Data Mining—Correlations

The Food and Agriculture Administration of the United Nations estimates that there are 925 million people who are hungry. Food insecurity is thought by many to be the world's most solvable problem, yet it continues to be a major global issue. This is highlighted by the fact that "Eradicating Extreme Poverty and Hunger" is the first all of the Millennium Goals set by the United Nations.

When considering issues of global health, a data-driven approach is necessary to understand the relationships that may be at play. In this Activity you will analyze the data of 42 different countries. These nations were selected for study because their data for the ***Percentage of Underweight Children*** was available. Please note that Germany and the United States are each missing two data points, and you will exclude those variables in your correlation calculations for those countries.

You will be analyzing *correlations,* not *causation*. It is important to understand that correlation means that two variables may be related, but it **does not mean** "cause and effect."

In this assignment, you will use the "**42 Countries for Data Analysis"** data to compute the correlation coefficient for each of the variables as they compare the correlations between the ***Percentage of Underweight Children*** (those who weigh less than two standard deviations below average for their age) and the other given variables. Through this analysis you may uncover interesting relationships that can be analyzed to determine what is related to reduced food scarcity and to determine what may help eradicate extreme poverty. By using a large number of countries, the results will be more statistically significant. Dx—fjg

The *Correlation Coefficient,* denoted by the variable "*r*", is a measurement that determines how two variables are related to each other. This does not mean that one variable causes the other, but it implies that they are related. The value of a correlation coefficient is always between -1 and +1 (-1< *r* <+1). You will use the following table to evaluate the relative strength or weakness of the correlation.

-.99 to -.4	-.39 to -.20	-.19 to .19	.20 to .39	.4 to .99
Strong Negative Relationship	Weak Negative Relationship	No or Negligible Relationship	Weak Positive Relationship	Strong Positive Relationship
(Increase in one variable predicts a decreases in the other)				**(Increase in one variable predicts an increases in the other)**

While TI graphing calculators and Microsoft Excel can make this calculation in seconds, the following formula can be used. (Note: ∑ x means the sum of all x's)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Task

In groups, divide the responsibilities of the different variables, calculate the correlation coefficient using the formula given above (program into TI-Graphing Calculators), and answer the questions.

1. Complete the following table to determine the strength of the correlations.

Correlations of Variables with Underweight Children Statistics		
Variable	**Correlation Coefficient**	**Strength of Correlation**
Gross Domestic Product		
Total Population		
GDP per Capita		
Infant Mortality		
Life Expectancy at Birth		
% of Population that Earn Below $2 a day		
Human Development Index		
Food Supply		
Agriculture percentage of GDP		
Internet Users		
Average Years of School for Girls		

2. Rank the variables by strength of correlation. Which variables appear to have strong correlations with *Percent of Underweight Children*?

3. Sometimes correlations can be deceptively high because the variables both measure the same thing. Which, if any, of the correlations could be explained by this phenomenon?

4. Based on the strength of these correlations, which of these variables appear to be most related to underweight children? (Remember that this is not a cause and effect relationship.)

5. Use the data and your findings to suggest possible solutions that a government could use to alleviate extreme poverty and hunger.

6. Reflect on your findings. What was the most surprising result? Explain.

Data Mining—Correlations

In groups, divide the responsibilities of the different variables, calculate the correlation coefficient using the formula given above (program into TI-Graphing Calculators), and answer the questions.

1. Complete the following table to determine the strength of the correlations.

Correlations of Variables with Underweight Children Statistics		
Variable	**Correlation Coefficient**	**Strength of Correlation**
Gross Domestic Product		
Total Population		
GDP per Capita		
Infant Mortality		
Life Expectancy at Birth		
% of Population that Earn Below $2 a day		
Human Development Index		
Food Supply		
Agriculture percentage of GDP		
Internet Users		
Average Years of School for Girls		

2. Rank the variables by strength of correlation. Which variables appear to have strong correlations with *Percent of Underweight Children*?

3. Sometimes correlations can be deceptively high because the variables both measure the same thing. Which, if any, of the correlations could be explained by this phenomenon?

4. Based on the strength of these correlations, which of these variables appear to be most related to underweight children? (Remember that this is not a cause and effect relationship.)

5. Use the data and your findings to suggest possible solutions that a government could use to alleviate extreme poverty and hunger.

6. Reflect on your findings. What was the most surprising result? Explain.

42 Countries for Data Analysis

Country	Gross Domestic Product GDP (in millions $)	Total Population	GDP per Capita (Income per person)	**Under-weight for age % of children**	Infant Mortality (per 1000 births)	Life Expectancy at Birth	% of Population that Earn Below $2 a Day	Human Development Index HDI	Food Supply Average Calories per Day	Agriculture (% of GDP)	Internet Users (per 100 people)	Average Years of School for Girls
	2010	*2010*	*2010*	***Most recent year 00-09***	*2010*	*2011*	*Most recent year 00-10*	*2011*	*2007*	*2011*	*2010*	*2009*
Argentina	716,500	40,412,376	$10,749	**2.3**	10.52	75.901	1.87	0.80	2,941	9.09	36.00	11.4
Armenia	17,970	3,092,072	$1,327	**4.2**	18.21	74.241	12.43	0.72	2,279.9	20.66	44.00	11.4
Bangladesh	283,500	148,692,131	$558	**41.3**	48.99	68.944	76.54	0.50	2,281.2	18.43	3.70	4.7
Belarus	141,800	9,595,421	$2,738	**1.3**	3.70	70.349	0.19	0.76	3,145.6	8.13	32.05	12.4
Bolivia	50,940	9,929,849	$1,233	**4.5**	40.94	66.618	24.89	0.66	2,064.1	11.75	20.00	9.0
Brazil	2,294,000	194,946,470	$4,699	**2.2**	20.50	73.488	10.82	0.72	3,112.5	5.46	40.65	9.0
Cambodia	33,820	14,138,255	$558	**28.8**	54.08	63.125	53.27	0.52	2,267.6	36.02	1.26	4.7
Chile	299,500	17,113,688	$6,334	**0.5**	7.40	79.12	2.72	0.81	2,920.4	3.40	45.00	11.9
China	11,300,000	1,341,335,152	$2,425	**4.6**	15.62	73.456	29.79	0.69	2,980.5	10.04	34.38	8.5
Republic of the Congo	18,070	4,042,899	$1,253	**11.8**	74.22	57.379	74.40	0.53	2,511.9	3.75	5.00	7.7
Cote d'Ivoire	36,070	19,737,800	$591	**29.4**	63.20	55.377	46.34	0.40	2,527.5	24.31	2.60	3.4

Croatia	79,300	4,403,330	$6,338	**1.0**	6.06	76.64	0.09	0.80	2,989.9	5.47	60.12	11.4
Dominican Republic	93,380	9,927,320	$4,049	**3.4**	21.30	73.396	9.88	0.69	2,295.1	6.10	39.53	9.9
Egypt	519,000	81,121,077	$1,976	**6.8**	24.23	73.235	15.43	0.64	3,194.6	13.95	26.74	8.0
El Salvador	44,580	6,192,993	$2,557	**6.6**	19.66	72.196	16.94	0.67	2,589.6	12.71	15.90	8.0
Ethiopia	94,850	82,949,541	$221	**34.6**	60.90	59.274	77.63	0.36	1,979.7	41.87	0.75	2.1
Germany	3,114,000	82,302,465	$25,306	**1.1**	3.51	80.414	—	0.91	3,547	—	82.53	12.2
Ghana	75,660	24,391,823	$359	**14.3**	40.90	64.228	51.84	0.54	2,907	27.27	9.55	6.9
Honduras	35,700	7,600,524	$1,392	**8.6**	19.85	73.126	29.84	0.63	2,623.4	12.40	11.09	7.6
India	4,421,000	1,224,614,327	$787	**43.5**	46.07	65.438	68.72	0.55	2,351.9	17.22	7.50	5.7
Indonesia	1,125,000	239,870,937	$1,144	**19.6**	27.00	69.366	46.12	0.62	2,538.4	16.88	9.90	8.3
Jamaica	24,560	2,741,052	$3,665	**1.9**	14.30	73.127	5.44	0.73	2,851.6	5.87	26.48	11.6
Jordan	36,940	6,187,227	$2,534	**1.9**	15.83	73.403	1.59	0.70	3,015.4	3.33	38.88	11.6
Kazakhstan	216,800	16,026,367	$2,482	**4.9**	23.06	67.017	1.12	0.75	3,490.1	5.26	33.38	12.2
Kenya	71,210	40,512,682	$469	**16.4**	43.61	57.134	67.21	0.51	2,089.3	23.13	25.90	8.5
Lesotho	3,723	2,171,318	$496	**16.6**	53.44	48.196	62.25	0.45	2,476.2	7.76	3.86	9.5
Mauritania	7,115	3,459,773	$609	**15.9**	58.93	58.582	47.69	0.45	2,841.1	16.26	3.00	3.3
Mexico	1,667,000	113,423,047	$6,105	**3.4**	16.77	76.954	5.19	0.77	3,266.3	3.73	31.05	9.8
Morocco	163,500	31,951,412	$1,844	**9.9**	26.49	72.15	14.03	0.58	3,236	15.09	49.00	4.3
Mozambique	24,000	23,390,765	$390	**18.3**	76.85	50.239	81.77	0.32	2,066.6	31.96	4.17	2.8
Pakistan	488,400	173,593,383	$669	**31.3**	61.27	65.437	60.19	0.50	2,292.8	21.62	16.78	4.2

Paraguay	40,640	6,454,548	$1,621	**3.4**	22.24	72.477	13.22	0.67	2,634.4	21.99	19.80	9.2
Peru	302,000	29,076,512	$3,180	**4.5**	21.50	73.99	12.74	0.73	2,457	7.82	34.30	10.3
Philippines	391,100	93,260,798	$1,383	**20.7**	18.75	68.749	41.53	0.64	2,564.9	13.04	25.00	10.4
Senegal	25,150	12,433,728	$562	**14.5**	55.16	59.318	60.36	0.46	2,347.7	17.84	16.00	2.7
South Africa	555,000	50,132,817	$3,746	**8.7**	42.67	52.797	31.33	0.62	2,998.5	2.40	12.33	10.0
Tanzania	67,900	44,841,226	$456	**16.7**	46.50	58.199	87.87	0.47	2,032.4	27.11	11.00	5.9
Thailand	602,200	69,122,234	$2,713	**7.0**	15.90	74.126	4.59	0.68	2,538.6	12.37	21.20	8.9
Tunisia	100,000	10,480,934	$3,165	**3.3**	24.98	74.515	8.06	0.70	3,326.5	8.32	36.56	7.1
Ukraine	329,300	45,448,329	$1,037	**0.9**	8.38	68.494	0.17	0.73	3,223.7	8.29	44.59	12.9
United States	15,080,000	310,383,948	$37,491	**1.3**	6.00	78.531	—	0.91	3,748.4	—	74.25	13.5
Yemen	57,970	24,052,514	$610	**43.1**	53.50	65.493	46.60	0.46	2,067.6	7.70	12.35	1.9