

What Are Pearson Residuals? (Definition & Example)

Authored by
stats writer

May 6, 2024

RECOMMENDED CITATION

stats writer (2024). *What Are Pearson Residuals? (Definition & Example)*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143271>

Pearson residuals are a statistical measure used to evaluate the goodness of fit for a regression model. They are calculated by taking the difference between the observed and predicted values and dividing it by the square root of the expected value. This measure helps to identify any patterns or outliers in the data that may affect the accuracy of the model. For example, if a regression model is used to predict housing prices, Pearson residuals can show if there are any significant differences between the predicted and actual prices, indicating potential flaws in the model. Overall, Pearson residuals are an essential tool in assessing the reliability and accuracy of regression models.

What Are Pearson Residuals? (Definition & Example)

Pearson residuals are used in a to analyze the difference between observed cell counts and expected cell counts in a contingency table.

The formula to calculate a Pearson residual is:

$$r_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}}$$

where:

r_{ij} : The Pearson residual for the cell in the i th column and j th row
 O_{ij} : The observed value for the cell in the i th column and j th row
 E_{ij} : The expected value for the cell in the i th column and j th row

A similar metric is the Standardized (adjusted) Pearson residual, which is calculated as:

$$r_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}(1-n_{i+})(1-n_{+j})}$$

where:

r_{ij} : The Pearson residual for the cell in the i th column

and j th row

O_{ij} : The observed value for the cell in the i th

column and j th row

E_{ij} : The expected value for the cell in

the i th column and j th row

n_{i+} : The row total divided by

the grand total

n_{+j} : The column total divided by the

grand total

Standardized Pearson residuals are normally

distributed with a mean of 0 and standard deviation of

1. Any standardized Pearson residual with an absolute

value above certain thresholds (e.g. 2 or 3) indicates a

lack of fit.

The following example shows how to calculate Pearson

residuals in practice.

Example: Calculating Pearson Residuals

Suppose researchers want to use a Chi-Square Test of

Independence to determine whether or not gender is

associated with political party preference.

They decide to take a simple random sample of 500 voters and survey them on their political party preference.

The following contingency table shows the results of the survey:

	Republican	Democrat	Independent	Total
Male	120	90	40	250
Female	110	95	45	250
Total	230	185	85	500

Before we calculate the Pearson residuals, we must first calculate the expected counts for each cell in the contingency table. We can use the following formula to do so:

For example, the expected value for Male Republicans is: $(230 \times 250) / 500 = 115$.

We can repeat this formula to obtain the expected value for each cell in the table:

	Republican	Democrat	Independent	Total
Male	115	92.5	42.5	250
Female	115	92.5	42.5	250
Total	230	185	85	500

Next, we can calculate the Pearson residual for each cell in the table.

For example, the Pearson residual for the cell that contains Male Republicans would be calculated as:

$$r_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}} \quad r_{ij} = (120 - 115) / \sqrt{115} \quad r_{ij} = 0.466$$

We can repeat this formula to obtain the Pearson residual for each cell in the table:

	Republican	Democrat	Independent
Male	0.446	-0.259	-0.383
Female	-0.446	0.259	0.383

Next, we can calculate the Standardized Pearson residual for each cell in the table.

For example, the Standardized Pearson residual for the cell that contains Male Republicans would be calculated as:

$$r_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}(1-p_i)(1-p_j)} \quad r_{ij} = (120 - 115) / \sqrt{115(1-250/500)(1-230/500)} \quad r_{ij} = 0.897$$

We can repeat this formula to obtain the Standardized Pearson residual for each cell in the table:

	Republican	Democrat	Independent
Male	0.897	-0.463	-0.595
Female	-0.897	0.463	0.595

We can see that none of the Pearson Standardized Residuals have an absolute value greater than 3, which indicates that none of the cells contribute to a significant lack of fit.

If we use to perform a Chi-Square Test of Independence, we'll find that the p-value of the test is **0.649198**.

	Group 1	Group 2	Group 3	Group 4	Group 5
Category 1	120	90	40		
Category 2	110	95	45		
Category 3					
Category 4					
Category 5					

CALCULATE

χ^2 Test Statistic: **0.864035**

p-value: **0.649198**

Since this p-value is not less than .05, we do not have sufficient evidence to say that there is an association between gender and political party preference.

The following tutorials explain how to perform a Chi-Square Test of Independence using different statistical software:

ARABPSYCHOLOGY.COM