

```
1 -----
2 Laboratorio 01
3 -----
4 > Ciliegi<-read.table("Cherry.dat",col.names=c("diam","alt","vol")
5 > Ciliegi
6   diam alt  vol
7 1  8.3  70 10.3
8 2  8.6  65 10.3
9 3  8.8  63 10.2
10 4 10.5  72 16.4
11 5 10.7  81 18.8
12 6 10.8  83 19.7
13 7 11.0  66 15.6
14 8 11.0  75 18.2
15 9 11.1  80 22.6
16 10 11.2  75 19.9
17 11 11.3  79 24.2
18 12 11.4  76 21.0
19 13 11.4  76 21.4
20 14 11.7  69 21.3
21 15 12.0  75 19.1
22 16 12.9  74 22.2
23 17 12.9  85 33.8
24 18 13.3  86 27.4
25 19 13.7  71 25.7
26 20 13.8  64 24.9
27 21 14.0  78 34.5
28 22 14.2  80 31.7
29 23 14.5  74 36.3
30 24 16.0  72 38.3
31 25 16.3  77 42.6
32 26 17.3  81 55.4
33 27 17.5  82 55.7
34 28 17.9  80 58.3
35 29 18.0  80 51.5
36 30 18.0  80 51.0
37 31 20.6  87 77.0
38 > attach(Ciliegi)
39 > vol
40 [1] 10.3 10.3 10.2 16.4 18.8 19.7 15.6 18.2 22.6 19.9 24.2 21.0 21.4 21.3 19.1
41 [16] 22.2 33.8 27.4 25.7 24.9 34.5 31.7 36.3 38.3 42.6 55.4 55.7 58.3 51.5 51.0
42 [31] 77.0
43 > dim(Ciliegi)
44 [1] 31 3
45 > dim(Ciliegi)[1]
46 [1] 31
47 > Ciliegi[7,1]
48 [1] 11
49 > Ciliegi[7,]
50   diam alt  vol
51 7  11  66 15.6
52 > Ciliegi[,2]
53 [1] 70 65 63 72 81 83 66 75 80 75 79 76 76 69 75 74 85 86 71 64 78 80 74 72 77
54 [26] 81 82 80 80 80 87
55 > Ciliegi$alt
56 [1] 70 65 63 72 81 83 66 75 80 75 79 76 76 69 75 74 85 86 71 64 78 80 74 72 77
57 [26] 81 82 80 80 80 87
58 > summary(alt)
59   Min. 1st Qu.  Median      Mean 3rd Qu.     Max.
60    63       72       76       76      80      87
61 > var(alt)
62 [1] 40.6
63 > hist(alt)
64 > help(hist)
65 > boxplot(alt)
66 > # Analisi grafica
67 > #
68 > plot(diam,vol)
69 > cor(diam,vol)
70 [1] 0.9671194
71 > # Y = alpha + beta X + epsilon, epsilon : E(epsilon)=0, Var(epsilon)=sigma^2
72 >
73 > # Stima del modello
74 >
75 > # --> Metodo dei minimi quadrati
76 > # beta=cov(X,Y)/Var(X)
77 > # alpha=mean(Y)-beta*mean(X)
78 >
```

```
79 > beta<-(sum(diam*vol)/length(diam)-mean(diam)*mean(vol))/(mean(diam^2)-mean(diam)^2)
80 > n<-length(diam)
81 > n
82 [1] 31
83 > beta<-(sum(diam*vol)/n-mean(diam)*mean(vol))/(mean(diam^2)-mean(diam)^2)
84 > beta
85 [1] 5.065856
86 > beta<-cov(diam,vol)/((n-1)*var(diam)/n)
87 > beta
88 [1] 5.065856
89 > alpha<-mean(vol)-beta*mean(diam)
90 > alpha
91 [1] -36.94346
92 > abline(alpha,beta)
93 > abline(h=40)
94 > abline(v=14,lty="dashed")
95 >
96 >
97 > # Step 3: Analisi dei residui
98 > # res = y - y.stimati
99 >
100 > y.stimati<-alpha+beta*diam
101 > y.stimati
102 [1] 5.103149  6.622906  7.636077 16.248033 17.261205 17.767790 18.780962 18.780962
103 [9] 19.287547 19.794133 20.300718 20.807304 20.807304 22.327061 23.846818 28.406089
104 [17] 28.406089 30.432431 32.458774 32.965360 33.978531 34.991702 36.511459 44.110244
105 [25] 45.630001 50.695857 51.709028 53.735371 54.241956 54.241956 67.413183
106 > points(diam,y.stimati,pch="X")
107 > residui<-vol-y.stimati
108 > residui
109 [1] 5.1968508 3.6770939 2.5639226 0.1519667 1.5387954 1.9322098 -3.1809615
110 [8] -0.5809615 3.3124528 0.1058672 3.8992815 0.1926959 0.5926959 -1.0270610
111 [15] -4.7468179 -6.2060887 5.3939113 -3.0324313 -6.7587739 -8.0653595 0.5214692
112 [22] -3.2917021 -0.2114590 -5.8102436 -3.0300006 4.7041430 3.9909717 4.5646292
113 [29] -2.7419565 -3.2419565 9.5868168
114 > plot(residui)
115 > abline(h=0)
116 > summary(residui)
117   Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
118 -8.065e+00 -3.107e+00  1.520e-01  3.209e-15  3.495e+00  9.587e+00
119 > boxplot(residui)
120 > hist(residui)
121 >
122 > # R2 = 1 - Var(residui)/Var(Y)
123 >
124 > r2<-1-var(residui)/var(vol)
125 > r2
126 [1] 0.9353199
127 >
128 > # ipotesi di normalita' y ~ N(alpha + beta*X,sigma^2)
129 > # epsilon ~ N(0,sigma^2)
130 >
131 > alpha
132 [1] -36.94346
133 > beta
134 [1] 5.065856
135 > sigma2<-sum(residui^2)/n
136 > sigma2
137 [1] 16.91299
138 > s2<-n*sigma2/(n-2)
139 > s2
140 [1] 18.07940
141 > # NB: s2 e' la varianza corretta
142 >
143 > # (alpha-alpha)/sqrt(var(alpha)) ~ t_(n-2)
144 >
145 > var.alpha<-s2*(1/n+mean(diam)^2/sum((diam-mean(diam))^2))
146 > var.alpha
147 [1] 11.3242
148 > var.beta<-s2/(sum((diam-mean(diam))^2))
149 > var.beta
150 [1] 0.06119536
151 > # alpha +- t_(n-2) * sqrt(Var(alpha))
152 > al<-alpha-qt(.975,n-2)*sqrt(var.alpha)
153 > al
154 [1] -43.82595
155 > a2<-alpha+qt(.975,n-2)*sqrt(var.alpha)
156 > a2
```

```
157 [1] -30.06096
158 > # beta +- t_(n-2) * sqrt(Var(beta))
159 > bl<-beta-qt(.975,n-2)*sqrt(var.beta)
160 > bl
161 [1] 4.559914
162 > b2<-beta+qt(.975,n-2)*sqrt(var.beta)
163 > b2
164 [1] 5.571799
165 >
166 > # test
167 > # H0 : beta = 0 , H1 : beta != 0
168 > # t = (beta-0)/sqrt(var(beta)) ~H0 t_(n-2)
169 >
170 > test.t<-(beta-0)/sqrt(var.beta)
171 > test.t
172 [1] 20.47829
173 > qt(.975,n-2)
174 [1] 2.045230
175 > # a_oss = 2*min(P(T<test.t),P(T>test.t))
176 > 2*min(pt(test.t,n-2),1-pt(test.t,n-2))
177 [1] 0
178 > # rifiuto H0
179 >
180 > # test
181 > # H0 : alpha = 0 , H1 : alpha != 0
182 > # t = (alpha-0)/sqrt(var(alpha)) ~H0 t_(n-2)
183 >
184 > test.t<-(alpha-0)/sqrt(var.alpha)
185 > test.t
186 [1] -10.97827
187 > 2*min(pt(test.t,n-2),1-pt(test.t,n-2))
188 [1] 7.621449e-12
189 > # rifiuto H0
190 >
191 > # Rifai l'esercizio con la trasformazione logaritmica dei dati
192 > # ln(vol)=alpha+beta*ln(diam)
193 > # ln(e^alpha * diam^beta)
194 > # vol=e^alpha * diam^beta
195 >
196 > x <- log(diam)
197 > y <- log(vol)
198 > plot(x,y)
199 > beta<-cov(x,y)/var(x)
200 > alpha<-mean(y)-beta*mean(x)
201 > alpha
202 [1] -2.353325
203 > beta
204 [1] 2.19997
205 >
206 > plot(diam,vol)
207 > abline(alpha,beta)
208 > val.predetti<-exp(alpha)*diam^beta
209 > lines(diam,val.predetti)
210 > res<-vol-val.predetti
211 > var(res)
212 [1] 10.56006
213 > var(residui)
214 [1] 0.01277479
215 >
216 > detach(Ciliegi)
217 -----
218 Laboratorio 02
219 -----
220 > brainbod<-read.table('brainbod.dat',header=T)
221 > brainbod
222   species bodywt brainwt
223 1 afeleph 6654.00 5712.00
224 2 cow 465.00 423.00
225 3 donkey 187.00 419.00
226 4 man 62.00 1320.00
227 5 graywolf 36.33 119.50
228 6 redfox 4.24 50.40
229 7 narmadillo 3.50 10.80
230 8 echidna 3.00 25.00
231 9 phalanger 1.62 11.40
232 10 guineapig 1.04 5.50
233 11 euredhog 0.79 3.50
234 12 chinchilla 0.43 4.00
```

```
235 13 ghamster 0.12 1.00
236 14 amole 0.06 1.00
237 15 lbbat 0.01 0.25
238 >
239 > attach(brainbod)
240 >
241 > # Step1: Analisi preliminare dei dati
242 >
243 > summary(brainwt)
244   Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
245    0.25     3.75     11.40    540.40    269.30    5712.00
246 > summary(bodywt)
247   Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
248    0.01     0.61     3.00    494.60    49.17    6654.00
249 > par(mfrow=c(2,2))
250 > hist(bodywt)
251 > boxplot(bodywt)
252 > hist(brainwt)
253 > boxplot(brainwt)
254 >
255 >
256 > cor(bodywt,brainwt)
257 [1] 0.975509
258 > plot(bodywt,brainwt)
259 > identify(bodywt,brainwt,species)
260 [1] 1 2 3 4 5 6 7 8 9 12 15
261 >
262 > x<-log(bodywt)
263 > y<-log(brainwt)
264 >
265 > summary(x)
266   Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
267 -4.6050 -0.5398  1.0990  1.4400  3.8600  8.8030
268 > summary(y)
269   Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
270 -1.386  -1.320  2.434  3.174  5.411  8.650
271 > hist(x)
272 > hist(y)
273 > boxplot(x)
274 > boxplot(y)
275 > cor(x,y)
276 [1] 0.9732277
277 > # y = alpha + beta x + epsilon?????????
278 >
279 > # Stima dei parametri
280 >
281 > n<-length(x)
282 > beta<-cov(x,y)/((n-1)*var(x)/n)
283 > beta
284 [1] 0.838046
285 > alpha<-mean(y)-beta*mean(x)
286 > alpha
287 [1] 1.96775
288 > plot(x,y)
289 > abline(alpha,beta)
290 >
291 > # Analisi dei residui
292 > val.pred<-alpha+beta*x
293 > res<-y-val.pred
294 > summary(res)
295   Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
296 -1.068e+00 -4.156e-01 -1.909e-01  7.253e-16  3.602e-01  1.759e+00
297 > boxplot(res)
298 > max(res)
299 [1] 1.758908
300 > brainbod[4,]
301   species bodywt brainwt
302 4 man 62 1320
303 > plot(x,y)
304 > r2<-1-var(res)/var(y)
305 > r2
306 [1] 0.9423396
307 >
308 > qqnorm(res)
309 > qqline(res)
310 >
311 > # intervallo di confidenza
312 > sigma2<-sum(res^2)/n
```

```

313 > sigma2
314 [1] 0.4492044
315 > s2<-sum(res^2)/(n-2)
316 > s2
317 [1] 0.5183127
318 > var.alpha<-s2*(1/(n+mean(x)^2/sum((x-mean(x))^2))
319 > var.beta<-s2/sum((x-mean(x))^2)
320 > var.alpha
321 [1] 0.07924495
322 > var.beta
323 [1] 0.004435418
324 > alpha.lower<-alpha-qt(0.975,n-2)*sqrt(var.alpha)
325 > alpha.upper<-alpha+qt(0.975,n-2)*sqrt(var.alpha)
326 > alpha.lower
327 [1] 1.359596
328 > alpha.upper
329 [1] 2.575904
330 > beta.lower<-beta-qt(0.975,n-2)*sqrt(var.beta)
331 > beta.upper<-beta+qt(0.975,n-2)*sqrt(var.beta)
332 > beta.lower
333 [1] 0.6941678
334 > beta.upper
335 [1] 0.9819243
336 > # verifica di ipotesi: (beta=0)/sqrt(var.beta)
337 > beta/sqrt(var.beta)
338 [1] 15.65707
339 > qt(.975,n,2)
340 [1] 2.160369
341 > # 15 > 2 --> accetto l'ipotesi nulla
342 >
343 >
344 > #####
345 > # UOMO
346 > # elimino il dato scomodo
347 > xl<-x[-4]
348 > xl
349 [1] 8.80297346 6.14203741 5.23110862 3.59264385 1.44456327 1.25276297 1.09861229
350 [8] 0.48242615 0.03922071 -0.23572233 -0.84397007 -2.12026354 -2.81341072 -4.60517019
351 > y1<-y[-4]
352 > y1
353 [1] 8.650325 6.047372 6.037871 4.783316 3.919991 2.379546 3.218876 2.433613
354 [9] 1.704748 1.252763 1.386294 0.000000 0.000000 -1.386294
355 >
356 > cor(x,y)
357 [1] 0.9732277
358 >
359 > cor(x1,y1)
360 [1] 0.989434
361 >
362 > n1<-length(x1)
363 > betal<-cov(x1,y1)/((n1-1)*var(x1)/14)
364 > betal
365 [1] 0.80816
366 > alpha1<-mean(y1)-betal*mean(x1)
367 > alpha1
368 [1] 1.879402
369 >
370 > plot(x,y)
371 > abline(alpha,beta)
372 > abline(alpha1,betal,lty='dashed')
373 >
374 > res1<-y1-alpha1-betal*x1
375 > par(mfrow=c(1,2))
376 > boxplot(res1)
377 > qqnorm(res1)
378 > qqline(res1)
379 >
380 -----
381 Laboratorio 03
382 -----
383 # Y = alpha + beta X + epsilon , epsilon ~ N(0,sigma^2)
384 > # costruiamo il nostro campione
385 >
386 > nc=30
387 > xs=1:30
388 > x
389 [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
390 [29] 29 30

```

```

391 > error<-rnorm(n,0,4) # sigma^2=16
392 > error
393 [1] 0.04406192 -2.78513488 -3.82533130 -7.47414116 -2.55828588 -8.81128405 10.57306139
394 [8] -1.00157483 -3.13954695 6.84054944 1.73122246 -2.20881372 -2.77174140 -5.27932086
395 [15] -1.87458175 1.09956605 -6.85715244 -1.71112577 -0.06596067 4.91131049 4.76435622
396 [22] -1.04205041 -1.11789092 -3.23884442 1.96330506 0.43448044 4.65260502 0.84634824
397 [29] 0.33454760 2.94081540
398 > y<-5+3*x+error # alpha=5, beta=3
399 > y
400 [1] 8.044062 8.214865 10.174669 9.525859 17.441714 14.188716 36.573061 27.998425
401 [8] 28.860453 41.840549 39.731222 38.791186 41.228259 41.720679 48.125418 54.095666
402 [15] 49.142848 57.288874 61.934039 69.913130 63.235644 69.957950 72.882109 73.761156
403 [22] 81.963305 83.434480 90.652605 89.846348 92.334548 97.940815
404 >
405 > # step 1
406 > cor(x,y)
407 [1] 0.989372
408 > plot(x,y)
409 >
410 >
411 > # step 2
412 > beta<-cov(x,y)/var(x)
413 > beta
414 [1] 3.120524
415 > alpha<-mean(y)-beta*mean(x)
416 > alpha
417 [1] 2.326758
418 >
419 > abline(alpha,beta)
420 >
421 > # step 3 analisi modello
422 >
423 > res<- y-alpha-beta*x
424 > boxplot(res)
425 > qqnorm(res)
426 > qqline(res)
427 >
428 >
429 > # aumento sigma^2
430 > error<-rnorm(n,0,10) # sigma^2=100
431 > y<-5+3*x+error # alpha=5, beta=3
432 > cor(x,y)
433 [1] 0.940608
434 > plot(x,y)
435 > beta<-cov(x,y)/var(x)
436 > beta
437 [1] 2.946698
438 > alpha<-mean(y)-beta*mean(x)
439 > alpha
440 [1] 7.626219
441 >
442 >
443 > # diminuisco sigma^2
444 > error<-rnorm(n,0,.25) # sigma^2=0.0625
445 > y<-5+3*x+error # alpha=5, beta=3
446 > cor(x,y)
447 [1] 0.999932
448 > plot(x,y)
449 > beta<-cov(x,y)/var(x)
450 > beta
451 [1] 3.003701
452 > alpha<-mean(y)-beta*mean(x)
453 > alpha
454 [1] 4.899634
455 >
456 >
457 > #-----#
458 > # Studio di simulazione #
459 > #-----#
460 >
461 > nc=30
462 > alpha<-5
463 > beta<-3
464 > sigma2<-16
465 >
466 > xs=1:30
467 >
468 > error<-rnorm(n,0,sqrt(sigma2))

```

```

469 > y<-5+3*x+error
470 >
471 > beta <-cov(x,y)/var(x)
472 > beta
473 [1] 3.017296
474 > error<-rnorm(n,0,sqrt(sigma2))
475 > y<-5+3*x+error
476 > beta <-cov(x,y)/var(x)
477 > beta
478 [1] 2.948699
479 >
480 > # numero tentativi = 1000
481 >
482 > beta.sim<-rep(0,1000)
483 > for (i in 1:1000) {
484 +   error<-rnorm(n,0,sqrt(sigma2))
485 +   y<-5+3*x+error
486 +   beta.sim[i]<-cov(x,y)/var(x)
487 + }
488 >
489 > mean(beta.sim)
490 [1] 2.997770
491 >
492 > # numero tentativi = 10000
493 >
494 > beta.sim<-rep(0,10000)
495 > for (i in 1:10000) {
496 +   error<-rnorm(n,0,sqrt(sigma2))
497 +   y<-5+3*x+error
498 +   beta.sim[i]<-cov(x,y)/var(x)
499 + }
500 > mean(beta.sim) # NB: vicinissima al vero valore del parametro
501 [1] 3.00149
502 >
503 > hist(beta.sim)
504 > boxplot(beta.sim)
505 >
506 > # sigma^2 / sum(x_i - var(x))^2
507 >
508 > var.beta<-sigma2/sum((x-mean(x))^2)
509 > var.beta
510 [1] 0.007119021
511 >
512 > # stima della densita'
513 > hist(beta.sim,prob=T)
514 > lines(density(beta.sim))
515 > lines(seq(2.7,3.3,.01),dnorm(seq(2.7,3.3,.01),3,sqrt(var.beta)))
516 > lines(seq(2.7,3.3,.01),dnorm(seq(2.7,3.3,.01),3,sqrt(var.beta))),col=2)
517 >
518 >
519 > #-----#
520 > # Livello di copertura degli IC per beta #
521 > #-----#
522 >
523 > nc=30
524 > xs=1:30
525 > error<-rnorm(n,0,4)
526 > y<-5+3*x+error
527 >
528 > # beta +- t_(n-2) * sqrt(var(beta))
529 >
530 > beta<-cov(x,y)/var(x)
531 > alpha<-mean(y)-beta*mean(x)
532 > res<-y-alpha-beta*x
533 > s2<-sum(res^2)/(n-2)
534 > beta.inf<-beta-qt(.975,n-2)*sqrt(var.beta)
535 > beta.inf
536 [1] 2.875947
537 > beta.sup<-beta+qt(.975,n-2)*sqrt(var.beta)
538 > beta.sup
539 [1] 3.21878
540 > # il 3 è compreso tra i due estremi
541 >
542 >
543 > #-----#
544 > # studio simulazione su livello copertura per beta #
545 > #-----#
546 >

```

```

547 > nc=30
548 > xs=1:30
549 >
550 > nt<-1000
551 >
552 > #valore iniziale
553 > beta.ic<-matrix(NA,ncol=2,nrow=nt) # NA sta per matrice vuota
554 >
555 > # simulazione
556 > for (i in 1:nt) {
557 +   error<-rnorm(n,0,4)
558 +   y<-5+3*x+error
559 +   beta<-cov(x,y)/var(x)
560 +   alpha<-mean(y)-beta*mean(x)
561 +   res<-y-alpha-beta*x
562 +   s2<-sum(res^2)/(n-2)
563 +   var.beta<-s2/sum((x-mean(x))^2)
564 +   beta.ic[i,1]<-beta-qt(.975,n-2)*sqrt(var.beta)
565 +   beta.ic[i,2]<-beta+qt(.975,n-2)*sqrt(var.beta)
566 + }
567 >
568 > # numero(beta.i<beta<beta.s)/nt
569 >
570 > sum( (beta.ic[,1]<=3)&(3<=beta.ic[,2]) )
571 [1] 948
572 > sum( (beta.ic[,1]<=3)&(3<=beta.ic[,2]) ) /nt
573 [1] 0.948
574 >
575 >
576 > #-----#
577 > # studio simulazione su livello copertura per alpha #
578 > #-----#
579 >
580 > nc=30
581 > xs=1:30
582 >
583 > nt<-1000
584 >
585 > #valore iniziale
586 > alpha.ic<-matrix(NA,ncol=2,nrow=nt) # NA sta per matrice vuota
587 >
588 > # simulazione
589 >
590 >
591 > for (i in 1:nt) {
592 +   error<-rnorm(n,0,4)
593 +   y<-5+3*x+error
594 +   beta<-cov(x,y)/var(x)
595 +   alpha<-mean(y)-beta*mean(x)
596 +   res<-y-alpha-beta*x
597 +   s2<-sum(res^2)/(n-2)
598 +   var.alpha<-s2*(1/(n+mean(x)^2)/sum((x-mean(x))^2))
599 +   alpha.ic[i,1]<-alpha-qt(.975,n-2)*sqrt(var.alpha)
600 +   alpha.ic[i,2]<-alpha+qt(.975,n-2)*sqrt(var.alpha)
601 + }
602 >
603 > sum( (alpha.ic[,1]<=5)&(5<=alpha.ic[,2]) )
604 [1] 953
605 > sum( (alpha.ic[,1]<=5)&(5<=alpha.ic[,2]) ) /nt
606 [1] 0.953
607 >
608 -----
609 Laboratorio 04
610 -----
611 > cement <- read.table("cement.dat",col.names=c("tempo", "resist"))
612 > attach(cement)
613 > cement
614 + tempo resist
615 1 1 13.0
616 2 1 13.3
617 3 1 11.8
618 4 2 21.9
619 5 2 24.5
620 6 2 24.7
621 7 3 29.8
622 8 3 28.0
623 9 3 24.1
624 10 3 24.2

```

```

625 11 3 26.2
626 12 7 32.4
627 13 7 30.4
628 14 7 34.5
629 15 7 33.1
630 16 7 35.7
631 17 28 41.8
632 18 28 42.6
633 19 28 40.3
634 20 28 35.7
635 21 28 37.3
636 > # Analisi preliminare
637 >
638 > plot(tempo,resist)
639 > cor(tempo,resist)
640 [1] 0.7862442
641 >
642 > # la relazione non sembra lineare
643 >
644 > # trasformazione
645 > # variabile esplicativa
646 >
647 > # log(x), exp(x), x^r, x^(-r)
648 >
649 > par(mfrow=c(2,2))
650 >
651 > plot(log(tempo),resist)
652 > cor(log(tempo),resist)
653 [1] 0.9396814
654 >
655 > plot(1/tempo,resist)
656 > cor(1/tempo,resist)
657 [1] -0.9499735
658 >
659 > plot(1/sqrt(tempo),resist)
660 > cor(1/sqrt(tempo),resist)
661 [1] -0.9732841
662 >
663 > plot(sqrt(tempo),resist)
664 > cor(sqrt(tempo),resist)
665 [1] 0.8613284
666 >
667 > x<-1/sqrt(tempo)
668 >
669 > # Stima del modello lineare
670 > # resist = beta_1 + beta_2*x + epsilon
671 > # var esplicativa: x, var risposta y
672 >
673 > fit <- lm(resist~x) # il nome "fit" in inglese significa "stima"
674 >
675 > summary(fit)
676
677 Call:
678 lm(formula = resist ~ x)
679
680 Residuals:
681 Min 1Q Median 3Q Max
682 -3.79469 -1.25666 -0.05666 1.89544 3.10531
683
684 Coefficients:
685 (Intercept) Estimate Std. Error t value Pr(>|t|)
686 45.655 1.023 44.63 < 2e-16 ***
687 x -32.599 1.764 -18.48 1.34e-13 ***
688 ---
689 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
690
691 Residual standard error: 2.133 on 19 degrees of freedom
692 Multiple R-Squared: 0.9473, Adjusted R-squared: 0.9445
693 F-statistic: 341.4 on 1 and 19 DF, p-value: 1.337e-13
694 >
695 > # NB: -----
696 > # Stima del modello lineare senza intercetta
697 > # resist = beta_2*x + epsilon
698 > # var esplicativa: x, var risposta y
699 > # fit <- lm(resist~x-1)
700 > # -----
701 >
702 > # resist = 45.655 - 32.599 * X

```

Page 9 of 45

```

703 > # H_0: alpha=0 vs H_1
704 > # sqrt(s^2)=2.133
705 > # R^2 = 1 - SOMMA_QUADRATI(res)/SOMMA_QUADRATI(y)
706 > # R^2 adjusted = 1 - (S_Q(res)/(n-2)) / (S_Q(y)/(n-1))
707 >
708 > # H_0: beta=0
709 >
710 > # coef, residuals, fitted, summary, deviance, predict, plot
711 >
712 > fit.val<-fit$fitted # valori stimati
713 > fit.val
714 1 2 3 4 5 6 7 8
715 13.05666 13.05666 13.05666 22.60456 22.60456 22.60456 26.83444 26.83444
716 9 10 11 12 13 14 15 16
717 26.83444 26.83444 26.83444 33.33414 33.33414 33.33414 33.33414 33.33414
718 17 18 19 20 21
719 39.49469 39.49469 39.49469 39.49469 39.49469 39.49469
720 > par(mfrow=c(1,1))
721 > plot(resist,fit.val) # plot tra valori originali e valori stimati
722 > abline(0,1) # piu' sono vicini alla bisettrice e piu' il modello e' buono
723 >
724 >
725 > # valutazione della bonta' del modello
726 > # analisi dei residui
727 >
728 > res<-fit$residuals # residui standardizzati (media=0 e var=1)
729 > res
730 1 2 3 4 5 6
731 -0.05665701 0.24334299 -1.25665701 -0.70456225 1.89543775 2.09543775
732 7 8 9 10 11 12
733 2.96555879 -2.73444127 -2.63444127 -0.63444127 -0.93413660
734 13 14 15 16 17 18
735 -2.93413660 1.16586340 -0.23413660 2.36586340 2.30530942 3.10530942
736 19 20 21
737 0.80530942 -3.79469058 -2.19469058
738 > # altro comando
739 > res<-rstandard(fit) # residui standardizzati (media=0 e var=1)
740 > res
741 1 2 3 4 5 6
742 -0.02984778 0.12819679 -0.66202607 -0.34305540 0.92289950 1.02028065
743 7 8 9 10 11 12
744 1.42684868 0.56079683 -1.31564884 -1.26753485 -0.30525502 -0.45196818
745 13 14 15 16 17 18
746 -1.41963860 0.56408577 -0.11328353 1.14468802 1.15297119 1.55308101
747 19 20 21
748 0.40276526 -1.89786623 -1.09764658
749 >
750 > # summary, boxplot, hist
751 > boxplot(res)
752 > # c'e' un po' di asimmetria dei residui
753 >
754 > hist(res)
755 > # non molto campanulare
756 >
757 > # grafico(res)
758 > plot(res)
759 > abline(h=0)
760 >
761 > # grafico(res,stimati)
762 > plot(res,fit.val)
763 >
764 > # grafico(x,res)
765 > plot(x,res)
766 >
767 > # varianza sigma^2 non sembra costante
768 >
769 > qqnorm(res)
770 > qqline(res)
771 >
772 > # -----
773 > # R^2, qqnorm/hist(res), grafici(res)
774 >
775 > help(lm)
776 > lm(resist~I(x^2)) # la "I" serve per specificare la formula del modello
777 >
778 Call:
779 lm(formula = resist ~ I(x^2))
780 >

```

Page 10 of 45

```

781 Coefficients:
782 (Intercept) I(x^2)
783 37.87 -26.92
784
785 >
786 > fit$coeff
787 (Intercept) x
788 45.65524 -32.59859
789 > plot(x,resist)
790 > abline(fit$coeff)
791 >
792 >
793 > # -----
794 > # -----
795 >
796 >
797 > windmill<-read.table("windmill.dat",header=T)
798 > windmill
799 wind dc
800 1 5.00 1.582
801 2 6.00 1.822
802 3 3.40 1.057
803 4 2.70 0.500
804 5 10.00 2.236
805 6 9.70 2.386
806 7 9.55 2.294
807 8 3.05 0.558
808 9 8.15 2.166
809 10 6.20 1.866
810 11 2.90 0.653
811 12 6.35 1.930
812 13 4.60 1.562
813 14 5.80 1.737
814 15 7.40 2.088
815 16 3.60 1.137
816 17 7.85 2.179
817 18 8.80 2.112
818 19 7.00 1.800
819 20 5.45 1.501
820 21 9.10 2.303
821 22 10.20 2.310
822 23 4.10 1.194
823 24 3.95 1.144
824 25 2.45 0.123
825 >
826 > attach(windmill)
827 >
828 > # analisi preliminare
829 >
830 > plot(wind,dc)
831 > cor(wind,dc)
832 [1] 0.9351434
833 > # mmhh... sembra esserci un legame lineare!
834 >
835 > # stima del modello
836 > fit<-lm(dc~wind)
837 >
838 > summary(fit)
839
840 Call:
841 lm(formula = dc ~ wind)
842
843 Residuals:
844 Min 1Q Median 3Q Max
845 -0.59869 -0.14099 0.06059 0.17262 0.32184
846
847 Coefficients:
848 Estimate Std. Error t value Pr(>|t|)
849 (Intercept) 0.13088 0.12599 1.039 0.31
850 wind 0.24115 0.01905 12.659 7.55e-12 ***
851 ---
852 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
853
854 Residual standard error: 0.2361 on 23 degrees of freedom
855 Multiple R-Squared: 0.8745, Adjusted R-squared: 0.869
856 F-statistic: 160.3 on 1 and 23 DF, p-value: 7.546e-12
857 >
858 >

```

Page 11 of 45

```

859 > # H_0: alpha=0
860 > qt(.975,23)
861 [1] 2.068658
862 >
863 > # dc = 0.24115*wind
864 > abline(fit$coeff)
865 >
866 >
867 > # -----
868 > # senza intercetta (come esercizio)
869 >
870 > # bonta del modello
871 > fit<-lm(dc~wind-1)
872 > summary(fit)
873
874 Call:
875 lm(formula = dc ~ wind - 1)
876
877 Residuals:
878 Min 1Q Median 3Q Max
879 -0.51276 -0.17156 0.08675 0.20282 0.36832
880
881 Coefficients:
882 Estimate Std. Error t value Pr(>|t|)
883 wind 0.25949 0.00715 36.29 <2e-16 ***
884 ---
885 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
886
887 Residual standard error: 0.2364 on 24 degrees of freedom
888 Multiple R-Squared: 0.9821, Adjusted R-squared: 0.9814
889 F-statistic: 1317 on 1 and 24 DF, p-value: < 2.2e-16
890 >
891 >
892 > res<-fit$res
893 > qqnorm(res)
894 > qqline(res)
895 >
896 > plot(wind,res)
897 >
898 > # R^2 elevato, qqnorm soddisfacente, plot(x,res) ha un andamento poco
899 > # soddisfacente
900 > # --> bisogna riformulare il modello e ristimarlo
901 >
902 > # trasformazione della variabile x
903 >
904 > par(mfrow=c(2,2))
905 >
906 > plot(log(wind),dc)
907 > plot(exp(wind),dc)
908 > plot(wind^2,dc)
909 > plot(1/wind,dc)
910 >
911 > cor(log(wind),dc)
912 [1] 0.978454
913 > cor(1/wind,dc)
914 [1] -0.989962
915 >
916 > fit1<-lm(dc~I(1/wind))
917 > summary(fit1)
918
919 Call:
920 lm(formula = dc ~ I(1/wind))
921
922 Residuals:
923 Min 1Q Median 3Q Max
924 -0.20547 -0.04941 0.01100 0.08352 0.12204
925
926 Coefficients:
927 Estimate Std. Error t value Pr(>|t|)
928 (Intercept) 2.9789 0.0449 66.34 <2e-16 ***
929 I(1/wind) -6.9345 0.2064 -33.59 <2e-16 ***
930 ---
931 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
932
933 Residual standard error: 0.09417 on 23 degrees of freedom
934 Multiple R-Squared: 0.98, Adjusted R-squared: 0.9792
935 F-statistic: 1128 on 1 and 23 DF, p-value: < 2.2e-16
936 >

```

Page 12 of 45

```
937 >
938 > # dc = 2.47 - 6.93/wind # nuovo modello
939 > # bonta' del modello
940 > res1<-fit$residuals
941 >
942 > par(mfrow=c(1,1))
943 > boxplot(res1)
944 >
945 > qqnorm(res1)
946 > qqline(res1)
947 >
948 > plot(1/wind, res1)
949 >
950 > # R^2, qqnorm
951 > plot(res1,fit$resfitted)
952 >
953 -----
954 Laboratorio 05
955 -----
956 > fruit<-read.table("Fruitfly.dat",col.names=c("RS","SS","NS"))
957 > fruit
958      RS      SS      NS
959 1  12.8  38.4  35.4
960 2  21.6  32.9  27.4
961 3  14.8  48.5  19.3
962 4  23.1  20.9  41.8
963 5  34.6  11.6  20.3
964 6  19.7  22.3  37.6
965 7  22.6  30.2  36.9
966 8  29.6  33.4  37.3
967 9  16.4  26.7  28.2
968 10 20.3  39.0  23.4
969 11 29.3  12.8  33.7
970 12 14.9  14.6  29.2
971 13 27.3  12.2  41.7
972 14 22.4  23.1  22.6
973 15 27.5  29.4  40.4
974 16 20.3  16.0  34.4
975 17 38.7  20.1  30.4
976 18 26.4  23.3  14.9
977 19 23.7  22.9  51.8
978 20 26.1  22.5  33.8
979 21 29.5  15.1  37.9
980 22 38.6  31.0  29.5
981 23 44.4  16.9  42.4
982 24 23.2  16.1  36.6
983 25 23.6  10.8  47.4
984 > attach(fruit)
985 > names(fruit)
986 [1] "RS" "SS" "NS"
987 >
988 > # Analisi preliminare (statistiche descrittivi/grafiche)
989 > summary(RS)
990      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
991    12.80   20.30   23.60   25.26  29.30   44.40
992 > summary(SS)
993      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
994    10.80   16.00   22.50   23.63  30.20   48.50
995 > summary(NS)
996      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
997    14.90   28.20   34.40   33.37  37.90   51.80
998 >
999 > boxplot(RS,SS,NS)
1000 >
1001 > # differenza (RS,SS) e NS
1002 > # differenza RS e SS
1003 >
1004 > ## Se Y1-F(theta1) e Y2-F(theta2), devo confrontare theta1 con theta2
1005 > ## H0: mu1=mu2
1006 >
1007 > RSS<-c(RS,SS)
1008 > RSS
1009 [1] 12.8 21.6 14.8 23.1 34.6 19.7 22.6 29.6 16.4 20.3 29.3 14.9 27.3 22.4 27.5
1010 [16] 20.3 30.2 38.7 39.0 12.8 14.6 12.2 23.1 29.4 16.0 20.1 23.3 22.9 22.5
1011 [31] 22.3 30.2 33.4 26.7 39.0 12.8 14.6 12.2 23.1 29.4 16.0 20.1 23.3 22.9 22.5
1012 [46] 15.1 31.0 16.9 16.1 10.8
1013 > # t-student mu(RSS) = mu(NS)
1014 > # normalita', sigma^2 omoschedasticita'
```

Page 13 of 45

```
1015 >
1016 > boxplot(RSS,NS)
1017 > hist(RSS,nclass=8)
1018 > hist(NS,nclass=8)
1019 >
1020 > qqnorm(RSS)
1021 > qqline(RSS)
1022 >
1023 > qqnorm(NS)
1024 > qqline(NS)
1025 >
1026 > var(RSS)
1027 [1] 77.00249
1028 > var(NS)
1029 [1] 79.9596
1030 >
1031 > # TRV (test rapporto di verosim. W(theta))
1032 > # H0: sigma^2(RSS) = sigma^2(NS)
1033 > # '---> H0: sigma^2(RSS)/sigma^2(NS)=1
1034 > # '---> H0: tau=1
1035 > # '---> H0: tau=S^2_1/S^2_2-F(n-1,m-1)=1
1036 > help(var.test)
1037 > var.test(RSS,NS)
1038
1039      F test to compare two variances
1040
1041 data:  RSS and NS
1042 F = 0.963, num df = 49, denom df = 24, p-value = 0.8933
1043 alternative hypothesis: true ratio of variances is not equal to 1
1044 95 percent confidence interval:
1045  0.4564086 1.8658070
1046 sample estimates:
1047 ratio of variances
1048      0.9630174
1049
1050 >
1051 > # test t H0: mu(RSS) = mu(NS)
1052 > # test t H0: mu(RSS) - mu(NS) = 0
1053 >
1054 > # t = (mean1 - mean2) - (mu1 - mu2) / sqrt(stima varianza)
1055 >
1056 > help(t.test)
1057 > t.test(RSS,NS,var.equal=T)
1058
1059      Two Sample t-test
1060
1061 data:  RSS and NS
1062 t = -4.1286, df = 73, p-value = 9.587e-05
1063 alternative hypothesis: true difference in means is not equal to 0
1064 95 percent confidence interval:
1065  -13.240812  -4.619188
1066 sample estimates:
1067 mean of x mean of y
1068    24.442    33.372
1069
1070 > # dal p-value --> rifiuto H0 --> le medie sono differenti
1071 >
1072 > boxplot(RS,SS,NS)
1073 >
1074 > # RS e NS sono uguali?
1075 >
1076 > qqnorm(RS);qqline(RS)
1077 >
1078 > var(RS)
1079 [1] 60.41007
1080 > var(NS)
1081 [1] 79.9596
1082 >
1083 > var.test(RS,NS)
1084
1085      F test to compare two variances
1086
1087 data:  RS and NS
1088 F = 0.755, num df = 24, denom df = 24, p-value = 0.4974
1089 alternative hypothesis: true ratio of variances is not equal to 1
1090 95 percent confidence interval:
1091  0.3329286 1.7144557
1092 sample estimates:
```

Page 14 of 45

```
1093 ratio of variances
1094 0.7555074
1095
1096 >
1097 > t.test(RS,NS,var.equal=T)
1098
1099      Two Sample t-test
1100
1101 data:  RS and NS
1102 t = -3.4251, df = 48, p-value = 0.001268
1103 alternative hypothesis: true difference in means is not equal to 0
1104 95 percent confidence interval:
1105  -12.880308  -3.351692
1106 sample estimates:
1107 mean of x mean of y
1108    25.256    33.372
1109
1110 >
1111 > # esercizio: SS-NS
1112 >
1113 > t.test(RS,NS)
1114
1115      Welch Two Sample t-test
1116
1117 data:  RS and NS
1118 t = -3.4251, df = 47.087, p-value = 0.001283
1119 alternative hypothesis: true difference in means is not equal to 0
1120 95 percent confidence interval:
1121  -12.882696  -3.349304
1122 sample estimates:
1123 mean of x mean of y
1124    25.256    33.372
1125
1126 >
1127 > boxplot(RS,SS,NS)
1128 >
1129 > # normalita'
1130 > qqnorm(SS);qqline(SS)
1131 >
1132 > # omoschedasticita'
1133 > var(RS)
1134 [1] 60.41007
1135 > var(SS)
1136 [1] 95.42293
1137 >
1138 > var.test(RS,SS)
1139
1140      F test to compare two variances
1141
1142 data:  RS and SS
1143 F = 0.6331, num df = 24, denom df = 24, p-value = 0.2698
1144 alternative hypothesis: true ratio of variances is not equal to 1
1145 95 percent confidence interval:
1146  0.2789774 1.4366273
1147 sample estimates:
1148 ratio of variances
1149    0.633077
1150
1151 >
1152 > t.test(RS,SS,var.equal=T)
1153
1154      Two Sample t-test
1155
1156 data:  RS and SS
1157 t = 0.6521, df = 48, p-value = 0.5175
1158 alternative hypothesis: true difference in means is not equal to 0
1159 95 percent confidence interval:
1160  -3.391875  6.647875
1161 sample estimates:
1162 mean of x mean of y
1163    25.256    23.628
1164
1165 >
1166 > boxplot(RS,SS,NS)
1167 > boxplot(log(RS),log(SS),log(NS))
1168 > qqnorm(log(RS));qqline(log(RS))
1169 >
1170 > detach()
```

Page 15 of 45

```
1171 >
1172 > -----
1173 > capto<-read.table("Capto.dat",header=T)
1174 > attach(capto)
1175 > capto
1176      Sp  Sd  Dp  Dd
1177 1  210 201 130 125
1178 2  169 165 122 121
1179 3  187 166 124 121
1180 4  160 157 104 106
1181 5  167 147 112 101
1182 6  176 145 101 85
1183 7  185 168 121 98
1184 8  206 180 124 105
1185 9  173 147 115 103
1186 10 146 136 102 98
1187 11 174 151 98 90
1188 12 201 168 119 98
1189 13 198 179 106 110
1190 14 148 129 107 103
1191 15 154 131 100 82
1192 >
1193 > # Y1 ~ N(m1,s1)
1194 > # Y2 ~ N(m2,s2)
1195 >
1196 > # D = Y1 - Y2 ~ N(m1-m2,s1+s2) # prima
1197 > # D = Y1 - Y2 ~ N(m1-m2,s1+s2-2*cov(Y1,Y2)) # adesso!
1198 > # D ~ N(m,s) , m=m1-m2 , s=s1+s2-2*cov(Y1,Y2)
1199 > # H0: m=0
1200 >
1201 > # mean(D)-m / stima(varianza)
1202 >
1203 > SD<-Sd-Sp
1204 > DD<-Dd-Dp
1205 >
1206 > boxplot(SD)
1207 > qqnorm(SD);qqline(SD)
1208 > qqnorm(DD);qqline(DD)
1209 >
1210 > t.test(SD)
1211
1212      One Sample t-test
1213
1214 data:  SD
1215 t = -8.1228, df = 14, p-value = 1.146e-06
1216 alternative hypothesis: true mean is not equal to 0
1217 95 percent confidence interval:
1218  -23.93258 -13.93409
1219 sample estimates:
1220 mean of x
1221  -18.93333
1222
1223 >
1224 > t.test(Sd,Sp,paired=T) # fa la stessa cosa che ho fatto prima
1225
1226      Paired t-test
1227
1228 data:  Sd and Sp
1229 t = -8.1228, df = 14, p-value = 1.146e-06
1230 alternative hypothesis: true difference in means is not equal to 0
1231 95 percent confidence interval:
1232  -23.93258 -13.93409
1233 sample estimates:
1234 mean of the differences
1235  -18.93333
1236
1237 >
1238 > # H0: m >= 0
1239 > # H1: m < 0
1240 >
1241 > # H0: m <= 0
1242 > # H1: m > 0
1243 >
1244 > t.test(Sd,Sp,paired=T,alternative="g")
1245
1246      Paired t-test
1247
1248 data:  Sd and Sp
```

Page 16 of 45

```
1249 t = -8.1228, df = 14, p-value = 1
1250 alternative hypothesis: true difference in means is greater than 0
1251 95 percent confidence interval:
1252 -23.03874      Inf
1253 sample estimates:
1254 mean of the differences
1255 -18.93333
1256
1257 >
1258 > # esercizio: test per la D
1259 > t.test(Dd,Dp,paired=T)
1260
1261      Paired t-test
1262
1263 data:  Dd and Dp
1264 t = -4.1662, df = 14, p-value = 0.000951
1265 alternative hypothesis: true difference in means is not equal to 0
1266 95 percent confidence interval:
1267 -14.037215 -4.496118
1268 sample estimates:
1269 mean of the differences
1270 -9.266667
1271
1272 >
1273 > t.test(Dd,Dp,paired=T,alternative="g")
1274
1275      Paired t-test
1276
1277 data:  Dd and Dp
1278 t = -4.1662, df = 14, p-value = 0.9995
1279 alternative hypothesis: true difference in means is greater than 0
1280 95 percent confidence interval:
1281 -13.18427      Inf
1282 sample estimates:
1283 mean of the differences
1284 -9.266667
1285
1286 -----
1287 Laboratorio 06
1288 -----
1289 > # caricamento dati
1290 > hook<-read.table('hook.dat',col.names=c('temp','press'))
1291 > hook
1292
1293   temp  press
1294 1 210.8 29.211
1295 2 210.2 28.559
1296 3 208.4 27.972
1297 4 202.5 24.697
1298 5 200.6 23.726
1299 6 200.1 23.369
1300 7 199.5 23.030
1301 8 197.0 21.892
1302 9 196.4 21.928
1303 10 196.3 21.654
1304 11 195.6 21.605
1305 12 193.4 20.480
1306 13 193.6 20.212
1307 14 191.4 19.758
1308 15 191.1 19.490
1309 16 190.6 19.386
1310 17 189.5 18.869
1311 18 188.8 18.356
1312 19 188.5 18.507
1313 20 185.7 17.267
1314 21 186.0 17.221
1315 22 185.6 17.062
1316 23 184.1 16.959
1317 24 184.6 16.881
1318 25 184.1 16.817
1319 26 183.2 16.385
1320 27 182.4 16.235
1321 28 181.9 16.106
1322 29 181.0 15.928
1323 30 181.0 15.919
1324 > attach(hook)
1325 >
1326 > # analisi preliminare: grafici
```

Page 17 of 45

```
1327 >
1328 > plot(press,temp)
1329 > cor(press,temp)
1330 [1] 0.9958803
1331 >
1332 > # stima del modello
1333 > fit<-lm(temp~press) # y in funzione di x --> y-x
1334 > summary(fit)
1335
1336 Call:
1337 lm(formula = temp ~ press)
1338
1339 Residuals:
1340      Min       1Q   Median       3Q      Max
1341 -1.6735 -0.6805  0.2203  0.5296  1.3976
1342
1343 Coefficients:
1344              Estimate Std. Error t value Pr(>|t|)
1345 (Intercept) 146.67290   0.77641  188.91 <2e-16 ***
1346 press       2.25260     0.03809   59.14 <2e-16 ***
1347 ---
1348 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1349
1350 Residual standard error: 0.806 on 29 degrees of freedom
1351 Multiple R-Squared: 0.9918,    Adjusted R-squared: 0.9915
1352 F-statistic: 3498 on 1 and 29 DF,  p-value: < 2.2e-16
1353
1354 >
1355 > # temp = 146.6 + 2.25*press    ## modello dei minimi quadrati
1356 >
1357 > # analisi dei residui
1358 >
1359 > # riassunti
1360 > # analisi grafiche
1361 > res<-fit$res
1362 >
1363 > # temp = 146.6 + 2.25*press    ## modello dei minimi quadrati
1364 >
1365 > # analisi dei residui
1366 >
1367 > # riassunti
1368 > # analisi grafiche
1369 > res<-fit$res
1370 > rm(res)
1371 > res<-fit$res
1372 > boxplot(res)
1373 > plot(res)
1374 >
1375 > plot(press,res)
1376 >
1377 > par(mfrow=c(2,2))
1378 > plot(fit)
1379 >
1380 > # lieve asimmetria dei residui
1381 > # x vs residui mostra un andamento di tipo parabolico
1382 >
1383 > # rimedio: riformulare il modello con una trasformata della x
1384 >
1385 > plot(press,res)
1386 >
1387 > # y= a + bx + cx^2
1388 > fit1<-lm(temp~press+I(press^2),x=T) # I: funz. indicatrice
1389 > fit1$x
1390 (Intercept) press I(press^2)
1391 1          1 29.211 853.2825
1392 2          1 28.559 815.6165
1393 3          1 27.972 782.4328
1394 4          1 24.697 609.9418
1395 5          1 23.726 562.9231
1396 6          1 23.369 546.1102
1397 7          1 23.030 530.3809
1398 8          1 21.892 479.2597
1399 9          1 21.928 480.8372
1400 10         1 21.654 468.8957
1401 11         1 21.605 466.7760
1402 12         1 20.480 419.4304
1403 13         1 20.212 408.5249
1404 14         1 19.758 390.3786
```

Page 18 of 45

```
1405 15         1 19.490 379.8601
1406 16         1 19.386 375.8170
1407 17         1 18.869 356.0392
1408 18         1 18.356 336.9427
1409 19         1 18.507 342.5090
1410 20         1 17.267 298.1493
1411 21         1 17.221 296.5628
1412 22         1 17.062 291.1118
1413 23         1 16.959 287.6077
1414 24         1 16.881 284.9682
1415 25         1 16.817 282.8115
1416 26         1 16.385 268.4682
1417 27         1 16.235 263.5752
1418 28         1 16.106 259.4032
1419 29         1 15.928 253.7012
1420 30         1 15.919 253.4146
1421 31         1 15.376 236.4214
1422 attr(,"assign")
1423 [1] 0 1 2
1424 >
1425 > summary(fit1)
1426
1427 Call:
1428 lm(formula = temp ~ press + I(press^2), x = T)
1429
1430 Residuals:
1431      Min       1Q   Median       3Q      Max
1432 -0.73906 -0.26314 -0.01578  0.25139  0.73891
1433
1434 Coefficients:
1435              Estimate Std. Error t value Pr(>|t|)
1436 (Intercept) 126.701623   2.112363  59.981 < 2e-16 ***
1437 press       4.157627   0.199069  20.885 < 2e-16 ***
1438 I(press^2)  -0.043754   0.004552  -9.612 2.29e-10 ***
1439 ---
1440 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1441
1442 Residual standard error: 0.3956 on 28 degrees of freedom
1443 Multiple R-Squared: 0.9981,    Adjusted R-squared: 0.998
1444 F-statistic: 7307 on 2 and 28 DF,  p-value: < 2.2e-16
1445
1446 > plot(temp,fit1$fitted)
1447 > abline(0,1)
1448 >
1449 > # analisi dei residui
1450 >
1451 > res1<-fit1$res
1452 > plot(res1)
1453 >
1454 > plot(press,res1)
1455 > plot(I(press^2),res1)
1456 >
1457 > qqnorm(res1);qqline(res1)
1458 >
1459 > detach()
1460 >
1461 >
1462 > # -----
1463 > # cherry.dat
1464 > # caricamento dati
1465 > cherry<-read.table('cherry.dat',col.names=c('diam','alt','vol'))
1466 > attach(cherry)
1467 >
1468 > # analisi preliminare
1469 > par(mfrow=c(2,2))
1470 > plot(vol~diam)
1471 > plot(vol~alt)
1472 > cor(vol,diam)
1473 [1] 0.9671194
1474 > cor(vol,alt)
1475 [1] 0.5982497
1476 > plot(diam,alt)
1477 > cor(diam,alt)
1478 [1] 0.5192801
1479 > pairs(cbind(vol,diam,alt)) # riassume tutti i grafici di dispersione
1480 >
1481 > # volume ~ diametro * altezza
1482 > # log(volume) ~ log(diametro) + log(altezza)
```

Page 19 of 45

```
1483 > pairs(cbind(log(vol),log(diam),log(alt)))
1484 > cor(log(vol),log(diam))
1485 [1] 0.976665
1486 > cor(log(vol),log(alt))
1487 [1] 0.6486377
1488 >
1489 > # ----- stima del modello -----
1490 > fit<-lm(log(vol)~log(diam)+log(alt))
1491 > summary(fit)
1492
1493 Call:
1494 lm(formula = log(vol) ~ log(diam) + log(alt))
1495
1496 Residuals:
1497      Min       1Q   Median       3Q      Max
1498 -0.168561 -0.048488  0.002431  0.063637  0.129223
1499
1500 Coefficients:
1501              Estimate Std. Error t value Pr(>|t|)
1502 (Intercept) -6.63162   0.79979  -8.292 5.06e-09 ***
1503 log(diam)    1.98265   0.07501  26.432 < 2e-16 ***
1504 log(alt)     1.11712   0.20444   5.464 7.81e-06 ***
1505 ---
1506 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1507
1508 Residual standard error: 0.08139 on 28 degrees of freedom
1509 Multiple R-Squared: 0.9777,    Adjusted R-squared: 0.9761
1510 F-statistic: 613.2 on 2 and 28 DF,  p-value: < 2.2e-16
1511
1512 >
1513 > # log(volume) ~ -6.63 + 1.98*log(diametro) + 1.11*log(altezza)
1514 > plot(log(vol),fit$fitt)
1515 > abline(0,1)
1516 >
1517 > # analisi dei residui
1518 > res<-fit$res
1519 > boxplot(res)
1520 >
1521 > plot(res,log(diam))
1522 > plot(log(diam),res)
1523 > plot(log(alt),res)
1524 >
1525 > par(mfrow=c(2,2))
1526 > plot(fit)
1527 >
1528 >
1529 > ynew<-log(vol)[-18]
1530 > ynew
1531 [1] 2.332144 2.332144 2.322388 2.797281 2.933857 2.980619 2.747271 2.901422
1532 [9] 3.117950 2.990720 3.186353 3.044522 3.063391 3.058707 2.949688 3.100092
1533 [17] 3.520461 3.246491 3.214868 3.540959 3.456317 3.591818 3.645450 3.751854
1534 [25] 4.014580 4.019980 4.065602 3.941582 3.931826 4.343805
1535 > length(ynew)
1536 [1] 30
1537 > x1new<-log(diam)[-18]
1538 > x2new<-log(alt)[-18]
1539 >
1540 > fit1<-lm(ynew ~ x1new + x2new)
1541 > summary(fit1)
1542
1543 Call:
1544 lm(formula = ynew ~ x1new + x2new)
1545
1546 Residuals:
1547      Min       1Q   Median       3Q      Max
1548 -0.17500 -0.05706  0.00624  0.05940  0.11383
1549
1550 Coefficients:
1551              Estimate Std. Error t value Pr(>|t|)
1552 (Intercept) -7.18549   0.78061  -9.205 8.14e-10 ***
1553 x1new        1.95816   0.07051  27.770 < 2e-16 ***
1554 x2new        1.26100   0.19984   6.310 9.39e-07 ***
1555 ---
1556 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1557
1558 Residual standard error: 0.07565 on 27 degrees of freedom
1559 Multiple R-Squared: 0.9814,    Adjusted R-squared: 0.98
1560 F-statistic: 712.3 on 2 and 27 DF,  p-value: < 2.2e-16
```

Page 20 of 45

```

1561
1562 >
1563 > plot(fit)
1564 >
1565 > # altro modo senza ridefinire le variabili
1566 > fit1<-lm(log(vol)-log(diam)+log(alt), subset=-c(18))
1567 > summary(fit1)
1568
1569 Call:
1570 lm(formula = log(vol) ~ log(diam) + log(alt), subset = -c(18))
1571
1572 Residuals:
1573     Min       1Q   Median       3Q      Max
1574 -0.17500 -0.05706  0.00624  0.05940  0.11383
1575
1576 Coefficients:
1577             Estimate Std. Error t value Pr(>|t|)
1578 (Intercept) -7.18549    0.78061  -9.205 8.14e-10 ***
1579 log(diam)    1.95816    0.07051   27.770 < 2e-16 ***
1580 log(alt)     1.26100    0.19984    6.310 9.39e-07 ***
1581 ---
1582 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1583
1584 Residual standard error: 0.07565 on 27 degrees of freedom
1585 Multiple R-Squared:  0.9814,    Adjusted R-squared:  0.98
1586 F-statistic: 712.3 on 2 and 27 DF,  p-value: < 2.2e-16
1587
1588 >
1589 > fit1<-lm(log(vol)-log(diam)+log(alt), subset=-c(11,15,16,17,18))
1590 > plot(fit1)
1591 > summary(fit1)
1592
1593 Call:
1594 lm(formula = log(vol) ~ log(diam) + log(alt), subset = -c(11,
1595     15, 16, 17, 18))
1596
1597 Residuals:
1598     Min       1Q   Median       3Q      Max
1599 -0.083857 -0.047953 -0.003806  0.039704  0.103097
1600
1601 Coefficients:
1602             Estimate Std. Error t value Pr(>|t|)
1603 (Intercept) -6.55062    0.63479  -10.319 4.21e-10 ***
1604 log(diam)    1.99231    0.05569   35.773 < 2e-16 ***
1605 log(alt)     1.09469    0.16394    6.678 8.21e-07 ***
1606 ---
1607 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1608
1609 Residual standard error: 0.05729 on 23 degrees of freedom
1610 Multiple R-Squared:  0.9907,    Adjusted R-squared:  0.9899
1611 F-statistic: 1223 on 2 and 23 DF,  p-value: < 2.2e-16
1612
1613 >
1614 > par(mfrow=c(1,1))
1615 > qqnorm(fit1$res)
1616 >
1617 -----
1618 Laboratorio 07
1619 -----
1620 > hills<-read.table('hills.dat',header=T)
1621 > hills
1622
1623   dist climb  time
1624 Greenmantle    2.5   650  16.083
1625 Carnethy       6.0  2500  48.350
1626 Craig Dunain   6.0   900  33.650
1627 Ben Rha       7.5   800  45.600
1628 Ben Lomond     8.0  3070  62.267
1629 Goatfell      8.0  2866  73.217
1630 Bens of Jura  16.0  7500 204.617
1631 Cairnpapple   6.0   800  36.367
1632 Scolty       5.0   800  29.750
1633 Traprain      6.0   650  39.750
1634 Lairig Ghru   28.0  2100 192.667
1635 Dollar        5.0  2000  43.050
1636 Lomonds       9.5  2200  65.000
1637 Cairn Table   6.0   500  44.133
1638 Eildon Two    4.5  1500  26.933
1639 Cairngorm    10.0  3000  72.250

```

Page 21 of 45

```

1639 Seven Hills    14.0  2200  98.417
1640 Knock Hill     3.0   350  78.650
1641 Black Hill     4.5  1000  17.417
1642 Cragg Beag     5.5   600  32.567
1643 Kildom Hill    3.0   300  15.950
1644 Meall Ant-Suidhe 3.5  1500  27.900
1645 Half Ben Nevis 6.0  2200  47.633
1646 Cow Hill       2.0   900  17.933
1647 N Berwick Law  3.0   600  18.683
1648 Cragg Dubh     4.0  2000  26.217
1649 Burnswark      6.0   800  34.433
1650 Largo Law      5.0   950  28.567
1651 Criffel        6.5  1750  50.500
1652 Acmony        5.0   500  20.950
1653 Ben Nevis      10.0  4400  85.583
1654 Knockfarrel   6.0   600  32.383
1655 Two Breweries  18.0  5200 170.250
1656 Cockleroi      4.5   850  28.100
1657 Moffat Chase  20.0  5000 159.833
1658 > attach(hills)
1659 >
1660 > # -----
1661 > # time = ??
1662 >
1663 > # ---- Analisi preliminare ----
1664 > pairs(hills)
1665 >
1666 > # Y = X*beta + epsilon
1667 > # "beta = (Xt X)^(-1) Xt y
1668 > # "beta = N(0,sigma^2 * (Xt X)^(-1))
1669 > # Se le colonne di (Xt X) sono esprimibili tramite le sue righe
1670 > # il rango della matrice non e' piu' pieno --> cioa' si sbaglia tutto
1671 > # se ci sono delle variabili in relazione tra di loro
1672 >
1673 > # R2=1 --> 'beta_i - 0
1674 >
1675 > cor(hills)
1676
1677   dist climb time
1678 1.0000000 0.6523461 0.9195892
1679 climb 0.6523461 1.0000000 0.8052392
1680 time 0.9195892 0.8052392 1.0000000
1681 > # Stima del modello
1682 >
1683 > fit<-lm(time-dist)
1684 > summary(fit)
1685
1686 Call:
1687 lm(formula = time ~ dist)
1688
1689 Residuals:
1690     Min       1Q   Median       3Q      Max
1691 -35.745  -9.037  -4.201   2.849  76.170
1692
1693 Coefficients:
1694             Estimate Std. Error t value Pr(>|t|)
1695 (Intercept)  -4.8407    5.7562  -0.841  0.406
1696 dist          8.3305    0.6196  13.446 6.08e-15 ***
1697 ---
1698 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1699
1700 Residual standard error: 19.96 on 33 degrees of freedom
1701 Multiple R-Squared:  0.8456,    Adjusted R-squared:  0.841
1702 F-statistic: 180.8 on 1 and 33 DF,  p-value: 6.084e-15
1703
1704 >
1705 > # il modello non sara' piu': time = alpha + beta * dist
1706 > # ma: time = beta * dist
1707 >
1708 > fit1<-lm(time-dist ~ 1) # tolgo l'intercetta
1709 > summary(fit1)
1710
1711 Call:
1712 lm(formula = time ~ dist ~ 1)
1713
1714 Residuals:
1715     Min       1Q   Median       3Q      Max
1716 -28.7646 -11.0285  -6.8327   0.5607  78.0847

```

Page 22 of 45

```

1717 Coefficients:
1718             Estimate Std. Error t value Pr(>|t|)
1719 dist    7.9083    0.3615   21.88 <2e-16 ***
1720 ---
1721 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1722
1723 Residual standard error: 19.87 on 34 degrees of freedom
1724 Multiple R-Squared:  0.9337,    Adjusted R-squared:  0.9317
1725 F-statistic: 478.6 on 1 and 34 DF,  p-value: < 2.2e-16
1726
1727 >
1728 > # NB: R2 e' cambiato --> R^2 = sum(y.stimati^2)/sum(y^2)
1729 > # modello: time = 7.90 * dist
1730 >
1731 > # ---- analisi bonta' del modello ----
1732 > # R^2
1733 > # analisi dei residui
1734 > # grafico (residui, variabile esclusa)
1735 >
1736 > res<-fit1$res
1737 > plot(res,climb)
1738 >
1739 > fit2<-lm(time-dist+climb~1) # oppure:
1740 > fit2<-update(fit1, . ~ . + climb)
1741 >
1742 > summary(fit2)
1743
1744 Call:
1745 lm(formula = time ~ dist + climb ~ 1)
1746
1747 Residuals:
1748     Min       1Q   Median       3Q      Max
1749 -18.089 -10.053  -5.539  -3.180  58.235
1750
1751 Coefficients:
1752             Estimate Std. Error t value Pr(>|t|)
1753 dist   5.605651    0.551046  10.173 1.05e-11 ***
1754 climb  0.010280    0.002118   4.853 2.84e-05 ***
1755 ---
1756 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1757
1758 Residual standard error: 15.41 on 33 degrees of freedom
1759 Multiple R-Squared:  0.9613,    Adjusted R-squared:  0.959
1760 F-statistic: 409.8 on 2 and 33 DF,  p-value: < 2.2e-16
1761
1762 >
1763 > # time = 5.60*dist + 0.01*climb
1764 > # R2: 0.93 - 0.96
1765 > # Analisi della variata # voglio vedere quale modello e' piu' efficace
1766 > # H0 : fit1
1767 > # H1 : fit2
1768 >
1769 > # F = ((SQR1-SQR2)/(n-p1-n+p2)) / (SQR2/(n-p2))
1770 >
1771 > anova(fit1,fit2)
1772 Analysis of Variance Table
1773
1774 Model 1: time ~ dist ~ 1
1775 Model 2: time ~ dist + climb ~ 1
1776      Res.DF  RSS Df Sum of Sq  F    Pr(>F)
1777 1      34 13423.2
1778 2      33  7832.5  1    5590.7 23.555 2.84e-05 ***
1779 ---
1780 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1781 > # rifiuto H0, conviene tener conto anche della seconda var esplicativa
1782 >
1783 > # time = 5.60*dist + 0.01*climb
1784 >
1785 > # analisi dei residui
1786 > res2<-fit2$res
1787 > boxplot(res2)
1788 >
1789 > qqnorm(res2);qqline(res2)
1790 >
1791 > plot(res2,fitted(fit2))
1792 > identify(res2,fitted(fit2),dimnames(hills)[[1]])
1793 [1] 7 11 33 35
1794 >

```

Page 23 of 45

```

1795 > fit3<-lm(time-dist+climb~1, subset=-c(7,11,33,35))
1796 > summary(fit3)
1797
1798 Call:
1799 lm(formula = time ~ dist + climb ~ 1, subset = -c(7, 11, 33,
1800     35))
1801
1802 Residuals:
1803     Min       1Q   Median       3Q      Max
1804 -14.161  -4.120  -2.137   1.415  59.386
1805
1806 Coefficients:
1807             Estimate Std. Error t value Pr(>|t|)
1808 dist   5.762763    0.779834   7.390 3.84e-08 ***
1809 climb  0.005645    0.002886   1.956  0.0602 .
1810 ---
1811 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1812
1813 Residual standard error: 12.29 on 29 degrees of freedom
1814 Multiple R-Squared:  0.9364,    Adjusted R-squared:  0.932
1815 F-statistic: 213.6 on 2 and 29 DF,  p-value: < 2.2e-16
1816
1817 >
1818 > fit3<-lm(time-dist+climb~1, subset=-c(7,11,33,35))
1819 > summary(fit3)
1820
1821 Call:
1822 lm(formula = time ~ dist + climb ~ 1, subset = -c(7, 11, 33,
1823     35))
1824
1825 Residuals:
1826     Min       1Q   Median       3Q      Max
1827 -14.161  -4.120  -2.137   1.415  59.386
1828
1829 Coefficients:
1830             Estimate Std. Error t value Pr(>|t|)
1831 dist   5.762763    0.779834   7.390 3.84e-08 ***
1832 climb  0.005645    0.002886   1.956  0.0602 .
1833 ---
1834 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1835
1836 Residual standard error: 12.29 on 29 degrees of freedom
1837 Multiple R-Squared:  0.9364,    Adjusted R-squared:  0.932
1838 F-statistic: 213.6 on 2 and 29 DF,  p-value: < 2.2e-16
1839
1840 > detach()
1841 >
1842 >
1843 >
1844 > # -----
1845 > # dati - gasoline
1846 > gasco<-read.table("gasoline.dat")
1847 > names(gasco)
1848 [1] "V1" "V2" "V3" "V4" "V5"
1849 > names(gasco)<-c('y','x1','x2','x3','x4')
1850 > ls()
1851 [1] "fit" "fit1" "fit2" "fit3" "gasco" "hills" "res" "res2"
1852 > attach(gasco)
1853 > y
1854 [1] 6.9 14.4 7.4 8.5 8.0 2.8 5.0 12.2 10.0 15.2 26.8 14.0 14.7 6.4 17.6
1855 [16] 22.3 24.8 26.0 34.9 18.2 23.2 18.0 13.1 16.1 32.1 34.7 31.7 33.6 30.4 26.6
1856 [31] 27.8 45.7
1857 >
1858 > # analisi preliminare
1859 > pairs(gasco)
1860 >
1861 > # (x2,x3)
1862 >
1863 > cor(gasco)
1864
1865   y      x1      x2      x3      x4
1866 y  1.0000000  0.2463260  0.3840706 -0.3150243  0.7152622
1867 x1 0.2463260  1.0000000  0.6205867 -0.7001539 -0.3216782
1868 x2 0.3840706  0.6205867  1.0000000 -0.9062248 -0.2979843
1869 x3 -0.3150243 -0.7001539 -0.9062248  1.0000000  0.4122466
1870 x4 0.7152622 -0.3216782 -0.2979843  0.4122466  1.0000000
1871 >
1872 > # stima del modello

```

Page 24 of 45

```
1873 > fit1<-lm(y~x4)
1874 > summary(fit1)
1875
1876 Call:
1877 lm(formula = y ~ x4)
1878
1879 Residuals:
1880      Min       1Q   Median       3Q      Max
1881 -14.75837  -6.27829   0.05255   5.16243  17.84805
1882
1883 Coefficients:
1884             Estimate Std. Error t value Pr(>|t|)
1885 (Intercept) -16.66206    6.68721  -2.492   0.0185 *
1886 x4           0.10937    0.01972   5.546 4.98e-06 ***
1887 ---
1888 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1889
1890 Residual standard error: 7.659 on 30 degrees of freedom
1891 Multiple R-Squared: 0.5063,    Adjusted R-squared: 0.4898
1892 >
1893 > # y = -16.66 + 0.10*x4
1894 >
1895 > fit11<-update(fit1, . ~ .+x1)
1896 > fit12<-update(fit1, . ~ .+x2)
1897 > fit13<-update(fit1, . ~ .+x3)
1898 >
1899 > anova(fit1,fit11)
1900 Analysis of Variance Table
1901
1902 Model 1: y ~ x4
1903 Model 2: y ~ x4 + x1
1904      Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1905 1         30  1759.69
1906 2         29   861.95    1    897.75 30.204 6.4e-06 ***
1907 ---
1908 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1909 > anova(fit1,fit12)
1910 Analysis of Variance Table
1911
1912 Model 1: y ~ x4
1913 Model 2: y ~ x4 + x2
1914      Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1915 1         30  1759.69
1916 2         29  369.87    1   1389.83 108.97 2.468e-11 ***
1917 ---
1918 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1919 > anova(fit1,fit13)
1920 Analysis of Variance Table
1921
1922 Model 1: y ~ x4
1923 Model 2: y ~ x4 + x3
1924      Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1925 1         30  1759.69
1926 2         29  170.61    1   1589.08 270.11 3.111e-16 ***
1927 ---
1928 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1929 >
1930 > # scelgo il trezo fit perche' ha il 'Sum of Sq' maggiore R2
1931 >
1932 > summary(fit3)
1933
1934 Call:
1935 lm(formula = time ~ dist + climb - 1, subset = ~c(7, 11, 33, 35))
1936
1937 Residuals:
1938      Min       1Q   Median       3Q      Max
1939 -14.161  -4.120  -2.137   1.415   59.386
1940
1941 Coefficients:
1942             Estimate Std. Error t value Pr(>|t|)
1943 dist    5.762763    0.779834    7.390 3.84e-08 ***
1944 climb  0.005645    0.002886    1.956  0.0602 .
1945 ---
1946 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1947
1948 Residual standard error: 12.29 on 29 degrees of freedom
```

Page 25 of 45

```
1951 Multiple R-Squared: 0.9364,    Adjusted R-squared: 0.932
1952 F-statistic: 213.6 on 2 and 29 DF,  p-value: < 2.2e-16
1953
1954 >
1955 > # -----
1956 >
1957 > # y = -18.46 + 0.155*x4 - 0.209*x3
1958 >
1959 > fit131<-update(fit13, . ~ .+x1)
1960 > fit132<-update(fit13, . ~ .+x2)
1961 >
1962 > anova(fit13,fit131)
1963 Analysis of Variance Table
1964
1965 Model 1: y ~ x4 + x3
1966 Model 2: y ~ x4 + x3 + x1
1967      Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1968 1         29  170.61
1969 2         28  146.00    1    24.61 4.7198 0.03844 *
1970 ---
1971 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1972 > anova(fit13,fit132)
1973 Analysis of Variance Table
1974
1975 Model 1: y ~ x4 + x3
1976 Model 2: y ~ x4 + x3 + x2
1977      Res.Df    RSS Df Sum of Sq    F    Pr(>F)
1978 1         29  170.612
1979 2         28  160.620    1    9.992 1.7419 0.1976
1980 >
1981 > # scelgo la fit131
1982 >
1983 > summary(fit131)
1984
1985 Call:
1986 lm(formula = y ~ x4 + x3 + x1)
1987
1988 Residuals:
1989      Min       1Q   Median       3Q      Max
1990 -3.5303  -1.3606  -0.2681   1.3911   4.7658
1991
1992 Coefficients:
1993             Estimate Std. Error t value Pr(>|t|)
1994 (Intercept)  4.032034    7.223341   0.558   0.5811
1995 x4           0.156527    0.006462  24.224 < 2e-16 ***
1996 x3          -0.186571    0.015922 -11.718 2.61e-12 ***
1997 x1           0.221727    0.102061   2.173   0.0384 *
1998 ---
1999 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2000
2001 Residual standard error: 2.283 on 28 degrees of freedom
2002 Multiple R-Squared: 0.959,    Adjusted R-squared: 0.9546
2003 F-statistic: 218.5 on 3 and 28 DF,  p-value: < 2.2e-16
2004
2005 >
2006 > # -----
2007 >
2008 > fit131bis<-update(fit131, . ~ .+1)
2009 > summary(fit131bis)
2010
2011 Call:
2012 lm(formula = y ~ x4 + x3 + x1 - 1)
2013
2014 Residuals:
2015      Min       1Q   Median       3Q      Max
2016 -3.6075  -1.3229  -0.3831   1.7549  4.9115
2017
2018 Coefficients:
2019             Estimate Std. Error t value Pr(>|t|)
2020 x4    0.157168    0.006283   25.017 < 2e-16 ***
2021 x3   -0.179328    0.009116 -19.672 < 2e-16 ***
2022 x1   0.274133    0.039548   6.932 1.28e-07 ***
2023 ---
2024 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2025
2026 Residual standard error: 2.256 on 29 degrees of freedom
2027 Multiple R-Squared: 0.9907,    Adjusted R-squared: 0.9898
2028 F-statistic: 1034 on 3 and 29 DF,  p-value: < 2.2e-16
```

Page 26 of 45

```
2029
2030 > # y = 0.15*x4 - 0.17*x3 + 0.27*x1
2031 >
2032 >
2033 >
2034 > # -----
2035 > # Altro metodo Hyper veloce!!
2036 > # y = alpha + b1*x1 + b2*x2 + b3*x3 + b4*x4
2037 >
2038 > fit<-lm(y~x1+x2+x3+x4)
2039 > summary(fit)
2040
2041 Call:
2042 lm(formula = y ~ x1 + x2 + x3 + x4)
2043
2044 Residuals:
2045      Min       1Q   Median       3Q      Max
2046 -3.5804  -1.5223  -0.1098   1.4237   4.6214
2047
2048 Coefficients:
2049             Estimate Std. Error t value Pr(>|t|)
2050 (Intercept) -6.820774    10.123152  -0.674   0.5062
2051 x1           0.227246    0.099937   2.274   0.0311 *
2052 x2           0.553726    0.369752   1.498   0.1458
2053 x3          -0.149536    0.029229  -5.116 2.23e-05 ***
2054 x4           0.154650    0.006446  23.992 < 2e-16 ***
2055 ---
2056 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2057
2058 Residual standard error: 2.234 on 27 degrees of freedom
2059 Multiple R-Squared: 0.9622,    Adjusted R-squared: 0.9566
2060 F-statistic: 171.7 on 4 and 27 DF,  p-value: < 2.2e-16
2061
2062 >
2063 > fit1<-lm(y~x1+x2+x3+x4-1)
2064 > summary(fit1)
2065
2066 Call:
2067 lm(formula = y ~ x1 + x2 + x3 + x4 - 1)
2068
2069 Residuals:
2070      Min       1Q   Median       3Q      Max
2071 -3.6693  -1.2920  -0.1271   1.2348   4.5478
2072
2073 Coefficients:
2074             Estimate Std. Error t value Pr(>|t|)
2075 x1    0.182249    0.073618   2.476   0.0196 *
2076 x2    0.375377    0.255639   1.468   0.1531
2077 x3   -0.167498    0.012061 -13.882 4.45e-14 ***
2078 x4    0.154725    0.006382  24.245 < 2e-16 ***
2079 ---
2080 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2081
2082 Residual standard error: 2.213 on 28 degrees of freedom
2083 Multiple R-Squared: 0.9914,    Adjusted R-squared: 0.9902
2084 F-statistic: 806.6 on 4 and 28 DF,  p-value: < 2.2e-16
2085
2086 >
2087 > # --> Elimino la x2 perche' l'alpha oss e' il piu' grande ed e' > 0.05
2088 >
2089 > fit2<-lm(y~x1+x3+x4-1)
2090 > summary(fit2)
2091
2092 Call:
2093 lm(formula = y ~ x1 + x3 + x4 - 1)
2094
2095 Residuals:
2096      Min       1Q   Median       3Q      Max
2097 -3.6075  -1.3229  -0.3831   1.7549  4.9115
2098
2099 Coefficients:
2100             Estimate Std. Error t value Pr(>|t|)
2101 x1    0.274133    0.039548   6.932 1.28e-07 ***
2102 x3   -0.179328    0.009116 -19.672 < 2e-16 ***
2103 x4    0.157168    0.006283   25.017 < 2e-16 ***
2104 ---
2105 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2106
```

Page 27 of 45

```
2107 Residual standard error: 2.256 on 29 degrees of freedom
2108 Multiple R-Squared: 0.9907,    Adjusted R-squared: 0.9898
2109 F-statistic: 1034 on 3 and 29 DF,  p-value: < 2.2e-16
2110
2111 >
2112 > # ora il modello va bene, viva la Ventura!
2113 >
2114 > # analisi dei residui
2115 > res<-fit2$res
2116 > boxplot(res)
2117 > qqnorm(res);qqline(res)
2118 >
2119 > par(mfrow=c(2,2))
2120 > plot(res,x1)
2121 > plot(res,x2)
2122 > plot(res,x3)
2123 > plot(res,x4)
2124 >
2125 -----
2126 Laboratorio 08
2127 -----
2128 > # sturdy.dat
2129 > x1<-scan('sturdy.dat')
2130 Read 26 items
2131 > x1
2132 [1] 2.34 2.46 2.83 2.04 2.69 2.64 3.00 3.19 3.83 2.61 2.07 2.80 2.58 2.98 2.30
2133 [16] 1.32 1.62 1.92 0.88 1.50 1.30 0.41 0.83 0.58 0.32 1.62
2134 >
2135 > groupe<-c(rep("A",5),rep("B",4),rep("C",6),rep("D",6),rep("E",5))
2136 > groupe
2137 [1] "A" "A" "A" "A" "A" "B" "B" "B" "B" "C" "C" "C" "C" "C" "C" "D" "D" "D" "D" "D" "E" "E" "E" "E" "E"
2138 [20] "D" "D" "E" "E" "E" "E" "E" "E"
2139 >
2140 > groupe<-factor(groupe) # fatorizzazione dei gruppi
2141 > groupe
2142 [1] A A A A B B B B C C C C C D D D D D E E E E E
2143 Levels: A B C D E
2144 >
2145 > # ---- Esistono delle differenze tra le marche ???
2146 >
2147 > # t_i ~ N(mu_i,sigma^2_i)
2148 >
2149 > # normalita', varianze uguali
2150 > # H0: mu_i = mu_j
2151 >
2152 > stu<-data.frame(x1,groupe) # creazione del dataframe = read.table ('file.dat')
2153 > stu
2154      x1 groupe
2155 1 2.34      A
2156 2 2.46      A
2157 3 2.83      A
2158 4 2.04      A
2159 5 2.69      A
2160 6 2.64      B
2161 7 3.00      B
2162 8 3.19      B
2163 9 3.83      B
2164 10 2.61      C
2165 11 2.07      C
2166 12 2.80      C
2167 13 2.58      C
2168 14 2.98      C
2169 15 2.30      C
2170 16 1.32      D
2171 17 1.62      D
2172 18 1.92      D
2173 19 0.88      D
2174 20 1.50      D
2175 21 1.30      D
2176 22 0.41      E
2177 23 0.83      E
2178 24 0.58      E
2179 25 0.32      E
2180 26 1.62      E
2181
2182 > attach(stu)
2183 > # analisi preliminare
2184 > plot(x1-groupe)
```

Page 28 of 45

```

2185 > # --> normalita' ok! (abbiamo pero' pochi dati)
2186 > # boxplot
2187 > var(xl[group=="A"]) # seleziono i valori di xl = 'A'
2188 [1] 0.09497
2189 > var(xl[group=="B"])
2190 [1] 0.2485667
2191 > var(xl[group=="C"])
2192 [1] 0.1089067
2193 > var(xl[group=="D"])
2194 [1] 0.1224667
2195 > var(xl[group=="E"])
2196 [1] 0.27317
2197 >
2198 > # ---- test TRV per verificare l'omoschedasticita'
2199 > # '--> test di Bartlett
2200 > help(bartlett.test)
2201 >
2202 > # '--> l'ipotesi di omoschedasticita' puo' essere assunta!
2203 >
2204 >
2205 > stu.lme<-lm(xl~group,x=T)
2206 > stu.lmGx
2207 (Intercept) groupB groupC groupD groupE
2208 1 1 0 0 0 0
2209 2 1 0 0 0 0
2210 3 1 0 0 0 0
2211 4 1 0 0 0 0
2212 5 1 0 0 0 0
2213 6 1 1 0 0 0
2214 7 1 1 0 0 0
2215 8 1 1 0 0 0
2216 9 1 1 0 0 0
2217 10 1 0 1 0 0
2218 11 1 0 1 0 0
2219 12 1 0 1 0 0
2220 13 1 0 1 0 0
2221 14 1 0 1 0 0
2222 15 1 0 1 0 0
2223 16 1 0 0 1 0
2224 17 1 0 0 1 0
2225 18 1 0 0 1 0
2226 19 1 0 0 1 0
2227 20 1 0 0 1 0
2228 21 1 0 0 1 0
2229 22 1 0 0 0 1
2230 23 1 0 0 0 1
2231 24 1 0 0 0 1
2232 25 1 0 0 0 1
2233 26 1 0 0 0 1
2234 attr(,"assign")
2235 [1] 0 1 1 1 1
2236 attr(,"contrasts")
2237 attr(,"contrasts")$group
2238 [1] "contr.treatment"
2239
2240 >
2241 > # x_i = alpha + b1*d1 + b2*d2 + b3*d3 + b4*d4
2242 > # alpha : E(A)
2243 > # alpha + b1 : E(B)
2244 > # alpha + b2 : E(C)
2245 > # alpha + b3 : E(D)
2246 > # alpha + b4 : E(E)
2247 >
2248 > summary(stu.lm)
2249
2250 Call:
2251 lm(formula = xl ~ group, x = T)
2252
2253 Residuals:
2254 Min 1Q Median 3Q Max
2255 -0.543333 -0.235500 0.005667 0.212667 0.868000
2256
2257 Coefficients:
2258 (Intercept) 2.47200 0.17929 13.788 5.40e-12 ***
2259 groupB 0.69300 0.26893 2.577 0.017583 *
2260 groupC 0.08467 0.24276 0.349 0.730733
2261 groupD -1.04867 0.24276 -4.320 0.000302 ***

```

Page 29 of 45

```

2263 groupE -1.72000 0.25355 -6.784 1.04e-06 ***
2264 ---
2265 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2266
2267 Residual standard error: 0.4009 on 21 degrees of freedom
2268 Multiple R-Squared: 0.8433, Adjusted R-squared: 0.8135
2269 F-statistic: 28.26 on 4 and 21 DF, p-value: 3.475e-08
2270
2271 >
2272 > # E(A)=2.47
2273 > # E(B)=2.47+0.69
2274 > # E(C)=E(A)+2.47 # perche' il coeff non e' significativo
2275 > # E(D)=2.47-1.04
2276 > # E(E)=2.47-1.72
2277 >
2278 > # statistica che verifica l'ipotesi H0: b1=b2=b3=b4=0
2279 > stu.aov<-aov(xl~group)
2280 > summary(stu.aov)
2281 Df Sum Sq Mean Sq F value Pr(>F)
2282 group 4 18.1681 4.5420 28.261 3.475e-08 ***
2283 Residuals 21 3.3751 0.1607
2284 ---
2285 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2286 >
2287 > detach()
2288 > # -----
2289 > grnc<-factor(c(rep("G1",15),rep("G2",11)))
2290 > grn
2291 [1] G1 G1 G1 G1 G1 G1 G1 G1 G1 G1 G1 G1 G1 G2 G2 G2 G2 G2 G2 G2
2292 [26] G2
2293 Levels: G1 G2
2294 > stul<-data.frame(xl,grn)
2295 > attach(stul)
2296 > plot(xl~grn)
2297 >
2298 > # test? --> t di student
2299 > # ipotesi: normalita', omoschedasticita'
2300 > qqnorm(xl[grn=="G1"]);qqline(xl[grn=="G1"])
2301 > qqnorm(xl[grn=="G2"]);qqline(xl[grn=="G2"])
2302 >
2303 > var(xl[grn=="G1"])
2304 [1] 0.2083495
2305 > var(xl[grn=="G2"])
2306 [1] 0.2934164
2307 >
2308 > var.test(xl[grn=="G1"],xl[grn=="G2"])
2309
2310 F test to compare two variances
2311
2312 data: xl[grn == "G1"] and xl[grn == "G2"]
2313 F = 0.7101, num df = 14, denom df = 10, p-value = 0.5422
2314 alternative hypothesis: true ratio of variances is not equal to 1
2315 95 percent confidence interval:
2316 0.1999999 2.2345278
2317 sample estimates:
2318 ratio of variances
2319 0.7100815
2320
2321 > t.test(xl[grn=="G1"],xl[grn=="G2"],var.equal=T,alternative="g")
2322
2323 Two Sample t-test
2324
2325 data: xl[grn == "G1"] and xl[grn == "G2"]
2326 t = 8.0229, df = 24, p-value = 1.500e-08
2327 alternative hypothesis: true difference in means is greater than 0
2328 95 percent confidence interval:
2329 1.237152 Inf
2330 sample estimates:
2331 mean of x mean of y
2332 2.690667 1.118182
2333
2334 >
2335 > detach()
2336 >
2337 > # -----
2338 >
2339 > topi<-read.table('rats.dat',header=T)
2340 > topi

```

Page 30 of 45

```

2341 tempo veleno trattamento
2342 1 0.31 I A
2343 2 0.82 I B
2344 3 0.43 I C
2345 4 0.45 I D
2346 5 0.45 I A
2347 6 1.10 I B
2348 7 0.45 I C
2349 8 0.71 I D
2350 9 0.46 A I
2351 10 0.88 I B
2352 11 0.63 I C
2353 12 0.66 I D
2354 13 0.43 I A
2355 14 0.72 I B
2356 15 0.76 I C
2357 16 0.62 I D
2358 17 0.36 II A
2359 18 0.92 II B
2360 19 0.44 II C
2361 20 0.56 II D
2362 21 0.29 II A
2363 22 0.61 II B
2364 23 0.35 II C
2365 24 1.02 II D
2366 25 0.40 II A
2367 26 0.49 II B
2368 27 0.31 II C
2369 28 0.71 II D
2370 29 0.23 II A
2371 30 1.24 II B
2372 31 0.40 II C
2373 32 0.38 II D
2374 33 0.22 III A
2375 34 0.30 III B
2376 35 0.23 III C
2377 36 0.30 III D
2378 37 0.21 III A
2379 38 0.37 III B
2380 39 0.25 III C
2381 40 0.36 III D
2382 41 0.18 III A
2383 42 0.38 III B
2384 43 0.24 III C
2385 44 0.31 III D
2386 45 0.23 III A
2387 46 0.29 III B
2388 47 0.22 III C
2389 48 0.33 III D
2390 >
2391 > topi$veleno<-factor(topi$veleno)
2392 > topi$tratt<-factor(topi$tratt)
2393 > attach(topi)
2394 >
2395 > # effetto del trattamento sul veleno II # W la Ventura
2396 > topi2<-data.frame(tempo=tempi$veleno=="II",trattamento=trattamento[veleno=="II"])
2397 > attach(topi2)
2398 >
2399 > plot(tempo,trattamento)
2400 > plot(tempo~trattamento)
2401 >
2402 > bartlett.test(tempo,trattamento)
2403
2404 Bartlett test for homogeneity of variances
2405
2406 data: tempo and trattamento
2407 Bartlett's K-squared = 9.5432, df = 3, p-value = 0.02288
2408
2409 >
2410 > # ipotesi non soddisfatte: cosa fare????
2411 > # '--> trasformato i dati con la potenza della Ventura
2412 > # log(tempo), exp(tempo), tempo^p, tempo^-p
2413 >
2414 > plot(1/tempo~trattamento)
2415 > bartlett.test(1/tempo~trattamento)
2416
2417 Bartlett test for homogeneity of variances
2418

```

Page 31 of 45

```

2419 data: 1/tempo by trattamento
2420 Bartlett's K-squared = 1.2807, df = 3, p-value = 0.7337
2421
2422 > # ---- ipotesi di normalita' soddisfatta
2423 > # '--> anche i qqnorm
2424 >
2425 > topi.aov<-aov(1/tempo~trattamento)
2426 > summary(topi.aov)
2427 Df Sum Sq Mean Sq F value Pr(>F)
2428 trattamento 3 9.1424 3.0475 7.3913 0.004594 **
2429 Residuals 12 4.9477 0.4123
2430 ---
2431 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2432 >
2433 > # H0: mu1=mu2=mu3=mu4 , alpha_oss<0.05 --> rifiuto H0
2434 >
2435 > detach()
2436 >
2437 >
2438 > mor<-read.table('Morley.dat',header=T)
2439 > mor$Expt<-factor(mor$Expt)
2440 > attach(mor)
2441 >
2442 > plot(Speed~Expt)
2443 > # c'e' un comportamento anomalo nel gruppo 3, medie simili
2444 >
2445 > # verifica della normalita'
2446 > par(mfrow=c(2,3))
2447 > qqnorm(Speed[Expt==1]);qqline(Speed[Expt==1])
2448 > qqnorm(Speed[Expt==2]);qqline(Speed[Expt==2])
2449 > qqnorm(Speed[Expt==3]);qqline(Speed[Expt==3])
2450 > qqnorm(Speed[Expt==4]);qqline(Speed[Expt==4])
2451 > qqnorm(Speed[Expt==5]);qqline(Speed[Expt==5])
2452 > # risultati poco soddisfacenti
2453 >
2454 > bartlett.test(Speed,Expt)
2455
2456 Bartlett test for homogeneity of variances
2457
2458 data: Speed and Expt
2459 Bartlett's K-squared = 11.5518, df = 4, p-value = 0.02102
2460
2461 > bartlett.test(1/Speed,Expt)
2462
2463 Bartlett test for homogeneity of variances
2464
2465 data: 1/Speed and Expt
2466 Bartlett's K-squared = 13.0758, df = 4, p-value = 0.01091
2467
2468 > bartlett.test(log(Speed),Expt)
2469
2470 Bartlett test for homogeneity of variances
2471
2472 data: log(Speed) and Expt
2473 Bartlett's K-squared = 11.7615, df = 4, p-value = 0.01922
2474
2475 > bartlett.test(Speed^2,Expt)
2476
2477 Bartlett test for homogeneity of variances
2478
2479 data: Speed^2 and Expt
2480 Bartlett's K-squared = 12.3319, df = 4, p-value = 0.01505
2481
2482 >
2483 > # non trovo una trasformata adatta!!
2484 > # si usa allora (grazie alla Ventura che ce lo ha detto!) un test non
2485 > # parametrico per il confronto delle medie
2486 > # '--> non assumo nessuna distrib per i dati
2487 > # ===> test di Kruskal Wallis
2488 >
2489 > kruskal.test(Speed,Expt)
2490
2491 Kruskal-Wallis rank sum test
2492
2493 data: Speed and Expt
2494 Kruskal-Wallis chi-squared = 15.0221, df = 4, p-value = 0.004656
2495
2496 >

```

Page 32 of 45


```
2497 -----
2498
2499 Riassunto:
2500 * Controllo le ipotesi: di normalita' e omoschedasticita'
2501 * Se sono verificate faccio i test altrimenti provo delle trasformate
2502 * se funziona e' tutto ok e faccio i test, altrimenti uso il test Kruskal
2503 -----
2504 Laboratorio 09
2505 -----
2506
2507 > ---- penicillin.dat
2508 > pen<-read.table("penicillin.dat",header=T)
2509 > pen
2510 miscela modo penicillina
2511 1 I A 89
2512 2 I B 88
2513 3 I C 97
2514 4 I D 94
2515 5 II A 84
2516 6 II B 77
2517 7 II C 92
2518 8 II D 79
2519 9 III A 81
2520 10 III B 87
2521 11 III C 87
2522 12 III D 85
2523 13 IV A 87
2524 14 IV B 92
2525 15 IV C 89
2526 16 IV D 84
2527 17 V A 79
2528 18 V B 81
2529 19 V C 80
2530 20 V D 88
2531 > # disegno fatto con 5x4 livelli complessivi, con 1 replicaz
2532 > attach(pen)
2533 > miscela
2534 [1] I I I I I II II II II III III III III IV IV IV V V V V
2535 [20] V
2536 Levels: I II III IV V
2537 > modo
2538 [1] A B C D A B C D A B C D A B C D A B C D
2539 Levels: A B C D
2540 > is.factor(miscela)
2541 [1] TRUE
2542 > is.factor(modo)
2543 [1] TRUE
2544 > # penicillina = miscela + errore
2545 > # penicillina = modo + errore
2546 > # penicillina = miscela + modo + errore
2547 >
2548 > # anova 1 fattore
2549 > # analisi preliminare
2550 > plot(penicillina-modo)
2551 >
2552 > # penicillina = modo + errore
2553 > # errore ~ N(0, sigma^2)
2554 >
2555 > # H0: mu_1 = mu_2 = mu_3 = mu_4
2556 >
2557 > bartlett.test(penicillina,modo)
2558
2559 Bartlett test for homogeneity of variances
2560
2561 data: penicillina and modo
2562 Bartlett's K-squared = 0.6901, df = 3, p-value = 0.8755
2563
2564 > # H0: mu_1 = mu_2 = mu_3 = mu_4
2565 >
2566 > # TRV funzione monotona crescente F
2567 > # F = [SS(tra i campioni)/(J-1)]/[SS(entro i campioni)/(n-J)] # SS e' la somma
2568 > # dei quadrati
2569 > # F - F_(J-1), (n-J) # F di Schnedecor
2570 >
2571 > # aov
2572 > summary(aov(penicillina-modo))
2573 Df Sum Sq Mean Sq F value Pr(>F)
2574 modo 3 70.00 23.33 0.7619 0.5318
```

Page 33 of 45

```
2575 Residuals 16 490.00 30.62
2576 >
2577 > # penicillina = modo + errore
2578 > # accettiamo H0
2579 >
2580 > # pen_ij = cost + alpha_j + err_ij
2581 >
2582 > fit<-lm(penicillina-modo,x=T)
2583 > fit$z
2584 (Intercept) modoB modoC modoD
2585 1 1 0 0 0
2586 2 1 1 0 0
2587 3 1 0 0 0
2588 4 1 0 0 1
2589 5 1 0 0 0
2590 6 1 1 0 0
2591 7 1 0 1 0
2592 8 1 0 0 1
2593 9 1 0 0 0
2594 10 1 1 0 0
2595 11 1 0 1 0
2596 12 1 0 0 0
2597 13 1 0 0 0
2598 14 1 1 0 0
2599 15 1 0 1 0
2600 16 1 0 0 1
2601 17 1 0 0 0
2602 18 1 1 0 0
2603 19 1 0 1 0
2604 20 1 0 0 1
2605 attr(,"assign")
2606 [1] 0 1 1 1
2607 attr(,"contrasts")
2608 attr(,"contrasts")$modo
2609 [1] "contr.treatment"
2610
2611 >
2612 > summary(fit)
2613
2614 Call:
2615 lm(formula = penicillina ~ modo, x = T)
2616
2617 Residuals:
2618 1Q Median 3Q Max
2619 -9.000e+00 -3.250e+00 3.002e-15 3.000e+00 8.000e+00
2620
2621 Coefficients:
2622 (Intercept) 84.000 2.475 33.941 2.45e-16 ***
2623 modoB 1.000 3.500 0.286 0.779
2624 modoC 5.000 3.500 1.429 0.172
2625 modoD 2.000 3.500 0.571 0.576
2626 ---
2627 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2628
2629 Residual standard error: 5.534 on 16 degrees of freedom
2630 Multiple R-squared: 0.125, Adjusted R-squared: -0.03906
2631 F-statistic: 0.7619 on 3 and 16 DF, p-value: 0.5318
2632
2633 > # H0: alpha_1 = alpha_2 = alpha_3 = 0
2634 >
2635 > # confronto tra i due modelli di regressione
2636 > anova(fit)
2637 Analysis of Variance Table
2638
2639 Response: penicillina
2640 Df Sum Sq Mean Sq F value Pr(>F)
2641 modo 3 70.00 23.33 0.7619 0.5318
2642 Residuals 16 490.00 30.62
2643 > # '--> stesso risultato di prima!
2644 >
2645 > plot(penicillina-miscela)
2646 > bartlett.test(penicillina,miscela)
2647
2648 Bartlett test for homogeneity of variances
2649
2650 data: penicillina and miscela
2651 Bartlett's K-squared = 2.3859, df = 4, p-value = 0.6652
2652
```

Page 34 of 45

```
2653 > # p-value elevato --> accetto l'ipotesi di omoschedasticita'
2654 > summary(aov(penicillina-miscela))
2655 Df Sum Sq Mean Sq F value Pr(>F)
2656 miscela 4 264.000 66.000 3.3446 0.03801 *
2657 Residuals 15 296.000 19.733
2658 ---
2659 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2660 >
2661 > # ----
2662 > # pen_(ijk) = cost + alpha_j + beta_k + errore_(ijk)
2663 >
2664 > summary(aov(penicillina-modo+miscela))
2665 Df Sum Sq Mean Sq F value Pr(>F)
2666 modo 3 70.000 23.333 1.2389 0.33866
2667 miscela 4 264.000 66.000 3.5044 0.04075 *
2668 Residuals 12 226.000 18.833
2669 ---
2670 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2671 > # e' come se avessimo fatto una scomposizione del tipo:
2672 > # Var(Y) = Var(modo) + Var(miscela) + Var(residua)
2673 >
2674 > fit1<-lm(penicillina-modo+miscela,x=T)
2675 > fit$z
2676 (Intercept) modoB modoC modoD
2677 1 1 0 0 0
2678 2 1 1 0 0
2679 3 1 0 1 0
2680 4 1 0 0 1
2681 5 1 0 0 0
2682 6 1 1 0 0
2683 7 1 0 1 0
2684 8 1 0 0 1
2685 9 1 0 0 0
2686 10 1 1 0 0
2687 11 1 0 1 0
2688 12 1 0 0 1
2689 13 1 0 0 0
2690 14 1 1 0 0
2691 15 1 0 1 0
2692 16 1 0 0 1
2693 17 1 0 0 0
2694 18 1 1 0 0
2695 19 1 0 1 0
2696 20 1 0 0 1
2697 attr(,"assign")
2698 [1] 0 1 1 1
2699 attr(,"contrasts")
2700 attr(,"contrasts")$modo
2701 [1] "contr.treatment"
2702
2703 >
2704 > summary(fit1)
2705
2706 Call:
2707 lm(formula = penicillina ~ modo + miscela, x = T)
2708
2709 Residuals:
2710 1Q Median 3Q Max
2711 -5.00 -2.25 -0.50 2.25 6.00
2712
2713 Coefficients:
2714 (Intercept) 90.000 2.745 32.791 4.1e-13 ***
2715 modoB 1.000 2.745 0.364 0.72194
2716 modoC 5.000 2.745 1.822 0.09351 .
2717 modoD 2.000 2.745 0.729 0.48018
2718 miscelaII -9.000 3.069 -2.933 0.01254 *
2719 miscelaIII -7.000 3.069 -2.281 0.04159 *
2720 miscelaIV -4.000 3.069 -1.304 0.21686
2721 miscelaV -10.000 3.069 -3.259 0.00684 **
2722 ---
2723 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2724
2725 Residual standard error: 4.34 on 12 degrees of freedom
2726 Multiple R-Squared: 0.5964, Adjusted R-squared: 0.361
2727 F-statistic: 2.534 on 7 and 12 DF, p-value: 0.07535
2728
2729 > # il modo non influenza i risultati
```

Page 35 of 45

```
2731 > # miscela I : Y=90
2732 > # miscela II : Y=90-9=81
2733 > # miscela III : Y=90-7=83
2734 > # miscela IV : Y=90-90 # il -4 non e' significativo!
2735 > # miscela V : Y=90-10=80
2736 >
2737 > # H0: alpha=beta=0
2738 >
2739 > anova(fit1)
2740 Analysis of Variance Table
2741
2742 Response: penicillina
2743 Df Sum Sq Mean Sq F value Pr(>F)
2744 modo 3 70.000 23.333 1.2389 0.33866
2745 miscela 4 264.000 66.000 3.5044 0.04075 *
2746 Residuals 12 226.000 18.833
2747 ---
2748 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2749 >
2750 > # ----
2751 > # la F sopra, deriva dal seguente test:
2752 > # F = [(-sigma - ^sigma)/(p-p0)] / [^sigma/(n-p)]
2753 >
2754 > # Var(modo)
2755 > # Y = cost + alpha + beta + errore
2756 > # H0: alpha = 0
2757 > # SQ(modo)
2758 >
2759 > # H0: beta = 0
2760 >
2761 > detach()
2762 > # ----
2763 >
2764 > topi<-read.table("rats.dat",header=T)
2765 > topi
2766 tempo veleno trattamento
2767 1 0.31 I A
2768 2 0.82 I B
2769 3 0.43 I C
2770 4 0.45 I D
2771 5 0.45 I A
2772 6 1.10 I B
2773 7 0.45 I C
2774 8 0.71 I D
2775 9 0.46 I A
2776 10 0.88 I B
2777 11 0.63 I C
2778 12 0.66 I D
2779 13 0.43 I A
2780 14 0.72 I B
2781 15 0.76 I C
2782 16 0.62 I D
2783 17 0.36 I A
2784 18 0.32 II B
2785 19 0.44 II C
2786 20 0.56 II D
2787 21 0.29 II A
2788 22 0.61 II B
2789 23 0.35 II C
2790 24 1.02 II D
2791 25 0.40 II A
2792 26 0.49 II B
2793 27 0.31 II C
2794 28 0.71 II D
2795 29 0.23 II A
2796 30 1.24 II B
2797 31 0.40 II C
2798 32 0.38 II D
2799 33 0.22 III A
2800 34 0.30 III B
2801 35 0.23 III C
2802 36 0.30 III D
2803 37 0.21 III A
2804 38 0.37 III B
2805 39 0.25 III C
2806 40 0.36 III D
2807 41 0.18 III A
2808 42 0.38 III B
```

Page 36 of 45

```
2809 43 0.24 III C
2810 44 0.31 III D
2811 45 0.23 III A
2812 46 0.29 III B
2813 47 0.22 III C
2814 48 0.33 III D
2815 > # disegno fattoriale 3x4, con 4 replicazioni
2816 >
2817 > attach(topi)
2818 > is.factor(veleno)
2819 [1] TRUE
2820 > is.factor(trattamento)
2821 [1] TRUE
2822 >
2823 > # -----
2824 > # analisi preliminare
2825 > plot(tempo-veleno+trattamento)
2826 Hit <Return> to see next plot:
2827 Hit <Return> to see next plot:
2828 >
2829 > help(interaction.plot)
2830 > interaction.plot(veleno, trattamento, tempo)
2831 > interaction.plot(trattamento, veleno, tempo)
2832 >
2833 > # -----
2834 > # Risp = Effetto(costante) + Effetto(trattamento) + Effetto(veleno) +
2835 > # Effetto(trattamento*veleno) + Effetto(residuo)
2836 > summary(aov(tempo-trattamento + veleno + trattamento:veleno))
2837
Df Sum Sq Mean Sq F value Pr(>F)
2838 trattamento 3 0.92121 0.30707 13.8056 3.777e-06 ***
2839 veleno 2 1.03301 0.51651 23.2217 3.331e-07 ***
2840 trattamento:veleno 6 0.25014 0.04169 1.8743 0.1123
2841 Residuals 36 0.80073 0.02224
2842 ---
2843 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2844 > # lo stesso e' con:
2845 > summary(aov(tempo-trattamento * veleno))
2846
Df Sum Sq Mean Sq F value Pr(>F)
2847 trattamento 3 0.92121 0.30707 13.8056 3.777e-06 ***
2848 veleno 2 1.03301 0.51651 23.2217 3.331e-07 ***
2849 trattamento:veleno 6 0.25014 0.04169 1.8743 0.1123
2850 Residuals 36 0.80073 0.02224
2851 ---
2852 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2853 >
2854 > # H0: trattamento non ha effetto
2855 > # y_ijk ~ N(mu_ijk, sigma^2)
2856 > # H0: mu1_jk = mu2_jk = mu3_jk
2857 > # i: trattamento # j: veleno # k: interazione tra trattamento e veleno
2858 >
2859 > # y_ijk = mu + alpha_i + beta_j + delta_ij + epsilon
2860 >
2861 > gc<-lm(tempo-trattamento * veleno)
2862 > summary(gc)
2863
Call:
lm(formula = tempo ~ trattamento * veleno)
2864
Residuals:
2865 Min 1Q Median 3Q Max
2866 -0.32500 -0.04875 0.00500 0.04312 0.42500
2867
Coefficients:
2868 (Intercept) 0.41250 0.07457 5.532 2.94e-06 ***
2869 trattamentoB 0.46750 0.10546 4.433 8.37e-05 ***
2870 trattamentoC 0.15500 0.10546 1.470 0.1503
2871 trattamentoD 0.19750 0.10546 1.873 0.0692 .
2872 velenoII -0.09250 0.10546 -0.877 0.3862
2873 velenoIII -0.20250 0.10546 -1.920 0.0628 .
2874 trattamentoB:velenoII 0.02750 0.14914 0.184 0.8547
2875 trattamentoC:velenoII -0.10000 0.14914 -0.671 0.5068
2876 trattamentoD:velenoII 0.15000 0.14914 1.006 0.3212
2877 trattamentoB:velenoIII -0.34250 0.14914 -2.297 0.0276 *
2878 trattamentoC:velenoIII -0.13000 0.14914 -0.872 0.3892
2879 trattamentoD:velenoIII -0.08250 0.14914 -0.553 0.5836
2880 ---
2881 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
2887
2888 Residual standard error: 0.1491 on 36 degrees of freedom
2889 Multiple R-Squared: 0.7335, Adjusted R-squared: 0.6521
2890 F-statistic: 9.01 on 11 and 36 DF, p-value: 1.986e-07
2891
2892 >
2893 > # il trattamento A e' quello di riferimento
2894 > # C e D si comportano come A --> non sono significativi
2895 > # il veleno significativo e' il II
2896 > # l'interazione significativa e' solo B:III
2897 > anova(g)
2898 Analysis of Variance Table
2899
Response: tempo
2900
Df Sum Sq Mean Sq F value Pr(>F)
2901 trattamento 3 0.92121 0.30707 13.8056 3.777e-06 ***
2902 veleno 2 1.03301 0.51651 23.2217 3.331e-07 ***
2903 trattamento:veleno 6 0.25014 0.04169 1.8743 0.1123
2904 Residuals 36 0.80073 0.02224
2905 ---
2906 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2907 >
2908 > # H0: alpha_i = 0
2909 > # H0: beta_i = 0
2910 > # H0: delta_ij = 0
2911 >
2912 >
2913 > # '--> y_ijk = mu + alpha_i + beta_j + epsilon
2914 > gc<-lm(tempo-trattamento+veleno)
2915 >
2916 > # A, I: Y = 0.45
2917 > # A, II: Y = 0.45 # xche' il veleno II non e' significativo
2918 > # A, III: Y = 0.45 - 0.34 = 0.11
2919 > # B, I: Y = 0.45 + 0.36 = 81
2920 > # ecc...
2921 >
2922 >
2923 > # -----
2924 > # analisi dei residui
2925 >
2926 > qqnorm(g$res);qqline(g$res)
2927 > plot(g$fitted,g$res)
2928 > # cosa fare wuando i residui non sono soddisfacenti???
2929 > # '--> riformulo il modello!
2930 >
2931 > gl<-lm(1/tempo-veleno * trattamento)
2932 > anova(gl)
2933 Analysis of Variance Table
2934
Response: 1/tempo
2935
Df Sum Sq Mean Sq F value Pr(>F)
2936 veleno 2 34.877 17.439 72.6347 2.310e-13 ***
2937 trattamento 3 20.414 6.805 28.3431 1.376e-09 ***
2938 veleno:trattamento 6 1.571 0.262 1.0904 0.3867
2939 Residuals 36 8.643 0.240
2940 ---
2941 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2942 >
2943 > qqnorm(gl$res);qqline(gl$res)
2944 > plot(gl$fitted,gl$res)
2945 >
2946 >
2947 -----
2948 Laboratorio 10
2949 -----
2950 > cats<-read.table('cats.dat',col.names=c('B','H','S'))
2951 > cats
2952
B H S
2953 1 2.3 9.6 1
2954 2 3.0 10.6 1
2955 3 2.9 9.9 1
2956 4 2.4 8.7 1
2957 5 2.3 10.1 1
2958 6 2.0 7.0 1
2959 7 2.2 11.0 1
2960 8 2.1 8.2 1
2961 9 2.3 9.0 1
2962 10 2.1 7.3 1
2963 11 2.1 8.5 1
2964 12 2.2 9.7 1
```

```
2965 13 2.0 7.4 1
2966 14 2.3 7.3 1
2967 15 2.2 7.1 1
2968 16 2.3 9.0 1
2969 17 2.1 7.6 1
2970 18 2.0 9.5 1
2971 19 2.9 10.1 1
2972 20 2.7 10.2 1
2973 21 2.6 10.1 1
2974 22 2.3 9.5 1
2975 23 2.6 8.7 1
2976 24 2.1 7.2 1
2977 25 2.9 9.4 2
2978 26 2.4 9.3 2
2979 27 2.2 7.2 2
2980 28 2.9 11.3 2
2981 29 2.5 8.8 2
2982 30 3.1 9.9 2
2983 31 3.0 13.3 2
2984 32 2.5 12.7 2
2985 33 3.4 14.4 2
2986 34 3.0 10.0 2
2987 35 2.6 10.5 2
2988 36 2.5 8.6 2
2989 37 2.8 10.0 2
2990 38 3.1 12.1 2
2991 39 3.0 13.8 2
2992 40 2.7 12.0 2
2993 41 2.8 12.0 2
2994 42 2.1 10.1 2
2995 43 3.3 11.5 2
2996 44 3.4 12.2 2
2997 45 2.8 13.5 2
2998 46 2.7 10.4 2
2999 47 3.2 11.6 2
3000 48 3.0 10.6 2
3001 > as.factor(cats$S)
3002 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2
3003 [39] 2 2 2 2 2 2 2 2 2 2 2
3004 Levels: 1 2
3005 > is.factor(cats$S)
3006 [1] FALSE
3007 > cats$S<-factor(cats$S)
3008 >
3009 > attach(cats)
3010 > # -----
3011 > # peso del cuore a seconda del sesso
3012 >
3013 > # analisi preliminare
3014 > plot(H-S)
3015 >
3016 > # confronto tra i due gruppi
3017 > # ipotesi: normalita
3018 > qqnorm(H[S==1]);qqline(H[S==1])
3019 > qqnorm(H[S==2]);qqline(H[S==2])
3020 >
3021 > # ipotesi: omoschedasticita'/variance
3022 > var(H[S==1])
3023 [1] 1.552446
3024 > var(H[S==2])
3025 [1] 3.234783
3026 > plot(H-S)
3027 >
3028 > # H0: sigma^2_1/sigma^2_2 = 1
3029 > # F = (s^2_1/(n1_1)) / (s^2_2/(n2_1))
3030 > var.test(H[S==1],H[S==2])
3031
3032 F test to compare two variances
3033
data: H[S == 1] and H[S == 2]
3034 F = 0.47989, num df = 23, denom df = 23, p-value = 0.08496
3035 alternative hypothesis: true ratio of variances is not equal to 1
3036 95 percent confidence interval:
3037 0.2076113 1.1094088
3038 sample estimates:
3039 ratio of variances
3040 0.4799227
3041
3042
```

```
3043 >
3044 > # t-Student
3045 > # H0: mu_1 = mu_2
3046 > # H1: mu_1 < mu_2
3047 >
3048 > t.test(H[S==1],H[S==2],alternative="less",var.equal=T)
3049
Two Sample t-test
3050
data: H[S == 1] and H[S == 2]
3051 t = -4.8419, df = 46, p-value = 7.455e-06
3052 alternative hypothesis: true difference in means is less than 0
3053 95 percent confidence interval:
3054 -Inf -1.412780
3055 sample estimates:
3056 mean of x mean of y
3057 8.8875 11.0500
3058
3059 > # -----
3060 > # mi aspetto che ci sia una relazione tra il peso ed il cuore
3061 > # proviamo:
3062 > plot(H-B)
3063 > cor(H,B)
3064 [1] 0.7335559
3065 >
3066 > # evidenziamo sul grafico i maschi e le femmine con colori diversi
3067 > points(B[S==1],H[S==1],col='red')
3068 > points(B[S==2],H[S==2],col='blue')
3069 >
3070 > # ---- H = alpha + alpha_j(S) + beta * B + beta_j * (S:B) + err
3071 > # S:B=interazione tra S e B
3072 >
3073 > # S = (1,0) # 1=F, 0=M
3074 > # (S:B) = (1,0)
3075 > #
3076 > # H_F = (alpha + alpha_1) + (beta+beta_1)*B
3077 > # H_M = alpha + beta*B
3078 >
3079 > # H0: alpha_1=beta_1=0
3080 > # F = ((-sigma^2 - sigma^2)/(p-p0))/ (sigma^2/(n-p))
3081 >
3082 > fit<-lm(H-S + B + S:B,x=S)
3083 > fit$
3084 (Intercept) S2 B S2:B
3085 1 0 2.3 0.0
3086 2 1 0 3.0 0.0
3087 3 1 0 2.9 0.0
3088 4 1 0 2.4 0.0
3089 5 1 0 2.3 0.0
3090 6 1 0 2.0 0.0
3091 7 1 0 2.2 0.0
3092 8 1 0 2.1 0.0
3093 9 1 0 2.3 0.0
3094 10 1 0 2.1 0.0
3095 11 1 0 2.1 0.0
3096 12 1 0 2.2 0.0
3097 13 1 0 2.0 0.0
3098 14 1 0 2.3 0.0
3099 15 1 0 2.2 0.0
3100 16 1 0 2.3 0.0
3101 17 1 0 2.1 0.0
3102 18 1 0 2.0 0.0
3103 19 1 0 2.9 0.0
3104 20 1 0 2.7 0.0
3105 21 1 0 2.6 0.0
3106 22 1 0 2.3 0.0
3107 23 1 0 2.6 0.0
3108 24 1 0 2.1 0.0
3109 25 1 1 2.9 2.9
3110 26 1 1 2.4 2.4
3111 27 1 1 2.2 2.2
3112 28 1 1 2.9 2.9
3113 29 1 1 2.5 2.5
3114 30 1 1 3.1 3.1
3115 31 1 1 3.0 3.0
3116 32 1 1 2.5 2.5
3117 33 1 1 3.4 3.4
3118 34 1 1 3.0 3.0
```

```
3121 35      1  1 2.6 2.6
3122 36      1  1 2.5 2.5
3123 37      1  1 2.8 2.8
3124 38      1  1 3.1 3.1
3125 39      1  1 3.0 3.0
3126 40      1  1 2.7 2.7
3127 41      1  1 2.8 2.8
3128 42      1  1 2.1 2.1
3129 43      1  1 3.3 3.3
3130 44      1  1 3.4 3.4
3131 45      1  1 2.8 2.8
3132 46      1  1 2.7 2.7
3133 47      1  1 3.2 3.2
3134 48      1  1 3.0 3.0
3135 attr(,"assign")
3136 [1] 0 1 2 3
3137 attr(,"contrasts")
3138 attr(,"contrasts")$S
3139 [1] "contr.treatment"
3140
3141 >
3142 > # ----- H = alpha + alpha_j(S) + beta * B + beta_j * (S:B) + err
3143 > # H_F = alpha + beta * B
3144 > # H_M = alpha + alpha_1*I + beta * B + beta_1*B
3145 > summary(fit)
3146
3147 Call:
3148 lm(formula = H ~ S + B + S:B, x = T)
3149
3150 Residuals:
3151      Min       1Q   Median       3Q      Max
3152 -1.9813 -0.9589 -0.1629  0.8573  2.6277
3153
3154 Coefficients:
3155             Estimate Std. Error t value Pr(>|t|)
3156 (Intercept)  2.9318     2.1105   1.389  0.17178
3157 S2           -0.2849     3.0313  -0.094  0.92554
3158 B            2.5525     0.8975   2.844  0.00674 **
3159 S2:B         0.4177     1.1784   0.354  0.72466
3160 ---
3161 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3162
3163 Residual standard error: 1.28 on 44 degrees of freedom
3164 Multiple R-Squared:  0.5664, Adjusted R-squared:  0.5368
3165 F-statistic: 19.16 on 3 and 44 DF, p-value: 4.269e-08
3166
3167 > fit<-lm(H~S * B,x=T) # stessa cosa
3168 > summary(fit)
3169
3170 Call:
3171 lm(formula = H ~ S * B, x = T)
3172
3173 Residuals:
3174      Min       1Q   Median       3Q      Max
3175 -1.9813 -0.9589 -0.1629  0.8573  2.6277
3176
3177 Coefficients:
3178             Estimate Std. Error t value Pr(>|t|)
3179 (Intercept)  2.9318     2.1105   1.389  0.17178
3180 S2           -0.2849     3.0313  -0.094  0.92554
3181 B            2.5525     0.8975   2.844  0.00674 **
3182 S2:B         0.4177     1.1784   0.354  0.72466
3183 ---
3184 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3185
3186 Residual standard error: 1.28 on 44 degrees of freedom
3187 Multiple R-Squared:  0.5664, Adjusted R-squared:  0.5368
3188 F-statistic: 19.16 on 3 and 44 DF, p-value: 4.269e-08
3189
3190 > # '----> H = 2.55*B
3191 > fit1<-lm(H~B ~ 1)
3192 > summary(fit1)
3193
3194 Call:
3195 lm(formula = H ~ B ~ 1)
3196
3197 Residuals:
3198      Min       1Q   Median       3Q      Max
```

Page 41 of 45

```
3199 -2.0372 -0.9521 -0.0969  0.8620  3.0732
3200
3201 Coefficients:
3202 Estimate Std. Error t value Pr(>|t|)
3203 B  3.8507     0.0714   53.93 <2e-16 ***
3204 ---
3205 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3206
3207 Residual standard error: 1.292 on 47 degrees of freedom
3208 Multiple R-Squared:  0.9841, Adjusted R-squared:  0.9838
3209 F-statistic: 2909 on 1 and 47 DF, p-value: < 2.2e-16
3210
3211 > # '----> H = 3.85*B
3212 >
3213 > anova(fit1,fit)
3214 Analysis of Variance Table
3215
3216 Model 1: H ~ B ~ 1
3217 Model 2: H ~ S + B + S:B
3218      Res.Df  RSS Df Sum of Sq    F Pr(>F)
3219 1      47 78.488
3220 2      44 72.073  3    6.415 1.3055 0.2846
3221 >
3222 > qqnorm(fit1$res);qqline(fit1$res)
3223 > plot(fit1$fitted,fit1$res)
3224 >
3225 > detach()
3226 >
3227 > # -----
3228 > insu<-read.table('insulate.dat',col.names=c('quando','temp','cons'))
3229 > insu
3230      quando temp cons
3231 1      prima -0.8  7.2
3232 2      prima -0.7  6.9
3233 3      prima  0.4  6.4
3234 4      prima  2.5  6.0
3235 5      prima  2.9  5.8
3236 6      prima  3.2  5.8
3237 7      prima  3.6  5.6
3238 8      prima  3.9  4.7
3239 9      prima  4.2  5.8
3240 10     prima  4.3  5.2
3241 11     prima  5.4  4.9
3242 12     prima  6.0  4.9
3243 13     prima  6.0  4.3
3244 14     prima  6.0  4.4
3245 15     prima  6.2  4.5
3246 16     prima  6.3  4.6
3247 17     prima  6.9  3.7
3248 18     prima  7.0  3.9
3249 19     prima  7.4  4.2
3250 20     prima  7.5  4.0
3251 21     prima  7.5  3.9
3252 22     prima  7.6  3.5
3253 23     prima  8.0  4.0
3254 24     prima  8.5  3.6
3255 25     prima  9.1  3.1
3256 26     prima 10.2  2.6
3257 27     dopo -0.7  4.8
3258 28     dopo  0.8  4.6
3259 29     dopo  1.0  4.7
3260 30     dopo  1.4  4.0
3261 31     dopo  1.5  4.2
3262 32     dopo  1.6  4.2
3263 33     dopo  2.3  4.1
3264 34     dopo  2.5  4.0
3265 35     dopo  2.5  3.5
3266 36     dopo  3.1  3.2
3267 37     dopo  3.9  3.9
3268 38     dopo  4.0  3.5
3269 39     dopo  4.0  3.7
3270 40     dopo  4.2  3.5
3271 41     dopo  4.3  3.5
3272 42     dopo  4.6  3.7
3273 43     dopo  4.7  3.5
3274 44     dopo  4.9  3.4
3275 45     dopo  4.9  3.7
3276 46     dopo  4.9  4.0
```

Page 42 of 45

```
3277 47     dopo  5.0  3.6
3278 48     dopo  5.3  3.7
3279 49     dopo  6.2  2.8
3280 50     dopo  7.1  3.0
3281 51     dopo  7.2  2.8
3282 52     dopo  7.5  2.6
3283 53     dopo  8.0  2.7
3284 54     dopo  8.7  2.8
3285 55     dopo  8.8  1.3
3286 56     dopo  9.7  1.5
3287 >
3288 > is.factor(insu$quando)
3289 [1] TRUE
3290 > attach(insu)
3291 >
3292 > # analisi preliminare
3293 > plot(cons~temp)
3294 > cor(cons,temp)
3295 [1] -0.6832545
3296 >
3297 > # evidenziamo i punti a seconda del quando:
3298 > points(temp[quando=="prima"],cons[quando=="prima"],col='red')
3299 > points(temp[quando=="dopo"],cons[quando=="dopo"],col='blue')
3300 >
3301 > cor(temp[quando=="prima"],cons[quando=="prima"])
3302 [1] -0.9714978
3303 > cor(temp[quando=="dopo"],cons[quando=="dopo"])
3304 [1] -0.9017078
3305 >
3306 > # -----
3307 > # cons = alpha + alpha_1*quando + beta*temp + beta_1*(quando~temp) + err
3308 > fit<-lm(cons~quando * temp,x=T) # si considerano gli effetti marginali del
3309 >                                # quando e della temperatura
3310 > fit$x
3311      (Intercept) quandoprima temp quandoprima:temp
3312 1      1      1      1  -0.8      -0.8
3313 2      1      1      1  -0.7     -0.7
3314 3      1      1      1   0.4      0.4
3315 4      1      1      1   2.5      2.5
3316 5      1      1      1   2.9      2.9
3317 6      1      1      1   3.2      3.2
3318 7      1      1      1   3.6      3.6
3319 8      1      1      1   3.9      3.9
3320 9      1      1      1   4.2      4.2
3321 10     1      1      1   4.3      4.3
3322 11     1      1      1   5.4      5.4
3323 12     1      1      1   6.0      6.0
3324 13     1      1      1   6.0      6.0
3325 14     1      1      1   6.0      6.0
3326 15     1      1      1   6.2      6.2
3327 16     1      1      1   6.3      6.3
3328 17     1      1      1   6.9      6.9
3329 18     1      1      1   7.0      7.0
3330 19     1      1      1   7.4      7.4
3331 20     1      1      1   7.5      7.5
3332 21     1      1      1   7.5      7.5
3333 22     1      1      1   7.6      7.6
3334 23     1      1      1   8.0      8.0
3335 24     1      1      1   8.5      8.5
3336 25     1      1      1   9.1      9.1
3337 26     1      1      1  10.2     10.2
3338 27     1      1      1  -0.7     -0.7
3339 28     1      1      1   0.8      0.0
3340 29     1      1      1   1.0      0.0
3341 30     1      1      1   1.4      0.0
3342 31     1      1      1   1.5      0.0
3343 32     1      1      1   1.6      0.0
3344 33     1      1      1   2.3      0.0
3345 34     1      1      1   2.5      0.0
3346 35     1      1      1   2.5      0.0
3347 36     1      1      1   3.1      0.0
3348 37     1      1      1   3.9      0.0
3349 38     1      1      1   4.0      0.0
3350 39     1      1      1   4.0      0.0
3351 40     1      1      1   4.2      0.0
3352 41     1      1      1   4.3      0.0
3353 42     1      1      1   4.6      0.0
3354 43     1      1      1   4.7      0.0
```

Page 43 of 45

```
3355 44      1      0  4.9      0.0
3356 45      1      0  4.9      0.0
3357 46      1      0  4.9      0.0
3358 47      1      0  5.0      0.0
3359 48      1      0  5.3      0.0
3360 49      1      0  6.2      0.0
3361 50      1      0  7.1      0.0
3362 51      1      0  7.2      0.0
3363 52      1      0  7.5      0.0
3364 53      1      0  8.0      0.0
3365 54      1      0  8.7      0.0
3366 55      1      0  8.8      0.0
3367 56      1      0  9.7      0.0
3368 attr(,"assign")
3369 [1] 0 1 2 3
3370 attr(,"contrasts")
3371 attr(,"contrasts")$quando
3372 [1] "contr.treatment"
3373
3374 >
3375 > # '----> cons_prima = alpha+alpha_1 + (beta+beta_1)*temp
3376 > # '----> cons_dopo = alpha + beta*temp
3377 >
3378 > # H0: alpha_1=beta_1=0
3379 > summary(fit)
3380
3381 Call:
3382 lm(formula = cons ~ quando * temp, x = T)
3383
3384 Residuals:
3385      Min       1Q   Median       3Q      Max
3386 -0.97802 -0.18011  0.03757  0.20930  0.63803
3387
3388 Coefficients:
3389             Estimate Std. Error t value Pr(>|t|)
3390 (Intercept)  4.72385     0.11810  40.000 < 2e-16 ***
3391 quandoprima  2.12998     0.18009  11.827 2.32e-16 ***
3392 temp        -0.27793     0.02292 -12.124 < 2e-16 ***
3393 quandoprima:temp -0.11530     0.03211  -3.591 0.00073 ***
3394 ---
3395 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3396
3397 Residual standard error: 0.323 on 52 degrees of freedom
3398 Multiple R-Squared:  0.9277, Adjusted R-squared:  0.9235
3399 F-statistic: 222.3 on 3 and 52 DF, p-value: < 2.2e-16
3400
3401 >
3402 > # cons_prima = alpha+alpha_1 + (beta+beta_1)*temp = 6.84-0.38*temp
3403 > # cons_dopo = 4.72 - 0.27 * temp
3404 >
3405 > fit1<-lm(cons~temp)
3406 > anova(fit1,fit)
3407 Analysis of Variance Table
3408
3409 Model 1: cons ~ temp
3410 Model 2: cons ~ quando + temp + quando:temp
3411      Res.Df  RSS Df Sum of Sq    F Pr(>F)
3412 1      54 39.995
3413 2      52 5.425  2    34.570 165.67 < 2.2e-16 ***
3414 ---
3415 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3416 >
3417 > abline(6.84,-0.38)
3418 > abline(4.72,-0.27)
3419 >
3420 > # analisi dei residui
3421 > qqnorm(fit$res);qqline(fit$res)
3422 > plot(fit$fitted,fit$res)
3423 >
3424 > # -----
3425 > fit2<-lm(cons~temp,subset=(quando=="prima"))
3426 > summary(fit2)
3427
3428 Call:
3429 lm(formula = cons ~ temp, subset = (quando == "prima"))
3430
3431 Residuals:
3432      Min       1Q   Median       3Q      Max
```

Page 44 of 45

```

3433 -0.62020 -0.19947 0.06068 0.16770 0.59778
3434
3435 Coefficients:
3436             Estimate Std. Error t value Pr(>|t|)
3437 (Intercept)  6.85383    0.11842   57.88 <2e-16 ***
3438 temp       -0.39324    0.01959  -20.08 <2e-16 ***
3439 ---
3440 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3441
3442 Residual standard error: 0.2813 on 24 degrees of freedom
3443 Multiple R-Squared: 0.9438,    Adjusted R-squared: 0.9415
3444 F-statistic: 403.1 on 1 and 24 DF,  p-value: < 2.2e-16
3445
3446 >
3447 > fit2<-lm(cons-temp,subset=(quando=='dopo'))
3448 > summary(fit2)
3449
3450 Call:
3451 lm(formula = cons ~ temp, subset = (quando == "dopo"))
3452
3453 Residuals:
3454      Min       1Q   Median       3Q      Max
3455 -0.97802 -0.11082  0.02672  0.25294  0.63803
3456
3457 Coefficients:
3458             Estimate Std. Error t value Pr(>|t|)
3459 (Intercept)  4.72385    0.12974   36.41 < 2e-16 ***
3460 temp       -0.27793    0.02518  -11.04 1.05e-11 ***
3461 ---
3462 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
3463
3464 Residual standard error: 0.3548 on 28 degrees of freedom
3465 Multiple R-Squared: 0.8131,    Adjusted R-squared: 0.8064
3466 F-statistic: 121.8 on 1 and 28 DF,  p-value: 1.046e-11
3467
3468 >
3469 > # H0: beta_1 = beta_2
3470 >
3471

```