**Answer Key**

1. A graph with dots and lines

   Description automatically generated Fit the following models and create scatter plots with regression lines for each of them. Record their R2 values and comment on the linearity.
   * 1. Linearity:

Strong linearity with slight variation.

```{r}

mod\_3 <- lm(Rank4 ~ Rank1, data = ski\_speeds)

summary(mod\_3)

plot(Rank4 ~ Rank1, data = ski\_speeds)

abline(mod\_3)

```

* + 1. A graph with dots and numbers

       Description automatically generatedR2: 83.7%
    2. Linearity:

Poor linearity with a slight s-curve shape.

```{r}

mod\_1 <- lm(Rank4 ~ Time1, data = ski\_speeds)

summary(mod\_1)

plot(Rank4 ~ Time1, data = ski\_speeds)

abline(mod\_1)

```

* + 1. A graph with dots and numbers

       Description automatically generatedR2: 77.7%
    2. Linearity:

Strong linearity.

```{r}

mod\_2 <- lm(Time4 ~ Time1, data = ski\_speeds)

summary(mod\_2)

plot(Time4 ~ Time1, data = ski\_speeds

abline(mod\_2)

```

* + 1. R2: 85.8%
    2. A graph with dots and lines

       Description automatically generatedLinearity:

Bad linearity, very clear s-curve shape.

```{r}

mod\_4 <- lm(Time4 ~ Rank1, data = ski\_speeds)

summary(mod\_4)

plot(Time4 ~ Rank1, data = ski\_speeds)

abline(mod\_4)

```

* + 1. R2: 81.9%

1. Which one of the models from Question 1 do you think is best? Why?

is the best model as it has both the highest R2 value, and the scatterplot shows a clear and strong linear relationship.

1. Interpret the slope coefficient for the model you just chose in Question 2. Remember to contextualize.

For every additional place in Rank1, Rank4 increases by 0.88 places.

```{r}

summary(mod\_3)

```

1. If a skier is ranked #14 after the first split, what rank would you predict for her finish?

The predicted Rank4 for a skier with a Rank1of 14, is 15.25 which can be rounded to 15th place at the finish.

```{r}

mod\_3 <- lm(Rank4 ~ Rank1, data = ski\_speeds)

summary(mod\_3)

Rank1 <- 14

Rank4 <- 2.95897 + 0.87807\*Rank1

Rank4

```

1. Interpret the slope coefficient for the best of these models to predict Time4. Remember to contextualize.

For every additional minute Time1 increases, Time4 increases by 4.66 minutes.

```{r}

summary(mod\_2)

```

1. If a skier had a time of 5.00 minutes after the first split, what time would you predict for her finish?

The predicted Time4 for a skier with a Time1 of 5 minutes is 30.07 minutes.

```{r}

mod\_2 <- lm(Time4 ~ Time1, data = ski\_speeds)

summary(mod\_2)

Time1 <- 5

Time4 <- 6.7561 + 4.6618\*Time1

Time4

```

1. Can it be concluded that the time and/or rank of a skier after the first distance interval has a substantial impact on their finishing rank and/or time? Explain your answer.

Yes, these models show a strong relationship between either Time4 and Time1 or Rank1 and Rank4. They both show that the if Rank1 or Time1 were to increase, so would Time4 or Rank4. This means that if a skier performs poorly in the beginning of a 10k, they are unlikely to improve their performance by the end of the race.