An Evolutionary Analysis of Industrial Entrepreneurship and R&D Resources in Turkiye

Burcu ALTUNOGLU, Mustafa Erdem BAKIR

Abstract

This study examines the relationship between the number of industrial enterprises in Turkiye and the distribution of R&D personnel by educational attainment between 2009 and 2023. Analyses based on TUIK data show that the number of industrial enterprises and R&D employment, especially at the undergraduate level, have increased steadily over the years. Correlation and regression analyses reveal that there are significant and strong relationships between the number of enterprises and the demand for highly educated personnel. However, there is a limited increase in the proportion of PhD-level and female staff. The findings point to the importance of industrial policies focused on qualified human resources.

1. Introduction

The increase in entrepreneurship in the industrial sector is not only an indicator of economic growth, but also plays a critical role in terms of a country's scientific production capacity, technological competence and innovation potential. In this context, understanding the relationship between structural transformations in the number of industrial enterprises and the composition of human resources employed in research and development (R&D) is of strategic importance for both the effectiveness of science policies and the guidance of industrial strategies.

This study analyzes the effects of the change in the number of industrial enterprises in Turkey between 2009 and 2022 on the educational level and gender distribution of R&D personnel. Within the scope of the research, not only the total R&D employment but also the quality of this employment, i.e. the academic qualification levels and gender-based distribution of the personnel, are taken into account. Thus, it is aimed to assess the causal relationship between industrial growth and scientific labor force and to develop feasible policy recommendations in this context. The study also aims to contribute to decision support systems from an industrial engineering perspective by quantitatively demonstrating this relationship through multivariate linear regression models.

2. Data

The datasets used in this study are based on comprehensive statistical sources published annually by the Turkish Statistical Institute (TUIK) and made publicly available. Two main datasets are used in this study: Industry Enterprise Data and R&D Activity Data.

The Industrial Enterprise Data reports the number of industrial enterprises operating in Turkey by year and has been published regularly since 2009. This data allows for monitoring the quantitative transformation of entrepreneurship activities.

R&D Activity Data contains detailed statistics on the human resources in research and development of enterprises operating in industry and services sectors. This dataset is disaggregated by education level (e.g. doctorate, master's, bachelor's) and gender distribution of R&D personnel and provides critical information for structural analysis of the scientific workforce.

Both data sets are suitable for modeling studies as they are provided at the annual level and have time series characteristics. In addition, the data provided at the national level allows valid inferences to be made across Turkey. All data sets were provided in .xlsx format and converted into analyzable form using the R programming

language. During the matching of variables, only overlapping years for the period 2009-2022 were taken into account and missing observations were systematically flagged and integrated into the analysis process.

Thanks to this structured data approach, the relationships between the number of industrial enterprises and the academic qualification and gender-based structure of R&D human resources could be analyzed holistically with statistical methods.

With this structure, the project will enable a detailed statistical investigation of the following relationships during the analysis phase:

- The relationship between the number of industrial enterprises and the educational level of the R&D workforce,

- The relationship between the number of industrial enterprises and the gender distribution (female/male ratio) of R&D personnel.

3. Analysis

The two datasets used in this project consist of quantitative variables that are well-suited for time-series analysis in terms of both content and structure.

Industrial Enterprise Data contains the annual total number of industrial enterprises operating across Turkiye. For example, in 2009, the number of active enterprises nationwide was recorded at approximately 3.2 million. The data are published regularly on a yearly basis and provide a macro-level representation of the dynamics of the industrial sector.

R&D Data includes the annual number of R&D personnel employed by industrial and service sector enterprises across Turkiye. The distribution of this variable over the years ranges approximately from 10,000 to 40,000 individuals. Given that R&D personnel data directly reflect scientific capacity, the variable holds strategic importance.

Both variables are quantitative and structured in a yearly time-series format. This structure allows for the application of both visual and model-based analytical techniques.

Total Number of Industrial Enterprises Over Time
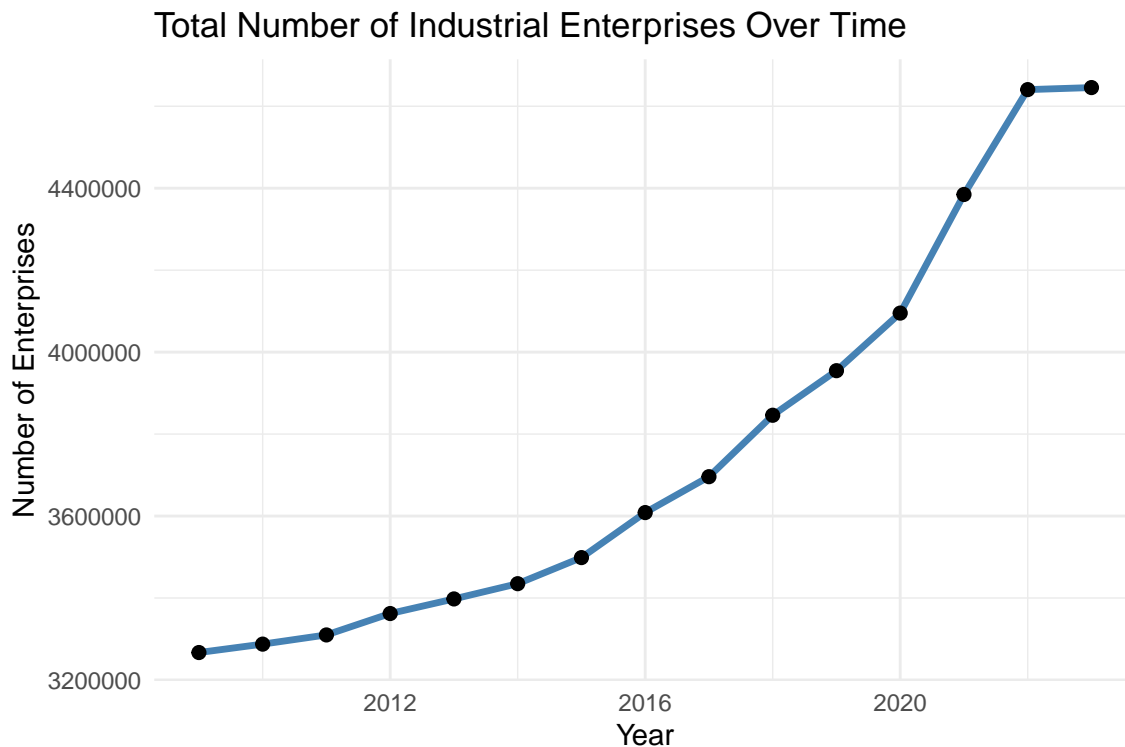
The figure shows a steady upward trend in the total number of industrial enterprises in Turkiye between 2009 and 2022. The growth, which was limited in the 2009-2015 period, gained momentum after 2016; a significant increase was observed especially between 2020-2021. This rapid rise can be attributed to the transformation in

production structures and digitalization during the pandemic period. However, the stagnation in 2022 points to the impact of market saturation or macroeconomic constraints. Overall, there is a quantitative growth in industrial entrepreneurship, but the sustainability and qualitative effects of this growth should be evaluated separately.

Distribution of R&D Personnel by Education Level (2009–2023)
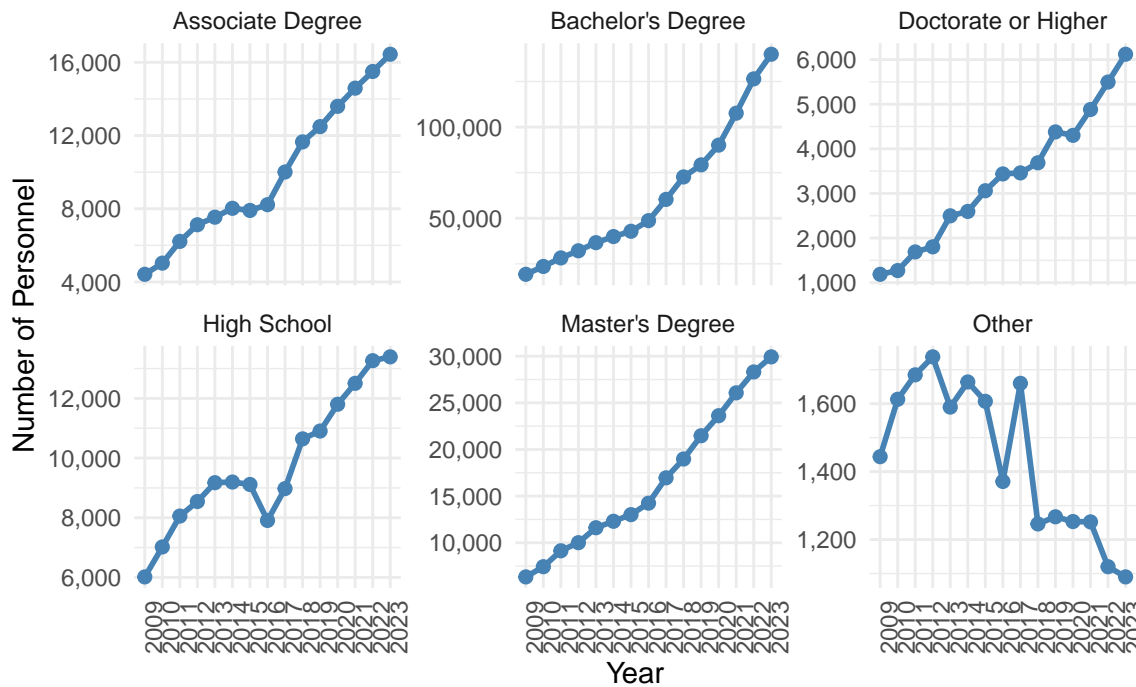
An analysis of the distribution of R&D personnel in Turkiye by educational level between 2009 and 2023 reveals that the number of bachelor's degree graduates is by far the highest and has been increasing rapidly over the years. While there is also an increase in the number of master's and associate's degree graduates, it is noteworthy that the number of personnel with a doctorate degree and above remains relatively limited. This indicates that there is an increasing need for highly qualified human resources in R&D activities, but this need has not yet been adequately met. While the proportion of high school graduates tends to decrease over time, the number of personnel in the "other" category fluctuates. In general, it can be said that there is a transformation in the R&D personnel structure towards higher levels of education.

## Distribution of R&D Personnel by Gender (2009–2023)
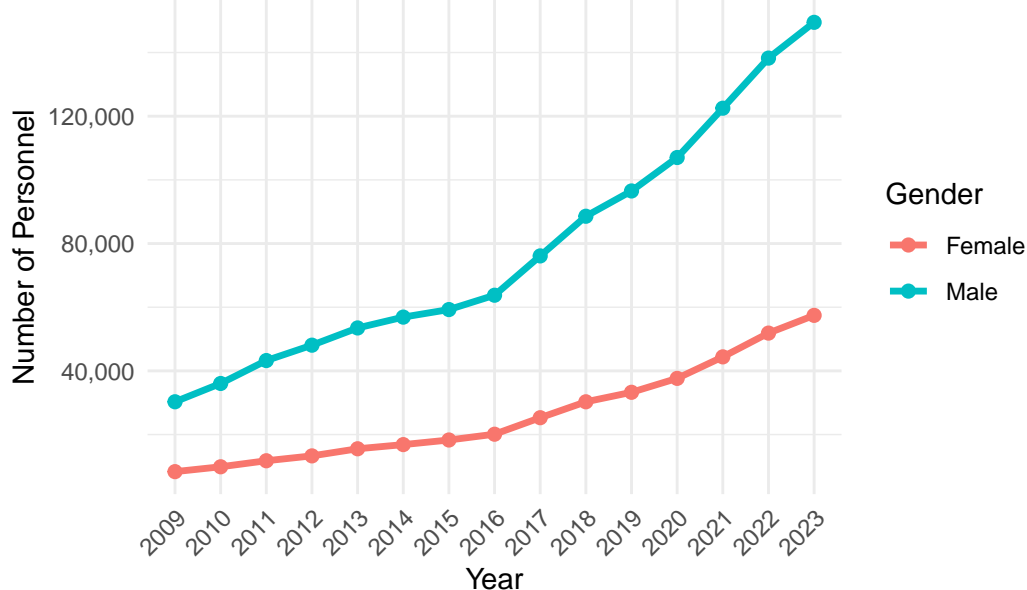
An analysis of the distribution of personnel employed in R&D in Turkiye between 2009 and 2023 by gender

shows that the number of male personnel increased steadily every year, reaching over 135,000 by 2023. The number of female personnel has shown a similar upward trend; however, this increase has been more limited compared to men, reaching approximately 60,000 in 2023. This indicates that a gender-based imbalance persists in the R&D sector.

It is noteworthy that the increase in the number of female personnel gained momentum especially after 2016. This increase shows that women's participation in R&D activities has strengthened and that there has been a positive development in terms of gender equality. However, the current picture reveals the need for more comprehensive and sustainable policies to increase the representation of women in R&D.

Female Share in R&D Personnel (2009–2023)
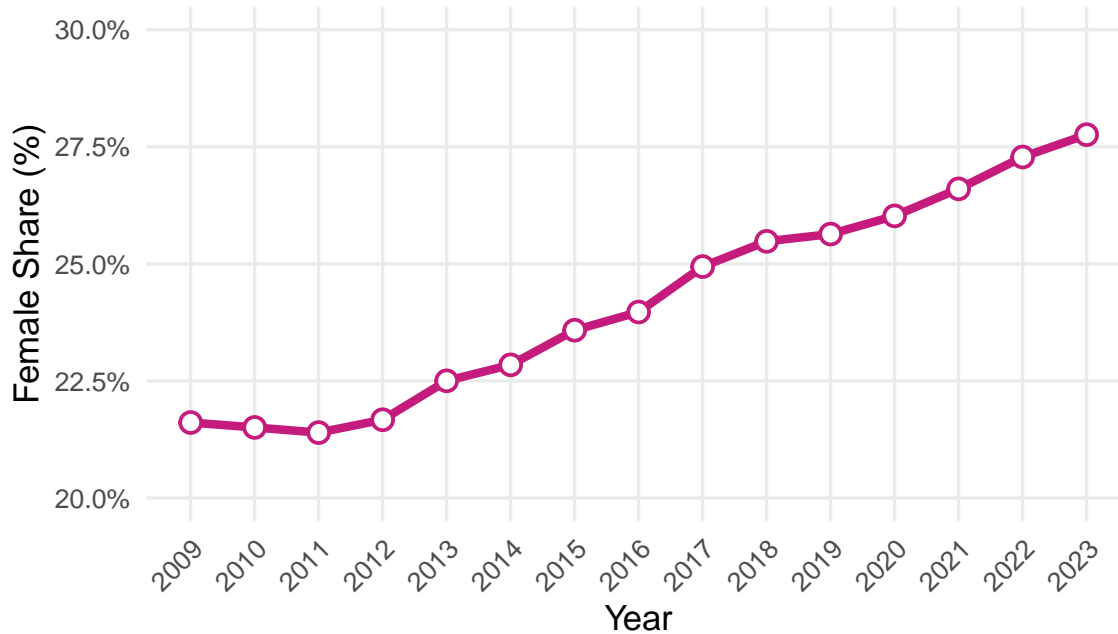
Between 2009 and 2023, the ratio of female employees in the R&D sector in Turkiye increased steadily. From 21.5% in 2009, the proportion of female employees increased to 27.8% by 2023. This increase shows that women's participation in R&D activities is strengthening and that there is a positive trend towards gender equality. However, the ratio is still below 30%, indicating that policies to increase the representation of women in R&D need to be sustained and further supported.

3.1 Exploratory Data Analysis

Correlation Analysis

Positive and strong correlation coefficients are observed for all education levels (except "Other"). In particular, there is an almost perfect linear relationship between bachelor's degree graduates (r = 1.00, p < 0.001) and the number of industrial enterprises. This suggests that the number of R&D personnel employed at the bachelor's level increases systematically with the increase in the number of enterprises. Similarly, master's (r = 0.98) and associate degree (r = 0.97) graduates also exhibit very strong and highly statistically significant relationships. Strong positive correlations are also obtained for PhD and high school graduates (r = 0.96 and r = 0.94, respectively), suggesting that industrial development generally increases the demand for R&D personnel from all educational levels.

The notable exception is the category "Other". A negative correlation coefficient (r = -0.87, p < 0.001) was calculated in this group. This suggests that the number of staff with "Other" educational attainment decreases as the number of enterprises increases, suggesting that industrialization has pushed this group of staff out of the system or that employment patterns have changed. This result suggests that the definitions of education level should be clarified and the transformation in this group should be examined in detail.

3.2 Trend Analysis

The relationship between the number of industrial enterprises and the employment of R&D personnel at different educational levels between 2009 and 2023 in Turkiye is analyzed. First of all, Pearson correlation analysis was used to determine the linear relationships between the number of industrial enterprises and the number of R&D personnel, and the direction, strength and statistical significance of these relationships were evaluated. According to the results obtained, it has been determined that the number of personnel at "Bachelor's Degree", "Master's Degree", "Associate Degree", "Doctorate or Higher" and "High School" levels show highly positive and statistically significant relationships with the number of industrial enterprises.

In line with these findings, simple linear regression analyses were conducted for these five education levels. With these analyses, the effect of the increase in the number of industrial enterprises on the number of R&D personnel at the relevant level was modeled and the significance and fitting power of the model were evaluated by calculating basic parameters such as slope coefficient, constant term, p-value, R^2 and F-statistic for each regression model. In addition, the trends are supported by graphs that visually present the regression lines and data points.

```
==================================================
Regression Summary for: Bachelor's Degree
==================================================

Intercept ( )    :     -233577.2
Slope ( )        :       0.07888
R²               :          0.9917
p-value (slope)  :       6.47e-15
F-statistic      :        1553.89


==================================================
Regression Summary for: Master's Degree
==================================================

Intercept ( )    :     -42553.22
Slope ( )        :       0.01573
R²               :          0.9696
p-value (slope)  :       3.00e-11
F-statistic      :         415.21


==================================================
Regression Summary for: Associate Degree
==================================================

Intercept ( )    :     -19501.87
Slope ( )        :       0.00782
R²               :           0.949
p-value (slope)  :       8.87e-10
F-statistic      :         241.69


==================================================
Regression Summary for: Doctorate or Higher
==================================================

Intercept ( )    :      -8040.64
Slope ( )        :       0.00302
R²               :          0.9247
p-value (slope)  :       1.12e-08
F-statistic      :         159.75


==================================================
Regression Summary for: High School
==================================================

Intercept ( )    :      -6685.04
Slope ( )        :       0.00437
R²               :          0.8888
p-value (slope)  :       1.44e-07
```
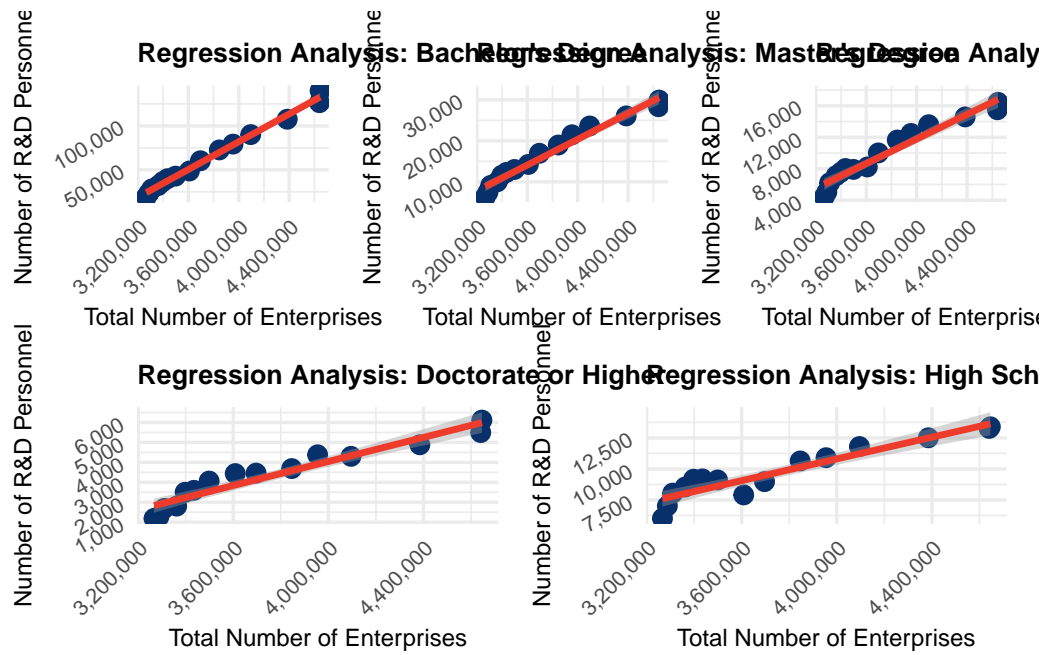
Model Summaries and Interpretation

Bachelor's Degree

- **Model**: *Personnel = –233,577.20 + 0.07888 × Enterprises*
- **R² = 0.9917**, indicating that 99.17% of the variation is explained by the model.
- **p-value (slope) = 6.472e–15**, which confirms high statistical significance.
- **Interpretation**: For every 1,000 new enterprises, the number of bachelor-level R&D personnel increases by approximately 79 people.

Master's Degree

- **Model**: *Personnel = –42,553.22 + 0.01573 × Enterprises*
- **R² = 0.9696**, showing a strong model fit.
- **p-value (slope) = 3.001e–11**
- **Interpretation**: A 1,000-unit increase in enterprises is associated with an increase of around 16 master's-level personnel.

Associate Degree

- **Model**: *Personnel = –19,501.87 + 0.00782 × Enterprises*
- **R² = 0.9490**
- **p-value (slope) = 8.872e–10**
- **Interpretation**: The number of associate-level personnel shows a strong and significant increase in parallel with enterprise growth.

Doctorate or Higher

- **Model**: *Personnel = –8,040.64 + 0.00302 × Enterprises*
- **R² = 0.9247**
- **p-value (slope) = 1.119e–08**
- **Interpretation**: An increase of 1,000 enterprises results in approximately 3 additional doctorate-level R&D personnel.

High School

- **Model**: *Personnel = –6,685.04 + 0.00437 × Enterprises*

- **R² = 0.8888**

- **p-value (slope) = 1.441e–07**

- **Interpretation**: Although relatively lower than other education levels, the relationship remains positive and statistically significant.

All models exhibit high R² values and statistically significant slope coefficients, indicating that the number of industrial enterprises is a strong predictor of R&D personnel employment across all education levels. The strongest relationships are observed at the bachelor's and master's levels, while the high school and doctorate levels also show substantial, though relatively smaller, effects. These findings emphasize that industrial growth drives the demand for qualified human capital in the R&D sector, highlighting the importance of aligning educational planning with industrial expansion.

3.3 Model Fitting

Multivariate Linear Regression Analysis

This section presents multivariate linear regression models to estimate the number of doctorate-level R&D personnel in Turkiye between 2009 and 2023, using various explanatory variables such as total number of enterprises, gender distribution, and other education levels.

Base Model: Enterprises + Gender

```
Call:
lm(formula = `Doctorate or Higher` ~ `Total Number of Enterprises` +
    Male + Female, data = veri_model)

Residuals:
    Min      1Q  Median      3Q     Max
-407.45 -199.81  -24.95   89.87  689.49

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   3.428e+03  5.939e+03   0.577   0.5754
`Total Number of Enterprises` -1.324e-03  1.937e-03  -0.684   0.5084
Male                          8.882e-02  4.527e-02   1.962   0.0756 .
Female                       -7.792e-02  1.372e-01  -0.568   0.5814
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 323.6 on 11 degrees of freedom
Multiple R-squared:  0.9643,    Adjusted R-squared:  0.9545
F-statistic: 98.96 on 3 and 11 DF,  p-value: 3.058e-08
```

Regression Equation:

$$\text{PhD} = 3428 - 0.00132 \cdot \text{Enterprises} + 0.0888 \cdot \text{Male} - 0.0779 \cdot \text{Female}$$

Interpretation of Coefficients:

Variable	Coefficient	p-value	Significance
Intercept	3428	0.575	Not significant
Total Enterprises	-0.00132	0.508	Not significant

Variable	Coefficient	p-value	Significance
Male	+0.0888	0.076	Marginally significant (p 0.0756)
Female	-0.0779	0.0581	Not significant

Model Performance:

- $R^2 = 0.9643$

- Adjusted $R^2 = 0.9545$

- F-statistic = 98.96, p < 0.001

- Residual standard error = 323.6

The multiple linear regression model is constructed to estimate the number of R&D personnel at doctoral level. The number of industrial enterprises, number of male and female personnel are used as explanatory variables in the model. The overall fit of the model is quite strong ($R^2 = 0.9643$) and most of the variance can be explained. However, when individual variables are analyzed, only the variable "number of male staff" is found to be borderline significant at the 10% significance level ($p = 0.0756$). This result may indicate a multicollinearity problem due to the inclusion of the number of men and women in the model together.

Extended Model: Education Levels + Female Ratio

This model is constructed to analyze how the number of R&D personnel at the PhD level is related to the number of industrial enterprises, the number of personnel at other levels of education and the share of women.

```
Call:
lm(formula = `Doctorate or Higher` ~ `Total Number of Enterprises` +
    `Master's Degree` + `Bachelor's Degree` + `Associate Degree` +
    Other + Female_Ratio, data = veri_model)

Residuals:
    Min      1Q  Median      3Q     Max
-351.65 -103.67    6.41  142.64  280.30

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    8.832e+03  8.749e+03   1.010   0.3423
`Total Number of Enterprises` -3.611e-03  1.919e-03  -1.882   0.0966 .
`Master's Degree`              5.107e-01  2.281e-01   2.239   0.0555 .
`Bachelor's Degree`            3.393e-02  2.437e-02   1.392   0.2013
`Associate Degree`            -6.003e-01  2.956e-01  -2.031   0.0768 .
Other                          3.568e-03  6.697e-01   0.005   0.9959
Female_Ratio                   1.401e+04  2.001e+04   0.700   0.5036
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 246.3 on 8 degrees of freedom
Multiple R-squared:  0.9849,    Adjusted R-squared:  0.9737
F-statistic: 87.23 on 6 and 8 DF,  p-value: 7.522e-07
```

Regression Equation:

$$\text{PhD} = 8832 - 0.0036 \cdot \text{Enterprises} + 0.5107 \cdot \text{Master's}$$
$$+ 0.0339 \cdot \text{Bachelor's} - 0.6003 \cdot \text{Associate}$$
$$+ 0.0036 \cdot \text{Other} + 14014 \cdot \text{Female}$$

Interpretation of Coefficients:

Variable	Coefficient	p-value	Significance
Intercept	8,832	0.342	Not significant
Enterprises	−0.0036	0.0966	Marginally significant (10%)
Master's Degree	+0.5107	0.0555	Marginally significant (5%)
Bachelor's Degree	+0.0339	0.201	Not significant
Associate Degree	−0.6003	0.0768	Marginally significant (10%)
Other	+0.0036	0.996	Not significant
Female Ratio	+14,014	0.504	High effect size but NS

Model Performance:

- $R^2 = 0.9849$
- Adjusted $R^2 = 0.9737$
- F-statistic = 87.23, p < 0.001
- Residual standard error = 246.3

This model analyzes the relationship between the number of PhD-level R&D personnel and entrepreneurship and personnel structure. The model has a high explanatory power with an R² value of 98.5%. While master's degree graduates have a positive and borderline significant effect, the effect of associate degree graduates is negative and borderline significant. Although the female ratio has a high effect coefficient, it is not statistically significant. This suggests that the interactions between education levels are determinant in the transition to doctoral level and that the gender effect is not direct.

Multivariate analysis confirms that PhD-level personnel are influenced more by internal structural factors such as the number of master's degree holders, rather than external macro indicators like the total number of enterprises. To improve interpretability, future models could apply variance inflation factor (VIF) analysis or dimensionality reduction techniques to mitigate multicollinearity effects.

3.4 Results

Two different multiple linear regression models were constructed in the analysis:

- Model 1: Basic model where the number of PhD-level employees is explained by the number of industrial enterprises and gender (male/female) distribution.

- Model 2: Expanded model including education levels and the share of women.

```
Call:
lm(formula = `Doctorate or Higher` ~ `Total Number of Enterprises` +
    Female + Male, data = veri_model)

Residuals:
    Min      1Q  Median      3Q     Max
-407.45 -199.81  -24.95   89.87  689.49

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   3.428e+03  5.939e+03   0.577   0.5754
`Total Number of Enterprises` -1.324e-03  1.937e-03  -0.684   0.5084
Female                        -7.792e-02  1.372e-01  -0.568   0.5814
Male                           8.882e-02  4.527e-02   1.962   0.0756 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 323.6 on 11 degrees of freedom
Multiple R-squared:  0.9643,    Adjusted R-squared:  0.9545
F-statistic: 98.96 on 3 and 11 DF,  p-value: 3.058e-08
```

```
Call:
lm(formula = `Doctorate or Higher` ~ `Total Number of Enterprises` +
    `Master's Degree` + `Bachelor's Degree` + `Associate Degree` +
    Other + Female_Ratio, data = veri_model)

Residuals:
    Min      1Q  Median      3Q     Max
-351.65 -103.67    6.41  142.64  280.30

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    8.832e+03  8.749e+03   1.010   0.3423
`Total Number of Enterprises` -3.611e-03  1.919e-03  -1.882   0.0966 .
`Master's Degree`              5.107e-01  2.281e-01   2.239   0.0555 .
`Bachelor's Degree`            3.393e-02  2.437e-02   1.392   0.2013
`Associate Degree`            -6.003e-01  2.956e-01  -2.031   0.0768 .
Other                          3.568e-03  6.697e-01   0.005   0.9959
Female_Ratio                   1.401e+04  2.001e+04   0.700   0.5036
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 246.3 on 8 degrees of freedom
Multiple R-squared:  0.9849,    Adjusted R-squared:  0.9737
F-statistic: 87.23 on 6 and 8 DF,  p-value: 7.522e-07


             df      AIC
model_multi1  5 221.3032
model_multi2  8 214.3404
```

Table 3: Comparison of Regression Model Performance

Model	R^2	Adjusted R^2	AIC
Model 1	0.9643	0.9545	221.3032
Model 2	0.9849	0.9737	214.3404

As a result of the multiple linear regression analyses, it is observed that both models exhibit high performance in explaining the number of R&D personnel at doctoral level. Model 1 is based on the number of industrial enterprises and gender variables and stands out as a strong base model with an R² value above 96%. However, Model 2 includes additional structural variables such as variables related to education (bachelor's degree, master's degree, associate's degree, etc.) and the proportion of female employees; thus, it offers higher values in terms of both R² and Adjusted R², while also reaching a lower (i.e. more preferable) value in terms of the AIC criterion. This indicates that Model 2 is more explanatory not only because it includes more variables but also because these variables contribute significantly to the model.

The lower AIC value of Model 2 reveals that the estimates are more reliable in terms of generalizability. These results suggest that the R&D staff structure at the PhD level should be evaluated not only with macro indicators such as the level of entrepreneurship or the total number of staff, but also with microstructural factors such as gender balance and education level. Therefore, it is recommended to prefer Model 2 in order to build more realistic, consistent and politically oriented models for strategic planning and decision-making processes.

4. Results and Key Takeaways

This study analyzed the distribution of R&D personnel working in the Turkish industrial sector according to their education levels using multiple linear regression models and revealed the determinants of the structure of PhD-level personnel. As a result of the comparison of two different models, the extended model (Model 2), which includes education-based variables and gender ratio, stands out with higher explanatory power and

lower information loss. This suggests that R&D personnel should be considered not only with macro indicators (e.g. number of enterprises) but also with micro structural factors (e.g. education level, gender ratio).

The model results show that especially master's degree graduates have a decisive effect on the transition to PhD level. It is also noteworthy that PhD-level employment may be higher in institutional structures with a higher proportion of women. In light of these analyses, it is necessary to increase R&D incentives at the doctoral level, develop interface policies to facilitate the integration of master's degree graduates into the industry, define norms for qualified R&D personnel per enterprise, and create structures to support the career transformation of bachelor's degree graduates.

Overall, the findings suggest that the Turkish industry still relies on an undergraduate-based R&D workforce structure, but this structure has limitations in terms of technological breakthrough and high value-added production. In this context, policies aimed at both increasing the quality of human capital and strengthening the structural interaction between industry and academia are critical for building sustainable R&D capacity.

References

- Fox, J. (2015). *Applied regression analysis and generalized linear models* (3rd ed.). Sage Publications.

- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.