$2^{η}$ Σειρά Ασκήσεων(2021-2022)

Φοιτητές:
Ειρήνη Χρυσικοπούλου(3180208)
Παναγιώτης Παναγιώτου(3180139)

1.

a.

Πρέπει αν ελέξουμε αν η τιμές ακολουθούν την κανονική κατανομή.

Stemplot:

0 | 44444

0 | 55556688899

1 | 013

1 |

2 |

2 | 8

Παρατηρούμε ότι υπάρχει μια ατυπική τιμή,γεγονός που σημαίνει οτι ίσως ο πληθυσμός να μην είναι κανονικά κατανεμημένος.Όμως,το μέγεθος(n=20>15) του δείγματος είναι τέτοιο που ακόμα και αν ο πληθυσμός δεν ακολουθεί την κανονική κατανομή μπορούμε να εφαρμόσουμε τις μεθοδους στατιστικής συμπερασματολογίας.

Ο τρόπος λήψεις του δείγματος είναι επίσης σωστός,καθώς πρόκειται για λήψη τυχαίων τιμών από τον πληθυσμό(SRS).

b. Η δειγματική μέση τιμή ισούται με $\bar{\chi}$=77.4

Η δειγματική τυπική απόκλιση ισούται με σ$\bar{\chi}$=55.52467

Ό τύπος για το διάστημα εμπιστοσύνης για την μέση τιμή είναι:

$\bar{\chi}$±t* σ$\bar{\chi}/\sqrt{n}$

Επομένως το διάστημα εμπιστοσύνης 95% είναι το:
77.4±2.093024$\frac{55.52467}{\sqrt{20}}$=[51.41365, 103.3863]

(για τον υπολογισμό του t* χρησιμοποιήθηκαν 19 βαθμοί ελευθερίας)

2.

a.

Η τυπική απόκλιση απόκλιση του δειγματικού μέσου ισούται με σ/√n , όπου σ η τυπική απόκλιση του πληθυσμού και n το μέγεθος του δείγματις.Επομένως η τυπική απόκλιση του δειγματικού μέσου είναι 12/√20.

b.

Η μηδενική υπόθεση αφορά κάποια παράμετρο του πληθυσμού(οχι του δείγματος).Επομένως ο δειγματικός μέσος δεν μπορεί να χρησιμοποιηθεί για την διατύπωση υποθέσεων.

c.

Για $\bar{\chi}$=45 το στατιστικό t έχει τιμή<0.Άρα για το p value για την εναλλακτική υπόθεση θα πάρει κάποια μεγάλη τιμή.Επομένως δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση.

d. Για να απορίψουμε την μηδενική υπόθεση πρέπει να γνωρίζουμε το επίπεδο σημαντικότητας α.
Επιπλέον το ότι το p value λαμβάνει μεγάλη τιμή σημαίνει ότι δεν απορρίπτεται η μηδενική υπόθεση.

3.

a.

Αναζητούμε την πιθανότητα το z να πάρει τιμές μεγαλύτερες ή ίσες του 1.34,επομένως

p value=1-Φ(1.34)= 0.09012267

b. .Αναζητούμε την πιθανότητα το z να πάρει τιμές μικρότερες ή ίσες του 1.34,επομένως

p value=Φ(1.34)= 0.9098773

c. Αναζητούμε την πιθανότητα το z να πάρει τιμές μεγαλύτερες από το 1.34 και μικρότερες από το -1.34,επομένως p value=2Φ(-1.34)=0.1802453

4.

a. Η τιμή 30 περιέχεται στο 95% διάστημα εμπιστοσύνης αν και μόνο αν η μηδενική υπόθεση δεν απορρίπτεται για επίπεδο σημαντικότητας α=100%-95%=5%=0.05.
Όμως p value=0.04≤α, συνεπώς η τιμή 30 δεν περιέχεται στο 95% διάστημα εμπιστοσύνης.

b. Η τιμή 30 περιέχεται στο 90% διάστημα εμπιστοσύνης αν και μόνο αν η μηδενική υπόθεση δεν απορρίπτεται για επίπεδο σημαντικότητας α=100%-90%=0.1.
Όμως p value=0.04≤α, συνεπώς η τιμή 30 δεν περιέχεται στο 90% διάστημα εμπιστοσύνης.

5.

Δημιούγουμε πίνακα με τα δεδομένα(αφαιρούμε την περίπτωση 14 όπου το βάρος είναι 6).
Το γεγονός ότι το δείγμα έχει ληφθεί με τυχαίο τρόπο(SRS) το καθιστά κατάλληλο για τις μεθόδους στατιστικής συμπερασματολογίας που γνωρίζουμε.

a.

Ελέγχουμε αν το βάρος είναι κανονικά κατανεμημένο παρατηρώντας το stemplot:

5 | 459
6 | 5789
7 | 012233357
8 | 012336
9 | 12

Παρατηρούμε οτι η κατανομή είναι αρκετά συμμετρική και το μέγεθος του δείγματος(n=24) μεγάλο επομένως μπορούμε να εφαρμόσουμε τις μεθόδους στατιστικής συμπερασματολογίας που γνωρίζουμε.

Δειγματικός μέσος= 73.79167

Δειγματική Τυπική Απόκλιση=9.978146

t*= 2.068658 (df=23)
Επομένως το διάστημα επιστοσύνης είναι το: 73.79167±4.2134023=[69.5782677, 78.0050723]

b.

Από τα stemplot προκύπτει ότι τα δεδομένα είναι κατάλληλα για τις μεθόδους συμπερασματολογίας που γνωρίζουμε καθώς φαίνονται αρκετά συμμετρικά κατανεμημένα,χωρίς ατυπικά σημεία.

Δημιουργούμε τους πίνακες female και male και υπολογίζουμε τον δειγματικό μέσο και την δειγματική τυπική απόκλιση για κάθε ομάδα δεδομένων.

Για τις γυναίκες ισχύουν τα ακόλουθα:

Stemplot:

5 | 459
6 | 579
7 | 013
8 | 23

```
> mean(female$weight)//δειγματικός μέσος(x1)
[1] 68
```

```
> sd(female$weight)//δειγματική τυπική απόκλιση(s1)
[1] 9.570789
```

Πλήθος γυναικών(n1)=11

Για τους άνδρες ισχύουν τα ακόλουθα:

Stemplot:

6 | 8
7 | 2233
7 | 57
8 | 013
8 | 6
9 | 12

> mean(male$weight))//δειγματικός μέσος(x2)
[1] 78.69231

> sd(male$weight) //δειγματική τυπική απόκλιση(s2)
[1] 7.598077

Πλήθος ανδρών(n2=13)

Επομένως το διάστημα εμπιστοσύνης είναι:

χ2-χ1±t* $\sqrt{\frac{s1^2}{n1}+\frac{s2^2}{n2}}$=10.69231±4.903152=[5.789158, 15.59546]

(Για τον υπολογισμό του t* ,df=10)

c.

Δημιουργούμε τους πίνακες smokers και non-smokers αντίστοιχα.Όπως θα παρατηρήσουμε από τα stemplots τα δεδομένα και για τις δύο ομάδες δεδομένων φαίνονται αρκετά συμμετρικά κατανεμημένα, επομένως μπορούμε να εφαρμόσουμε μεθόδους για κανονικά κατανεμημένα δεδομένα.

Θα ελέξουμε αν το μέσο βάρος των καπνιστών απέχει σημαντικά από το μέσο βάρος των μη καπνιστών:

Για το βάρος των καπνιστών έχουμε:

Stemplot:

5 | 9
6 | 5
7 | 137
8 | 0236
9 | 2

m1<-mean(smokers$weight)//Μέση τιμή μ1
> m1
[1] 76.8

> s1<-sd(smokers$weight) //Τυπική απόκλιση s1
> s1
[1] 9.975526

Πλήθος καπνιστών(n1)=10

Για το βάρος των μη καπνιστών έχουμε:

Stemplot:

5 | 45
6 | 789
7 | 022335
8 | 13
9 | 1

> m2<-mean(non_smokers$weight)// Μέση τιμή μ2
> m2
[1] 71.64286

> s2<-sd(non_smokers$weight) // Τυπική απόκλιση s2
> s2
[1] 9.76341

Πλήθος μη καπνιστών(n2)=14

Στατιστικό t=$\frac{\mu 1-\mu 2}{\sqrt{\frac{s1^2}{n1}+\frac{s2^2}{n2}}}$= 1.259715

Εύρεση p value (df=min{n1-1,n2-1}=9)

> p_value<-2*pt(-abs(t),9)
> p_value
[1] 0.2394574

Επομένως p value= 0.2394574=23,95%

Το p value είναι αρκετά μεγάλο,επομένως δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση,ότι δηλαδή οι καπνιστές έχουν το ίδιο βάρος με τους μη καπνιστές.

6.

a.Το stemplot είναι:

4 | 6999
5 | 012334444
5 | 67
6 | 0334
6 | 9

Παρατηρούμε ότι τα δεδομένα,μπορεί να μην κατανεμημένα με απόλυτη συμμετρία(συγκέντρωση γύρω από τον μέσο),είναι όμως αρκετά στο πλήθος και χωρίς σημαντικές αποκλίσεις.Δεδομένου ότι και το δείγμα λήφθηκε με τυχαίο τρόπο τα δεδομένα είναι κατάλληλα για τις μεθόδους συμπερασματολογίας που γνωρίζουμε.

b.

> mean(data) // Μέση τιμή($\bar{\chi}$)
[1] 5.5
> sd(data) //Τυπική απόκλιση(σ)
[1] 0.6008766

c.

Το διάστημα είναι το $\bar{\chi}$ ±t*$\frac{\sigma}{\sqrt{n}}$=[5.218781, 5.781219],

t* =2.093024,df=19.

7.

Παρατηρούμε ότι οι εκτιμήσεις του εμπειρογνώμονα και του συνεργείου δεν είναι ανεξάρτητες,επομένως δεν μπορούμε να χρησιμοποίησουμε τις γνωστές μας μεθόδους.

Θα θεωρήσουμε την διαφορά της εκτίμησης του συνεργείου απο την εκτίμηση του εμπειρογνώμονα για να δούμε πόσο «κοντά» στην εκτίμηση του εμπειρογνώμονα είμαστε.

Το διάνυσμα των διαφορών είναι: data<-c(100,50,-50,0,-50,200,250,200,150,300)

Το stemplot είναι:
-0 | 55
 0 | 05
 1 | 05
 2 | 005
 3 | 0

Παρατηρούμε ότι το δείγμα αν και μικρό δεν παρουσιάζει ατυπικές τιμές και τα δεδομένα έχουν μια σχετική συμμετρία,επομένως μπορούμε να εφαρμόσουμε τις μεθόδους που γνωρίζουμε,χωρίς βέβαια απόλυτη ακρίβεια.Επίσης έχουμε:

> x<-mean(data)
> x
[1] 115

```
> s<-sd(data)
> s
[1] 124.8332
```

Θα ελέξουμε την μηδενική υπόθεση H_0 :μ=0 (δηλαδή ότι το συνεργείο δεν υπερκεκτιμά τις ζημιές) και την εναλλακτική υπόθεση H_α:μ>0(δηλαδή ότι το συνεργείο όντως υπερκτιμά τις ζημιές).

Για την μηδενική υπόθεση:

$t=\frac{x-0}{s/\sqrt{n}}=2.913182$,

Επομένως p value=1-Φ(2.913182)= 0.008611

Το p value είναι μικρό,άρα μπορούμε να απορρίψουμε την μηδενική υπόθεση.Επομένως το συνεργείο υπερκτιμά την ζημιά.

8.

a. Αρχικά,τα δεδομένα είναι κατάλληλα για τις γνωστές μας μεθόδους συμπερασματολογίας,γιατί τα δείγματα των υψών των δύο πληθυσμών έχουν ληφθεί με τυχαίο τρόπο και είναι κοντά στην κανονική κατανομή.Επίσης τα δείγματα είναι ανεξάρτητα,επομένως μπορούμε να κάνουμε t.test.(Διαγράφτηκαν γραμμές με κενά)

```
> t.test(m$height,fe$height)

        Welch Two Sample t-test

data:  m$height and fe$height
t = 8.8041, df = 48.324, p-value = 1.303e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1226416 0.1952212
sample estimates:
mean of x mean of y
 1.804265  1.645333
```

Όπως προκύπτει από το t-test το C είναι το [0.1226416, 0.1952212].

b. Έστω μ_α και μ_γ ο μέσος βαθμός στις πιθανότητες για τον πληθυσμό των ανδρών και των γυναικων αντίστοιχα.Θα κάνουμε τον έλεγχο: $H_0:\mu_\alpha= \mu_\gamma, H_\alpha: \mu_\alpha> \mu_\gamma$

```
> t.test(fe$prob,m$prob,alternative="greater")

        Welch Two Sample t-test
data:  fe$prob and m$prob
t = -1.0147, df = 57.807, p-value = 0.8428
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.199158       Inf
sample estimates:
mean of x mean of y
 5.650000  6.102941
```

Όπως προκύπτει από το t test το p value=0.8428>0.05,επομένως δεν πρέπει απορρίπτεται η μηδενική υπόθεση.Συμπεραίνουμε επομένως ότι οι άντρες και οι γυναίκες επιτυγχάνουν κατά μέσο όρο τον ίδιο βαθμό στις πιθανότητες.

c.

Επειδή τα δείγματα βαθμών δεν είναι ανεξάρτητα μεταξύ τους θα εξετάσουμε την διαφορά των μέσων $\mu=\mu_{\pi}-\mu_{\mu}$,όπου μ_{π} και μ_{μ} οι μέσοι βαθμοί του πληθυσμού των φοιτητών για τις πιθανότητες και τα μαθηματικά αντίστοιχα,και θα κάνουμε τον δίπλευρο έλεγχο $H_0:\mu=0$.

```
> t.test(dif)

        One Sample t-test

data:  dif
t = -2.6213, df = 100, p-value = 0.01013
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.8697422 -0.1203568
sample estimates:
 mean of x
-0.4950495
```

Παρατηρούμε οτι p-value = 0.01013,επομένως απορρίπτεται η μηδενική υπόθεση.