L3. Correlation and Regression

Example 1: Plastic Bags

A Canadian grocery store's plan to shame customers out of using plastic bags backfired spectacularly. Vancouver's East West Market printed embarrassing phrases like "Dr. Toews' Wart Ointment Wholesale", hoping to deter use—but instead, people loved them and flocked to collect them. If guilt and humiliation aren't enough to change consumer behaviour, then Kenya's especially draconian ban on plastic bags might help; anyone caught selling, producing, or even carrying a plastic bag faces a $38000 fine or four years in prison.

The decomposition time for biodegradable plastic bags depends on moisture levels. The table below shows how long they take to break down based on environmental humidity.

Moisture in Environment (%)	2	3	4	7	8	9	12	13	14	16	17
Time to Decompose (days)	39	36	36	34	33	31	28	27	26	25	23

a. Make a scatter plot of the data.

b. Calculate the sample covariance between the moisture level in the environment and the number of days it takes for the bag to decompose.

c. Calculate the coefficient of correlation, r, for the given data. Comment on the strength and direction of the relationship.

d. Calculate the coefficient of determination, R^2, for the given data. Comment on the amount of variation that is accounted for by this data.

e. Calculate the slope of the least squares line and interpret it in the context of the problem.

f. Calculate the y-intercept of the least squares line and interpret it in the context of the problem.

g. Estimate how long it would take for a bag to decompose if there is 18% moisture in the air. Comment on the reliability of this estimate.

h. Predict how long it will take for a bag to fully decompose in an environment that contains 5% moisture. Comment on the reliability of this estimate.

i. Scientists observed that it took 33 days for a bag to fully decompose in an environment that had 8.5% moisture in it. Calculate the residue, and comment on the performance of the model

Solution

Example 2: Stealthy Starbucks

With over 29,000 locations across the U.S., Starbucks is everywhere—including inside the CIA headquarters in Langley, Virginia. This ultra-secret café, known as "Stealthy Starbucks", is so classified that it doesn't show up on GPS or on Google Maps. But the secrecy doesn't stop there. Receipts are titled 'Store Number 1' and no names—real or fake—are written on cups. Despite its clandestine nature, Stealthy Starbucks is the busiest in the world, as CIA agents rarely leave the premises. Its top-selling items? Lemon pound cake and Frappuccinos.

The amount of sugar and caloric count for some of Starbucks' more popular drinks are shown in the table below:

Sugar (g)	42	35	40	18	47	54	30	36
Calories (kcal)	290	250	260	180	300	290	190	230

a. Make a scatter plot of the data.

b. Calculate the coefficient of correlation, r, for the given data. Comment on the strength and direction of the relationship.

c. Calculate the coefficient of determination, R^2, for the given data. Comment on the amount of variation that is accounted for by this data.

d. Calculate the slope of the least squares line and interpret it in the context of the problem.

e. Calculate the y-intercept of the least squares line and interpret it in the context of the problem.

f. Make a prediction for a drink that contains 60 g of sugar. Is this estimate trustworthy?

g. Make a prediction for a drink that contains 10 g of sugar. Is this estimate reliable?

h. A Starbucks coffee beverage that has $37g$ of sugar, actually contains 250 calories. Calculate the residual and interpret the result.

Solution

Example 3: Hot vs. Cold

You're more likely to devour a meal if it's cold—because, apparently, hot food tricks your brain into thinking you're full faster. So if you've ever wondered why ice cream disappears faster than a bowl of oatmeal, science has your answer. To put this to the test, researchers served participants identical portions of soup at different temperatures and measured how much they actually ate. The results are shown in the table below.

Temperature (C)	75	65	55	45	35	25	15	5
Average Consumption (g)	250	270	300	340	390	450	510	580

a. Make a scatter plot of the data.

b. Calculate the coefficient of correlation, r, for the given data. Comment on the strength and direction of the relationship.

c. Calculate the coefficient of determination, R^2, for the given data. Comment on the amount of variation that is accounted for by this data.

d. Calculate the slope of the least squares line and interpret it in the context of the problem.

e. Calculate the y-intercept of the least squares line and interpret it in the context of the problem.

f. Predict how much soup would be consumed if it were served at $10°C$. Comment on the reliability of the estimate.

g. Predict how much soup would be consumed if it were served at $85°C$. Comment on the reliability of the estimate.

h. Suppose that soup served at $20°C$ resulted in an average consumption of $475\,g$ of soup, calculate the residual and comment on the result.

Solution