USDA Food Data

IC 2024 - *Group 24055*: Sushil Shrestha, Bridgetta Ulrich, Meghana Padmanabhan, Rachel Weiss

Table of Contents

Background

- USDA's Agricultural Research Service (ARS): Analyzed foods for over a century, determining nutrient/component values for U.S. consumption.
- USDA's FoodData Central (FDC)
  - Launched in 2019 to provide transparent analytical data to Americans.
  - Integrated data system offers expanded nutrient/food component data in one location.
  - Includes five distinct data types, each with a unique purpose and possessing unique attributes.

Data

- Foundation Foods (FF): CURRENT data with extensive metadata, including sample details and analytical approaches.
- SR Legacy (SR): HISTORIC data that offers comprehensive values derived from analyses, imputations, and literature, but will not receive further updates.
- With scarce resources, FDC explores new ways to incorporate data due to the high cost of food analysis.

- **The project will focus on comparing two data types: FF and SR.**
  - Historical data analysis helps us understand nutrient/component changes over time, aiding in better resource allocation.

Understanding the Dataset

- NDB Number Classification:
  - Foods are identified by **SR_NDB and FF_NDB numbers.** With a unified NDB number linking both datasets.
- Nutrient Identification:
  - Each nutrient is assigned a specific code **(Nutrient_id)**. A separate table links nutrient codes to their descriptions and identifies the nutrient description.
- Ranking System:
  - Nutrients are ordered by rank, grouping similar types (e.g., proximates or vitamins).
- Nutrient Data Specifics:
  - Includes detailed data (mean, min, max, median).
  - Some nutrients are calculated, not measured, providing only mean values.
  - *Calculated Nutrients Examples: Energy, carbohydrate by difference, fatty acids, total content of fats.*

Goals

- **Adapt to declining resources** and limited staffing by finding new methods to integrate data within FDC.
- **Address the impracticality** of analyzing every nutrient for a single food item, especially when analysis costs surpass $50,000.
- **Gain insights into the changes** of comparable foods over time to determine which components require continued analysis and which remain relatively constant, potentially reducing unnecessary analysis during food updates.

Can we use this data to improve our understanding of analytical sampling?

Yes, analytical sampling allows for inferences to be made in regards to a larger population or data set. Using this data has allowed us to clearly define variables and then decide how to interpret the data and what steps are necessary to find answers to the questions given our assumptions.

This data set consists of **4,395 data records** which is just a snapshot to the wide variety of foods and nutrients that the USDA's Food Central database covers.

Have the food components changed over time for equivalent food?

We observed that **221** food components changed.

Many of these changes are **the way the components are recorded** rather than the change of the components in the foods themselves.

(e.g Vitamin D (D2 + D3), International Units vs Vitamin D)

Formulas used in Excel:

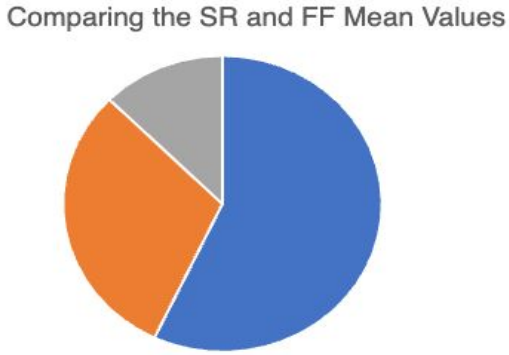
```
=IF(EXACT(TRIM(LOWER(C2)),
TRIM(LOWER(D2))), "No Change",
"Changed")
=COUNTIF(E:E, "Changed")
```

Vitamin D (D2 + D3), International Units	Vitamin D	Changed
Retinol	Retinol	No Change
SFA 14:0	14:00	Changed
Potassium, K	Potassium, K	No Change

FF_NDB	SR_NDB	FF_Component	SR_Component	Changed or No Change	Amount that Changed
16158	16158	Magnesium, Mg	Magnesium, Mg	No Change	221
16158	16158	Fatty acids, total saturated	Fatty acids, total saturated	No Change	
16158	16158	Iron, Fe	Iron, Fe	No Change	
16158	16158	Water	Water	No Change	
16158	16158	Selenium, Se	Selenium, Se	No Change	
16158	16158	Energy	Energy	No Change	
16158	16158	Carbohydrate, by difference	Carbohydrate, by difference	No Change	
16158	16158	Pantothenic acid	Pantothenic acid	No Change	

When comparing nutrients from the datasets, does the mean value for SR fall between the Min/Max values of FF?

Amount of foods that do: **2506**
Amount of foods that do not: **1336**
Amount of foods that are NA: **553**



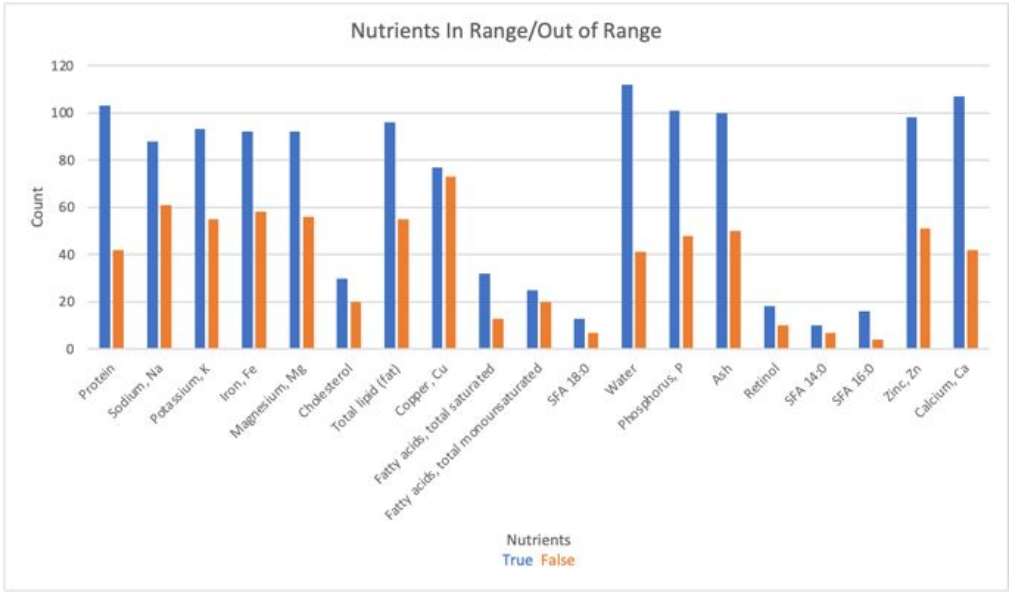
Formulas used:

```
=IF(OR(ISBLANK(Z2),ISBLANK(AA2)),"NA",IF(AND(Y2>=Z2,Y2<=AA2),"TRUE", "FALSE"))
```

SR Mean per 100g	FF Min	FF Max	In Range?	SUM OF FALSE	SUM OF TRUE	SUM OF NA
71	56.6	82	TRUE	0	1	0
1.437			NA	0	0	1
2.44	1.87	2.96	TRUE	0	1	0
66.59	56.1	65.6	FALSE	1	0	0
2.6	0	32.3	TRUE	0	1	0

Which nutrients fall out of range the most and which fall in range more frequently?

Protein	Sodium, Na	Potassium, K	Iron, Fe	Magnesium, Mg	Cholesterol	Total lipid (fat)	Copper, Cu	Fatty acids, total saturate	Fatty acids, total monounsaturate	SFA 18:0	Water	Phosphorus, P	Ash	Retinol	SFA 14:0	SFA 16:0	Zinc, Zn	Calcium, Ca
103	88	93	92	92	30	96	77	32	25	13	112	101	100	18	10	16	98	107
42	61	55	58	56	20	55	73	13	20	7	41	48	50	10	7	4	51	42



Formulas used: (example):

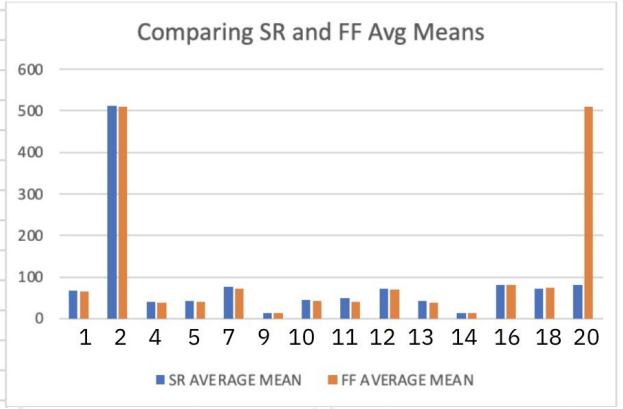
```
=COUNTIFS (B:B, PROTEIN, C:C,
TRUE)
COUNTIFS (B:B, PROTEIN, C:C,
FALSE)
```

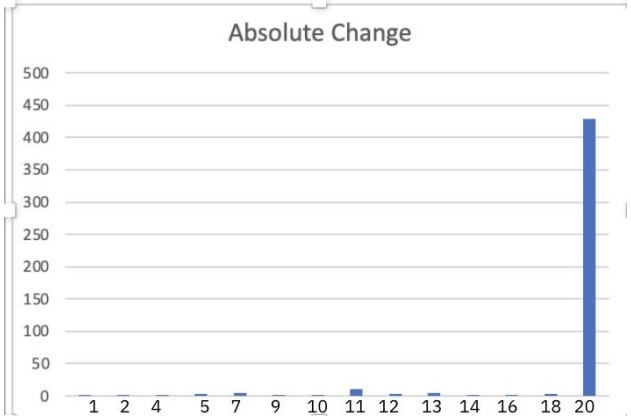
Which food groups have had the biggest change?

Food Group 20 had the largest change from historic SR Legacy dataset to current Foundation Foods dataset. Food Group 20 consists of Cereal Grains and Pasta

Steps:

1. Found the average of the SR Avg Means and FF Avg Means for each food group
   `=AVERAGEIF(C:C, "1", D:D)`
2. Found the absolute change between the SR and FF Avg Means
   `=ABS(I2-H2)`





Conclusion

Adapt to declining resources and limited staffing by finding new methods to integrate data within FDC.

→ **Exploring methods of data integration within the FDC platform: algorithms, machine learning models**

Address the impracticality of analyzing every nutrient for a single food item, especially when analysis costs surpass $50,000.

→ **Conduct a Cost-Benefit Analysis of every food item, compare them to weigh potential benefits and significance of the nutrient. Identifying nutrient are the most crucial way to make decisions**

→ **Seeing which nutrients pass the $50,000 threshold to focus on analyzing the nutrients that are cost-effective and most essential**

Gain insights into the changes of comparable foods over time to determine which components require continued analysis and which remain relatively constant, potentially reducing unnecessary analysis during food updates.

→ **Using Data Visualization to make it easier to understand trends and patterns.**

→ **Highlight which components had significant changes throughout the analysis.**

Appendix:

INFO CHALLENGE

Our Excel File

Link to the Case

USDA

Thank you for listening!

Questions?