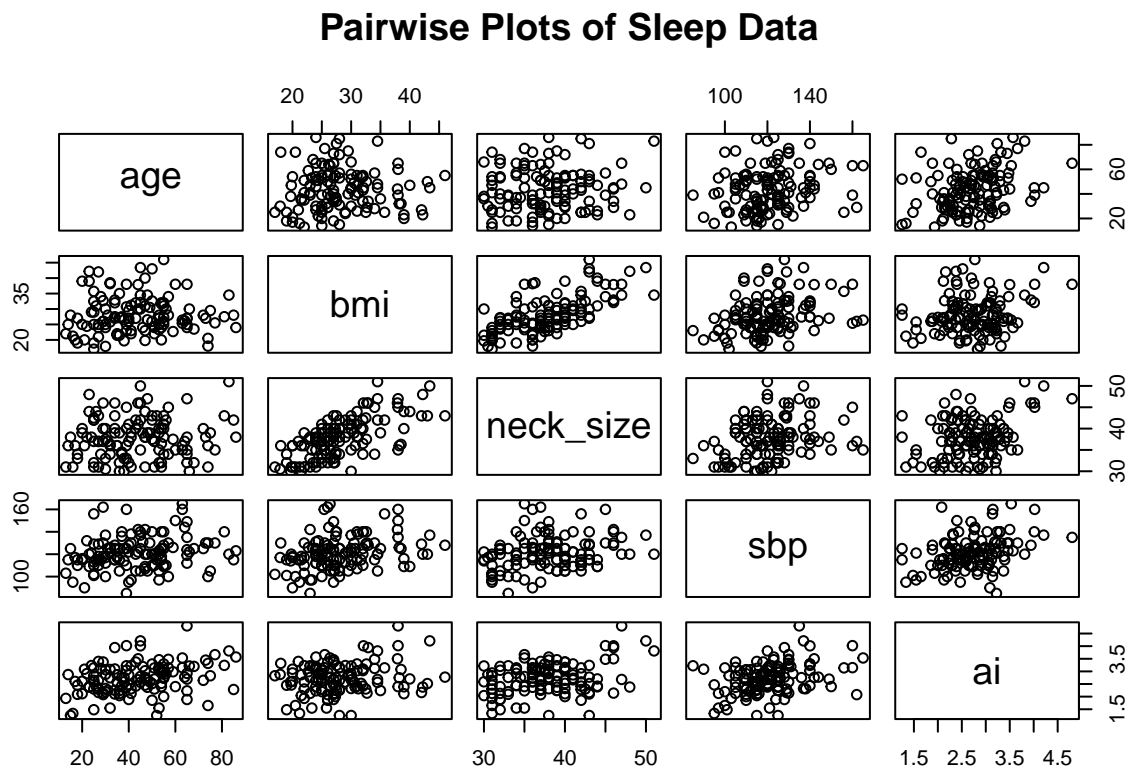# STAT2170/STAT6180 Assignment 2025

Akib Bin Anowar - 48682462

Question 1

(a) Correlation and Plot

```
sleep <- read.csv("sleep.csv")
pairs(sleep, main="Pairwise Plots of Sleep Data")
```


Pairwise Plots of Sleep Data

```
round(cor(sleep), 2)
```

```
##            age  bmi neck_size  sbp   ai
## age       1.00 0.02      0.08 0.20 0.32
## bmi       0.02 1.00      0.67 0.31 0.19
```

```
## neck_size 0.08 0.67      1.00 0.25 0.33
## sbp        0.20 0.31      0.25 1.00 0.35
## ai         0.32 0.19      0.33 0.35 1.00
```

We observe moderate correlation between neck size and arousal index, as well as some positive correlation with age and systolic blood pressure.

(b) Full Regression Model & Neck Size Impact

```r
model_full <- lm(ai ~ age + bmi + neck_size + sbp, data=sleep)
summary(model_full)
```

```
##
## Call:
## lm(formula = ai ~ age + bmi + neck_size + sbp, data = sleep)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.67136 -0.32269  0.01491  0.35778  1.47595
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.159406   0.518207  -0.308  0.75893
## age          0.008789   0.002964   2.965  0.00367 **
## bmi         -0.009852   0.011312  -0.871  0.38557
## neck_size    0.040627   0.014208   2.859  0.00503 **
## sbp          0.010218   0.003555   2.875  0.00481 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5417 on 117 degrees of freedom
## Multiple R-squared:  0.2471, Adjusted R-squared:  0.2213
## F-statistic: 9.598 on 4 and 117 DF,  p-value: 9.54e-07
```

```r
confint(model_full, "neck_size", level = 0.95)
```

```
##                 2.5 %     97.5 %
## neck_size 0.01248892 0.06876571
```

Each cm increase in neck size is associated with an estimated 0.041 increase in log-scaled arousal index (CI: [0.012, 0.069]).

(c) ANOVA and Overall Model Test

Model:
$$ai = \beta_0 + \beta_1 \cdot age + \beta_2 \cdot bmi + \beta_3 \cdot neck\_size + \beta_4 \cdot sbp + \varepsilon$$

Hypotheses:
H_0: All $\beta = 0$ (no relationship)
H_1: At least one $\beta \neq 0$

```
anova(model_full)
```

```
## Analysis of Variance Table
##
## Response: ai
##             Df Sum Sq Mean Sq F value    Pr(>F)
## age          1  4.591  4.5911 15.6440 0.0001314 ***
## bmi          1  1.605  1.6045  5.4674 0.0210727 *
## neck_size    1  2.646  2.6460  9.0159 0.0032731 **
## sbp          1  2.425  2.4251  8.2633 0.0048069 **
## Residuals  117 34.337  0.2935
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
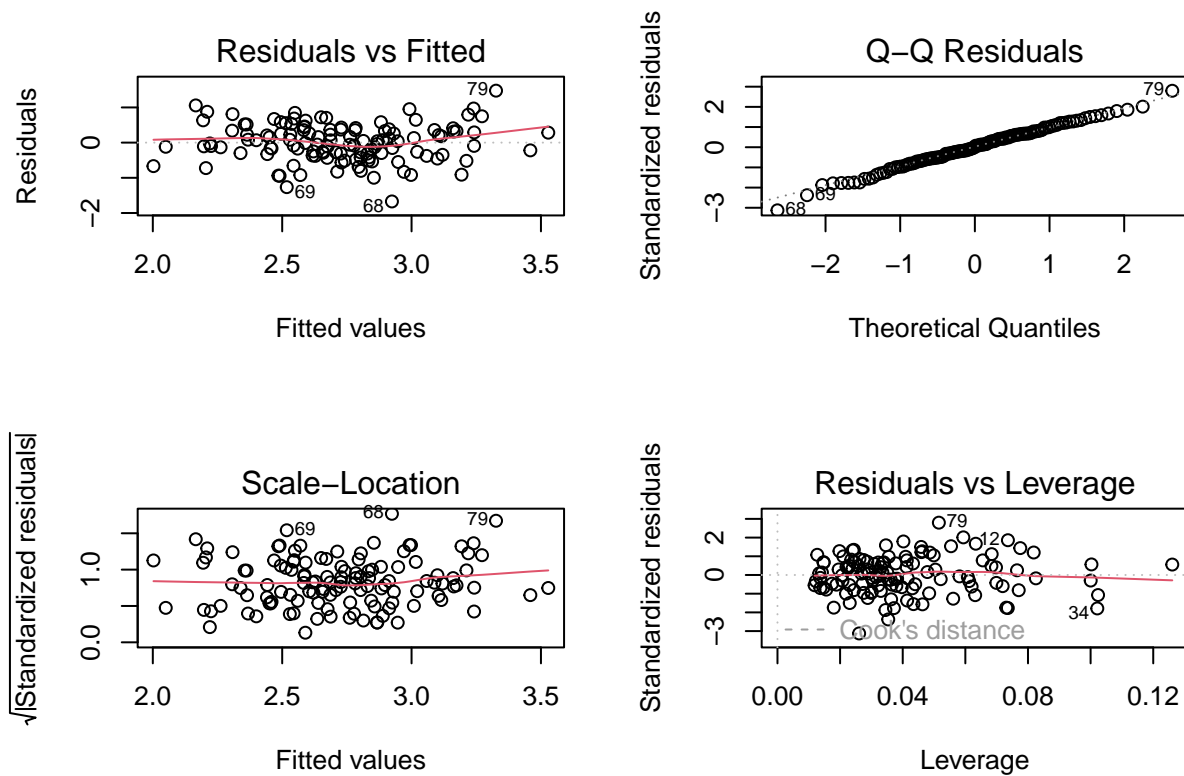
The F-statistic follows $F_{4,117}$ under H_0 and the p-value indicates strong evidence against H_0. Thus, at least one predictor significantly explains variation in ai.

(d) Model Diagnostics

```
par(mfrow=c(2,2))
plot(model_full)
```

The residual plots suggest assumptions of linearity, normality, and homoscedasticity are reasonably met.

(e) R-squared

```r
summary(model_full)$r.squared
```

```
## [1] 0.2470586
```

About 24.7% of the variation in arousal index is explained by the predictors.

(f) Model Selection

```r
step_model <- step(model_full)
```

```
## Start:  AIC=-144.67
## ai ~ age + bmi + neck_size + sbp
```

```
##
##            Df Sum of Sq    RSS     AIC
## - bmi       1   0.22261 34.559 -145.88
## <none>                   34.337 -144.67
## - neck_size 1   2.39958 36.736 -138.43
## - sbp       1   2.42507 36.762 -138.35
## - age       1   2.57977 36.916 -137.84
##
## Step:  AIC=-145.88
## ai ~ age + neck_size + sbp
##
##            Df Sum of Sq    RSS     AIC
## <none>                   34.559 -145.88
## - sbp       1    2.2261 36.785 -140.27
## - neck_size 1    2.6580 37.217 -138.84
## - age       1    2.7292 37.288 -138.61
```

summary(step_model)

```
##
## Call:
## lm(formula = ai ~ age + neck_size + sbp, data = sleep)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.65415 -0.35334  0.04008  0.37534  1.45627
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.066166   0.506509  -0.131  0.89629
## age          0.009007   0.002951   3.053  0.00280 **
## neck_size    0.032630   0.010831   3.013  0.00317 **
## sbp          0.009579   0.003475   2.757  0.00676 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5412 on 118 degrees of freedom
## Multiple R-squared:  0.2422, Adjusted R-squared:  0.2229
## F-statistic: 12.57 on 3 and 118 DF,  p-value: 3.452e-07
```

Final model: ai ~ age + neck_size + sbp

(g) Model Comparison

The adjusted R² of the reduced model is nearly identical to the full model, suggesting the excluded predictor (bmi) adds little value.

Question 2

(a) Balanced Design Check

```
energy <- read.csv("energy.csv")
table(energy$range, energy$menu)
```
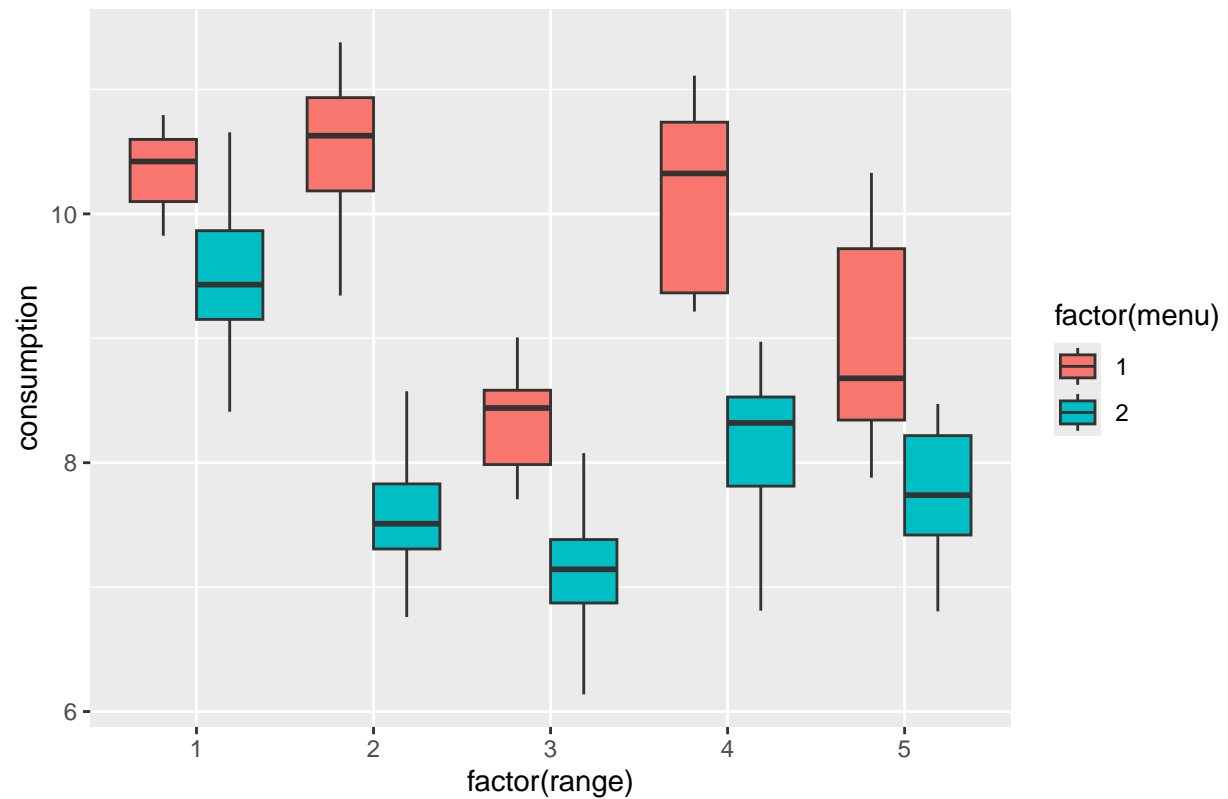
```
##
##      1 2
##    1 8 8
##    2 8 8
##    3 8 8
##    4 8 8
##    5 8 8
```

Each cell has equal observations → the design is **balanced**.

(b) Graphs

```
library(ggplot2)
ggplot(energy, aes(x=factor(range), y=consumption, fill=factor(menu))) +
    geom_boxplot(position=position_dodge()) +
    labs(title="Consumption by Range and Menu")
```

Consumption by Range and Menu


```
interaction.plot(x.factor=energy$range, trace.factor=energy$menu,
                 response=energy$consumption,
                 fun=mean, type="b", col=c("red","blue"),
                 legend=TRUE, xlab="Range", ylab="Consumption")
```
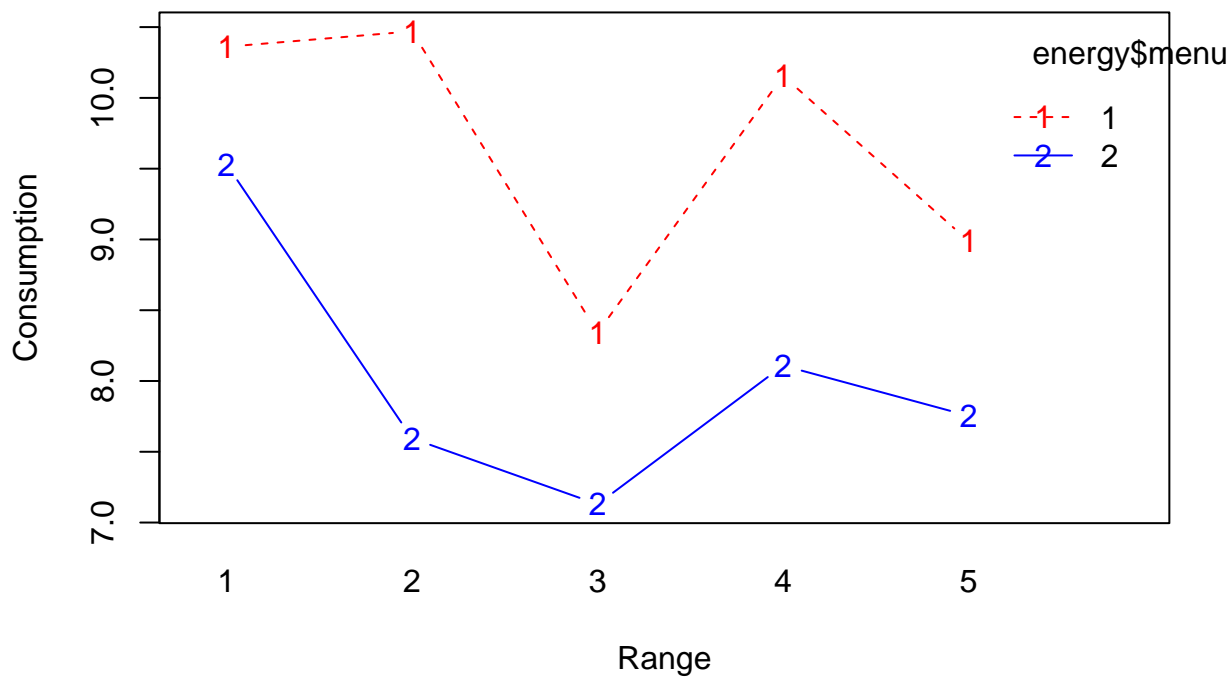
(c) Interaction Model

$consumption = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$
Where α_i = range effect, β_j = menu effect, and $(\alpha\beta)_{ij}$ = interaction.

(d) ANOVA

```
interaction_model <- lm(consumption ~ factor(range)*factor(menu), data=energy)
anova(interaction_model)
```

```
## Analysis of Variance Table
##
## Response: consumption
##                            Df Sum Sq Mean Sq  F value     Pr(>F)
## factor(range)               4 44.580  11.145  25.4400 4.935e-13 ***
## factor(menu)                1 53.772  53.772 122.7414 < 2.2e-16 ***
## factor(range):factor(menu)  4 10.781   2.695   6.1522 0.0002663 ***
## Residuals                  70 30.667   0.438
```
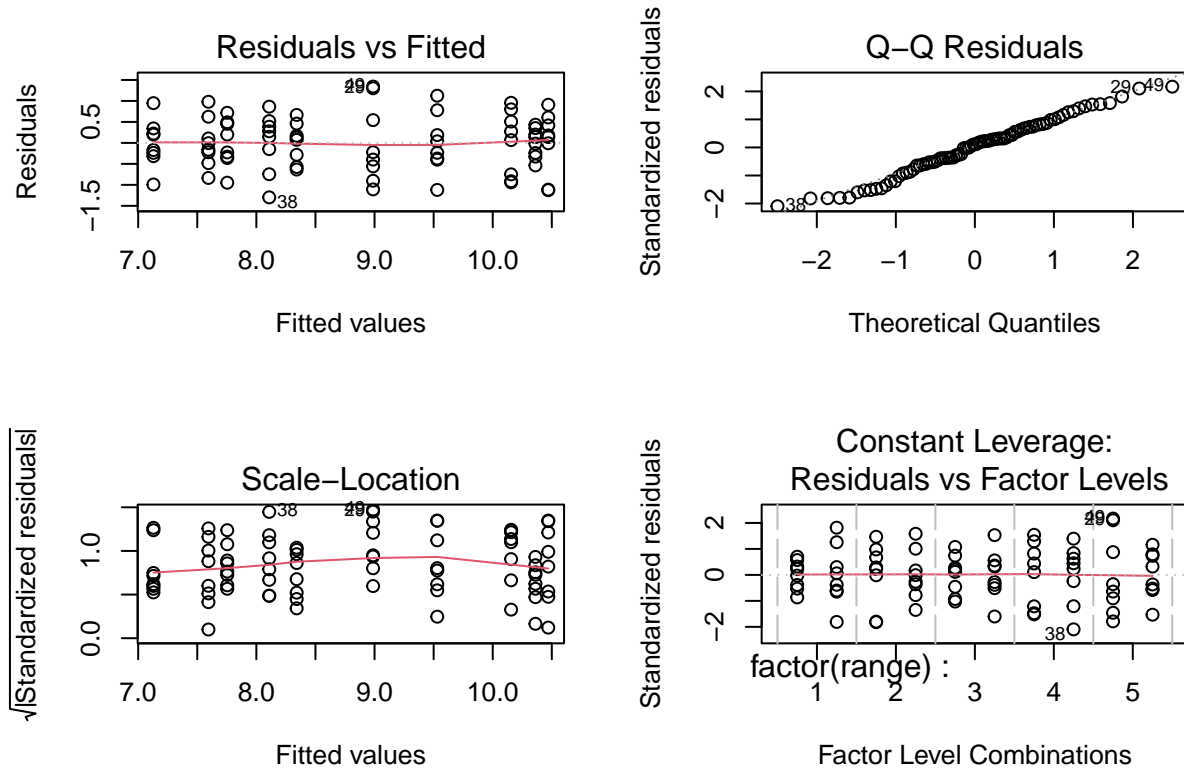
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significant interaction found ($p < 0.001$). Model diagnostics are below.

```
par(mfrow=c(2,2))
plot(interaction_model)
```


Conclusion: Both range, menu, and their interaction significantly affect energy consumption.