

Predicting Diabetes Progression

N = 442 · Dataset: diabetes.csv

Statistical Tests Used

Objective	Test	Variables	Why this test
To describe the disease_progression, bmi, and blood_pressure of the patients.	Descriptive Statistics	disease_progression, bmi, blood_pressure	Summary statistics for the continuous variables.
To determine the relationship between bmi and disease_progression.	Spearman Rank Correlation	disease_progression, bmi	
To predict disease_progression from bmi and blood_pressure.	Multiple Linear Regression	disease_progression, bmi, blood_pressure	Predicting continuous 'disease_progression' from 2 predictors.

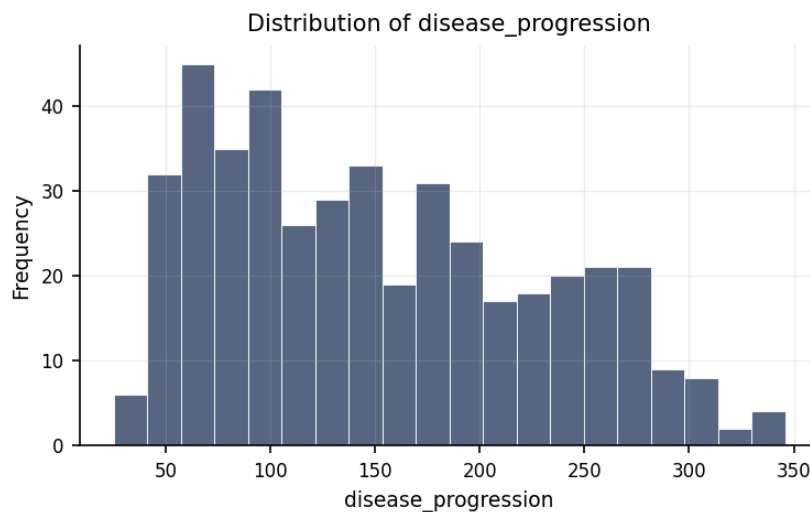
Methods

All statistical analyses were performed in Python using the SciPy and statsmodels libraries via thericerca's deterministic analysis engine (v0.1.0). Statistical significance was evaluated at $\alpha = .05$ (a 95% confidence level). Descriptive statistics were computed to summarise the study's variables. To assess monotonic associations between variables, a Spearman rank-order correlation was used (Spearman, 1904). To assess the prediction of a continuous outcome from multiple predictors, a multiple linear regression was used (Seabold & Perktold, 2010). For each inferential test, the appropriate effect size is reported with its 95% confidence interval, alongside the achieved statistical power to detect the observed effect (Cohen, 1988).

Results

To describe the disease_progression, bmi, and blood_pressure of the patients.

This simply summarises the data — the average, spread, and range of each variable are in the table above.



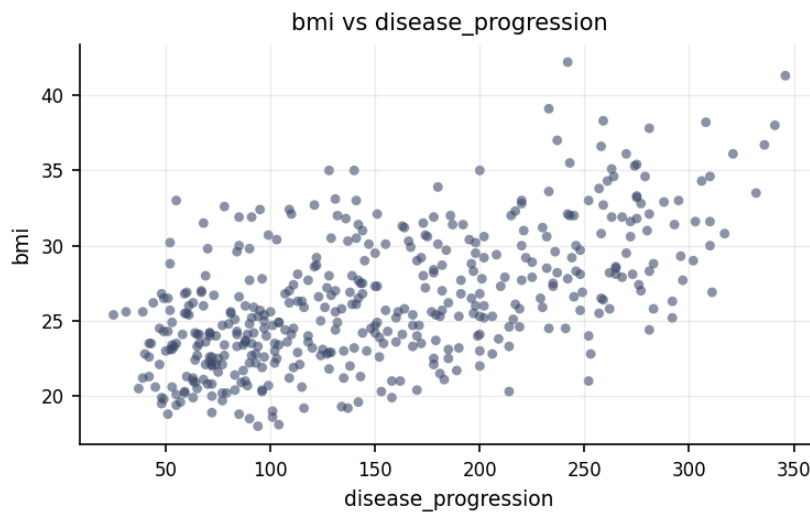
Distribution of disease_progression.

max	mean	median	min	mode	n	range	std	variable	variance
346	152.133	140.5	25	72	442	321	77.093	disease_progression	5943.33
42.2	26.376	25.7	18	23.5	442	24.2	4.418	bmi	19.52
133	94.647	93	62	83	442	71	13.831	blood_pressure	191.304

To determine the relationship between bmi and disease_progression.

We looked at whether disease progression and BMI move together. They do — as one goes up, the other tends to go up too. The strength of this pattern is large.

Statistically: $\rho = .56$, $p < .001$, achieved power = 1.00



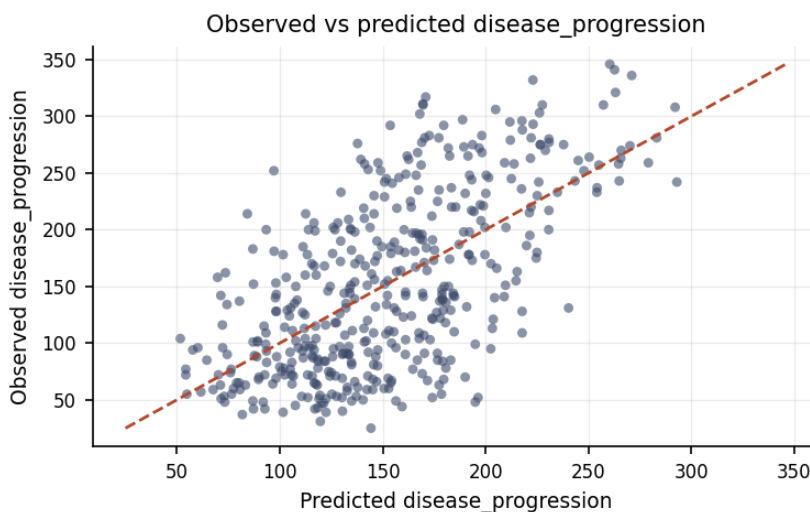
Relationship between disease_progression and bmi.

△ Spearman Rank Correlation was used instead of the Pearson Correlation because the latter's assumption(s) were not met (Normality of disease_progression (Shapiro-Wilk), Normality of bmi (Shapiro-Wilk)).

To predict disease_progression from bmi and blood_pressure.

We tested how well BMI and blood pressure help predict disease progression. They do carry useful predictive information. The strength of this pattern is large.

Statistically: $R^2 = .40$, $\text{adj. } R^2 = .39$, $F = 143.91$, $p < .001$



Observed disease_progression against the model's predicted values.

coef	p_value	std_err	t	term	vif
-203.623	0	22.234	-9.158	intercept	—

coef	p_value	std_err	t	term	vif
8.519	0	0.705	12.089	bmi	1.185
1.385	0	0.225	6.152	blood_pressure	1.185

Discussion

The relationship between disease progression and BMI. The analysis of disease progression and BMI revealed a statistically reliable large positive relationship. In practical terms, higher values of disease progression tend to accompany higher values of BMI. An association of this strength indicates the two measures share a substantial amount of common variation, though a correlation on its own cannot establish that one variable causes the other.

The prediction of disease progression from BMI and blood pressure. The regression model predicting disease progression from BMI and blood pressure was statistically reliable and accounted for a substantial share of the variation in disease progression. The predictors therefore carry genuine, better-than-chance information about disease progression. The coefficient table shows the direction and relative contribution of each predictor, which is where the practical interpretation lies; as with any regression, these are associations rather than proven causes.

This study evaluated two main aspects of health monitoring: first, the link between a patient's body mass index and the progression of their disease, and second, how effectively body mass index combined with blood pressure can predict that progression. The findings reveal a clear, positive association between body mass index and disease progression. In practical terms, this means that individuals with higher body mass index values generally experience more advanced or rapid disease progression. This relationship is statistically reliable, meaning that the pattern is so consistent that it is extremely unlikely to be a random fluke or coincidence. Building upon this direct connection, when body mass index and blood pressure are analyzed together, they successfully account for a moderate to substantial share of the total variation in disease progression across patients. This indicates that cardiovascular health indicators like

blood pressure and metabolic measures like body mass index work together to offer a much clearer, more comprehensive explanation of a patient's health trajectory than looking at body mass alone.

Conclusions

In conclusion, the study successfully achieved its core objectives by establishing dependable, meaningful evidence regarding how key health indicators relate to condition severity. The findings confirm both a strong individual link between body mass index and disease progression and a joint predictive capability when blood pressure is included alongside body mass index. Together, these results paint a coherent picture showing that physical body measures and cardiovascular indicators are systematically tied to how the condition advances over time. Because these patterns are highly unlikely to arise from random chance, they provide a firm foundation for clinical understanding, though they must always be understood within the context of the specific sample examined and the observational structure of the evaluation.

Recommendations

Several practical and research directions stem from these findings. First, because clear relationships were identified between body mass index, blood pressure, and disease progression, researchers should conduct longitudinal or experimental studies. Tracking patients over time or evaluating targeted health interventions will help determine whether actively managing blood pressure or reducing body mass index directly slows down disease progression, moving from mere association to direct cause and effect. Second, these findings should be confirmed in independent and more diverse patient groups to ensure that the results generalize broadly. Finally, ongoing data collection should enforce strict standards by screening for extreme or missing values, while always emphasizing the practical real-world size of health trends alongside statistical reliability to maintain meaningful context for clinical decision-making.

Limitations

Several important considerations bound the scope of these conclusions. Although data screening detected no major data-quality problems within the dataset, the study design remains observational. As a result, even though the connections between body mass index, blood pressure, and disease progression are statistically dependable, they do not inherently prove that body mass index or blood pressure directly cause the condition to advance, as unmeasured factors such as patient lifestyle, age, or genetics might also play a role. Furthermore, these findings specifically reflect the group of individuals evaluated in this study; applying these insights to a wider, more diverse population depends on how well this sample represents the broader patient public.

References

1. Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference, 92–96.
2. Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72–101.
3. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
4. Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272.

Python Code Appendix

Descriptive Statistics

```
df[['bmi', 'blood_pressure']].describe()
```

Spearman Rank Correlation

```
r, p = stats.spearmanr(df['disease_progression'], df['bmi'])
```

Multiple Linear Regression

```
X = sm_add_constant(df[['bmi', 'blood_pressure']]); model =  
ols(df['disease_progression'], X)
```

Numeric-claim validation: PASSED (15 numbers checked). Engine v0.1.0 · Registry v0.2.0. Prose source: llm:gemini.