

Μάθημα e-Learning του Οικονομικού Πανεπιστημίου Αθηνών

«Σχεδιασμός και Ανάλυση Κλινικών Δοκιμών με χρήση της R»

Ημερομηνία: 01/01/2024

Ονοματεπώνυμο: Αλέξανδρος Στεφανίδης

ΑΣΚΗΣΗ 1:

Πριν ξεκινήσουμε την ανάλυση δεδομένων, μετατρέπουμε τις κατηγορικές μεταβλητές sex, ae2 και dose1 από numeric (default) σε factor.

```
> str(dat1)
'data.frame':      105 obs. of  10 variables:
 $ age   : int   73 73 74 78 77 70 80 81 77 71 ...
 $ sex   : Factor w/ 2 levels "male","female": 2 1 2 2 2 1 2 2
 2 1 ...
 $ ian1  : int   73 72 72 76 76 64 79 78 75 69 ...
 $ ian2  : int   73 73 73 78 76 66 79 78 76 70 ...
 $ ian3  : int   73 73 73 78 76 66 79 78 76 70 ...
 $ dose1: Factor w/ 3 levels "8","16","24": 3 1 3 3 2 1 2 1 3
 1 ...
 $ mmse1: int   19 23 16 22 23 21 25 22 20 19 ...
 $ mmse2: int   22 27 20 24 23 22 29 22 23 22 ...
 $ ae2   : Factor w/ 2 levels "No Adverse Reaction",...: 1 1 1 2
 1 1 1 2 1 1 ...
 $ mmse3: int    3 4 4 2 0 1 4 0 3 3 ...
```

Το σύνολο δεδομένων περιέχει 105 παρατηρήσεις και 10 μεταβλητές, όπου ονομάσαμε τα επίπεδα της μεταβλητής sex σε “male” και “female”, και τα επίπεδα της μεταβλητής ae2 σε “No Adverse Reaction (τιμή 0)” και “Adverse Reaction” (τιμή 1). Επίσης θέτουμε ως mmse3 (βελτίωση στην τιμή του mmse) την διαφορά mmse2 – mmse1.

```
> dat1$mmse3 <- dat1$mmse2 - dat1$mmse1
```

Διερευνητική Ανάλυση Δεδομένων.

Για τα χαρακτηριστικά των μεταβλητών έχουμε τα εξής:

```
> summary(age); sd(age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 61.00   70.00   75.00   74.72   79.00   93.00
[1] 6.613644
```

Η μέση τιμή για τη μεταβλητή age είναι 74.72, η διάμεσος 75 χρόνια, η ελάχιστη τιμή 61 και η μέγιστη τιμή 93 χρόνια, ενώ το πρώτο και το τρίτο τεταρτημόριο είναι 70 και 79 χρόνια αντίστοιχα. Η τυπική απόκλιση είναι 6.613.

Με παρόμοιο τρόπο παίρνουμε και τα χαρακτηριστικά για τις υπόλοιπες μεταβλητές:

Για τη μεταβλητές ian1, ian2 και ian3 :

```
> summary(ian1);sd(ian1)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 55.00  68.00   72.00   71.86  76.00   90.00
[1] 6.53015
```

Η μεταβλητή ian1 εκφράζει την ηλικία εμφάνισης πρώτων συμπτωμάτων άνοιας. Έτσι η μέση ηλικία εμφάνισης συμπτωμάτων είναι 71.86 (δηλαδή 72 χρόνια), η διάμεσος είναι 72 χρόνια, η ελάχιστη και η μέγιστη τιμή είναι 55 και 90 χρόνια αντίστοιχα και τέλος το πρώτο και το τρίτο τεταρτημόριο είναι 68 και 76 χρόνια αντίστοιχα. Η τυπική απόκλιση είναι 6.5.

Για τη μεταβλητή ian2 (ηλικία διάγνωσης άνοιας):

```
> summary(ian2);sd(ian2)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 56.00  68.00   73.00   72.77  77.00   90.00
[1] 6.60748
```

Η μέση ηλικία διάγνωσης άνοιας είναι 72.77 χρόνια, η ελάχιστη ηλικία είναι 56 χρόνια και η μέγιστη 90 χρόνια. Η τυπική απόκλιση είναι 6.607.

Για τη μεταβλητή ian3:

```
> summary(ian3);sd(ian3)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 56.00  69.00   73.00   73.15  78.00   90.00
[1] 6.684915
```

Η μέση ηλικία έναρξης χορήγησης αρχικής θεραπείας για την άνοια είναι περίπου 73 χρόνια. Η ελάχιστη ηλικία είναι 56 χρόνια ενώ η μέγιστη 90 χρόνια. Η διάμεσος είναι 73 χρόνια και τέλος η τυπική απόκλιση είναι 6.69.

Για τη μεταβλητή mmse1:

```
> summary(mmse1);sd(mmse1)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 14.00  17.00   20.00   19.57  22.00   25.00
[1] 2.582698
```

Η μέση τιμή του έλεγχου minimal στην επίσκεψη 1 είναι 19.57 μονάδες ενώ η ελάχιστη και η μέγιστη τιμή του είναι 14 και 25 μονάδες αντίστοιχα. Η διάμεσος είναι 20 μονάδες και η τυπική απόκλιση 2.59.

Για τη μεταβλητή mmse2:

```
> summary(mmse2);sd(mmse2)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 15.00  20.00   22.00   21.95  24.00   29.00
[1] 3.142087
```

Για τη μεταβλητή mmse3:

```
> summary(mmse3);sd(mmse3)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
-3.000  1.000  2.000  2.381  4.000 12.000  
[1] 2.301258
```

Τώρα για τις κατηγορικές μεταβλητές του συνόλου δεδομένων dat1 έχουμε τα εξής:

Για τα ποσοστά της μεταβλητής sex:

```
> summary(dat1$sex)  
male female  
  41     64
```

Στο σύνολο δεδομένων που έχουμε, 41 άτομα/ασθενείς είναι άνδρες και 64 άτομα είναι γυναίκες.

Για τα ποσοστά της μεταβλητής dose1:

```
> summary(dat1$dose1)  
 8 16 24  
35 35 35
```

35 άτομα πήραν δόση φαρμάκου 8 mg, 35 άτομα πήραν δόση 16 mg, και τέλος 35 άτομα πήραν δόση 24 mg.

Για τα ποσοστά της μεταβλητής ae2:

```
> summary(dat1$ae2)  
No Adverse Reaction    Adverse Reaction  
          90             15
```

Δηλαδή 90 ασθενείς δεν εμφάνισαν ανεπιθύμητες ενέργειες από την επίσκεψη 1, ενώ 15 ασθενείς εμφάνισαν ανεπιθύμητες ενέργειες.

Επίλυση πρωτεύον ερωτήματος:

Για να απαντήσουμε τώρα στο πρωτεύον ερώτημα, αρκεί να εξετάσουμε εάν η μέση τιμή της μεταβλητής mmse3 (βελτίωση στην τιμή του mmse) διαφέρει (στατιστικώς) στα τρία επίπεδα της μεταβλητής dose1. Επειδή δεν γνωρίζουμε εάν η μεταβλητή mmse ακολουθεί κανονική κατανομή, θα αρκεστούμε στον μη-παραμετρικό έλεγχο των Kruskal-Wallis με μηδενική υπόθεση την εξής:

$$H_0 : \text{Οι κατανομές των τριων επιπέδων της dose1 είναι ίδιες}$$

```
> kruskal.test(dat1$mmse3 ~ dat1$dose1)  
  
kruskal-wallis rank sum test  
  
data:  dat1$mmse3 by dat1$dose1  
kruskal-wallis chi-squared = 8.6417, df = 2, p-value = 0.01329
```

Πήραμε p-value=0.013 < 0.05 άρα σε επίπεδο σημαντικότητας 5% απορρίπτουμε την μηδενική υπόθεση, που πρακτικά σημαίνει ότι η βελτίωση στην τιμή mmse μεταξύ των επισκέψεων 2 και 1, διαφέρει μεταξύ των τριών δόσεων του φαρμάκου ($\alpha=5\%$).

Απαντήσεις στα δευτερεύοντα ερωτήματα:

Απάντηση στο ερώτημα 1:

Για να απαντήσουμε στο πρώτο ερώτημα, θα πρέπει να κάνουμε συγκρίσεις ανά δυο ομάδες. Θα εφαρμόσουμε τα pairwise Wilcoxon tests.

```
> pairwise.wilcox.test(dat1$mmse3, dat1$dose1, p.adj = "bonferroni", exact = F)
```

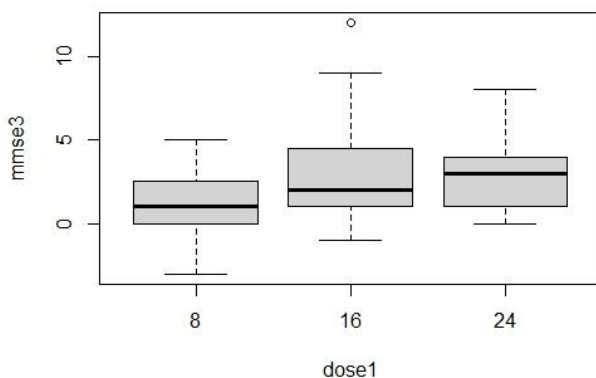
Pairwise comparisons using wilcoxon rank sum test with continuity correction

data: dat1\$mmse3 and dat1\$dose1

	8	16
16	0.1260	-
24	0.0095	1.0000

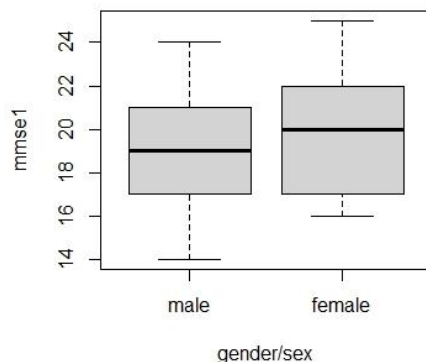
P value adjustment method: bonferroni

Παρατηρούμε ότι για τις δόσεις των 8 mg και 24 mg, πήραμε $p\text{-value}=0.01<0.05$ άρα υπάρχει στατιστικά σημαντική διαφορά για αυτές τις δύο ομάδες ($\alpha=5\%$). Συνεπώς για τις δόσεις των 8 και 24 mg παίρνουμε διαφορετικές τιμές βελτίωσης.



Από τα boxplots παρατηρούμε επίσης ότι η δόσεις των 8mg και 16 mg, καθώς και των 8 mg και 24 mg φαίνεται να διαφέρουν ως προς την βελτίωση του mmse.

Απάντηση στο ερώτημα 2:



Από το διάγραμμα boxplot που πήραμε, βλέπουμε ότι διάμεσος και η κατανομή της mmse1 για τους άνδρες και τις γυναίκες είναι περίπου ίδιες.

Για να εξετάσουμε εάν η μέση τιμή του mmse στη baseline σχετίζεται με τη μεταβλητή sex, θα εφαρμόσουμε τον έλεγχο των Mann-Whitney

```
> wilcox.test(dat1$mmse1 ~ dat1$sex)

    wilcoxon rank sum test with continuity correction

data:  dat1$mmse1 by dat1$sex
W = 1215, p-value = 0.5233
alternative hypothesis: true location shift is not equal to 0
```

Πήραμε $p\text{-value} = 0.523 > 0.05$, άρα δεν απορρίπτουμε την μηδενική υπόθεση σε ε.σ. 5%. Δηλαδή η μέση τιμή του mmse στη baseline δεν σχετίζεται με το φύλο των ασθενών ($\alpha=5\%$).

Θα εξετάσουμε τώρα τη συσχέτιση της μεταβλητής mmse με τις συνεχείς μεταβλητές age, ian1, ian2 και ian3.

Για τις μεταβλητές mmse και age:

```
> cor.test(dat1$mmse1, age)

Pearson's product-moment correlation

data:  dat1$mmse1 and age
t = -2.9023, df = 103, p-value = 0.004531
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.44325609 -0.08791754
sample estimates:
      cor 
-0.2749497
```

Παρατηρούμε ότι για την μηδενική υπόθεση $H_0 : \rho_{XY} = 0$ πήραμε ότι το διάστημα εμπιστοσύνης δεν περιέχει το μηδέν, και ακόμα $p\text{-value} = 0.0045 < 0.05$ άρα απορρίπτουμε την μηδενική υπόθεση, δηλαδή οι μεταβλητές mmse1 και age σχετίζονται σε ε.σ. 5%.

Για τις μεταβλητές mmse1 και ian1:

```
> cor.test(dat1$mmse1, ian1)

Pearson's product-moment correlation

data: dat1$mmse1 and ian1
t = -1.8554, df = 103, p-value = 0.0664
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.35912075  0.01225399
sample estimates:
      cor
-0.1798336
```

Πήραμε $p\text{-value} = 0.0664 > 0.05$ και επίσης το διάστημα εμπιστοσύνης $(-0.35, 0.12)$ περιέχει το μηδέν, άρα σε επίπεδο σημαντικότητας 5% ο έλεγχος minimal (mmse1) δεν σχετίζεται με την ηλικία εμφάνισης πρώτων συμπτωμάτων άνοιας.

Για τις μεταβλητές mmse1 και ian2:

```
> cor.test(dat1$mmse1, ian2)

Pearson's product-moment correlation

data: dat1$mmse1 and ian2
t = -1.9827, df = 103, p-value = 0.05007
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -3.698099e-01 -7.217418e-05
sample estimates:
      cor
-0.1917348
```

Πήραμε $p\text{-value} = 0.05$ άρα δεν απορρίπτουμε την μηδενική υπόθεση, δηλαδή οι μεταβλητή mmse1 και η ηλικία διάγνωσης άνοιας δεν σχετίζονται.

Τέλος για τις μεταβλητές mmse1 και ian3:

```
> cor.test(dat1$mmse1, ian3)

Pearson's product-moment correlation

data: dat1$mmse1 and ian3
t = -2.3017, df = 103, p-value = 0.02336
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.39605455 -0.03082027
sample estimates:
      cor
-0.221179
```

Πήραμε $p\text{-value} = 0.023 < 0.05$ άρα σε ε.σ. 5% απορρίπτουμε την μηδενική υπόθεση, δηλαδή οι μεταβλητές mmse1 και η ηλικία έναρξης χορήγησης θεραπείας σχετίζονται ($\alpha=5\%$).

Απάντηση στο ερώτημα 3:

Αρκεί να εξετάσουμε την υπόθεση $H_0 : \mu_1 = \mu_2 = \mu_3$, για την μεταβλητή mmse1 στα τρία επίπεδα της μεταβλητής dose1 (8, 16 και 24 mg). Επειδή δεν έχουμε γνώση της κατανομής της mmse1 θα εφαρμόσουμε τον μη-παραμετρικό έλεγχο Kruskal-Wallis:

```
> kruskal.test(dat1$mmse1 ~ dat1$dose1)

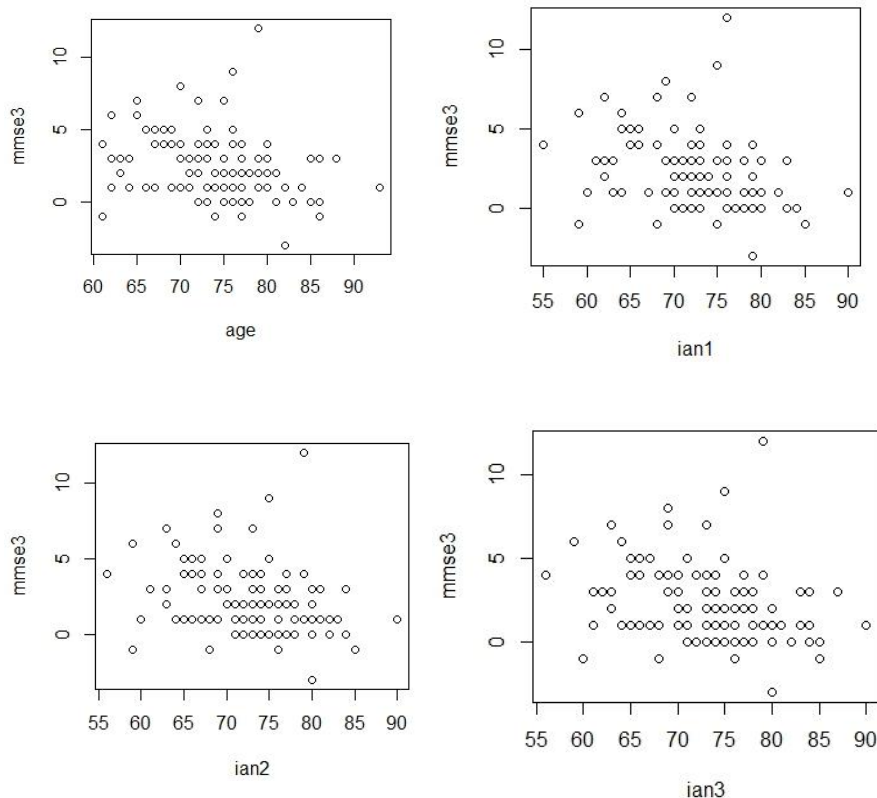
kruskal-wallis rank sum test

data:  dat1$mmse1 by dat1$dose1
kruskal-wallis chi-squared = 2.4996, df = 2, p-value = 0.2866
```

Πήραμε $p\text{-value}=0.286 > 0.05$ άρα δεν απορρίπτουμε την H_0 ($\alpha=5\%$) και συνεπώς η μέση τιμή του mmse στη baseline δεν σχετίζεται με την θεραπεία που πήραν οι ασθενείς.

Απάντηση στο ερώτημα 4:

Για να εξετάσουμε την συσχέτιση της μεταβλητής mmse3 με τις συνοδευτικές μεταβλητές θα εφαρμόσουμε τα εξής: Αρχικά κάνουμε τα διαγράμματα (plots), υπολογίζουμε τους συντελεστές συσχέτισης και εφαρμόζουμε γραμμικά μοντέλα.



Από τα παραπάνω διαγράμματα για τις συνεχείς μεταβλητές, παρατηρούμε ότι η mmse3 δεν έχει κάποια γραμμική σχέση με τις υπόλοιπες μεταβλητές baseline και ότι δεν υπάρχει κάποιο pattern στα σημεία που να υποδηλώνει κάποια φανερή σχέση μεταξύ των μεταβλητών.

Εργαζόμενοι όπως και στο ερώτημα (2) θα βρούμε τους συντελεστές συσχέτισης:

```
> cor.test(dat1$mmse3, age)

Pearson's product-moment correlation

data: dat1$mmse3 and age
t = -3.2072, df = 103, p-value = 0.001786
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4660731 -0.1163815
sample estimates:
      cor
-0.3013253

> cor.test(dat1$mmse3, ian1)

Pearson's product-moment correlation

data: dat1$mmse3 and ian1
t = -3.4139, df = 103, p-value = 0.0009169
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4810945 -0.1354394
sample estimates:
      cor
-0.3188282

> cor.test(dat1$mmse3, ian2)

Pearson's product-moment correlation

data: dat1$mmse3 and ian2
t = -3.299, df = 103, p-value = 0.001333
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4727870 -0.1248677
sample estimates:
      cor
-0.3091345

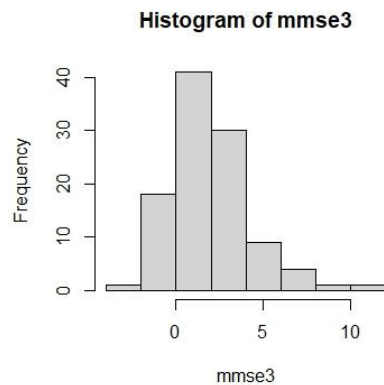
> cor.test(dat1$mmse3, ian3)

Pearson's product-moment correlation

data: dat1$mmse3 and ian3
t = -3.2732, df = 103, p-value = 0.001448
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4709129 -0.1224938
sample estimates:
      cor
-0.3069525
```


Έχουμε ισχυρές ενδείξεις ότι οι συντελεστές συσχέτισης είναι όλοι διάφοροι του μηδενός σε ε.σ. 5%, επομένως παίρνουμε ότι η βελτίωση της mmse σχετίζεται με όλες τις μεταβλητές baseline.

Για να εξετάσουμε τη συσχέτιση της mmse3 με τη μεταβλητή φύλο/sex αρκεί να κάνουμε τον έλεγχο υπόθεσης $H_0 : \mu_1 = \mu_2$ για τα δύο επίπεδα της μεταβλητής sex.



Το ιστόγραμμα δεν δείχνει ένδειξη ότι η mmse3 ακολουθεί κανονική κατανομή.

```
> ks.test(mmse3, pnorm)

Asymptotic one-sample Kolmogorov-Smirnov test

data:  mmse3
D = 0.66039, p-value < 2.2e-16
alternative hypothesis: two-sided

Warning message:
In ks.test.default(mmse3, pnorm) :
ties should not be present for the Kolmogorov-Smirnov test
```

Ο έλεγχος Kolmogorov-Smirnov έδειξε ότι απορρίπτουμε την μηδενική υπόθεση, επομένως η mmse3 δεν ακολουθεί κανονική κατανομή (οπότε πάμε σε μη παραμετρικό έλεγχο).

```
> wilcox.test(mmse3 ~ sex, data = dat1)

Wilcoxon rank sum test with continuity correction

data:  mmse3 by sex
W = 1354.5, p-value = 0.7799
alternative hypothesis: true location shift is not equal to 0
```

Ο έλεγχος Mann-Whitney έδειξε ότι $p\text{-value}=0.78 > 0.05$ άρα δεν απορρίπτουμε την μηδενική υπόθεση συνεπώς η βελτίωση της μέσης τιμής mmse δεν σχετίζεται με τη μεταβλητή sex.

Απάντηση στο ερώτημα 5:

Για να εξετάσουμε τη συσχέτιση μεταξύ της μεταβλητής ae2 και της μεταβλητής sex θα προχωρήσουμε σε χ^2 έλεγχο ανεξαρτησίας.

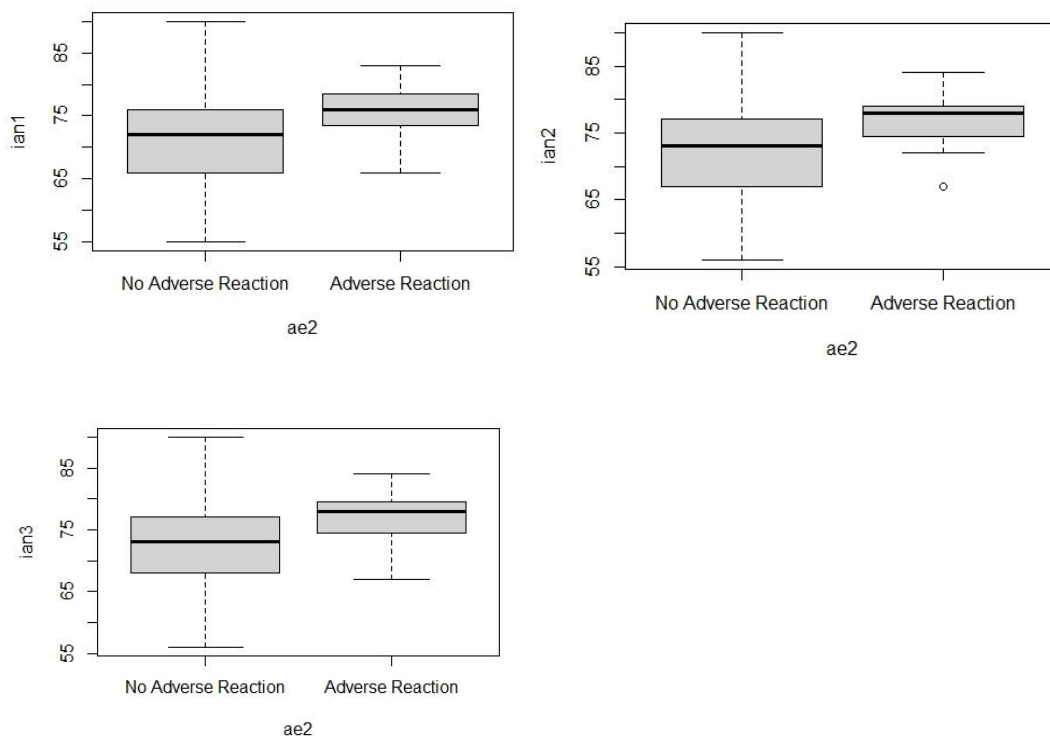
```
> chisq.test(dat1$ae2, dat1$sex)

Pearson's Chi-squared test with Yates' continuity correction

data:  dat1$ae2 and dat1$sex
X-squared = 7.2392e-31, df = 1, p-value = 1
```

Ο έλεγχος με μηδενική υπόθεση H_0 : οι μεταβλητές *ae2* και *sex* είναι ανεξάρτητες έδωσε $p\text{-value}=1$ άρα δεν απορρίπτουμε την μηδενική υπόθεση. Συνεπώς η εμφάνιση ανεπιθύμητων ενεργειών δεν σχετίζεται με το φύλο των ασθενών.

Τώρα για να δούμε αν σχετίζεται η μεταβλητή *ae2* με τις συνεχείς μεταβλητές ηλικίας *ian1*, *ian2*, *ian3*, θα δούμε αρχικά τα boxplots.



Από τα διαγράμματα παίρνουμε ότι δεν φαίνεται να υπάρχουν σημαντικές διαφορές της μεταβλητής *ian1*, *ian2* και *ian3* στα επίπεδα της μεταβλητής *ae2*.

Για να ελέξουμε τώρα με μεταβλητή απόκρισης (response) την *ae2* θα εφαρμόσουμε τρία μοντέλα λογιστικής παλινδρόμησης της μορφής $\text{logit}(p) = \beta_0 + \beta_1 X + \varepsilon$

```
> model51 <- glm (ae2 ~ ian1, binomial, data=dat1);summary(mod
el51)

Call:
glm(formula = ae2 ~ ian1, family = binomial, data = dat1)
```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.68830    3.68422  -2.901  0.00372 **
ian1         0.12095    0.04895   2.471  0.01348 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 86.124  on 104  degrees of freedom
Residual deviance: 79.166  on 103  degrees of freedom
AIC: 83.166

Number of Fisher Scoring iterations: 5

```

Παρατηρούμε ότι για τον συντελεστή $\beta_1 = 0.12$ πήραμε $p\text{-value} = 0.014 < 0.05$ άρα σε ε.σ. 5% απορρίπτουμε την μηδενική υπόθεση $H_0 : \beta_1 = 0$ συνεπώς η μεταβλητή ae2 μπορεί να προβλεφθεί με χρήση της ηλικίας εμφάνισης πρώτων συμπτωμάτων άνοιας.

Με όμοιο τρόπο παίρνουμε και τα υπόλοιπα μοντέλα.

```

> model52 <- glm (ae2 ~ ian2, binomial, data= dat1); summary(m
ode152)

Call:
glm(formula = ae2 ~ ian2, family = binomial, data = dat1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.10015    3.82140  -2.905  0.00368 **
ian2         0.12486    0.05008   2.493  0.01266 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 86.124  on 104  degrees of freedom
Residual deviance: 78.877  on 103  degrees of freedom
AIC: 82.877

Number of Fisher Scoring iterations: 5

```

Πήραμε $p\text{-value} = 0.012 < 0.05$ άρα σε ε.σ. 5% απορρίπτουμε την μηδενική υπόθεση. Συνεπώς, μπορούμε να προβλέψουμε την ύπαρξη ανεπιθύμητων ενεργειών με χρήση της ηλικίας διάγνωσης άνοιας.

Τέλος για την ae2 και ian3:

```

> model53 <- glm (ae2 ~ ian3, binomial, data = dat1); summary(
model53)

Call:
glm(formula = ae2 ~ ian3, family = binomial, data = dat1)

```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.28320    3.65953  -2.810  0.00495 **
ian3         0.11348    0.04781   2.374  0.01761 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 86.124  on 104  degrees of freedom
Residual deviance: 79.753  on 103  degrees of freedom
AIC: 83.753

Number of Fisher Scoring iterations: 5

```

Πήραμε $p\text{-value} = 0.018 < 0.05$, άρα απορρίπτουμε την μηδενική υπόθεση σε ε.σ. 5%. Επομένως, η ύπαρξη ανεπιθύμητων ενεργειών μπορεί να προβλεφθεί με logit μοντέλο από την ηλικία έναρξης χορήγησης αρχικής θεραπείας.

Απάντηση στο ερώτημα 6:

Για να εξετάσουμε την σχέση μεταξύ της εμφάνισης ανεπιθύμητων ενεργειών (response variable η ae2) και της τιμής της mmse1 (independent variable) πριν δοθεί το φάρμακο, θα εφαρμόσουμε ένα μοντέλο λογιστικής παλινδρόμησης και θα εξετάσουμε εάν είναι στατιστικά σημαντικό.

```

> model6 <- glm(ae2 ~ mmse1, binomial, data = dat1) ; summary(
model6)

Call:
glm(formula = ae2 ~ mmse1, family = binomial, data = dat1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.19955    2.13425  -0.562    0.574
mmse1       -0.03037    0.10891  -0.279    0.780

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 86.124  on 104  degrees of freedom
Residual deviance: 86.046  on 103  degrees of freedom
AIC: 90.046

Number of Fisher Scoring iterations: 4

```

Ελέγχουμε το μοντέλο $\text{logit}(p) = \beta_0 + \beta_1 X + \varepsilon$

Για τον συντελεστή της μεταβλητής mmse1 πήραμε $p\text{-value}=0.780$ άρα σε επίπεδο σημαντικότητας 5% δεν απορρίπτουμε την μηδενική υπόθεση $H_0 : \beta_1 = 0$, και συνεπώς η εμφάνιση ανεπιθύμητων ενεργειών μεταξύ των επισκέψεων 1 και 2, δεν σχετίζεται με την τιμή της mmse πριν δοθεί το φάρμακο.

Απάντηση στο ερώτημα 7:

Αφού πρόκειται για δύο κατηγορικές μεταβλητές, θα εφαρμόσουμε τον έλεγχο χ^2 για την ανεξαρτησία των μεταβλητών με μηδενική υπόθεση:

$$H_0 : \text{Οι δύο μεταβλητές είναι ανεξάρτητες}$$

```
> chisq.test(dat1$ae2, dat1$dose1)

Pearson's Chi-squared test

data:  dat1$ae2 and dat1$dose1
X-squared = 6.0667, df = 2, p-value = 0.04815
```

Πήραμε p-value=0.05 άρα δεν απορρίπτουμε την μηδενική υπόθεση σε ε.σ. 5%, δηλαδή η εμφάνιση ανεπιθύμητων ενεργειών μεταξύ των επισκέψεων 1 και 2 δεν σχετίζεται με τη δόση του φαρμάκου που πήραν οι ασθενείς.

ΑΣΚΗΣΗ 2 (Υπολογισμός μεγέθους δείγματος):

Θέτουμε groupmeans = c(1.5, 3, 3) , var = 5, α = 0.05, power=0.8, και τέλος groups = 3.

```
> groupmeans = c(1.5, 3, 3)
> power.anova.test(groups = 3, between.var = var(groupmeans),
+                   within.var = 5, sig.level = 0.05, power=0.8
+ )

Balanced one-way analysis of variance power calculation

      groups = 3
         n = 33.1349
between.var = 0.75
within.var  = 5
sig.level   = 0.05
power       = 0.8

NOTE: n is number in each group
```

Πήραμε ότι για τα δεδομένα που έχουμε, για να επιτύχουμε ισχύ του ελέγχου ίση με 80%, το μέγεθος σε κάθε ομάδα πρέπει να είναι 33 ασθενείς, άρα συνολικά $3 \times 33 = 99$ άτομα.