

L'obiettivo prefissato è quello di studiare la situazione lavorativa degli ex-studenti a quattro anni dalla laurea. Si utilizzerà quindi un modello di regressione per capire se e quali caratteristiche individuali portino con più o meno facilità ad essere occupati dopo tale periodo. Data la natura binaria della variabile d'interesse sono stati presi in considerazione Modelli Lineari Generalizzati di tipo logistico.

Dei 6707 individui appartenenti al campione il 73% si dichiara occupato a 4 anni dalla laurea. La Figura 1 ci mostra il tasso di occupazione dopo tale periodo per i diversi tipi di facoltà, dove il colore grigio chiaro indica l'essere occupato. Appare evidente che per alcune lauree ci sia un tasso di occupazione maggiore (difesa e sicurezza) rispetto ad altre (Geo-Biologico).

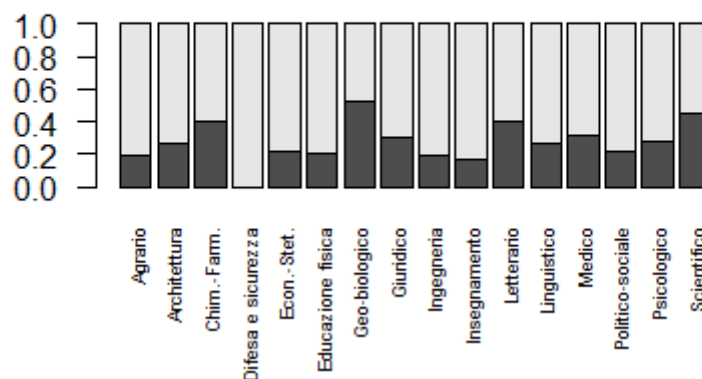


Figura 1. Tasso di occupazione per facoltà

Avendo un gran numero di variabili ausiliarie e volendo capire se esiste una relazione statisticamente significativa tra esse e la variabile di interesse, si ricorre appunto a modelli lineari generalizzati di tipo logistici. Il primo modello testato è un modello che chiameremo completo, poiché prende in considerazione tutte le variabili ausiliarie ritenute di interesse rispetto alla variabile dipendente, che in dettaglio sono:

- GRUPPO di Laurea (GL)
- Partecipazione a programmi volti a promuovere la mobilità studentesca in ambito europeo o extraeuropeo (ERASMUS)
- Voto di laurea conseguito (VOTOL)
- Attualmente cerca lavoro (CERCAL)
- Regione di attuale domicilio o estero (REGIONEDOM)
- Voto del diploma (se conseguito in Italia) (VD1)

Al fine di migliorare l'interpretabilità dei coefficienti del modello si è scelto di riclassificare le variabili ausiliarie più complesse. Nello specifico sono state ricodificate le variabili VOTOL, VD1, REGIONEDOM e GL.

La variabile numerica VOTOL è stata categorizzata in 4 modalità (<80, tra 81 e 95, tra 96 e 104, >105), la variabile numerica VD1 è stata categorizzata in 5 modalità (<75, tra 76 e 85, tra 86 e 95, >96, diplomati all'estero), la variabile REGIONEDOM ha permesso la creazione delle ripartizioni a 4 modalità (nord, centro, sud-isole, stranieri) ed infine GL è stata categorizzata secondo 3 classi di laurea (STEM, sanitaria, umanistica). Il codice sottostante mostra il modello con le nuove variabili ricodificate.

```
mod_n<-glm(SL~CL+ERASMUS+VOTOLCLASSI+VDCLASSI+CERCAL+RIP,family = binomial(link = "logit"),data = dt)
```

Data la struttura complessa delle variabili ausiliarie il modello completo a variabili ricodificate rende difficile l'interpretazione dei risultati. A tal fine si è proceduto ad effettuare una regressione stepwise per identificare un modello minimale e un test ANOVA per motivarne statisticamente la scelta.

```
mod_sn<-step(mod_n,direction = "both")
```

```
anova(mod_n,mod_sn, test="Chisq")
```

Analysis of Deviance Table

Model 1: SL ~ CL + ERASMUS + VOTOLCLASSI + VDCLASSI + CERCAL + RIP

Model 2: SL ~ CL + CERCAL + RIP

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	6692	7258.2			
2	6700	7265.5	-8	-7.285	0.5062

Dall'output della ANOVA risulta chiaro come le varianze residue tra il modello completo e il modello minimale derivato dalla stepwise non siano statisticamente differenti e quindi per il principio della parsimonia il modello minimale è da preferire rispetto quello completo.

Individuato un modello soddisfacente, possiamo ora procedere all'interpretazione dei coefficienti per capire quali fattori aiutino o meno a possedere un lavoro a quattro anni dal conseguimento della laurea. Dalla summary del modello e dalla Figura 2 si evince immediatamente che tutti i coefficienti di regressione sono statisticamente significativi. Inoltre, possiamo distinguere in base ai segni dei coefficienti, le variabili che aumentano la probabilità di essere occupati da quelle che la diminuiscono. Essendo stata svolta una regressione multipla tutte le considerazioni che si faranno di seguito varranno supponendo costanti le altre covariate.

```
> summary(mod_sn)
```

Call:

```
glm(formula = SL ~ CL + CERCAL + RIP, family = binomial(link = "logit"),
    data = dt)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2286	-1.2152	0.5902	0.7571	1.3961

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.06365	0.14633	14.103	< 2e-16	***
CLSTEM	0.20161	0.08842	2.280	0.022601	*
CLUmanistica	0.33262	0.08896	3.739	0.000185	***
CERCALsi	-1.27095	0.05809	-21.877	< 2e-16	***
RIP2	-0.70434	0.12870	-5.473	4.43e-08	***
RIP3	-0.73672	0.14374	-5.125	2.97e-07	***
RIP4	-1.29335	0.17100	-7.563	3.93e-14	***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 7847.8 on 6706 degrees of freedom
Residual deviance: 7265.5 on 6700 degrees of freedom
AIC: 7279.5

Number of Fisher Scoring iterations: 4

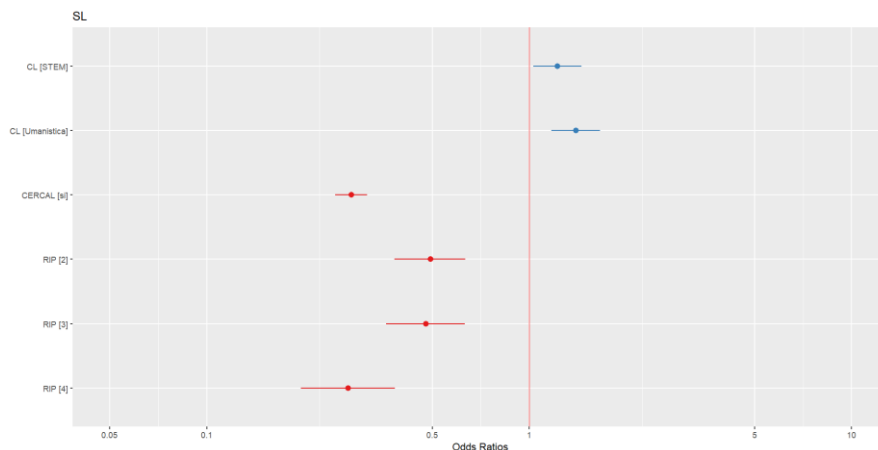


Figura 2. OR del modello di regressione

Nella classe di coefficienti che aumentano la probabilità di essere occupati rientrano per la variabile CL la classe di laurea STEM e quella Umanistica.

La classe di Laurea Sanitaria è quella di riferimento e quindi rispetto ad essa uno studente che ha conseguito una laurea STEM o Umanistica ha una probabilità (maggiore) di essere occupato a 4 anni della laurea rispettivamente di $\exp(0.20161)=1.22$ volte e $\exp(0.33262)= 1.394617$ volte. Tale risultato è da considerarsi verosimile se si tiene conto del fatto che molti studenti della classe Sanitaria sono studenti di medicina. Tali studenti alla fine del primo ciclo di studi iniziano spesso un secondo ciclo di studi di specializzazione che può tenerli lontani per più di 4 anni dal mondo del lavoro.

Nella classe di coefficienti che diminuiscono la probabilità di essere occupati rientrano le variabili Cerca Lavoro e Ripartizione territoriale.

La modalità di riferimento della variabile Cerca lavoro è "No". Il coefficiente quindi legato alla modalità Cerca lavoro "Sì" ci mostra come la probabilità di essere occupati è di $\exp(-1.27095)= 0.280565$ volte (quindi minore) rispetto a chi non ricerca lavoro. Anche tale indicazione ci sembra verosimile identificando più probabilmente negli ex-studenti che cercano lavoro, ragazzi alla ricerca di una prima occupazione rispetto a lavoratori in cerca di un miglior posto lavorativo.

Per quanto concerne la variabile Ripartizione territoriale, la modalità di riferimento è il "Nord". I coefficienti legati alle modalità Centro, Sud-Isole e Stranieri mostrano come gli ex-studenti residenti in tali ripartizioni/estero abbiano una probabilità di essere occupati minore rispetto a quelli del Nord e nel dettaglio pari a $\exp(-0.70434)= 0.4944348$ volte, $\exp(-0.73672)= 0.4786814$ volte e $\exp(-1.29335)=0.2743502$ volte quelli residenti nel Nord. Anche tale evidenza ci sembra ragionevole considerando le differenze territoriali in relazione alla possibilità per un giovane laureato di trovare un lavoro in Italia.

In conclusione, dalle analisi svolte risulta chiaro che la facoltà presso cui si consegue la laurea ha un importante impatto su quella che è la probabilità di trovare un lavoro nei successivi 4 anni.

Non risulta esserci nessuna relazione tra i voti di laurea e diploma conseguiti rispetto al trovare o meno lavoro dopo la laurea, così come non sembra esserci una relazione tra aver effettuato progetti formativi all'estero durante gli studi rispetto a trovare o meno occupazione dopo la laurea.

Risulta infine significativa la residenza degli ex-studenti, mostrando un divario forte a livello territoriale tra il nord e resto d'Italia, continuando ad identificare in tale regioni il territorio trainante per l'occupazione giovanile e nello specifico per i laureati.