

Esempio di prova scritta per i Frequentanti

Quesiti iniziali. In riferimento a quanto trattato nel Capitolo ... del libro di testo di Martin Bland, la/o studente/essa esponga

in queste domande, gli studenti danno prova della loro preparazione teorica in quanto sono richiesti di elencare le definizioni e le proprietà degli argomenti trattati nei capitoli 4 (Statistica descrittiva), 5 (Rappresentare i dati), 6 (Distribuzioni di probabilità e le variabili aleatorie; la distribuzione binomiale), 7 (La distribuzione normale)

Quesiti intermedi. Con il seguente comando di R ..

*in queste domande, sia a risposta aperta che a risposta multipla, gli studenti danno prova della loro conoscenza e comprensione degli argomenti da essi stessi svolti nelle cosiddette **Attività di studio individuale**, ossia 'i compiti per casa'.*

Quesiti intermedi. In base al seguente grafico ..

gli studenti danno prova di saper interpretare le informazioni numeriche fornite da un grafico. E viceversa, date alcune informazioni numeriche e alcune statistiche descrittive gli studenti sanno riconoscere il grafico appropriato.

Quesiti finali. In una ricerca di laboratorio si vuole determinare se un certo **risultato** sia legato alla misura di una certa **variabile** quantitativa. La popolazione da cui è tratto il campione in esame è suddivisa in **gruppi** (variabile qualitativa) ed i dati raccolti nel foglio elettronico danno origine ad un dataset di tre colonne. Ecco le statistiche descrittive del dataset ed i modelli lineari presi in esame:

```
> summary(risultato)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 22.00  26.25   31.00   32.25  34.00   72.00
```

```
> table(gruppi)
gruppi
  A    B
 51   51
```

```
> summary(variabile)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 11.00  18.00   20.50   23.21  24.00   58.00
```

```
> modello1 = lm (risultato ~ gruppi * variabile)
> summary(modello1)
```

```
Call:
lm(formula = risultato ~ gruppi * variabile)
```

```
Residuals:
  Min      1Q  Median      3Q      Max
-11.645  -5.831  -0.610   2.879  33.321
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    29.62706     6.78452   4.367 3.13e-05 ***
gruppiB        -0.43124     7.33749  -0.059  0.953
variabile       0.02667     0.33686   0.079  0.937
gruppiB:variabile 0.16684     0.35039   0.476  0.635
```

```
Residual standard error: 8.015 on 98 degrees of freedom
Multiple R-squared:  0.1006, Adjusted R-squared:  0.07306
F-statistic: 3.654 on 3 and 98 DF,  p-value: 0.01516
```

```
> modello2 = lm (risultato ~ gruppi + variabile)
> summary(modello2)
```

```
Call:
lm(formula = risultato ~ gruppi + variabile)
```

```
Residuals:
  Min      1Q  Median      3Q      Max
-11.299  -5.586  -0.708   2.915  33.213
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    26.56403     2.14749  12.370 <2e-16 ***
gruppiB         2.96704     1.69713   1.748  0.0835 .
variabile       0.18088     0.09231   1.959  0.0529 .
```

```
Residual standard error: 7.983 on 99 degrees of freedom
Multiple R-squared:  0.09851, Adjusted R-squared:  0.0803
F-statistic: 5.409 on 2 and 99 DF,  p-value: 0.005895
```

```
> anova(modello1, modello2)
Analysis of Variance Table
```

```
Model 1: risultato ~ gruppi * variabile
Model 2: risultato ~ gruppi + variabile
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     98 6294.8
2     99 6309.4 -1   -14.563 0.2267  0.635
```

```
> modello3 = lm (risultato ~ gruppi)
> summary(modello3)
```

Call:

```
lm(formula = risultato ~ gruppi)
```

Residuals:

Min	1Q	Median	3Q	Max
-9.333	-5.333	-0.333	2.843	37.667

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	30.157	1.134	26.602	<2e-16 ***
gruppiB	4.176	1.603	2.605	0.0106 *

Residual standard error: 8.096 on 100 degrees of freedom

Multiple R-squared: 0.06355, Adjusted R-squared: 0.05419

F-statistic: 6.787 on 1 and 100 DF, p-value: 0.01059

```
> anova(modello3, modello2)
```

Analysis of Variance Table

Model 1: risultato ~ gruppi

Model 2: risultato ~ gruppi + variabile

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	100	6554.1				
2	99	6309.4	1	244.69	3.8394	0.05287 .

```
> modello4 = lm (risultato ~ variabile)
```

```
> summary(modello4)
```

Call:

```
lm(formula = risultato ~ variabile)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.623	-4.998	-1.097	2.283	35.002

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	26.68546	2.16832	12.307	< 2e-16 ***
variabile	0.23958	0.08687	2.758	0.00692 **

Residual standard error: 8.065 on 100 degrees of freedom

Multiple R-squared: 0.07068, Adjusted R-squared: 0.06139

F-statistic: 7.606 on 1 and 100 DF, p-value: 0.006918

```
> anova(modello4, modello2)
```

Analysis of Variance Table

Model 1: risultato ~ variabile

Model 2: risultato ~ gruppi + variabile

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	100	6504.2				
2	99	6309.4	1	194.79	3.0565	0.08352 .

In base alle pagine precedenti decidete

- quale è la dimensione del campione?

Inoltre, quale potrebbe essere il modello statistico minimale adeguato da tenere in considerazione?

- il modello 1
- il modello 2
- il modello 3
- il modello 4

Giustificate la vostra scelta: ...