μόλις, ολοκληρωθεί η τοποθέτηση των σημείων στο σχήμα σχηματίζεται αυτόματα η ευθεία γραμμικής παλινδρόμησης, η οποία τροποποιείται κάθε φορά που προστίθεται κάποιο σημείο ή κάποιο από τα σημεία μετακινείται ή διαγράφεται. Ταυτόχρονα υπολογίζονται οι συντελεστές της ευθείας παλινδρόμησης, και ο συντελεστής συσχέτισης. Η διαδικασία ολοκληρώνεται με το πάτημα του κουμπιού End. Ένα στιγμιότυπο της παραπάνω προσομοίωσης παρατίθεται στο επόμενο σχήμα.

