



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΑΝΟΙΚΤΑ ΑΚΑΔΗΜΑΪΚΑ ΜΑΘΗΜΑΤΑ



Ηλεκτρονικοί Υπολογιστές I

Κατανομές και έλεγχοι υποθέσεων με τη
γλώσσα R

Διδάσκων: Επίκουρος Καθηγητής
Αθανάσιος Σταυρακούδης



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Υπολογιστική Στατιστική με τη γλώσσα R

Αθανάσιος Σταυρακούδης

<http://stavrakoudis.econ.uoi.gr>

13 Δεκεμβρίου 2012



Απλές πράξεις με την R

```
> x <- c (1,2,3,4)
> min(x)
[1] 1
> max(x)
[1] 4
> sum(x)
[1] 10
> mean(x)
[1] 2.5
> median(x)
[1] 5
> var(x)
[1] 1.666667
> sqrt(x)
[1] 1.000000 1.414214 1.732051 2.000000
> x^2
[1] 1 4 9 16
```



Κανονική Κατανομή

```
> rnorm(30)
> rnorm(30, mean=15, sd=3)
> x <- rnorm(30, mean=15, sd=3)
```

Ομοιόμορφη Κατανομή

```
> runif(50)
> runif(50, min=0, max=4)
> x <- runif(50, min=0, max=4)
```



Άσκηση

Να κατασκευαστούν 100 τυχαίοι αριθμοί από κανονική κατανομή $(0,1)$, να δοθούν τα περιγραφικά στατιστικά, και να κατασκευαστεί ιστόγραμμα συχνοτήτων

```
> x <- rnorm(100, mean=0, sd=1)
```

```
> min(x)
```

```
[1] -2.502098
```

```
> max(x)
```

```
[1] 2.683640
```

```
> mean(x)
```

```
[1] -0.1715795
```

```
> sqrt(var(x))
```

```
[1] 0.9961516
```

```
> hist(x, 11)
```

library(moments)

- > **library** (moments)
- > kurkosis (x)
- > skewness (x)

$$kurt = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

$$skew = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}}$$



Νέο γράφημα

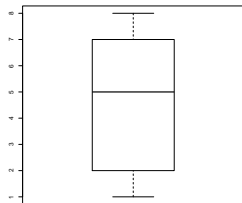
boxplot

```
> x <- c (1,5,2,7,8)
```

```
> summary(x)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.0	2.0	5.0	4.6	7.0	8.0

```
> boxplot(x)
```



Συνδυακόμενη και συσχέτιση

```
> x <- c (1,5,2,7,8)
```

```
> y <- c (2,9,3,6,10)
```

```
> var(x,y)
```

```
[1] 9.25
```

```
> cov(x,y)
```

```
[1] 9.25
```

```
> cor(x ,y)
```

```
[1] 0.857917
```



Η συνάρτηση `dnorm`

Το ύψος της καμπύλης κανονικής κατατομής

```
> dnorm(0)
[1] 0.3989423

> dnorm(1)
[1] 0.2419707

> dnorm(-1)
[1] 0.2419707

> dnorm(1,10,5)
[1] 0.01579003

> v <- c(-1,0,1)
> dnorm(v)
[1] 0.2419707 0.3989423 0.2419707
```



Η συνάρτηση pnorm

Το ολοκλήρωμα της συνάρτησης πιθανότητας

```
> pnorm(0)
[1] 0.5

> pnorm(1, mean=0, sd=1)
[1] 0.8413447

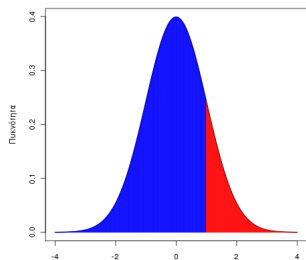
> pnorm(-1, mean=0)
[1] 0.1586553

> pnorm(1, mean=10, sd=5)
[1] 0.03593032

> v <- c(-1,0,1)
> pnorm(v)
[1] 0.1586553 0.5000000 0.8413447
```



Διαγραμματική επεξήγηση της συνάρτησης pnorm



Το εμβαδόν (ολοκλήρωμα) της περιοχής $(-\infty, 1)$ ισούται με 0.8413447

Το εμβαδόν (ολοκλήρωμα) της περιοχής $(1, \infty)$ ισούται με 0.1586553

```
> pnorm(1)
[1] 0.8413447
> 1-pnorm(1)
[1] 0.1586553
```



Η συνάρτηση qnorm

Σε ποιο x η πιθανότητα είναι p ;

```
> qnorm(0.5)
```

```
[1] 0
```

```
> qnorm(0.84, mean=0)
```

```
[1] 0.9944579
```

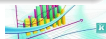
```
> qnorm(0.157, mean=0, sd=2)
```

```
[1] -2.013729
```

```
> v <- c(0.2, 0.5, 0.9)
```

```
> qnorm(v, mean=1, sd=1)
```

```
[1] 0.1583788 1.0000000 2.2815516
```



Συναρτήσεις κανονικής κατανομής

4 βασικές συναρτήσεις

`dnorm` Η πυκνότητα πιθανότητας **d** στο σημείο **x**

`pnorm` Η πιθανότητα **p** στο σημείο **x**

`qnorm` Σε ποιο σημείο **x** η πυκνότητα πιθανότητας είναι **d**

`rnorm` Τυχαίοι αριθμοί από κανονική κατανομή

Το ίδιο μοτίβο για άλλες κατανομές

`dt` Η πυκνότητα πιθανότητας **d** στο σημείο **x** της κατανομής **t**

`dbinom` Η πιθανότητα **p** στο σημείο **x** της διωνυμικής κατανομής

`qchisq` Σε ποιο σημείο **x** η πυκνότητα πιθανότητας της κατανομής χ^2 είναι **d**

`rgamma` Τυχαίοι αριθμοί από τη γάμμα κατανομή



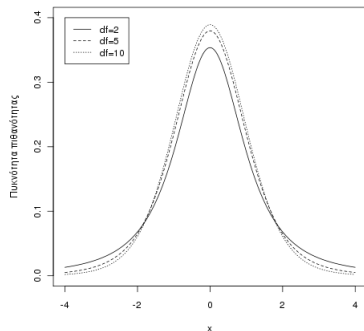
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

4 βασικές συναρτήσεις

- rt** Αποδίδει τυχαίους αριθμούς από την t κατανομή
- dt** Υπολογίζει το ύψος της καμπύλης της t κατανομής
- pt** Υπολογίζει το ολοκλήρωμα της t κατανομής
- qt** Υπολογίζει το σημείο στο οποίο η t κατανομή έχει δεδομένο ολοκλήρωμα, δηλαδή είναι η αντίστροφη της συνάρτησης **pt**.



Η κατανομή t διαγραμματικά



```
> x <- seq(-4, 4, by=0.01); y1 <- dt(x, df=2)
> y2 <- dt(x, df=5); y3 <- dt(x, df=10)
> plot(x, y1, ylim=c(0,0.5), lty=1, type="l", ylab="Density",
> lines(x, y2, lty=2); lines(x, y3, lty=3)
> legend(-4, 0.4, c("df=2", "df=5", "df=10"),
        lty=c(1, 2, 3))
```

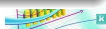


Το ύψος (που ακολουθεί κανονική κατανομή) 40 φοιτητών έχει μέσο 172 και τυπική απόκλιση 4. Ποια είναι η πιθανότητα κάποιος να έχει ύψος άνω του 178;

```
> pnorm(178, mean=172, sd=4, lower.tail=FALSE)
[1] 0.0668072
```

Το ύψος (που ακολουθεί κανονική κατανομή) 40 φοιτητών έχει μέσο 172 και τυπική απόκλιση 4. Ποια είναι η πιθανότητα κάποιος να έχει ύψος μέχρι 178;

```
> pnorm(178, mean=172, sd=4)
[1] 0.9331928
```



$$\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} < \mu < \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}$$

```
> x <- c(3,5,8,2,5,6,2,9)
> m <- mean(x)
[1] 5
> sd <- sqrt(var(x))
[1] 2.618615
> n <- length(x)
[1] 8
> e <- qt(0.975, df=n-1)*sd/sqrt(n)
[1] 2.189217
> m-e
[1] 2.810783
> m+e
[1] 7.189217
```



Ερώτημα

Να βρεθεί αν το διάνυσμα τιμών x έχει μέσο μικρότερο του 5.

Υπολογισμός z

```
> x <- c(5, 2, 7, 4, 5)
> m <- mean(x)
> sd <- sqrt(var(x))
> n <- length(x)
> (m-5)/(sd/sqrt(n))
[1] -0.492366

> pnorm(-0.492366)
[1] 0.3112303
```

Μη απόρριψη

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$H_0 : \bar{x} < 5$$

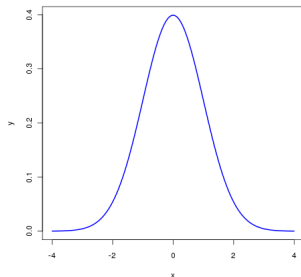
$$H_\alpha : \bar{x} \geq 5$$



Γράφημα τυπικής κανονικής κατανομής

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

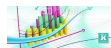
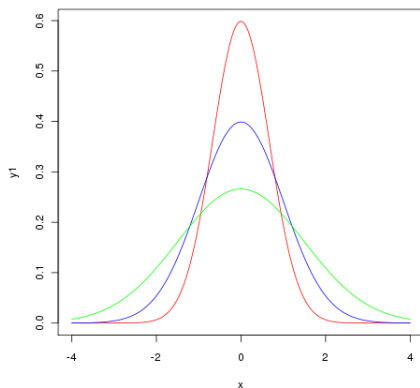
```
> x <- seq(-4,4,length=201)
> y <- 1/sqrt(2*pi)*exp(-x^2/2)
> plot(x,y,type="l")
> png("normal1.png")
> plot(x,y,type="l",col="blue")
> dev.off()
```



Πολλές γραμμές στο ίδιο γράφημα

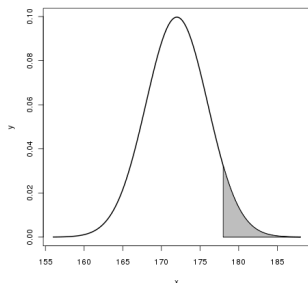
```
> x <- seq(-4, 4, by=0.01)
> y1 <- dnorm(x, mean=0,
              sd=2/3)
> y2 <- dnorm(x, mean=0,
              sd=1)
> y3 <- dnorm(x, mean=0,
              sd=3/2)

> png("normal2.png")
> plot(x, y1, type="l",
       col="red")
> lines(x, y2, type="l",
        col="blue")
> lines(x, y3, type="l",
        col="green")
> dev.off()
```



Σκίαση περιοχών

```
> x <- seq(156, 188, length=501)
> y <- dnorm(x, mean=172, sd=4)
> x1 <- seq(178, 188, length=101)
> y1 <- dnorm(x1, mean=172, sd=4)
> png("normal3.png")
> plot(x, y, type="l", lwd=2)
> polygon(c(178,x1,188), c(0,y1,0), col="gray")
> dev.off()
```



Έλεγχος του μέσου του πληθυσμού με γνωστή διακύμανση

Έστω ένα δείγμα 12 παρατηρήσεων που προέρχεται από κανονική κατανομή έχει μέσο 50. Η τυπική απόκλιση του πληθυσμού είναι ίση με 4. Ας κάνουμε την υπόθεση πως ο μέσος είναι μικρότερος του 52.

$$H_0 : \mu \geq 52$$

$$H_\alpha : \mu < 52$$

```
> n      <- 12
> xbar   <- 50
> sigma <- 4
> mu     <- 52
> z      <- (xbar - mu) / (sigma / sqrt(n))
> p      <- pnorm(z)
> p
[1] 0.04163226
```

Η τιμή $pval=0.0416$ μας λέει ότι απορρίπτεται η H_0 σε επίπεδο σημαντικότητας 5%, αλλά όχι σε επίπεδο σημαντικότητας 1%.



Έλεγχος του μέσου του πληθυσμού με γνωστή διακύμανση

Ας υποθέσουμε πως έχουμε ένα δείγμα που προέρχεται από πληθυσμό, που ακολουθεί την κανονική κατανομή αλλά δεν ξέρουμε τη διακύμανση. Έστω ότι θέλουμε να ελέγξουμε αν ο μέσος του πληθυσμού είναι μικρότερος του 52:

$$H_0 : \mu \geq 52$$

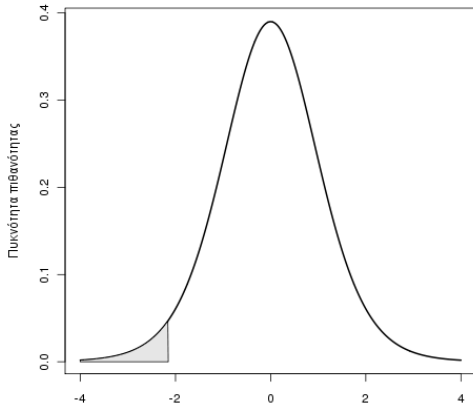
$$H_a : \mu < 52$$

```
> x      <- c(45, 48, 50, 52, 48, 51, 55, 50, 53, 54, 49, 45)
> n      <- length(x)
> mx     <- mean(x)
> sigma  <- sd(x)
> mu     <- 52
> z      <- (mx-mu) / (sigma/sqrt(n))
> p      <- pt(z, df=n-1)
> p
[1] 0.02722336
```

Η τιμή $p=0.027$ μας λέει πως απορρίπτεται η H_0 σε επίπεδο σημαντικότητας 5%, αλλά όχι σε επίπεδο σημαντικότητας 1%.



Έλεγχος του μέσου του πληθυσμού με γνωστή διακύμανση



Η σκιασμένη περιοχή στο παρακάτω γράφημα δείχνει την περιοχή απόρριψης



Σας ευχαριστώ για την προσοχή σας

Είμαι στη διάθεσή σας για σχόλια, απορίες και ερωτήσεις



Τέλος Ενότητας



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Ιωαννίνων**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Έχουν προηγηθεί οι κάτωθι εκδόσεις:

- Έκδοση 1.0 διαθέσιμη εδώ.

<http://ecourse.uoi.gr/course/view.php?id=1064>.

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Ιωαννίνων, Διδάσκων:
Επίκουρος Καθηγητής Αθανάσιος
Σταυρακούδης. «Ηλεκτρονικοί Υπολογιστές IV.
Κατανομές και έλεγχοι υποθέσεων με τη
γλώσσα R». Έκδοση: 1.0. Ιωάννινα 2014.
Διαθέσιμο από τη δικτυακή διεύθυνση:
<http://ecourse.uoi.gr/course/view.php?id=1064>.

Σημείωμα Αδειοδότησης

- Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά Δημιουργού - Παρόμοια Διανομή, Διεθνής Έκδοση 4.0 [1] ή μεταγενέστερη.



- [1] <https://creativecommons.org/licenses/by-sa/4.0/>.