

```

# Menyiapkan paket
library(readr)
library(dplyr)
library(lmtest)
library(knitr)

# memuat data dan filter
data_new <- data[, c("pidlink", "pef", "age", "height", "sex", "Asthma")]
data_filtered <- data_new[!is.na(data_new$Asthma), ]
data_clean <- data_filtered[complete.cases(data_filtered[, c("pef", "age", "height", "sex"))], ]
cat("\nJumlah baris data awal:", nrow(data_new), "\n")
cat("Jumlah baris setelah memfilter Asthma non-NA:", nrow(data_filtered), "\n")
cat("Jumlah baris setelah pembersihan lengkap:", nrow(data_clean), "\n")

# Memeriksa distribusi Sex dan Asthma
cat("\nDistribusi nilai Sex:\n")
print(table(data_clean$sex, useNA = "ifany"))
cat("\nDistribusi nilai Asthma:\n")
print(table(data_clean$Asthma, useNA = "ifany"))

# Memeriksa Sex dan Asthma sebagai faktor
data_clean$sex <- as.factor(data_clean$sex)
data_clean$Asthma <- as.factor(data_clean$Asthma)

# Pemeriksaan jumlah kategori
sex_categories <- length(unique(na.omit(data_clean$sex)))
asthma_categories <- length(unique(na.omit(data_clean$Asthma)))
if (sex_categories != 2 || asthma_categories != 2) {
  cat("\nPeringatan: Analisis regresi linier memerlukan dua kategori untuk variabel kategorik.\n")
  if (sex_categories != 2) cat("Sex memiliki", sex_categories, "kategori:",
    paste(unique(na.omit(data_clean$sex)), collapse = ", "), "\n")
  if (asthma_categories != 2) cat("Asthma memiliki", asthma_categories, "kategori:",
    paste(unique(na.omit(data_clean$Asthma)), collapse = ", "), "\n")
  stop("Silakan perbaiki data sebelum melanjutkan.")
}

# Membuat beberapa model regresi linier
model1 <- lm(pef ~ age + height + sex + Asthma, data = data_clean) # Full model
model2 <- lm(pef ~ age + height + sex, data = data_clean) # Tanpa Asthma
model3 <- lm(pef ~ age + height + Asthma, data = data_clean) # Tanpa Sex
model4 <- lm(pef ~ height + sex + Asthma, data = data_clean) # Tanpa Age
model5 <- lm(pef ~ height + Asthma, data = data_clean) # Hanya Height dan Asthma

# MengHitung AIC untuk setiap model
aic_values <- data.frame(
  Model = c("Model 1: Full (age + height + sex + Asthma)",
    "Model 2: age + height + sex",
    "Model 3: age + height + Asthma",

```

```

    "Model 4: height + sex + Asthma",
    "Model 5: height + Asthma"),
  AIC = c(AIC(model1), AIC(model2), AIC(model3), AIC(model4), AIC(model5))
)
aic_values$AIC <- round(aic_values$AIC, 2)
aic_values <- aic_values[order(aic_values$AIC), ] # Urutkan berdasarkan AIC

# Menampilkan tabel perbandingan AIC
cat("\nTabel Perbandingan AIC untuk Berbagai Model:\n")
print(kable(aic_values, format = "markdown"))

# membandingkan model terbaik (AIC terendah)
best_model_name <- aic_values$Model[1]
best_model <- switch(which(aic_values$Model == best_model_name),
  `1` = model1, `2` = model2, `3` = model3, `4` = model4, `5` = model5)
cat("\nModel Terbaik (AIC Terendah):", best_model_name, "\n")

summary_best <- summary(best_model)
cat("\nRingkasan Model Terbaik:\n")
print(summary_best)

# Menentukan Tabel koefisien model terbaik
coefs <- as.data.frame(summary_best$coefficients)
coefs <- cbind(Variable = rownames(coefs), coefs)
colnames(coefs) <- c("Variable", "Estimate", "Std.Error", "t.value", "P.Value")
coefs$P.Value <- round(coefs$P.Value, 4)
coefs$Significance <- ifelse(coefs$P.Value < 0.05, "Significant", "Not Significant")
cat("\nTabel Koefisien Model Terbaik:\n")
print(kable(coefs, format = "markdown"))

# Pemeriksaan asumsi untuk model terbaik
# Memeriksa Normalitas residu
residu <- residuals(best_model)
shapiro_residu <- shapiro.test(residu)
cat("\nUji Normalitas Residu (Shapiro-Wilk) untuk Model Terbaik:\n")
cat("p-value:", shapiro_residu$p.value, "\n")
cat(ifelse(shapiro_residu$p.value < 0.05, "Residu tidak normal (p < 0.05).\n", "Residu normal (p >= 0.05).\n"))

# Memeriksa Homoskedastisitas
bp_test <- bptest(best_model)
cat("\nUji Homoskedastisitas (Breusch-Pagan) untuk Model Terbaik:\n")
cat("p-value:", bp_test$p.value, "\n")
cat(ifelse(bp_test$p.value < 0.05, "Varians tidak konstan (p < 0.05).\n", "Varians konstan (p >= 0.05).\n"))

# Membuat Visualisasi asumsi untuk model terbaik
par(mfrow = c(2, 2))

```

```
plot(best_model, which = 1:2) # Residuals vs Fitted, Q-Q Plot
plot(best_model, which = 3) # Scale-Location (homoskedastisitas)
plot(best_model, which = 5) # Residuals vs Leverage

# Memeriksa Interpretasi R-squared untuk model terbaik
cat("\nR-squared dan Adjusted R-squared untuk Model Terbaik:\n")
cat("R-squared:", summary_best$r.squared, "\n")
cat("Adjusted R-squared:", summary_best$adj.r.squared, "\n")
cat("Interpretasi: ", round(summary_best$r.squared * 100, 2), "% variasi dalam PEF dapat dijelaskan oleh model.\n")
```