**ΑΣΚΗΣΗ 1| ΕΡΓΑΣΙΑ 1**

Φτιάχνουμε το παρακάτω διάνυσμα(ν) με τις τιμές του πλαισίου :

v<-c(73,6,77,81,91,101,135,61,65,68,18,20,23,12,14,18,23,26,26,27,2,3,3,40,41,41,6,8,11,12,37,38,38,6,73,6,51)

**Ερώτηση (i)**

**>mean(v)** :Μέση Τιμή

[1] 37.32432

**>median(v)** :Διάμεσος

[1] 26

**>max(v)-min(v)** :Δειγματική Έκταση ή Εύρος

[1] 133

**>var(v)** :Διασπορά

[1] 1047.836

**>as<-sqrt(var(v))^(-3)\*(mean((v[]-mean(v))^3)); as** :Συντελεστής Ασσυμετρίας

[1] 1.020921

**>quantile(v,0.25) ; quantile(v,0.75)** :Πρώτο & Τρίτο Τεταρτημόριο

25%

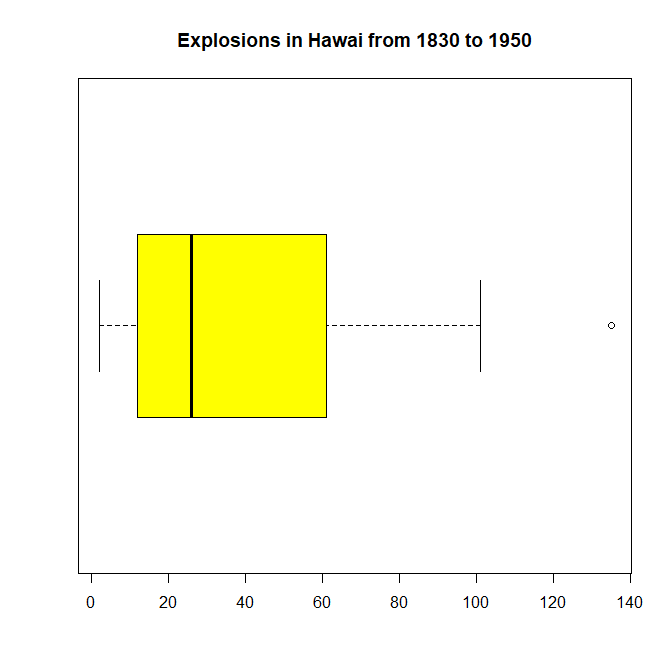
12

75%

61

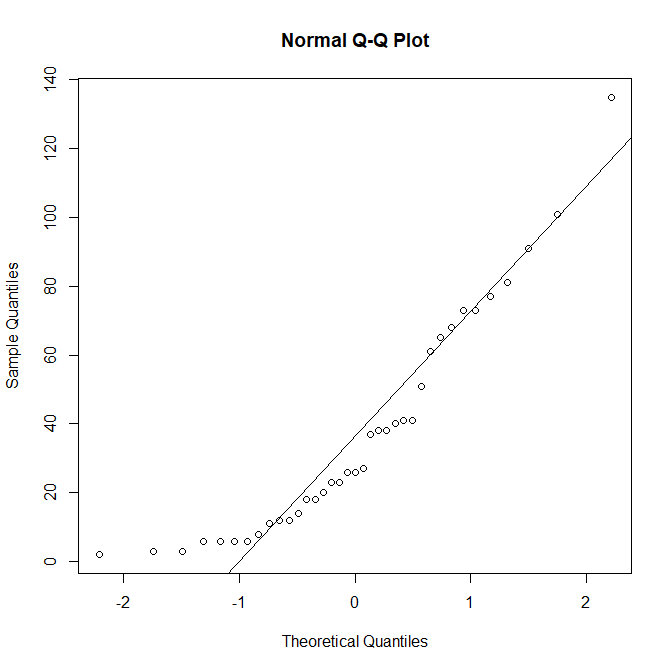
**Ερώτηση (ii)**

**>boxplot(v,main="Explosions in Hawain from 1830 to 19550",col=7,horizontal=T)** :Θηκόγραμμα



Από το παραπάνω θηκόγραμμα συμπεραίνουμε ότι τα δεδομένα δεν προέρχονται από κανονικό πληθυσμό αφού η διάμεσος τείνει προς το πρώτο τεταρτημόριο ,το εύρος των τιμών στα δύο ακραία τεταρτημόρια διαφέρει ενώ ταυτόχρονα παρατηρούμε την ύπαρξη ακραίας τιμής.

**>qqnorm(v) ; qqline(v)** :Q-Q Διάγραμμα



\*Οι αποκλίσεις των παρατηρήσεων δεν είναι πολύ μεγάλες αλλά ταυτόχρονα ούτε πολύ μικρές από την qqline οπότε δεν είναι ξεκάθαρο ότι τα δεδομένα προέρχονται από κανονικό πληθυσμό.

**Ερώτηση (iii)**

**>library(nortest)**

**> lillie.test(v)**

Lilliefors (Kolmogorov-Smirnov) normality test

data: v

D = 0.16566, p-value = 0.01178

**> shapiro.test(v)**

Shapiro-Wilk normality test

data: v

W = 0.88729, p-value = 0.001332

* Σύμφωνα με το **Lillie Test** το δείγμα είναι κανονικό αφού **p-value > a=0.01**
* Σύμφωνα με το **Shapiro Test** δεν είναι κανονικό το δείγμα αφού **p-value < a=0.01**

**Ερώτηση (iv)**

Έχουμε : Ε(Χ)=ab (1) & E(X2)=ab2 + a2b2 (2)

Επίσης ξέρουμε ότι : Var(X) = E(X2) – (E(X))2  (3)

1. ⬄ E(X)=ab⬄(E(X))2=a2b2

Από (1) και (2) έχουμε : Var(X) = ab2 + a2b2-a2b2 = ab2 ⬄ 1047.836=ab2

Επίσης έχουμε: E(X)=ab ⬄ 37.32=ab ⬄ a = 37.32/b

Άρα 1047.836=37.32/b \* b2 ⬄ 1047.836 = 37.32\*b ⬄ **b = 28,077**

Για b=28,077 έχουμε από την (1) ⬄ **a=1,33**

**> class<-seq(0,160,20)**

**> hist(v,breaks=class,prob=TRUE)** :Κατασκευή Ιστογράμματος

**> a<-1.3**

**> b<-28.07**

**> k<-rgamma(v,shape=a,scale=b);k** :Προσαρμογή κατάλληλης Γάμμα κατανομής

[1] 2.505366 10.585107 21.844865 77.568683 32.803134 76.178986 16.629462 64.323728 25.900052 9.927396 38.717604

[12] 21.267661 2.421219 61.593472 45.029441 53.040752 81.220729 42.734144 36.398283 7.782932 114.775343 29.822177

[23] 23.428128 15.922063 46.321339 28.909592 17.317530 51.045795 9.728319 95.347971 35.899332 18.634491 22.211704

[34] 20.637531 44.779196 30.211804 30.914781

**> lines(density(k),col="green")** :Καμπύλη Πυκνότητας Πιθανότητας

****

**Ερώτηση (v)**

**> l<-0.028**

**> x1<-qexp(ppoints(v),l)**

**> qqplot(x1,v,main="Exponential Distribution with rate=0.028")**

**> a1<-qexp(0.25,l) ; b1<-quantile(v,0.25)**

**> a2<-qexp(0.75,l) ; b2<-quantile(v,0.75)**

**> abline(b1-a1\*((b2-b1)/(a2-a1)),(b2-b1)/(a2-a1))**



Σύμφωνα με το διάγραμμα **qq** τα δεδομένα ακολουθούν την εκθετική κατανομή.

**> ks.test(v,"pexp",rate=l,exact=TRUE**)

One-sample Kolmogorov-Smirnov test

data: v

D = 0.10459, p-value = 0.7745

alternative hypothesis: two-sided

**> ks.test(v,"pexp",rate=l,exact=FALSE)**

One-sample Kolmogorov-Smirnov test

data: v

D = 0.10459, p-value = 0.8131

alternative hypothesis: two-sided

Για α=0.05 : και στους δύο ελέγχους γίνεται δεκτή η υπόθεση ότι τα δεδομένα προέρχονται από **εκθετική κατανομή με παράμετρο λ=0.028** αφού p-value > a .

**Ερώτηση (vi)**

**> wilcox.test(v,mu=20,exact=FALSE,correct=FALSE)**

Wilcoxon signed rank test

data: v

V = 500, p-value = 0.008671

alternative hypothesis: true location is not equal to 20

**> library(BSDA)**

**> SIGN.test(v,md=20)**

One-sample Sign-Test

data: v

s = 22, p-value = 0.243

alternative hypothesis: true median is not equal to 20

95 percent confidence interval:

18.00000 40.94273

sample estimates:

median of x

26

Achieved and Interpolated Confidence Intervals:

Conf.Level L.E.pt U.E.pt

Lower Achieved CI 0.9011 18 40.0000

Interpolated CI 0.9500 18 40.9427

Upper Achieved CI 0.9530 18 41.0000

Σύμφωνα με το **Wilcox.test** η διάμεσος δεν είναι ίση με 20 ενώ σύμφωνα με το **SΙGN.test** είναι .

**ΑΣΚΗΣΗ 2 | ΕΡΓΑΣΙΑ 1**

**> p<-0.1**

**> n<-20**

**> k<-0:n**

**> pM0<-0.05**

**Ερώτηση Α(i)**

**> f<-dgeom(k,prob=p) ; f** : Συνάρτηση πιθανότητας για τη γεωμετρική κατανομή

[1] 0.10000000 0.09000000 0.08100000 0.07290000 0.06561000 0.05904900 0.05314410 0.04782969 0.04304672 0.03874205 0.03486784

[12] 0.03138106 0.02824295 0.02541866 0.02287679 0.02058911 0.01853020 0.01667718 0.01500946 0.01350852 0.01215767

**> cT<-1/(1-f[1]) ; pT<-c(0,cT\*f[-1]); pT** :Αποκομμένη στο 0

[1] 0.00000000 0.10000000 0.09000000 0.08100000 0.07290000 0.06561000 0.05904900 0.05314410 0.04782969 0.04304672 0.03874205

[12] 0.03486784 0.03138106 0.02824295 0.02541866 0.02287679 0.02058911 0.01853020 0.01667718 0.01500946 0.01350852

**Ερώτηση Α(ii)**

**cM<-(1-pM0)/(1-f[1]);pM<-c(pM0,cM\*f[-1]);pM** : Τροποποιημένη στο 0

[1] 0.05000000 0.09500000 0.08550000 0.07695000 0.06925500 0.06232950 0.05609655 0.05048690 0.04543821 0.04089438 0.03680495

[12] 0.03312445 0.02981201 0.02683081 0.02414773 0.02173295 0.01955966 0.01760369 0.01584332 0.01425899 0.01283309

**>m<-cbind(k,pT,pM,f)** : Κατασκευή Πίνακα

**> rownames(m)<-rep("",nrow(m))**

**> colnames(m)<-c("k","Truncated Geometric","Modified Geometric","Geometric")**

**> round(m,digits=7)** : Στρογγυλοποίηση στα 7 δεκαδικά ψηφία

k Truncated Geometric Modified Geometric Geometric

0 0.0000000 0.0500000 0.1000000

1 0.1000000 0.0950000 0.0900000

2 0.0900000 0.0855000 0.0810000

3 0.0810000 0.0769500 0.0729000

4 0.0729000 0.0692550 0.0656100

5 0.0656100 0.0623295 0.0590490

6 0.0590490 0.0560966 0.0531441

7 0.0531441 0.0504869 0.0478297

8 0.0478297 0.0454382 0.0430467

9 0.0430467 0.0408944 0.0387420

10 0.0387420 0.0368049 0.0348678

11 0.0348678 0.0331245 0.0313811

12 0.0313811 0.0298120 0.0282430

13 0.0282430 0.0268308 0.0254187

14 0.0254187 0.0241477 0.0228768

15 0.0228768 0.0217330 0.0205891

16 0.0205891 0.0195597 0.0185302

17 0.0185302 0.0176037 0.0166772

18 0.0166772 0.0158433 0.0150095

19 0.0150095 0.0142590 0.0135085

20 0.0135085 0.0128331 0.0121577

**> x<-seq(0,20,1)**

**>plot(x,xlab="x",ylab="f(x)",xlim=c(0,20),ylim=c(0,0.1),main="Geometric,Truncated and Modified Geometric,p=0.1,pM0=0.05")** : Άξονες

**> points(pT,pch=3)** : Σημεία αποκομμένης

**> points(pM,pch=4)** : Σημεία Τροποποιημένης

**> points(f)** : Σημεία Γεωμετρικής

**> legend(15,0.08,c("Truncated","Modified","Geometric"),pch=c(3,4,1))**

****

**Ερώτηση Β**

**> fx1<-pM**

**> fx2<-convolve(fx1,rev(fx1),type="o")**

**> fx3<-convolve(fx1,rev(fx2),type="o")**

**> fx4<-convolve(fx1,rev(fx3),type="o")**

**> fx5<-convolve(fx1,rev(fx4),type="o")**

**> fx6<-convolve(fx1,rev(fx5),type="o")**

**> round(fx6,digits=7)**

[1] 0.0000000 0.0000002 0.0000010 0.0000038 0.0000110 0.0000263 0.0000544 0.0001009 0.0001717 0.0002728 0.0004101 0.0005886

[13] 0.0008128 0.0010860 0.0014105 0.0017874 0.0022167 0.0026973 0.0032269 0.0038024 0.0044199 0.0050747 0.0057613 0.0064735

[25] 0.0072042 0.0079449 0.0086865 0.0094193 0.0101332 0.0108181 0.0114645 0.0120633 0.0126065 0.0130871 0.0134992 0.0138385

[37] 0.0141015 0.0142864 0.0143923 0.0144198 0.0143703 0.0142461 0.0140507 0.0137881 0.0134630 0.0130809 0.0126481 0.0121711

[49] 0.0116568 0.0111122 0.0105444 0.0099601 0.0093659 0.0087681 0.0081722 0.0075836 0.0070067 0.0064457 0.0059039 0.0053844

[61] 0.0048894 0.0044207 0.0039797 0.0035672 0.0031836 0.0028290 0.0025030 0.0022051 0.0019341 0.0016891 0.0014687 0.0012713

[73] 0.0010956 0.0009400 0.0008027 0.0006824 0.0005774 0.0004862 0.0004075 0.0003398 0.0002820 0.0002329 0.0001913 0.0001563

[85] 0.0001271 0.0001028 0.0000827 0.0000661 0.0000525 0.0000415 0.0000326 0.0000254 0.0000197 0.0000152 0.0000116 0.0000088

[97] 0.0000066 0.0000049 0.0000037 0.0000027 0.0000020 0.0000014 0.0000010 0.0000007 0.0000005 0.0000003 0.0000002 0.0000002

[109] 0.0000001 0.0000001 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000

[121] 0.0000000

**> sum(fx6[1:7])**

[1] 9.678496e-05

**ΑΣΚΗΣΗ 3 | ΕΡΓΑΣΙΑ 1**

**> a<-read.table("Askisi3.txt",header=TRUE) ; a** : Εισαγωγή αρχείου Άσκηση3

Y X1 X2 X3

1 0.90 72.4 76.3 29.18

2 0.91 41.6 70.3 29.35

3 0.96 34.3 77.1 29.24

4 0.89 35.1 68.0 29.27

5 1.00 10.7 79.0 29.78

6 1.10 12.9 67.4 29.39

7 1.15 8.3 66.8 29.69

8 1.03 20.1 76.9 29.48

9 0.77 72.2 77.7 29.09

10 1.07 24.0 67.7 29.60

11 1.07 23.2 76.8 29.38

12 0.94 47.4 86.6 29.35

13 1.10 31.5 76.9 29.63

14 1.10 10.6 86.3 29.56

15 1.10 11.2 86.0 29.48

16 0.91 73.3 76.3 29.40

17 0.87 75.4 77.9 29.28

18 0.78 96.6 78.7 29.29

19 0.82 107.4 86.8 29.03

20 0.95 54.9 70.9 29.37

**> attach(a)**

**Ερώτηση (i)**

**> lm(Y~X1+X2+X3)**

Call:

lm(formula = Y ~ X1 + X2 + X3)

Coefficients:

(Intercept) X1 X2 X3

-3.5077781 -0.0026250 0.0007989 0.1541550

Άρα η προσαρμοσμένη εξίσωση παλινδρόμησης είναι :

**Y = -3.5077781 – 0.0026250X1 + 0.0007989X2 + 0.1541550X3**

**Ερώτημα (ii)**

**> m<-lm(Y~X1+X2+X3)**

**> summary(m)**

Call:

lm(formula = Y ~ X1 + X2 + X3)

Residuals:

Min 1Q Median 3Q Max

-0.11799 -0.02526 0.01345 0.04103 0.06523

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.5077781 3.0048641 -1.167 0.26017

X1 -0.0026250 0.0006549 -4.008 0.00101 \*\*

X2 0.0007989 0.0020451 0.391 0.70121

X3 0.1541550 0.1013675 1.521 0.14784

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05617 on 16 degrees of freedom

Multiple R-squared: 0.8005, Adjusted R-squared: 0.763

F-statistic: 21.4 on 3 and 16 DF, p-value: 7.609e-06

Επομένως αφού **p-value = 7.609e-06 < a = 0.0**5 η παλινδρόμηση είναι στατιστικά σημαντική.

**Ερώτηση (iii)**

Για επίπεδο σημαντικότητας **a=0.15** έχουμε :

Από την ερώτηση (ii) παίρνουμε από τα αποτελέσματα της εντολής **summary(m)** τη στήλη **Pr(>|t|)** η οποία δίνει την **p-value** για κάθε μεταβλητή.

Έτσι έχουμε ότι η Χ1 και Χ2 είναι στατιστικά σημαντικές αφού p-value = 0.7 > a=0.15.

**Ερώτηση (iv)**

**> confint(m,level=0.99) :** Υπολογισμός διαστημάτων εμπιστοσύνης

0.5 % 99.5 %

(Intercept) -12.284329874 5.268773592

X1 -0.004537701 -0.000712280

X2 -0.005174475 0.006772357

X3 -0.141917182 0.450227242

**Ερώτηση (v)**

**> SST<-sum((Y-mean(Y))^2) ;SST**

[1] 0.25298

**> SSR<-sum((predict(m)-mean(Y))^2);SSR**

[1] 0.202501

**> SSE<-SST-SSR;SSE**

[1] 0.05047896

**Ερώτηση (vi)**

**> N<-20**

**> P<-3**

**> S2<-SSE/(N-P-1);S2 :** Εκτίμηση Διασποράς - Διακύμανσης

[1] 0.003154935

**Ερώτηση (vii)**

**>predict(m,newdata=data.frame(X1=75,X2=70,X3=29.5),interval="confidence",level=0.90) :** Μέση πρόβλεψη διαστήματος εμπιστοσύνης

fit lwr upr

1 0.8988468 0.8353685 0.9623251

**Ερώτηση (viii)**

**> library(MASS)**

**> mod1<-stepAIC(m,direction="both",scope=(~X3+X2+X1),k=2)**

Start: AIC=-111.64

Y ~ X1 + X2 + X3

Df Sum of Sq RSS AIC

- X2 1 0.000481 0.050960 -113.449

<none> 0.050479 -111.639

- X3 1 0.007296 0.057775 -110.939

- X1 1 0.050693 0.101172 -99.733

Step: AIC=-113.45

Y ~ X1 + X3

Df Sum of Sq RSS AIC

<none> 0.050960 -113.45

- X3 1 0.007284 0.058244 -112.78

+ X2 1 0.000481 0.050479 -111.64

- X1 1 0.050462 0.101422 -101.68

Σύμφωνα με τη συνάρτηση AIC για κ=2 το καταλληλότερο γραμμικό μοντέλο είναι το : **Y = βο + β1Χ1 + β2Χ3 + ε**

**> mod1<-stepAIC(m,direction="both",scope=(~X3+X2+X1),k=log(20))**

Start: AIC=-107.66

Y ~ X1 + X2 + X3

Df Sum of Sq RSS AIC

- X2 1 0.000481 0.050960 -110.462

- X3 1 0.007296 0.057775 -107.951

<none> 0.050479 -107.656

- X1 1 0.050693 0.101172 -96.746

Step: AIC=-110.46

Y ~ X1 + X3

Df Sum of Sq RSS AIC

- X3 1 0.007284 0.058244 -110.785

<none> 0.050960 -110.462

+ X2 1 0.000481 0.050479 -107.656

- X1 1 0.050462 0.101422 -99.692

Step: AIC=-110.79

Y ~ X1

Df Sum of Sq RSS AIC

<none> 0.058244 -110.785

+ X3 1 0.007284 0.050960 -110.462

+ X2 1 0.000469 0.057775 -107.951

- X1 1 0.194736 0.252980 -84.408

Σύμφωνα όμως με τη συνάρτηση AIC για κ=log(20) το καταλληλότερο γραμμικό μοντέλο είναι το : **Y = βο + β1Χ1 + ε**

\*Θα επιλέξουμε το μοντέλο με το χαμηλότερο **AIC** δηλαδή το **Υ ~ Χ1**