Section VI: Statistical Analysis

SPSS Analysis

The data gathered through the household survey was compiled into an Excel database. The data was then imported into SPSS (Statistical Package for the Social Sciences) to conduct data examination and cleansing processes and to maintain a high level of data integrity throughout the analysis. Each variable was analyzed for consistency and reasonableness and, where necessary, values were imputed for records that were either missing or well beyond the range of normal variation. The analysis processes followed for each variable are outlined below. By conducting the household analysis in SPSS, the study team was also able to merge community information and several identifiers that allowed aggregation of the household responses into various groupings for reporting. These groupings include:

- The 19 community groupings from the 1985 differential study conducted by McDowell Group.
- The 19 community groupings defined in the 1994 update.
- The 40 Alaska House districts.
- The 20 Alaska Senate districts.
- An 18-block grouping developed by the McDowell Group based on geographic and commuting factors in Alaska.

Data Examination and Cleansing

Two data examination processes were conducted in SPSS to ensure that the data input into the Excel database from the household survey forms were properly recorded. A simple random sample equaling approximately 1.0 percent of all surveys was drawn and the values recorded for each variable for each survey record were compared against the paper survey forms received from the contractor overseeing the household survey.[1] No errors were found in any of the randomly drawn survey records.

As a second check on the accuracy of the data, descriptive statistics for each numeric variable were calculated and all records with extreme values were compared to the survey forms. In total, approximately 40 survey records were checked against the physical survey forms and values were corrected for three records.

[1] It was assumed that the occurrence of an error in the recording of the survey data would be a rare event. Developing a statistically valid sample to test for the occurrence of a rare event is generally very expensive due to the large sample size required to develop high relative precision. In essence, nearly every record would require re-examination. The selection of approximately 1.0 percent of records for thorough comparison against the physical survey forms was intended to confirm that no *systematic* errors were made in compiling the data into Excel.

Examination of Individual Variables and Data Imputation

Income

Respondents to the household survey were asked for the household's total pre-tax income from all sources in 2007. Most households provided this information. Those who refused to answer the direct income question were asked the category that best describes their household's income. The midpoint of each of the associated category was used to impute income for the household. Household income was set to *missing* for those households that refused to provide income information.

Demographic Information

Four variables were created based on the demographic information provided by the respondents:

- Average age of adults living in the household
- Average age of children living in the household
- The count of adults living in the household
- The count of children living in the household.

Housing

On average, housing and related costs are the largest component of household expenses. Because of its relative importance in the household budget, particular attention was paid to housing in both the survey instrument and in the examination of responses to housing questions. This is summarized in the following steps:

1. Survey responses were segmented based on whether or not the household reported owning or renting their home.
2. For those households that reported owning their home, monthly mortgage payments reported to be more than four standard deviations greater than the median monthly mortgage amount were truncated at this amount.[2]
3. For those households that reported owning their home, annual property tax payments reported to be more than four standard deviations greater than the median annual property tax payments were truncated at this amount.[3]
4. For those households that reported owning their home, but did not report property tax or property insurance payment amounts, values were imputed based on the statistical relationship between monthly house payments and annual property tax and insurance payments.[4]

[2] The median, rather than the mean of the distribution of mortgage payment amounts was chosen because it is a better representation of the central tendency of the distribution. Under the assumption of a symmetric distribution, which mortgage payments are not, four standard deviations from the median would encompass more than 99 percent of all mortgage payments. By truncating extreme values at this level, we reduce the degree of influence that the very small portion of the population has on the mean and other parametric statistical estimates.
[3] Ibid.
[4] Statistical relationship based on those households that reported both mortgage payments and property tax or property insurance amounts.

5. For those households that described their home as a condo, but did not report monthly condo fee, values were imputed based on the statistical relationship between monthly mortgage payments and monthly condo fees.[5]

6. For those households that described their home as a mobile home, but did not report monthly space rental cost, values were imputed based on the statistical relationship between monthly mortgage payments and monthly rental cost.[6]

7. Housing shelter cost was computed for each record.

    a. **Homeowner Shelter Cost** = (monthly mortgage payment) + (annual property tax/12) + (annual property insurance/12) + (monthly condo fees) + (monthly mobile home space rent)

    b. **Renter Shelter Cost** = monthly rental cost

8. For all households, home sizes reported to be more than four standard deviations greater than the median home size were truncated at this point.[7] Calculations were done separately for owners and renters.

9. For all households missing home size, values were imputed as the average empirical home size for each of the geographic regions identified in 1985 and 2008. Separate calculations were made for owners and renters.

10. For all households, missing values for home energy consumption costs were imputed based on statistical models of energy costs regressed on home size and indicators for alternative fuel types (e.g. natural gas, electric, wood, etc.). Separate models were estimated for each fuel type.

11. Home energy consumption costs were examined for extreme outliers.

12. Total home energy consumption was calculated as the sum of energy costs across all fuel types consumed by the household.

13. Total home utilities costs were calculated for each household. Any household with total utility costs greater than four standard deviations from the mean were truncated at that value.

14. Total housing costs were computed for each household.

    a. **Homeowner Total Housing Cost** = Homeowner shelter cost + total heating cost + total utility cost

    b. **Renter Total Housing Cost** = Renter shelter cost + total heating cost + total utility cost

15. Per foot shelter costs and total housing costs were calculated for each household.

    a. **Per Foot Homeowner Shelter Cost** = Homeowner shelter cost / home size

    b. **Per Foot Renter Shelter Cost** = Renter shelter cost / home size

[5] Ibid.
[6] Ibid.
[7] Ibid.

c. **Per Foot Homeowner Total Housing Cost** = Homeowner total housing cost / home size

d. **Per Foot Renter Total Housing Cost** = Renter total housing cost / home size

Vehicles

1. For those households that reported owning a car, monthly car payments reported to be more than four standard deviations greater than the median monthly car payment amount were truncated at this amount.

2. Monthly payment amount for other vehicle types (e.g. snow machine, boat, etc.) were examined for extreme outliers; no values were truncated for these variables.

3. Any record monthly vehicle fuel cost (for all vehicles) reported to be more than four standard deviations greater than the median monthly fuel cost amount was truncated at this amount.

4. Any record with monthly vehicle maintenance cost (for all vehicles) reported to be more than four standard deviations greater than the median monthly vehicle maintenance cost amount was truncated at this amount.

5. Any record with monthly vehicle insurance cost (for all vehicles) reported to be more than four standard deviations greater than the median monthly vehicle insurance cost amount was truncated at this amount.

In-State and Out-of-State Plane Travel

1. Any record with in-state air travel cost reported to be more than four standard deviations greater than the median in-state air travel cost amount was truncated at this amount.

2. Any record with out-of-state air travel cost reported to be more than four standard deviations greater than the median out-of-state air travel cost amount was truncated at this amount.

Food

1. Any record with weekly spending on groceries reported to be more than four standard deviations greater than the median reported spending on groceries was truncated at this amount.

2. Any record with weekly [food item][8] spending amount reported to be more than four standard deviations greater than the median [food item] spending amount was truncated at this amount.

3. Food away from home was analyzed for extreme values. None were found.

Clothing

1. Average monthly spending on clothing was analyzed for extreme values. None were found.

[8] Food Item categories include: Meats, poultry, and fish; Cereals and bread; Dairy products; Fruits and vegetables; Soups, frozen meals, and snacks; Nonalcoholic beverages other than milk.

2. Local monthly spending on clothing was computed as average monthly spending on clothes multiplied by the reported percent purchased in local area.

3. Non-local monthly spending on clothing was computed as average monthly spending on clothes multiplied by the reported percent purchased non-locally.

Medical

1. Any record with monthly spending on medical insurance reported to be more than four standard deviations greater than the median monthly spending amount was truncated at this amount.

2. Any record with monthly spending on medical expenses reported to be more than four standard deviations greater than the median monthly spending amount was truncated at this amount.

Household Budget

Survey respondents were asked to estimate the percent of their total household income that was spent on the following four categories:

- Housing and utilities
- Groceries and dining out
- Transportation and travel
- All other expenses, including clothing, recreation, entertainment, medical, education, taxes, and savings.

When necessary, the survey administrator assisted the respondent to assure that the sum of the four percentages equaled 100 percent.

Federal Taxes

To derive disposable income (total income – taxes) for each household, it was necessary to approximate the household's federal tax obligation. This was done through the following five steps, based on the 1040 tax form.

Step 1: Filing status: *married* or *head of household*.

- If the household consisted of only one adult, then it was assumed that the appropriate tax status is *head of household*.
- If the household consisted of two or more adults, then it was assumed that the appropriate tax status is *married filing jointly*.

Step 2: Compute exemptions and deductions.

- Head of household: $11,500 + total child count * $3,500.
- Married filing jointly: $17,900 + (total child count + total adult count – 2) * $3,500.

Step 3: Compute adjusted gross income: Total Income – exemptions and deductions.

Step 4: Compute federal taxes.

a. Head of Household:

- If (0 < adj. gross inc. <= $11,450) fed tax = 0 + 0.10 * (adj. gross inc. – 0).

- If ($11,450 < adj. gross inc. <= $43,650) fed tax = 0 + 0.15 * (adj. gross inc. – $11,450).
- If ($43,650 < adj. gross inc. <= $112,650) fed tax = 0 + 0.25 * (adj. gross inc. – $43,650).
- If ($112,650 < adj. gross inc. <= $182,400) fed tax = 0 + 0.28 * (adj. gross inc. – $112,650).
- If ($182,400 < adj. gross inc. <= $357,700) fed tax = 0 + 0.33 * (adj. gross inc. – $182,400).
- If ($357,700 < adj. gross inc.) fed tax = 0 + 0.35 * (adj. gross inc. – $357,700).

b. Married filing jointly:

- If (0 < adj. gross inc. <= $16,050) fed tax = 0 + 0.10 * (adj. gross inc. – 0).
- If ($16,050 < adj. gross inc. <= $65,100) fed tax = 0 + 0.15 * (adj. gross inc. – $16,050).
- If ($65,100 < adj. gross inc. <= $131,450) fed tax = 0 + 0.25 * (adj. gross inc. – $65,100).
- If ($131,450 < adj. gross inc. <= $200,300) fed tax = 0 + 0.28 * (adj. gross inc. – $131,450).
- If ($200,300 < adj. gross inc. <= $357,700) fed tax = 0 + 0.33 * (adj. gross inc. – $200,300).
- If ($357,700 < adj. gross inc.) fed tax = 0 + 0.35 * (adj. gross inc. – $357,700).

Step 5: Compute after tax income = household income – federal tax.

Calculation of Spending as a Percent of Income

For each component of household spending, the study team calculated spending as a percent of after tax household income.[9]

Alternative Aggregations

Alternative aggregations of the household data were developed based on (1) community, (2) the 1985 GDP study, (3) the 1994 GDP study, (4) the 2009 Alaska House of Representative districts, (5) the 2009 Alaska Senate districts, and (6) an aggregation developed by the study team.

Development of Approximate Standard Errors

Typically, the estimation of cost of living and other indices based on survey data does not include the development of standard errors. The reason for this is that survey data are generally "complex" in that the surveyed households often do not perfectly represent the population of interest and, therefore, parameter estimates must be developed inclusive of some sort of weighting scheme. Methods that allow for the calculation of (near) exact standard errors do exist, but they are beyond the ability of typical statistical software packages (e.g. SPPS, SAS, Stata) and developing individual standard errors for each parameter of interest requires extensive analyst and computer time. This would certainly be the case for the Alaska GDS, which includes the surveying of more than 2,500 Alaska households across 74 communities regarding their income and spending patterns. It would be prohibitively expensive in terms of time and budget to attempt to calculate exact standard error of each of the many cost of living parameter estimates derived from the household and retail surveys.

[9] Note 1: Any household with spending less than 30 percent or greater than 120 percent of after tax income was set to *missing* for the purpose calculating spending as a percent of household income.
Note 2: If spending at the major household sector (e.g. housing, transportation, etc.) was missing for any record, the value was imputed based on the respondents stated spending on the particular sector as a percent of household income (if available).

Generalized Variance Function Method

Fortunately, there is an alternative to the development of exact standard errors. The generalized variance function (GVF) method is a procedure for estimating *approximate* standard errors for estimated means, proportions, ratios, indexes, and the difference between sample estimates. GVF estimates can be developed relatively quickly using standard statistical software (e.g. SPSS, SAS, Stata). Although the specification of the GVF function varies based the parameter of interest, available data, and the needs of the analyst, the following characteristics are typical parts of GVF procedure:

1. Reliance on the *central limit theorem*, which states that the means from sufficiently large samples repeatedly drawn from any distribution will be normally distributed.

2. Monte Carlo-styled resampling of the survey data to develop an empirical distribution for the particular parameter.

3. The estimation of a statistical function that relates the parameter of interest to its variance (or standard deviation).

4. The construction of a formula based on the parameter estimates in Step 3 that allows for the calculation of the approximate standard error (and by extension, confidence intervals).

Development of Generalized Variance Function and the "*A*" and "*B*" Parameters

An SPSS routine was written to perform the resampling routines, obtain the parameter estimates of interest, and estimate the statistical models. The generalized steps of the routine are as follows:

Step 1: Draw 300 samples of approximately 100 cases per sample. Each sample is drawn using a Bernouli random variable, with each of the approximately 2,500 records having an equal probability (1 chance in 25) of being selected in each of the 100 samples.

Step 2: Compute the mean and variance of each continuous variable for each of the 100 samples.

Step 3: (Natural) log-transform each of the mean and variance estimates.

Step 4: Estimate the following statistical regression model:

$$\ln\left(\hat{\sigma}_x^2\right) = \beta_o + \beta_1 \ln\left(\bar{x}\right)$$

Where:

$\ln\left(\hat{\sigma}_x^2\right)$ is the vector of standard deviation estimates for the variable of interest "x" transformed by the natural log function.

$\ln\left(\bar{x}\right)$ is the vector of mean estimates for the variable of interest "x" transformed by the natural log function.

Step 5: Save the Y-intercept and slope coefficient from each statistical model. These parameter estimates will be used as the "*A*" and "*B*" parameters for calculating the approximate standard error for each of the variables of interest.

Calculating an Approximate Standard Error using the *A* and *B* Coefficients

The *A* and *B* coefficients estimated in the SPSS routine are simply the y-intercept and slope coefficients from the simple statistical models relating the mean value of each parameter estimate to its standard deviation. Estimating the standard error and the lower and upper confidence bounds of any of the parameter estimates of interest is a straightforward 3-step process.

Step 1: Estimate the approximate variance of the parameter of interest:

$$\tilde{\sigma}_x^2 = e^{A*\ln(\bar{x})+B}$$

Where:

$\tilde{\sigma}_x^2$ is the approximate variance of the parameter of interest.

$e^{A*\ln(\bar{x})+B}$ is the A and B function raised to the exponential function e.[10]

Step 2: Compute the approximate standard error from the variance:

$$se_x = \sqrt{\tilde{\sigma}_x^2} \Big/ \sqrt{n}$$

Step 3: Compute the lower and upper 95 percent confidence bounds.[11]

Lower Bound = $\bar{x} - 2 * se_x$

Upper Bound = $\bar{x} + 2 * se_x$

The *A* & *B* Table, Sample Sizes, and an Example Calculation

The *A* and *B* coefficients necessary to calculate approximate standard errors and confidence intervals are shown in Table VI-1. The sample sizes for each of the regional aggregations and individual communities, which are also necessary for calculating an approximate standard error, are shown in Tables VI-2, VI-3, VI-4, and VI-5.

Presented below are two examples of how to use the *A* and *B* coefficients to calculate the approximate standard error and confidence intervals for a parameter of interest. Only one example is actually necessary to demonstrate the process; however, by presenting two examples, the impact of sample size on the size of the approximate standard errors is shown and, by extension, the precision of the confidence interval.

Juneau Cost of Living Differential

Based on the previously described analysis, the cost of living differential for Juneau is 1.11. To calculate the standard error on this estimate, the study team utilized the *A* coefficient of 2.236 and the *B* coefficient of -3.042 from row 30 of Table VI-1 and the sample size of 300 from row 7 of Table VI-2.

[10] Note the parameter of interest is log-transformed within the exponential function.

[11] Based on the assumption that two standard deviations on either side of the sample parameter estimate constitutes a 95 percent confidence interval. In fact, for very small sample sizes, the number of standard deviations increases.

Step 1: Estimate the approximate variance of the parameter of interest:

$$\tilde{\sigma}_x^2 = e^{A*\ln(\bar{x})+B} = \tilde{\sigma}_{Juneau}^2 = e^{2.236*\ln(1.11)-3.042} \approx 0.0603$$

Step 2: Compute the approximate standard error from the variance:

$$se_x = \sqrt{\tilde{\sigma}_x^2} \Big/ \sqrt{n} = se_{Juneau} = \sqrt{0.0603} \Big/ \sqrt{300} \approx 0.014$$

Step 3: Compute the lower and upper confidence bounds[12]

$$\text{Lower Bound} = \bar{x} - 2 * se_x = 1.11 - 2 * 0.014 \approx 1.082$$

$$\text{Upper Bound} = \bar{x} + 2 * se_x = 1.11 + 2 * 0.014 \approx 1.138$$

Aleutians Cost of Living Differential

Based on our analysis, the cost of living differential for the Aleutians is 1.50. To calculate the standard error on this estimate, the study team utilized the *A* coefficient of 2.236 and the *B* coefficient of -3.042 from row 30 of Table VI-1, but the sample size for the Aleutians is only 77 (from row 17 of Table VI-2).

Step 1: Estimate the approximate variance of the parameter of interest:

$$\tilde{\sigma}_x^2 = e^{A*\ln(\bar{x})+B} = \tilde{\sigma}_{Aleutians}^2 = e^{2.236*\ln(1.50)-3.042} \approx 0.1182$$

Step 2: Compute the approximate standard error from the variance:

$$se_x = \sqrt{\tilde{\sigma}_x^2} \Big/ \sqrt{n} = se_{Aleutians} = \sqrt{0.1182} \Big/ \sqrt{77} \approx 0.0392$$

Step 3: Compute the lower and upper 95% confidence bounds[13]

$$\text{Lower Bound} = \bar{x} - 2 * se_x = 1.50 - 2 * 0.0392 \approx 1.42$$

$$\text{Upper Bound} = \bar{x} + 2 * se_x = 1.50 + 2 * 0.0392 \approx 1.58$$

As the two examples demonstrate, not only can the cost of living differentials (or any other parameter) differ greatly across the state, so can the relative variance of the estimated distributions (i.e., all else being equal, there is greater variation associated with larger parameter values) and as sample size goes up, the standard error of the estimated parameter goes down. Because of this, for Juneau, with its low cost of living differential and large sample size (relative to the Aleutians), the approximate standard error is relatively small and the precision of the estimate is relatively high. The approximate 95 percent confidence interval for Juneau extends from 1.08 up to 1.14. Comparatively, the approximate 95 percent confidence interval for the Aleutians is relatively wide, extending from 1.42 up to 1.58. This is due to the larger cost of living differential and smaller sample size.

[12] Based on the assumption that two standard deviations on either side of the sample parameter estimate constitutes a 95 percent confidence interval.

[13] Based on the assumption that two standard deviations on either side of the sample parameter estimate constitutes a 95 percent confidence interval.

Table VI-1: *A* and *B* Coefficients for Computing Approximate Standard Errors and Confidence Intervals for Results from the 2008 Alaska Differential Study

Row	Parameter of Interest	*A*	*B*
1	Shelter Cost as a Percent of Income—Owner	0.897	-2.462
2	Shelter Cost as a Percent of Income—Renter	1.069	-2.986
3	All Housing Costs as a Percent of Income—Owner	2.169	-1.074
4	All Housing Costs as a Percent of Income—Renter	1.069	-2.718
5	Vehicle Fuel Cost as a Percent of Income	2.600	1.517
6	Vehicle Maintenance Cost as a Percent of Income	2.190	1.338
7	Vehicle Insurance Cost as a Percent of Income	2.303	0.736
8	Automobile Payment as a Percent of Income	1.220	1.504
9	All Other Vehicle Payments as a Percent of Income	1.593	1.713
10	Spending on Food as a Percent of Income	2.398	-0.309
11	Spending on Groceries as a Percent of Income	2.536	0.131
12	Spending on Food Away from Home as a Percent of Income	2.438	1.739
13	Spending on Meat as a Percent of Income	2.547	1.607
14	Spending on Cereals and Bread as a Percent of Income	3.435	5.488
15	Spending on Dairy as a Percent of Income	3.987	7.515
16	Spending on Fruit as a Percent of Income	3.112	3.631
17	Spending on Soup as a Percent of Income	2.210	0.938
18	Spending on Beverage as a Percent of Income	2.512	2.634
19	Spending on Clothing as a Percent of Income	2.275	1.230
20	Local Spending on Clothes as a Percent of Income	2.117	1.574
21	Non-local Spending on Clothing as a Percent of Income	2.041	0.909
22	All Medical Spending as a Percent of Income	2.061	0.636
23	Spending on Medical Insurance as a Percent of Income	1.851	0.361
24	Spending on Medical Expenses as a Percent of Income	2.168	1.692
25	Spending on Travel as a Percent of Income	1.570	-1.186
26	Spending on In-State Travel as a Percent of Income	1.523	-0.924
27	Spending on Out-of-State Travel as a Percent of Income	1.556	-1.106
28	Household income	2.565	-6.886
29	After Tax Income	1.989	-0.657
30	**Cost of Living Differential Relative to Anchorage**	**2.236**	**-3.042**

Table VI-2: Sample Sizes for Regional Blocks

Sample Block #	Regional Blocks	Sample Size
1	Anchorage	300
2	Fairbanks	300
3	Parks/Elliott/Steese Highways	65
4	Glennallen Region	50
5	Delta Junction/Tok Region	76
6	Roadless Interior	51
7	Juneau	300
8	Ketchikan/Sitka	200
9	Southeast Mid-Size Communities	105
10	Southeast Small Communities	51
11	Mat-Su	187
12	Kenai Peninsula	200
13	Prince William Sound	100
14	Kodiak	104
15	Arctic Region	153
16	Bethel/Dillingham	151
17	Aleutian Region	77
18	Southwest Small Communities	77

Table VI-3: Sample Sizes for 1985 GDS Groupings

District #	1985 GDS Groupings	Sample Size
1	Ketchikan/Prince of Wales	153
2	Petersburg/Wrangell	49
3	Sitka	80
4	Juneau	300
5	Icy Strait/Lynn Canal	74
6	Cordova/Valdez	150
7	Palmer/Wasilla	216
8	Anchorage	300
9	Seward	24
10	Kenai/Cook Inlet	176
11	Kodiak	104
12	Aleutian Islands	79
13	Bristol Bay	56
14	Bethel	111
15	Yukon/Kuskokwim	78
16	Fairbanks/Fort Yukon	398
17	Barrow/Kotzebue	100
18	Nome	70
19	Wade Hampton	29

Table VI-4: Sample Sizes for 1994 GDS Groupings

District #	1994 GDS Groupings	Sample Size
1	Ketchikan/Prince of Wales	153
2	Wrangell/Petersburg	49
3	Sitka	80
4	Juneau	300
5	Icy Strait/Lynn Canal	74
6A	Cordova/Valdez (excluding Valdez Duty Station)	90
6B	Cordova/Valdez (Valdez Duty Station)	60
7	Palmer/Wasilla	216
8	Anchorage	300
9	Seward	24
10	Kenai/Cook Inlet	176
11	Kodiak	104
12	Aleutian Islands	79
13	Bristol Bay	56
14	Bethel	111
15A	Yukon/Kuskokwim (excluding Nenana Duty Station)	71
15B	Yukon/Kuskokwim (Nenana Duty Station)	7
16A	Fairbanks/Fort Yukon (South of Arctic Circle)	378
16B	Fairbanks/Fort Yukon (North of Arctic Circle)	20
17	Barrow/Kotzebue	100
18	Nome	70
19	Wade Hampton	29

Table VI-5: Sample Sizes for Individual Communities

Sample Block and Community	Sample Size
1: Anchorage	300
2: Fairbanks North Star Borough	300
3: Healy	22
3: Cantwell	4
3: Central	2
3: Nenana	7
3: Manley Hot Springs	1
3: Talkeetna	29
4: Glennallen	41
4: Chitina	3
4: Paxson	1
4: Slana	3
4: Tazlina	2
5: Delta Junction	53
5: Tok	21
5: Eagle	1
5: Northway	1
6: Galena	21
6: Fort Yukon	20
6: McGrath	10
7: Juneau	300
8: Ketchikan	120
8: Sitka	80
9: Craig	13
9: Haines	23
9: Klawock	7
9: Metlakatla	12
9: Petersburg	30
9: Wrangell	19
10: Hoonah	14
10: Skagway	14
10: Yakutat	10
10: Elfin Cove	1
10: Gustavus	8
10: Pelican	3
10: Tenakee Springs	2
11: Palmer	37
11: Wasilla	82
11: Willow	24
11: Other Mat-Su Borough	44

Table VI-5 cont'd: Sample Sizes for Individual Communities

Sample Block and Community	Sample Size
12: Seward	22
12: Kasilof	4
12: Kenai	77
12: Nikiski	6
12: Soldotna	23
12: Sterling	10
12: Homer	26
12: Anchor Point	6
12: Cooper Landing	2
12: Ninilchik	4
12: Seldovia	6
12: Other Kenai Peninsula	14
13: Cordova	37
13: Valdez	60
13: Whittier	3
14: Kodiak	104
15: Barrow	56
15: Kotzebue	44
15: Nome	48
15: Teller	5
16: Bethel	106
16: Dillingham	45
17: Adak	2
17: Cold Bay	1
17: King Cove	10
17: Sand Point	13
17: Unalaska/Dutch Harbor	51
18: Aniak	11
18: Anvik	2
18: Chignik	2
18: Emmonak	17
18: Goodnews Bay	5
18: Iliamna	2
18: King Salmon	9
18: Saint Mary's	12
18: Unalakleet	17

The state could use the results of this study to set a unique differential for each individual community or regional block. However, such an approach may prove to be administratively inefficient, and could lead to a de facto assumption regarding the precision of the estimated differentials that does not exist – especially for small communities. Alternatively, a differential could be set that would apply to subsets of communities. There are numerous options for grouping communities based on such factors as geographic region, political boundaries, community size, or the size of the estimated differential. There are likely positive and negative aspects to each of these grouping methods.

Among the many alternatives for grouping communities is one based purely on the statistical similarity between pairs of estimated differentials. As noted earlier, each differential represents the mean value of the difference in the cost of living between the particular community and Anchorage. For each differential, an *approximate* standard error was estimated based on Monte Carlo simulation and the development of generalized variance functions. Using this information, along with the sample size for each regional block or individual community, the Fisher Least Significant Difference (FLSD) method was used to group communities and regional blocks based on individual pair-wise comparisons between the respective differentials.

The FLSD is a procedure for comparing the means of multiple (more than two) populations based on individual pair-wise comparisons. The two-step procedure begins with a standard one-way analysis of variance (ANOVA) test to determine if the differentials of all of the regional blocks and individual communities are jointly equal (null hypothesis), versus the alternative hypothesis that at least one differential is different from the others. If the null hypothesis is rejected, as it was in this analysis, then the second step of the FLSD procedure is taken. The second step consists of applications of two-sample *t*-tests between every pair of means.[14] Two differentials are placed in the same group if results of the *t*-test indicate there is not a statistically significant difference between the two differentials.[15]

Results of the FLSD Analysis

The results of the FLSD analysis indicate that the 18 regional blocks and 11 communities examined can be pooled into four groups (see VI-6). One community, Valdez, could be placed in either of two groupings.[16]

[14] The total number of two-sample combinations is equal to "24 choose 2" or $\left(\frac{24!}{2!(24-2)!}\right)$, which results in 276 combinations.

[15] For more information on the Fisher Least Significant Difference Method, please see: Koopmans, L.H., Introduction to Contemporary Statistical Methods, Second Edition, PWS Publishers, 1987.

[16] The estimated differential for Valdez, 1.08, is closer to the differentials of regional blocks and communities in Group 2. However, its relatively small sample size (60 observations) results in it also being part of Group 1, the "Anchorage" block.

Table VI-6: Community Groupings Based on FLSD Method

Sample Block/Community		Differential	Group #
Sample Blocks			
1	Anchorage	1.00	1
2	Fairbanks	1.03	1
3	Parks/Elliott/Steese Highways	1.00	1
4	Glennallen Region	0.97	1
5	Delta Junction/Tok Region	1.04	1
6	Roadless Interior	1.31	3
7	Juneau	1.11	2
8	Ketchikan/Sitka	1.09	2
9	Southeast Mid-Size Communities	1.05	1
10	Southeast Small Communities	1.02	1
11	Mat-Su	0.95	1
12	Kenai Peninsula	1.01	1
13	Prince William Sound	1.08	2
14	Kodiak	1.12	2
15	Arctic Region	1.48	4
16	Bethel/Dillingham	1.49	4
17	Aleutian Region	1.50	4
18	Southwest Small Communities	1.44	4
Communities			
	Barrow	1.50	4
	Bethel	1.53	4
	Cordova	1.13	2
	Dillingham	1.37	3
	Homer	1.03	1
	Ketchikan	1.04	1
	Kotzebue	1.61	5
	Nome	1.39	3
	Sitka	1.17	2
	Unalaska/Dutch Harbor	1.58	5
	Valdez	1.08	2

Table VI-7 shows the minimum and maximum differentials for each community group and the number of communities within that group.

Table VI-7: Community Grouping Statistics

2008 GDP #	Sample Blocks and/or Communities	Minimum Differential	Maximum Differential
1	Anchorage, Delta Junction/Tok Region, Fairbanks, Glennallen Region, Kenai Peninsula, Ketchikan, Mat-Su, Parks/Elliott/Steese Highways, Southeast Mid-size Communities, Southeast Small Communities	.95	1.05
2	Cordova, Juneau, Kodiak, Sitka, Valdez	1.08	1.17
3	Dillingham, Nome, Roadless Interior	1.31	1.39
4	Barrow, Bethel, Aleutians (other than Unalaska/ Dutch Harbor), Southwest Small Communities	1.44	1.53
5	Kotzebue, Unalaska/Dutch Harbor	1.58	1.61