Το Youth Risk Behavior Surveillance System (YRBSS) ξεκίνησε να αναπτύσεται το 1990 στις ΗΠΑ ώστε να δίνει τη δυνατότητα κάποιας εποπτείας σχετικά με συμπεριφορές-συνήθειες οι οποίες μακρυπρόθεσμα ( ή και όχι ) μπορούν να οδηγήσουν σε προβλήματα υγείας , παραβατική συμπεριφορά , αναπηρία , θάνατο ή και άλλα κοινωνικά προβλήματα .Ο υπό εξέταση πληθυσμός είναι το σύνολο των μαθητών που φοιτούνε από την ένατη έως και τη δωδέκατη βαθμίδα της δευτεροβάθμιας εκπαίδευσης των ΗΠΑ . Σε αντιστοιχία με την Ελλάδα η συγκεκριμένη έρευνα θα αφορούσε τους μαθητές που φοιτούνε από την Γ' Γυμνασίου έως και τη Γ' Λυκείου .

Τέτοιες συμπεριφορές μπορεί να αποτελούνε η χρήση βίας ατομικά ή και απλά η ύπαρξή της στο στενό κοινωνικό περίγυρο , η κατανάλωση αλκοόλ , ναρκωτικών ουσιών , το κάπνισμα , η ανεπαρκής σωματική άσκηση και άλλα πολλά .

Έχει καθιερωθεί η συγκεκριμένη έρευνα να πραγματοποιείται κάθε δύο χρόνια με σημείο εκκίνησης να αποτελεί το έτος 1991 . Στο παρών θα ασχοληθούμε με το έτος 2013 . Πιο συγκεκριμένα σε εθνικό επίπεδο επιλέχτηκε ένα δείγμα 193 σχολείων ( high schools ) από τα οποία τελικά συμμετείχαν στην έρευνα τα 148 ( ποσοστό ανταπόκρισης της τάξης του 77 %) . Λόγοι μη συμμετοχής των υπολοίπων είναι για παράδειγμα η άρνηση της πλειοψηφίας του συλλόγου γονέων και κηδεμόνων για τη συμμετοχή του συγκεκριμένου σχολείου στην έρευνα .

Στα 148 σχολεία που εν τέλει δέχτηκαν να συμμετάσχουν φοιτούσαν συνολικά 15480 μαθητές από τους οποίους τελικά στην έρευνα συμμετείχαν οι 13583 ( ποσοστό ανταπόκρισης 88 % ) . Επομένως , το συνολικό ποσοστό ανταπόκρισης ειίναι της τάξης του 77%*88%=68% . Το τελευταίο είναι κάτι που ενδιαφέρει αφού εξ' αρχής για τη συγκεκριμένη έρευνα έχει οριστεί ως κατώτατο όριο για το συνολικό ποσοστό ανταπόκρισης το 68%.

Λίγο πιο συγκεκριμένα η διαδικασία που ακολουθήθηκε για την επιλογή του δείγματος είναι η εξής :

Αρχικά, έγινε στρωματοποίηση των σχολικών μονάδων με βάση γεωγραφικά κριτήρια .Σε κάθε στρώμα επιλέχθηκε ένας αριθμός σχολείων με πιθανότητα ανάλογη με τον αριθμό των μαθητών που φοιτούσαν στο σχολείο.Στη συνέχεια, με απλή τυχαία δειγματοληψία επιλέχθηκαν τάξεις από τις οποίες ο κάθε μαθητής είχε τη δυνατότητα να συμμετάσχει στην έρευνα με τη σύμφωνη γνώμη των γονιών του.Βέβαια , είναι πολύ σύνηθες σε έρευνες μεγάλης κλίμακας να δίνονται μόνο τα clusters και η στρωματοποίηση που έγινε στο πρώτο στάδιο δειγματοληψίας ή ακόμα να δίνονται ψευδοπρωτεύουσες δειγματοληπτικές μονάδες οι οποίες δίνουν καλές προσεγγίσεις αλλά δεν ανταποκρίνονται στο ακριβές δειγματοληπτικό σχέδιο .Πέρα από απλούστευση στους υπολογισμούς , αυτή η πρακτική κάνει τα πράγματα δύσκολα σε ότι αφορά την οποιαδήποτε ταυτοποίηση ατόμων , η οποία ίσως να ήταν πραγματοποιήσιμη αν τα ακριβή στοιχεία για κάθε στάδιο της δειγματοληπτικής έρευνας ήταν δημοσιεύσιμα . (Lumley , Complex Surveys A guide to analysis using R σελ. 42)

Η έρευνα με την οποία ασχολούμαστε δεν αποτελεί εξαίρεση , σε σχέση με όσα αναφέρθηκαν προηγουμένως , και άρα η συνάρτηση που θα επιλεγεί να ακολουθήσει το συγκεκριμένο δειγματοληπτικό σχέδιο είναι η

svydesign( ~ psu, strata = ~ stratum , data = year2013 , weights = ~ weight ,nest=TRUE ) **(1)**

, αφού μας δίνονται οι πρωτεύουσες δειγματοληπτικές μονάδες μόνο αλλά και η αρχική στρωματοποίηση .Επίσης η μεταβλητή weights μας υποδηλώνει την ύπαρξη στάθμισης .

Η διαδικασία που θα ακολουθηθεί για την εργασία είναι η εξής :

Θα επιλεγούνε κάποιες μεταβλητές από το dataset που αφορά το έτος 2013 και δουλεύοντας με αυτές θα απαντηθούνε τα ζητούμενα 1.1 , 1.2 , 1.3 .

Υποθέτωντας ότι η (1) μας περιγράφει το δειγματοληπτικό σχέδιο με στάθμιση,θα δώ σουμε κάποια δειγματοληπτικά αποτελέσματα για τη μεταβλητή bmipct η οποία είναι μια συνάρτηση του βάρους και του ύψους των μαθητών. Πρώτα ας δούμε μια εκτίμη ση για τον πληθυσμό χρησιμοποιώντας την εντολή svytotal( ~ one , yr2013weighted ) ,(η μεταβλητή one έχει την τιμή 1). Τα αποτελέσματα είναι

```
        total     SE
one 13583 892.18  ,
```

Χρησιμοποιώντας την svymean( ~ bmipct , yr2013weighted , na.rm = TRUE ) παίρ-νουμε ως αποτέλεσμα μια εκτίμηση για τη μέση τιμή του πληθυσμού

```
              mean     SE
bmipct    63.482     0.5144
```

Για να βρούμε ένα 95% διάστημα εμπιστοσύνης για την προηγούμενη εκτίμηση χρη-σιμοποιούμε την confint(MeanbmipctWeighted, level = 0.95, df = dfyr2013weighted) και έχουμε

```
                2.5 %  97.5 %
bmipct 62.44309 64.5207
```

Τώρα ,χρησιμοποιώντας την εντολή svyratio θα πάμε να υπολογίσουμε ένα λόγο.Θα χρησιμοποιήσουμε τη μεταβλητή q36 του dataset . Για τη συγκεκριμένη μεταβλητή οι τιμές μεγαλύτερες του ένα αφορούνε τους μαθητές που κάπνισαν τσιγάρο τουλά-χιστον μια φορά σε σχολικό χώρο τις τελευταίες τριάντα μέρες ενώ η τιμή ένα αφορά αυτούς που δεν κάπνισαν .Επομένως τα αποτελέσματα που έχουμε είναι

```
Ratio estimator: svyratio.survey.design2(numerator = ~notSmoked,denominator=~S
moked, yr2013weighted, na.rm = TRUE)
Ratios=
                Smoked
notSmoked 25.10935
SEs=
                Smoked
notSmoked 2.923842     ,
```

Σε κάθε περίπτωση αυτό που παρατηρούμε είναι ότι η εκτίμηση για αυτούς που δεν κάπνιζαν είναι αριθμιτικά μεγαλύτερη από εκείνη των καπνιστών . Επίσης μπορούμε να θέσουμε στη συνάρτηση αντίστροφα τις εντολές numerator και denominator και να πάρουμε ένα λόγο πολύ μικρότερο της μονάδας.

Συνεχίζοντας, θα δώσουμε κάποια δειγματοληπτικά αποτελέσματα για υποπληθυ-σμούς. Θα χωρίσουμε τη μεταβλητή q68 του dataset σε δυο υποπληθυσμούς. Κα-ταρχάς η μεταβλητή q68 παίρνει την τιμή 1 για τους μαθητές οι οποίοι τις τελευταίες 30 ημέρες έμειναν τουλάχιστον για μια ολόκληρη μέρα νηστικοί προκειμένου να χά-σουν βάρος. Επίσης, παίρνει την τιμή 2 για τους μαθητές οι οποίοι δεν υπέβαλαν τον εαυτό τους στην προηγούμενη διαδικασία .Οπότε, η μεταβλητή θα χωριστεί σύμφω-

να με τις προηγούμενες δύο περιπτώσεις χρησιμοποιώντας τις εντολές
subset( yr2013weighted , q68 == 1 )
subset( yr2013weighted , q68 == 2 )
αντίστοιχα.Στον κάθε υποπληθυσμό θα πάμε να βρούμε μια εκτίμηση για τη μέση τιμή της μεταβλητής bmipct .Τα αποτελέσματα είναι

```
          mean     SE
bmipct 68.656 0.9794
```

, για όσους δοκίμασαν να χάσουν βάρος
ενώ

```
          mean     SE
bmipct 62.684 0.5696
```

, για όσους δε δοκίμασαν να χάσουν βάρος

Παρατηρούμε δηλαδή ότι αυτοί που δοκίμασαν να χάσουν βάρος ήταν όντως και οι μαθητές με τα περισσότερα κιλά (μεγάλο bmipct σημαίνει μαθητής με αρκετά κιλά) . Για τις προηγούμενες εκτιμήσεις τα 95% διαστήματα εμπιστοσύνης είναι

```
             2.5 %   97.5 %
bmipct 66.73604 70.57518
```

και

```
             2.5 %   97.5 %
bmipct 61.56731 63.80025
```

αντίστοιχα.

Τώρα,θα πραγματοποιηθεί ένα t-test.Οι εμπλεκόμενες μεταβλητές θα είναι η q69 και η bmipct (έχει σημειωθεί προηγουμένως τι αναπαριστά ).Η μεταβλητή q69 παίρνει την τιμή 1 αν ο μαθητής προσπάθησε να χάσει βάρος τις τελευταίες τριάντα ημέρες καταναλώνοντας υγρά (για να κόψει όσο είναι δυνατόν το αίσθημα πείνας ), και την τιμή 2 αν όχι .Οπότε αυτό που θέλουμε να δούμε είναι αν υπάρχει διαφορά ανάμεσα στις μέσες τιμές της μεταβλητής bmipct μεταξύ εκείνων που προσπάθησαν να χάσουν βάρος καταναλώνοντας υγρά και εκείνων που δεν το επιχείρησαν. Συνεπώς χρησιμοποιώντας την ακόλουθη εντολή παίρνουμε τα αποτελέσματα .

```
svyttest(bmipct~q69,yr2013weighted )
```

```
        Design-based t-test

data:  bmipct ~ q69
t = 3.6479, df = 39, p-value = 0.000772
alternative hypothesis: true difference in mean is not equal to 0
95 percent confidence interval:
  3.771517 12.529954
sample estimates:
difference in mean 8.150735
```

Αυτό που προκύπτει είναι ότι η pvalue του ελέγχου είναι μικρότερη του 0.05 .Αυτό σημαίνει ότι σε στάθμη σημαντικότητας 95% η αρχική υπόθεση απορρίπτεται και το συμπέρασμα είναι ότι όντως υπάρχει διαφορά για τη μέση τιμή της bmipct μεταξύ εκείνων που προσπάθησαν να χάσουν βάρος και εκείνων που δεν δοκίμασαν , κάτι που είναι και λογικό. Το προηγούμενο αφορούσε το δειγματοληπτικό σχέδιο με την ύπαρξη στάθμισης . Η αντίστοιχη εντολή που περιγράφει το δειγματοληπτικό σχέδιο χωρίς στάθμιση είναι η svydesign( ~ psu, strata = ~ stratum , data = year2013 ,nest=TRUE ) στην οποία απουσιάζει η παράμετρος weights . Επομένως,εφαρμόζοντας την εντολή svyttest(bmipct~q69,yr2013unweighted ) τα αποτελέσματα που παίρνουμε είναι

Design-based t-test

data:  bmipct ~ q69
t = 4.5912, df = 39, p-value = 4.507e-05
alternative hypothesis: true difference in mean is not equal to 0
95 percent confidence interval:
 4.893398 12.183374
sample estimates:
difference in mean
        8.538386

Τέλος, θα εφαρμόσουμε ένα χ τετράγωνο τεστ ανάμεσα στις μεταβλητές q2 και q15. Η μεταβλητή q2 περιγράφει το φύλο του μαθητή ( 1 για κορίτσι , 2 για αγόρι) ενώ η q15 περιγράφει τη συχνότητα που κάποιος μαθητής έφερε κάποιου είδους οπλισμό στο σχολείο τις τελευταίες τριάντα ημέρες.Τα παρακάτω αποτελέσματα είναι για το δειγματοληπτικό σχέδιο με στάθμιση.

```
Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~q2 + q15, yr2013weighted)
F = 3.9133, ndf = 2.058, ddf = 84.379, p-value = 0.02269
```

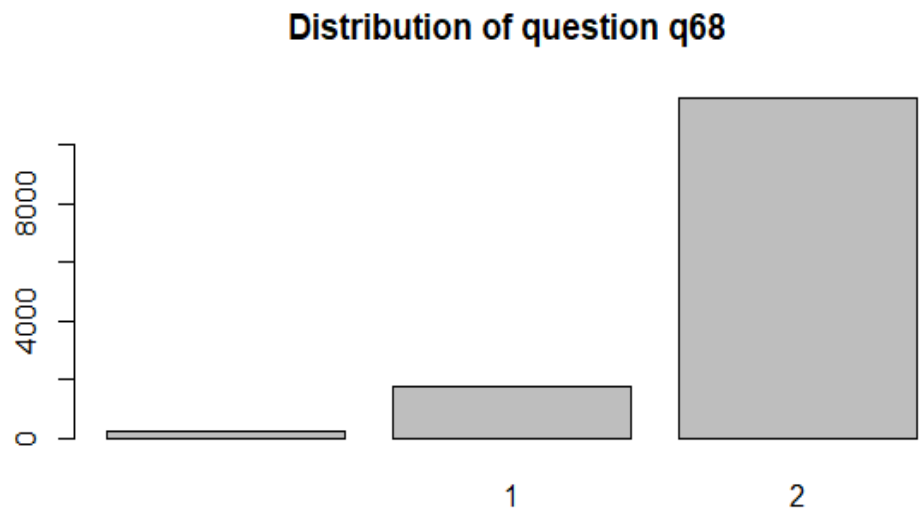
Η pvalue είναι μικρότερη του 0.05 οπότε μπορεί να απορριφθεί η αρχική υπόθεση της ανεξαρτησίας άρα υπάρχει εξάρτηση μεταξύ φύλου και οπλοκατοχής .

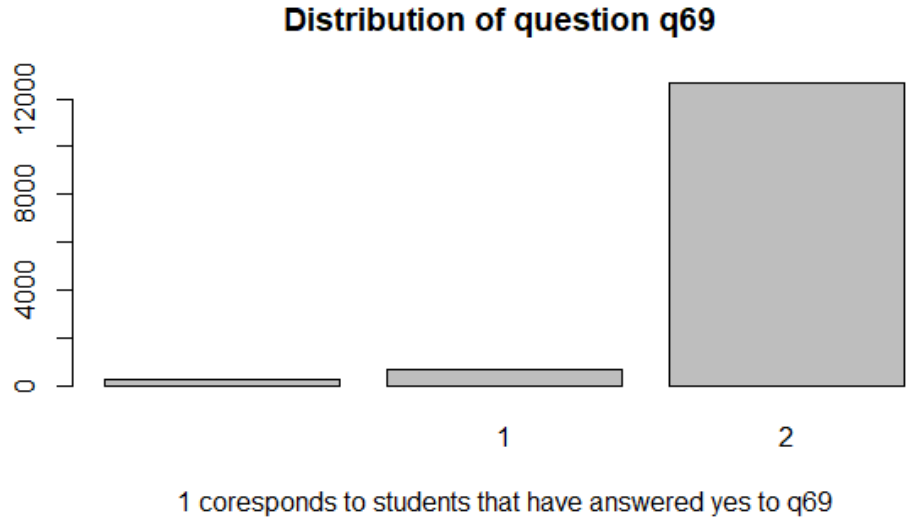
Παρακάτω ειναι ο αντίστοιχος έλεγχος χωρίς στάθμιση

```
Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~q2 + q15, yr2013unweighted)
F = 6.5952, ndf = 2.3393, ddf = 95.9106, p-value = 0.00119
```

Ενδεικτικά δίνονται και δύο γραφικές παραστάσεις για κάποια από τα δεδομένα που χρησιμοποιήθηκαν.




Η πρώτη στήλη αφορά εκείνους που επέλεξαν να μην απαντήσουν στην ερώτηση.