

Εφαρμοσμένη στατιστική: Ανάλυση Διακύμανσης: Παράδειγμα 1.

Ξεκινάμε το πρώτο παράδειγμα ανάλυσης διακύμανσης με τα δεδομένα PlantGrowth που υπάρχουν ήδη στην R. Καλό είναι να δημιουργήσουμε ένα αντίγραφο ασφαλείας των δεδομένων και να τα φέρουμε σε μορφή έτοιμη για ανάλυση:

```
plant.df <- PlantGrowth
plant.df$group <- factor(plant.df$group, labels = c("Control", "Treatment
1", "Treatment 2"))
```

Το όρισμα labels δίνει τα επίπεδα του παράγοντα group.

Ένας πρώτος γραφικός έλεγχος των δεδομένων γίνεται με boxplot:

```
boxplot(weight~group, data=plant.df) # boxplot
```

Boxplot of Plant Growth by Treatment Group

Το boxplot δείχνει να υπάρχουν διαφορές αλλά δεν μπορούμε να πούμε αν οι διαφορές είναι στατιστικά σημαντικές ή όχι. Για να εξερευνήσουμε στατιστικά τη διαφορά θα χρησιμοποιήσουμε την ανάλυση της διακύμανσης. Ξεκινάμε με την συνάρτηση `lm`:

```
plant.mod1 = lm(weight ~ group, data = plant.df)
```

οώσαμε το μοντέλο σε μια μεταβλητή προκειμένου να μπορούμε να κάνουμε πράξεις.

```
> summary(plant.mod1)
Call:
lm(formula = weight ~ group, data = plant.df)

Residuals:
    Min       1Q   Median       3Q      Max
-1.0710 -0.4180 -0.0060  0.2627  1.3690

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.0320     0.1971  25.527 <2e-16 ***
groupTreatment 1  -0.3710     0.2788  -1.331  0.1944
groupTreatment 2   0.4940     0.2788   1.772  0.0877 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ΤΟ ΜΟΝΤΕΛΟ ΔΙΝΕΙ ΤΗΝ ΕΚΤΙΜΗΣΗ ΤΟΥ μ_1 , (ΠΡΩΤΗ ΓΡΑΜΜΗ), $\mu_2-\mu_1$, (ΔΕΥΤΕΡΗ ΓΡΑΜΜΗ), $\mu_3-\mu_1$, (ΤΡΙΤΗ ΓΡΑΜΜΗ),

```
Residual standard error: 0.6234 on 27 degrees of freedom
Multiple R-squared:  0.2641,    Adjusted R-squared:  0.2096
F-statistic: 4.846 on 2 and 27 DF,  p-value: 0.01591
```

Το μοντέλο δείχνει κάποια ένδειξη διαφοράς στην κατά μέσο όρο αύξηση για το 2^ο παράγοντα σε σχέση με τον πρώτο (control). Η ανάλυση διακύμανσης μπορεί να γίνει με την εντολή

```
> anova(plant.mod1)
Analysis of Variance Table

Response: weight
      Df Sum Sq Mean Sq F value Pr(>F)
group   2  3.7663  1.8832  4.8461 0.01591 *
Residuals 27 10.4921  0.3886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F test – σελίδα 11,
εβδομάδα 11.

Ο πίνακας επιβεβαιώνει την ύπαρξη διαφορών μεταξύ των group, όπως είδαμε και στο μοντέλο. Με τη συνάρτηση `confint` μπορούμε να υπολογίσουμε διαστήματα εμπιστοσύνης για τη μέση τιμή και τις διαφορές από τα υπόλοιπα επίπεδα του παράγοντα:

Εφαρμοσμένη στατιστική: Ανάλυση Διακύμανσης: Παράδειγμα 1.

```
> confint(plant.mod1)
                2.5 %    97.5 %
(Intercept)    4.62752600 5.4364740
groupTreatment 1 -0.94301261 0.2010126
groupTreatment 2 -0.07801261 1.0660126
```

Για να απεικονίσουμε γραφικά τους δείκτες του μοντέλου:

```
plant.mod = data.frame(Fitted = fitted(plant.mod1), Residuals =
  resid(plant.mod1), Treatment = plant.df$group)
and then produce the plot:

plot(plant.mod$Fitted, plant.mod$Residuals)
Residuals diagnostic plot
```

Αυτό που δείχνει το γράφημα είναι ότι δεν υπάρχει κάποιο βασικό πρόβλημα με τα υπόλοιπα του μοντέλου, αλλά κάποια ένδειξη για διαφορές στις μεταβολές των υπολοίπων στα 3 group.