

Part 1: Knowledge Sync — The Surveyor’s License (5 Minutes)

Complete these four checks from *Chapter 5 of r.statypus.org* to ensure your tools are calibrated.

1. The _____ (or r^2) measures the percentage of variation in y explained by x .
2. The vertical distance from an observed point to our line is the _____.
3. The Best-Fit line minimizes the Sum of the _____ Residuals.
4. The regression line is mathematically guaranteed to pass through the coordinates: (____, ____).

Part 2: Manual Sighting — The “Ruler Trick” (10 Minutes)

Run the following code in R to prepare your data. Then, use R to find your anchor point and follow the steps below to replicate the survey in the box at the bottom of this page.

```
# Load the datasets and select the Beef hot dogs
load(url("https://statypus.org/files/StatypusCh5.RData"))
Beef <- hot_dogs[hot_dogs$type == "Beef", ]
```

Instructions:

1. **Generate the Map:** Review **Section 5.1** to use the `plot()` function for Calories (as x) vs. Sodium (as y).
2. **Sketch the Scatterplot:** Look at your R plot. In the box below, sketch the points as accurately as you can.
3. **Mark the Anchor:** Use R to find the exact values for \bar{x} and \bar{y} . Mark that coordinate on your sketch with a **colored pen or pencil**.
4. **Sighting the Path:** Place your ruler on your colored mark and tilt it until the line **looks like the best possible fit** for the points. Draw this line across the entire plot area.

Part 3: The Digital Survey (10 Minutes)

Manual sighting with a ruler is a great start, but to be a professional surveyor, we need the mathematical precision of the Regression Line. In Chapter 5, we learned that technology can find the exact slope and intercept that minimizes the “noise” in our data.

Use the hot dog data already loaded in R to calculate the official Regression Line for this survey.

Analysis Check:

1. **The Official Line:** Review **Section 5.7.1** to use the `LinearRegression()` function to redo the Calories vs. Sodium plot from Part 2. Once you see the output, go back to your sketch on **Page 1**. Using a **different colored pen or pencil**, draw the official Regression Line shown in your R plot.
2. **The Visibility:** Your R output provides the correlation coefficient (r). Find the **Coefficient of Determination**:

$$r^2 \approx \underline{\hspace{2cm}} \%$$

Based on this percentage, why did it feel easy or difficult to “aim” your ruler?

3. **The Prediction:** Use the estimates from your output to write the full equation for the **Regression Line** in the form $\hat{y} = a + bx$:

$$\hat{y} = \underline{\hspace{10cm}}$$

4. **The Edge Effect:** If you pivot your ruler slightly at your colored anchor point (\bar{x}, \bar{y}) , where does the predicted value \hat{y} move the most: near the center or at the far edges of the plot?

Reflections from the Field: The “Hourglass” Effect

Thinking about your answer to Question 4, why is our estimate of the path less “stable” as we move further away from the (\bar{x}, \bar{y}) anchor? If we were to repeat this survey with a new sample, where would you expect the most variation in our predictions?