1. (5 points) Evaluate:
Homogeneity of variance / covariance matrices across groups.
Multicollinearity among the predictors

I conducted a discriminate analysis to predict group membership based on how people answered 6 questions. Box's M=43.172, p=.457, which is not significant so there is no evidence that the homogeneity of variance assumption was violated.

I then conducted a multiple regression analysis to look for any issues with multicollinearity. As you can see on the coefficient table below, none of the survey questions seem to have a VIF higher than 1.5, and the condition index is 19.006, indicating there are no issues with multicollinearity.

Test Results

Box's M		43.172
F	Approx.	1.008
	df1	42
	df2	508107.438
	Sig.	.457

Tests null hypothesis of equal population covariance matrices.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	1.139	.230		4.957	.000		
	I prefer bitter over sweet flavors.	-.075	.034	-.094	-2.205	.028	.976	1.025
	I enjoy the warm beverages.	-.170	.035	-.212	-4.903	.000	.943	1.060
	I almost always get 8 hours of sleep at night.	.129	.033	.168	3.859	.000	.935	1.070
	I dislike sweet foods and drinks.	-.137	.034	-.173	-3.980	.000	.933	1.072
	I feel refreshed waking up in the morning.	.169	.034	.217	4.947	.000	.921	1.085
	I love dogs.	.064	.034	.080	1.874	.062	.972	1.029

a. Dependent Variable: Caffeine consumption method

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	I prefer bitter over sweet flavors.	I enjoy the warm beverages.	I almost always get 8 hours of sleep at night.	I dislike sweet foods and drinks.	I feel refreshed waking up in the morning.	I love dogs.
1	1	6.495	1.000	.00	.00	.00	.00	.00	.00	.00
	2	.145	6.686	.00	.05	.14	.13	.14	.14	.06
	3	.098	8.133	.00	.44	.10	.04	.01	.02	.37
	4	.090	8.481	.00	.23	.21	.28	.00	.01	.34
	5	.078	9.135	.00	.07	.02	.27	.18	.61	.06
	6	.076	9.258	.00	.05	.34	.17	.58	.11	.00
	7	.018	19.006	1.00	.16	.19	.10	.09	.11	.17

a. Dependent Variable: Caffeine consumption method

2. (10 points) Run the analysis. Use "Compute from Group Sizes" as your prior probabilities option. Please determine how many discriminant functions are retained and each function's predictive strength using:
   Canonical correlations
   Significance tests on the functions

Based on the Wilks' Lambda statistic, both functions 1 and 2 are significant, with function 2 adding to the predictive power of the analysis. Therefore, I will retain both functions. When I review the canonical correlations, I can see that function one predicts 23.52% of the variance($.485^2$) and function two predicts 10.56% of the variance ($.325^2$) over and beyond function one.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.308[a]	72.3	72.3	.485
2	.118[a]	27.7	100.0	.325

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	.684	165.031	12	.000
2	.895	48.425	5	.000

3. (10 points) Interpret each discriminant function, including evaluations of:
a. Discriminant loadings (Note: In "real life" you may be expected to analyze weights as well, however we only want to look at loadings for this assignment)

a. When I review the rotated structure matrix, I see that "I almost always get 8 hours of sleep at night", "I feel refreshed waking up in the morning", and "I love dogs" load together on function 1. On function 2, "I enjoy the warm beverages", "I dislike sweet foods and drinks", and "I prefer bitter over sweet flavors" all load together. With the exception of the dog question, these items intuitively make sense for why they clustered together. I'm going to leave the dog question in the analysis, however, because there doesn't seem to be any collinearity with function 2, so perhaps dog people do fit appropriately with function 1. Function 1 seems to best be labeled as "people who are well-rested" and function 2 seems to best be labeled as "people who like warm, bitter things".

Rotated Structure Matrix

	Function	
	1	2
I almost always get 8 hours of sleep at night.	.758*	.056
I feel refreshed waking up in the morning.	.614*	-.201
I love dogs.	.439*	.040
I enjoy the warm beverages.	.007	.756*
I dislike sweet foods and drinks.	.049	.696*
I prefer bitter over sweet flavors.	-.151	.187*

Rotated pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

b. Territorial Map (Note: You do not need to paste your territorial map into your word document. Simply explain your interpretation clearly and send your output file)

b. When I look at the territorial map (as well as the group centroid table), function 1 seems to be separating people who need caffeine from people who do not need caffeine. Function 2 seems to be separating people who drink coffee from people who do not drink coffee.

Functions at Group Centroids

Caffeine consumption method	Function 1	2
Coffee drinker	-.281	.736
Energy drink drinker	-.394	-.246
Feels no need for caffeine	.609	-.325

Unstandardized canonical discriminant functions evaluated at group means

c. Univariate ANOVA's with post-hoc tests (do these only for discriminant functions and NOT the predictors)

c. I conducted a one-way ANOVA looking at the mean differences between each discriminant function, and the ANOVA summary table (below) indicates both functions are significant (p<.001). I used a post hoc test (Bonferroni) to look closer at the mean differences of each group on each function, and there are significant mean differences between people who don't need caffeine and coffee drinkers, as well as people who don't need caffeine and energy drink drinkers on function 1. On function 2, there are significant mean differences between people who drink coffee and people who drink energy drinks, as well as people who drink coffee and people who don't need caffeine.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Discriminant Scores from Function 1 for Analysis 1	Between Groups	93.114	2	46.557	46.557	.000
	Within Groups	437.000	437	1.000		
	Total	530.114	439			
Discriminant Scores from Function 2 for Analysis 1	Between Groups	92.928	2	46.464	46.464	.000
	Within Groups	437.000	437	1.000		
	Total	529.928	439			

Multiple Comparisons

Bonferroni

Dependent Variable	(I) Caffeine consumption method	(J) Caffeine consumption method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Discriminant Scores from Function 1 for Analysis 1	Coffee drinker	Energy drink drinker	.11354911	.12024639	1.000	-.1754309	.4025291
		Feels no need for caffeine	-.88915182*	.12008076	.000	-1.1777338	-.6005698
	Energy drink drinker	Coffee drinker	-.11354911	.12024639	1.000	-.4025291	.1754309
		Feels no need for caffeine	-1.00270093*	.11233175	.000	-1.2726602	-.7327416
	Feels no need for caffeine	Coffee drinker	.88915182*	.12008076	.000	.6005698	1.1777338
		Energy drink drinker	1.00270093*	.11233175	.000	.7327416	1.2726602
Discriminant Scores from Function 2 for Analysis 1	Coffee drinker	Energy drink drinker	.98182721*	.12024639	.000	.6928472	1.2708072
		Feels no need for caffeine	1.06058702*	.12008076	.000	.7720050	1.3491690
	Energy drink drinker	Coffee drinker	-.98182721*	.12024639	.000	-1.2708072	-.6928472
		Feels no need for caffeine	.07875981	.11233175	1.000	-.1911995	.3487191
	Feels no need for caffeine	Coffee drinker	-1.06058702*	.12008076	.000	-1.3491690	-.7720050
		Energy drink drinker	-.07875981	.11233175	1.000	-.3487191	.1911995

*. The mean difference is significant at the 0.05 level.

d. Please write a conclusion explaining what the information in questions a through c tell you about the relationship between the predictors and group membership

e. Based on the results of each statistical test, there does seem to be support for my initial conclusion: one set of predictors relates to caffeine preference, and that function differentiates between people who need caffeine and people who do not need caffeine. The other set of predictors relates to coffee preference, and that function differentiates between people who drink coffee and people who do not drink coffee.

4. (5 points). Discuss prediction accuracy using the classification tables. How well would we do classifying the cases on the basis of the functions? Does this classification do better than chance (use the proportional chance criterion and cross-validated hit rate to answer this question, and make sure you are explicit about what base rate / cutoff you used for this criterion)?

To understand how well we will do classifying group membership based on these two functions, I looked at the cross-validated hit rate and calculated the proportional chance criterion to get a base rate for comparison.

The cross-validated hit rate is 55.7%. To get the base rate, I did the following for each group:
N= 440
Coffee drinkers: n=123, $(123/440)^2 = .0781$
Energy drink drinkers: n=158, $(158/440)^2 = .1289$
No caffeine needed: n=159, $(159/440)^2 = .1306$
.0781+.1289+.1306 = .3376
Base rate = .3376, hit rate = .3376*1.25 = .422

The cutoff to ensure we are doing better than chance is 42.2%, and we are well above that. Therefore, we will be able to reliably use these functions to predict group membership.

Classification Results[a,c]

		Caffeine consumption method	Predicted Group Membership: Coffee drinker	Energy drink drinker	Feels no need for caffeine	Total
Original	Count	Coffee drinker	62	32	29	123
		Energy drink drinker	23	83	52	158
		Feels no need for caffeine	15	34	110	159
	%	Coffee drinker	50.4	26.0	23.6	100.0
		Energy drink drinker	14.6	52.5	32.9	100.0
		Feels no need for caffeine	9.4	21.4	69.2	100.0
Cross-validated[b]	Count	Coffee drinker	62	32	29	123
		Energy drink drinker	23	79	56	158
		Feels no need for caffeine	19	36	104	159
	%	Coffee drinker	50.4	26.0	23.6	100.0
		Energy drink drinker	14.6	50.0	35.4	100.0

	Feels no need for caffeine	11.9	22.6	65.4	100.0

a. 58.0% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 55.7% of cross-validated grouped cases correctly classified.