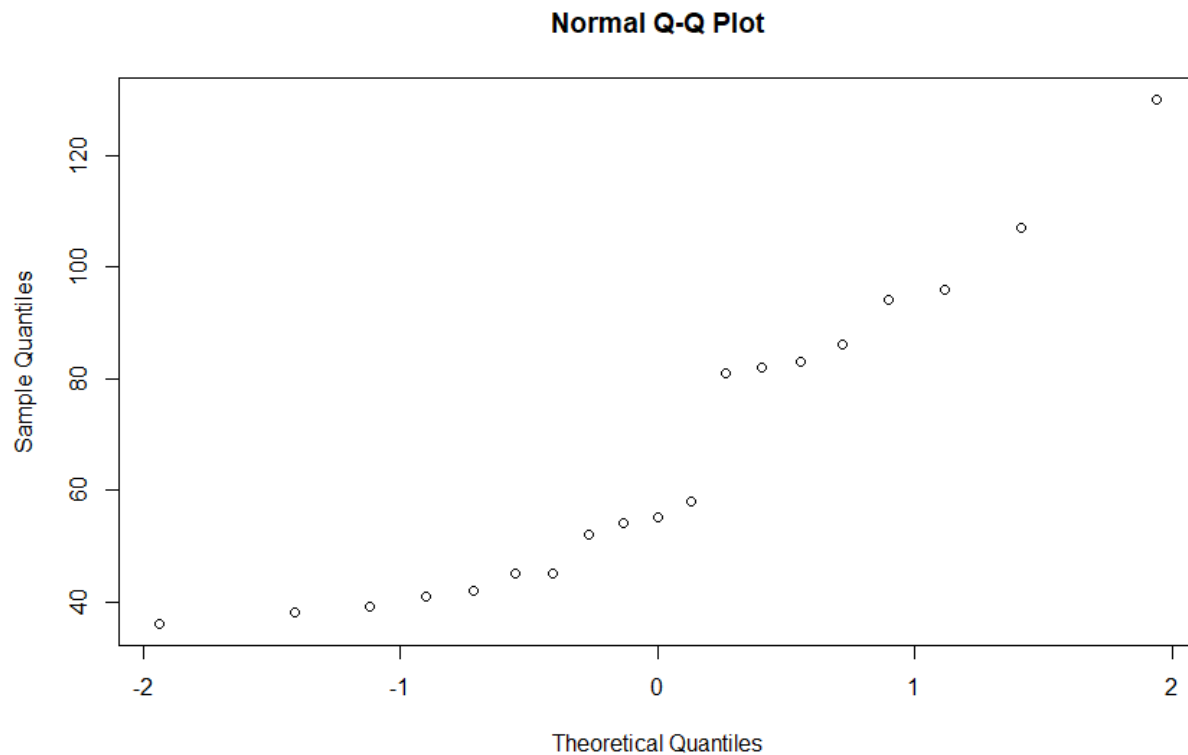


1)

a. Επειδή τα δεδομένα βρίσκονται σε κανονική κατανομή μπορούμε να εφαρμόσουμε τις μεθόδους συμπερασματολογίας που ξέρουμε. Επίσης είναι τυχαία επιλεγμένα από ένα πληθυσμό υπό την προϋπόθεση ότι ο υπολογιστής δεν είναι μεροληπτικός. Ωστόσο παρατηρούμε την ύπαρξη ενός outlier το οποίο αφαιρούμε.

**b.**

```
> x <- data$V1
> x
[1] 82 55 58 94 86 45 42 36 41 130 96 39 107 52 54 45 81 83
38
> n = length(x)
> n
[1] 19
> tstar <- -qt(0.025, df = length(x)-1)
> tstar
[1] 2.093024
```

```
> mean(x) + c(-1,1) * tstar * sd(x)/sqrt(length(x))  
[1] 53.25508 79.79755
```

Διάστημα εμπιστοσύνης 95% = [53.25508, 79.79755] Αφαιρέσαμε το outlier.

2)

a. Ο τύπος εύρεσης της τυπικής απόκλισης του δείγματος είναι η τυπική απόκλιση του πληθυσμού διά την **ρίζα** του μεγέθους του δείγματος.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

b. Δεν μπορούμε να υποθέσουμε το μέσο όρο ενός δείγματος γιατί αυτό είναι δεδομένο.

c. Απορρίπτεται η H_0 όταν $H_0: \mu < 54$, αφού για $\bar{x} = 45$ θα ήταν πιο πιθανό να ισχύει για το πληθυσμό $\mu < 54$ για λογικό αριθμό δείγματος και τυπικής απόκλισης.

d. Συνήθως το ποσοστό πρέπει να είναι πολύ μικρότερο για να απορριφθεί η H_0 . Το 52% σημαίνει ότι η H_0 είναι αρκετά πιθανό να ισχύει. Πρέπει να γνωρίζουμε το διάστημα εμπιστοσύνης, δηλαδή και το επίπεδο σημαντικότητας α . Το διάστημα εμπιστοσύνης είναι συνήθως πάνω από 80%.

3)

$$z = 1.34$$

$$\Phi(1.34) = 1 - \Phi(-1.34) = 1 - 0.0901 = 0.9099$$

a. $p\text{-value} = 1 - \Phi(1.34) = 1 - 0.9099 = 0.0901 = 9.01\%$

b. $p\text{-value} = \Phi(1.34) = 0.9099 = 90.99\%$

c. $p\text{-value} = 2 * \Phi(-1.34) = 0.1802 = 18.02\%$

4)

a. Για 95% διάστημα εμπιστοσύνης, προκύπτει ότι το επίπεδο σημαντικότητας α είναι ίσο με 5%. Εφόσον $p\text{-value} = 0.04 = 4\% < 5\%$, η υπόθεση H_0 απορρίπτεται. Αυτό σημαίνει πως το $\mu = 30$ δεν ανήκει μέσα στο διάστημα εμπιστοσύνης 95% των δεδομένων.

b. Για 90% διάστημα εμπιστοσύνης, προκύπτει ότι το επίπεδο σημαντικότητας α είναι ίσο με 10%. Εφόσον $p\text{-value} = 0.04 = 4\% < 10\%$, η υπόθεση H_0 απορρίπτεται. Αυτό σημαίνει πως το $\mu = 30$ δεν ανήκει μέσα στο διάστημα εμπιστοσύνης 90% των δεδομένων

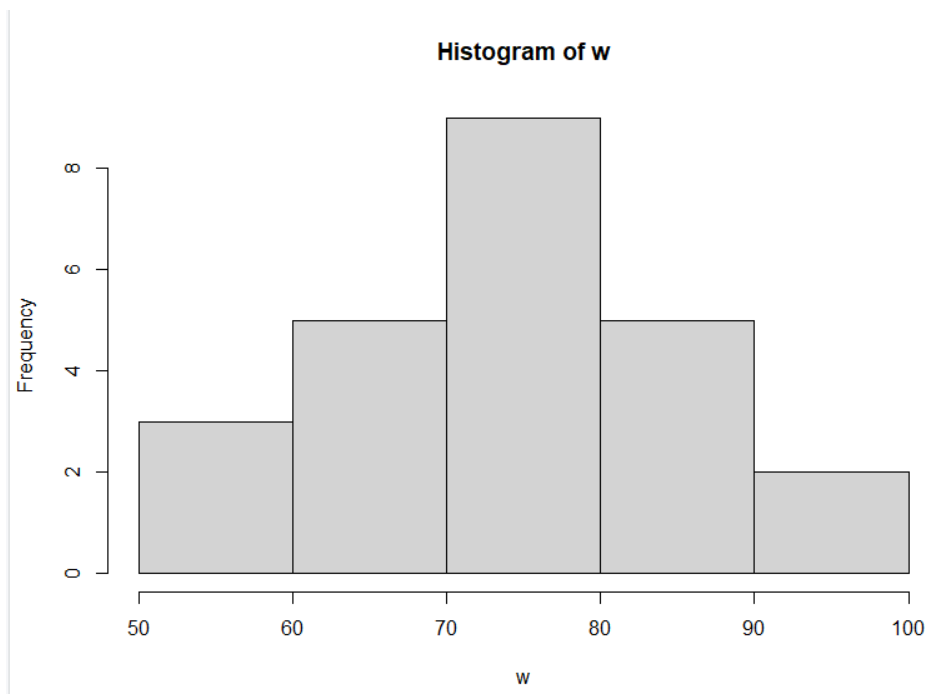
5)

a.

```
> w
[1] 80 81 75 83 71 73 65 67 54 77 55 83 91 92 86 73 82 69 73 70 59 68 72 72
> tstar <- -qt(0.025, df = length(w) - 1)
> tstar
[1] 2.068658
> mean(w) + c(-1,1) * tstar * sd(w)/sqrt(length(w))
[1] 69.57826 78.00507
```

Διάστημα εμπιστοσύνης 95% = [69.57826, 78.00507] Αφαιρέσαμε το outlier.

Μπορούμε να χρησιμοποιήσουμε την κατανομή t αφού έχουμε $n > 15$ και το δείγμα μας προσεγγίζει την κανονική κατανομή.



b.

```
> wF
[1] 54 55 59 65 67 69 70 71 73 82 83
> wM
[1] 68 72 72 73 73 75 77 80 81 83 86 91 92
> tstar <- -qt(0.1, df = length(wF) - 1)
> tstar
[1] 1.372184
> difference <- mean(wM) - mean(wF)
> difference
[1] 10.69
> c80 <- difference + c(-1,1) * tstar * sqrt(sd(wM)^2/length(wM) +
sd(wF)^2/length(wF))
> c80
[1] 5.786848 15.593152
```

Διάστημα εμπιστοσύνης διαφοράς μέσων βαρών 80% = [5.786848, 15.593152]

c. Παίρνουμε την μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ και την εναλλακτική υπόθεση $H_1: \mu_1 \neq \mu_2$, όπου μ_1 το μέσο βάρος των καπνιστών και μ_2 το μέσο βάρος των μη-καπνιστών του πληθυσμού.

Διαλέγουμε 95% διάστημα εμπιστοσύνης, υπολογίζουμε τον εκτιμητή t για να συγκρίνουμε τα 2 δείγματα (καπνιστές και μη καπνιστές).

```
> sW
[1] 80 83 71 73 65 77 92 86 82 59
> nW
[1] 81 75 67 54 55 83 91 73 69 73 70 68 72 72
> t <- ((mean(sW) - mean(nW)) / (sqrt(sd(sW)^2/length(sW) +
sd(nW)^2/length(nW))))
> t
[1] 1.259715
```

$t = 1.259715$

Με χρήση λογισμικού για $t = 1.259715$ και $df = 9$ για $\alpha = 5\%$

The p -value is .239463.

The result is *not* significant at $p < .05$.

Η μηδενική υπόθεση $H_0: \mu_1 = \mu_2$ δεν απορρίπτεται, αφού $p\text{-value} > \alpha$.

Άρα δεν δεχόμαστε την $H_1: \mu_1 \neq \mu_2$, δηλαδή δεν υπάρχει διαφορά μεταξύ του βάρους των καπνιστών και μη-καπνιστών (με βάση το δείγμα που μας έχει δοθεί).

6)

a. Επειδή τα δεδομένα βρίσκονται αρκετά κοντά στην κανονική κατανομή μπορούμε να εφαρμόσουμε τις μεθόδους συμπερασματολογίας που ξέρουμε. Επίσης είναι τυχαία επιλεγμένα από ένα πληθυσμό υπό την προϋπόθεση ότι η επιλογή μας δεν είναι μεροληπτική. Το δείγμα μας είναι αρκετά μεγάλο σε πλήθος.

b.

```
> x
[1] 5.7 4.6 6.4 6.3 6.9 5.2 4.9 5.4 4.9 5.6 5.4 5.3 4.9 5.1 5.0 6.0 6.3 5.4
5.3 5.4
> mean(x)
[1] 5.5
> sd(x)
[1] 0.6008766
```

mean = 5.5

sd = 0.6008766

c.

C = 95%

```
> tstar <- -qt(0.025, df = length(x)-1)
> tstar
[1] 2.093024
```

t* = 2.093024

```
> mean(x) + c(-1,1) * tstar * sd(x)/sqrt(length(x))
[1] 5.218781 5.781219
```

Διάστημα εμπιστοσύνης 95% = [5.218781, 5.781219]

7.

a. $H_0: \mu_1 \leq 0$ $H_1: \mu_1 > 0$, με μ_1 μέσος όρος διαφοράς της εκτίμησης (συνεργείο – εμπειρογνώμονας)

```
> s
[1] 500 1550 1250 1300 750 1000 1250 1300 800 2500
> e
[1] 400 1500 1300 1300 800 800 1000 1100 650 2200
> dif <- s - e
> dif
[1] 100 50 -50 0 -50 200 250 200 150 300
> t <- mean(dif) / (sd(dif) / sqrt(length(dif)))
> t
[1] 2.913182
> tstar <- -qt(0.05, df = length(dif)-1)
> tstar
[1] 1.833113
```

$|t| = 2.913182$

$t^* = 1.833113$

Εφόσον $|t| > t^*$ η μηδενική υπόθεση $H_0: \mu_1 \leq 0$ απορρίπτεται, αφού $p\text{-value} \leq \alpha$.

Άρα δεχόμαστε την $H_1: \mu_1 > 0$, δηλαδή το συνεργείο είναι πολύ πιθανό να κάνει υπερεκτίμηση της ζημιάς. Ο αριθμός του δείγματος είναι οριακός, με ένα μεγαλύτερο δείγμα θα ήμασταν πιο σίγουροι για το συμπέρασμα.

8.

a. Αφαιρέσαμε το outlier με βάση το ύψος, φτιάξαμε δύο πίνακες με τα ύψη όλων των γυναικών και τα ύψη όλων των ανδρών.

```
> tstar <- -qt(0.025, df = min(length(hM$height), length(hF$height)) - 1)
> tstar
[1] 2.026192
> ci <- mean(hM$height) - mean(hF$height) + c(-1,1) * tstar *
(sqrt(sd(hM$height)^2/length(hM$height) +
sd(hF$height)^2/length(hF$height)))
> ci
[1] 0.09840989 0.15563221
```

Υπολογίζουμε το tstar και κατασκευάζουμε το διάστημα εμπιστοσύνης για $C = 95\%$.

Διάστημα εμπιστοσύνης διαφοράς ύψους $95\% = [0.09840989, 0.15563221]$

(Ύψος άντρα – Ύψος γυναίκας) για $df = \text{length}(hF\$height) - 1 = 37$

b.

Επίπεδο σημαντικότητας $\alpha = 5\%$

Παίρνουμε την μηδενική υπόθεση $H_0: \mu_1 > \mu_2$ και την εναλλακτική υπόθεση $H_1: \mu_1 \leq \mu_2$

Με μ_1 μέσο βαθμό ανδρών και μ_2 μέσο βαθμό γυναικών.

Υπολογίζουμε t, tstar.

```
> t <- (mean(probM) - mean(probF)) / (sqrt(sd(probM)^2/length(probM) +
sd(probF)^2/length(probF))))
> t
[1] -1.184079
> tstar <- -qt(0.05, df = min(length(probM), length(probF)) - 1)
> tstar
[1] 1.69236
```

$|t| = 1.18$

$t^* = 1.69$

Εφόσον $|t| < t^*$ η μηδενική υπόθεση $H_0: \mu_1 \leq \mu_2$ δεν απορρίπτεται, αφού $p\text{-value} > \alpha$.

Άρα δεν δεχόμαστε την $H_1: \mu_1 \leq \mu_2$, δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα για τους βαθμούς των δύο μαθημάτων από το συγκεκριμένο δείγμα.

c.

Κρατάμε μόνο τους φοιτητές που έχουν βαθμό και στα δύο μαθήματα.

$H_0: \mu = 0$ $H_1: \mu \neq 0$, με μ μέσος όρος διαφοράς των βαθμών (μαθηματικά – πιθανότητες)

```
> t <- mean(difgrade) / (sd(difgrade) * sqrt(length(difgrade)))  
> t  
[1] 0.9707747  
> tstar <- -qt(0.05, df = length(difgrade)-1)  
> tstar  
[1] 1.660881
```

$|t| = 0.9707747$

$t^* = 1.66$

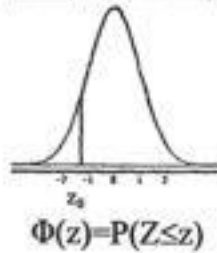
Εφόσον $|t| < t^*$ η μηδενική υπόθεση $H_0: \mu_1 = 0$ δεν απορρίπτεται, αφού

$p\text{-value} = 33\% > \alpha = 5\%$.

Άρα δεν δεχόμαστε την $H_1: \mu_1 > 0$, δηλαδή δεν μπορούμε να εξάγουμε κάποιο συμπέρασμα με αυτό το δείγμα.

Γενικά όπου δεν αναφέρεται, το πλήθος των δειγμάτων είναι ικανοποιητικό για να εφαρμόσουμε τις μεθόδους συμπερασματολογίας που γνωρίζουμε. Η κατανομή των δεδομένων είναι αρκετά κοντά στην κανονική κατανομή.

Τυποποιημένη Κανονική Κατανομή



Τα στοιχεία του Πίνακα εκφράζουν τις πιθανότητες $\Phi(z) = P(Z \leq z)$ που παριστάνονται από το εμβαδόν κάτω από την καμπύλη της τυποποιημένης κανονικής κατανομής αριστερά από το z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0227	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0186	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1921	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Pvalue calculation: <https://www.socscistatistics.com/pvalues/tdistribution.aspx>