STA 108 HW 1

1.2

y = 2x + 300 This is a functional relation because there is no randomness involved. At x visits, we know 100% what y would be.

1.8

Yes, E{Y} will still be 104 because the expected value is calculated over many observations, not on one observation. No, the Y value does not have to be 108, because the sampling is random so the error (\mathcal{E}_i) might be different.

1.12

- a. Observational data, because the predictor variables are uncontrolled and it didn't come from an experimental study.
- b. Validity is questionable because it claims a direct causation between number of colds and time exercised, but we only know that there is a correlation.
- c. Underlying illness, mobility issues, daily environment.
- d. Add more explanatory variables, and make the study a completely randomized experiment.

1.27

See appendix for R code.

- a. My plot supports that musscle mass decreases as age increases because the regression line has a negative slope.
- b. 1. The slope $(\beta_1 = -1.19)$ is the point estimate of difference in mean muscle mass for women differing in age by one year.
 - 2. $\hat{y}_{x=60} = 84.94683$
 - 3. $\mathcal{E}_8 = 4.443252$

1.30

The implication is that the dependent variable is not affected by the independent variable ($\beta_1 = 0$). The regression function ($y = \beta_0 + \mathcal{E}_i$) would be a horizontal line on a graph with a y-intercept of β_0 .

Appendix

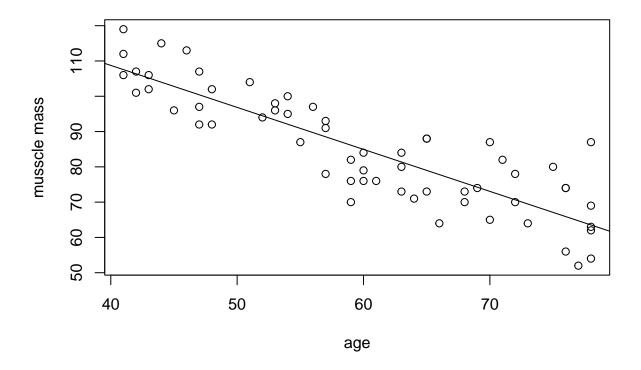
1.27(a)

```
file_path <- "CH01PR27.txt"
data <- read.table(file_path)
y <- data$V1
x <- data$V2</pre>
```

```
model <- lm(y ~ x, data = data)
intercept <- coef(model)[1]
slope <- coef(model)[2]
cat("y = ", round(intercept, 2), " + ", round(slope, 2), "x\n", sep = "")

## y = 156.35 + -1.19x

plot(x, y, xlab = "age", ylab = "musscle mass")
abline(model)</pre>
```



1.27(b)

1

slope

x ## -1.189996

 $\mathbf{2}$

```
new_age <- 60
predict(model, newdata = data.frame(x = new_age))

##     1
## 84.94683</pre>
3
```

model\$residuals[8]

8 ## 4.443252