

Output For Takehome Project Sarah Cummings

First, I use the descriptives feature to standardize the variables.

```
GET
  FILE='C:\Users\scummi10\Downloads\BurgersOriginal.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
DESCRIPTIVES VARIABLES=Calories TotalFat_g Sodium_mg CaloriesFromFat SaturatedFat_g TransFat_g
  Cholesterol_mg Carbohydrates_g Fiber_g Sugars_g Protein_g
  /SAVE
  /STATISTICS=MEAN STDDEV MIN MAX.
```

➔ Descriptives

[DataSet1] C:\Users\scummi10\Downloads\BurgersOriginal.sav

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Calories	190	140	1280	525.32	206.630
Total Fat (g)	190	3	92	26.96	15.065
Sodium (mg)	190	310	2920	1202.47	434.996
Calories From Fat	164	35	830	246.74	135.540
Saturated Fat (g)	190	1.0	36.0	8.421	6.2842
Trans Fat (g)	190	.0	4.0	.703	.9510
Cholesterol (mg)	190	5	260	75.47	43.822
Carbohydrates (g)	190	4	126	42.86	15.649
Fiber (g)	190	0	10	2.65	1.586
Sugars (g)	190	0	24	6.75	4.033
Protein (g)	190	9	72	27.73	11.393
Valid N (listwise)	164				

Next, I perform k-means cluster analysis with k= 3 using all of our z-score variables except calories from fat, since we do not have calories from fat for our Wendy's sandwiches.

➔ Quick Cluster

[DataSet1] C:\Users\scummi10\Desktop\spss data\BurgersOriginal.sav

Iteration History ^a			
Iteration	Change in Cluster Centers		
	1	2	3
1	4.503	3.451	3.021
2	.390	.289	.000
3	.097	.075	.000
4	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 4. The minimum distance between initial centers is 10.130.

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(Calories)	.61296	-.64546	2.38086
Zscore: Total Fat (g)	.46991	-.58637	2.72404
Zscore: Sodium (mg)	.67620	-.53294	.86788
Zscore: Saturated Fat (g)	.30056	-.50903	3.05772
Zscore: Trans Fat (g)	.17939	-.40291	2.79797
Zscore: Cholesterol (mg)	.37166	-.54326	2.93496
Zscore: Carbohydrates (g)	.64448	-.42579	.02070
Zscore: Fiber (g)	.53318	-.32062	-.29355
Zscore: Sugars (g)	.44681	-.33915	.44586
Zscore: Protein (g)	.51312	-.58924	2.47328

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(Calories)	67.012	2	.294	187	227.943	.000
Zscore: Total Fat (g)	67.218	2	.292	187	230.368	.000
Zscore: Sodium (mg)	35.712	2	.629	187	56.799	.000
Zscore: Saturated Fat (g)	68.622	2	.277	187	247.940	.000
Zscore: Trans Fat (g)	52.966	2	.444	187	119.237	.000
Zscore: Cholesterol (mg)	68.218	2	.281	187	242.685	.000
Zscore: Carbohydrates (g)	24.538	2	.748	187	32.793	.000
Zscore: Fiber (g)	16.117	2	.838	187	19.225	.000
Zscore: Sugars (g)	14.392	2	.857	187	16.798	.000
Zscore: Protein (g)	61.740	2	.350	187	176.211	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Number of Cases in each Cluster

Cluster	1	71.000
	2	108.000
	3	11.000
Valid		190.000
Missing		.000

Next, I use LDA with the saved cluster membership and all of my Z-score variables as the independent variables (except that calories from fat variable)

➔ Discriminant

Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		190	100.0
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	0	.0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	0	.0
Total		190	100.0

Analysis 1

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	4.158 ^a	92.0	92.0	.898
2	.360 ^a	8.0	100.0	.514

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.143	355.480	20	.000
2	.735	56.077	9	.000

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
Zscore: Sodium (mg)	.210	.486
Zscore: Saturated Fat (g)	.259	-.459
Zscore: Trans Fat (g)	.243	-.099
Zscore: Cholesterol (mg)	-.013	.064
Zscore: Carbohydrates (g)	-.089	.526
Zscore: Fiber (g)	.065	.165
Zscore: Sugars (g)	.170	.424
Zscore: Protein (g)	.406	.081
Zscore(Calories)	-.063	-.761
Zscore: Total Fat (g)	.474	.715

Structure Matrix

	Function	
	1	2
Zscore: Saturated Fat (g)	.787 [*]	-.458
Zscore: Cholesterol (mg)	.786 [*]	-.270
Zscore: Total Fat (g)	.770 [*]	.000
Zscore(Calories)	.757 [*]	.396
Zscore: Protein (g)	.672 [*]	.155
Zscore: Trans Fat (g)	.535 [*]	-.483
Zscore: Carbohydrates (g)	.176	.787 [*]
Zscore: Sodium (mg)	.321	.705 [*]
Zscore: Fiber (g)	.099	.677 [*]
Zscore: Sugars (g)	.166	.424 [*]

Functions at Group Centroids

Cluster Number of Case	Function	
	1	2
1	1.130	.695
2	-1.410	-.311
3	6.551	-1.431

Unstandardized canonical discriminant functions evaluated at group means

Classification Statistics

Classification Processing Summary

Processed	190
Excluded	
Missing or out-of-range group codes	0
At least one missing discriminating variable	0
Used in Output	190

Prior Probabilities for Groups

Cluster Number of Case	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1	.333	71	71.000
2	.333	108	108.000
3	.333	11	11.000
Total	1.000	190	190.000

Box's Test of Equality of Covariance Matrices

Log Determinants

Cluster Number of Case	Rank	Log Determinant
1	10	-16.751
2	10	-21.327
3	10	-40.485
Pooled within-groups	10	-18.415

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results

Box's M	415.773
F	Approx. 2.818
df1	110
df2	2228.567
Sig.	.000

Tests null hypothesis of equal population covariance matrices.

Next, I will create a scatterplot based on my discriminant scores and my cluster membership

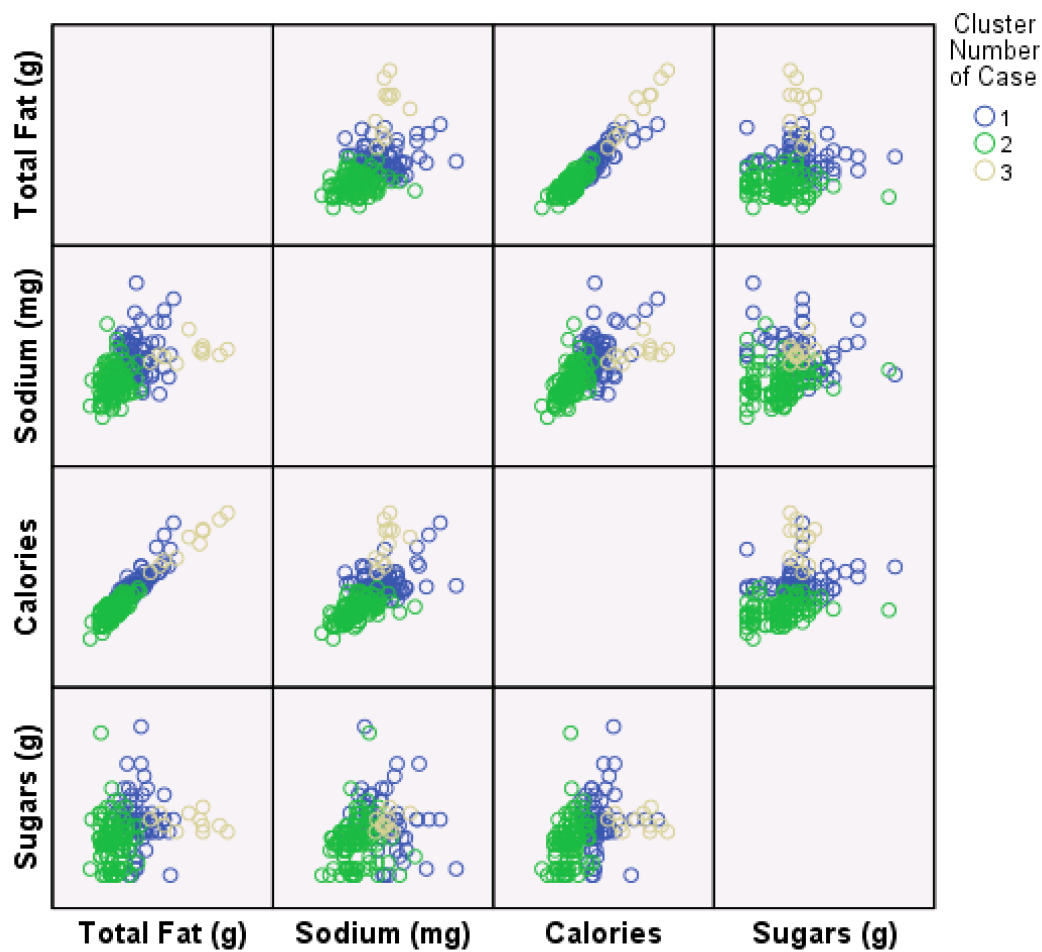
GRAPH

```
/SCATTERPLOT(BIVAR)=Dis2_1 WITH Dis1_1 BY QCL_2  
/MISSING=LISTWISE.
```

→ Graph



I also created a matrix scatterplot of a few of the variables and the cluster memberships.



Then, I wanted to grasp a better understanding of the overall nutritional value of each cluster so I generated a table of means

Report											
Cluster Number of Case		Calories	Total Fat (g)	Sodium (mg)	Carbohydrate s (g)	Fiber (g)	Sugars (g)	Protein (g)	Saturated Fat (g)	Trans Fat (g)	Cholesterol (mg)
1	Mean	651.97	34.04	1496.62	52.94	3.49	8.55	33.58	10.310	.873	91.76
	N	71	71	71	71	71	71	71	71	71	71
	Std. Deviation	128.781	8.873	420.815	16.973	1.698	4.167	5.879	3.9590	.7593	21.976
2	Mean	391.94	18.13	970.65	36.19	2.14	5.38	21.02	5.222	.319	51.67
	N	108	108	108	108	108	108	108	108	108	108
	Std. Deviation	89.989	6.310	298.458	11.462	1.329	3.580	6.696	2.2921	.5136	20.336
3	Mean	1017.27	68.00	1580.00	43.18	2.18	8.55	55.91	27.636	3.364	204.09
	N	11	11	11	11	11	11	11	11	11	11
	Std. Deviation	178.835	16.162	179.053	2.136	.603	1.368	11.344	6.2012	.8090	47.844
Total	Mean	525.32	26.96	1202.47	42.86	2.65	6.75	27.73	8.421	.703	75.47
	N	190	190	190	190	190	190	190	190	190	190
	Std. Deviation	206.630	15.065	434.996	15.649	1.586	4.033	11.393	6.2842	.9510	43.822

Finally, I created some cross tabs to show the break down of each cluster

Cluster Number of Case * Restaurant Crosstabulation

Count

		Restaurant						Total	
		Arby's	Burger King	Chick-fil-a	Dairy Queen	McDonald's	Sonic		Wendy's
Cluster Number of Case	1	14	9	3	7	14	16	8	71
	2	15	22	7	10	24	15	15	108
	3	0	1	0	1	1	6	2	11
Total		29	32	10	18	39	37	25	190

Cluster Number of Case * Meat Crosstabulation

Count

	Meat							Total
	Beef	Chicken	Fish	Ham	Other	Turkey	Veggie	
Cluster Number of Case 1	30	34	1	2	0	4	0	71
2	39	54	4	3	2	2	4	108
3	11	0	0	0	0	0	0	11
Total	80	88	5	5	2	6	4	190