JMP013: Per Capita Income

One-way ANOVA and Kruskal-Wallis

Produced by

Dr. DeWayne Derryberry
Idaho State University, Department of Mathematics

Per Capita Income

One-Way ANOVA and Kruskal-Wallis

Key ideas

ANOVA, Kruskal-Wallis Test, Transformations, Tests for Unequal Variances, Samples Versus Populations, Randomization (Permutation) Tests.

Background

Which countries are the richest? Which are the poorest? Are there wealth disparities between the different regions of the world? A common measure of the wealth of a nation is per capita income. Data on per capita income, in US dollars, is readily available for countries around the world.

Per capita income data for a subset of nations was taken from *The World Factbook*

(cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html).

The Task

Answer the following questions: Do regions differ in overall wealth? Which regions are richest and which are poorest?

The Data Per Capita Income.jmp

The data table includes 2012 estimated per capita incomes for 169 of the 229 countries listed on *The World Factbook*.

Macro Region	Asia, Africa, Europe, Latin America, North America, and Oceania
Region	The subregion
Country	Name of the country
Per Capita Income in US$	Per capita income (total income/total population) in US$

Notes: For some countries, only older estimates were available. The number of countries listed can change within a given year. One of the main issues discussed is whether the data are a sample or an entire population, and what conclusions are possible if the data is a population.

Analysis

We are interested in comparing the average wealth of several regions. The formal analysis method for comparing several groups at once is analysis of variance (ANOVA). In this analysis, the hypotheses are:

Ho: All regions have, except for random variation, the same average per capita income.

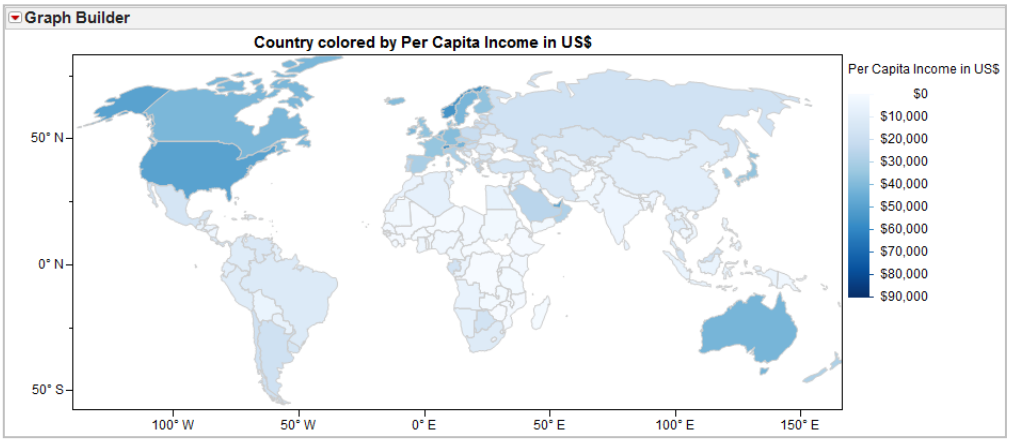
Ha: Some regions have very different average per capita incomes than other regions.

The first step in the analysis, as usual, is to examine assumptions. Ideally, the data are roughly normally distributed, the groups have equal variances, the data for the different regions is independent, and each observation within a region is also independent.

We start by exploring the data. From the geographic map of per capita income by country (Exhibit 1), we can see evidence that there are indeed regional differences.

North America and Europe, for example, appear to have higher average-income levels than Africa and Latin America.

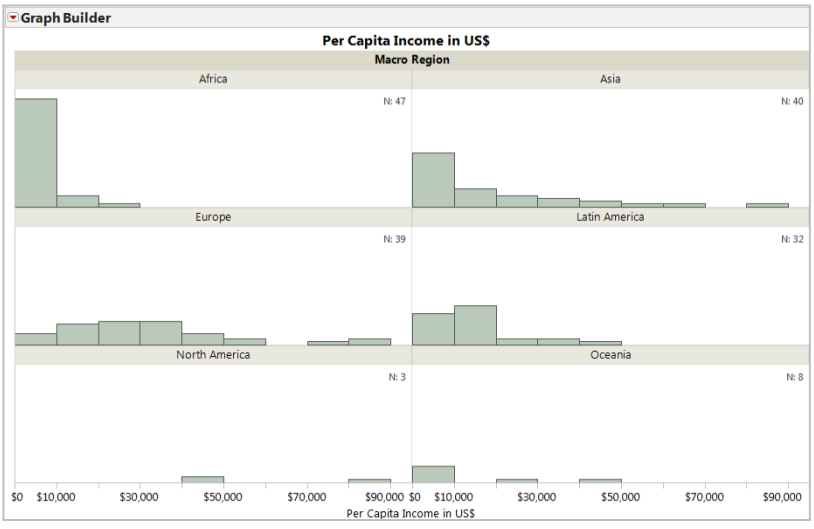
Exhibit 1 Map of Per Capita Income by Country



*(Graph > Graph Builder; Drag **Country** to the Shape zone (at the bottom left corner of graph frame), and drag **Per Capita Income in US$** to the Color zone (to the right of the graph frame). Double-click on the color key to change the color gradient, and click the Done button to close the Control Panel.)*

Now we look at the distributions of per capita income by region (Exhibit 2).

Exhibit 2 **Distribution of Per Capita Income by Macro Region**



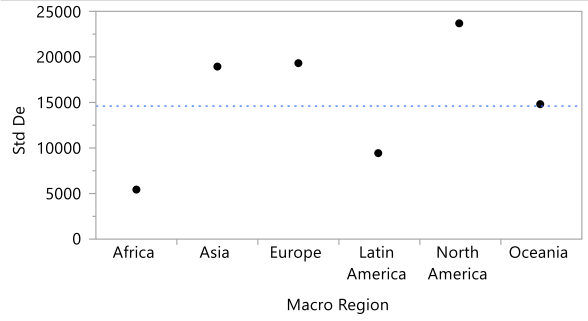
*(Graph > Graph Builder; Drag **Per Capita Income in US$** to the X drop Zone, and drag **Macro Region** to the Wrap zone (top right). Click on the histogram icon above the graph. To display the number of observations, drag the Caption icon onto the graph. Then, under the Caption Box on the bottom left, change the summary statistic from Mean to N.)*

The data displays right skew in most cases. Notice that the sample sizes are so small for North America and Oceania (bottom panels in Exhibit 2) that information about shape is extremely limited.

Next, we consider the equal spread assumption (Exhibit 3).

Exhibit 3 Test for Equal Spread in Per Capita Income

Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
Africa	47	5346.36	3684.47	2963.83
Asia	40	18909.36	14604.50	13007.50
Europe	39	19293.26	14560.55	14269.23
Latin America	32	9344.09	6578.52	6509.38
North Americ	3	23662.84	17933.33	17600.00
Oceania	8	14736.88	11506.25	8637.50

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	3.1162	5	163	0.0103 *
Brown-Forsyth	5.9576	5	163	<.0001 *
Levene	9.6479	5	163	<.0001 *
Bartlett	14.6168	5	.	<.0001 *

Warning: Small sample sizes. Use Caution.

*(Analyze > Fit Y by X; Select **Per Capita Income in US$** as Y, Response and **Macro Region** as X, Factor and click OK. Then, select Unequal Variances from the red triangle.)*

Four tests of equal variance are conducted. Each of the tests is based on the following hypotheses:

Ho: All regions have similar spreads.

Ha: At least one region has a different spread than the others.

All of the tests produce the same result: All p-values are <0.0001, indicating clear evidence of unequal spread. We consider Levene's test the "best" of these, following the advice of *The Statistical Sleuth* (Ramsey and Schafer). In other words, on the rare occasions these tests do not all agree, we will ignore the others and use Levene's test.

So far, we are 0 for 2 in terms of meeting assumptions for conducting ANOVA – the data for each of the groups does not appear to be normal, and the variances are not equal. We have two options in terms of how to proceed:

1. We can consider the logarithmic transformation of the data to improve shape and reduce the variation in spread, and perform ANOVA on the transformed data.
2. We can use a nonparametric test, the Kruskal-Wallis test.

A Couple of Interesting Points

The small sample sizes for North America, and to some extent Oceania, limit the reliability of all the tests for equal variance. How does this affect the analysis? Even when we cannot speak on an absolute scale, we can often speak on a relative scale. If the logarithmic transformation substantially equalizes standard deviations and increases the p-values for the test of equal variance, the analysis using the logarithmic transformation is better than the analysis using the original data, even if some caution must be used when drawing conclusions from tests of equal variance (assuming the logarithmic transformation improves shape, which it does in this case).

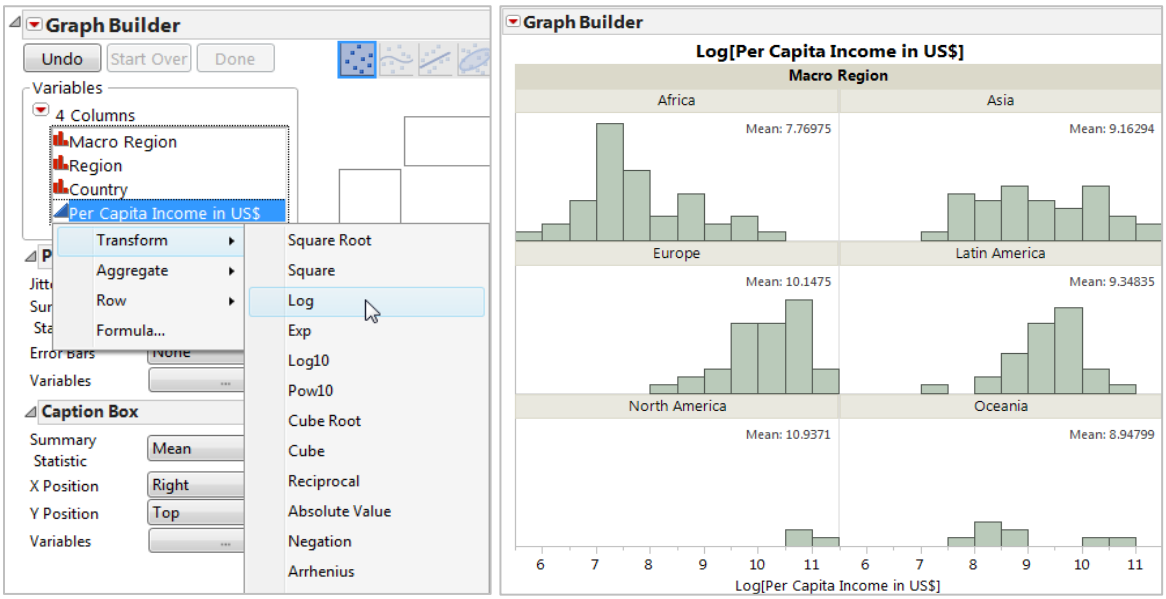
The nonparametric test is also aided by the logarithmic transformation. Although nonparametric tests do not have the assumption of normal data, such tests work best when samples being compared come from populations that differ only in center, having similar shapes and spread. On the other hand nonparametric methods, such as the one we will be using, produce identical results (p-values) for the original data and

for any monotone transformation of the data (i.e., for any transformation preserving the order of the data). This is a very subtle point: The mere fact that a monotone transformation of the data causes the samples to appear much more similar in shape and spread strengthens the results for the nonparametric test, even though actually enlisting that transformation would not change the nonparametric test results!

The Log Transformation

We start by exploring the log-transformed data. The shape of the data for the regions (Exhibit 4) is now much less skewed, and there doesn't appear to be any outliers.

Exhibit 4 Exploring a Logarithmic Transformation

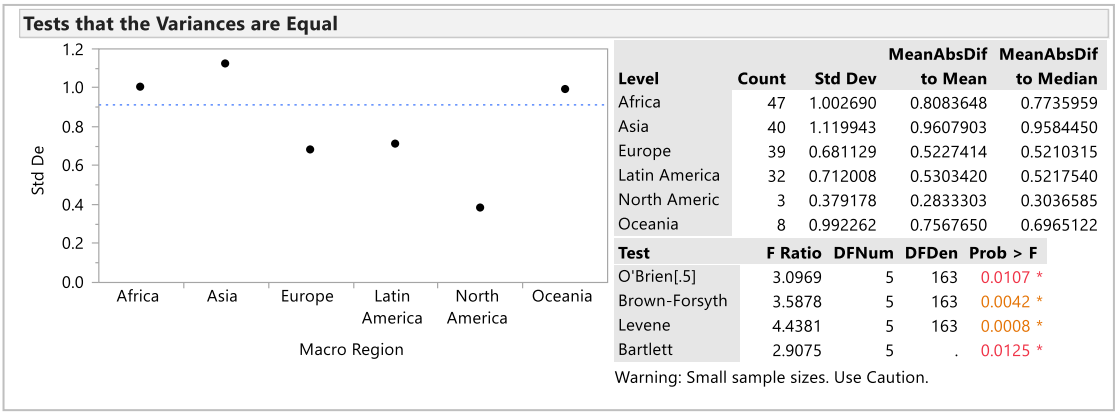


*(Graph > Graph Builder; Right-click on **Per Capita Income in US$** and select Transform > Log. Drag the transformed column to the X drop Zone, and drag **Macro Region** to the Wrap zone (top right corner). Click on the histogram icon above the graph. To save the transformed data to the data table, right-click on the transformed variable, and select Add to Data Table.*

Note that in JMP 10 and earlier versions, the transformed column can only be created using the Transcendental > Log function in the Formula Editor.)

On the logarithmic scale, there is still evidence of unequal spread (Exhibit 5), but the evidence is not quite as pronounced.

Exhibit 5 Test for Equal Spread in Log(Per Capita Income)



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
Africa	47	1.002690	0.8083648	0.7735959
Asia	40	1.119943	0.9607903	0.9584450
Europe	39	0.681129	0.5227414	0.5210315
Latin America	32	0.712008	0.5303420	0.5217540
North Americ	3	0.379178	0.2833303	0.3036585
Oceania	8	0.992262	0.7567650	0.6965122

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	3.0969	5	163	0.0107 *
Brown-Forsyth	3.5878	5	163	0.0042 *
Levene	4.4381	5	163	0.0008 *
Bartlett	2.9075	5	.	0.0125 *

Warning: Small sample sizes. Use Caution.

*(Analyze > Fit Y by X; Select **Log(Per Capita Income in US$)** as Y, Response and **Macro Region** as X, Factor and click OK. Then, select Unequal Variances from the red triangle.)*

The most problematic region is North America. Otherwise the standard deviations for the different regions are fairly close. Generally, standard deviations should be compared using ratios (rather than differences). The ratio of the largest to smallest standard deviation, on the log scale, is 1.1199/0.3792 = 2.93. The ratios on the original scale were generally larger, with the largest ratio (North America vs. Africa) of 4.4. A rule of thumb is that ratios of less than 2.0 are considered quite small (see *The Basic Practice of Statistics*, 5th Edition. 2010. David S Moore, W.H. Freeman. Page 645).

Overall, the logarithmic scale is much better than the original scale for analyzing this data with ANOVA (we'll give it a B versus a C-). However, we will use both the logarithmic transformation and a nonparametric procedure for analysis of this data.

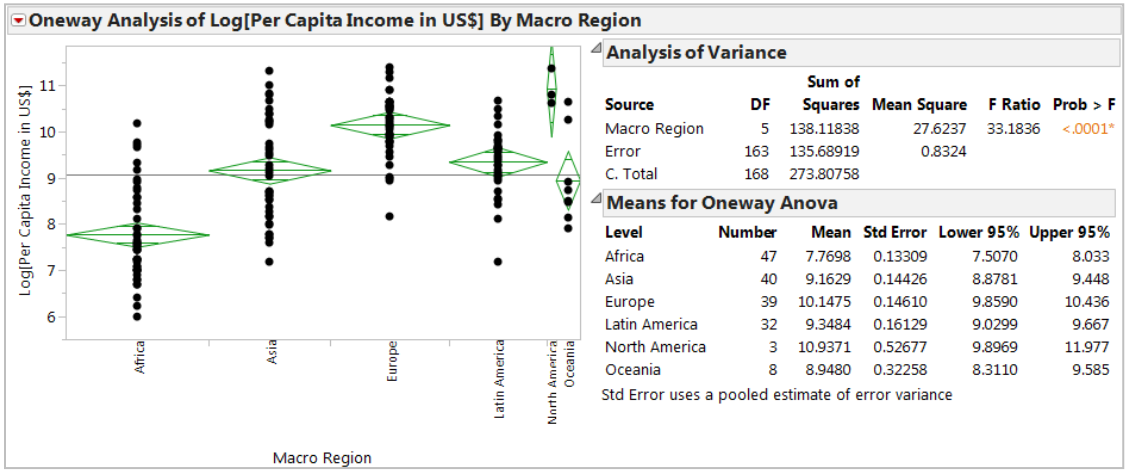
ANOVA on the Transformed Data

We can see that the North American region has the highest average per capita income, Europe is second highest, while Africa lags behind the rest of the world (Exhibit 6). The averages are given in the Means for One-Way ANOVA table (bottom right), and are displayed graphically as means diamonds – the center line of each diamond is the group average.

But, are these differences due to chance variation, or is more going on?

The test statistic, the F Ratio (top right in Exhibit 6) is 33.1836 with a p-value of less than 0.0001 (under Prob > F). This is convincing evidence that average per capita incomes vary from region to region. At least some of the differences we are seeing are likely real.

Exhibit 6 Analysis for the Regions on the Log Scale



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Macro Region	5	138.11838	27.6237	33.1836	<.0001*
Error	163	135.68919	0.8324		
C. Total	168	273.80758			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Africa	47	7.7698	0.13309	7.5070	8.033
Asia	40	9.1629	0.14426	8.8781	9.448
Europe	39	10.1475	0.14610	9.8590	10.436
Latin America	32	9.3484	0.16129	9.0299	9.667
North America	3	10.9371	0.52677	9.8969	11.977
Oceania	8	8.9480	0.32258	8.3110	9.585

Std Error uses a pooled estimate of error variance

*(Analyze > Fit Y by X; Select **Log(Per Capita Income in US$)** as Y, Response and **Macro Region** as X, Factor and click OK. Then, select Means/ANOVA from the red triangle.)*

Since we saw evidence of unequal variances, we can use the Welch ANOVA test (Exhibit 7) for further evidence that there are significant differences in per capita income.

Exhibit 7 The Welch ANOVA Test

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
39.2533	5	18.831	<.0001*

(Displays by default when Unequal Variances is selected – see Exhibit 5.)

A Nonparametric Test

Because the assumptions for the parametric test (ANOVA) are well met using the log-transformed data, the nonparametric test will produce very similar results.

Since we have more than two groups, the Kruskal-Wallis test will be performed. Note that this test, which involves transforming data to ranks, will produce identical results whether used on the original or transformed data.

The nonparametric test (Exhibit 8) also provides convincing evidence that different regions have different median incomes, with a p-value < 0.0001 (under Prob>ChiSq).

Exhibit 8 Kruskal-Wallis Test (Wilcoxon Test for More than Two Groups)

Wilcoxon / Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
Africa	47	1768.50	3995.00	37.628	-7.811
Asia	40	3501.00	3400.00	87.525	0.372
Europe	39	5006.00	3315.00	128.359	6.308
Latin America	32	2986.50	2720.00	93.328	1.067
North America	3	479.000	255.000	159.667	2.661
Oceania	8	624.000	680.000	78.000	-0.411

1-way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
82.8698	5	<.0001*

(In the One-way analysis window, select Nonparametric > Wilcoxon Test from the top red triangle.)

However, the presentation of the nonparametric test provides some additional insight into the analysis.

The nonparametric test produces a z-score for each group relative to the overall average (the last column in table in Exhibit 8). For example, Latin America has a z-score of 1.067, meaning that Latin America is, on average, above the average median per capita income.

If all regions have similar per capita incomes, the z-scores will follow the well-known rule: They will be between -1 and 1 about 68% of the time, between -2 and 2 about 95% of the time, and between -3 and 3 almost always (about 99.73% of the time). The large negative z-score for Africa (-7.811) and the large positive z-score for Europe (6.308) are immediate evidence of real underlying differences. If all regions do in fact have similar incomes, these z-scores are hard to explain.

It is interesting that North America, which has the highest median per capita income, does not produce the largest z-score. This is due to small sample size. Although North America has the greatest median income, with a sample size of only 3, it DOES NOT provide the strongest overall evidence of income disparities. The z-score is based on how far above the median a region is and how large a sample size is involved. Europe, which has a high median per capita income and a sample size of 39, provides the strongest evidence that at least one region is well above the average median income.

Scope of Inference: Permutation Analysis

It is often just assumed that data comes from random samples. However, we might question whether this is the case in this example. Since our inference is to world regions, we might consider whether we have random samples of countries from each region of the world. A quick Google search will verify there are about 230 countries in the world (this number often varies within a year). Our data set is made up of 169 countries. This is not a sample of countries around the world, it is a majority of the countries in the world (roughly 73%). This is not inference from a sample to a population, we have (almost) all of the population.

So, is any formal statistical analysis (i.e., hypothesis testing) necessary? It might seem that the data already says it all, and that inferential methods are not required to address our question. After all, nearly the entire population is there - a difference is a difference. However, the question whether all regions have the same average per capita income is more subtle than that.

Clearly, even if there are no intrinsic differences from region to region, observed average per capita incomes will differ, just due to random variation. There are 169 discrete per capita incomes, so any random assignment of incomes to countries will produce some differences in average income from region to region.

This is the basis for an analysis first proposed by R. A. Fisher. Consider the hypotheses:

> Ho: Any differences in average income, from region to region, can easily be explained by chance variation.
>
> Ha: The differences in average income are too great to be easily explained by chance variation.

So what it this “chance variation”?

It’s just the random assignment of per capita incomes to countries. Under the null hypothesis, every per capita income is equally likely to be assigned to any country. Considering all possible assignments of incomes to countries, we can determine how many produce apparent disparities as great as, or greater than, those found in the observed arrangement.

In this case, a true permutation test would involve enumerating all possible assignments of the incomes to countries and computing some statistic that measures disparities in income from region to region (such as the F statistic). After this procedure is completed, we can compute a frequency-based number:

> p = (the cases where the computed statistic is as great as or greater than the statistic for the observed arrangement)/(the number of possible arrangements).

This is the chance (probability) we could have gotten such an unusual and disparate arrangement of incomes to regions just by chance, if the null hypothesis is true.

This procedure can be quite involved, but it is not necessary. When the assumptions of the standard ANOVA are met, the p-value from the ANOVA (or a nonparametric alternative, if needed) is a good approximation to this value “p.” (This is discussed in *The Statistical Sleuth* by Ramsey and Schafer. For those with a further interest, entire books have been written on permutation [or randomization] tests).

Summary

Statistical Insights

What we should conclude from this analysis is that the differences from region to region in per capita income cannot be explained by random variation. North America and Europe are clearly doing very well, and Africa is clearly doing poorly (no surprises here).

Which analysis is best, the logarithmic transformation or the nonparametric test? This would be a reasonable question to ask a student taking their oral preliminary exam in a graduate program in statistics!

When the assumptions for an F-test are met, this would be the better test (particularly if no transformations were required). The nonparametric test, which has fewer assumptions, will generally be less powerful. However, the logarithmic transformation did not fully address the problem of unequal spread. Would a different transformation perform better?

In this case, both approaches are a reasonable way to analyze the data, and both lead to the same conclusion: The disparities between regions are quite stark. Both analyses, when taken together, will be stronger than either analysis taken alone.

The nonparametric analysis, as we saw, has an additional feature in that z-scores are computed for each region as deviations from the overall median. These z-scores account for both raw differences in the medians and the role of sample size to produce a number that indicates not just which regions are above or below average, but also how unusual that difference seems to be.

Implications

We performed what appeared to be a standard statistical analysis, under the assumption that our data were from a sample. However, our data actually represented nearly the entire population. The importance of this distinction cannot be underestimated.

In a number of circumstances, analysis involves either populations or highly non-random samples. In these cases, the use of permutation (or randomization) tests allow for the computation of a meaningful p-value. Fortunately, when the assumptions of the standard ANOVA are met, the p-value from ANOVA (or a nonparametric alternative, if needed) is a good approximation for a p-value from a randomization test - standard methods can be applied and results can be interpreted.

Note that, in a case like this, we also need to be aware of potential non-random missing data problems. Out of the 230 countries in the world, which countries have not been included? Have smaller or newer countries been omitted? Have some countries withheld information? Non-random missing data can lead to biased inferences.

JMP® Features and Hints

In this case study we used the Graph Builder to create a geographic map of the per capita incomes for each of the countries in the data set. We then used the Graph Builder to explore the shapes of the distributions for incomes for each of the regions, and used the Transform feature to explore the shape of log-transformed data and to create a new transformed variable in the data table.

We used Fit Y by X to test for unequal variances for both the original and the transformed data, to conduct a one-way ANOVA on the transformed data, to perform a Welch's ANOVA test, and to perform a Kruskal-Wallis nonparametric test.

Exercises

1. Cost of Living

Consider the cost of living for the largest cities in the United States (with populations greater than 50,000). The Cost of Living Index measures relative price levels for housing, utilities, transportation, grocery items, health care, and miscellaneous goods and services. A cost of living score of 100 is the national average. The further a score falls below 100, the lower the cost of living.

Data are based on the Council for Community and Economic Research's calculations of living expenses. Population and median household income data are from the US Census Bureau.

The Data Cost of Living.jmp

Metro Area	The metropolitan area
State	50 states and the District of Columbia
Region	N, W, E, and S
Population	The population
Cost of Living	The cost of living index
Med HH Income	The median household income

Perform an analysis of this data. (Note: Data for the New York metro area has been hidden and excluded.)

a. Display the data and assess all relevant assumptions.
b. Test for differences from region to region. State a conclusion.
c. Perform a nonparametric test. State a conclusion.
d. Which analysis would you report, and why?
e. Which region is doing best? Which is doing the worst?
f. Unexclude and unhide data for New York, and redo your analysis. What impact does this have on your test results and your conclusions?

2. Census at School

Consider a sample of data from the Census at School. This is an international project for students in grades 4-12 in which students complete an online survey, analyze results for their class, and compare their class with random samples of students from across the globe. A random sample of 500 12th grade students from the US, collected in 2013, was downloaded from: amstat.org/censusatschool/RandomSampleExport.cfm.

The Data Census at School 12th.jmp

The data set includes the student's state, region, and responses on 56 survey questions (many of these have been hidden and excluded). We are interested in testing for regional differences in the average responses for the following questions:

Doing_Homework_Hours
Watching_TV_Hours

a. For each response, perform an analysis following steps a-d above, and summarize your conclusions for each response.
b. This data is the result of survey responses. Are there any potential issues with drawing conclusions using this data? Discuss.