TUGAS UJIAN TENGAH SEMESTER
MATA KULIAH BIOSTATISTIK LANJUT

Dosen Pengampu: dr. Dwi Agustian, MPH., PhD.

Disusun oleh :
Staviera A. (131520240001)

PROGRAM STUDI MAGISTER EPIDEMIOLOGI

FAKULTAS KEDOKTERAN

UNIVERSITAS PADJADJARAN

BANDUNG

2025

INTRODUCTION

Peak expiratory flow rate (PEFR) is a crucial parameter in assessing lung function. PEFR measures the rate at which an individual can expel air from their lungs after a maximal inspiration(Mrindha et al., 2012). However, peak expiratory flow rate is not solely determined by lung condition, but is also influenced by a multitude of other factors. Several factors might influence peak expiratory flow rate, including individual characteristics such as age, sex, and height, as well as health conditions like asthma(Dominelli & Molgat-Seon, 2022; Melén et al., 2019; Soremekun et al., 2023).

Understanding the factors that influence peak expiratory flow rate is essential for the diagnosis, management, and prevention of respiratory diseases(Bae et al., 2021; Cen et al., 2019; Donaldson et al., 2012; Louis et al., 2022). By recognizing these factors, healthcare professionals can provide more effective care and treatment to patients, ultimately improving their quality of life.

Data management and analysis held important part in conducting health research. Incorrect data processing might lead to biased result, or even held no meaningful result at all. Effective data processing enables researchers to cleanse data of errors and inaccuracies, transforming raw data into a format suitable for analysis. In this case, the use of statistical software such as R studio should have been very useful to reduce inaccuracies, especially in managing large number of data(Batko & Ślęzak, 2022; Subrahmanya et al., 2022).

In order to understand and become familiar with this process, we are encouraged to use R-studio, to conduct a data-processing and data-analyzing exercise using raw, big data that contains PEF data and other variables that might contribute as factors associated with PEF.

METHODS

1. Data Preparation

The dataset used in this analysis was derived from the merging of two files: *pef* and *w5*, as previously generated during the Midterm Examination (UTS) for the Data Management course. The final dataset included the following variables: `pidlink` (participant identifier), `pef` (peak expiratory flow), `age`, `height`, `sex`, and `Asthma` status.

2. GitHub Account and Repository Setup

A new repository was created in personal GitHub account to manage and store all R scripts and relevant files used in the Advanced Biostatistics Midterm project. This repository ensures reproducibility, proper version control, and traceability of all analysis steps.

3. Univariable Analysis

Univariable analyses were conducted to examine the individual relationships between the dependent variable `pef` and each of the independent variables: `age`, `height`, `sex`, and `Asthma`. The choice of statistical tests was based on the type and distribution of the variables.

Pearson or Spearman correlation coefficients were used for continuous variables (`age`, `height`), while independent t-tests were used for mean comparisons in binomial categoric variables (`sex`, `Asthma`). Correlation and comparison were stated as significant if p value < 0,05. Analysis was conducted in R-studio software, and R codes that was used for analysis was saved and uploaded to GitHub repository.

4. Tabulation of Univariable Results

The results of the univariable analyses were summarized and presented in a single table formatted according to standard scientific journal conventions. The table includes estimates of association (correlation coefficients for correlation, mean differences for independent t-test), 95% confidence intervals, and p-values to assess statistical significance.

5. Multivariable Linear Regression Modeling

A multivariable linear regression analysis was performed with `pef` as the dependent variable and `age`, `height`, `sex`, and `Asthma` as independent variables. All variables were included in the initial model to assess their adjusted effects on `pef`. Analysis was also conducted in R-studio software, and R codes that was used for analysis was saved and uploaded to GitHub repository.

6. Model Selection and Evaluation

Several candidate regression models were constructed using different combinations of predictor variables. On this attempt, four models were constructed : (1)pef ~ age + sex + height, (2) pef ~ sex + height + Asthma , (3) pef ~ age + sex + Asthma , (4) pef ~ age + sex + height + Asthma. The performance of each model was evaluated using the Akaike Information Criterion (AIC). The model with the lowest AIC value was selected as the final model, indicating the best trade-off between model fit and complexity.

RESULT

1. Univariate Analysis Results

Table 1. Pearson's Product-Moment Correlation Results : Variable Height & PEF

Variables	Correlation (r)	95% CI	t	df	p-value
Height & PEF	0,607	0,599 to 0,616	111,8	21391	< 2.2e-16 (0,0000)

Table 2. Pearson's Product-Moment Correlation Results : Variable Age & PEF

Variables	Correlation (r)	95% CI	t	df	p-value
Age & PEF	-0,3411376	-0,3529246 to -0,3292423	-53,078	21391	< 2.2e-16 (0,0000)

Table 3. Independent Samples t-Test Results Comparing PEF Means in No-Asthma and Yes-Asthma Groups

Group	N	Mean	95% CI of Difference	t	df	p-value
No-Asthma		375.53				
Yes-Asthma		304.38	61.27 to 81.03	14.14	600.28	< 2.2e-16 (0,000)

Table 4. Independent Samples t-Test Results Comparing PEF Means in Sex Groups

Group	N	Mean	95% CI of Difference	t	df	p-value
Female		304,440				
Male		450,682	-148,793 to -143,691	-112,37	17631	< 2.2e-16 (0,000)

2. Multivariate Linear Regression Results

Table 5. Model 1: pef ~ age + sex + height

Predictor	Estimate (β)	Std. Error	t-value	p-value	Significance
(Intercept)	-41.34	15.45	-2.68	0.00746	**
Age	-2.59	0.04	-59.30	< 0.001	***
Sex (Male)	111.24	1.60	69.34	< 0.001	***
Height	3.03	0.10	31.00	< 0.001	***

Model Statistics: pef = -41,32+(-2,59)Age+111,24(if Male)+3,03(height)+e

Residual Standard Error: 81.75 (df = 21,273)

R-squared: 0.5219

Adjusted R-squared: 0.5219

F-statistic: 7741 on 3 and 21,273 df, p < 0.001

116 observations deleted due to missingness

Table 6. Model 2: pef ~ sex + height + Asthma

Predictor	Estimate (β)	Std. Error	t-value	p-value	Significance
(Intercept)	-425.58	15.01	-28.36	< 0.001	***
Sex (Male)	89.28	1.67	53.37	< 0.001	***
Height	4.86	0.10	48.84	< 0.001	***
Asthma (Yes)	-67.57	3.72	-18.17	< 0.001	***

Model Statistics: pef = -425,58+89,28(if Male)+4,86(height)+(-67,57)if Asthma Yes + e

Residual Standard Error: 87.58 (df = 21,273)

R-squared: 0.4514

Adjusted R-squared: 0.4513

F-statistic: 5835 on 3 and 21,273 df, p < 0.001

116 observations deleted due to missingness

Table 7. Model 3: pef ~ age + sex + Asthma

Predictor	Estimate (β)	Std. Error	t-value	p-value	Significance
(Intercept)	435.38	1.97	221.20	< 0.001	***
Age	-3.01	0.04	-71.57	< 0.001	***
Sex (Male)	146.60	1.14	128.76	< 0.001	***
Asthma (Yes)	-65.39	3.52	-18.57	< 0.001	***

Model Statistics: pef = 435,58 + (-3,01)Age + 146,6(if Male) + (-65,39) if Asthma Yes + e

Residual Standard Error: 82.91 (df = 21,273)

R-squared: 0.5083

Adjusted R-squared: 0.5082

F-statistic: 7330 on 3 and 21,273 df, p < 0.001

116 observations deleted due to missingness

Table 8. Model 4: pef ~ age + sex + height + Asthma

Predictor	Estimate (β)	Std. Error	t-value	p-value	Significance
(Intercept)	-40.64	15.32	-2.65	0.008	**
Age	-2.59	0.04	-59.63	< 0.001	***
Sex (Male)	111.01	1.59	69.78	< 0.001	***
Height	3.04	0.10	31.32	< 0.001	***
Asthma (Yes)	-65.71	3.44	-19.09	< 0.001	***

Model Statistics: -40,64 + (-2,59)Age + 111,01 (if Male) + 3,04 (height) + (-65,71) if Asthma Yes + e

Residual Standard Error: 81.07 (df = 21,272)

R-squared: 0.530

Adjusted R-squared: 0.5299

F-statistic: 5996 on 4 and 21,272 df, p < 0.001

116 observations deleted due to missingness

Table 9. Result of AIC Model Comparison

Model	Degrees of Freedom (df)	AIC
Model 1	5	247,783.7
Model 2	5	250,710.7
Model 3	5	248,381.6
Model 4	**6**	**247,424.3**

DISCUSSION

Results from univariate analysis presenting significant findings. Correlation between height and PEF demostrated positive and strong correlation (**Table 1**). This implies that taller individuals tend to have higher PEF values, likely reflecting larger lung volumes. In accordance with existing theory, which comprises that increasing height facilitate larger size of thoracic cavity, larger volume of lung and stronger respiratory muscle. This finding is also similar with previous studies, consistent findings both in children (Kyejo et al., 2024), or in adult population(Dharamshi et al., 2015)

Conversely, **Table 2** shows that age was negatively correlated with PEF, (Pearson's $r = -0.341$). This inverse relationship suggests a decline in pulmonary function with increasing age. A large study in China came across similar finding(Ren et al., 2012), and also in India(Sandhu et al., 2016). This decrease might be caused by degenerative changes in the musculoskeletal system of thoracoabdominal compartment leading to decrease in respiratory muscle strength with associated decrease in joint mobility and lung elasticity.

Individuals without asthma had significantly higher PEF values (Mean = 375.53) compared to those with asthma (Mean = 304.38). The difference in means was statistically significant, with p value lower than 0.05. This finding indicates that the presence of asthma is associated with a substantial reduction in pulmonary function as measured by PEF. Theoretically, presence of asthma might lead to small airway obstruction induced by inflammation, therefore causing lower PEF. However, statistical findings worldwide depict different results(Fonseca et al., 2006). This might be caused by different asthma control status in each study, and also reversible inflammation in asthma pathology which might lead to full recovery of airway obstruction, especially in mild and controlled asthma patients.

Similarly, the comparison of PEF values between sexes revealed a highly significant difference (Table 4). Males exhibited a higher mean PEF (Mean = 450.682) compared to females (Mean = 304.440) with p value lower than 0.05. These results suggest that biological sex plays an important role in determining pulmonary capacity, similar with previous study findings(Jahan et al., 2022). Males have larger lungs and stronger inspiratory muscles than females, which led to higher lung function(Berad et al., 2016). Other than that, number of alveoli per unit area in males are higher than females, with larger size and greater compliance compared to females(Dikshit et al., 2005).

Multivariable analysis As presented in Table 8, all predictors in the model were significantly associated with PEF ($p < 0.001$). Height remained a positive predictor of PEF (β = 3,04), suggesting that each additional centimeter in height corresponded to a 3.04 unit increase in PEF. Age was also remained as negative predictor for PEF (β = –2.59), indicating that for each additional year of age, PEF decreased by approximately 2.59 units, holding other variables constant. Sex also showed a significant effect; males had significantly higher PEF values compared to females (β = 111.01). Additionally, the presence of asthma was associated with a marked reduction in PEF (β = –65.71)

These findings confirmed that PEF is influenced by a combination of demographic and clinical characteristics, with older age and asthma being associated with lower pulmonary function, while male sex and greater height are associated with higher PEF.

CONCLUSION

Data analysis from this study was able to depict significant associations between PEF and several demographic and clinical variables using both univariable and multivariable statistical approaches. Univariable analyses revealed that individuals with asthma and females had significantly lower PEF values compared to their counterparts, while PEF was positively correlated with height and negatively correlated with age. These results are evidences that demonstrating the ability of statistical analysis to support theoretical and clinical framework in healthcare needs. Further exploration and exercise were needed in order to minimize limitation, increase effectivity and precision in using R studio for data processing and analyzing.

REFERENCES

Bae, W. D., Kim, S., Park, C. S., Alkobaisi, S., Lee, J., Seo, W., Park, J. S., Park, S., Lee, S., & Lee, J. W. (2021). Performance improvement of machine learning techniques predicting the association of exacerbation of peak expiratory flow ratio with short term exposure level to indoor air quality using adult asthmatics clustered data. *PLoS ONE*, *16*(1 January), 1–16. https://doi.org/10.1371/journal.pone.0244233

Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-021-00553-4

Berad, A., Sneha, P., & Prasad, D. (2016). *Original article Gender variation in Lung function parameters in Indian Adolescents*. *6*(2), 5414–5416.

Cen, J., Ma, H., Chen, Z., Weng, L., & Deng, Z. (2019). Monitoring peak expiratory flow could predict COPD exacerbations: A prospective observational study. *Respiratory Medicine*, *148*(October 2018), 43–48. https://doi.org/10.1016/j.rmed.2019.01.010

Dharamshi, H. A., Faraz, A., Ashraf, E., Alam, S. S., Ali, A., Shakeel, O., Abidi, S. M. A., Rizvi, S. S., Fatima, Z., Wasy, H. A., Fatima, F., Mahar, M., & Naqvi, T. (2015). Variation of PEFR with height, weight and waist-hip ratio in medical students. *International Archives of Medicine*, *8*(1). https://doi.org/10.3823/1683

Dikshit, M. B., Raje, S., & Agrawal, M. J. (2005). Lung functions with spirometry: An Indian perspective-I. Peak expiratory flow rates. *Indian Journal of Physiology and Pharmacology*, *49*(1), 8–18.

Dominelli, P. B., & Molgat-Seon, Y. (2022). Sex, gender and the pulmonary physiology of exercise. *European Respiratory Review*, *31*(163). https://doi.org/10.1183/16000617.0074-2021

Donaldson, G. C., Seemungal, T. A. R., Hurst, J. R., & Wedzicha, J. A. (2012). Detrended fluctuation analysis of peak expiratory flow and exacerbation frequency in COPD. *European Respiratory Journal*, *40*(5), 1123–1129. https://doi.org/10.1183/09031936.00180811

Fonseca, A. C. C. F., Fonseca, M. T. M., Rodrigues, M. E. S. M., Lasmar, L. M. L. B. F., & Camargos, P. A. M. (2006). Peak expiratory flow monitoring in asthmatic children. *Jornal de Pediatria*, *82*(6), 465–469. https://doi.org/10.2223/JPED.1566

Jahan, I., Begum, M., Sumi, S. A., Jahan, N., Islam, M. Z., & Haque, M. (2022). Correlation of Gender and Anthropometric Measurement with Pulmonary Function Tests Among Undergraduate Medical Students. *Advances in Human Biology*, *12*(3), 316–321. https://doi.org/10.4103/aihb.aihb_119_22

Kyejo, W., Matillya, N., Ismail, N., Gachocha, G., Salum, H., Iseme, R., & Noorani, M. (2024). The anthropometric determinants of peak expiratory flow rate among children in Dar Es Salaam, Tanzania. *BMC Pediatrics*, *24*(1), 1–8. https://doi.org/10.1186/s12887-023-04520-1

Louis, R., Satia, I., Ojanguren, I., Schleich, F., Bonini, M., Tonia, T., Rigau, D., Brinke, A. ten, Buhl, R., Loukides, S., Kocks, J. W. H., Boulet, L. P., Bourdin, A., Coleman, C., Needham, K., Thomas, M., Idzko, M., Papi, A., Porsbjerg, C., … Usmani, O. S. (2022). European Respiratory Society guidelines for the diagnosis of asthma in adults. *European Respiratory Journal*, *60*(3).

https://doi.org/10.1183/13993003.01585-2021

Melén, E., Guerra, S., Hallberg, J., Jarvis, D., & Stanojevic, S. (2019). Linking COPD epidemiology with pediatric asthma care: Implications for the patient and the physician. *Pediatric Allergy and Immunology*, *30*(6), 589–597. https://doi.org/10.1111/pai.13054

Mrindha, M. A.-A., Amin, M. R., & Kabir, A. L. (2012). Peak Expiratory Flow Rate (PEFR)-A Simple Ventilatory Lung Function Test. *Journal of Shaheed Suhrawardy Medical College*, *3*(2), 44–47. https://doi.org/10.3329/jssmc.v3i2.12078

Ren, W. ying, Li, L., Zhao, R. ya, & Zhu Dr., L. (2012). Age-associated changes in pulmonary function: A comparison of pulmonary function parameters in healthy young adults and the elderly living in shanghai. *Chinese Medical Journal*, *125*(17), 3064–3068. https://doi.org/10.3760/cma.j.issn.0366-6999.2012.17.018

Sandhu, P. K., Bajaj, D., & Mehta, K. (2016). Correlation of peak expiratory flow rate with age and anthropometric parameters in elderly (>65 years). *National Journal of Physiology, Pharmacy and Pharmacology*, *6*(1), 89–92. https://doi.org/10.5455/njppp.2015.5.0104201584

Soremekun, S., Heaney, L. G., Skinner, D., Bulathsinhala, L., Carter, V., Chaudhry, I., Hosseini, N., Eleangovan, N., Murray, R., Tran, T. N., Emmanuel, B., Garcia Gil, E., Menzies-Gow, A., Peters, M., Lugogo, N., Jones, R., & Price, D. B. (2023). Asthma exacerbations are associated with a decline in lung function: A longitudinal population-based study. *Thorax*, *78*(7), 643–652. https://doi.org/10.1136/thorax-2021-217032

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., Naik, N., & Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish Journal of Medical Science*, *191*(4), 1473–1483. https://doi.org/10.1007/s11845-021-02730-z