The Effects of Aging Rate and Urbanization Rate on Birth Rate in China

Research Topic

This project analyzes the association between aging rate and birth rate as well as urbanization rate and birth rate.

The research uses the multiple linear regression model.

MATH 3544 W03

Probability and Math Statistics

郑皓予 HaoyuZheng 1337149

Research Content, Data Source, and Variables

Data Source

Source: National Bureau of Statistics of China

Data: provincial-level population data for 2023 and 2024

Final n in sample after deleting missing values: n = 62 observations

Variable Table

Variable	Meaning	Role
Birth Rate	Births per 1,000 population	Y / Dependent
Aging Rate	Share of population aged 65 and above	X_1 / Independent
Urbanization Rate	Urban population share	X_2 / Independent

Variable Calculation

Aging Rate

Population aged 65 and above/Sample population

Urbanization Rate

Urban population/Resident population

Method: Multiple Linear Regression

Model Specification

$$\text{Birth Rate} = \beta_0 + \beta_1 * \text{Aging Rate} + \beta_2 * \text{Urbanization Rate} + \varepsilon$$

Least Squares

Estimate parameters: estimate and figure out estimate values of β0, β1, β2

Aims: to minimize the value of

$SSE = \Sigma (Y_i - \hat{Y}_i)^2$

For accuracy, reduce the errors in prediction

t-tests

To test the coefficients in the model by t test respectively

H0: $\beta j = 0$

The p value <0.05 is significant

F-test and R^2

To test the F value in the overall regression

R^2 measures the explanation force or power of the regression

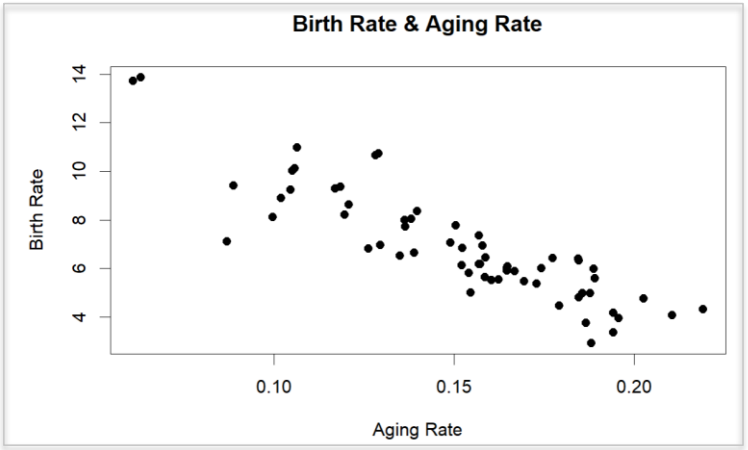
The adjusted R^2 considers the number of predictors

Data Visualization and Preliminary Patterns

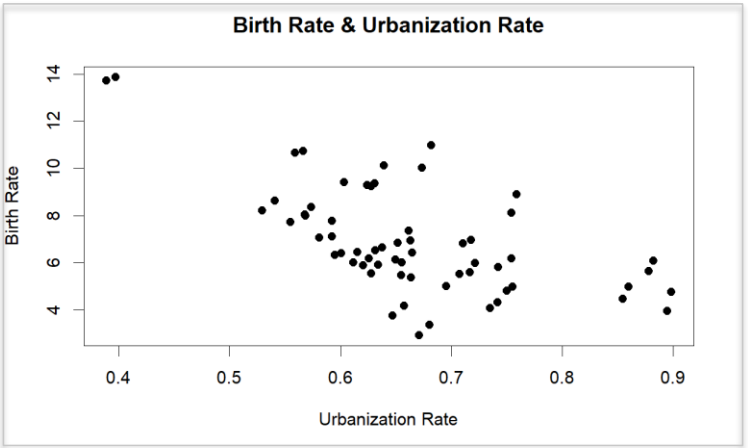
Data Preview

地区年份	人口出生率(‰)	城镇人口(万人)	65岁及以上人口数(抽样调查)(人	人口数(抽样调查)(人	常住人口(万人)	城市化率	老龄化率(抽样口径
北京市2024	6.09	1926	3703	22470	2183	0.882272103	0.164797508
北京市2023	5.63	1920	3650	23018	2186	0.87831656	0.158571553
天津市2024	4.99	1173	2606	14037	1364	0.859970674	0.185652205
天津市2023	4.47	1166	2572	14362	1364	0.85483871	0.179083693
河北省2024	5.9	4679	12498	75943	7378	0.634182705	0.164570797
河北省2023	5.54	4641	12640	77846	7393	0.627755985	0.162371862
山西省2024	6.94	2285	5598	35463	3446	0.663087638	0.157854665
山西省2023	6.13	2252	5549	36496	3466	0.649740335	0.15204406
内蒙古自治区2024	5.52	1689	3942	24578	2388	0.707286432	0.160387338
内蒙古自治区2023	5	1667	3899	25229	2396	0.695742905	0.154544374
辽宁省2024	4.32	3082	9366	42763	4155	0.741756919	0.219021116
辽宁省2023	4.06	3074	9273	44035	4182	0.735054998	0.210582491
吉林省2024	4.17	1524	4629	23845	2317	0.657747087	0.194128748
吉林省2023	3.76	1514	4594	24629	2339	0.647285165	0.186528077
黑龙江省2024	3.35	2061	6051	31174	3029	0.680422582	0.194104061
黑龙江省2023	2.92	2055	6065	32242	3062	0.67112998	0.188108678
上海市2024	4.75	2228	5169	25520	2480	0.898387097	0.202547022
上海市2023	3.95	2225	5123	26187	2487	0.894652191	0.19563142
江苏省2024	4.98	6440	16478	87756	8526	0.755336617	0.187770637
江苏省2023	4.81	6398	16569	89776	8526	0.750410509	0.184559348
浙江省2024	6.17	5033	10789	68651	6670	0.754572714	0.157157215
浙江省2023	5.8	4919	10756	69780	6627	0.742266486	0.154141588
安徽省2024	6.17	3831	9885	63020	6123	0.625673689	0.156854967
安徽省2023	6.45	3765	10228	64452	6121	0.615095573	0.15869174
福建省2024	6.95	3011	5591	43162	4193	0.718101598	0.129535239
福建省2023	6.81	2972	5561	44046	4183	0.71049486	0.12625437
江西省2024	6.65	2871	6435	46339	4502	0.63771657	0.138867908
江西省2023	6.52	2850	6411	47541	4515	0.631229236	0.134852022
山东省2024	6.42	6701	18400	103752	10080	0.664781746	0.177345979
山东省2023	6.01	6634	18577	106592	10123	0.655339326	0.174281372
河南省2024	7.78	5795	15144	100717	9785	0.59223301	0.150361905
河南省2023	7.06	5701	15389	103349	9815	0.580845644	0.148903231
湖北省2024	5.38	3871	10385	60048	5834	0.663524169	0.172944977
湖北省2023	5.48	3822	10418	61472	5838	0.654676259	0.169475534
湖南省2024	5.89	4059	11228	67303	6539	0.620737116	0.16682763
湖南省2023	6	4017	11386	69159	6568	0.611601705	0.164635116
广东省2024	8.89	9701	13399	131542	12780	0.759076682	0.101861003
广东省2023	8.12	9583	13330	133790	12706	0.754210609	0.099633754
广西壮族自治区2024	8.37	2877	7208	51594	5013	0.57390784	0.139706167
广西壮族自治区2023	8.04	2854	7310	52933	5027	0.567734235	0.138099106
海南省2024	9.37	661	1278	10785	1048	0.630725191	0.118497914
海南省2023	9.28	651	1284	10982	1043	0.624161074	0.116918594
重庆市2024	5.99	2301	6194	32828	3190	0.721316614	0.188680395
重庆市2023	5.58	2287	6353	33600	3191	0.716703228	0.189077381
四川省2024	6.41	5027	15878	86091	8364	0.601028216	0.184432751
四川省2023	6.32	4978	16268	88112	8368	0.594885277	0.184628654
贵州省2024	10.74	2187	5127	39732	3860	0.566580311	0.129039565
贵州省2023	10.65	2162	5216	40697	3865	0.559379043	0.128166695
云南省2024	8.62	2519	5783	47918	4655	0.541138561	0.120685337
云南省2023	8.22	2473	5880	49205	4673	0.529210357	0.119500051
西藏自治区2024	13.87	147	240	3812	370	0.397297297	0.062959077
西藏自治区2023	13.72	142	234	3843	365	0.389041096	0.06088993
陕西省2024	7.36	2615	6381	40691	3953	0.661522894	0.156816004
陕西省2023	6.83	2575	6338	41613	3952	0.651568826	0.152308173
甘肃省2024	8	1397	3449	25300	2458	0.568348251	0.136324111
甘肃省2023	7.71	1368	3542	25956	2465	0.554969574	0.136461704
青海省2024	10.11	379	646	6108	593	0.639123103	0.105762934
青海省2023	9.25	373	654	6255	594	0.627946128	0.104556355
宁夏回族自治区2024	10.97	497	798	7499	729	0.68175583	0.106414189
宁夏回族自治区2023	10.02	491	806	7676	729	0.673525377	0.105002606
新疆维吾尔自治区2024	9.42	1583	2393	26995	2623	0.603507434	0.088646046
新疆维吾尔自治区2023	7.1	1539	2376	27356	2598	0.592378753	0.086854803

Birth Rate & Aging Rate



Birth Rate & Urbanization Rate



Preliminary Observations

The birth rate and aging rate show a clear negative pattern.

The birth and urbanization have negative pattern, but they are more scattered.

The patterns confirm the further research of regression analysis.

Regression Results and Significance Tests

$$Y = 17.3164 - 49.3954(\text{Aging Rate}) - 4.4767(\text{Urbanization Rate})$$

Regression Output Table

Variable	Estimate	Std. Error	t value	p-value	Result
Intercept	17.3164	0.8769	19.747	< 2e-16	Significant
Aging Rate	-49.3954	4.4550	-11.088	4.67e-16	Significant
UrbanizationRate	-4.4767	1.5472	-2.893	0.00533	Significant

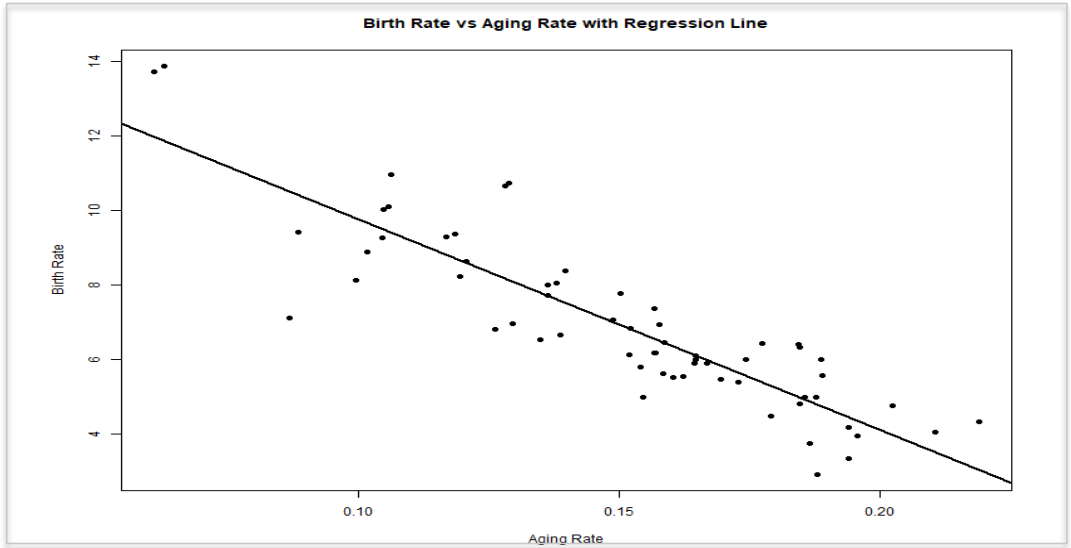
Model Fit

Statistic	Value
Residual Standard Error	1.033
R-squared	0.8008
Adjusted R-squared	0.7940
F-statistic	118.6
P-value(overall)	< 2.2e-16

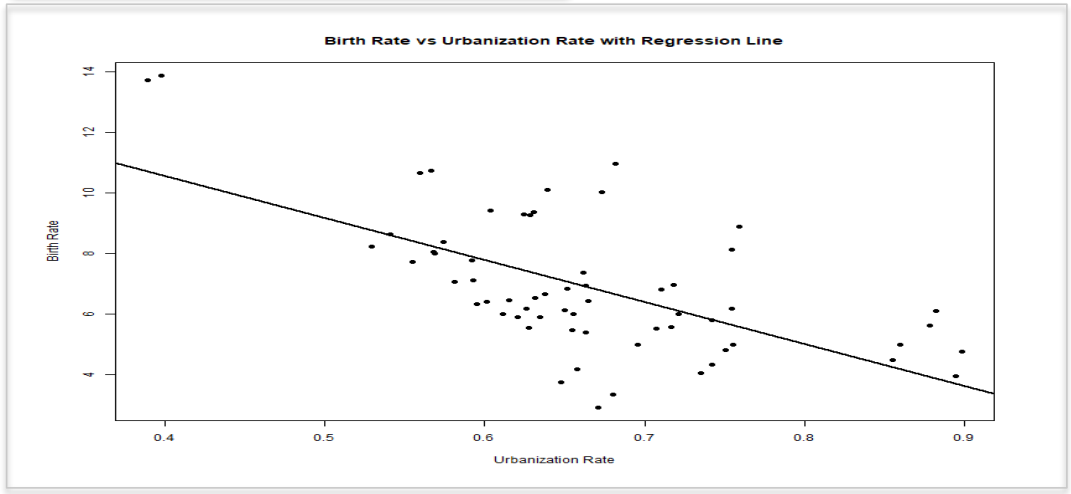
The coefficients are negative and significant; $R^2 = 0.8008$ and the overall p-value is < 2.2e-16.

Fitted Graphs: Trends and Regression Plane

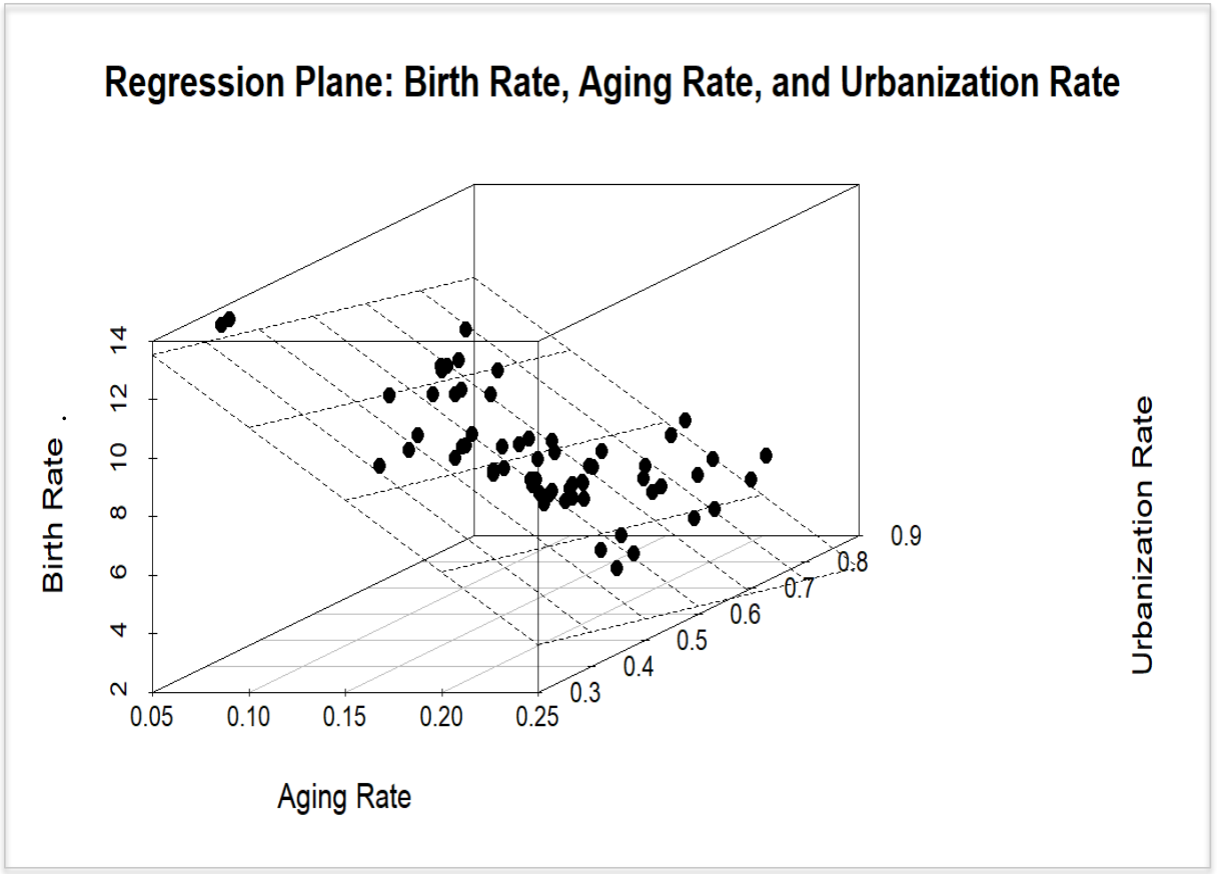
Aging rate fitted line



Urbanization rate fitted line

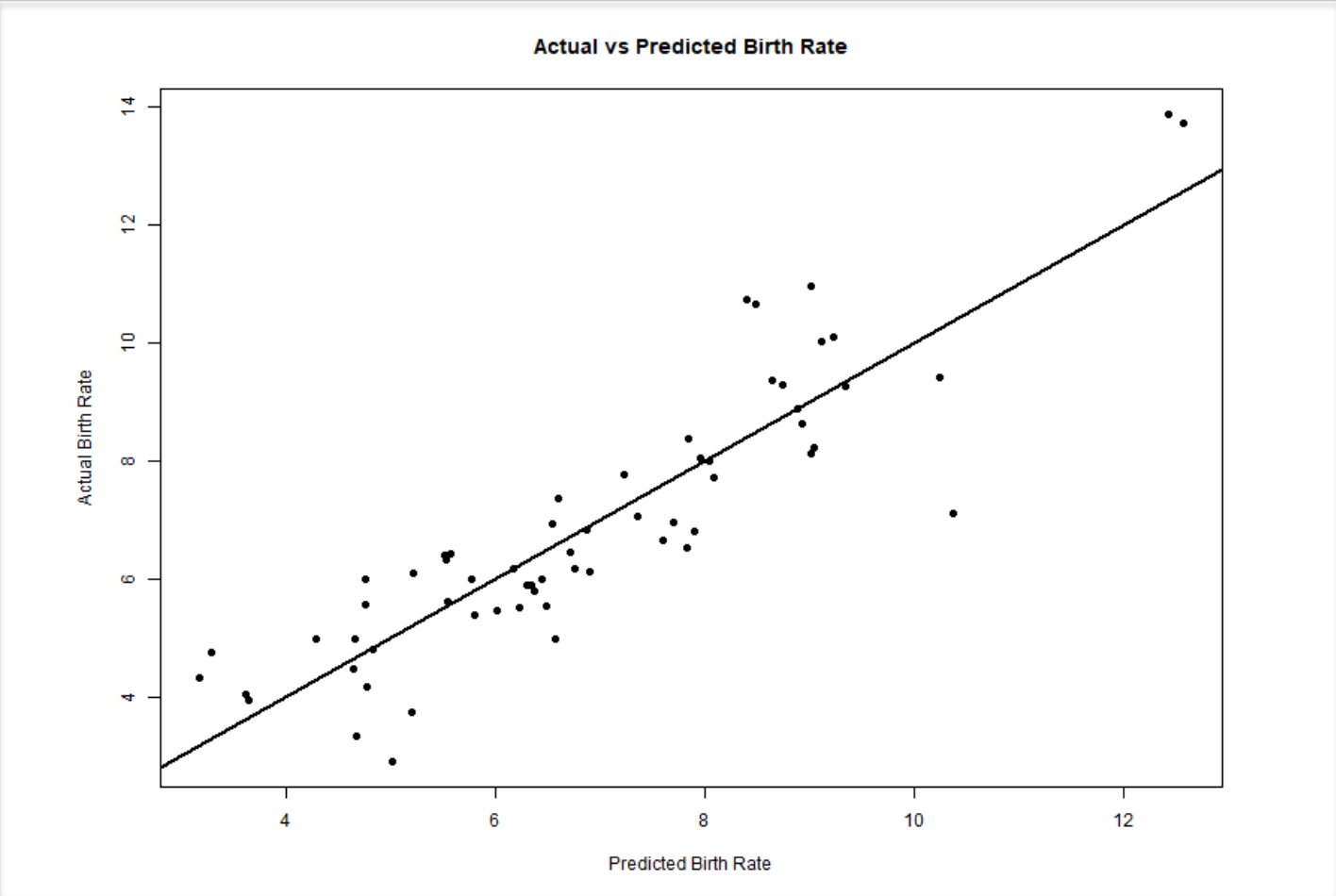


Multiple regression plane



Model Fit: Actual vs Predicted Birth Rate

Actual vs Predicted Birth Rate



Fit Summary

The points are mainly close to the fitted line.

The value of R^2 is 0.8008, which means that it explains very well.

Residual Error

The residual standard error is 1.033.

The model captures the major overall trend, but not all of the variation.

Model Decision

The F-test p <0.001.

Reject H0: $\beta 1 = \beta 2 = 0$

Conclusion

Main Findings

Aging rate and urbanization rate have a negative association with the birth rate.

Aging rate is the stronger predictor with 0.494 decrease in the birth rate associated with a 0.01 increase in aging rate.

Urbanization rate is also an important predictor of birth rate, which there is 0.045 decrease in the birth rate with a 0.01 increase in urbanization rate.

The R^2 is 0.8008, demonstrating the variable combination is capable of explaining 80% of the variation in birth rate.

Overall, it is the higher the aging rate and urbanization rate, the lower the birth rate.
Moreover, aging rate seems to be the stronger predictor.

Code

```r
# Scatter plot: Birth Rate and Aging Rate
ggplot(data, aes(x = AgingRate, y = BirthRate)) +
  geom_point(size = 3) +
  labs(
    title = "Birth Rate & Aging Rate",
    x = "Aging Rate",
    y = "Birth Rate"
  ) +
  theme_minimal()

# Scatter plot: Birth Rate and Urbanization Rate
ggplot(data, aes(x = UrbanizationRate, y = BirthRate)) +
  geom_point(size = 3) +
  labs(
    title = "Birth Rate & Urbanization Rate",
    x = "Urbanization Rate",
    y = "Birth Rate"
  ) +
  theme_minimal()

# Fitted line: Aging Rate
ggplot(data, aes(x = AgingRate, y = BirthRate)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    title = "Aging Rate Fitted Line",
    x = "Aging Rate",
    y = "Birth Rate"
  ) +
  theme_minimal()

# Fitted line: Urbanization Rate
ggplot(data, aes(x = UrbanizationRate, y = BirthRate)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    title = "Urbanization Rate Fitted Line",
    x = "Urbanization Rate",
    y = "Birth Rate"
  ) +
  theme_minimal()
```

```r
# Actual vs Predicted Birth Rate
ggplot(data, aes(x = Predicted, y = BirthRate)) +
  geom_point(size = 3) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  labs(
    title = "Actual vs Predicted Birth Rate",
    x = "Predicted Birth Rate",
    y = "Actual Birth Rate"
  ) +
  theme_minimal()
# Multiple linear regression
model <- lm(BirthRate ~ AgingRate + UrbanizationRate, data = data)

# Regression summary
reg_summary <- summary(model)

# Coefficient table: Estimate, Std. Error, t value, p-value
coef_table <- as.data.frame(coef(reg_summary))

# Rename columns
colnames(coef_table) <- c("Estimate", "Std_Error", "t_value", "p_value")

# Add significance result
coef_table$Result <- ifelse(coef_table$p_value < 0.05,
                            "Significant",
                            "Not Significant")

# Show coefficient table
coef_table
```

References

National Bureau of Statistics of China. (n.d.). *National Data*. National Bureau of Statistics of China. Retrieved from https://data.stats.gov.cn/

Zhang, E. (2026). *Chapter 9: Linear Regression Analysis*. Math 3544 Probability and Math Statistics, Wenzhou-Kean University.