

# BIOSTAT702 Midterm 1 – Coding Portion

Jiaqi Wang

2025-10-09

**By submitting an exam, you are formally agreeing to the terms below and acknowledging that you have neither given nor received unauthorized aid in the completion of the exam.**

## Instructions

For the following set of tasks, you will be using the Hypoxia Dataset. Start by reading the Hypoxia Data Dictionary and Dataset Introduction. Optionally, you can also look at the Hypoxia Paper.

Type your answers and code where prompted. When you are done, click the knit button (knitting needle!) at the top of the Rstudio screen. It should create an .html file which will be your submission file to Canvas.

For this portion of the midterm, you may use any technology resources available to you, including notes, exercises, Internet searches, generative AI tools, etc. However, please be aware that you must submit answers to the questions written in your own words. This means that you should not quote phrases from other sources, including AI tools, even with proper attribution. This portion of the midterm must be done independently, without human assistance. **This portion of the exam is worth 30 points. Each question will be worth 5 points.**

## Load in the Dataset

```
# the csv file MUST be in the same folder as this Rmd file in order for this code to run properly as is.
```

```
Hypoxia = read.csv("hypoxia.csv")
```

## Question 1

Use the dataset introduction to come up with an *inference-focused* study question that you can answer with simple linear regression.

Study question: Among adults with obstructive sleep apnea (OSA) undergoing laparoscopic bariatric surgery, is lower minimum nocturnal oxygen saturation (Min SaO<sub>2</sub>) associated with lower intraoperative time-weighted average mean arterial pressure (TWA MAP)?

## Question 2

What is your predictor? What is your outcome? (Note: there are two predictors you could use based on the dataset intro. Just pick one.)

Predictor (X): minimum nocturnal oxygen saturation (Min SaO<sub>2</sub>) (%)

Outcome (Y): intraoperative time-weighted average mean aetrial pressure (TWA MAP)(mmHg)

### Question 3

Look at the structure of your data, and the distributions of your variables of interest. Describe what you see. Is there any missing data? If so, how did you handle it?

```
# CODE HERE -- press the green arrow to run all the code in this "chunk"
#Look at structure
names(Hypoxia)

##   [1] "Age"           "Female"         "Race"
##   [4] "BMI"           "Sleeptime"      "Min.Sao2"
##   [7] "AHI"           "Smoking"        "Diabetes"
##  [10] "Hyper"         "CAD"            "Preop.AntiHyper.Med"
##  [13] "CPAP"          "Type.Surg"      "Duration.of.Surg"
##  [16] "Duration.of.Surg1" "Duration.of.Surg2" "TWA.MAP"
##  [19] "TWA.MAP1"      "TWA.MAP2"       "TWA.HR"
##  [22] "TWA.HR1"       "TWA.HR2"        "Intraop.AntiHyper.Me
d"
##  [25] "Vasopressor"   "Ephedrine"      "Ephedrine.Amt"
##  [28] "Epinephrine"   "Epinephrine.Amt" "Phenylephrine"
##  [31] "Phenylephrine.Amt" "MAC"            "Propofol.Induction"
##  [34] "IV.Morphine.Eq" "Crystalloids"   "Colloids"

str(Hypoxia[, c("Min.Sao2", "TWA.MAP")])

## 'data.frame':   281 obs. of  2 variables:
##  $ Min.Sao2: num   90 94 76 52 95 51 86 56 85 68 ...
##  $ TWA.MAP : num   80.1 80.6 85.3 78.9 82.5 ...

#summary stats
summary(Hypoxia[, c("Min.Sao2", "TWA.MAP")])

##      Min.Sao2      TWA.MAP
##  Min.   :28.00   Min.    : 66.00
## 1st Qu.:74.00   1st Qu.: 81.24
## Median :82.00   Median : 88.55
```

```
## Mean :78.71 Mean : 89.01
## 3rd Qu.:86.00 3rd Qu.: 95.94
## Max. :95.00 Max. :127.20

#check missing
colSums(is.na(Hypoxia[, c("Min.Sao2", "TWA.MAP")]))

## Min.Sao2 TWA.MAP
## 0 0
```

Both variables of interest are continuous numeric variable.

X (Min SaO<sub>2</sub>) ranges roughly from 28-95% with a mean of 78.7% and median 82.0%. Most patients have minimum nocturnal oxygen saturation between 74.0% and 86.0%.

TWA MAP ranges from 66.0 mmHg to 127.0 mmHg, with a mean of 89.0 mmHg and median 88.6 mmHg, approximately symmetric. The dataset includes 281 observations and both columns have no missing values.

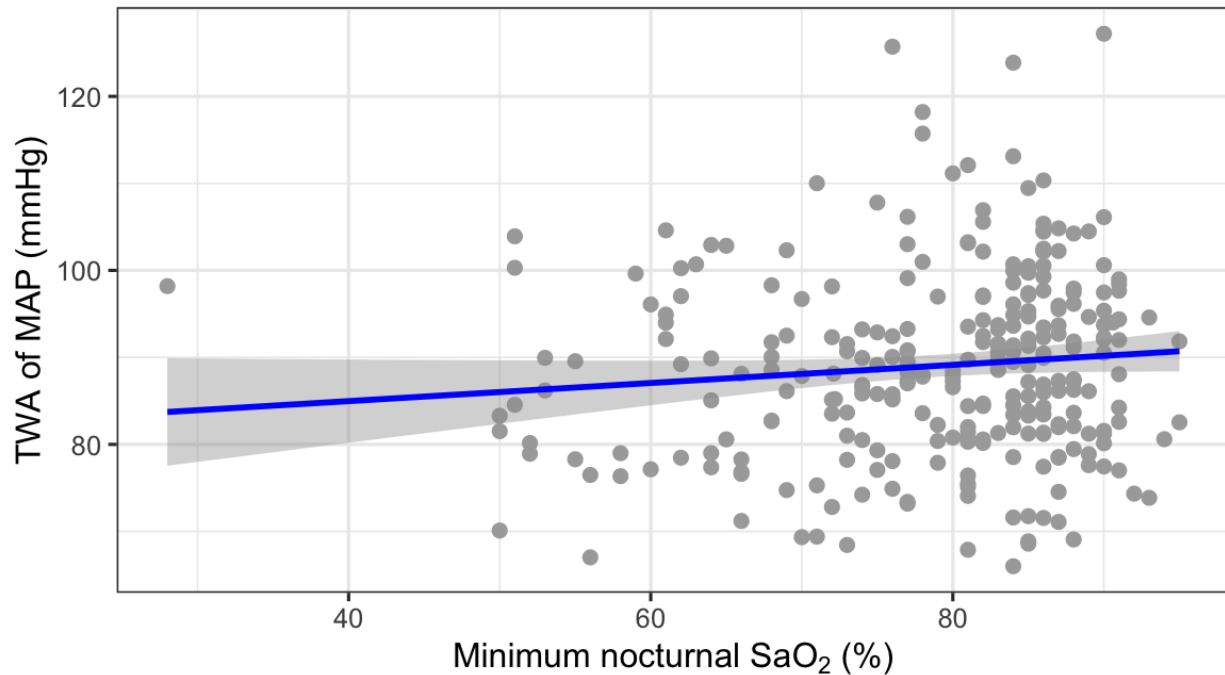
## Question 4

Create a scatterplot of your variables of interest and append a best fit line.

```
# Scatterplot with best-fit line (simple linear regression)
ggplot(Hypoxia, aes(x = Min.Sao2, y = TWA.MAP)) +
  geom_point(color = "darkgray", size = 2) + # individual data points
  geom_smooth(method = "lm", se = TRUE, color = "blue") + # regression line with CI band
  labs(
    title = expression("Relationship between Minimum " * "SaO"[2] * " and Intraoperative MAP"),
    x = expression("Minimum nocturnal " * "SaO"[2] * " (%)"),
    y = "TWA of MAP (mmHg)"
  ) +
  theme_bw(base_size = 12)

## `geom_smooth()` using formula = 'y ~ x'
```

## Relationship between Minimum SaO<sub>2</sub> and Intraoperative MAP



The scatterplot below shows the relationship between patients' minimum nocturnal oxygen saturation and their intraoperative time-weighted average MAP. There is a slight positive trend which indicates that patients with higher minimum SaO<sub>2</sub> tend to have higher MAP. But the relationship appears weak and the data are widely scattered.

### Question 5

Run a simple linear regression and output the summary.

```
# simple linear regression
fit <- lm(`TWA.MAP` ~ `Min.Sao2`, data = Hypoxia)
# summary
summary(fit)

##
## Call:
## lm(formula = TWA.MAP ~ Min.Sao2, data = Hypoxia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.559  -7.892  -0.190   6.541  37.016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  80.80858    4.80699   16.811  <2e-16 ***
## Min.Sao2      0.10418    0.06052    1.721   0.0863 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.75 on 279 degrees of freedom
## Multiple R-squared:  0.01051,    Adjusted R-squared:  0.006961
## F-statistic: 2.963 on 1 and 279 DF,  p-value: 0.08631

# 95% CI for coefficients
confint(fit)

##              2.5 %      97.5 %
## (Intercept) 71.3459947 90.2711577
## Min.Sao2    -0.0149624  0.2233136
```

We fit a simple linear regression model to examine whether minimum nocturnal oxygen saturation Min SaO<sub>2</sub> is associated with intraoperative mean arterial pressure (TWA MAP).

The fitted model is

$$\text{TWA MAP}_i = \beta_0 + \beta_1 \text{Min SaO}_{2i} + \varepsilon_i,$$

where  $\varepsilon_i \sim N(0, \sigma^2)$ .

The estimated regression equation is:

$$\widehat{\text{TWA MAP}} = 80.81 + 0.10 \text{Min SaO}_2,$$

$$\beta_1 = 0.10 \text{ (95\%CI: -0.015, 0.22),}$$

$$p = 0.086.$$

The coefficient of determination was  $R^2 = 0.0105$ , indicating that minimum nocturnal oxygen saturation explains approximately 1% of the variation in intraoperative MAP.

Because the 95% confidence interval for  $\beta_1$  includes 0 and  $p > 0.05$ , there is no statistically significant evidence of a linear association between minimum nocturnal oxygen saturation and intraoperative mean arterial pressure.

The simple linear regression model examined whether minimum nocturnal oxygen saturation (Min SaO<sub>2</sub>) predicts intraoperative time-weighted average mean arterial pressure (TWA MAP). The estimated slope ( $\beta_1$ ) is 0.10 mmHg per 1% SaO<sub>2</sub> (95% CI: -0.02 to 0.22,  $p = 0.09$ ): this means that for every 1% higher minimum oxygen saturation, the mean intraoperative MAP is estimated to be about 0.10 mmHg higher. The 95% confidence interval includes 0, and the  $p$ -value (0.086) is  $> 0.05$ , which indicates not statistically significant. The model  $R^2 = 0.0105$ , indicating that Min SaO<sub>2</sub> explains about 1% of the variability in MAP. The association between minimum nocturnal SaO<sub>2</sub> and intraoperative MAP was positive but not statistically significant ( $\beta = 0.10$  mmHg per 1% SaO<sub>2</sub>, 95% CI -0.02 to 0.22,  $p = 0.09$ ). Minimum oxygen saturation explained very little variation in intraoperative MAP ( $R^2 \approx 0.01$ ). These findings suggest that nocturnal hypoxemia is not meaningfully related to intraoperative blood pressure.

## Question 6

Answer your study question using inference techniques (i.e., report the point estimate, confidence interval, and p-value). Interpret the results in the context of the problem.

Using the simple linear regression model

$$\text{TWA MAP}_i = \beta_0 + \beta_1 \text{ Min SaO}_{2_i} + \varepsilon_i$$

we tested whether minimum nocturnal oxygen saturation Min SaO<sub>2</sub> is linearly associated with intraoperative mean arterial pressure (TWA MAP).

The estimated slope was  $\hat{\beta}_1 = 0.104 \text{ mmHg per } 1\% \text{ SaO}_2$ , 95%CI: (-0.015, 0.223),

$p = 0.086$ .

### Interpretation:

For each 1% increase in minimum nocturnal SaO<sub>2</sub>, the mean intraoperative MAP is estimated to increase by approximately 0.10 mmHg.

However, the 95% confidence interval includes 0 and the p-value (0.086) is greater than the typical significance level of 0.05, so there is no statistically significant evidence of a linear relationship between minimum nocturnal SaO<sub>2</sub> and intraoperative MAP.

The coefficient of determination was  $R^2 = 0.0105$ , indicating that only about 1 % of the variability in intraoperative MAP is explained by minimum nocturnal SaO<sub>2</sub>.

In the context of this study, these findings suggest that the severity of nocturnal hypoxemia is **not a meaningful predictor** of intraoperative blood pressure among OSA patients undergoing bariatric surgery.