Quiz 1 Practice Solutions

You can expect about half this number of questions on Quiz #1

Conceptual Exercises

In Sleuth (Chap 9, p. 261) 1, 2, 3, 4, 5, 8

Question One

The weekly gas consumption (in 1000 cubic feet) and the average outside temperature (in degrees Celsius) was recorded for 26 weeks before and 30 weeks after cavity-wall insulation had been installed in a house in south-east England in the 1960s. The house thermostat was set at 20C throughout.

The following model is proposed

$$\mu\{\textit{Gas Consumption}|\textit{Before, Temperature}\} = \beta_0 + \beta_1 \textit{before} + \beta_2 \textit{Temperature}$$

where *before* is an indicator variable for before insulation was installed.

1. What is the model for the mean gas consumption for a week **before** insulation was installed, in terms of the parameters?

$$\beta_0 + \beta_1 + \beta_2 \text{Temperature}$$

2. What is the model for the mean gas consumption for a week **after** insulation was installed, in terms of the parameters?

$$\beta_0 + \beta_2 \text{Temperature}$$

3. Which parameter describes the change in mean gas consumption before and after insulation installation, holding temperature constant?

The change in mean consumption before and after insulation holding temperature constant, is the difference in the above two means. Hence, β_1 captures that change.

Question Two

FEV (forced expiratory volume) is an index of pulmonary function that measures the volume of air expelled after one second of constant effort.

It is of interest whether being a smoker affects FEV, but it is also known the gender and height also affect FEV.

The following model is found to fit well:

$$\mu\{\textit{FEV}|\textit{Height, Female, Smoker}\} = \beta_0 + \beta_1 \textit{height} + \beta_2 \textit{smokerCurrent} + \beta_3 \textit{female} + \beta_4 \textit{smokerCurrent} \times \textit{female}$$

where *height* is height in inches, *smokerCurrent* is an indicator variable for being a currently being a smoker, and *female* is an indicator for being female.

1. What is the effect of *height*, in terms of the parameters?

$$\begin{aligned}\text{Effect of height} &= \mu\{FEV \mid H+1, F, S\} - \mu\{FEV \mid H, F, S\} \\ &= (\beta_0 + \beta_1(H+1) + \beta_2 S + \beta_3 F + \beta_4 S \times F) - (\beta_0 + \beta_1 H + \beta_2 S + \beta_3 F + \beta_4 S \times F) \\ &= \beta_1\end{aligned}$$

2. What is the difference in mean FEV between a **female smoker** and a **female non-smoker** of the same height, in terms of the parameters?

$$\begin{aligned}&\mu\{FEV \mid H, F=1, S=1\} - \mu\{FEV \mid H, F=1, S=0\} \\ &= (\beta_0 + \beta_1 H + \beta_2 1 + \beta_3 1 + \beta_4 1 \times 1) - (\beta_0 + \beta_1 H + \beta_2 0 + \beta_3 1 + \beta_4 0 \times 1) \\ &= (\beta_0 + \beta_1 H + \beta_2 + \beta_3 + \beta_4) - (\beta_0 + \beta_1 H + \beta_3) \\ &= \beta_2 + \beta_4\end{aligned}$$

3. The model fitted to 654 people and the results are shown below. Using the fitted model, what is the fitted mean FEV for a **male non-smoker** 60 inches tall?

```
Call:
lm(formula = FEV ~ Height + Sex + Smoker + Sex:Smoker, data = fev)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6771 -0.2496  0.0026  0.2424  2.0904

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              -5.2284     0.1924  -27.17   <2e-16 ***
Height                    0.1294     0.0031   41.72   <2e-16 ***
SexFemale                -0.1074     0.0356   -3.01   0.0027 **
SmokerCurrent             0.1626     0.0893    1.82   0.0691 .
SexFemale:SmokerCurrent  -0.2160     0.1135   -1.90   0.0575 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.426 on 649 degrees of freedom
Multiple R-squared:  0.76,  Adjusted R-squared:  0.759
F-statistic:  514 on 4 and 649 DF,  p-value: <2e-16
```

$$\begin{aligned}\mu\{FEV \mid H=60, F=0, S=0\} &= \beta_0 + \beta_1 60 + \beta_2 0 + \beta_3 0 + \beta_4 0 \times 0 \\ &= \beta_0 + \beta_1 60 \\ \hat{\mu}\{FEV \mid H=60, F=0, S=0\} &= \hat{\beta}_0 + \hat{\beta}_1 60 \\ &= -5.2284 + 0.1294 \times 60 = 2.5356\end{aligned}$$

Question Three

A researcher is interested in the relationship between weight and body length for three species of rats: black rat (*Rattus rattus*), brown rat (*Rattus norvegicus*) and the Maori rat (*Rattus exulans Peale*).

Let *brown* be an indicator variable for a brown rat, *black* be an indicator variable for a black rat, and *maori* be an indicator for a Maori rat.

1. Write down a model where mean body weight depends linearly on body length with possibly different slopes and intercepts for each species of rat.

$$\mu\{\text{Body weight} \,|\, \text{Length, Species}\} = \beta_0 + \beta_1 Length + \beta_2 black + \beta_3 brown + \beta_4 black \times length + \beta_5 brown \times length$$

(or any other model with just two of the three indicators including the interaction with length)

2. Write down a model where mean body weight depends linearly on body length with the same slope and intercept for all species of rat.

$$\mu\{\text{Body weight} \,|\, \text{Length, Species}\} = \beta_0 + \beta_1 Length$$

3. How many extra parameters are in model 1?

4

Question Four

Consider the following regression model

$$\mu\{Y|X, Z\} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \times Z$$

where X is a continuous variable and Z is also a continuous variable.

1. What is the effect of X?

$$\begin{aligned}\mu\{Y|X+1, Z\} - \mu\{Y|X, Z\} \\ &= (\beta_0 + \beta_1(X+1) + \beta_2 Z + \beta_3(X+1) \times Z) - (\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \times Z) \\ &= (\beta_0 + \beta_1 X + \beta_1 + \beta_2 Z + \beta_3 X \times Z + \beta_3 \times Z) - (\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 X \times Z) \\ &= \beta_1 + \beta_3 \times Z\end{aligned}$$

The effect of X depends on the value of Z.

Question Five

The following model is estimated in R

$$\mu\{\textit{weight}|\textit{height, waist, male}\} = \beta_0 + \beta_1 \textit{male} + \beta_2 \textit{height} + \beta_3 \textit{waist} + \beta_4 \textit{male} \times \textit{waist}$$

where *male* is an indicator variable that the subject is male, *height* is the height of the subject in cm, and *waist* is their waist girth in centimeters, measured at the narrowest part of torso.

1. Which parameter captures the difference in the mean weight for a male compared to a female of the same height and waist?

Mean weight for male:

$$\begin{aligned}\mu\{\textit{weight}|\textit{height, waist, male = 1}\} &= \beta_0 + \beta_1 1 + \beta_2 \textit{height} + \beta_3 \textit{waist} + \beta_4 1 \times \textit{waist} \\ &= \beta_0 + \beta_1 + \beta_2 \textit{height} + (\beta_3 + \beta_4)\textit{waist}\end{aligned}$$

Mean weight for female:

$$\begin{aligned}\mu\{\textit{weight}|\textit{height, waist, male = 0}\} &= \beta_0 + \beta_1 0 + \beta_2 \textit{height} + \beta_3 \textit{waist} + \beta_4 0 \times \textit{waist} \\ &= \beta_0 + \beta_2 \textit{height} + \beta_3 \textit{waist}\end{aligned}$$

Difference in mean:

$$\begin{aligned}&\mu\{\textit{weight}|\textit{height, waist, male = 1}\} - \mu\{\textit{weight}|\textit{height, waist, male = 0}\} \\ &= (\beta_0 + \beta_1 + \beta_2 \textit{height} + \beta_3 \textit{waist} + \beta_4 \textit{waist}) - (\beta_0 + \beta_2 \textit{height} + \beta_3 \textit{waist}) \\ &= \beta_1 + \beta_4 \textit{waist}\end{aligned}$$

The difference depends on the waist size. (*Bad question: difference in mean doesn't depend on a single parameter)*

2. Which parameter captures the difference in the relationship between mean weight and waist for a male compared to a female, keeping height constant?

The relationship between mean weight and waist is captured by the slope with respect to waist. From above, the slope on waist is $\beta_3 + \beta_4$ for males and β_3 for females. Therefore the difference in the relationship is captured by β_4.

3. The results from the fit are shown below. Interpret the estimate for β_2.

```
##
## Call:
## lm(formula = weight ~ male + waist + height + waist:male, data = bdims)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.714  -2.848  -0.161   2.768  20.115
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -96.5079     5.1890  -18.60   <2e-16 ***
## male          8.3477     3.7111    2.25   0.0249 *
## waist         1.0253     0.0361   28.36   <2e-16 ***
## height        0.5188     0.0289   17.92   <2e-16 ***
## male:waist   -0.1489     0.0481   -3.09   0.0021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.39 on 502 degrees of freedom
## Multiple R-squared:  0.892,  Adjusted R-squared:  0.892
## F-statistic: 1.04e+03 on 4 and 502 DF,  p-value: <2e-16
```

β_2 multiplies *height* in the regression model. Since *height* only occurs once on the RHS of the regression model, β_2 is the effect of height. Read the estimate off the `height` line in the output: 0.5188.

It is estimated that each 1cm increase in height is associated with an increase in mean weight of 0.52 kgs, after accounting for waist size and gender.

Question Six

Consider the following model of total length to head length for possums in Australia:

$$\mu\{\textit{totalL}|\ \textit{headL, sex}\} = \beta_0 + \beta_1 headL + \beta_2 sexm$$

where *totalL* is the total length of a possum in mm, *headL* is the length of it's head in mm, and *sexm* is an indicator variable for the possum being male.

1. Write down the model for the mean total length as a function of the head length for a **female possum**.

For a female possum, *sexm* = *0*:

$$\mu\{\textit{totalL}|\ \textit{headL, sex}\} = \beta_0 + \beta_1 headL + \beta_2 0 = \beta_0 + \beta_1 headL$$

2. Write down the model for the mean total length as a function of the head length for a **male possum**.

For a male possum, *sexm* = *1*:

$$\mu\{\textit{totalL}|\ \textit{headL, sex}\} = \beta_0 + \beta_1 headL + \beta_2 1 = (\beta_0 + \beta_2) + \beta_1 headL$$

3. What kind of model is this? (i.e. equal lines, parallel lines, separate lines)

Same slope, different intercept. Parallel lines.

4. The model is fit in R and the results are shown below, write a sentence interpreting β_2.

```
##
## Call:
## lm(formula = totalL ~ headL + sex, data = possum, na.action = na.exclude)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.469 -2.065  0.583  1.777  8.447
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.2610     7.6149    1.08  0.28057
## headL         0.8643     0.0825   10.48  < 2e-16 ***
## sexm         -2.0646     0.5957   -3.47  0.00078 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.97 on 101 degrees of freedom
## Multiple R-squared:  0.533,  Adjusted R-squared:  0.524
## F-statistic: 57.7 on 2 and 101 DF,  p-value: <2e-16
```

β_2 is the difference in the intercept between male and female possums (or equivalently the effect of being male). Find $\hat{\beta}_2$ on the sexm line, -2.0646.

It is estimated that mean total length for male possums is 2.06 units lower than the mean total length for female possums with the same head length.