

15.4.1: Practice with Nutrition

Let's try a second example using the same [nutrition data set from fast food restaurants](#) from [OpenIntro.org](#) as the [chapter on correlations](#). If you grab your calculations and Sum of Products table from that [practice with the nutrition data](#), you'll be ahead of the game!

Scenario

When Dr. MO was on a low-carb diet, she noticed that things that were low in carbohydrates tended to be high in fat, and vice versa. Our prior analysis found a strong correlation between fat and carbs, but not in the direction that Dr. MO expected! Now, we will use our knowledge of this strong correlation to see if we can use the Total Fat in our sample chicken dishes from one fast food restaurant to predict the Total Carbs in those same dishes.

Step 1: State the Hypotheses

Because of trying to simplify regression, our hypotheses are still a little clunky, but hopefully you'll see the value of regression analyses in the end.

✓ Example 15.4.1.1

What could be a research hypothesis for this scenario? State the research hypothesis in words and symbols.

Solution

The research hypothesis should probably be something like, "There will be a positive slope in the regression line for Total Fat and Total Carbs."

Symbols: $\beta > 0$

In other words, we're trying to construct a statistically significant (as determined by an ANOVA) regression line equation that could use to predict Total Carbs from Total Fat.

✓ Example 15.4.1.2

What is the null hypothesis for this scenario? State the null hypothesis in words and symbols.

Solution

The null hypothesis is that there is no relationship between Total Fat and Total Carbs, so "The slope in the regression line for Total Fat and Total Carbs will be zero." Neither variable can predict or explain the other variable.

Symbols: $\beta = 0$

Step 2: Find the Critical Value

Our critical value for regression will come from the [Table of Critical Values of F](#), which shows that the appropriate critical value at $p=0.05$ is $F_{Critical}(1, 11) = 4.84$.

Step 3: Calculate the Test Statistic

Okay, here's where the tedious part starts! Use the data in Table 15.4.1.1 to start your calculations!

Table 15.4.1.1: Raw Scores in Empty Sum of Products Table

Total Fat	Difference: Fat - Mean	Fat Difference Squared	Total Carbs	Difference: Carbs - Mean	Carbs Difference Squared	Fat Diff * Carb Diff
11			19			
12			30			
13			31			
17			28			

Total Fat	Difference: Fat - Mean	Fat Difference Squared	Total Carbs	Difference: Carbs - Mean	Carbs Difference Squared	Fat Diff * Carb Diff
24			53			
24			26			
26			52			
28			47			
31			66			
31			56			
35			53			
44			81			
46			53			
$\sum = 342.00$	$\sum = ?$	$\sum = ?$	$\sum = 595.00$	$\sum = ?$	$\sum = ?$	$\sum = ?$

To start filling in the Table 15.4.1.1, we'll need to find the means again for Total Fat and Total Carbs. Since we did that in the last chapter, let's just put them in a Table 15.4.1.2

Table 15.4.1.2- Descriptive Statistics of Nutrition Information

	Mean	Standard Deviation	N
Total Fat	26.31	11.32	13
Total Carb	45.77	17.89	13

Now that we have the means, we can fill in the complete Sum of Products table.

? Exercise 15.4.1.1

Fill in Table 15.4.1.1 by finding the differences of each score from that variable's mean, squaring the differences, multiplying them, then finding the sums of each of these.

Answer

Completed Sum of Products table:

Table 15.4.1.3: Completed Sum of Products Table

Total Fat	Difference: Fat - Mean	Fat Difference Squared	Total Carbs	Difference: Carbs - Mean	Carbs Difference Squared	Fat Diff * Carb Diff
11	-15.31	234.40	19	-26.77	716.63	409.85
12	-14.31	204.78	30	-15.77	248.69	225.67
13	-13.31	177.16	31	-14.77	218.15	196.59
17	-9.31	86.68	28	-17.77	315.77	165.44
24	-2.31	5.34	53	7.23	52.27	-16.70
24	-2.31	5.34	26	-19.77	390.85	45.67
26	-0.31	0.10	52	6.23	38.81	-1.93
28	1.69	2.86	47	1.23	1.51	2.08
31	4.69	22.00	66	20.23	409.25	94.88
31	4.69	22.00	56	10.23	104.65	47.98

Total Fat	Difference: Fat - Mean	Fat Difference Squared	Total Carbs	Difference: Carbs - Mean	Carbs Difference Squared	Fat Diff * Carb Diff
35	8.69	75.52	53	7.23	52.27	62.83
44	17.69	312.94	81	35.23	1241.15	623.22
46	19.69	387.70	53	7.23	52.27	142.36
$\Sigma = 342.00$	$\Sigma = 0.00$	$\Sigma = 1536.77$	$\Sigma = 595.00$	$\Sigma = -0.01$	$\Sigma = 3842.31$	$\Sigma = 1997.92$

We can do a computation check to see if our difference scores (each score minus the mean for that variable) were calculated correctly by seeing if they each sum to nearly zero, which they do; yay! We could use the sum of those squares to calculate the standard deviation for each variable, but that was already provided in Table 15.4.1.2

So, we have what we need to complete this activity. But what is this activity? Ultimately, our goal is to predict Total Carbs using Total Fat with a regression line equation ($\hat{y} = a + bx$). So, we need to find the slope (b) and the intercept (a).

Here's the formula for slope:

$$b = \frac{(Diff_x \times Diff_y)}{Diff_x^2} = \frac{(Diff_F \times Diff_C)}{Diff_F^2}$$

What this equation is telling us to do is pretty simple since we have the Sum of Products table filled out.

✓ Example 15.4.1.3

Calculate the slope for this scenario.

Solution

$$b = \frac{1997.92}{1539.77} = 1.30$$

The result means that as Total Fat (X) changes by 1 unit, Total Carb (Y) will change by 1.30. This is a positive relation.

Next, we use this slope (b = 1.30), along with the means of each variable, to compute the intercept:

$$a = \bar{X}_y - (b \times \bar{X}_x) = \bar{X}_C - (b \times \bar{X}_F)$$

✓ Example 15.4.1.4

Using the means and the slope (b) that we just calculated, calculate the intercept (a).

Solution

$$a = 45.77 - (1.30 \times 26.31)$$

$$a = 45.77 - (34.20)$$

$$a = 11.57$$

Now that we have all of our parameters estimated, we can give the full equation for our line of best fit: $\hat{y} = a + bx$

✓ Example 15.4.1.5

Construct the regression line equation for to predict Total Carbs (\hat{y}).

Solution

$$\hat{y} = 11.57 + 1.30x$$

Let's look at that regression line on our scatterplot generated using statistical software.

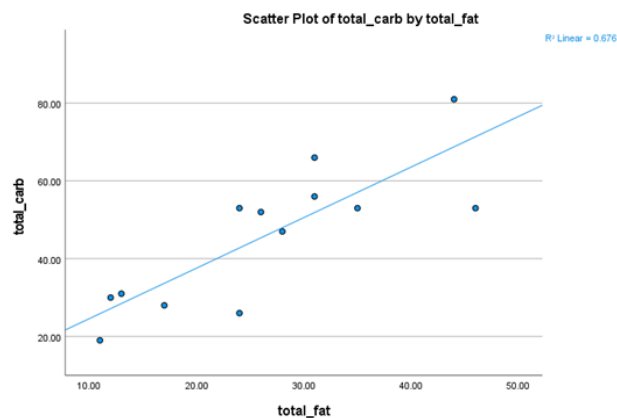


Figure 15.4.1.1: Scatterplot and Regression Line of Fast Food Nutrition Data (CC-BY-SA Michelle Oja via data from OpenIntro.org)

It looks like our regression line is a pretty good match for our data! But to make sure, let's make sure that we have a statistically significant model using this regression line. Table 15.4.1.4 shows an ANOVA Summary Table with the Sum of Squares already included from an analysis conducted using statistical software.

Table 15.4.1.4: ANOVA Summary Table with SS

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Model	1038.88			
Error	497.89			
Total				

✓ Example 15.4.1.6

Fill in the rest of the ANOVA Summary Table from Table 15.4.1.4

Solution

Table 15.4.1.5: ANOVA Summary Table

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Model	1038.88	1	$MS_M = \frac{SS_M}{df_M} = \frac{1038.88}{1}$	$F = \frac{MS_M}{MS_E} = \frac{1038.88}{45.26} = 22.95$
Error	497.889	$N - 2 = 13 - 2 = 11$	$MS_E = \frac{SS_E}{df_E} = \frac{497.89}{11}$	N/A
Total	173.99	$N - 1 = 13 - 1 = 14$	N/A	N/A

We can do a computation check to make sure that our Degrees of Freedom are correct since the sum of the Model's *df* and the Error's *df* should equal the Total *df*. And since $1 + 11 = 12$, we're on the right track.

This gives us an obtained *F* statistic of 22.95, which we will now use to test our hypothesis.

Step 4: Make the Decision

We now have everything we need to make our final hypothesis testing decision. Our calculated test statistic was $F_{Calc} = 22.95$ and our critical value was $F_{Crit} = 4.84$. Since our calculated test statistic is greater than our critical value, we can reject the null hypothesis because this is *still* true:

Note

(Critical < Calculated) = Reject null = There is a linear relationship. = $p < .05$
 (Critical > Calculated) = Retain null = There is not a linear relationship. = $p > .05$

Write-Up: Reporting the Results

Let's use the [four components needed for reporting results](#) to organize our conclusions so far.

✓ Example 15.4.1.7

Add text here Describe the four components for reporting results for this scenario in complete sentences.

Solution

1. For our sample of 13 chicken dishes from one fast food restaurant, the average Total Fat was 26.31 and the average Total Carb was 45.77.
2. The research hypothesis was that there will be a positive slope in the regression line for Total Fat and Total Carbs.
3. The ANOVA results were $F(1,11)=22.95$, $p<.05$, showing that our regression line could be used to make predictions.
4. The results are statistically significant, and the positive Sum of Products shows a positive slope so the research hypothesis is supported.

Using Regression

But really, what does that even mean? How is any of that useful? Well, it turns out that because our regression equation was found to be statistically significant, we can use it to make predictions.

✓ Example 15.4.1.8

Imagine that you went to the fast food restaurant that this data is from, and they had a new chicken meal with even lower fat than anything currently on their menu (10 grams of Total Fat). Use the regression line equation to estimate how many Total Carbs would be in this new meal. Don't forget to end with a complete sentence.

Solution

With this regression line equation:

$$\hat{y} = 11.57 + 1.30x$$

We would replace the "x" placeholder with our new value of 10 grams of Total Fat to find:

$$\hat{y} = 11.57 + (1.30 * 10)$$

$$\hat{y} = 11.57 + (13.00)$$

$$\hat{y} = 24.57$$

The estimated Total Carbs of this new chicken meal with 10 Total Fat is 24.57 grams of carbohydrates.

That's not what we hypothesized, but this is the whole point of regression, to use a variable (or variables) that we have to predict variables that we don't have. Knowing this, let's try a final conclusion.

In our sample of 13 chicken dishes from one fast food restaurant, we found the average Total Fat to be 26.31 and the average Total Carbs to be 45.77. This data was used to create a statistically significant regression equation ($F(1, 11)=22.95$, $p<.05$), which supports the research hypothesis that there would be a positive slope. If the fast food restaurant came out with a new low-fat chicken dish (10 grams of fat), we can predict that the Total Carbs of that dish would be 24.57 grams.

Accuracy in Prediction

Wanna know how accurate that prediction is? Square root the Mean Square of the Error term from the ANOVA Summary Table. As this formula shows, this is similar to the standard deviation formula because it follows the same logic:

$$S_{(Y-\hat{Y})} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{N - 2}}$$

But since we have the MS of the Error in Table 15.4.1.5 we can just square root 42.26:

$$s_{(Y-\hat{Y})} = \sqrt{MS_E} = \sqrt{45.26} = 6.73$$

So on average, our predictions will be almost 7 (6.73) grams away from the actual values. There are no specific cutoffs or guidelines for how big our standard error of the estimate can or should be. What do you think? Is a variation around the prediction of 7 grams of carbohydrates precise enough? There's no right or wrong answer, just your thoughts and opinion!

Hopefully you saw how regression equations can be useful for making predictions about values that we don't have yet based on samples of two variables that we do already have data for. In the next section, you'll learn how we can combine multiple variables to predict one variable.

Contributors and Attributions

- [Foster et al.](#) (University of Missouri-St. Louis, Rice University, & University of Houston, Downtown Campus)
- [Dr. MO](#) (Taft College)

This page titled [15.4.1: Practice with Nutrition](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).