

Studio statistico sulla meteorologia marziana



Massimiliano Galvagno

Luca Imeneo

Giulia Tosetti

Introduzione

Il seguente studio statistico è volto all'analisi metereologica dello scorso anno marziano.

Un anno marziano è diviso in dodici periodi non uguali fra loro, infatti la durata media di uno di essi è una quantità variabile tra 46-67 giorni. La caratteristica inclinazione assiale del pianeta ($25,19^\circ$), simile alla Terra, fa sì che si percepiscano cambiamenti climatici stagionali. Nell'emisfero sud a partire dell'autunno, il cui inizio coincide con l'inizio dell'anno marziano, si susseguono le stagioni, come sul nostro pianeta. In particolare, il progetto si propone di confrontare tra loro l'inverno e l'estate dell'emisfero meridionale.

I dati utilizzati sono stati raccolti giornalmente dal rover Curiosity presso il cratere Gale ⁽¹⁾, il quale lavora alla missione della NASA *Mars Science Laboratory* (MSL). Ad oggi il dataset è reperibile sul sito rover environmental monitoring station ⁽²⁾.

Per l'analisi del campione è stato utilizzato il software applicativo *R*, indicato particolarmente per lo studio statistico.

Il lavoro è articolato in una prima analisi descrittiva dei dati, volta ad acquisire una visione completa del record, grazie alla quale è stato possibile scegliere quali aspetti di quest'ultimo analizzare in dettaglio. Sulla base delle evidenze osservate dalla statistica descrittiva sono state elaborate delle affermazioni circa l'ambiguità della suddivisione delle stagioni marziane e alcune loro caratteristiche. Infine, grazie alle conoscenze acquisite tramite l'analisi eseguita, è stato possibile supporre e successivamente testare, l'esistenza di leggi che descrivono l'evoluzione temporale o mettono in correlazione, i diversi parametri del Pianeta Rosso.

Presentazione del campione

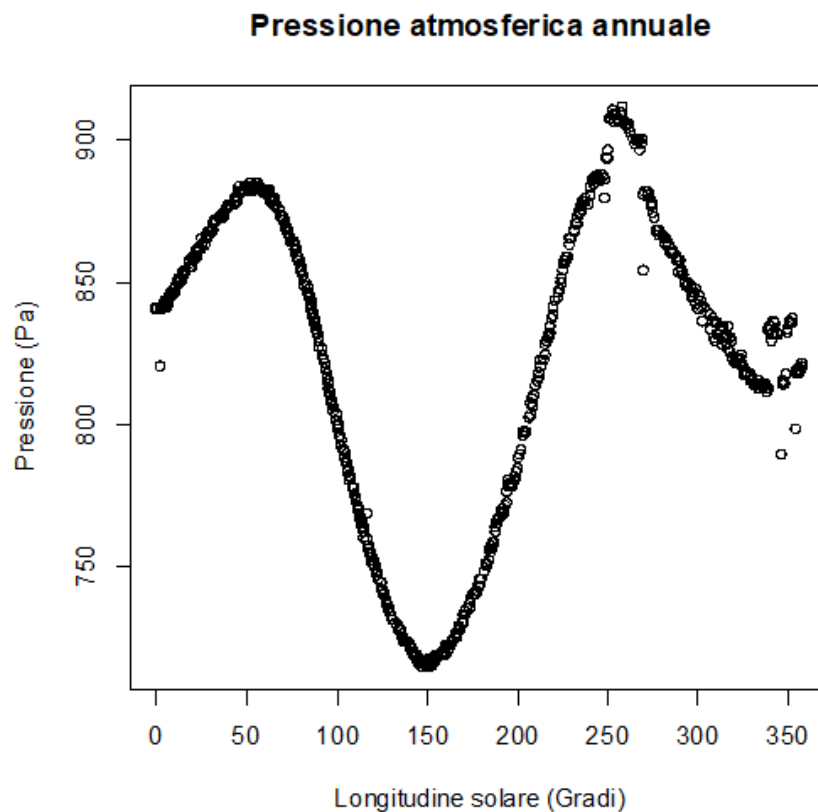
I dati forniti dalla sonda dalla sonda spaziale Curiosity, utilizzati nel progetto, sono:

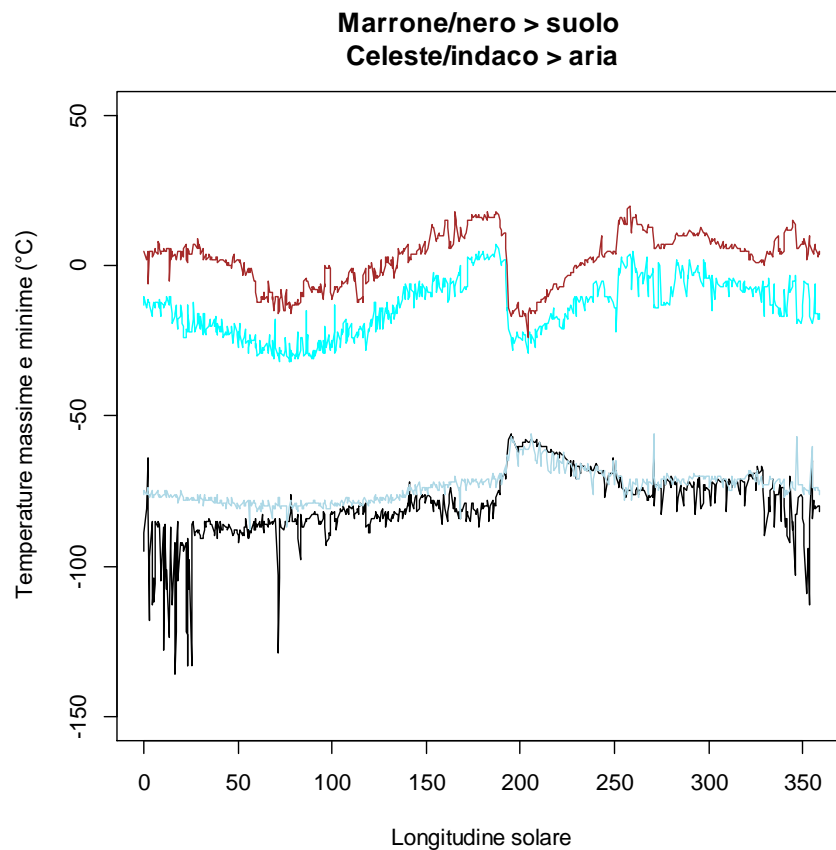
- Temperatura massima dell'aria;
- Temperatura minima dell'aria;
- Temperatura massima del suolo;
- Temperatura minima del suolo;
- Pressione atmosferica;
- Indice UV.

È inoltre nota la longitudine solare di ogni giorno, sulla quale è definito come mese il periodo di tempo in cui il pianeta spazia un angolo di 30° nella sua orbita.

Al fine di poter lavorare opportunamente sul data-set, i parametri devono essere considerabili variabili aleatorie tra loro identicamente distribuite e indipendenti.

Studio variabili temperatura e pressione:





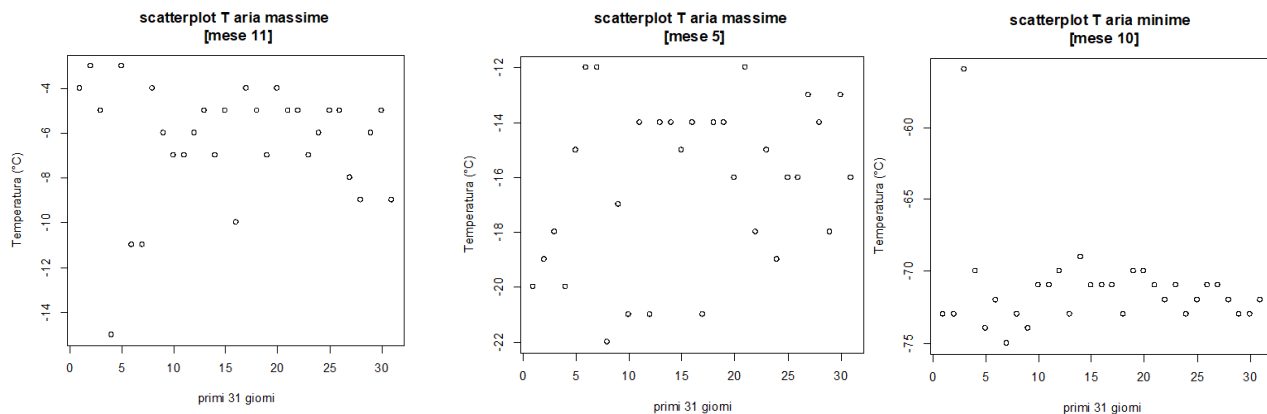
Le variabili aleatorie temperatura e pressione giornaliere, a causa dei trend climatici, non sono considerabili identicamente distribuite né su un arco di tempo annuale, né su uno mensile. Dall'andamento dal loro plot annuale si può supporre un andamento periodico. In particolare, la pressione sembra avere un trend sinusoidale.

È interesse di questo studio soffermarsi sui parametri di temperatura. Non essendo questi i.i.d. , occorre ridurre l'intervallo di giorni considerato di ogni mese, fino alla durata di un mese terrestre (31 giorni).

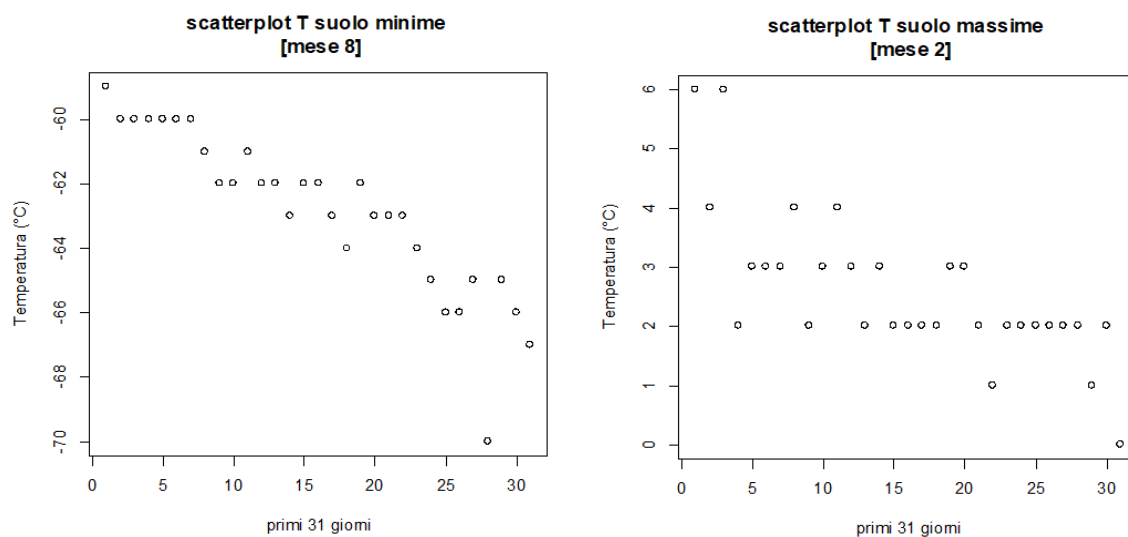
Per poter verificare l'ipotesi d'indipendenza è necessario studiare l'andamento degli scatterplot dei campioni scelti: un riscontro positivo si ottiene quando il grafico non presenta alcun trend definibile.

Si è quindi **studiato** il grafico temporale di tutte le **variabili** prese in esame, relative ai **primi 31 giorni di ogni mese**.

Scatterplot come questi sono stati considerati idonei, non sono individuabili tendenze particolari, al netto al più di eventuali outliers.



Alcuni campioni tuttavia sono stati scartati, per le chiare eteroschedasticità presentate. Se ne riporta qualcuno a titolo d'esempio.



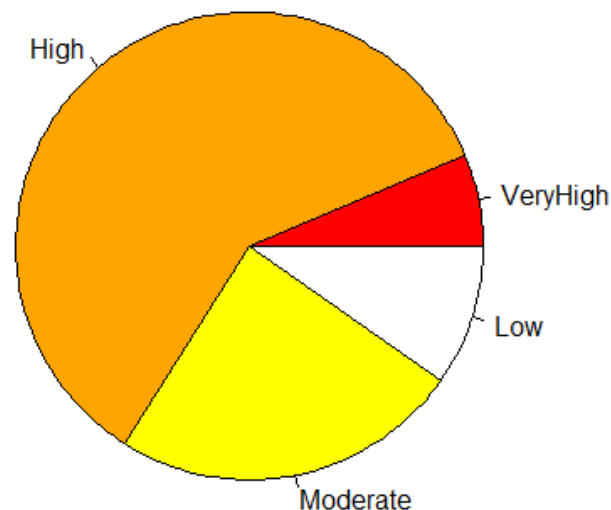
Dall'analisi di tutti i campioni è emerso che tra i dati che rispettano le ipotesi del modello statistico figurano tutte le **temperature aria estive** (mesi 10-11-12) ed **invernali** (mesi 4-5-6). I mesi autunnali e primaverili presentano leggeri trend ascendenti o discendenti: l'intervento di alcune variabili non considerate nel progetto potrebbero esserne la causa.

È riportato in calce al documento l'insieme completo degli scatterplot esaminati.

Studio variabile UV:

Con “indice UV” si indica il livello di radiazioni ultravioletta presente sul Pianeta Rosso. La variabile si presenta in forma categorica: essa è divisa in quattro classi “Low”, “Moderate”, “High”, “Very High”.

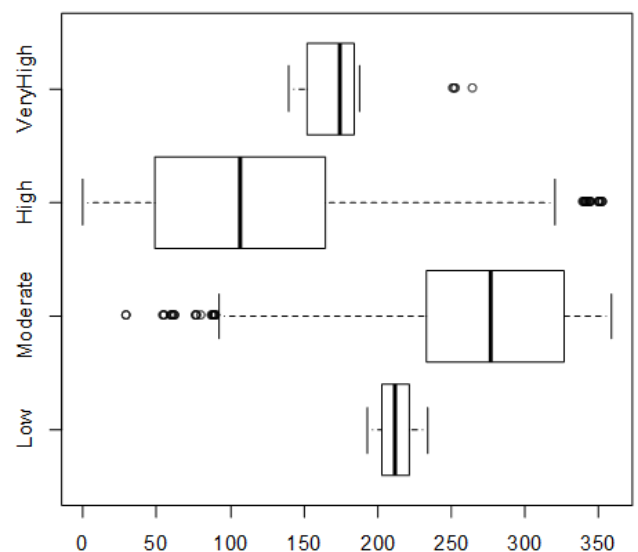
Distribuzione annuale radiazioni



summary(radiazioni)

Low	Moderate	High	Very High
65	161	397	42

Dal grafico a torta si osserva che sul suolo marziano le radiazioni assumono per la maggiore il livello High, poi a seguire Moderate, Low e Very High. Dai boxplot relativi all'intero anno si osserva che questo parametro assume determinati livelli quasi esclusivamente in certi periodi dell'anno, ciò fa pensare a un andamento periodico, andamento che non può essere predetto data la categoricità della variabile

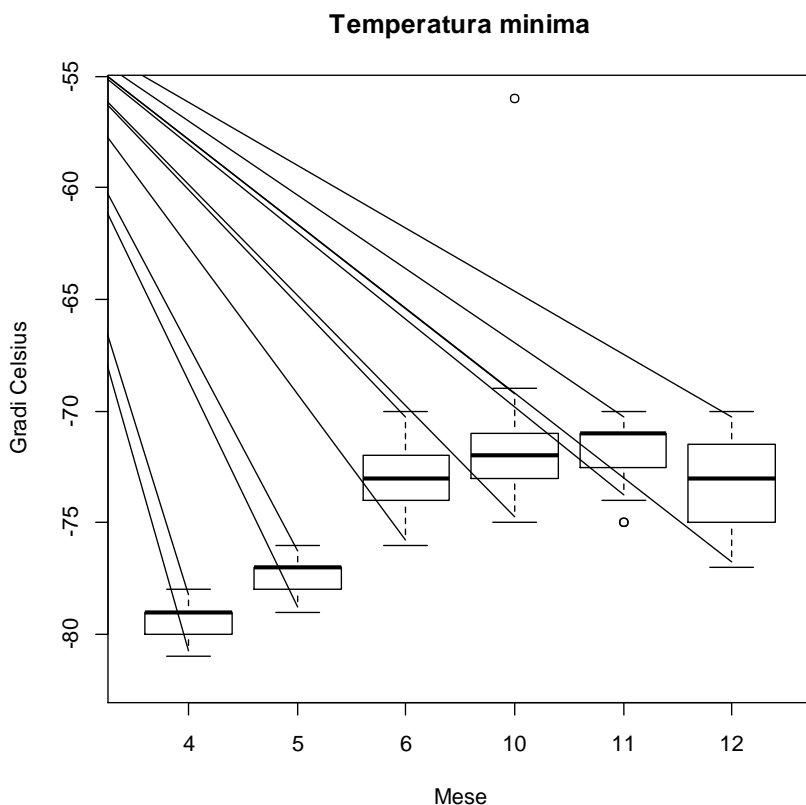


Alla luce delle suddette conclusioni sul campione si decide di proseguire nello studio della meteorologia del Pianeta Rosso concentrandosi prevalentemente alle temperature dell'aria.

Statistica descrittiva

Si prosegue con lo studio delle temperature dell'aria, parametro d'elezione per la restante parte del lavoro.

Analisi temperatura dell'aria:



Inverno

summary(m4)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-81.00	-80.00	-79.00	-79.39	-79.00	-78.00

summary(m5)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-79.00	-78.00	-77.00	-77.48	-77.00	-76.00

summary(m6)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-76.00	-74.00	-73.00	-73.19	-72.00	-70.00

Estate

summary(m10)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-75.00	-73.00	-72.00	-71.32	-71.00	-56.00

summary(m11)

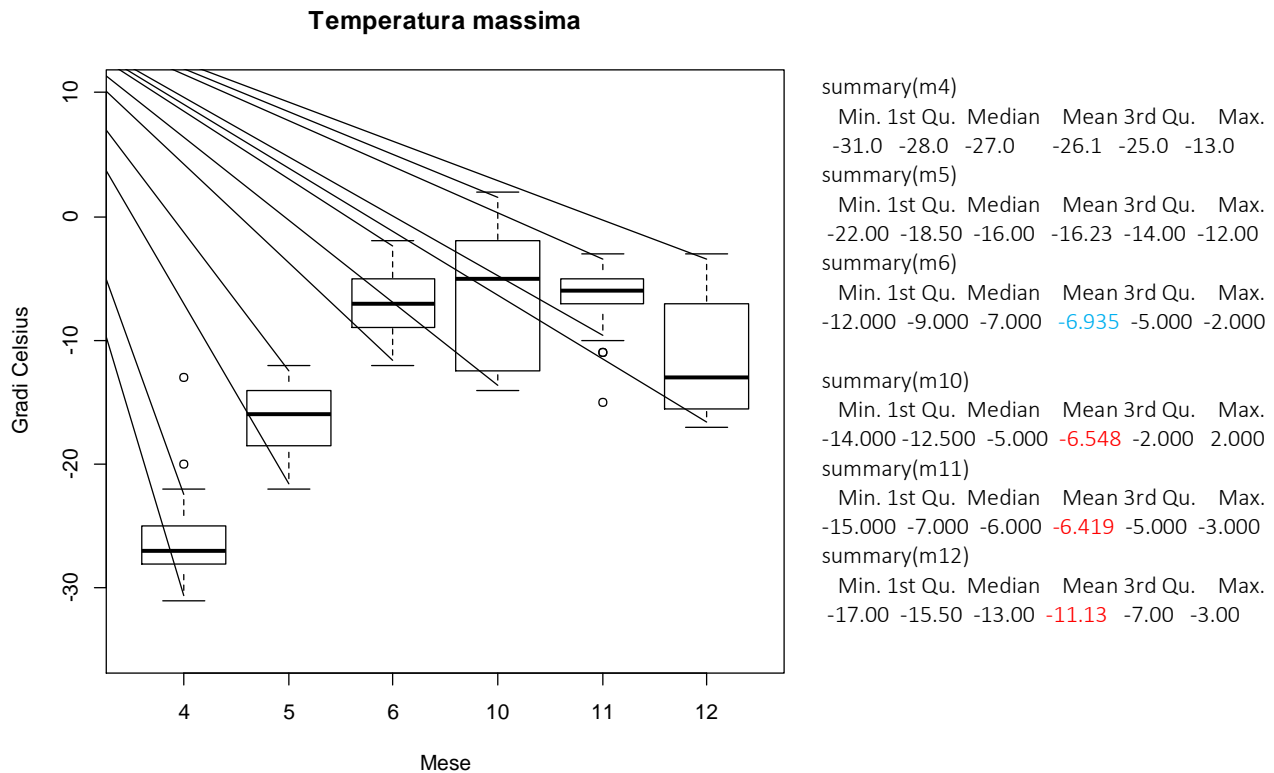
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-75.00	-72.50	-71.00	-71.65	-71.00	-70.00

summary(m12)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-77.00	-75.00	-73.00	-73.06	-71.50	-70.00

Sono raffigurati i boxplot delle temperature minime per i tre mesi di ogni stagione. I dati presentano una variabilità relativamente contenuta, media e mediana vicine seppur all'interno di range interquartilici ridotti. È presente una leggera tendenza ascendente nell'inverno. Il campione è quasi esente da outliers; nella statistica inferenziale il valore erratico del mese 10 è stato rimosso.

Tuttavia, si notano temperature medie molto vicine (si vedano i mesi 6, 12). In questo senso la demarcazione tra estate e inverno, seppur presente, è tutt'altro che netta.

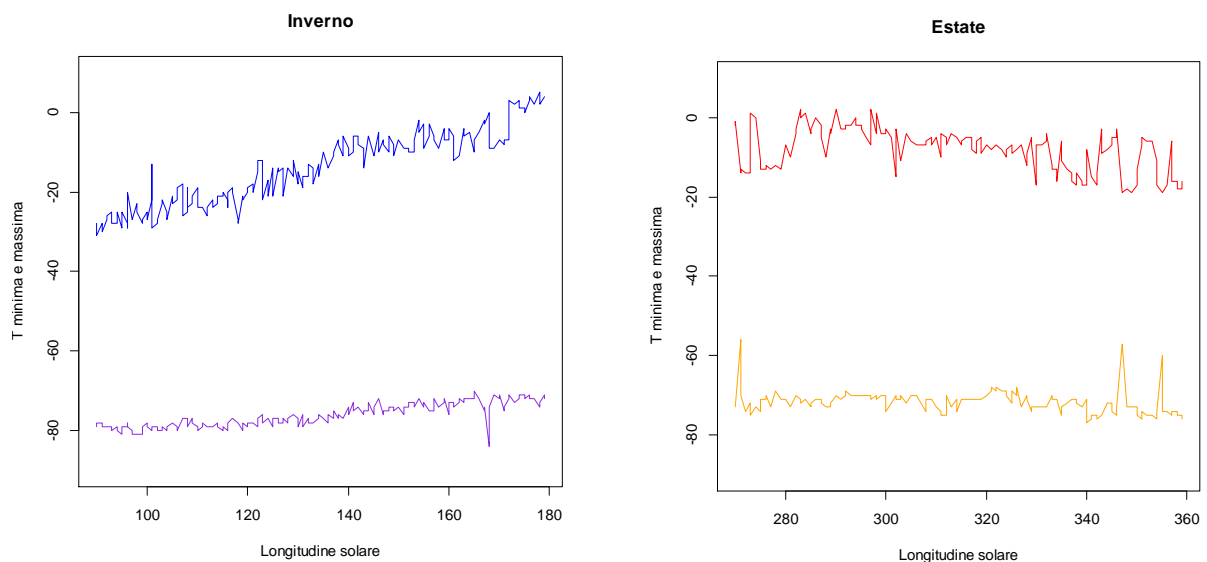


Le temperature massime mostrano invece maggiore variabilità con un IQR significativamente più ampio. Si accentua leggermente la presenza di code, soprattutto nei mesi invernali.

Si riconferma la tendenza ascendente precedentemente osservata e la forte varianza di temperature, che non mostrano una dipendenza rigida dalla stagione di appartenenza, al contrario di quanto avviene sul Pianeta Terra.

Questo aspetto verrà ripreso e commentato nell'ambito della statistica inferenziale.

Per esaminare meglio l'andamento delle temperature sopracitate nel corso di estate e inverno, si è scelto di riportarne un grafico integrale, rispetto alla posizione orbitale marziana:



Le temperature massime invernali aumentano chiaramente nel tempo, confermando quanto prima osservato nei boxplot.

Gli altri parametri non manifestano tendenze marcate; tuttavia, si riscontra che la differenza tra le temperature massime e minime giornaliere si attesta in entrambe le stagioni intorno ai 50-60 °C.

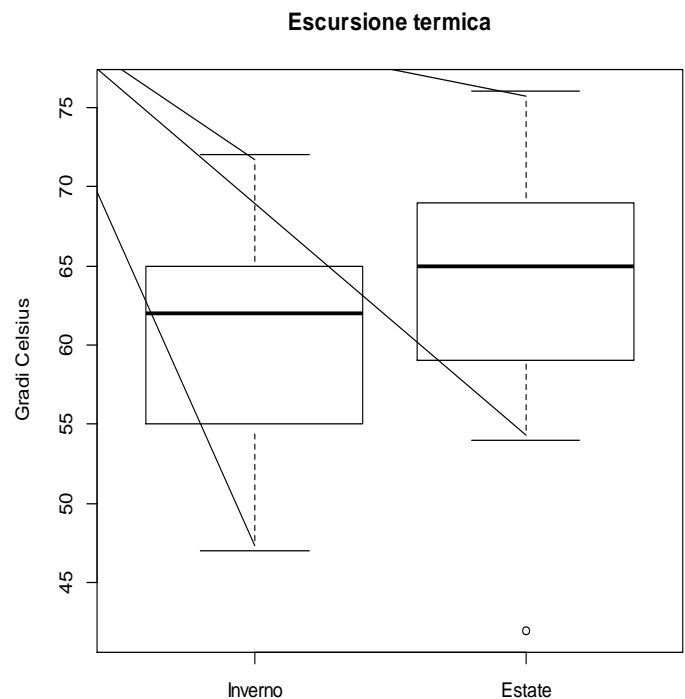
Analisi escursione termica:

Si riporta un boxplot per il parametro giornaliero:

$$\text{escursione termica} := t_{\max} - t_{\min}$$

I grafici mostrano la zona di maggior concentrazione dei valori assunti dal parametro nel corso di entrambe le stagioni.

I due grafici a scatola piuttosto simmetrici evidenziano valori sostanzialmente simili, leggermente maggiori per l'estate.

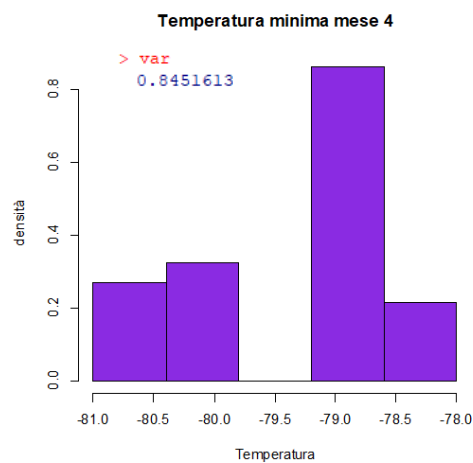


Anche la somiglianza tra le escursioni termiche nelle due stagioni verrà esaminata nel seguito del lavoro.

L'illustrazione dei sei campioni in esame prosegue attraverso lo studio della loro distribuzione ed eventuale gaussianità.

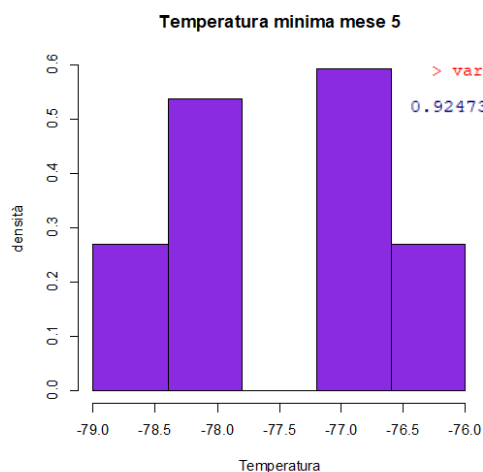
Analisi normalità campioni:

Vengono presentati in sequenza gli istogrammi relativi alle temperature massime e minime dei primi trentun giorni dei 6 mesi considerati, per fornirne un'idea qualitativa, accompagnati dalle rispettive varianze. La gaussianità è confermata o confutata da altrettanti test di Shapiro-Wilks, particolarmente indicati per campioni di questa numerosità ($n = 31$). Si inizia dai mesi invernali:



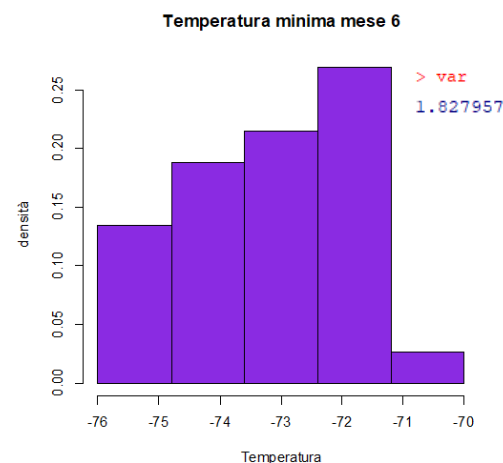
Shapiro-Wilk normality test

data: x
W = 0.84092, p-value = 0.0003264



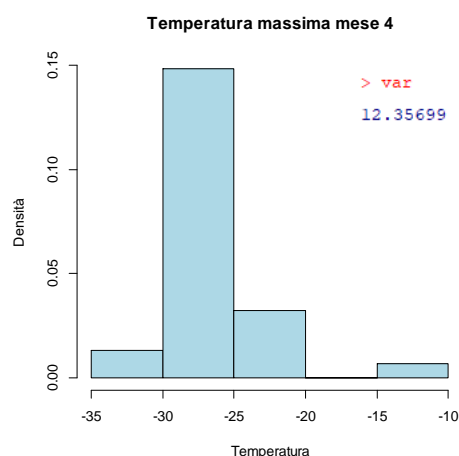
Shapiro-Wilk normality test

data: x
W = 0.88451, p-value = 0.003025



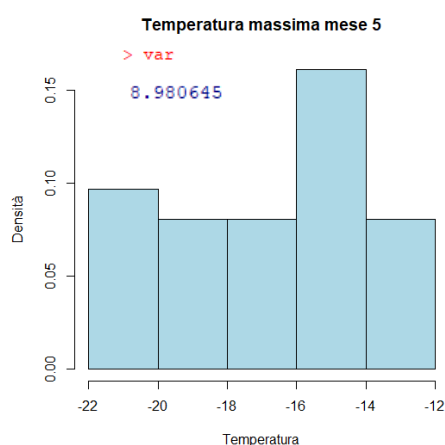
Shapiro-Wilk normality test

data: x
W = 0.92188, p-value = 0.02648



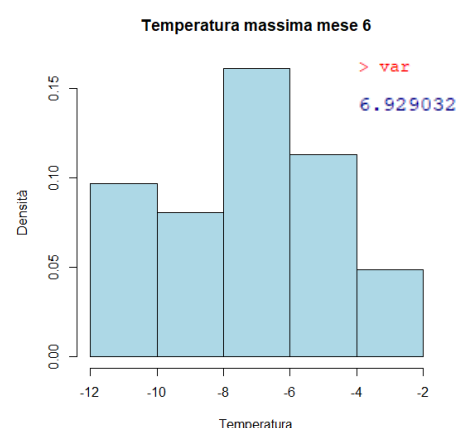
Shapiro-Wilk normality test

data: x
W = 0.83448, p-value = 0.0002408



Shapiro-Wilk normality test

data: x
W = 0.92821, p-value = 0.0392

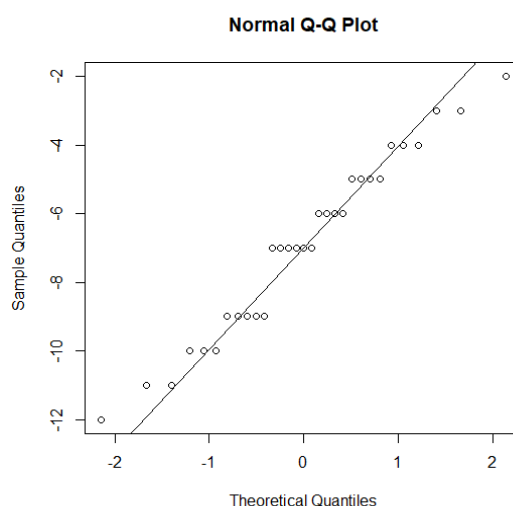


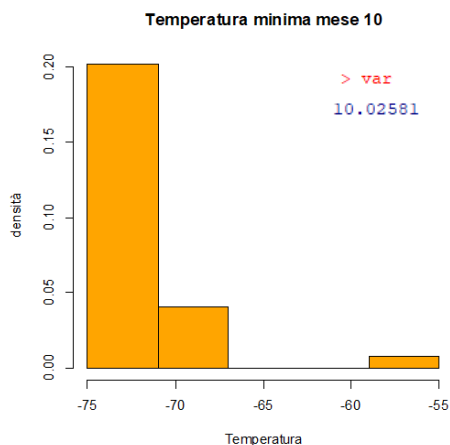
Shapiro-Wilk normality test

data: x
W = 0.96538, p-value = 0.4018

I campioni non presentano densità normale nei mesi invernali, con eccezione della temperatura massima nel mese 6. Con un p-value del 40% non è possibile rifiutare l'ipotesi nulla di gaussianità.

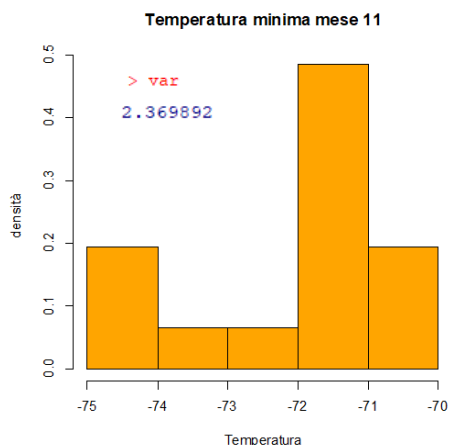
Si riporta il Q-Q plot corrispettivo per completezza. Nonostante l'aspetto "a gradini", imputabile al fatto che il campione assume solo valori interi, il trend generale è buono. Si riconferma il risultato dello Shapiro-Test





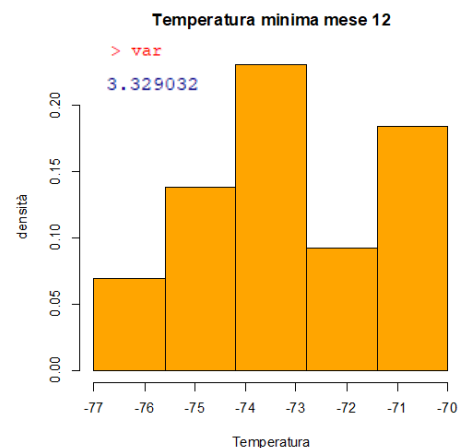
Shapiro-Wilk normality test

data: x
W = 0.58553, p-value = 3.472e-08



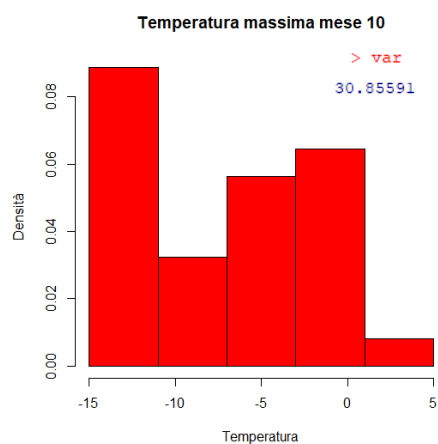
Shapiro-Wilk normality test

data: x
W = 0.80728, p-value = 7.061e-05



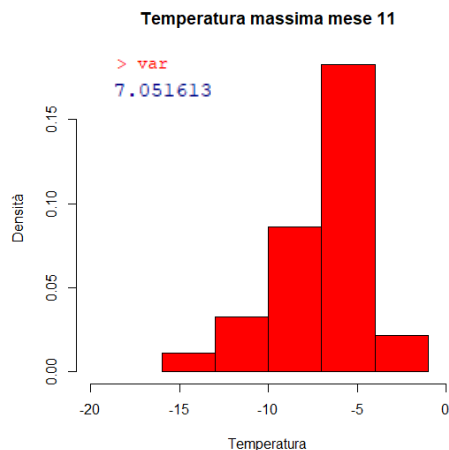
Shapiro-Wilk normality test

data: x
W = 0.92873, p-value = 0.04049



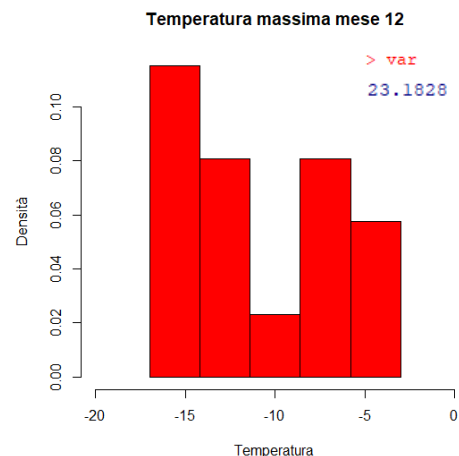
Shapiro-Wilk normality test

data: x
W = 0.88749, p-value = 0.003563



Shapiro-Wilk normality test

data: x
W = 0.87327, p-value = 0.001654



Shapiro-Wilk normality test

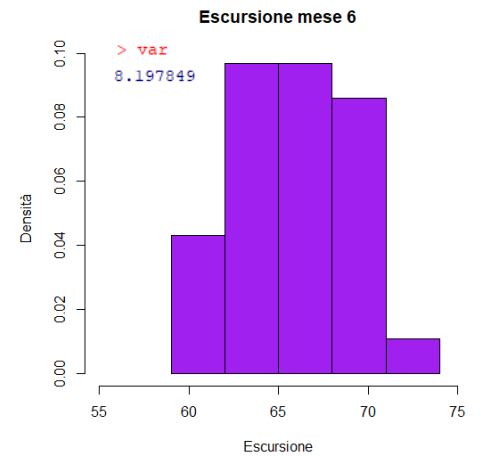
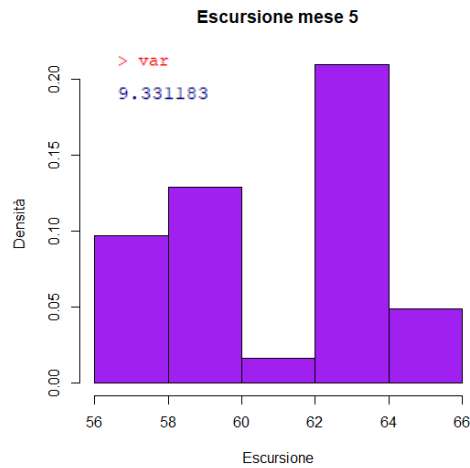
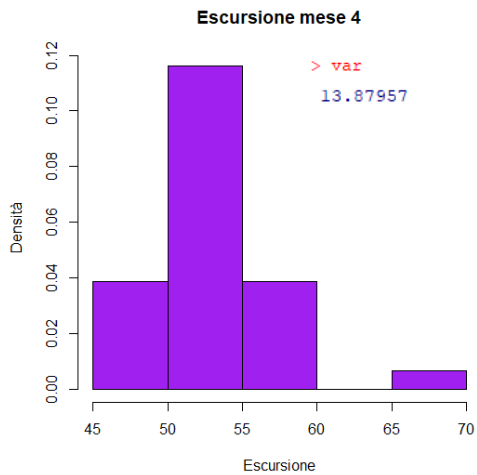
data: x
W = 0.8972, p-value = 0.006141

Nei mesi estivi i parametri considerati non evidenziano particolari tracce di gaussianità. I dati seguono distribuzioni bimodali, asimmetriche o irregolari, come per i mesi invernali.

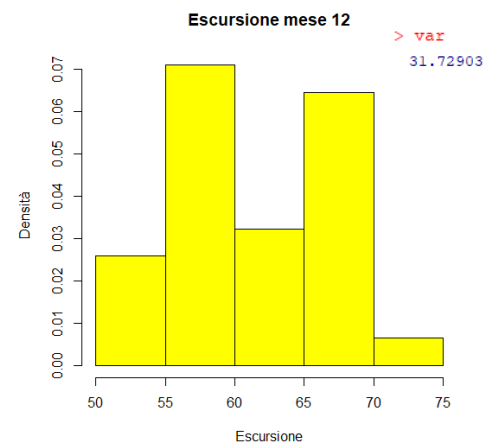
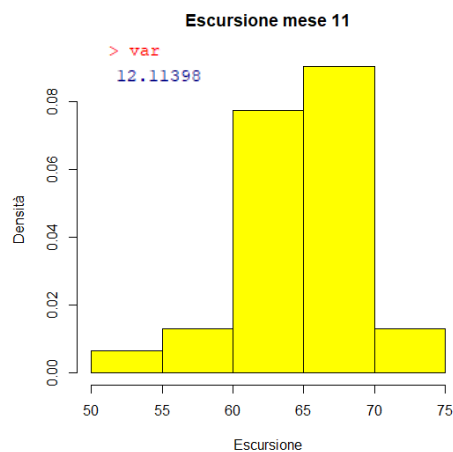
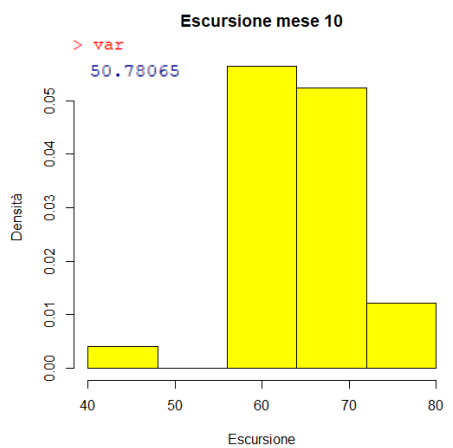
Nel seguito del progetto ci si limita a trattare i campioni di temperature massime e minime solamente come numerosi ($n > 30$) e non gaussiani, mancando evidenze di distribuzione normale per la maggior parte dei campioni.

L'analisi termina esaminando il nuovo parametro definito: l'escursione termica.

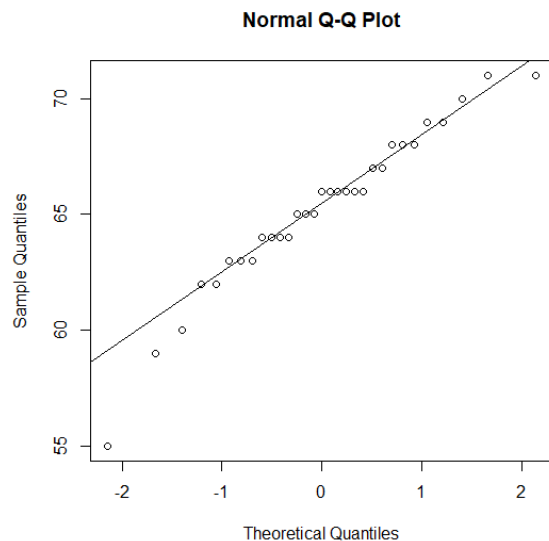
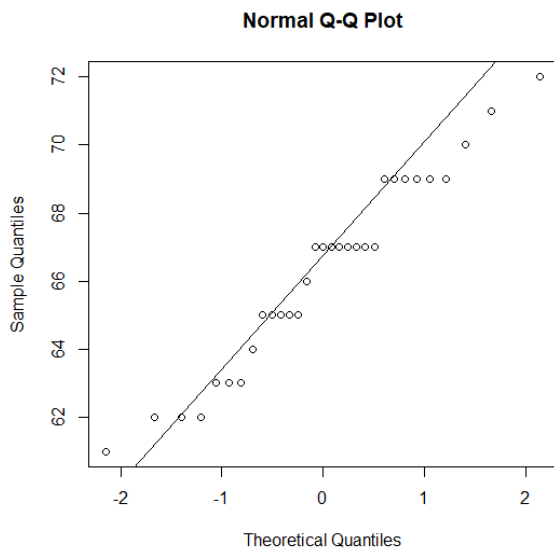
Inverno



Estate



Anche le escursioni termiche, come il resto dei dati, non appaiono in generale gaussiani, con due eccezioni. Completiamo l'analisi dei mesi 6 e 11:



I Q-Q plot rimangono buoni e ricordano ancora che il campione maneggiato è discreto. Il mese 6 si comporta meno bene del 12, ma ancora in accordo col test di Shapiro-Wilks.

Test d'ipotesi

Anormalità del mese 6:

Si vuole scoprire se su Marte i mesi invernali ed estivi seguono la concezione stagionale comune: si consideri quindi l'inverno la stagione più fredda dell'anno e l'estate quella più calda. Sebbene sulla Terra sia impensabile che un mese estivo possa avere temperatura *media* inferiore a uno invernale, dall'analisi del campione appare proprio questo.

Si propone quindi un test d'ipotesi per verificare che il mese invernale 6 sia più caldo del 12.

- Si è scelto di definire più caldo di un altro un mese che presenta valore atteso delle temperature massime maggiore.

È noto che dove un campione non presenti densità gaussiana, ma sia numeroso, è possibile applicare il TLC, che garantisce in modo approssimato la gaussianità della media campionaria \bar{X}_{31} .

Non si conosce la varianza e il campione è numeroso si sceglie quindi di effettuare uno Z-test, dove

$$Z_0 = \frac{\bar{X}_m - \bar{Y}_n - \delta_0}{\sqrt{\frac{S_x^2}{m} + \frac{S_y^2}{n}}}$$

e $\delta_0=0$.

“Il mese 12 è più freddo del mese 6?”

$H_0: \mu_6 < \mu_{12}$

$H_1: \mu_6 \geq \mu_{12}$

$RC = Z_0 > Z_{1-pvalue}$

```
> z.test(x,y,alternative="greater",mu=0,sigma.x=2.63,sigma.y=4.81,conf.level=0.95)
```

Two-sample z-Test

Il test fornisce un p-value pari allo 0.00102%.

data: x and y

La conclusione risulta forte: l'ultimo mese

z = 4.2591, p-value = 1.026e-05

invernale è più caldo dell'ultimo mese estivo.

Si conferma dunque un'anomalia nel clima dei mesi 6 e 12 rispetto alla stagione di cui fanno parte.

Il mese 6 ha inoltre temperature massime medie molto vicine a quelle degli altri due mesi estivi. Si propongono intervalli di confidenza al 95% per la differenza tra i valori del mese 6 e quelli dei mesi 10 e 11.

$IC(0.95) \mu_6 - \mu_{11}$

(-2,552850 : 1,778656)

$IC(0.95) \mu_6 - \mu_{10}$

(-1,836627 : 0.804369)

In entrambi i casi gli intervalli sono molto ristretti.

Il mese 6, invernale secondo la definizione, appare per certi versi assimilabile ad uno estivo: la sua temperatura massima media dista appena qualche grado da quelle dei primi due mesi estivi ed è addirittura maggiore di quella del terzo. Tutto con confidenza del 95% almeno.

La dipendenza poco rigida delle temperature di un mese dalla stagione di appartenenza è spiegabile osservando un importante parametro fisico marziano: l'atmosfera.

È noto che Marte presenti un'atmosfera sensibilmente diversa da quella terrestre: essa è notevolmente più rarefatta e provoca meno effetto serra, a causa della sua composizione e della bassa gravità del pianeta. Ciò può spiegare le particolarità riscontrate nelle stagioni marziane. Infatti, le variazioni di temperatura che avvengono nelle stagioni terrestri sono dovute all'incidenza dei raggi solari sull'atmosfera, che li riflette in misura diversa se la colpiscono in modo più o meno perpendicolare.

È ragionevole pensare che un'atmosfera labile come quella del Pianeta Rosso non riesca a provocare tale fenomeno con una magnitudine tale da impattare effettivamente sul corso delle stagioni. Si suppone che prendano invece il sopravvento altri fenomeni, non analizzati nel seguente progetto.

Lo scarso impatto dell'atmosfera sul clima spiega anche altri fenomeni emersi nel corso dello studio. Si ricordano l'assenza di effetto memoria tra le temperature di un giorno e il successivo, assieme alle escursioni termiche estreme e la loro somiglianza nel corso delle stagioni.

Somiglianza escursioni termiche:

Diventano ora oggetto d'attenzione l'escursione termica, l'altro parametro che incuriosiva nella statistica descrittiva.

È proposto un primo campione casuale di mesi posti a confronto, per ogni coppia si effettua il medesimo test di disuguaglianza (H_1) tra le rispettive escursioni termiche:

"L'escursione termica del mese 5 è diversa dal mese 12?"

```
> z.test(x,y,alternative="two.sided",mu=0,sigma.x=3.05,sigma.y=5.63,conf.level=0.95)
```

$H_0: \mu_5 = \mu_{12}$

Two-sample z-Test

$H_1: \mu_5 \neq \mu_{12}$

$RC = |Z_0| > z_{1-\frac{pvalue}{2}}$

data: x and y

z = -0.58905, p-value = 0.5558

“L’escursione termica del mese 10 è diversa dal mese 6?”

```
> z.test(x,y,alternative="two.sided",mu=0,sigma.x=2.86,sigma.y=7.12,conf.level=0.95)
```

$H_0: \mu_{10} = \mu_6$

Two-sample z-Test

$H_1: \mu_{10} \neq \mu_6$

$RC = |Z_0| > Z_{1-\frac{pvalue}{2}}$

```
data: x and y
z = 1.0768, p-value = 0.2816
```

L’escursione termica del mese 11 è diversa dal mese 10?”

```
> z.test(x,y,alternative="two.sided",mu=0,sigma.x=3.48,sigma.y=7.12,conf.level=0.95)
```

$H_0: \mu_{11} = \mu_{10}$

Two-sample z-Test

$H_1: \mu_{11} \neq \mu_{10}$

$RC = |Z_0| > Z_{1-\frac{pvalue}{2}}$

```
data: x and y
z = 0.31729, p-value = 0.751
```

“L’escursione termica del mese 5 è diversa dal mese 6?”

```
> z.test(x,y,alternative="two.sided",mu=0,sigma.x=3.05,sigma.y=2.86,conf.level=0.95)
```

$H_0: \mu_5 = \mu_6$

Two-sample z-Test

$H_1: \mu_5 \neq \mu_6$

$RC = |Z_0| > Z_{1-\frac{pvalue}{2}}$

```
data: x and y
z = 0.60138, p-value = 0.5476
```

Soffermendosi sui p-value, non si ha significatività a nessun livello ragionevole, per ora.

Viene presentato un ultimo test sulle escursioni dei mesi 6 ed 11, che nella statistica descrittiva avevano presentato distribuzione normale.

Per ottenere un’approssimazione migliore di quella fornita dal TLC si vuole effettuare un t-test. Esso richiede popolazioni normali con varianza incognita ma uguale, si procede effettuando un opportuno F-test per comprovare quest’ipotesi:

“L’escursione del mese 11 e del mese 6 hanno una varianza significativamente diversa?”

$H_0: \sigma^2_{11} = \sigma^2_6$

```
> var.test(x,y, alternative="two.sided")
```

$H_1: \sigma^2_{11} \neq \sigma^2_6$

F test to compare two variances

$RC = F_0 < \frac{f_{pvalue}}{2} (m-1, n-1); F_0 := \frac{S^2_x}{S^2_y}$

```
data: x and y
F = 1.4777, num df = 30, denom df = 30, p-value = 0.2904
```

Il p-value del test risulta alto: la probabilità di commettere un errore di I specie è elevata.

Dunque non esiste evidenza per cui bisogna rifiutare H_0 . La conclusione è debole, tuttavia consente di procedere con il t-test.

Il p-value del test risulta alto: la probabilità di commettere un errore di I specie è elevata. Dunque non esiste evidenza per cui bisogna rifiutare H_0 . La conclusione è debole, tuttavia consente di procedere con il t-test.

“L’escursione termica del mese 6 è diversa dal mese 11?”

$$T_0 = \frac{\bar{X}_m - \bar{Y}_n - \delta_0}{\sqrt{S_p^2 \left(\frac{1}{m} + \frac{1}{n} \right)}} \text{ con } S_p^2 := \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2}$$

$H_0: \mu_6 = \mu_{11}$

$H_1: \mu_6 \neq \mu_{11}$

$RC = |T_0| > t_{1-\frac{pvalue}{2}}(m+n-2)$

```
> t.test(x,y, alternative="two.sided", conf.level=0.95)
```

```
Welch Two Sample t-test
```

```
data: x and y
t = -1.2752, df = 57.85, p-value = 0.2073
```

Il riscontro positivo su questo primo campione casuale spinge una nuova domanda: è possibile che le escursioni termiche mensili seguano una stessa legge?

Regressione lineare

Si è ricercato prima un modello per i primi 3 mesi, verificando poi se questo fosse valido anche per gli altri 3.

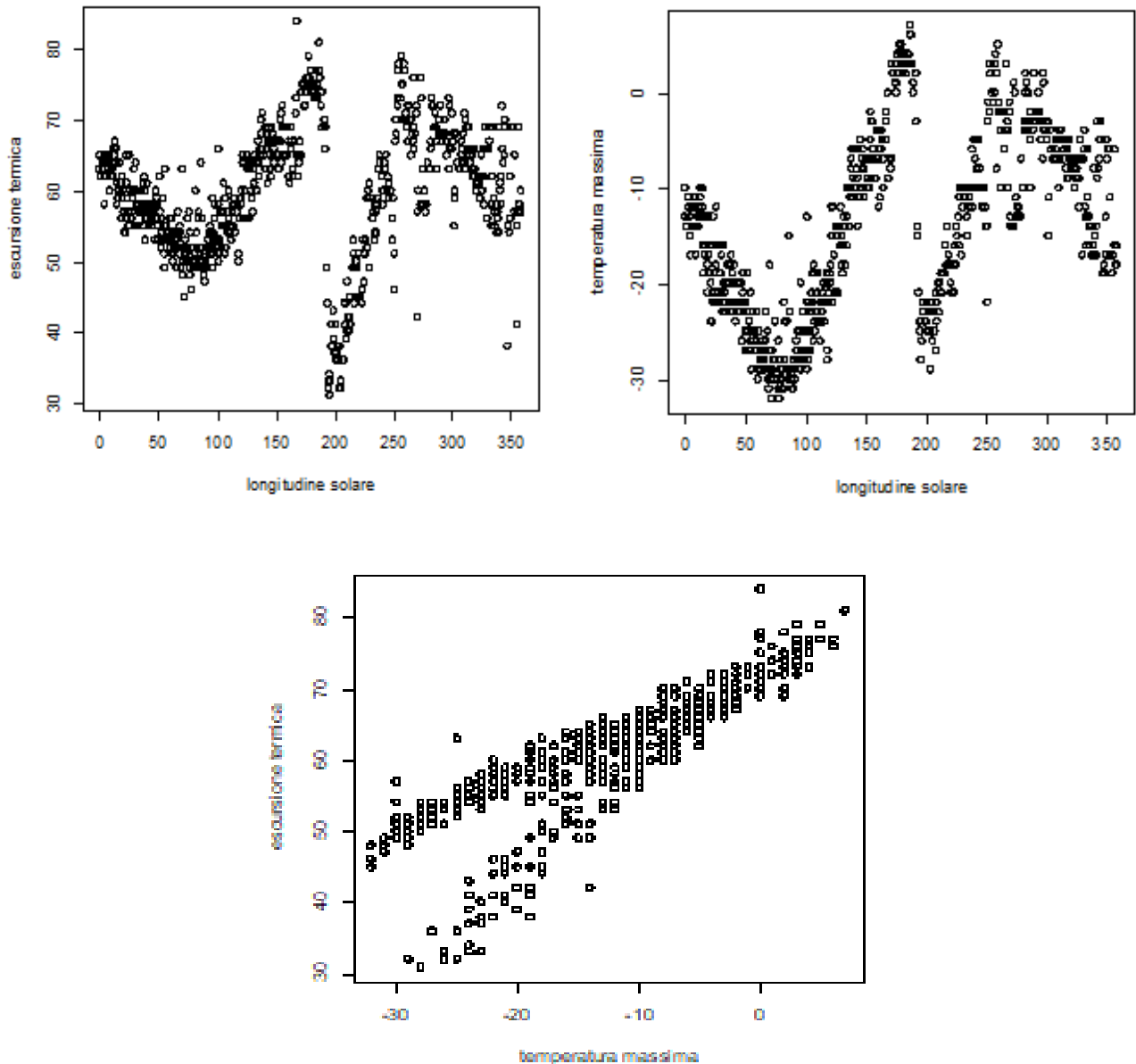
Al fine di poter effettuare una regressione su tutta una stagione, occorre raggruppare i singoli dataset dei tre mesi che la compongono in un unico record contenente i primi 31 giorni di ogni mese.

I dati al suo interno manterranno la proprietà di indipendenza, mentre, poiché appartengono a periodi di tempo molto lunghi, non è detto che siano tra loro identicamente distribuiti.

Tuttavia, ciò non è richiesto dalle ipotesi del modello lineare gaussiano.

Osservando l'andamento temporale dell'escursione termica si osserva una netta somiglianza con quello della temperatura massima dell'aria.

Effettuando alcuni scatterplot si nota un'apparente relazione tra i due dati.



Dai grafici si può supporre che l'escursione termica vari linearmente con la temperatura massima.

Considerando che lo studio è stato fatto su tre periodi di tempo non contigui, può essere opportuno aggiungere il regressore categorico mese, affinché la legge tenga conto separatamente dei tre periodi.

Inverno:

Viene dunque scelta come legge di regressione:

escursione termica ~ temperatura massima + temperatura massima : mese^[1]

```
> mod = lm( eta ~ tamax + tamax:me )
```

Call:

```
lm(formula = eta ~ tamax + tamax:me)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.3649	-0.8048	0.0113	0.9284	2.9284

Coefficients:

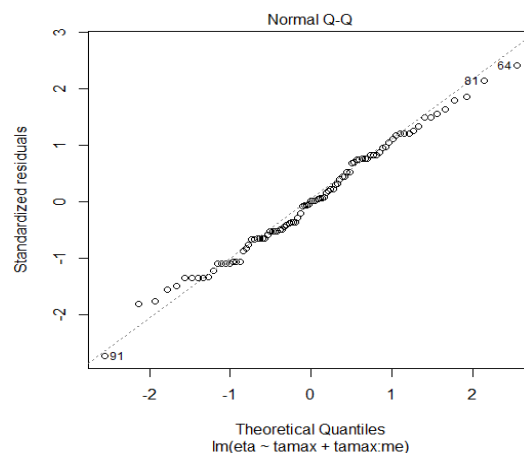
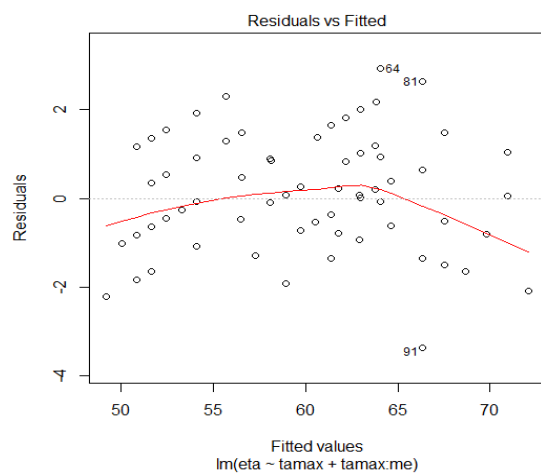
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	74.391416	0.542556	137.113	< 2e-16 ***
tamax	0.812347	0.022143	36.686	< 2e-16 ***
tamax:me5	0.002133	0.020027	0.107	0.915
tamax:me6	0.334300	0.057620	5.802	9.9e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.251 on 89 degrees of freedom

Multiple R-squared: 0.9612, Adjusted R-squared: 0.9599

F-statistic: 735.6 on 3 and 89 DF, p-value: < 2.2e-16



Shapiro-Wilk normality test data: mod\$residuals W = 0.99028, p-value = 0.7322

Sebbene il modello spieghi molto bene la variabilità dei dati (**Adjusted R-squared: 0.9439**), e sia globalmente molto significativo (**p-value: 2.2e-16**), esso non rispetta a pieno le ipotesi del modello lineare gaussiano a causa della eteroschedasticità dei residui. Il pattern ad U rovesciata dello scatterplot suggerisce la mancanza di un termine quadratico.

Poiché i residui non sono i.i.d., non occorre proseguire nell'analisi del modello.

Sulla base di quanto osservato, se ne ripropone un altro che presenti un termine quadratico della temperatura, e inoltre trascuri il regressore *mese* che prima era risultato poco significativo:

escursione termica ~ *temperatura massima* + (*temperatura massima*)²

Call:

lm(formula = eta ~ tamax + tamax2)

Residuals:

Min	1Q	Median	3Q	Max
-3.8027	-0.8272	-0.1241	1.0638	3.2251

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70.052406	0.618880	113.192	< 2e-16 ***
tamax	0.397009	0.085658	4.635	1.2e-05 ***
tamax2	-0.009605	0.002501	-3.840	0.000229 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

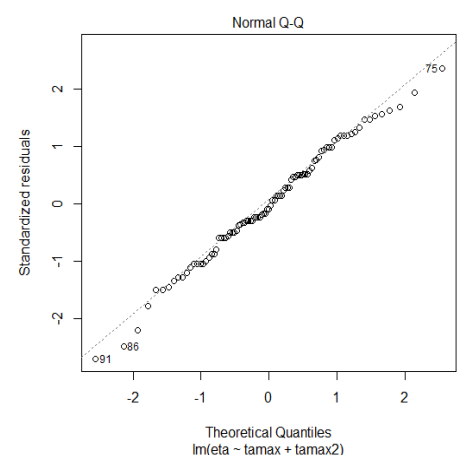
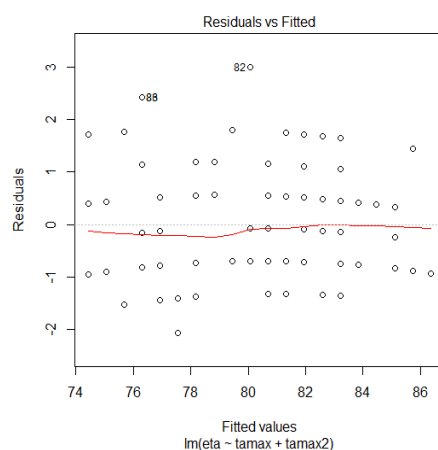
Residual standard error: 1.432 on 90 degrees of freedom

Multiple R-squared: 0.9486, Adjusted R-squared: 0.9475

F-statistic: 831.3 on 2 and 90 DF, p-value: < 2.2e-16

**Shapiro-Wilk
normality test**

data: mod\$residuals
W = 0.99013, p-value
= 0.7212



Questo nuovo modello rispetta le ipotesi del modello lineare gaussiano, infatti i residui sono omoschedastici, disposti a nuvola attorno all'asse orizzontale, e il 95% di essi è compreso tra +2 e -2. I residui presentano distribuzione approssimativamente gaussiana, come mostra sia il qq-plot, sia lo **Shapiro-Wilk** test, che ha **p-value = 0.7212** nonostante l'alta numerosità del campione.

Infine, dall'output di R si osserva che il modello spiega bene la variabilità dei dati (**Adjusted R-squared: 0.9475**) e sia i regressori che il modello sono significativi.

Predizioni:

Il modello risulta soddisfacente. Esso viene dunque testato registrando ulteriori dati dalla stessa fonte del dataset originario.

Si consideri ad esempio il giorno 12 del mese 6 (**ls= 156°**), anno terrestre 2012, ossia circa tre anni marziani prima di quanto considerato nello studio.

Applicando il modello per calcolare un intervallo di predizione al 95% si ottiene:

```
fit    lwr    upr
59.79424 56.91303 62.67546
```

L'escursione termica sperimentale, pari a 58°C, rientra effettivamente nell'intervallo predetto.





Si consideri ora lo stesso giorno ma a un anno marziano di distanza dal precedente:

$$IP(0.95) ; l_s = 155^\circ$$

fit	lwr	upr
63.26806	60.39159	66.14452

Con un'escursione termica di 64°C, la predizione è andata ancora a buon fine.

Per evitare che si possa pensare che il modello funzioni solo su un particolare giorno, si testa ora il giorno 6 del mese 5 dello stesso anno

fit	lwr	upr
54.99174	52.11769	57.86579

$$\text{escursione termica} = 55$$

Il modello prevede effettivamente le escursioni termiche dei mesi invernali. È stata trovata una legge che le determina.



Estate:

Dopo aver trovato un modello di regressione che descrive l'andamento dell'escursione termica nella stagione invernale, si procede verificando se lo stesso andamento sia presente nella stagione estiva.

Si ripropone dunque, per il nuovo record, il modello:

$$\text{escursione termica} \sim \text{temperatura massima} + (\text{temperatura massima})^2$$

Call:

```
lm(formula = eta ~ tamax + tamax2)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.1711	-1.2118	-0.2449	0.8809	4.8938

Coefficients:

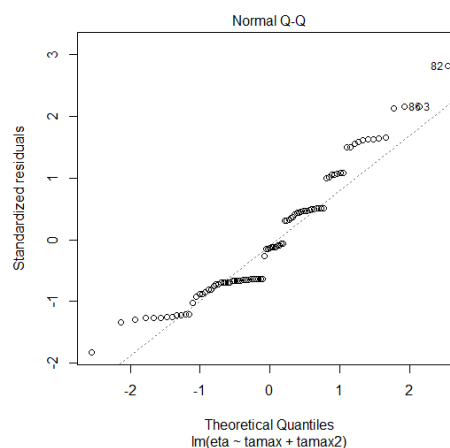
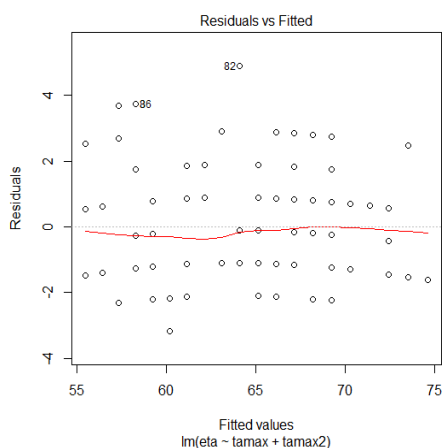
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	72.445336	0.494737	146.432	< 2e-16 ***
tamax	1.081486	0.131437	8.228	1.38e-12 ***
tamax2	0.004886	0.007432	0.657	0.513

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.76 on 90 degrees of freedom

Multiple R-squared: 0.8907, Adjusted R-squared: 0.8883

F-statistic: 366.8 on 2 and 90 DF, p-value: < 2.2e-16



**Shapiro-Wilk
normality test**

**data:
mod\$residuals
W = 0.93573, p-
value = 0.0001898**

Si osserva, sia dal qq-plot, sia dal test di Shapiro-Wilk, che i residui del modello non sono gaussiani, pertanto esso non si può definire valido per discutere l'andamento dell'escursione termica d'estate.

Non è possibile dunque affermare che il parametro studiato abbia lo stesso andamento per i sei mesi studiati.

Pressione annua:

Sebbene si discosti dall'obiettivo primario del lavoro, vista l'evidenza emersa nei dati è interessante studiare l'andamento della pressione nel tempo.

Al fine di poter effettuare una regressione più precisa, si è ridotto l'arco di tempo in esame all'intervallo di longitudine solare compreso tra 35° e 180°.

Il primo approccio è stato di regressione lineare semplice. Stimato il valore della pulsazione e della fase della sinusoide a partire dal plot Pressione ~ Longitudine, si ottiene:

$$pressione \sim \sin(\omega * t + \varphi)^{[2]}$$

Call:

lm(formula = p ~ reg)

Residuals:

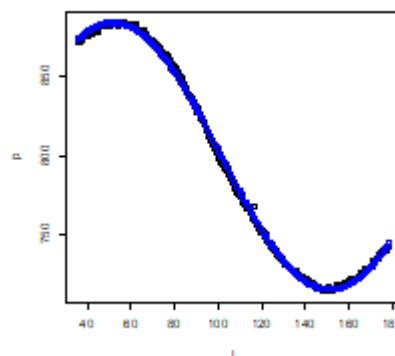
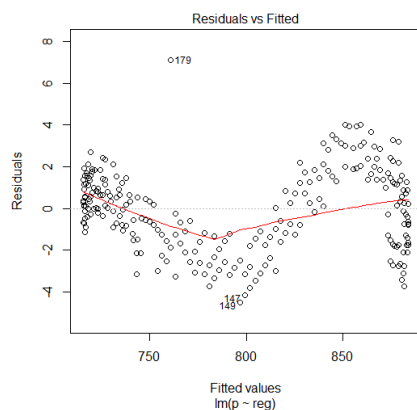
Min	1Q	Median	3Q	Max
-4.5006	-1.4058	0.1319	1.3754	7.1150

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	799.7003	0.1086	7363	<2e-16 ***
reg	84.0952	0.1411	596	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.875 on 296 degrees of freedom
Multiple R-squared: 0.9992, **Adjusted R-squared:** 0.9992
F-statistic: 3.552e+05 on 1 and 296 DF, **p-value:** < 2.2e-16



Come si può osservare sia dall' output di R che dalla sovrapposizione tra la curva reale e quella teorica fornita dal programma, il modello spiega estremamente bene come variano i dati (**Multiple R-squared: 0.9992**).

Tuttavia, i residui presentano un chiaro pattern eteroschedastico che non rispetta le ipotesi del modello lineare gaussiano.

Osservando un trend sinusoidale si è scelto di eliminare la fase stimata empiricamente ed utilizzare diverse sinusoidi con pulsazioni crescenti.

Procedendo a successive iterazioni, al fine di migliorare lo scatterplot ed ignorando i termini non significativi, si è giunti al seguente modello:

Call:

lm(formula = p ~ sin(w * t) + cos(w * t) + sin(2 * w * t) + cos(2 * w * t) + sin(4 * w * t) + cos(4 * w * t) + cos(6 * w * t) + cos(8 * w * t))

Residuals:

Min	1Q	Median	3Q	Max
-2.9721	-0.6745	-0.0358	0.6530	8.2859

Coefficients:

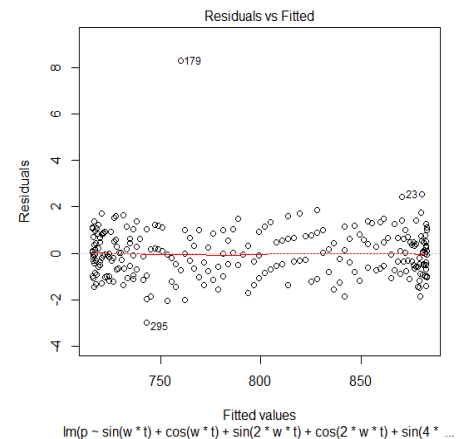
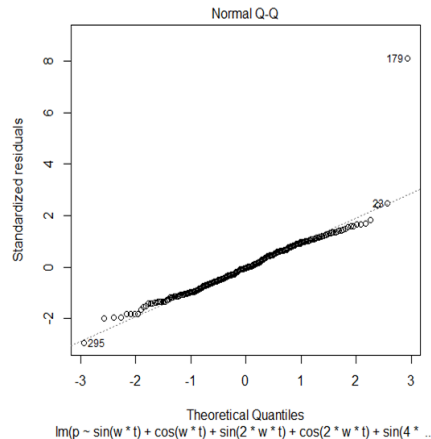
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	798.27853	0.10729	7440.261	< 2e-16 ***
sin(w * t)	83.03528	0.09193	903.288	< 2e-16 ***
cos(w * t)	-8.33696	0.17968	-46.399	< 2e-16 ***
sin(2 * w * t)	-2.11704	0.10611	-19.951	< 2e-16 ***
cos(2 * w * t)	-1.97759	0.13231	-14.947	< 2e-16 ***
sin(4 * w * t)	-0.52069	0.08842	-5.889	1.08e-08 ***
cos(4 * w * t)	-0.98341	0.08963	-10.971	< 2e-16 ***
cos(8 * w * t)	-0.38266	0.08709	-4.394	1.56e-05 ***
cos(6 * w * t)	-0.22327	0.08983	-2.485	0.0135 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.043 on 289 degrees of freedom
Multiple R-squared: 0.9997, Adjusted R-squared: 0.9997
F-statistic: 1.436e+05 on 8 and 289 DF, p-value: < 2.2e-16

Shapiro-Wilk normality test

data: mod\$residuals
W = 0.9037, p-value = 7.431e-13



Il modello multiplo spiega estremamente bene la variabilità della pressione nel tempo, ma questa volta i residui sono ragionevolmente omoschedastici. Inoltre, sono anche reputabili gaussiani poiché il qq-plot aderisce bene alla retta dei quantili, almeno per quanto riguarda la porzione centrale. Il risultato negativo dello Shapiro test può essere trascurato a causa della grande quantità di dati, con la quale, notoriamente, tale test pecca di eccessiva sensibilità.

Si effettua ora una verifica della legge tramite un intervallo di predizione di livello 0.99:

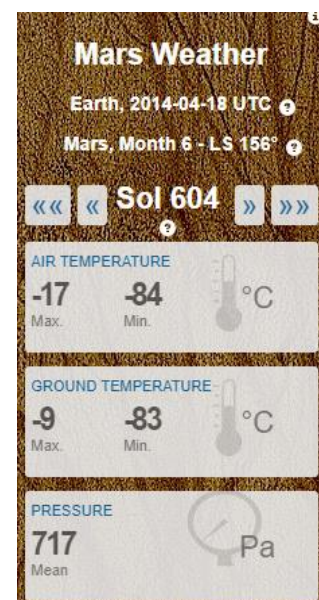
```
fit   lwr   upr
717.0557 714.3223 719.7891
```

L'intervallo ottenuto predice bene il valore. Anche questo modello risulta soddisfacente.

note:

1) Sono state definite sul software R le variabili
 eta := escursione termica dell'aria
 tamax := temperatura massima dell'aria
 tamax2 := tamax * tamax
 me := mese (regressore categorico di livelli 4-5-6 per l'inverno e 10-11-12 per l'estate)

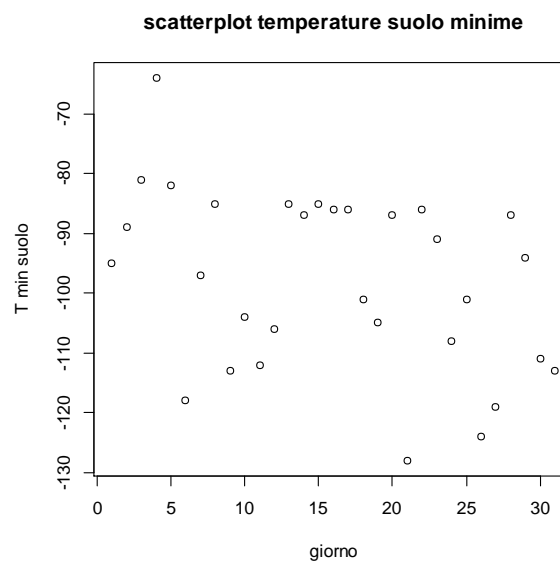
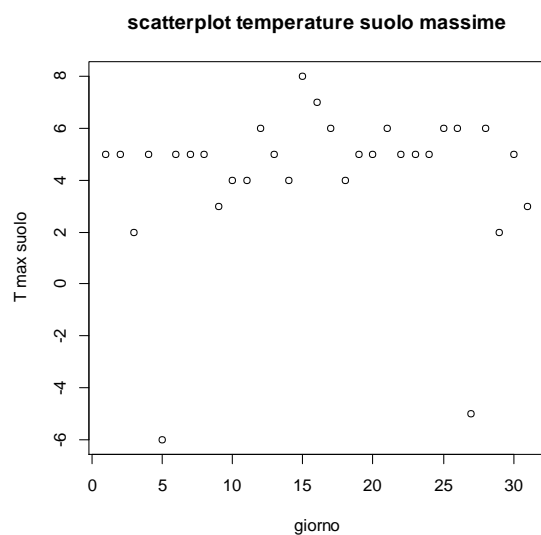
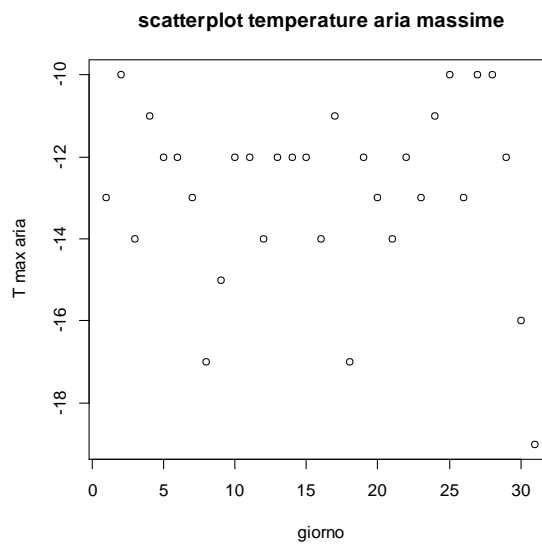
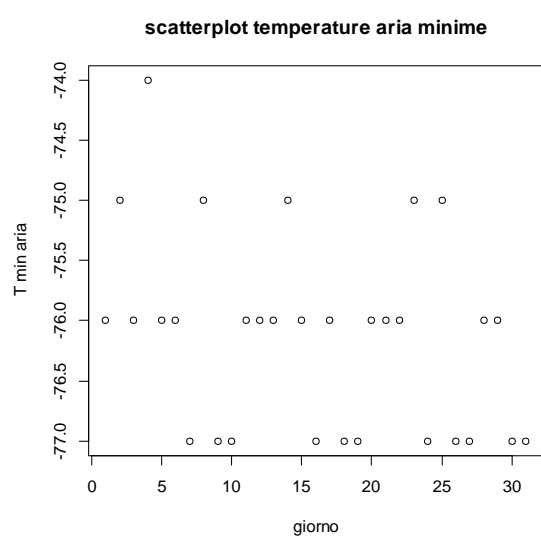
2) Sono state definite sul software R le variabili
 t=ls[which(ls>35 & ls<180)]
 p=press[which(ls>35 & ls<180)]
 ω := 0.03155
 φ := -0.07



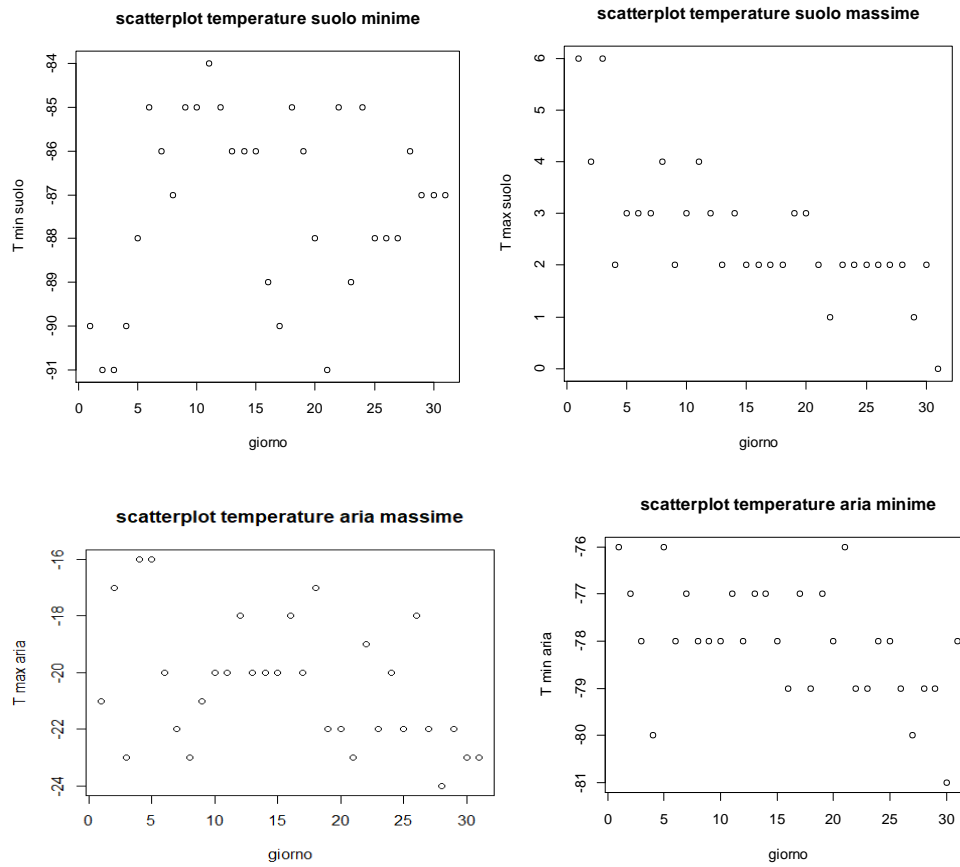
Risorse utilizzate

Scatterplot dati mensili, in alcuni campioni sono stati successivamente rimossi outliers.

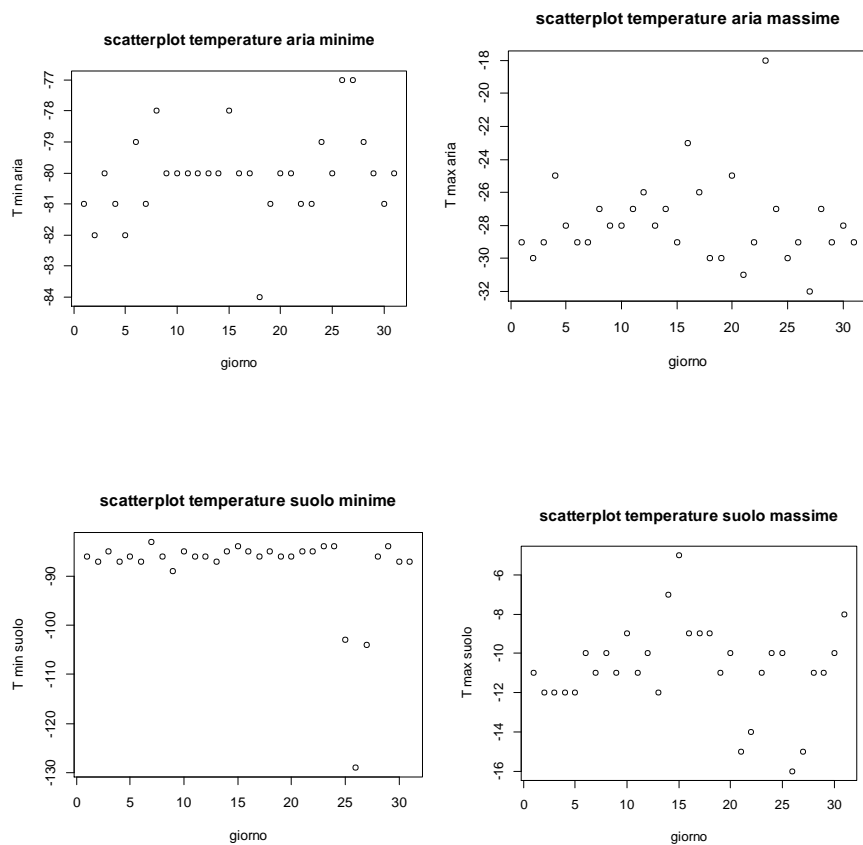
Mese 1



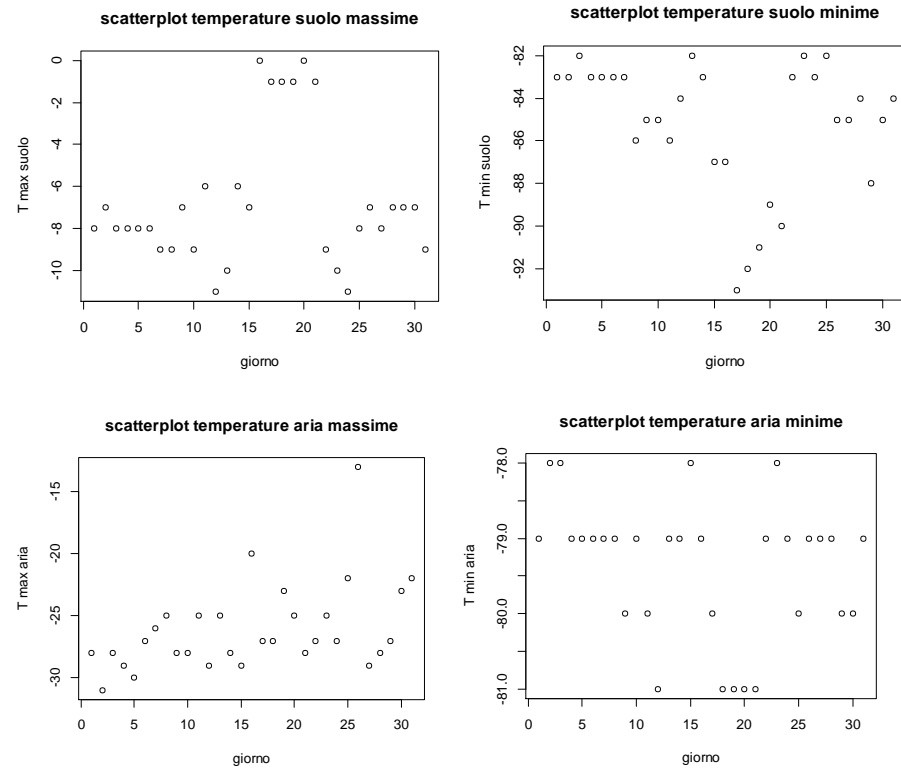
Mese 2



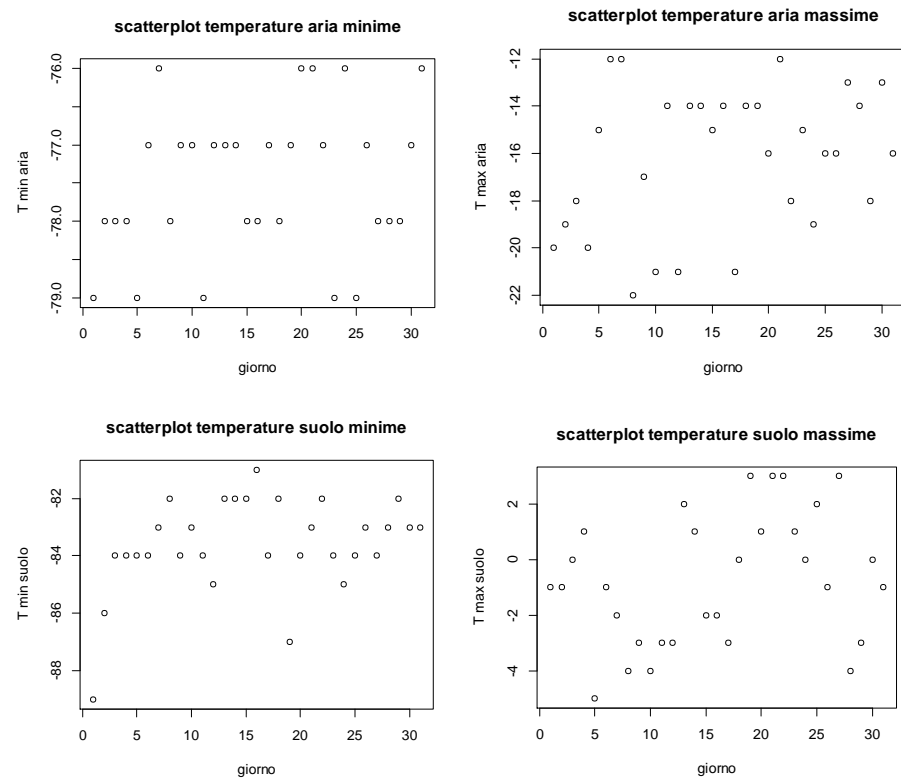
Mese 3



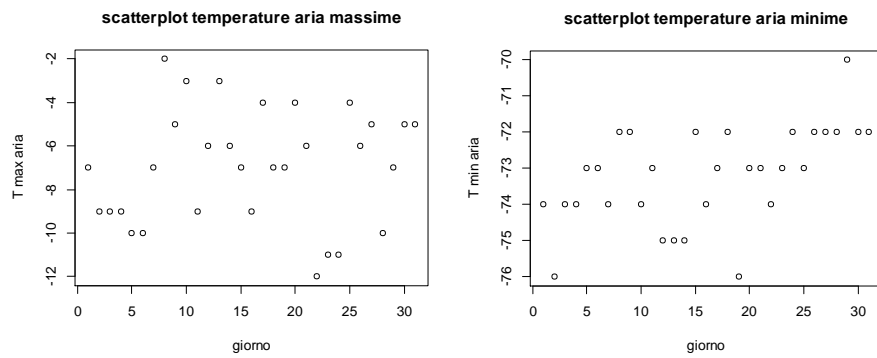
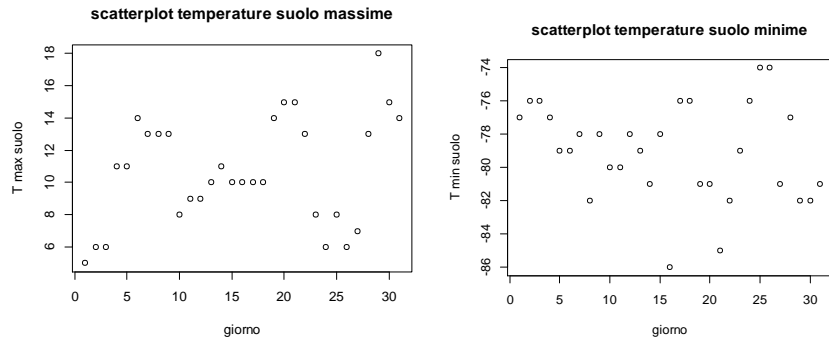
Mese 4



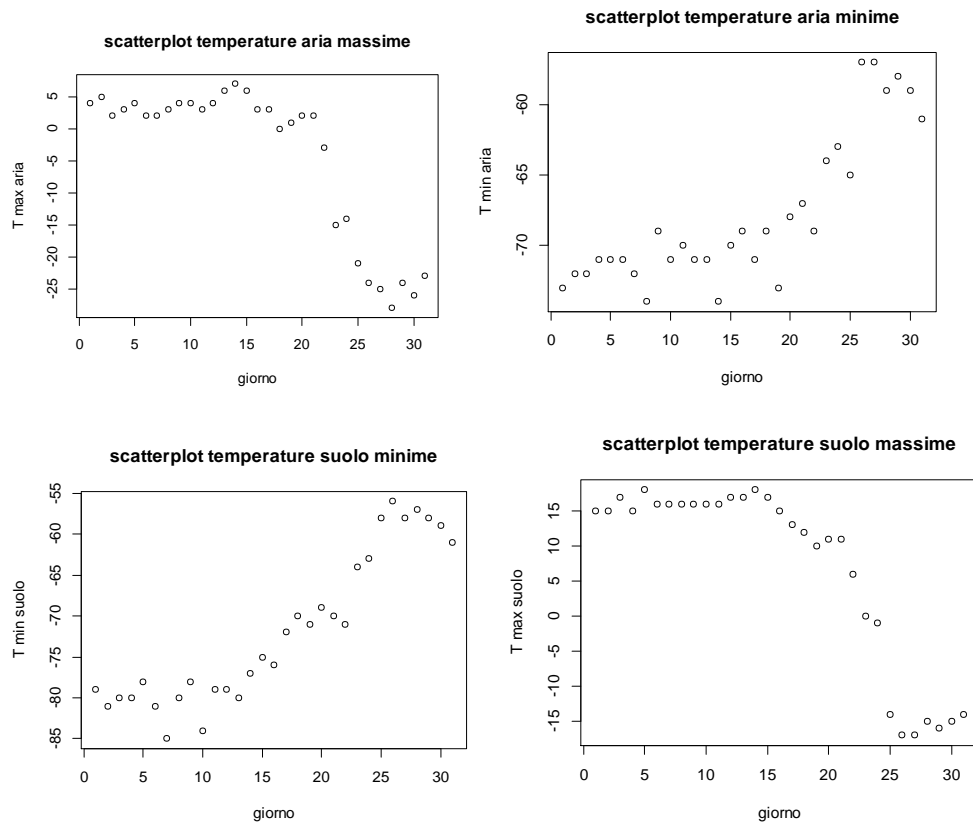
Mese 5



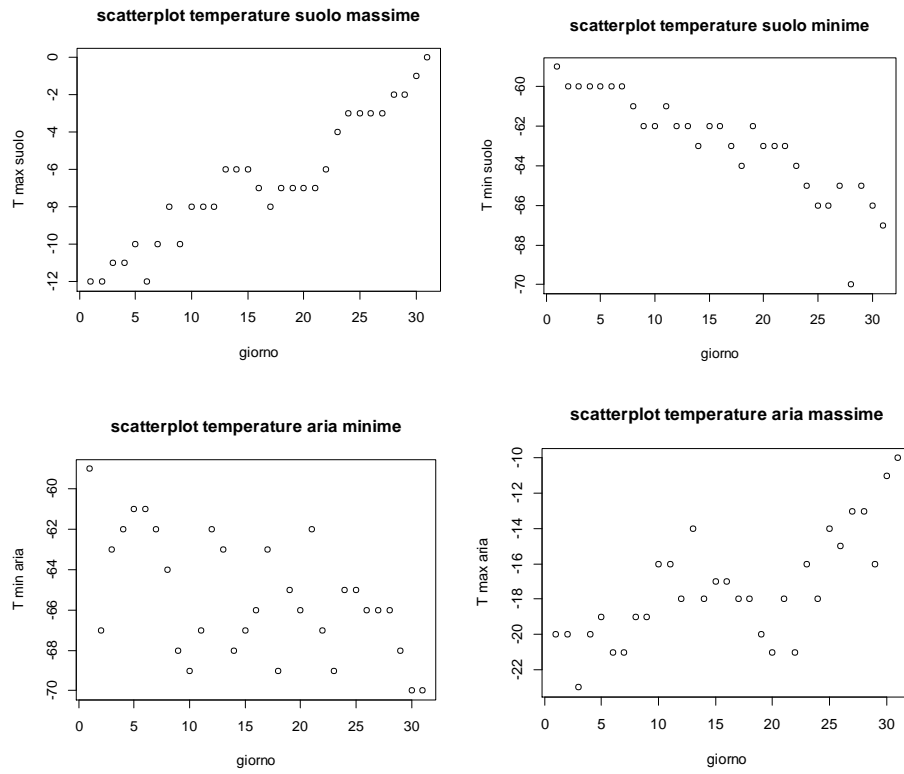
Mese 6



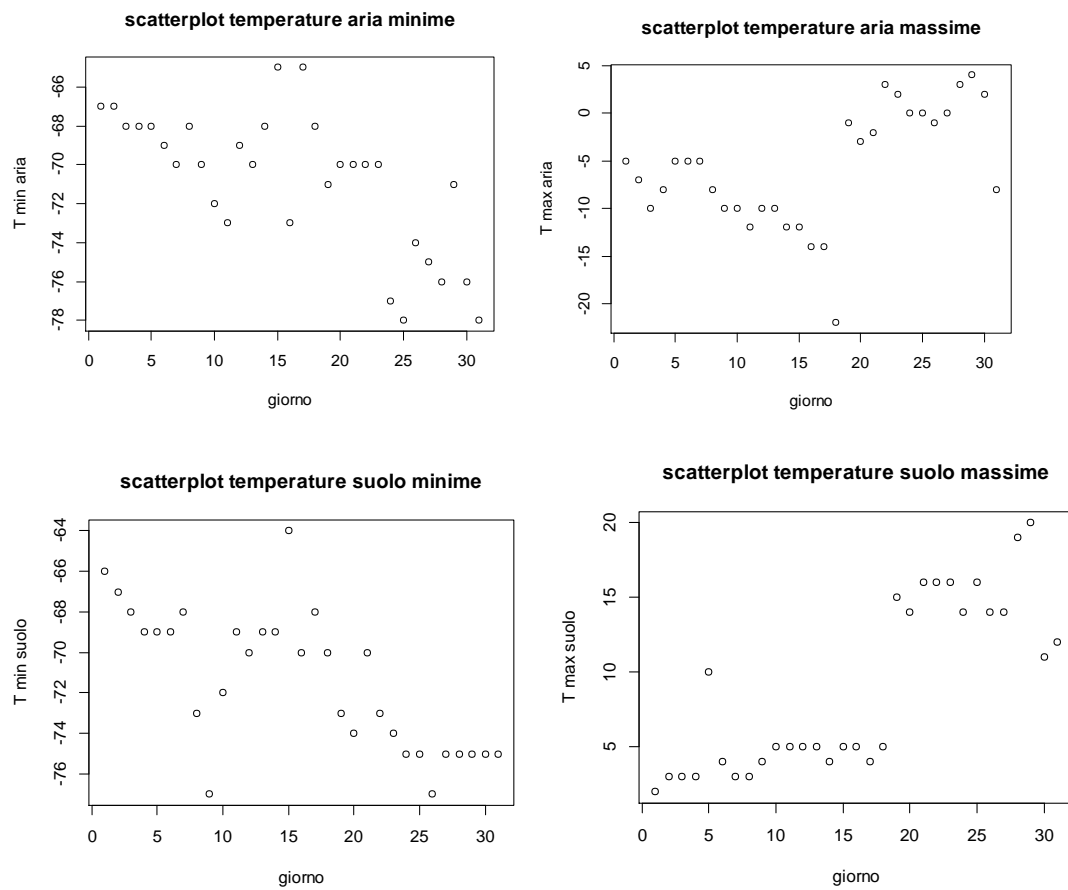
Mese 7



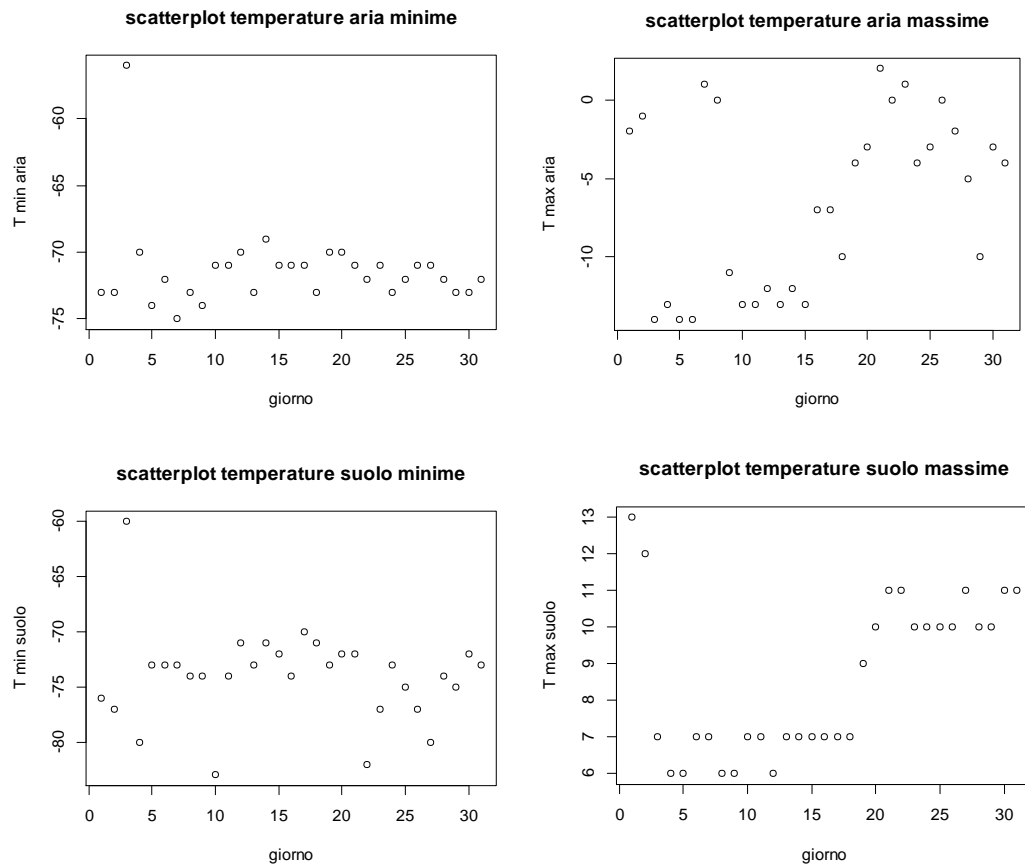
Mese 8



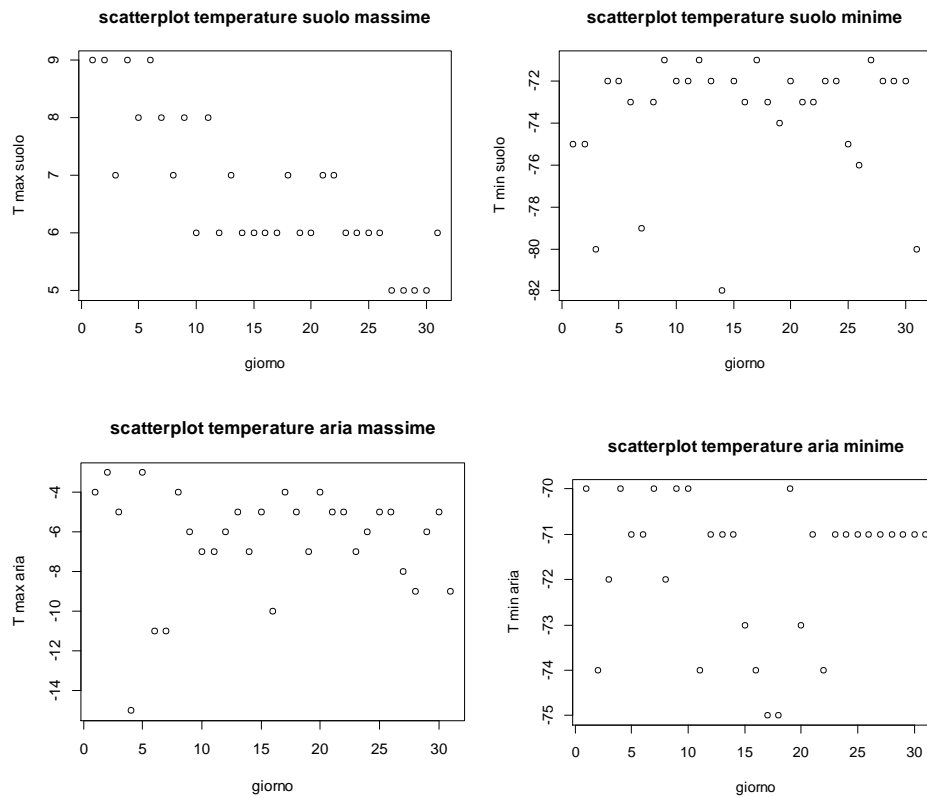
Mese 9



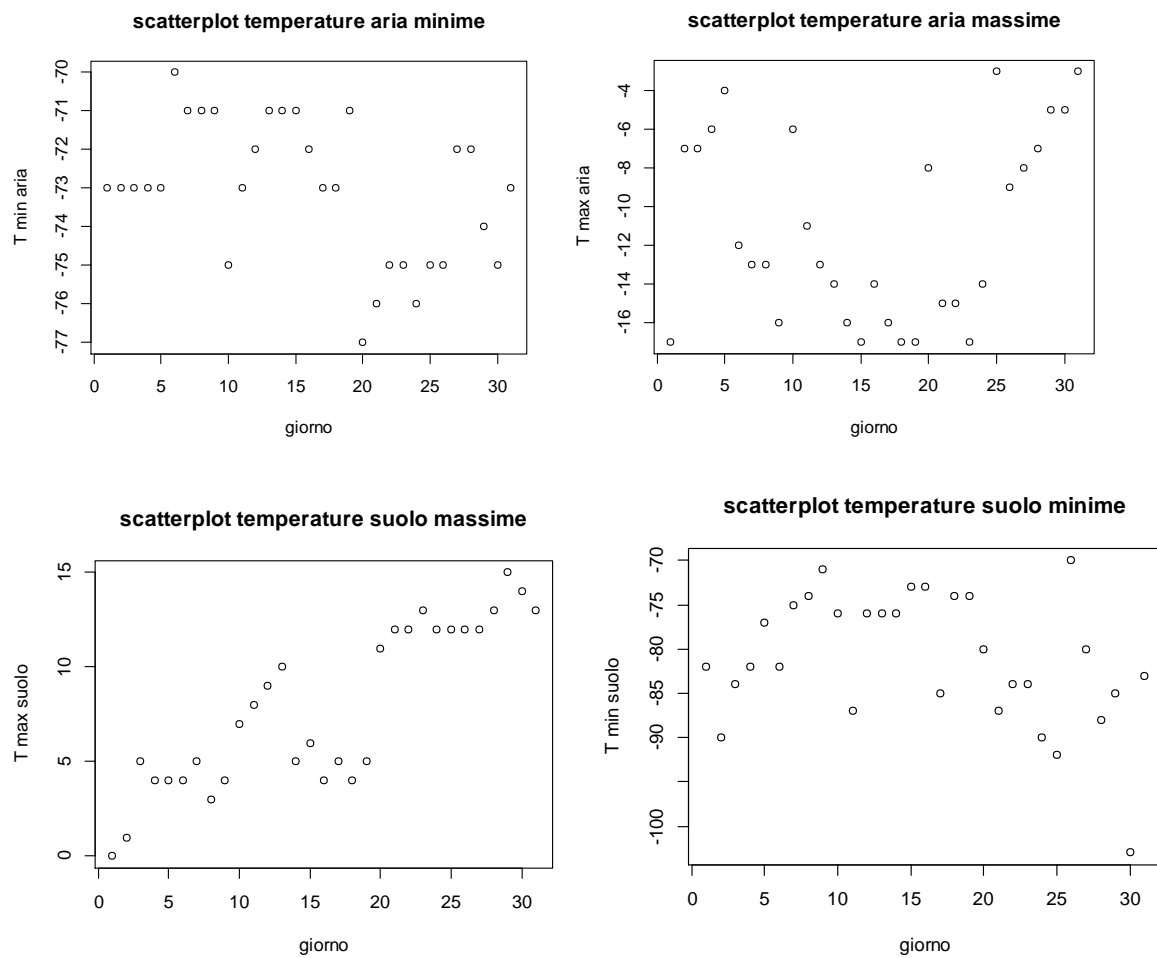
Mese 10



Mese 11



Mese 12



Sitografia

<http://cab.inta-csic.es/remis/en#previous-sol>

[https://it.wikipedia.org/wiki/Marte_\(astronomia\)](https://it.wikipedia.org/wiki/Marte_(astronomia))

https://it.wikipedia.org/wiki/Atmosfera_di_Marte

Bibliografia

, *The Darian Calendar*, Thomas Gangale, 1999

, *Atmospheric Imaging Results from the Mars Exploration Rovers: Spirit and Opportunity*, M.T. Lemmon