Χρήσιμες Εντολές της R

Εξισώσεις με λέξεις

Y = X + άλλοι παράγοντες

Βασικά

```
print("Hello world!")

# ανάθεση τιμής σε αντικείμενο
my_number <- 5

# διάνυσμα
 my_vector <- c(1, 2, 3)

# πρώτο στοιχείο διανύσματος
my_vector[1]

# ταξινόμηση τιμών σε διάνυσμα
 sort(my_vector)

# αριθμητικές πράξεις
sum(1, 2, 100), +, -, *, /
sqrt(157)
abs(data_set$Y)

# λογικοί τελεστές
>, <, >=, <=, ==, !=, |, &

# το αποτέλεσμα είναι μια νέα
μεταβλητή με τιμές TRUE ή FALSE
data_set$C <- data_set$A > data_set$B
```

Κατανομή Πιθανότητας

```
# υπολογισμός P( X <= 65.1)
xpnorm(65.1, data_set$mean, data_set$sd)

# κατανομή τιμών z
zscore(data_set$Y)

# επιστρέφει την τιμή του ελέγχου t
qt(.975, df = 999)
# επιστρέφει την τιμή του ελέγχου F
qf(.95, df1 = 1, df2 = 100)

# διάστημα εμπιστοσύνης της κατανομής t
confint(empty_model)

# υπολογίζει τιμή p για την κατανομή F
xpf(sample_f, df1 = 2 , df2 = 10)
```

Πίνακες Σύνοψης Τιμών

```
# υπολογίζει την σύνοψη των πέντε αριθμών
favstats(~ Y, data = data_set)

# δημιουργεί πίνακα συχνοτήτων
table(data_set$Y)

# δημιουργεί πίνακα σχετικών συχνοτήτων
prop.table(table(data_set$Y))*100

#  δημιουργεί πίνακα διπλής εισόδου
table(data_set$Y, data_set$X)
```

Στατιστικοί δείκτες

```
mean(data_set$Y)
var(data_set$Y)
sd(data_set$Y)

cohensD(Y ~ X, data = data_set)
cor(Y ~ X, data = data_set)

b1(Y ~ X, data = data_set)
b1(one_model)

pre(Y ~ X, data = data_set)
f(Y ~ X, data = data_set)

f(Y ~ X1 + X2,
data = data_set, predictor = ~X2)
```

Πλαίσιο Δεδομένων

```
# δομή πλαισίου δεδομένων
str(data_set)

# εμφάνιση πρώτων και τελευταίων έξι γραμμών
head(data_set)
tail(data_set)

# επιλογή μεταβλητών
select(data_set, Y1, Y2)

# πρώτες έξι γραμμές επιλεγμένων
μεταβλητών
head(select(data_set, Y1, Y2))

# πρόσβαση σε μεταβλητή
data_set$Y

# επιλογή γραμμών που ικανοποιούν μια συνθήκη
data_set[data_set$Y > 40]
filter(data_set, Y > 300)

# επιλογή γραμμών χωρίς απούσες τιμές (NA)
filter(data_set, is.na(Y) == FALSE)
filter(data_set, !is.na(Y))

# ταξινόμηση γραμμών με βάση μεταβλητή
arrange(data_set, Y)

# δημιουργία πλαισίου δεδομένων από αρχείο
data_set <- read.csv("file_name", header = TRUE)

# μετατροποή ποσοτικής μεταβλητής
# σε ποιοτική
factor(data_set$Y)
factor(data_set$Y,
  levels = c(1,2),
  labels = c("A", "B"))

# αλλαγή κωδικοποίησης
recode(data_set$Y, "0" = 0, "1" = 50, "2" = 100)

# δημιουργία δύο ομάδων ίδιου μεγέθους
ntile(data_set$Y, 2)

# μετατροπή ποιοτικής μεταβλητής
# σε ποσοτική
as.numeric(data_set$Y)
```

Χρήσιμες Εντολές της R

Προσομοίωση

```
# δειγματοληψία με επανατοποθέτηση
sample(data_set, 6)

# δειγματοληψία με επανατοποθέτηση
resample(data_set, 10)

do(3) * resample (data_set, 10)

# τυχαίο ανακάτεμα τιμών
shuffle(data_set$Y)

# προσομοίωση 10000 τιμών
# από την κανονική κατανομή
sim_Y <- rnorm(10000, Y_stats$mean,
Y_stats$sd)

# αποθήκευση προσομοιωμένων τιμών
σε πλαίσιο δεδομένων
data_set<-data.frame(sim_Y)

# προσομοίωση
# δειγματοληπτικής κατανομής μέσων
όρων
sdom_sim <- do(10000) * mean(rnorm(157,
Y_stats$mean, Y_stats$sd))

# προσομοίωση bootstrap
# δειγματοληπτικής κατανομής
μέσων όρων
sdom_boot <- do(10000) *
mean(resample(data_set$Y, 157))

# τυχαιοποίηση της δειγματοληπτικής
κατανομής των b1, κεντραρισμένη στο 0
sdob1 <- do(1000) *
  b1(shuffle(Y) ~ X, data = data_set)

# δειγματοληπτική κατανομή bootstrap των
b1 κεντραρισμένη στο δειγματικό b1
sdob1_boot <- do(1000) *
  b1(Y ~ X, data = resample(data_set))

# καταμέτρηση του αριθμού των
ακραίων τιμών b1
tally(sdob1$b1 > sample_b1 |
sdob1$b1 < -sample_b1)

#  επιστροφή  TRUE για  το  μεσαίο 95% της
κατανομής
 middle(sdob1$b1, .95)

# τυχαιοποίηση δειγματοληπτικής κατανομής PRE
sdopre <- do(1000) * pre(shuffle(Y) ~ X, data =
data_set)

# randomize sampling distribution of Fs
sdof <- do(1000) *
  f(shuffle(Y) ~ X, data = data_set)

# καταμέτρηση ακραίων τιμών F
sample_f <- f(shuffle(Y) ~ X, data = data_set)
tally(~f > sample_f, data = sdof)
```

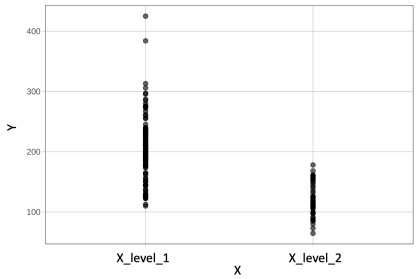
Προσαρμογή και Αξιολόγηση Μοντέλων

```
# κενό μοντέλο
empty_model <- lm(Y ~ NULL,
data = data_set)

# χρήση μιας ανεξάρτητης μεταβλητής
one_model <- lm(Y ~ X, data = data_set)

# χρήση περισσότερων ανεξάρτητων μεταβλητών
# σε πολυμεταβλητό μοντέλο
multi_model <- lm(Y ~ X1 + X2, data = data_set)

# όλες οι συγκρίσεις μοντέλων που μπορού να
# γίνουν σε σχέση με το πολυμεταβλητό μοντέλο
generate_models(multi_model)

# προβλέψεις μοντέλου και υπόλοιπα
data_set$empty_predict <- predict(empty_model)
data_set$empty_resid <- resid(empty_model)

# πίνακας ANOVA
anova(empty_model)
supernova(one_model)

# έλεγχος t, με συγκεντρωτική (pooled) διακύμανση
t.test(Tip ~ Condition, data = data_set,
var.equal=TRUE)

# διόρθωση πολλαπλών συγκρίσεων:
# "Tukey","Bonferroni","none"
pairwise(one_model, correction = "none")
```

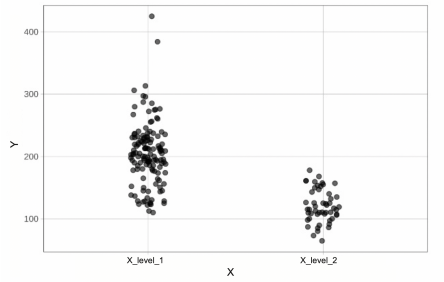
Χρήσιμες Εντολές της R

Διαγράμματα

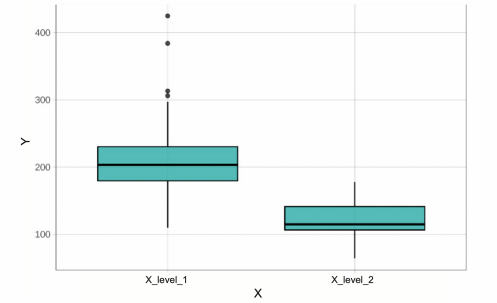
```
gf_point(Y ~ X, data = data_set)
```



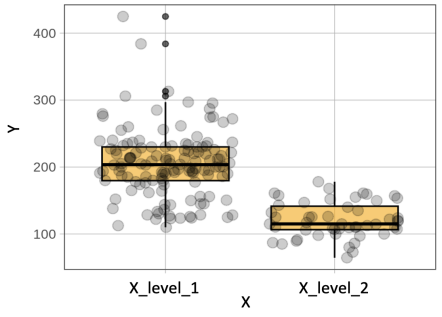

```
gf_jitter(Y ~ X, data = data_set)
```



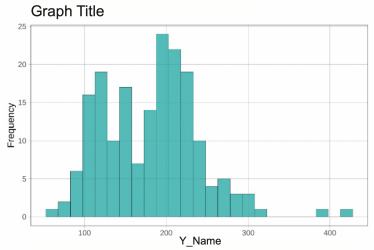

```
gf_boxplot(Y ~ X, data = data_set)
```



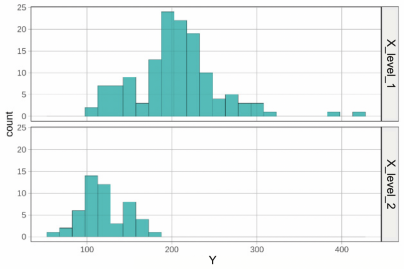

```
gf_boxplot(Y ~ X, data = data_set, fill =
"orange") %>%
 gf_jitter(height = 0, alpha = .2, size = 3)
```



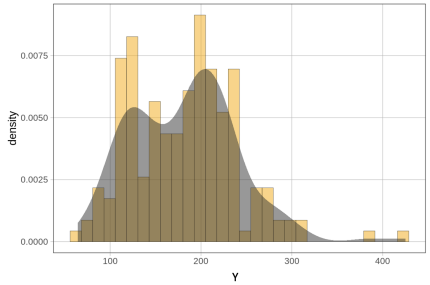

```
gf_histogram(~ Y, data = data_set) %>%
  # αλλαγή τίτλων
  gf_labs(title = "Graph Title",
  x = "Y_Name", y = "Frequency")
```



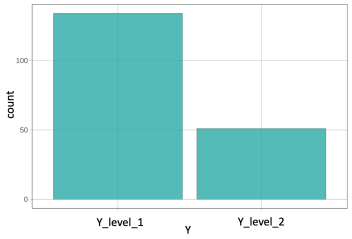

```
# διαιρεμένο ιστόγραμμα
gf_histogram(~ Y, data = data_set) %>%
  gf_facet_grid(X ~ .)
```



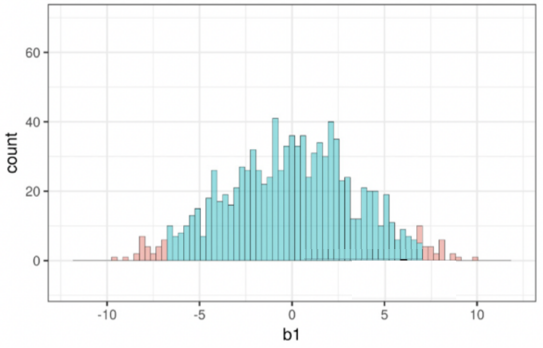

```
gf_dhistogram(~ Y, data = data_set,
fill = "orange") %>%
  gf_density()
```



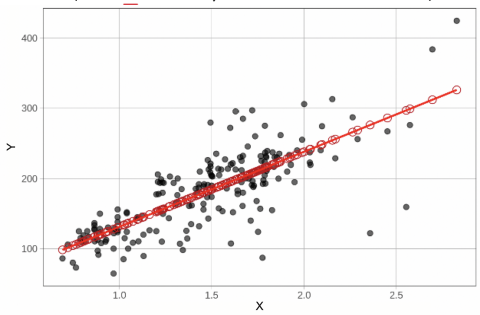

```
gf_bar( ~ Y, data = data_set)
```




```
# δειγματοληπτική κατανομή b1
gf_histogram(~b1, data = sdob1,
 fill = ~middle(b1, .95)) %>%
 # τροποποίηση αξόνων x και y
 gf_lims(x = c(-12, 12), y = c(0, 70))
```




```
gf_point(Y ~ X, data = data_set) %>%
 # προβολή προβλέψεων μοντέλου
 gf_point(Y ~ X , shape = 1, size = 3,
color = "firebrick") %>%
 # προσαρμογή μοντέλου με κόκκινη γραμμή
 gf_model(one_model, color = "red")
```




```
pairwise(one_model, plot = TRUE)
```