Example – Jackknife Estimation Method

This is a simple example showing how the Jackknife estimates are calculated using the Jackknife estimation method. The example consists of 2018 Survey of International Air Travelers (SIAT) and Department of Homeland Security (DHS) I-94 data to calculate the variance for Kazakhstan residents with a main destination to the state of New York. The SIAT is used to collect up-to-date information about international air travels to and from the Unites States. The I-94 is the official U.S. government database for arrival counts into the United States by foreign country. The results calculated in this example pertain to Table 23 of the National Travel and Tourism Office (NTTO)'s annual publication, "Overseas Travelers to the United States 2018." In the report, Table 23 displays responses to question 3c, which asks the respondent their main destination for their current trip.

Example - Jackknife Estimation Method

Example: Residents of Kazakhstan with a main destination to the state of New York – 2018 SIAT data

*Please note: This is a simple example to show how the Jackknife estimation method works. Drawing conclusions off a sample size of eleven is not adequate to produce reliable estimates.

1) **Weighting the data**
   - Once the information is collected from the sample data, a weighting step is used for the final analysis to calculate a variable called '**FINALWT**'. The weights are constructed for each response and the sum of the weights represent the number of passengers in the population. The weight for each record is the number of passengers represented by the sample respondent.

     The weighting step can be called post-stratification using I-94 information. The post-stratum variable is the cross classification of "country of residence" and "port of customs information." The step-by-step description of the weighting procedure can be described as follows:

     Step 1. The initial weight **(WTINIT)** has a value greater than one, depending on the number of children and the size of the travel party. (Children are not surveyed. The initial weight proportions the number of children to the adults in the party)

     WTINIT = 1 + Child/Adult

     Step 2. Both the I-94 data and Survey responses are sorted and summarized by country of residence and the U.S. port of customs.

     Step 3. The weight computed for individual Survey responses is the result of directly proportioning the I-94 data to the Surveys by country of residence and the U.S. port of customs.

     FINALWT is calculated as follows:

     For matchleg = 1; when the record contains both a I-94 count and IFS count**:**

     - I94count ≠ missing & WTinsum ≠ missing
       -> **FINALWT** = I94count×WTinit / WTinsum

     For matchleg = 2; when the record is missing the I-94 country of residence and/or U.S. port of entry is missing**:**

- I94count = missing & WTinsum ≠ missing
  -> **FINALWT** = WTinit / WTinsum

For **matchleg = 3;** when there is no corresponding IFS data, (i.e. U.S. port of customs and country of residence)**:**

- -> **FINALWT** = I94 count

Variable definitions:

WTINIT is the initial weight,
WTinsum is the sum of initial weight by country of residence and port of customs,
I94count is the sum of country of residence and port of entry pairs,
FINALWT is the Final weight,
matchleg is the definition of the comparison of Survey records to I-94 counts:

2) Subset the data and define the variables

CIC ID Number – A unique ID given to each respondent who participates in the Survey of International Air Traffic.

Q1. Airline Code – The numerical code used to distinguish airlines throughout the coding process.

Q1. Flight Number – A one to four-digit number used to identify the flight.

Q1. Month of Flight – The month the flight took place.

Q1. Day of Flight – The day the flight took place.

Q3a. Country of Residence – The country of residence recorded by the respondent.

Q3b. Main Destination – The main destination recorded by the respondent.

FINALWT – This is a weight factor applied to each respondent. (Described in Part 1)

flightid – This is a concatenation of 6 variables to create a unique flight identifier. The variables that make up flightid are air flag, airline, flight number, year of flight, month of flight, and day of flight.

y – This is a flag variable. This flags the set of observations that are the target of the Jackknife variance estimates.

- First, subset the data for all respondents who recorded Kazakhstan as their country of residence.
- Once the data has been subset for residents of Kazakhstan, flag all Kazakhstan residents whose main destination is the state of New York.
- In this case of this example, all the respondents who recorded their country of residence as Kazakhstan **AND** recorded the state of New York as their main destination will be flagged. That means the **y** variable will be equal to 1 for these observations and 0 for the rest.
- Once the data is subset and flagged, there is a total of 11 records and 5 are flagged for New York as the main destination,

	CIC ID Number	Q1. Airline Code	Q1. Flight Number	Q1. Month of Flight	Q1. Day of Flight	Q3a. Country of Residence	Q3b. Main Destination on this trip	FINALWT	flightid	count	y	st
1	931055	41	0907	6	28	73102	26055	13	10410907180628	1	0	26
2	912154	23	0405	5	3	73102	36061	175	20230405180503	1	1	36
3	993469	23	0441	9	20	73102	29213	316	20230441180920	1	0	29
4	950236	67	0080	7	27	73101	4013	507	20670080180727	1	0	4
5	831465	93	0101	1	16	73102	36061	1009	20930101180116	1	1	36
6	870826	93	0101	3	15	73102	36061	1009	20930101180315	1	1	36
7	912106	93	0101	5	3	73101	36061	2752	20930101180503	1	1	36
8	951076	93	0101	7	24	73102	36029	1520	20930101180724	1	1	36
9	991307	93	0101	9	10	73102	6000	507	20930101180910	1	0	6
10	991306	93	0101	9	10	73102	55021	18	20930101180910	1	0	55
11	991244	93	0101	9	17	73101	51810	507	20930101180917	1	0	51

- The table below shows how many **UNIQUE** flights each airline has. The flights are color coded to help distinguish unique flights.

Airline	Flight #	Month of Flight	Day of Flight	# of Unique Flights
41	907	6	23	**1**
23	405	5	3	**2**
	441	9	20	
67	80	7	27	**1**
93	101	1	16	**6**
	101	3	15	
	101	5	3	
	101	7	24	
	101	9	10	
	101	9	10	
	101	9	17	

The number of **UNIQUE** flights an airline has, is an integral component of the Jackknife computation. If an airline only has **1 UNIQUE** flight, then the airline and all records associated with it are removed from the data set. The technique is invalid otherwise and will not function properly.

- After removing airlines 41 and 67 to ensure the Jackknife technique will work, we are left with 9 records, 5 of which are flagged for New York as the main destination **(y)**, and 8 **UNIQUE** flights. As mentioned before, the number of **UNIQUE** flights is important. Variable *nhk* represents the number of **UNIQUE** flights for the airline associated with that record. It is not a variable that is part of the original data set but created after the initial subset takes place. It is a check to make sure that airlines with one unique flight do not get past this part of the process.

	CIC ID Number	Q1. Airline Code	Q1. Flight Number	Q1. Month of Flight	Q1. Day of Flight	Q3a. Country of Residence	Q3b. Main Destination on this trip	FINALWT	flightid	count	y	st	nhk
1	912154	23	0405	5	3	73102	36061	175	20230405180503	1	1	36	2
2	993469	23	0441	9	20	73102	29213	316	20230441180920	1	0	29	2
3	831465	93	0101	1	16	73102	36061	1009	20930101180116	1	1	36	6
4	870826	93	0101	3	15	73102	36061	1009	20930101180315	1	1	36	6
5	912106	93	0101	5	3	73101	36061	2752	20930101180503	1	1	36	6
6	951076	93	0101	7	24	73102	36029	1520	20930101180724	1	1	36	6
7	991307	93	0101	9	10	73102	6000	507	20930101180910	1	0	6	6
8	991306	93	0101	9	10	73102	55021	18	20930101180910	1	0	55	6
9	991244	93	0101	9	17	73101	51810	507	20930101180917	1	0	51	6

3) **The tricky part, math**
   - First look at the formulas and try to make as much sense of the notation as we can.

Jackknife Estimation Method

Each stratum must have at least two PSUs to use the Jackknife estimation method. We will only consider the case when a PSU is deleted for Jackknife. (Delete-a-group Jackknife method can be developed if necessary, but we will cover only the classical Jackknife method where a PSU is deleted.)

Strata/Stratum: airline

Primary sampling unit (PSU), cluster: flight

Let K be the total number of PSUs.

Let h_k be the stratum from which the deleted PSU for the kth replicate belongs to.

Let n_{h_k} be the total number of PSUs in the stratum h_k.

The Jackknife variance estimator can be written as

$$\hat{V}(\hat{\theta}) = \sum_{k=1}^{K} \alpha_k (\hat{\theta}^{(k)} - \hat{\theta})^2$$

where the Jackknife coefficients are defined as $\alpha_k = \dfrac{n_{h_k}-1}{n_{h_k}}, \ k = 1, \cdots, K$,

$\hat{\theta}^{(k)}$ is the estimate from the kth Jackknife replicate subsample by using replicate weights, and $\hat{\theta}$ is the estimate from the full sample for θ, which is a population parameter of interest,

i.e. $\hat{\theta}^{(k)} = \dfrac{\sum w_{ij}^{(k)} y_{ij}}{\sum w_{ij}^{(k)}}$ and $\hat{\theta} = \dfrac{\sum w_{ij} y_{ij}}{\sum w_{ij}}$

where $w_{ij}^{(k)}$ is the k-th replicate weight for the jth member of the ith PSU, and w_{ij} is the variable FINALWT. The replication weights are computed by

$$w_{ij}^{(k)} = \begin{cases} w_{ij} & if \ i \ th \ PSU \ is \ not \ in \ the \ stratum \ h_k \\ w_{ij}/\alpha_k & if \ i \ th \ PSU \ is \ in \ the \ stratum \ h_k \end{cases}$$

for $i \neq k$. Also, $w_{ij}^{(k)} = 0$ for $i = k$.

Reference

SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc., chapter 85, p6500.

- What we know:

K = Total # of Primary Sampling Units (PSUs) = # of **UNIQUE FLIGHTS**.
h_k = The stratum(airline) from which the **DELETED** Primary Sampling Unit (**UNIQUE FLIGHT**) for the kth replicate belongs to.
n_{h_k} = The total # of Primary Sampling Units (**UNIQUE FLIGHTS)** within each Strata (**Airline**).
$\hat{\theta}^{(k)}$ = The estimate from the kth Jackknife replicate subsample by using replicate weights, which is a estimate of the population parameter of interest.

$\hat{\theta}$ = The estimate from the full sample for θ, which is a population parameter of interest.

w_{ij} = Final weight for the jth member of the ith PSU = Final weight for each observation within each flight.

$w_{ij}^{(k)}$ = The kth replicate weight for the jth member of the ith PSU = Replicate weights for each observation within each airline.
y_{ij} = This is the flag variable, from part one, to denote the respondent's main destination as New York. The ith and jth notation refer to the ith PSU (flight) and jth observation within that PSU(flight).
α_k = The Jackknife coefficients for each Strata (Airline).

For this example, we know

K = **8**
h_k = Airline 23-2 flights and Airline 93-6 flights
n_{h_k} = 2 and 6
$\hat{\theta} = \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}}$ and $\hat{\theta}^{(k)} = \frac{\sum w_{ij}^{(k)} y_{ij}}{\sum w_{ij}^{(k)}}$; more on this below.
w_{ij} = This is the final weight associated with each observation. The final weight is part of the Survey of International Air Traffic data processing and part of the original dataset.

$$w_{ij}^{(k)} = \begin{cases} w_{ij} & \text{if } i \text{ th PSU is not in the stratum } h_k \\ w_{ij}/\alpha_k & \text{if } i \text{ th PSU is in the stratum } h_k \end{cases}$$

for $i \neq k$. Also, $w_{ij}^{(k)} = 0$ for $i = k$. More on this below.

y_{ij} = 1 for every observation where New York is the main destination and 0 otherwise.

$$\alpha_k = \frac{n_{h_k} - 1}{n_{h_k}}, \; k = 1, \cdots, K,$$

Lets solve for α_k Since we know n_{h_k} we can figure out α_k for each airline(strata):

n_{h_k} = **2** flights for airline 23 and **6** flights for airline 93
$$\alpha_k = \frac{n_{h_k} - 1}{n_{h_k}}, \; k = 1, \cdots, K,$$
$$\alpha_k = \frac{(2-1)}{2} = \mathbf{1/2}$$
$$\alpha_k = \frac{(6-1)}{6} = \mathbf{5/6}$$

- Update the formula to see what is know,

$$\hat{V}(\hat{\theta}) = \sum_{k=1}^{K} \alpha_k (\hat{\theta}^{(k)} - \hat{\theta})^2$$

We know that $K = 8$, which means we will have to sum this part of the equation **8** times

$$\alpha_k (\hat{\theta}^{(k)} - \hat{\theta})^2$$

We know for the 2 flights on airline 23 that $\alpha_k = \frac{1}{2}$ and the 6 flight on airline 93 that $\alpha_k = \frac{5}{6}$

To expand on the equation for each iteration,

When $K=1 \rightarrow \frac{1}{2} (\hat{\theta}^{(1)} - \hat{\theta})^2 +$
$\quad\quad K=2 \rightarrow \frac{1}{2} (\hat{\theta}^{(2)} - \hat{\theta})^2 +$
$\quad\quad K=3 \rightarrow \frac{5}{6} (\hat{\theta}^{(3)} - \hat{\theta})^2 +$
$\quad\quad K=4 \rightarrow \frac{5}{6} (\hat{\theta}^{(4)} - \hat{\theta})^2 +$
$\quad\quad K=5 \rightarrow \frac{5}{6} (\hat{\theta}^{(5)} - \hat{\theta})^2 +$
$\quad\quad K=6 \rightarrow \frac{5}{6} (\hat{\theta}^{(6)} - \hat{\theta})^2 +$
$\quad\quad K=7 \rightarrow \frac{5}{6} (\hat{\theta}^{(7)} - \hat{\theta})^2 +$
$\quad\quad K=8 \rightarrow \frac{5}{6} (\hat{\theta}^{(8)} - \hat{\theta})^2 = \hat{V}(\hat{\theta})$

We know $\hat{\theta} = \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}}$ and $\hat{\theta}^{(k)} = \frac{\sum w_{ij}^{(k)} y_{ij}}{\sum w_{ij}^{(k)}}$, so when we plug it in the equation looks as below,

$K=1 \rightarrow \frac{1}{2} \left(\frac{\sum w_{ij}^{(1)} y_{ij}}{\sum w_{ij}^{(1)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=2 \rightarrow \frac{1}{2} \left(\frac{\sum w_{ij}^{(2)} y_{ij}}{\sum w_{ij}^{(2)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=3 \rightarrow \frac{5}{6} \left(\frac{\sum w_{ij}^{(3)} y_{ij}}{\sum w_{ij}^{(3)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=4 \rightarrow \frac{5}{6} \left(\frac{\sum w_{ij}^{(4)} y_{ij}}{\sum w_{ij}^{(4)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=5 \rightarrow \frac{5}{6} \left(\frac{\sum w_{ij}^{(5)} y_{ij}}{\sum w_{ij}^{(5)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=6 \rightarrow \frac{5}{6} \left(\frac{\sum w_{ij}^{(6)} y_{ij}}{\sum w_{ij}^{(6)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=7 \rightarrow \frac{5}{6} \left(\frac{\sum w_{ij}^{(7)} y_{ij}}{\sum w_{ij}^{(7)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 +$

$K=8 \rightarrow \frac{5}{6} \left(\frac{\sum w_{ij}^{(8)} y_{ij}}{\sum w_{ij}^{(8)}} - \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} \right)^2 = \hat{V}(\hat{\theta})$

- Now solve $\hat{\theta} = \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}}$. This will be the same for every part of the summation because it's the estimate from the full sample

$$\hat{\theta} = \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}}$$

w_{ij} and y_{ij} are known from the original setup of the data. Look at the data set below,

CIC ID Number	Airline Code	flightid	nhk	FINALWT	y
912154	23	20230405180503	2	w_{11}	y_{11}
993469	23	20230441180920	2	w_{21}	y_{21}
831465	93	20930101180116	6	w_{31}	y_{31}
870826	93	20930101180315	6	w_{41}	y_{41}
912106	93	20930101180503	6	w_{51}	y_{51}
951076	93	20930101180724	6	w_{61}	y_{61}
991307	93	20930101180910	6	w_{71}	y_{71}
991306	93	20930101180910	6	w_{72}	y_{72}
991244	93	20930101180917	6	w_{81}	y_{81}

Remember, w_{ij}= Final weight for the jth member of the ith PSU**(UNIQUE FLIGHT)**

Notice the highlighted cells, the flightid is the same. These two respondents were on the same flight. The notation, w_{71} and w_{72}, distinguishes the unique flight(7) and the respondent on that flight(1, 2).

CIC ID Number	Airline Code	flightid	nhk	FINALWT	y
912154	23	20230405180503	2	175	1
993469	23	20230441180920	2	316	0
831465	93	20930101180116	6	1009	1
870826	93	20930101180315	6	1009	1
912106	93	20930101180503	6	2752	1
951076	93	20930101180724	6	1520	1
991307	93	20930101180910	6	507	0
991306	93	20930101180910	6	18	0
991244	93	20930101180917	6	507	0

Looking at the data $\hat{\theta}$ can be solved for,

$$\hat{\theta} = \frac{\sum w_{ij} y_{ij}}{\sum w_{ij}} = \frac{(175*1)+(316*0)+(1009*1)+(1009*1)+(2752*1)+(1520*1)+(507*0)+(18*0)+(507*0)}{(175+316+1009+1009+2752+1520+507+18+507)} = \frac{6465}{7813} = \mathbf{.8275}$$

- The only part of the equation left is $\hat{\theta}^{(k)} = \frac{\sum w_{ij}^{(k)} y_{ij}}{\sum w_{ij}^{(k)}}$, but in order to solve for $\hat{\theta}^{(k)}$ the replicate final weights need to be calculated first. The replicate weights formula is

$$w_{ij}^{(k)} = \begin{cases} w_{ij} & \text{if } i \text{ th PSU is not in the stratum } h_k \\ w_{ij}/\alpha_k & \text{if } i \text{ th PSU is in the stratum } h_k \end{cases}$$

for $i \neq k$. Also, $w_{ij}^{(k)} = 0$ for $i = k$.

- The data set below shows the notation for each replicate weight and where they reside for each observation and the one after that shows all the replicate weights,

CIC ID Number	Airline Code	flightid	nhk	FINALWT	y	Replicate #1	Replicate #2	Replicate #3	Replicate #4	Replicate #5	Replicate #6	Replicate #7	Replicate #8
912154	23	20230405180503	2	w_{11}	y_{11}	w_{11}^1	w_{11}^2	w_{11}^3	w_{11}^4	w_{11}^5	w_{11}^6	w_{11}^7	w_{11}^8
993469	23	20230441180920	2	w_{21}	y_{21}	w_{21}^1	w_{21}^2	w_{21}^3	w_{21}^4	w_{21}^5	w_{21}^6	w_{21}^7	w_{21}^8
831465	93	20930101180116	6	w_{31}	y_{31}	w_{31}^1	w_{31}^2	w_{31}^3	w_{31}^4	w_{31}^5	w_{31}^6	w_{31}^7	w_{31}^8
870826	93	20930101180315	6	w_{41}	y_{41}	w_{41}^1	w_{41}^2	w_{41}^3	w_{41}^4	w_{41}^5	w_{41}^6	w_{41}^7	w_{41}^8
912106	93	20930101180503	6	w_{51}	y_{51}	w_{51}^1	w_{51}^2	w_{51}^3	w_{51}^4	w_{51}^5	w_{51}^6	w_{51}^7	w_{51}^8
951076	93	20930101180724	6	w_{61}	y_{61}	w_{61}^1	w_{61}^2	w_{61}^3	w_{61}^4	w_{61}^5	w_{61}^6	w_{61}^7	w_{61}^8
991307	93	20930101180910	6	w_{71}	y_{71}	w_{71}^1	w_{71}^2	w_{71}^3	w_{71}^4	w_{71}^5	w_{71}^6	w_{71}^7	w_{71}^8
991306	93	20930101180910	6	w_{72}	y_{72}	w_{72}^1	w_{72}^2	w_{72}^3	w_{72}^4	w_{72}^5	w_{72}^6	w_{72}^7	w_{72}^8
991244	93	20930101180917	6	w_{81}	y_{81}	w_{81}^1	w_{81}^2	w_{81}^3	w_{81}^4	w_{81}^5	w_{81}^6	w_{81}^7	w_{81}^8

CIC ID Number	Airline Code	flightid	nhk	FINALWT	y	Replicate #1	Replicate #2	Replicate #3	Replicate #4	Replicate #5	Replicate #6	Replicate #7	Replicate #8
912154	23	20230405180503	2	175	1	0	350	175	175	175	175	175	175
993469	23	20230441180920	2	316	0	632	0	316	316	316	316	316	316
831465	93	20930101180116	6	1009	1	1009	1009	0	1210.8	1210.8	1210.8	1210.8	1210.8
870826	93	20930101180315	6	1009	1	1009	1009	1210.8	0	1210.8	1210.8	1210.8	1210.8
912106	93	20930101180503	6	2752	1	2752	2752	3302.4	3302.4	0	3302.4	3302.4	3302.4
951076	93	20930101180724	6	1520	1	1520	1520	1824	1824	1824	0	1824	1824
991307	93	20930101180910	6	507	0	507	507	608.4	608.4	608.4	608.4	0	608.4
991306	93	20930101180910	6	18	0	18	18	21.6	21.6	21.6	21.6	0	21.6
991244	93	20930101180917	6	507	0	507	507	608.4	608.4	608.4	608.4	608.4	0

- Let's solve for the first column of replicate weights. This consists of,

$$w_{11}^{(1)}, w_{21}^{(1)}, w_{31}^{(1)}, w_{41}^{(1)}, w_{51}^{(1)}, w_{61}^{(1)}, w_{71}^{(1)}, w_{72}^{(1)}, w_{81}^{(1)}$$

$$w_{11}^{(1)} = 0, \text{ because } i = k$$
$$w_{21}^{(1)} = w_{ij}/\alpha_k = \frac{316}{\left(\frac{1}{2}\right)} = 632$$

$$w_{31}^{(1)} = 1009$$
$$w_{41}^{(1)} = 1009$$
$$w_{51}^{(1)} = 2752$$
$$w_{61}^{(1)} = 1520$$
$$w_{71}^{(1)} = 507$$
$$w_{72}^{(1)} = 18$$
$$w_{81}^{(1)} = 507$$

With this is info we can solve for $\hat{\theta}^{(1)}$,

$$\hat{\theta}^{(1)} = \frac{(0*1)+(632*0)+(1009*1)+(1009*1)+(2752*1)+(1520*1)+(507*0)+(18*0)+(507*0)}{(0+632+1009+1009+2752+1520+507+18+507)} = \frac{6290}{7954} = .7908$$

- This process is needed 7 more times because there are **8 UNIQUE FLIGHTS**.

$$\hat{\theta}^{(2)} = \frac{(350*1)+(0*0)+(1009*1)+(1009*1)+(2752*1)+(1520*1)+(507*0)+(18*0)+(507*0)}{(350+0+1009+1009+2752+1520+507+18+507)} = \frac{6640}{7672} = \textbf{.8655}$$

$$\hat{\theta}^{(3)} = \frac{(175*1)+(316*0)+(0*1)+(1210.8*1)+(3302.4*1)+(1824*1)+(608.4*0)+(21.6*0)+(608.4*0)}{(175+316+0+1210.8+3302.4+1824+608.4+21.6+608.4)} = \frac{6512.2}{8066.6} = \textbf{.8073}$$

$$\hat{\theta}^{(4)} = \frac{(175*1)+(316*0)+(1210.8*1)+(0*1)+(3302.4*1)+(1824*1)+(608.4*0)+(21.6*0)+(608.4*0)}{(175+316+1210.8+0+3302.4+1824+608.4+21.6+608.4)} = \frac{6512.2}{8066.6} = \textbf{.8073}$$

$$\hat{\theta}^{(5)} = \frac{(175*1)+(316*0)+(1210.8*1)+(1210.8*1)+(0*1)+(1824*1)+(608.4*0)+(21.6*0)+(608.4*0)}{(175+316+1210.8+1210.8+0+1824+608.4+21.6+608.4)} = \frac{4420.6}{5975} = \textbf{.7398}$$

$$\hat{\theta}^{(6)} = \frac{(175*1)+(316*0)+(1210.8*1)+(1210.8*1)+(3302.4*1)+(0*1)+(608.4*0)+(21.6*0)+(608.4*0)}{(175+316+1210.8+1210.8+3302.4+0+608.4+21.6+608.4)} = \frac{5899}{7453.4} = \textbf{.7915}$$

$$\hat{\theta}^{(7)} = \frac{(175*1)+(316*0)+(1210.8*1)+(1210.8*1)+(3302.4*1)+(1824*1)+(0*0)+(0*0)+(608.4*0)}{(175+316+1210.8+1210.8+3302.4+1824+0+0+608.4)} = \frac{7723}{8647.4} = \textbf{.8931}$$

$$\hat{\theta}^{(8)} = \frac{(175*1)+(316*0)+(1210.8*1)+(1210.8*1)+(3302.4*1)+(1824*1)+(608.4*0)+(21.6*0)+(0*0)}{(175+316+1210.8+1210.8+3302.4+1824+608.4+21.6+0)} = \frac{7723}{8669} = \textbf{.8909}$$

- Now the variance can be solved for, $\hat{V}(\hat{\theta}) = \sum_{k=1}^{K} \alpha_k(\hat{\theta}^{(k)} - \hat{\theta})^2$,

K=1 → $\alpha_k(\hat{\theta}^{(1)} - \hat{\theta})^2$ → $\frac{1}{2}$ $(.7908 - .8275)^2$ → $\frac{1}{2}$ $(-.0367)^2$ = .00067

K=2 → $\alpha_k(\hat{\theta}^{(2)} - \hat{\theta})^2$ → $\frac{1}{2}$ $(.8655 - .8275)^2$ → $\frac{1}{2}$ $(.038)^2$ = .00072

K=3 → $\alpha_k(\hat{\theta}^{(3)} - \hat{\theta})^2$ → $\frac{5}{6}$ $(.8073 - .8275)^2$ → $\frac{5}{6}$ $(-.0202)^2$ = .00034

K=4 → $\alpha_k(\hat{\theta}^{(4)} - \hat{\theta})^2$ → $\frac{5}{6}$ $(.8073 - .8275)^2$ → $\frac{5}{6}$ $(-.0202)^2$ = .00034

K=5 → $\alpha_k(\hat{\theta}^{(5)} - \hat{\theta})^2$ → $\frac{5}{6}$ $(.7398 - .8275)^2$ → $\frac{5}{6}$ $(-.0877)^2$ = .00641

K=6 → $\alpha_k(\hat{\theta}^{(6)} - \hat{\theta})^2$ → $\frac{5}{6}$ $(.7915 - .8275)^2$ → $\frac{5}{6}$ $(-.036)^2$ = .00108

K=7 → $\alpha_k(\hat{\theta}^{(7)} - \hat{\theta})^2$ → $\frac{5}{6}$ $(.8931 - .8275)^2$ → $\frac{5}{6}$ $(.0656)^2$ = .00359

K=8 → $\alpha_k(\hat{\theta}^{(8)} - \hat{\theta})^2$ → $\frac{5}{6}$ $(.8909 - .8275)^2$ → $\frac{5}{6}$ $(.0634)^2$ = .00335

$$\hat{V}(\hat{\theta}) = .00067 + .00072 + .00034 + .00034 + .00641 + .00108 + .00359 + .00335 =$$

<u>.0165</u>

The variance has now been solved for and other tests can be performed with this estimate.

4) What does it mean?

- The variance is a measure that describes how much a random variable differs from its expected value. That means, the bigger the variance the bigger the dispersion or range of the data and vice versa.
- Once the Jackknife variance has be calculated calculations such as the standard error, significance testing, and confidence intervals can all be calculated.

- The standard error of a statistic is the standard deviation of its sampling distribution or an estimate of that standard deviation.

  The standard error is equal to square root of the variance,

  SE = $\sqrt{Jackknife\ Variance}$

  From the example above, the standard error would be equal to $\sqrt{.0165}$ = .1285.

- A 95% confidence interval is a range of values that you can be 95% certain contains the true mean of the population.

  $\bar{x} \pm 1.96$ (Standard Error) = 95% Confidence Interval
  $\hat{p} \pm 1.96$ (Standard Error) = 95% Confidence Interval

  The example above deals with a proportion but the formula still applies the same,

  $.7758 \pm 1.96(.1285) = .7758 \pm .2519 = (.5239, 1.0277) = (52.39\%, 102.77\%)$

  To interpret the results, we can safely assume that in 95% of all possible samples, the population proportion will fall between 24.14% and 26.80%.

  ***<u>**NOTE:**</u> Remember this is just a simple example that has 11 total observations! Nothing can be greater than 100%! This is a perfect example why conclusions should not be drawn using <u>**small data sets**</u>!

- A t-test is a form of significance testing where two unique point estimates can be compared. It is used to determine whether there is a significant difference between the point estimates of two groups. The t-value formula is show below.

  $t = (\bar{x}_1 - \bar{x}_2) / \sqrt{\widehat{V1} + \widehat{V2}}$,

  where $\bar{x}_1$ and $\bar{x}_2$ are equal to two mean estimates and $\widehat{V1}$ and $\widehat{V2}$ are equal to their variances.