Esercizio 1

$H_0 : \mu = \mu_0 = 100$
$H_1 : \mu \neq \mu_0$

Poiché n<30 e la deviazione standard non era nota a priori ma è stata stimata (s=14.4), la statistica test da utilizzare è t:

$$t_{oss} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{112.8 - 100}{\frac{14.4}{\sqrt{18}}} = 3.77$$

Sotto H_0, t si distribuisce come una t di student con n-1=17 g.d.l.

$t_{crit}=t_{17,0.05}=-2.11$
$t_{crit}=t_{17,1-0.05}=2.11$

$t_{oss} > t_{17,0.05}$ $\Longrightarrow$ Rifiuto H_0

Esistono evidenze che mostrano che il peso dei soggetti diabetici è mediamente diverso dal peso ideale.

Alternativamente era possibile determinare:
α_{oss} = p-value =Probabilità di ottenere, sotto H_0, un valore uguale o maggiore rispetto a quello realmente osservato = $P((t>3.77)\cup(t<-3.77))=0.002$
Poiché $\alpha_{oss}<\alpha=0.05$, RIFIUTO H_0

Esercizio 2

a) Sulla base della tabella precedente costruite la seguente nuova tabella:

	Asma Intermittente	Asma Persistente	Totale
Broncostruzione		n	
Sì	64	151	215
No	474	301	775
Totale	538	452	990

b) Valutate con opportuno test statistico se la proporzione di broncostruiti nel gruppo dei soggetti con asma <u>intermittente</u> è significativamente diversa da quella che si osserva nei soggetti con asma <u>persistente</u> (stabilite l'ipotesi da testare il tipo di test e la soglia critica, eseguite il test e se rifiutate l'ipotesi nulla stabilite l'errore di tipo alfa che commettete)

$H_0 : \pi_I = \pi_P$
$H_1 : \pi_I \neq \pi_P$

Test Z

$Z_{\text{critico}, \alpha=0.05} = \pm 1.96$

$p_I = 64/538 = 0.119$
$p_P = 151/452 = 0.334$

$$\hat{p} = \frac{p_I n_I + p_P n_P}{n_I + n_P} = \frac{64+151}{538+452} = \frac{215}{990} = 0.22$$

$$ES(p_I - p_P) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_I} + \frac{\hat{p}(1-\hat{p})}{n_P}} = \sqrt{0.22(1-0.22)\left(\frac{1}{538} + \frac{1}{452}\right)} =$$

$$\sqrt{0.1716(0.0019+0.0022)} = \sqrt{0.1716 * 0.0041 = \sqrt{0.00070356}} = 0.027$$

$$Z_{oss} = \frac{p_I - p_P}{ES(p_I - p_P)} = \frac{0.119 - 0.334}{0.027} = \frac{-0.215}{0.027} = -7.96$$

Z_{oss} cade nella Regione di Rifiuto, quindi **Rifiuto H_0**: la proporzione di broncostruiti nel gruppo dei soggetti con asma <u>intermittente</u> è significativamente diversa da quella che si osserva nei soggetti con asma <u>persistente</u>

p-value= P(Z<-7.96)+ P(Z>7.96) <0.001

Esercizio 3

Nell'ambito dell'indagine internazionale sui disturbi respiratori European Community Respiratory Health Survey (ECRHS) II, è stato somministrato un questionario a 527 soggetti italiani residenti a Verona, Pavia e Torino. Per gli stessi soggetti, personale opportunamente addestrato ha applicato all’esterno della finestra della cucina un campionatore passivo per rilevare il livello di concentrazione “*outdoor*” di NO_2 dell’abitazione.

Concentrazione di NO_2 outdoor ($\mu g/m^3$)	**Totale**	**Frequenze attese ***
[0-30)	95	101
[30-50)	155	
[50-60)	118	
[60-140)	159	243
Totale	527	

a) Assumendo che la distribuzione della concentrazione di NO_2 outdoor ($\mu g/m^3$) sia normale con la stessa media e la stessa d.s. del campione studiato (57 e 31 $\mu g/m^3$ rispettivamente), è stato calcolato il numero di soggetti attesi (*) per alcune delle classi riportate nella tabella; calcolate il numero di soggetti attesi nelle classi di concentrazione dell’NO_2 outdoor [30-50) e [50-60).

Assumendo i parametri della popolazione $\mu=57\ \mu g/m^3$ e $\sigma=31\ \mu g/m^3$ possiamo andare a calcolare i 3 valori di Z associati ai valori :

I) $[NO_2] = 30\ \mu g/m^3$
II) $[NO_2] = 50\ \mu g/m^3$
III) $[NO_2] = 60\ \mu g/m^3$

Andremo a sostituire questi tre valori nella formula:

$$z = \frac{x-\mu}{\sigma} = \frac{x - 57\frac{\mu g}{m^3}}{31\frac{\mu g}{m^3}}$$

Perciò

IV) $Z\_30\ \mu g/m^3 = -0.87$
V) $Z\_50\ \mu g/m^3 = -0.23$
VI) $Z\_60\ \mu g/m^3 = 0.10$

Ci serviranno questi 3 valori di Z poiché andremo a calcolare la probabilità compresa nell’intervallo [30-50) e nell’intervallo [50-60). Una volta che avremo calcolato queste due probabilità le moltiplicheremo per 527 ed avremo il numero di attesi per ogni categoria.
Avremo perciò

VII) P((Z>-0,87) and (Z<-0,23))=P((z=-0.23)-P(z=-0.87)=0.4090-0.1922=0.2168

VIII) P((Z>-0,23) and (Z<0,10))= P((z=0.10)-P(z=-0.23)=0.5398-0.4090=0,1308

Ora moltiplicheremo le probabilità calcolate per 527 e avremo il numero di attesi per classe:

IX) E[30-50)=0,2168*527=114.25
X) E[50-60)=0,1308*527=68,9

b) Valutate mediante un opportuno test statistico se esiste discrepanza tra i valori di concentrazione dell'NO_2 outdoor osservati e quelli attesi nell'ipotesi decritta al punto a). Stabilite il tipo di test, i gradi di libertà, la soglia critica e se rifiutate l'ipotesi nulla stabilite l'errore di tipo alfa che commettete. Interpretate a parole la decisione presa sulla base del test.

Utilizzeremo un test del chi quadro che va a valutare quanto un valore osservato (per ogni cella) si discosta da un valore atteso calcolato sotto l'ipotesi che la distribuzione di NO2 segua una normale (l'ipotesi nulla).

La statistica del test viene calcolata come $\chi^2 = \sum \frac{(O-E)^2}{E}$

Una volta calcolato il valore verrà confrontato con il valore di soglia e verrà rifiutata/non rifiutata l'ipotesi nulla.

Formalizzando:

I) formulazione ipotesi
   H0: il campione estratto proviene da una popolazione che ha distribuzione normale.
   H1: il campione estratto proviene da una popolazione che non ha distribuzione normale.

II) Scelta regione critica e gradi di libertà

Alfa viene settato pari a 0.05. I g.d.l. vengono calcolati come numero celle –numero parametri stimati per il calcolo degli attesi -1 e cioè 4-2-1=1.

$\chi^2_{\alpha=0,05;g.d.l=1} = 3.84$

Rifiuterò l'ipotesi nulla se il valore calcolato sarà maggiore di 3.84

III) Calcolo del chi quadro

Chiquadro =

$\frac{(95-101)^2}{101} + \frac{(155-114.25)^2}{114.25} + \frac{(118-68.9)^2}{68.9} + \frac{(159-243)^2}{243}$ =0,35+14.53+34.99+29.03

Chiquadro=78.91

IV) Poiché Chiquadro calcolato è maggiore della soglia critica (78.91>3.84) non accetto l'ipotesi nulla: il campione estratto proviene da una popolazione che non ha distribuzione normale.
V) Calcolo del p value: Con 1 grado di libertà, il valore 78.91 è maggiore del valore del Chiquadro che lascia a destra una probabilità pari a p=0.0001, cioè 15.
Perciò avrò una significatività p<0.0001.

Esercizio 4

a) Determinate l’errore standard della differenza delle medie nella distanza percorsa in 6 minuti per i 2 gruppi.

Errore std = $\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$

Errore std = 8.747

b) Secondo voi, la distanza percorsa in 6 minuti da un soggetto nel gruppo di coloro che riportano sintomi cronici di tosse/catarro è significativamente diversa da quella che percorre un soggetto nel gruppo di controllo?
b1. Stabilite l’ipotesi nulla

H0: $\mu_{casi} = \mu_{controlli}$

b2. Stabilite l’ipotesi alternativa

H1: $\mu_{casi} \neq \mu_{controlli}$

b3. Stabilite il tipo di test

test z

b4. Determinate la soglia critica ad un livello alpha=0.01

$z_{(0,01)} = 2.58$

b5. Eseguite il test

z = (584.4-604.2)/8.747 = 2.26

b6. In base ai risultati del test, rifiutate l’ipotesi nulla? Perché?

Non rifiuto l’ipotesi nulla perché il valore ottenuto dal test (z=2.26) è inferiore al valore critico di t calcolato in precedenza (z=2.58).

b7. In caso di rifiuto dell’ipotesi nulla, determinate la probabilità dell’errore di I tipo nel rifiutare H_0 (p-value).

p=0.0119*2=0.0238

Esercizio 5

Uno studio è stato condotto per valutare la relazione esistente tra la latitudine (x) e i tassi di mortalità per melanoma maligno (y) negli uomini nel periodo 1950-1959 negli Stati Uniti d'America. Nello studio sono stati considerati i tassi riferiti a 49 stati. La latitudine per gli stati inclusi nello studio va da 28° a 47.5° e la sua media è 39.5°; la media del tasso di mortalità è risultata 159.9 morti per 10 milioni.

a) Sapendo che $\sum(x_i - \bar{x})\ (y_i - \bar{y})$ = -6100.2, $\sum(y_i - \bar{y})^2$ = 53637.3, e che $\sum(x_i - \bar{x})^2$ = 1020.5, si calcolino i coefficienti di regressione della retta Y = a+bX.

b= -6100.2/1020.5= -5.978
a= 159.9-(-5.978*39.5)= 396.07

b) Sapendo che l'errore standard associato al parametro stimato b è pari a 0.598, si verifichi l'ipotesi che la stima del parametro della variabile indipendente sia diverso da zero con un opportuno test d'ipotesi (specificando l'ipotesi nulla, alternativa e la soglia critica). Si commentino i risultati.

H0: $\beta=0$ (assenza di associazione tra latitudine e mortalità per melanoma maligno)
H1: $\beta\neq 0$

Soglia critica: $t_{n-2} = t_{47} = \pm 2.01$ (circa)

t= -5.978/0.598= -9.996

Si rifiuta l'ipotesi nulla di assenza di associazione tra latitudine e mortalità per melanoma maligno: a latitudini inferiori, negli USA, corrispondono tassi di mortalità significativamente superiori.

c) Sulla base dei coefficienti di regressione stimati al punto a), qual è il valore atteso del tasso di mortalità per tumori maligni negli uomini che vivono ad una latitudine di 31°?

$Y_{(31)}$= 396.07-5.978*31= 210.71