

# Quanto incide il livello di istruzione sul reddito percepito?

Questo lavoro di ricerca si prefigge di studiare l'eventuale presenza di correlazioni tra il livello di istruzione e il reddito da lavoro

Tipo dato	Retribuzione lorda oraria per ora retribuita delle posizioni lavorative dipendenti in euro (mediana).											
Territorio	Italia											
Qualifica contrattuale	totale											
Classe di dipendenti	totale											
Seleziona periodo	2021											
Titolo di studio	RLOSS			RLOD			RLOL			RLOT		
Sesso	RLOMSS	RLOFSS	RLOSST	RLODM	RLODF	RLODT	RLOLM	RLOLF	RLOLT	RLOTM	RLOTF	RLOTT
fornitura di acqua reti fognarie, attività di gestione dei rifiuti e costruzioni	14,30	12,61	14,22	14,94	14,73	14,90	17,91	16,17	17,08	14,52	14,62	14,54
servizi (g-s, escluso o)	11,56	10,48	11,54	11,68	11,82	11,70	13,79	12,61	13,37	11,53	11,63	11,54
commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli trasporto e magazzinaggio	10,93	10,24	10,61	11,85	11,16	11,46	16,10	13,03	14,07	11,55	11,07	11,30
attività dei servizi di alloggio e di ristorazione	11,27	11,34	11,30	12,03	11,77	11,88	16,03	13,76	14,35	11,73	11,83	11,79
servizi di informazione e attività finanziarie e assicurative	12,09	10,88	11,98	13,28	14,38	13,54	15,34	15,25	15,29	12,42	13,32	12,53
attività immobiliari	10,38	10,29	10,33	10,68	10,39	10,52	10,97	10,75	10,82	10,47	10,37	10,42
attività professionali, scientifiche e tecniche	15,39	12,00	14,00	16,53	13,53	15,06	17,87	15,33	16,78	16,96	14,06	15,67
noleggio, agenzie di viaggio, servizi di supporto alle imprese	22,57	15,02	19,77	26,38	21,26	23,74	26,22	21,25	23,42	26,14	21,09	23,47
istruzione	11,19	10,52	10,87	11,95	11,32	11,48	16,17	12,22	13,13	11,81	11,27	11,45
sanità e assistenza sociale	10,98	9,99	10,46	12,91	11,57	11,96	17,54	13,58	15,11	13,99	11,95	12,61
attività artistiche, sportive, di intrattenimento e divertimento	10,15	9,46	9,85	10,66	10,38	10,53	12,17	11,86	11,97	10,40	10,11	10,27
altre attività di servizi	10,33	9,62	9,87	10,84	10,37	10,51	14,15	12,36	12,64	11,74	11,12	11,28
	10,39	10,29	10,31	10,83	10,52	10,57	13,23	11,50	11,76	11,24	10,73	10,81
Mean	10,24	9,90	10,11	10,82	10,44	10,65	11,53	11,08	11,24	10,68	10,46	10,58
Std	9,37	8,15	8,39	10,16	8,46	8,89	11,66	10,53	10,80	9,69	8,36	8,70
Cv												
Min	12,08	10,72	11,57	13,04	12,14	12,49	15,38	13,42	14,12	12,99	12,13	12,46
1st Quartile	3,31	1,59	2,73	4,08	3,02	3,54	3,81	2,76	3,26	4,08	2,94	3,49
Median	27%	15%	24%	31%	25%	28%	25%	21%	23%	31%	24%	28%
3rd Quartile	9,37	8,15	8,39	10,16	8,46	8,89	10,97	10,53	10,80	9,69	8,36	8,70
Max	10,36	9,95	10,21	10,83	10,42	10,55	12,70	11,68	11,87	10,96	10,60	10,70
Lower bound	10,98	10,29	10,61	11,85	11,32	11,48	15,34	12,61	13,37	11,73	11,27	11,45
Upper bound	11,83	11,11	11,76	13,10	12,68	12,75	16,86	14,51	15,20	13,21	12,64	12,57
Baffo inf.	22,57	15,02	19,77	26,38	21,26	23,74	26,22	21,25	23,42	26,14	21,09	23,47
Baffo sup.	8,15	8,20	7,89	7,42	7,03	7,25	6,47	7,44	6,86	7,59	7,54	7,88
	14,03	12,86	14,09	16,50	16,07	16,05	23,09	18,74	20,20	16,57	15,70	15,38
	9,37	9,46	8,39	10,16	8,46	8,89	10,97	10,53	10,80	9,69	8,36	8,70
	12,09	12,61	14,00	14,94	14,73	15,06	17,91	16,17	17,08	14,52	14,62	14,54

Le variabili sono:


- **RLOSS**: retribuzione lorda oraria senza titolo di studio, licenza di scuola elementare e media
- **RLOD**: retribuzione lorda oraria con diploma
- **RLOL**: retribuzione lorda oraria con laurea e post laurea
- **RLOT**: retribuzione lorda oraria totale

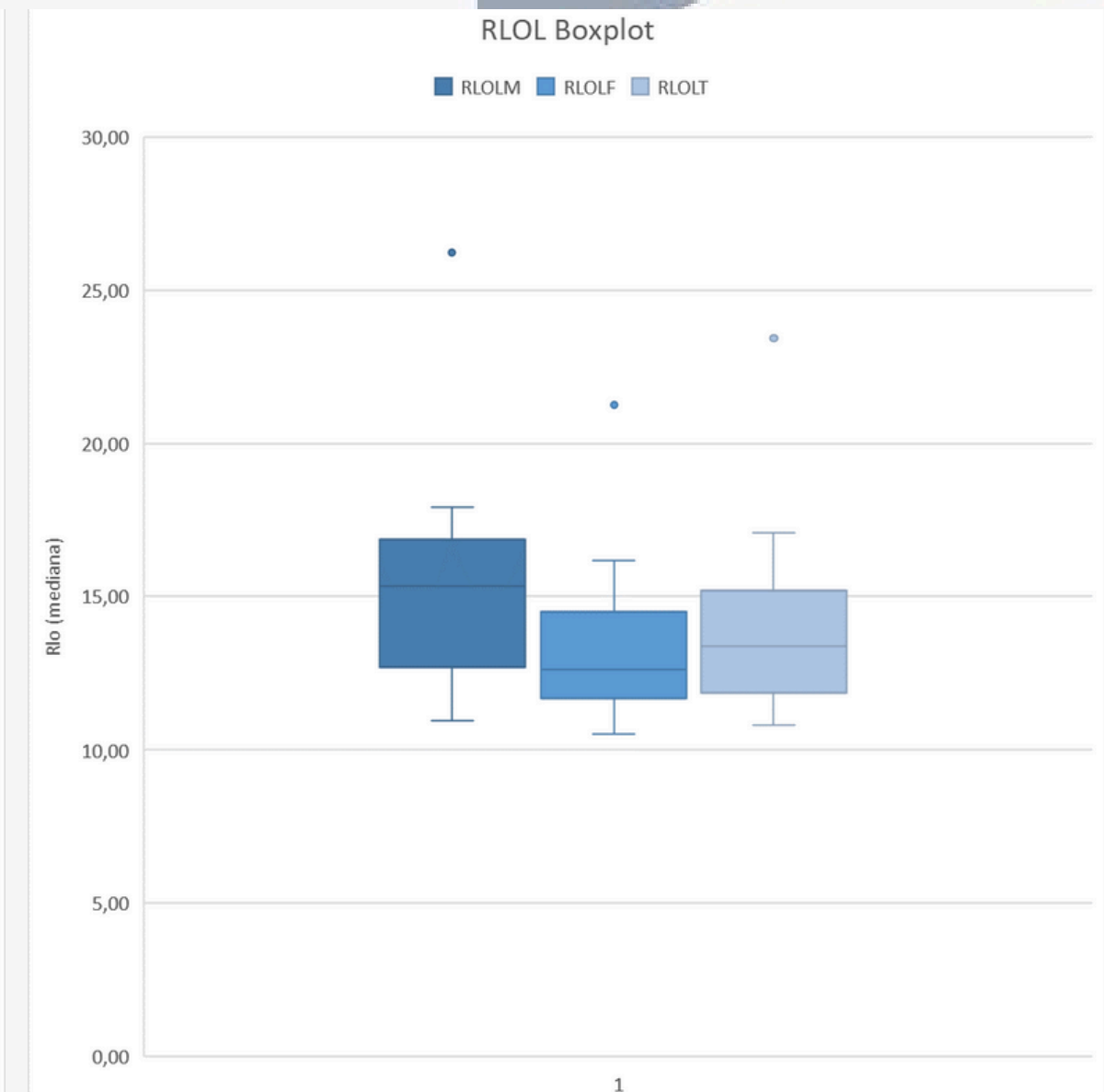
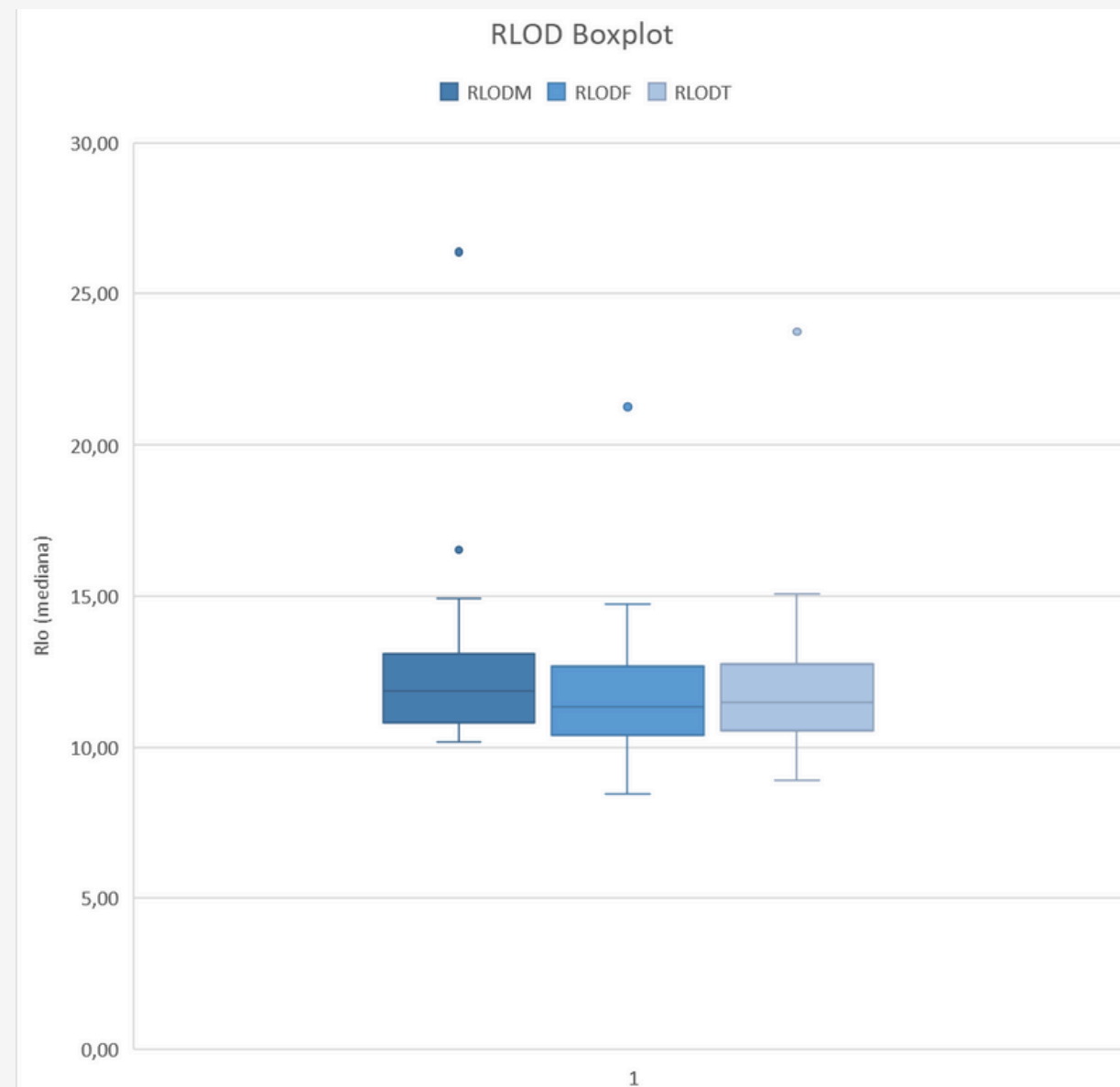
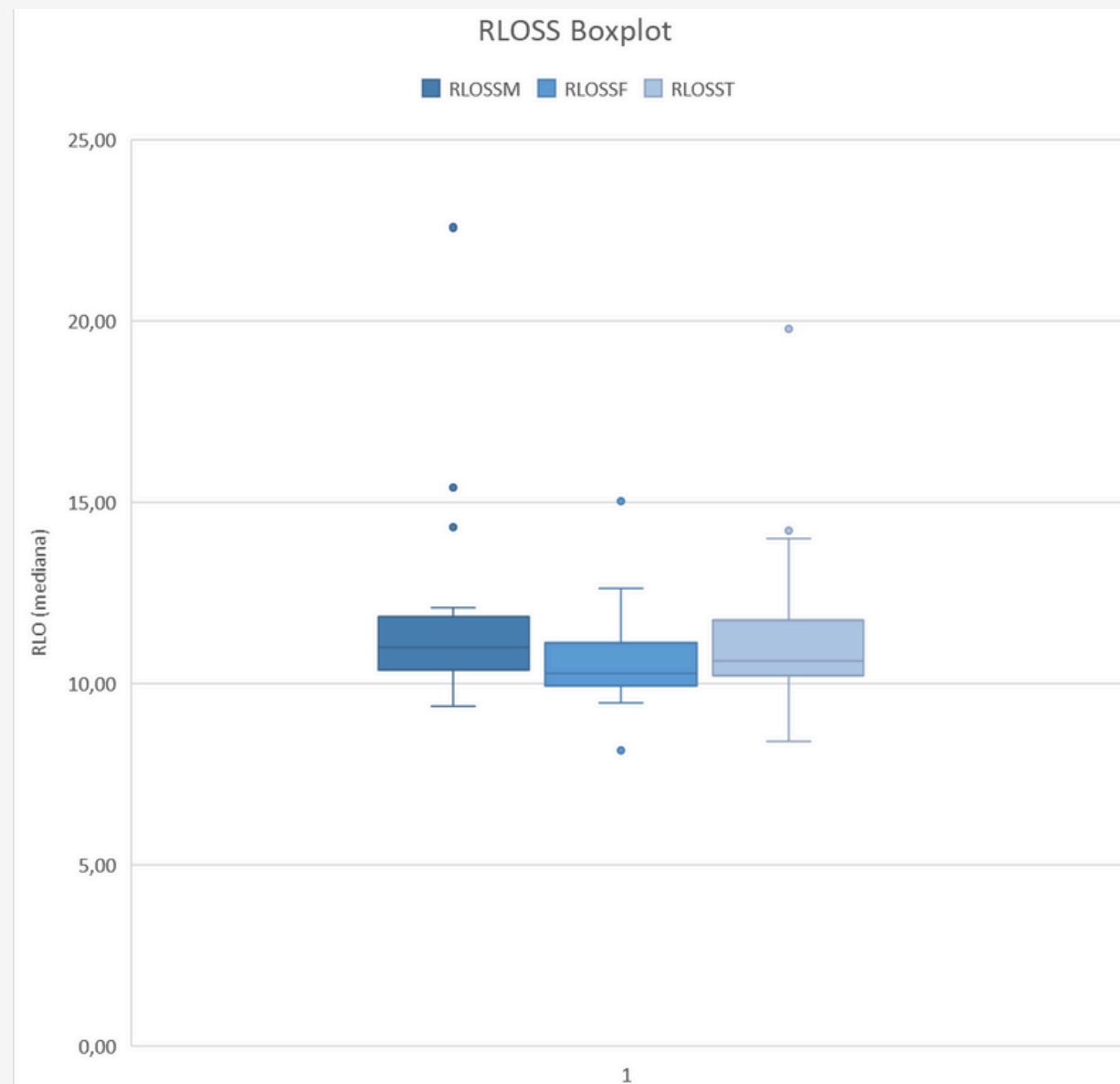
\* per ciascun livello di istruzione sono state rilevate le retribuzioni distintamente per maschi e per femmine.

Il seguente dataset fa riferimento alle rilevazioni ISTAT dell'anno 2021 relative alla retribuzione lorda oraria per ora retribuita delle posizioni lavorative dipendenti in euro (mediana), differenziate per livello di istruzione.



**Gli indici riportati nella slide precedente evidenziano:**

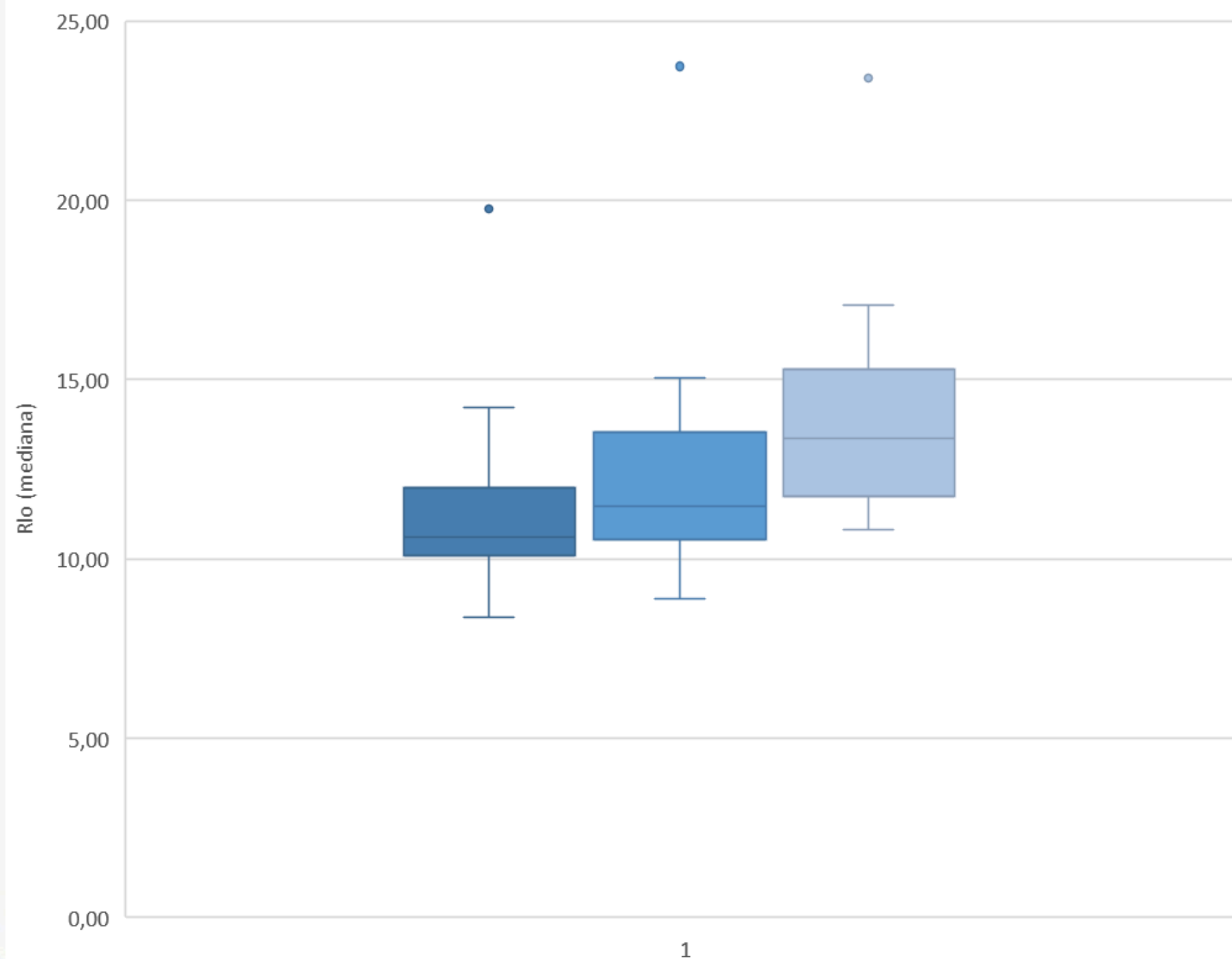
- **un andamento moderatamente crescente della retribuzione all'aumentare del livello di istruzione, come si può notare dalle mediane e dalle medie;**
  - **la media più alta corrisponde a 15.38 per RLOLM;**
  - **le distribuzioni delle retribuzioni mostrano una variabilità rispetto al valore medio intorno al 25% per le retribuzioni orarie totali;**
  - **questa variabilità è dovuta alla presenza di outliers come si può notare dalla massima retribuzione registrata di 26.38 nel caso di RLODM.**
- 



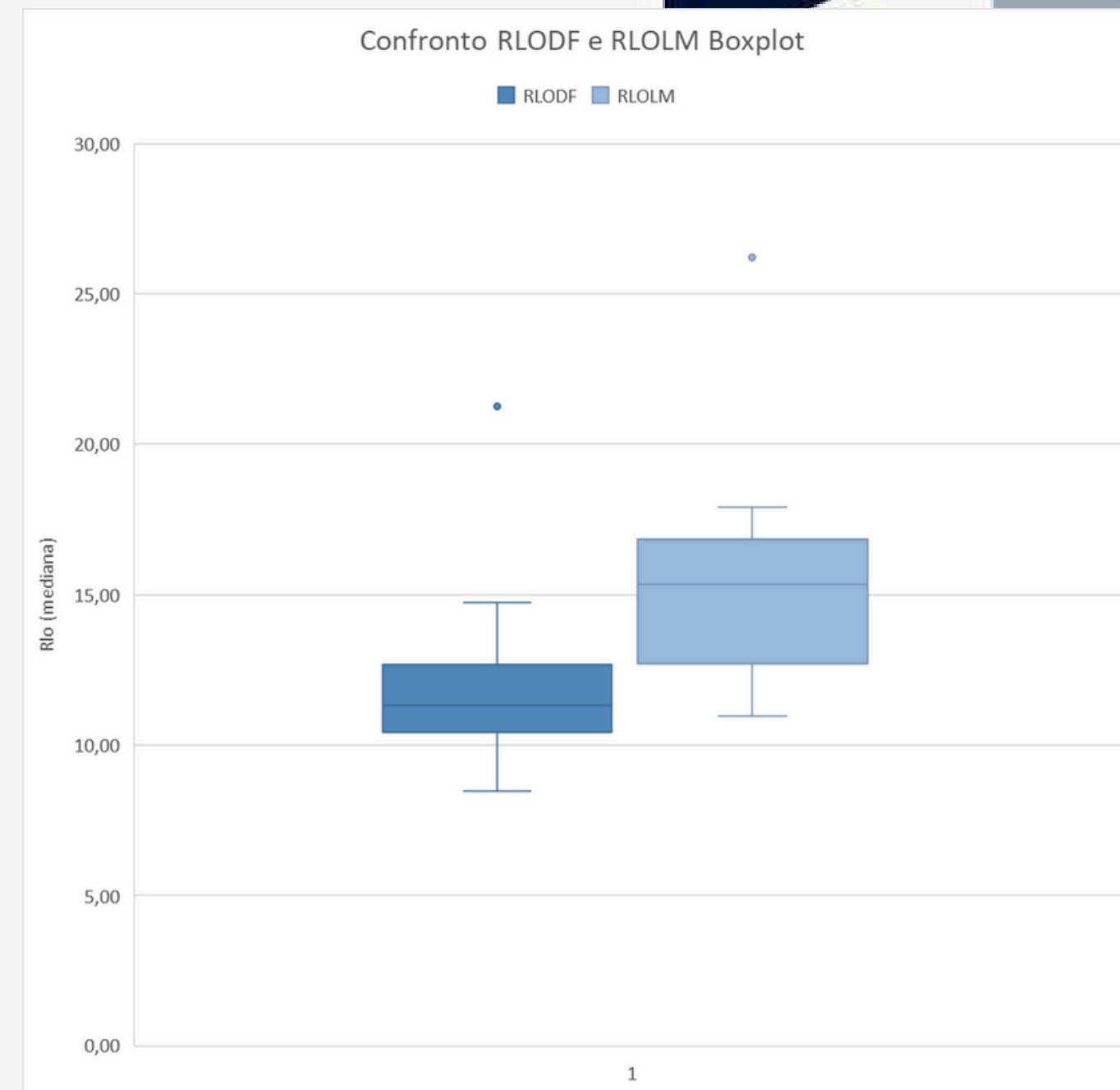
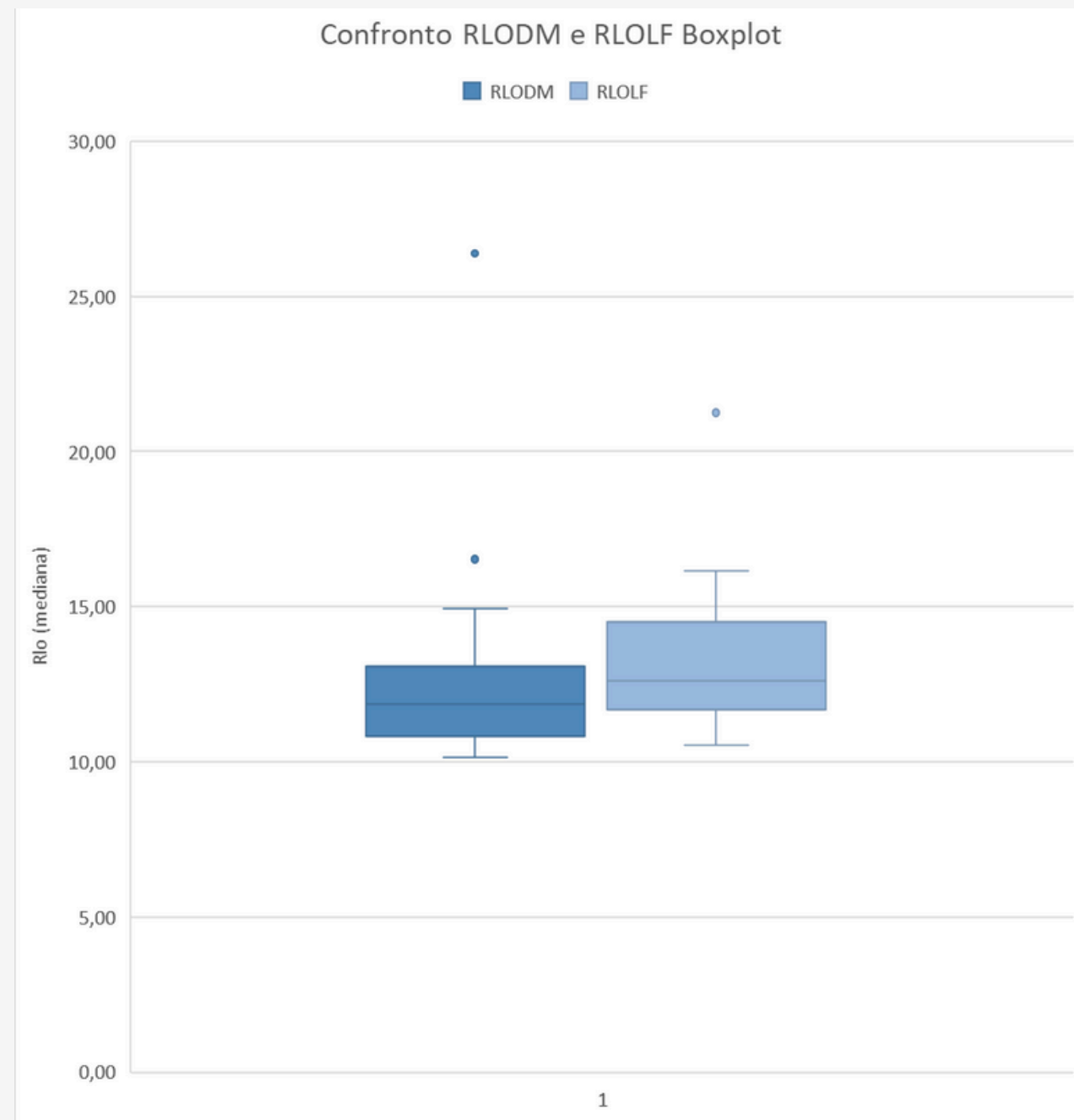
**I box plot dimostrano un'asimmetria positiva per tutte le distribuzioni in quanto la mediana è minore della media aritmetica confermando la presenza di outliers. Inoltre, possiamo notare che per ogni livello di istruzione la distribuzione delle RLO assume valori maggiori per i maschi, questo è ancora più evidente quando si ha un titolo di laurea.**

Confronto RLOT Boxplot

■ RLOSST ■ RLODT ■ RLOLT



**Le retribuzioni lorde  
orarie totali sono  
crescenti nel livello  
di studio**



**Confronto: la variabilità delle distribuzioni tra maschi con laurea e femmine con il diploma è maggiore rispetto al caso opposto**

## Analisi Anova RLOM

→ i tre gruppi per cui confrontiamo le medie sono  
RLOSSM, RLODM, RLOLM

Risultati ANOVA:

	sum_sq	df	F	PR(>F)
gruppo	86.586471	2.0	3.080174	0.056462
Residual	590.328893	42.0	NaN	NaN

ANOVA non significativa, non eseguiamo il test di Tukey.  
<ipython-input-35-88f6aefbb570>:17: FutureWarning: Series.  
if anova\_table['PR(>F)'][0] < 0.05:

con p-value 0.056 non possiamo rifiutare l'ipotesi nulla e affermare che le medie tra i gruppi siano diverse, tuttavia, essendo un valore molto vicino al livello di significatività 0.05, potrebbe essere un indizio di effettiva differenza tra le medie che andrebbe ulteriormente indagato (ad esempio allargando i campioni)

Possiamo tuttavia concludere che per i maschi il livello di studio non sembra essere incisivo sulla retribuzione oraria lorda.

## Analisi Anova RLOF →

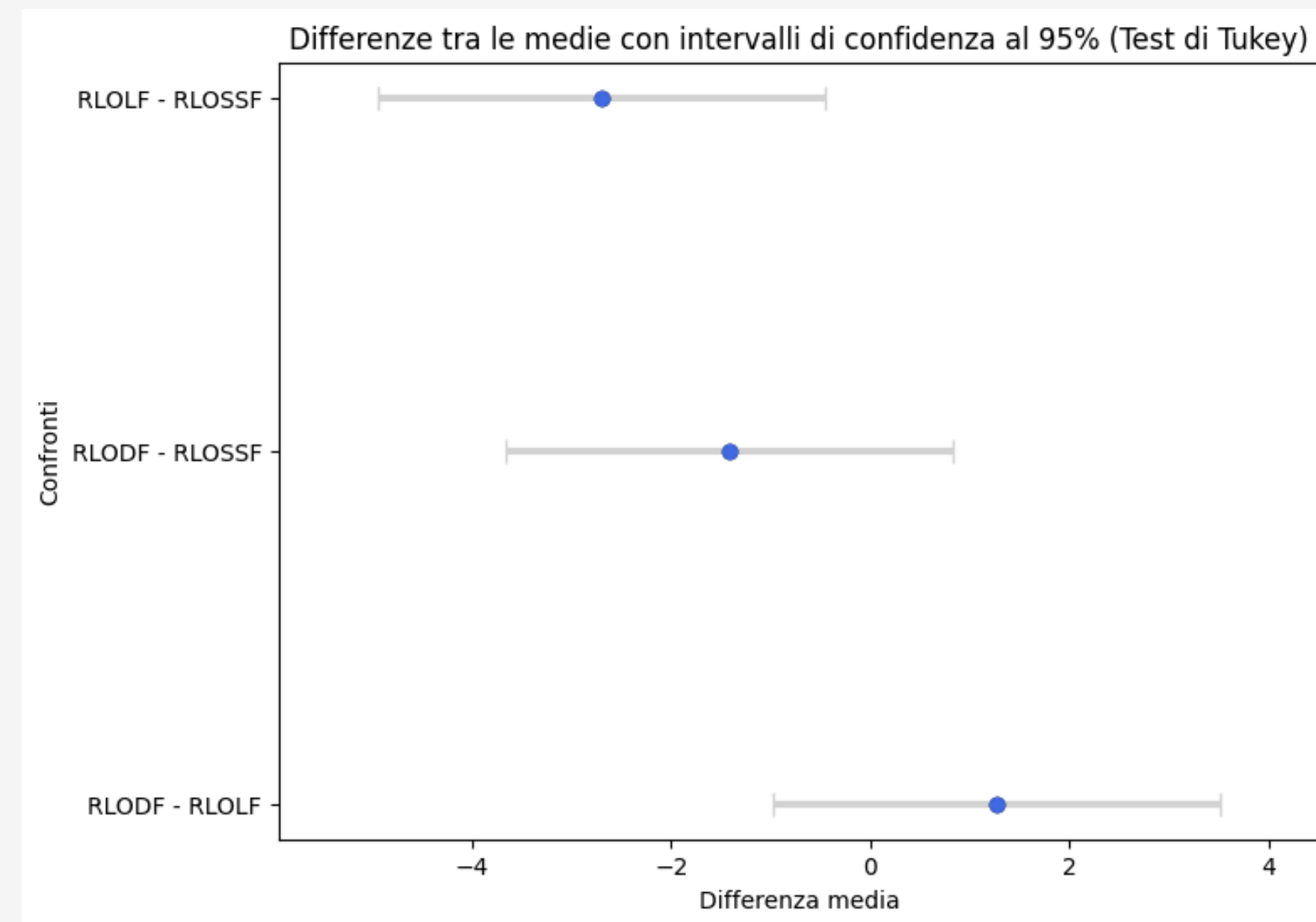
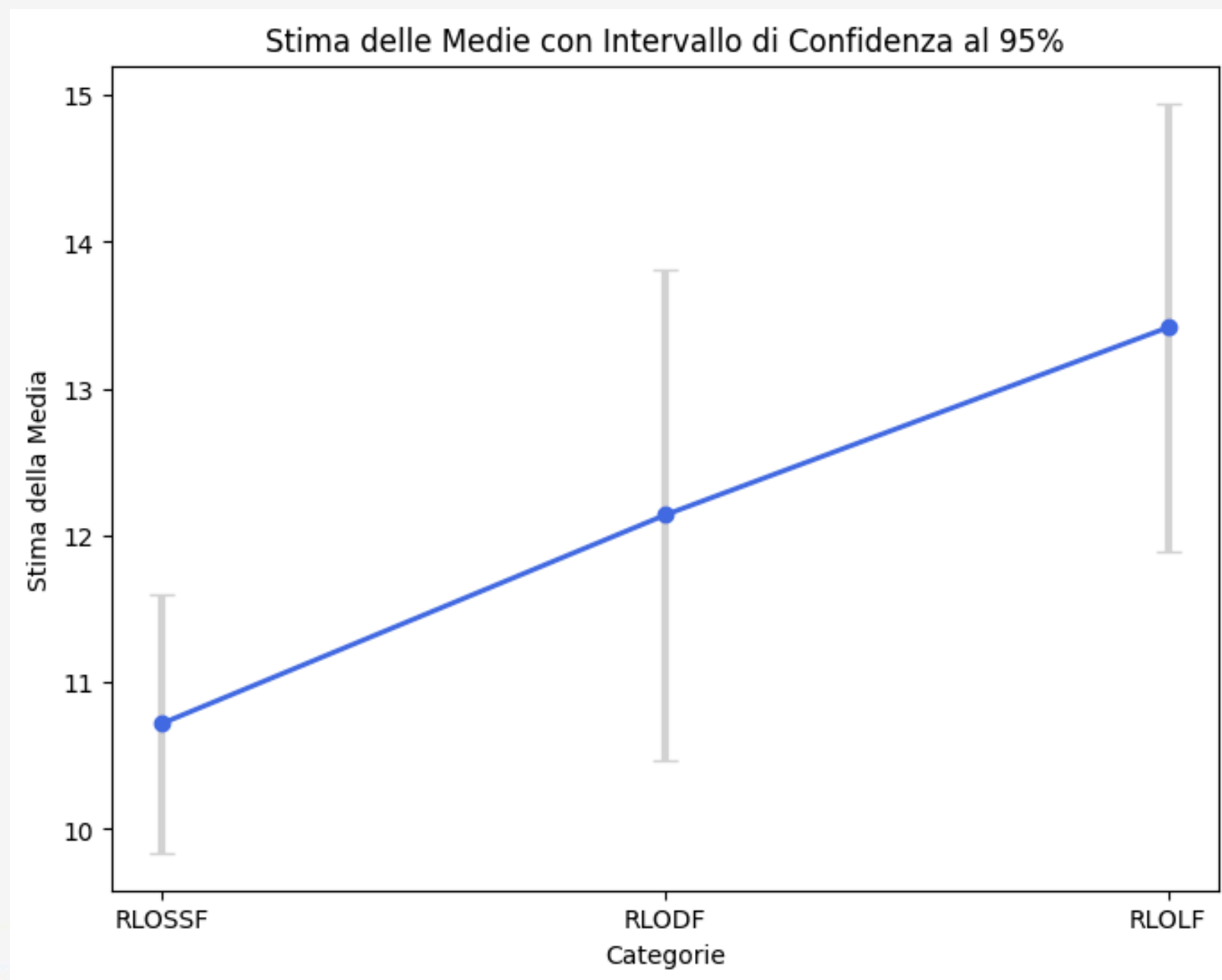
I tre gruppi per cui confrontiamo le medie sono RLOSSF, RLODF, RLOLF

```
Risultati ANOVA:
              sum_sq    df      F    PR(>F)
gruppo      54.698413    2.0  4.266123  0.020573
Residual  269.253067   42.0      NaN      NaN
<ipython-input-10-85bde93dd5c9>:17: FutureWarning: Series
if anova_table['PR(>F)'][0] < 0.05:
```

```
Risultati del test di Tukey:
Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====
group1 group2 meandiff p-adj  lower  upper  reject
-----
RLODF  RLOLF    1.2787 0.3588 -0.9675  3.5248  False
RLODF  RLOSSF  -1.4207 0.2845 -3.6668  0.8255  False
RLOLF  RLOSSF  -2.6993 0.0152 -4.9455 -0.4532  True
-----
```

Dal test ANOVA si evince una relazione tra il livello di studio e la RLOF, procediamo quindi con il test di Tukey. Da questo possiamo evincere che l'unica differenza significativa è RLOLF e RLOSSF: in media le donne senza titolo di studio guadagnano 2.70€ in meno





## Regressione lineare RLOM per titolo di studio

Seconda Regressione:

### OLS Regression Results

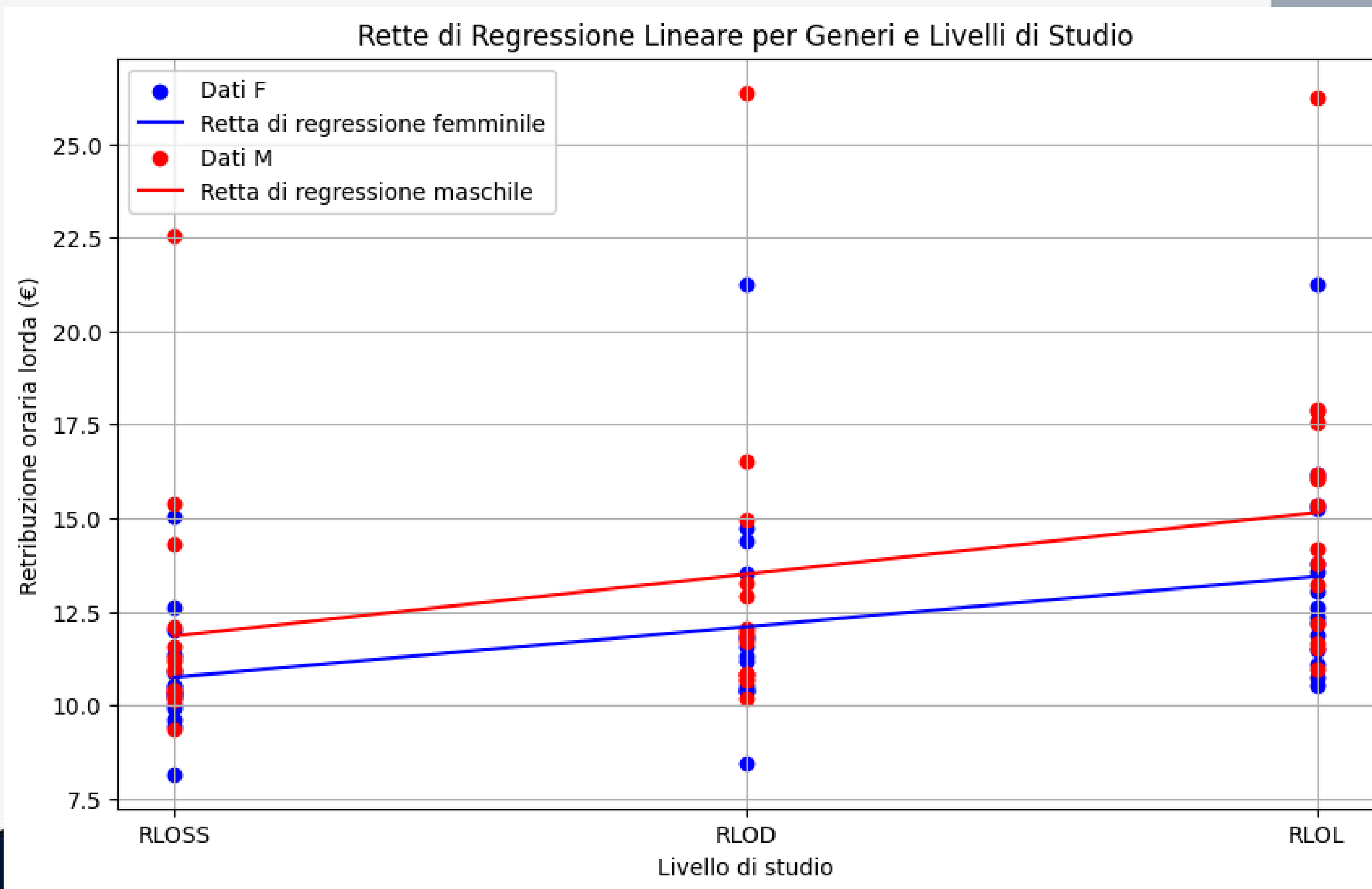
```
=====
Dep. Variable:          y      R-squared:          0.121
Model:                  OLS    Adj. R-squared:       0.100
Method:                 Least Squares    F-statistic:      5.911
Date:                   Wed, 27 Nov 2024    Prob (F-statistic): 0.0193
Time:                   18:06:02    Log-Likelihood:   -121.95
No. Observations:       45    AIC:              247.9
Df Residuals:           43    BIC:              251.5
Df Model:                1
Covariance Type:        nonrobust
=====
              coef    std err          t      P>|t|      [0.025     0.975]
-----
const         11.8456     0.877     13.509     0.000     10.077     13.614
x1             1.6513     0.679      2.431     0.019      0.282      3.021
=====
```

**l'intercetta ci dice che la RLOM di base è 11.86. Questa aumenta poi di 1.65€ per titolo di studio. I p-value sono significativi. Tuttavia  $R^2$  ci dice che il modello riesce a spiegare solo il 12% della variabilità.**

## Regressione lineare RLOF per titolo di studio

```
Prima Regressione:
                        OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.169
Model:                  OLS    Adj. R-squared:      0.149
Method:                 Least Squares    F-statistic:      8.726
Date:                  Wed, 27 Nov 2024    Prob (F-statistic): 0.00507
Time:                  18:06:02    Log-Likelihood:    -104.11
No. Observations:      45    AIC:              212.2
Df Residuals:          43    BIC:              215.8
Df Model:               1
Covariance Type:       nonrobust
=====
                        coef    std err          t      P>|t|      [0.025    0.975]
-----
const          10.7430      0.590     18.213     0.000      9.553     11.933
x1              1.3497      0.457      2.954     0.005      0.428      2.271
=====
```

**l'intercetta ci dice che la RLOF di base è 10.74. Questa aumenta poi di 1.35€ per titolo di studio. I p-value sono significativi. Tuttavia  $R^2$  ci dice che il modello riesce a spiegare solo il 16% della variabilità.**



# CONCLUSIONI

**Possiamo quindi affermare che il livello di istruzione ha un impatto significativo sulla retribuzione oraria lorda, con un aumento delle retribuzioni all'aumentare del titolo di studio. Questo effetto non è però uniforme tra i generi: le donne, a parità di livello di istruzione, continuano a percepire una retribuzione inferiore rispetto agli uomini.**

**Il divario retributivo risulta evidente sin dalle qualifiche più basse e tende ad amplificarsi con il crescere del livello di studio.**