Fertility and the Demographic Transition: An Analysis of Swiss Cantonal Fertility Patterns (1888)

Faculty of Economics and Business

University of Neuchâtel

Supervisor:

[redacted]

Program:

Bachelor of Science in Economics and Business

Authors and Student Numbers:

[redacted]

Table of content

Abstract 3
1. Introduction 3
2. Data 3
2.1 Descriptive overview: 3
3. Methodology 4
4. Results 5
4.1 Simple regressions..5
4.2 Multiple regressions 6
4. Conclusion...7
5. Bibliography...8
6. Appendix 9

FIGURE 1 : HISTOGRAM AND BOXPLOT OF FERTILITY 4
*FIGURE 2 : SIMPLE REGRESSION PLOT: FERTILITY VS EDUCATION.................. **ERROR! BOOKMARK NOT DEFINED.***
*FIGURE 3 : RESIDUALS VS FITTED: EDUCATION .... **ERROR! BOOKMARK NOT DEFINED.***
FIGURE 4 : DATA EXPLORATION 9
FIGURE 5 : HISTOGRAM AND BOXPLOT: EDUCATION 9
FIGURE 6 : HISTOGRAM AND BOXPLOT: EXAMINATION 9
FIGURE 7 : HISTOGRAM AND BOXPLOT : INFANT MORTALITY 10
FIGURE 8 : HISTOGRAM AND BOXPLOT : CATHOLIC 10
FIGURE 9 : RESIDUALS VS FITTED: URBANIZATION.. 11
FIGURE 10 : SCATTERPLOT: FERTILITY VS URBANIZATION 11
FIGURE 11 : RESIDUALS VS FITTED: EDUCATION 12
FIGURE 12 : SCATTERPLOT FERTILITY VS EDUCATION.. 12
FIGURE 13 : SCATTERPLOT FERTILITY VS EXAMINATION 13
FIGURE 14 : RESIDUALS VS FITTED: EXAMINATION 13
FIGURE 15 : RESIDUALS VS FITTED: CATHOLIC 14
FIGURE 16 : SCATTERPLOT FERTILITY VS CATHOLIC 14
FIGURE 17: RESIDUALS VS FITTED: MORTALITY 15
FIGURE 18 : SCATTERPLOT FERTILITY VS MORTALITY 15
FIGURE 19 : RESIDUALS VS FITTED: URB + EDU.. 16
FIGURE 20 : RESIDUALS VS FITTED: + EXA 17
FIGURE 21 : RESIDUALS VS FITTED: + CATHOLIC 18
FIGURE 22 : RESIDUALS VS FITTED: + FULL MODEL 19

Abstract

In this analysis we study the determinants of fertility across French-speaking Swiss cantons using historical data. Inspired by Gary Becker's study of the Demographic Transition, we analyze how several factors affect fertility rates across a set of cantons, with a justified focus on education and urbanization. Regression analysis reveals evidence of omitted variable bias when education is excluded. Our findings confirm the central role of education in Switzerland's early fertility decline.

1. Introduction

In this report, we will look at the determinants of fertility across French-speaking cantons of Switzerland during the demographic transition. This process represents a widespread demographic shift observed across most developed societies, typically marked by declining birth rates and changing family structures. Looking back to the year 1888, we analyze how rurality (determined by the share of the population working in agriculture) and education influenced fertility rates. By using a cross-sectional dataset from the R datasets package, we apply the ordinary least squares (OLS) regression to make an estimate of potential relationships. Several economic theories suggest that rising education levels increase the opportunity cost of having a child, thus provoking a reduction in fertility. As Becker argues in one of his major works on fertility, the decline of fertility could be explained by rational household decisions maximizing utility under constraints, where the cost of raising children increases with modernization. We also study whether excluding education from the model could possibly lead us to Omitted Variable Bias (OVB), specifically in estimating the effect of rurality on fertility.

2. Data

We obtained a dataset from R's datasets package containing the necessary data dated from 1888. The dataset consists of 47 observations on 6 variables. To conduct our analyses, we assumed that the data represents a random sample of the populations of 47 cities through the French-speaking cantons of Switzerland at about 1888. The 6 variables are all quantitative continuous variables and measure the population's participation share as a percentage from 0 to 100. We will be focusing our analysis on the standardized fertility measure and socioeconomic indicators which could have a potential explanation for the decline in fertility during the demographic transition. We gathered the dataset on these 5 variables to explore their potential influence on fertility, (our key variable) adding one after the other. To conduct our analysis, we decided to add another variable called Urbanization (not part of our dataset), which we constructed by using the variable Agriculture. All variables but Fertility give proportions of the population and they were all taken and computed by the Federal Statistical Office.

2.1 Descriptive overview:

- **Fertility (Y)** is a quantitative variable which represents the fertility level of a given population expressed as a percentage in the dataset (range: 35.00 – 92.5, mean: 70.14).
- **Urbanization (x_1)** is computed as 100 minus the share in agriculture (range: 10.30 – 98.8, mean: 49.34).

- **Education (x_2)** is the percentage of draftees with education beyond primary school (range: 1.00 – 53.00, mean: 10.98).
- **Examination (x_3)** measures the percentage of draftees in the military receiving the highest mark on army examination (range: 3.00 – 37.00, mean: 16.49).
- **Catholic (x_4)** is the share of the Catholic population (range: 2.15 – 100.00, mean: 41.14).
- **Infant Mortality (x_5)** measures infant deaths per 100 births (range: 10.80 – 26.60, mean: 19.94).
- **Agriculture** (not used directly) is the percentage of males of the population involved in agriculture as occupation. It is obtained by dividing the number of men employed in agriculture by the total number of active men (range: 1.20 – 89.70, mean: 50.66).

Distribution of Fertility Levels in 1888:

Most values lie between 60 and 80, with a slight right skew and a few low outliers.

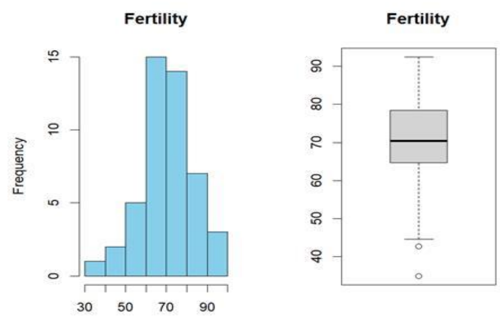

Figure 1: Histogram and Boxplot of Fertility

3. Methodology

The OLS estimator models the linear relationship between fertility (Y) and explanatory variables by minimizing the sum of squared residuals. It finds the best-fitting line through the data, estimating unknown population parameters from the sample, using the general equation for simple regression: $\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X_i + u_i$. Thus, we measure the partial effect of a given independent variable X1 on the dependent variable Y, captured by **β_1**, holding other variables constant: $\Delta Y = \beta_1 \cdot \Delta X_1$; when all other factors are fixed:

$$\hat{\beta}_1 = \frac{\Delta Y}{\Delta X}$$

Our model selection proceeds by gradually expanding the set of regressors and tracking how the estimated coefficients evolve across the addition and subtraction of independent variables.

Special emphasis is placed on the identification of **Omitted Variable Bias (OVB)**, defined by the equation: $\widehat{\beta_1} = \beta_1 + \rho X u \frac{\sigma_u}{\sigma_X}$. OVB is present when an important explanator variable is left out of the regression analysis. It must fulfill two conditions:

1. The included regressor (X) must be correlated with the omitted variable (Z):

2. the omitted variable Z must itself affect the dependent variable $Y: \beta_2 \neq 0$. OVB impacts the OLS estimator, rendering it **biased and inconsistent**

To correct for omitted variable bias, we include the omitted variable as an additional regressor in the linear model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + u_i$

4. Results

4.1 Simple regressions

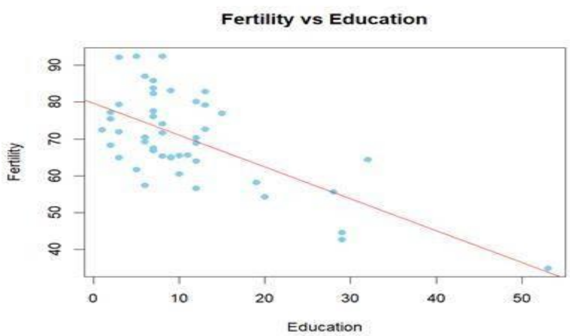

Figure 2: Simple Regression Plot: Fertility vs Education

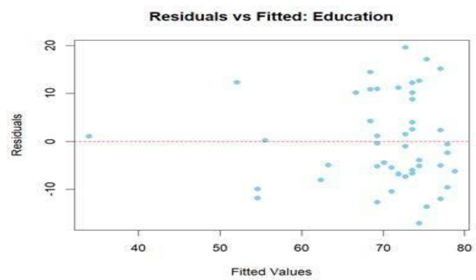

Figure 3: Residuals vs Fitted: Education

$$\widehat{Fertility_i} = 79.61 - 0.862 \cdot Education_i$$

$$\widehat{Fertility_i} = 86.82 - 1.011 \cdot Examination_i$$

$$\widehat{Fertility_i} = 64.43 + 0.139 \cdot Catholic_i$$

$$\widehat{Fertility_i} = 34.52 + 1.787 \cdot InfantMortality_i$$

Model	Coefficient	Significance	R²
Fertility ~ Urbanization	−0.194	$p = 0.015$ *	0.105
Fertility ~ Education	−0.862	$p < 0.001$ ***	0.428
Fertility ~ Examination	−1.011	$p < 0.001$ ***	0.404
Fertility ~ Catholic	+0.139	$p = 0.001$ ***	0.198
Fertility ~ Infant Mort.	+1.787	$p = 0.004$ **	0.155

With each model iteration, we observe:

1. Changes in the **magnitude and sign** of key coefficients (e.g., Urbanization)
2. Shifts in **statistical significance** (p-values)
3. Improvements in **model fit** (adjusted **R^2**): we will refer to the adjusted **R^2** as it penalizes the non-parsimonious addition of variables: $\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\left(\frac{SSR}{TSS}\right)$

Regarding OVB, it cannot simply be resolved by increasing the number of observations used to estimate β_1. Particularly, the estimated effect of X_1 on Y may be overstated or understated depending on the direction of the correlation. Moreover, OVB results in the violation of the first OLS assumption $E(u_i|X_i) = 0$. In effect, the error term captures all other factors that affect Y but are not included in the model.

4.2 Multiple regressions

Model 1: Urbanization + Education

$$Fertility_i = 77.43 + 0.066 \cdot Urbanization_i - 0.963 \cdot Education_i$$

Adding Education reverses the sign and removes the significance of the Urbanization coefficient, indicating omitted variable bias. Urbanization and Education are positively correlated, while Education negatively affects Fertility. As a result, the negative effect of Education was mistakenly captured by Urbanization in the simple model.

Model 2 (Full Model): Urbanization + Education + Examination + Catholic + Infant Mortality

$$Fertility = 49.70 + 0.172 \cdot Urbanization - 0.871 \cdot Education - 0.258 \cdot Examiation + 0.104 \cdot Catholic + 1.077\ InfantMortality$$

Model	Urbanization	Education	Examinatio n	Catholic	Infant Mortality	Adj. R^2
M1: Urban. + Education	+0.066	−0.963	—	—	—	0.424
	(p = 0.41)	(p < 0.001 ***)				
M2: + (Full Model)	+0.172	−0.871	−0.258	+0.104	+1.077	0.671
	(p = 0.019 *)	(p < 0.001 ***)	(p = 0.315)	(p = 0.005 **)	(p = 0.007 **)	

Additional regression models using three and four variables (e.g., Education + Urbanization + Catholic) are presented in the appendix, along with other specifications exploring alternative variable combinations (fig.26, 27, 28, 29, 30, 31, 32).

Final model interpretation:

(**Adjusted R^2 = 0.671**) The final model explains **67.1%** of the variation in fertility across Swiss cantons. This is a substantial improvement over the simple regressions, where R^2 ranged between 0.105 and 0.428, and also over intermediate multivariate models (e.g., Model 1 with adj. $R^2 = 0.424$). This improvement in explanatory power confirms that including all relevant socioeconomic and demographic factors significantly enhances model fit.

- **Education** has the strongest and most consistent effect: its coefficient is **−0.871** and highly significant (**$p < 0.001$**). This confirms that higher education levels are strongly linked to lower fertility. According to the final model, a 1 percentage point increase in education is associated with a 0.871-point decrease in the fertility index, holding other variables constant.
- **Urbanization**, which was negatively associated with fertility in the simple model, becomes **positive (+0.172, $p = 0.019$)** after controlling for education. This sign reversal reveals **omitted variable bias**: urban areas also tend to be more educated, and the effect of education was wrongly attributed to urbanization in simpler models.
- **Catholic** population share has a **positive effect (+0.104, $p = 0.005$)**, suggesting that more Catholic regions had higher fertility.
- **Infant Mortality** also has a **positive and significant effect (+1.077, $p = 0.007$)**, which may reflect compensatory fertility in high-mortality regions.

4. Conclusion

In conclusion, our analysis highlights that during the demographic transition education had a key role in the reduction of the fertility. At first, urbanization seemed to be an important factor, and we were able to see that after adding other variables, such as education, religion and infant mortality, a presence of omitted variable bias (OVB) appeared. This occurred even more when education is left out of the model.

Our findings support the theory that a higher level of education increases the opportunity cost of having children, leading to lower fertility rates. Moreover, the effect we observed on fertility can partly be explained by other socioeconomics variables, which can be seen in the prolonged models.

However, the analysis has certain limitations, such as the lack of income data. In fact, this is an essential part of the opportunity cost analysis, thus, not having these data limits our ability to fully assess the effects on fertility. Another limitation is the fact that we observed Heteroskedasticity throughout the models, which suggests that the variance of the residuals is not constant across all levels of the independent variable. This may affect the reliability of the estimated standard error.

Finaly, our analysis empathizes the importance of taking key socioeconomic factors into account to avoid biased interpretation and shows fundamentals economics sources on Switzerland's fertility regression.

5. Bibliography

Federal Statistical Office. (1888). *Swiss fertility and socioeconomic indicators (1888) data*. [Dataset]. Retrieved from https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/swiss.html

Bachfischer, M. (2021, March 16). Analysis of Swiss dataset in R. Matthias Bachfischer Blog. Retrieved from https://bachfischer.me/posts/2021/03/analysis_of_swiss_dataset_in_r

Dhar, P. (2017). *Comparative study of clustering techniques on Swiss dataset*. International Journal of Statistical Sciences. Retrieved from https://www.ripublication.com/ijss17/ijssv12n1_07.pdf

6. Appendix

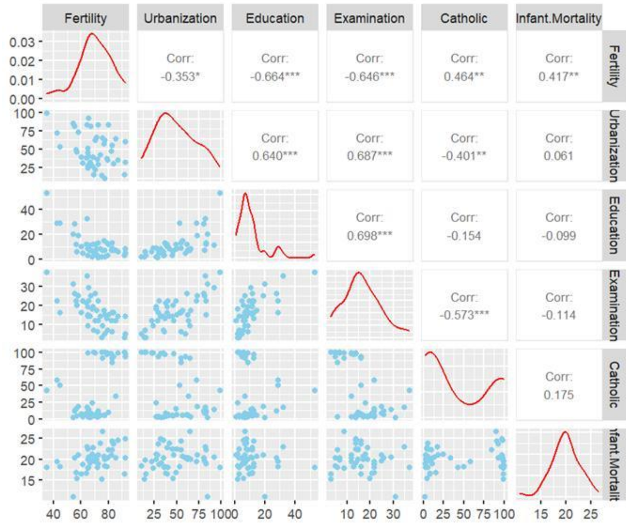

Figure 4: Data exploration

Dataset Structure and Variable Overview

Histograms:

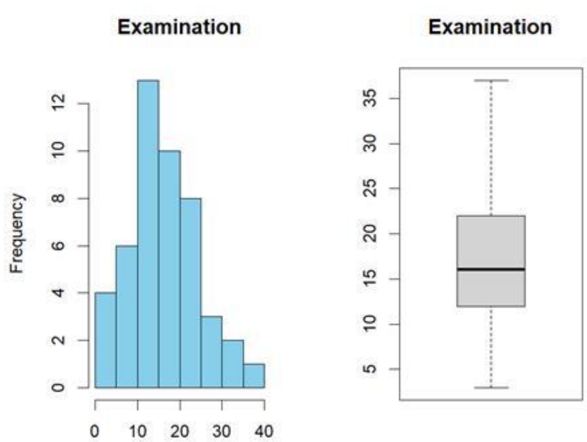

Figure 5 : Histogram and boxplot: Examination

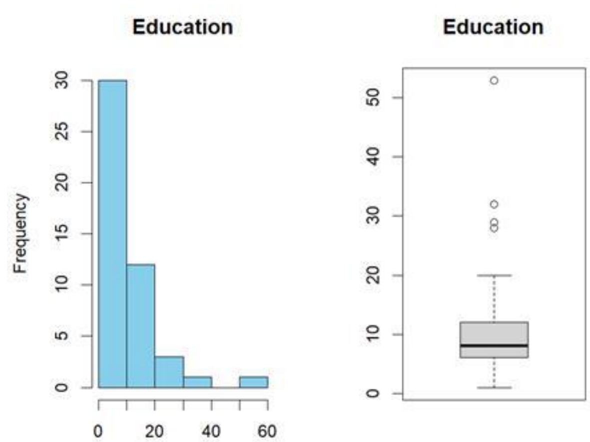

Figure 6 : Histogram and boxplot: Education

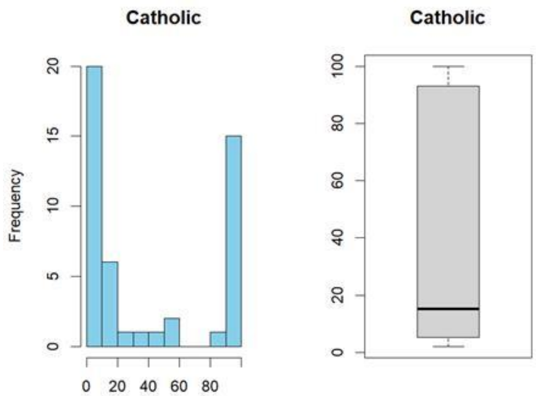

Figure 7 : Histogram and boxplot : Catholic

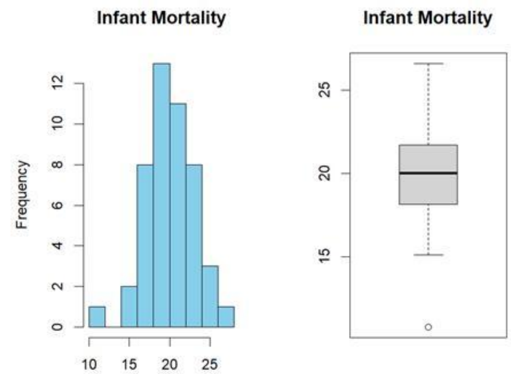

Figure 8 : Histogram and boxplot : Infant Mortality

SIMPLE REGRESSION OUTPUT

Fertility – Urbanization

```
lm(formula = Fertility ~ Urbanization, data = swiss)

Residuals:
     Min       1Q   Median       3Q      Max 
-25.5374  -7.8685  -0.6362   9.0464  24.4858 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  79.72455    4.15895  19.169   <2e-16 ***
Urbanization -0.19420    0.07671  -2.532   0.0149 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.82 on 45 degrees of freedom
Multiple R-squared:  0.1247,	Adjusted R-squared:  0.1052 
F-statistic: 6.409 on 1 and 45 DF,  p-value: 0.01492
```

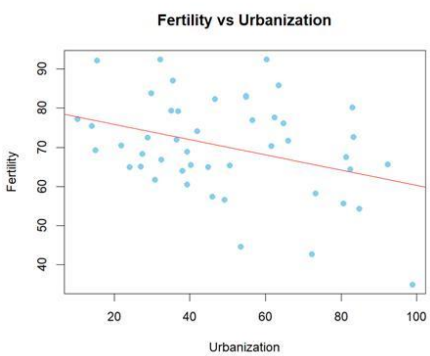

Figure 9 : Scatterplot: Fertility vs Urbanization

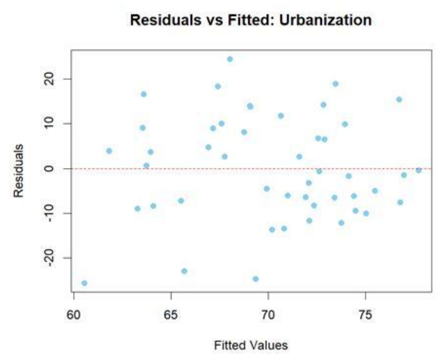

Figure 10 : Residuals vs Fitted: Urbanization

Fertility – Education

```
Call:
lm(formula = Fertility ~ Education, data = swiss)

Residuals:
    Min      1Q  Median      3Q     Max 
-17.036  -6.711  -1.011   9.526  19.689 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  79.6101     2.1041  37.836  < 2e-16 ***
Education    -0.8624     0.1448  -5.954 3.66e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.446 on 45 degrees of freedom
Multiple R-squared:  0.4406,	Adjusted R-squared:  0.4282 
F-statistic: 35.45 on 1 and 45 DF,  p-value: 3.659e-07
```

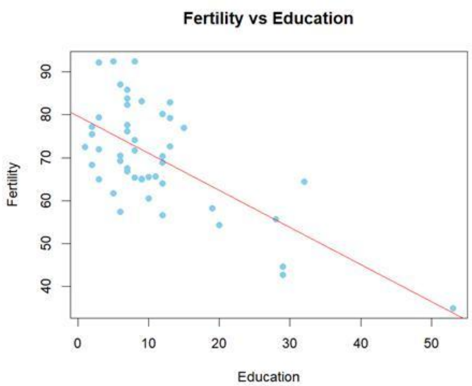

Figure 11 : Scatterplot Fertility vs Education

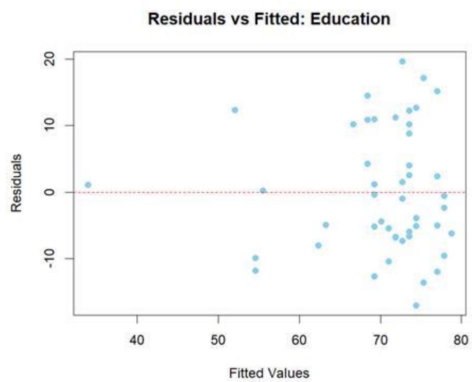

Figure 12 : Residuals vs Fitted: Education

Fertility – Examination

```
Call:
lm(formula = Fertility ~ Examination, data = swiss)

Residuals:
     Min       1Q   Median       3Q      Max 
-25.9375  -6.0044  -0.3393   7.9239  19.7399 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  86.8185     3.2576  26.651  < 2e-16 ***
Examination  -1.0113     0.1782  -5.675 9.45e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.642 on 45 degrees of freedom
Multiple R-squared:  0.4172,	Adjusted R-squared:  0.4042 
F-statistic: 32.21 on 1 and 45 DF,  p-value: 9.45e-07
```

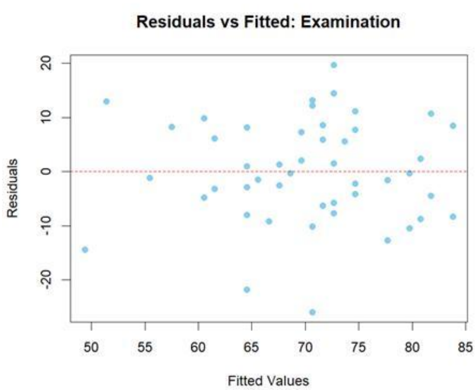

Figure 13 : Residuals vs Fitted: Examination

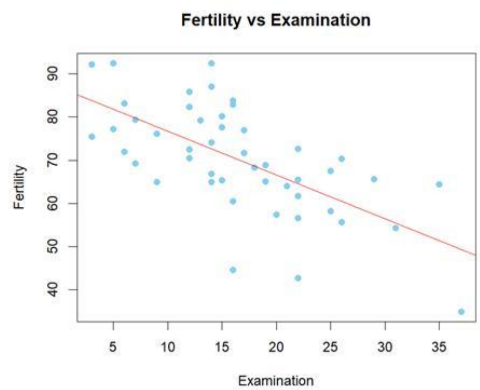

Figure 14 : Scatterplot Fertility vs Examination

Fertility – Catholic

```
Call:
lm(formula = Fertility ~ Catholic, data = swiss)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.309  -4.060   0.511   6.851  16.682 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 64.42826    2.30510  27.950  < 2e-16 ***
Catholic     0.13889    0.03956   3.511  0.00103 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.19 on 45 degrees of freedom
Multiple R-squared:  0.215,	Adjusted R-squared:  0.1976 
F-statistic: 12.33 on 1 and 45 DF,  p-value: 0.001029
```

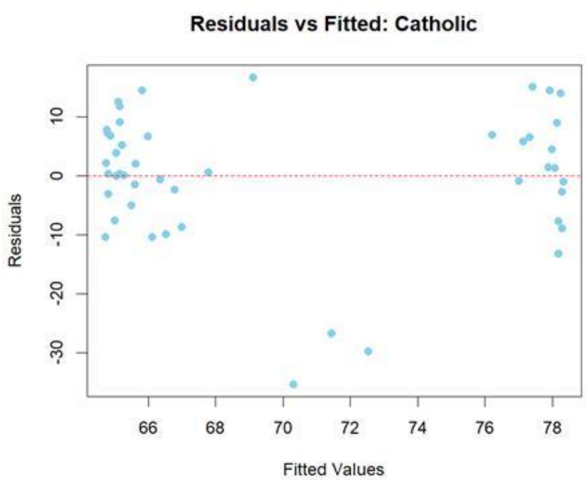

Figure 15 3: Residuals vs Fitted: Catholic

4

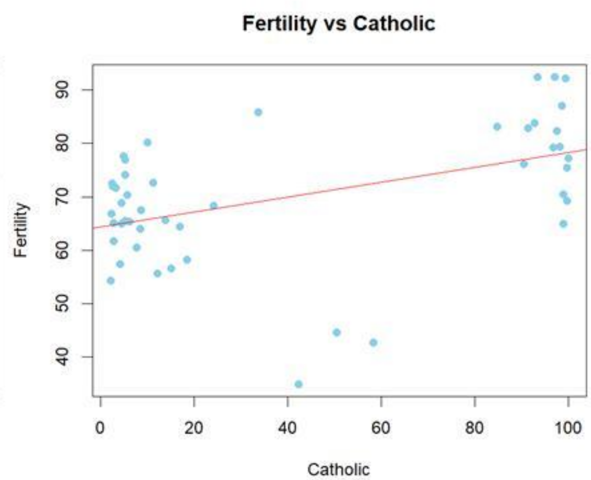

Figure 161 : Scatterplot Fertility vs Catholic

2

Fertility – Infant Mortality

```
Call:
lm(formula = Fertility ~ Infant.Mortality, data = swiss)

Residuals:
    Min      1Q  Median      3Q     Max 
-31.672  -5.687  -0.381   7.239  28.565 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)   
(Intercept)       34.5155    11.7113   2.947  0.00507 **
Infant.Mortality   1.7865     0.5812   3.074  0.00359 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.48 on 45 degrees of freedom
Multiple R-squared:  0.1735,	Adjusted R-squared:  0.1552 
F-statistic: 9.448 on 1 and 45 DF,  p-value: 0.003585
```

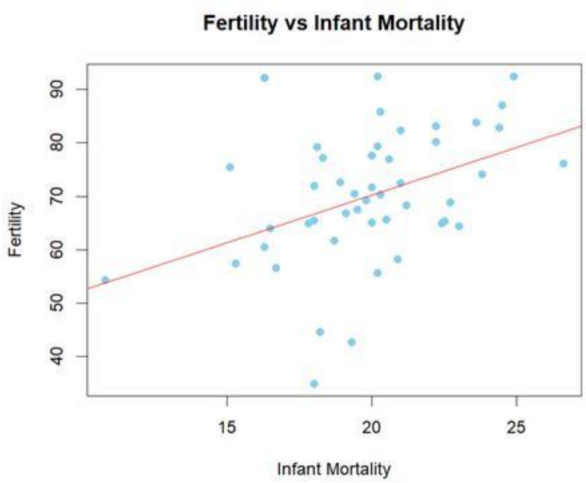

Figure 175 : Scatterplot Fertility vs Mortality

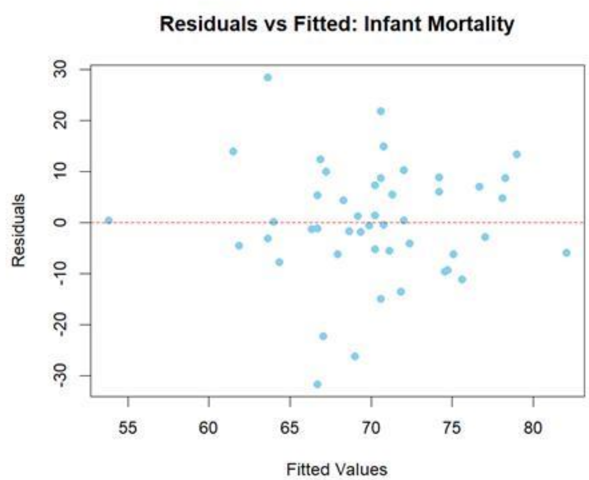

Figure 18: Residuals vs Fitted: Mortality

MULTIPLE REGRESSION OUTPUT

 # Model 1: Urbanization + Education

```
lm(formula = Fertility ~ Urbanization + Education, data = swiss)

Residuals:
     Min       1Q   Median       3Q      Max 
-17.3072  -6.6157  -0.9443   8.7028  20.5291 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  77.43255    3.36644  23.001  < 2e-16 ***
Urbanization  0.06648    0.08005   0.830    0.411    
Education    -0.96276    0.18906  -5.092  7.1e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.479 on 44 degrees of freedom
Multiple R-squared:  0.4492,    Adjusted R-squared:  0.4242 
F-statistic: 17.95 on 2 and 44 DF,  p-value: 2e-06
```

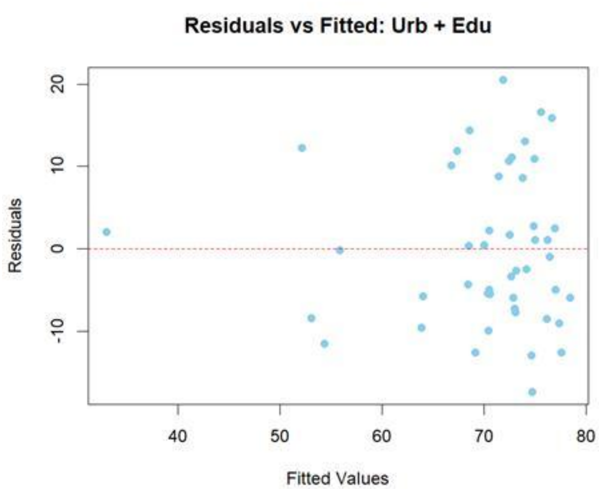

Figure 19: Residuals vs Fitted: Urb + Edu

Model 2: Urbanization + Education + Examination

```
lm(formula = Fertility ~ Urbanization + Education + Examination,
    data = swiss)

Residuals:
    Min      1Q  Median      3Q     Max 
-14.967  -4.978  -1.045   4.906  21.358 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  81.78413    3.33836  24.498  < 2e-16 ***
Urbanization  0.18017    0.08071   2.232  0.03084 *  
Education    -0.67242    0.19366  -3.472  0.00119 ** 
Examination  -0.79744    0.24679  -3.231  0.00237 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.601 on 43 degrees of freedom
Multiple R-squared:  0.5568,	Adjusted R-squared:  0.5259 
F-statistic: 18.01 on 3 and 43 DF,  p-value: 1.017e-07
```

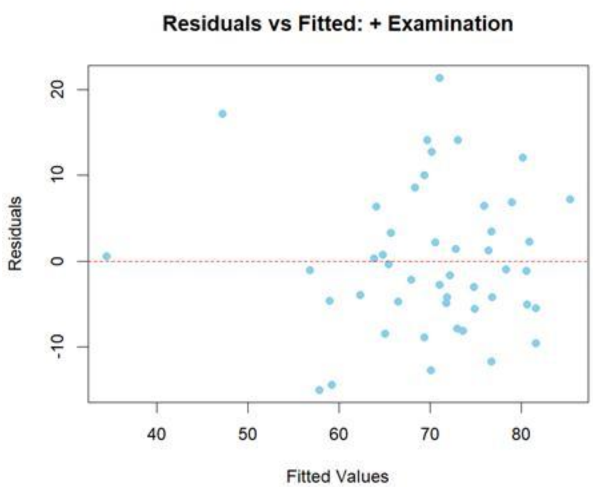

Figure 20: Residuals vs Fitted: + Exa

Model 3: Urbanization + Education + Examination + Catholic

```
lm(formula = Fertility ~ Urbanization + Education + Examination + 
    Catholic, data = swiss)

Residuals:
     Min       1Q   Median       3Q      Max 
-15.7813  -6.3308   0.8113   5.7205  15.5569 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  68.99087    4.86832  14.171  < 2e-16 ***
Urbanization  0.22065    0.07360   2.998  0.00455 ** 
Education    -0.96161    0.19455  -4.943 1.28e-05 ***
Examination  -0.26058    0.27411  -0.951  0.34722    
Catholic      0.12442    0.03727   3.339  0.00177 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.736 on 42 degrees of freedom
Multiple R-squared:  0.6498,	Adjusted R-squared:  0.6164 
F-statistic: 19.48 on 4 and 42 DF,  p-value: 3.95e-09
```

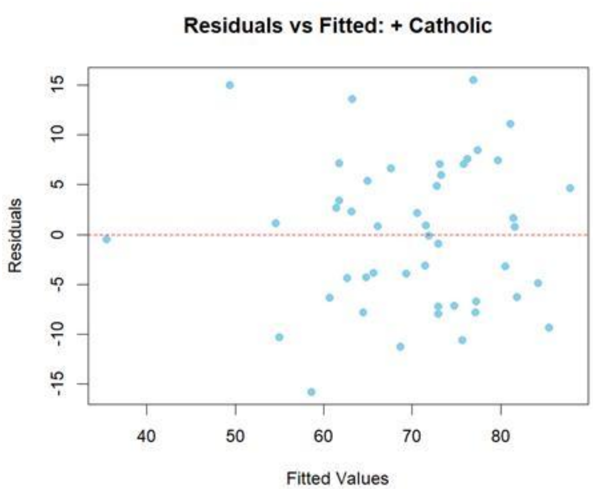

Figure 21: Residuals vs Fitted: + Catholic

Model 4: Full Model (Urbanization + Education + Examination + Catholic + Infant Mortality)

```
lm(formula = Fertility ~ Urbanization + Education + Examination +
    Catholic + Infant.Mortality, data = swiss)

Residuals:
     Min       1Q   Median       3Q      Max
-15.2743  -5.2617   0.5032   4.1198  15.3213

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      49.70378    8.18878   6.070 3.44e-07 ***
Urbanization      0.17211    0.07030   2.448  0.01873 *
Education        -0.87094    0.18303  -4.758 2.43e-05 ***
Examination      -0.25801    0.25388  -1.016  0.31546
Catholic          0.10412    0.03526   2.953  0.00519 **
Infant.Mortality  1.07705    0.38172   2.822  0.00734 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.165 on 41 degrees of freedom
Multiple R-squared:  0.7067,    Adjusted R-squared:  0.671
F-statistic: 19.76 on 5 and 41 DF,  p-value: 5.594e-10
```

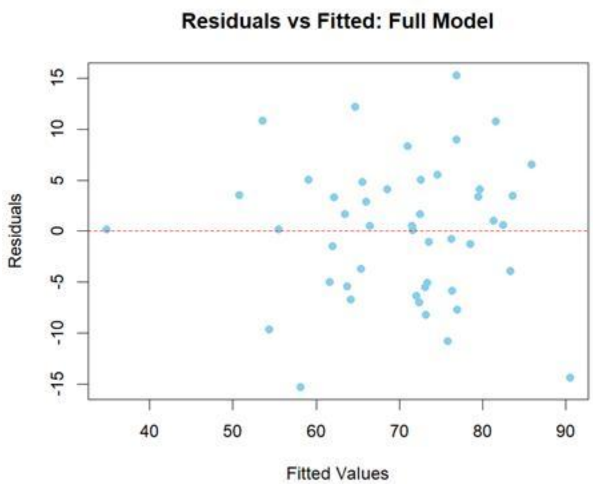

Figure 22: Residuals vs Fitted: + Full Model