Quesito 1. A barista is told that the optimal serving temperature for coffee is $180°F$. Five temperatures are taken of the served coffee: 175, 185, 170, 184, and 175 degrees. Find a 90% confidence interval of the form $(-\infty, b]$. La documentazione della funzione `t.test` riporta quanto segue:

```
t.test(x, ...)

## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)

## S3 method for class 'formula'
t.test(formula, data, subset, na.action, ...)

ARGUMENTS

x   a (non-empty) numeric vector of data values.

y   an optional (non-empty) numeric vector of data values.

alternative  a character string specifying the alternative hypothesis,
             must be one of "two.sided" (default), "greater" or "less".
             You can specify just the initial letter.
mu   a number indicating the true value of the mean (or difference in means
     if you are performing a two sample test).

paired         a logical indicating whether you want a paired t-test.

var.equal   a logical variable indicating whether to treat the two variances as being equal.
            If TRUE then the pooled variance is used to estimate the variance otherwise
            the Welch (or Satterthwaite) approximation to the degrees of freedom is used.

conf.level  confidence level of the interval.
```

1. Scrivere il comando completo per ottenere l'intervallo di cui sopra.

```
x=c(175, 185, 170, 184, 175)
t.test(x, alternative='less')
```

2. Che comando dovremo scrivere per valutare H_0 : la temperatura media è $180°F$ contro H_A : la temperatura media è diversa da $180°F$?

```
t.test(x, alternative='two.sided')
```

Quesito 2. Marie is getting married tomorrow at an outdoor ceremony in the desert. In recent years it has rained only 6 days each year. But the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 5% of the times. What is the probability that it will rain on the day of Marie's wedding?

Risposta

R — event: it rains on Marie's wedding

T_+ — event: the weatherman predicts rain

$\Pr(R) = 6/365 = 1.6\%$ — it rains 6 days out of 365

$\Pr(T_+|R) = 90\%$ — when it rains, rain is predicted

$\Pr(T_+|\neg R) = 5\%$ — when it does not rain, rain is predicted

$\Pr(T_+) = \Pr(T_+|R) \cdot \Pr(R) + \Pr(T_+|\neg R) \cdot \Pr(\neg R) = 6.4\%$

$$\Pr(R|T_+) = \frac{\Pr(R) \cdot \Pr(T_+|R)}{\Pr(T_+)} = 23.1\%$$

Risposta

Quesito 3. Ripetiamo 2 volte lo stesso Z-test (σ nota) a coda inferiore con campioni di dimensione crescente. Assumendo vera H_0, qual è la probabilità che in almeno uno di questi test il p-valore risulti ≥ 0.05 ?

Si scelga tra le seguenti opzioni la più opportuna.

1. La probabilità è $= \ldots$ (specificare)
2. La probabilità è $< \ldots$ (specificare)
3. La probabilità è $> \ldots$ (specificare)
4. Non ci sono sufficienti informazioni per stimare questa probabilità.

Risposta 1. La probabilità è $= 1 - (0.05)^2 = 0.9975$.

Quesito 4. Il comando `summary(lm(y ~ x))` produce il seguente output.

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1620 -2.1056 -0.2472  1.9249  5.2423

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.5413     0.7004  -0.773    0.443
x             1.1830     0.1312   9.016 1.25e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.919 on 58 degrees of freedom
Multiple R-squared:  0.5836,       Adjusted R-squared:  0.5764
F-statistic:  81.3 on 1 and 58 DF,  p-value: 1.246e-12
```

1. Che significato ha il numero `1.25e-12` nella nona riga?

È il p-valore del t-test fatto per valutare

H_0 : cefficiente angolare $= 0$ contro H_A cefficiente angolare $\neq 0$.

L'ipotesi nulla viene rifiutata, quindi la variabile x è significativa.

Quesito 5. Consider 15 subjects split at random into three groups. Each group is as- signed a month. For each group we record in the dataframe `data` the number of calories consumed on a randomly chosen day:

calories	month
2166	may
1568	may
2233	may
1882	may
2019	may
2279	sep
2075	sep
2131	sep
2009	sep
1793	sep
2226	d[illegible]
2154	dec
2583	dec
2010	dec
2190	dec

We assume that the amounts consumed are normally distributed with common variance but perhaps different means. Use one-way analysis of variance is to decide whether the difference in the sample means is indicative of a difference in the population means or is attributable to sampling variation.

```
aov(formula, data = NULL, ...)

ARGUMENTS

formula A formula specifying the model.

data    A data frame in which the variables specified in the formula
        will be found. If missing, the variables are searched for in
        the standard way.
```

1. Scrivere il comando completo per ottenere tramite la funzione `aov` testare l'ipotesi di cui sopra.

```
aov(calories ~ weight, data=data)
```

2. Se otteniamo il seguente risultato, cosa possiamo concludere?

```
            Df Sum Sq Mean Sq F value Pr(>F)
month        2 174664   87332   1.786  0.209
Residuals   12 586720   48893
```

È il p-valore dell'F-test *non* è sufficientemente piccolo per concludere che la differenza tra le medie campionarie è indicativa di una differenza tra le medie di popolazione.