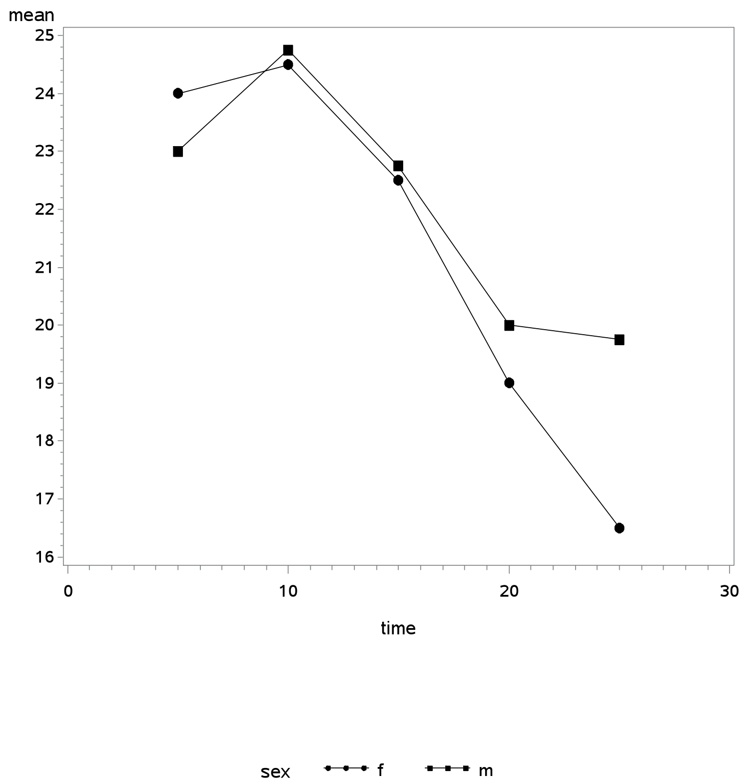
1. Eight individuals were given a drug to measure the effect of a soporific effect on alertness. They were measured 5, 10, 15, 20, and 25 minutes afterward. At each time point, each person's response time (in hundredths of a second) to a stimulus was recorded. Each person's sex was also recorded for consideration. Output for the code below is available in "ex2_drug.pdf".

```
proc glm data=drugs;
 class sex;
 model t1-t5 = sex;
 manova h=sex m=t2-t1,t3-t2,t4-t3,t5-t4;
 manova h=sex m=t1+t2+t3+t4+t5;
 run; quit;
```

   (a) With time on the horizontal axis and response value on the vertical axis, the profile plot below compares the two sexes. In words, what does interaction mean in this situation? If there were no interaction present, what would the profile plot above be expected to look like?



   Interaction means the effect of sex is not the same for all time points. Equivalently, the effect of time is not the same for both sexes. If there were no interaction, the profile plot would be expected to show parallel lines.

   (b) Use the output to carry out the test of interaction. State the test statistic and conclusion with $\alpha = .05$.

   The MANOVA model is $\mathbf{Y}_{ij} = \mu_i + \epsilon_{ij}$, where $\mathbf{Y}_{ij}$ is the vector of time responses for the jth person of the ith sex, μ_i is the mean vector for the ith sex, and ϵ_{ij} is multivariate normal with mean vector 0 and covariance matrix Σ (for each i and j). The test of H_0 : no interaction versus H_a : interaction has test statistic $F = 0.12$ with p-value 0.9673. So, we do not have significant evidence of interaction in this situation.

(c) Is there an effect due to sex? Use $\alpha = .05$ to conduct the appropriate test, based on the output. Also, comment on the appropriateness of this test in light of your results above in the test for interaction.

They hypotheses here are H_0 : no sex effect (common μ_i) versus H_a : sex effect (at least two μ_i different). The test statistic is $F = 0.05$ with p-value 0.8225. So, there is not significant evidence of a sex effect.

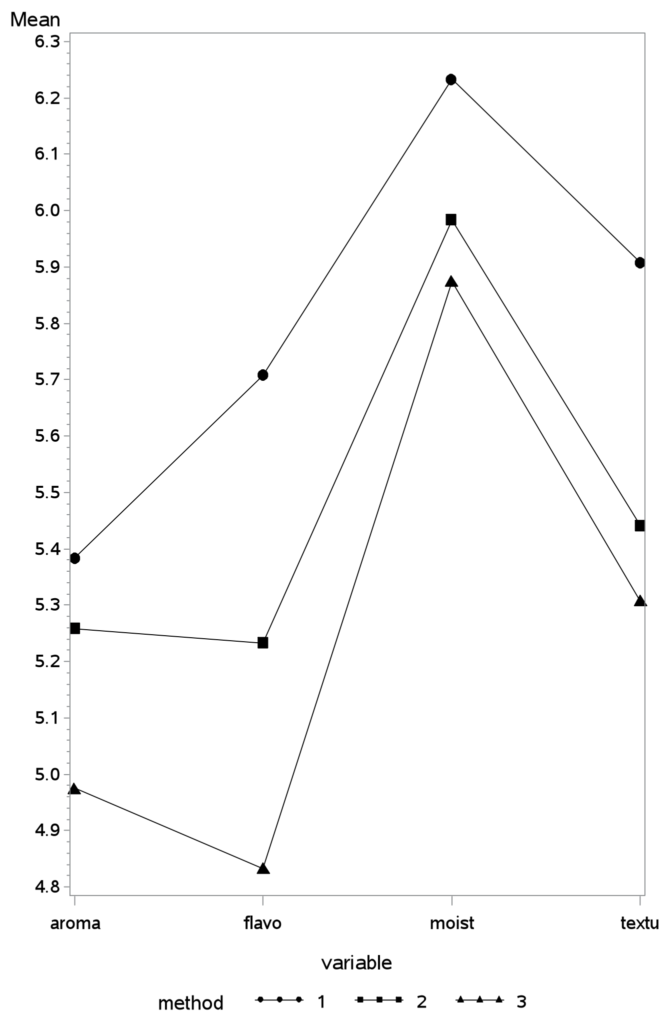
The test for sex effect here is appropriate because the interaction test was insignificant. If the test for interaction had been significant, the sex effect would depend on the time point, and combining time points (as the test here does) would not be appropriate.

2. In a study comparing attributes of fish preparation, three methods were compared. Variables recorded were aroma score, flavor score, texture score, moisture score. Twelve fish were prepared under each method.

   (a) State the MANOVA model for this situation. Specifically, what does Y_{ijk} represent in this situation?

   Y_{ijk} is the score for the ith method, jth fish, and kth variable. The MANOVA model states that the $[Y_{ij1}, \ldots, Y_{ij4}]'$ are independent multivariate normal with mean vector μ_i and covariance matrix Σ.

   (b) The profile plot below is for the method mean scores with attributes along the horizontal axis. Use it to comment on the evidence for a method effect. How would the lines be expected to look under the assumption of no method effect?



If there were no method effect, the scores would be the same for all methods. As it is, the first method seems to have the highest scores, and the third method seems to have to lowest. This is evidence of a method effect.

Output for the following code is available in the separate file "ex2_fish.pdf". Use it to answer parts c) and d).

```
proc discrim pool=test data=fish; *passes with small samples;
  class method;
  var aroma flavor texture moisture;
  run;
proc glm data=fish;
  class method;
  model aroma flavor texture moisture = method;
  manova h=method;
  run; quit;
```

(c) Are the methods significantly different for any attributes? Answer with the appropriate hypothesis test(s). State the hypotheses, test statistic(s), p-value(s), and conclusion(s) at the .05 level of significance.

The hypotheses are $H_0 : \mu_1 = \mu_2 = \mu_3$ versus H_a : at least one difference. The F test statistic is 4.24 with p-value .00044. This is significant evidence that the method mean scores are not all equal.

Consideration of the individual variables allows for a more specific follow-up. With a Bonferroni correction of 4 (number of individual ANOVA tests considered), we find that only the flavor variable exhibits a significant method effect (p-value 0.0006).

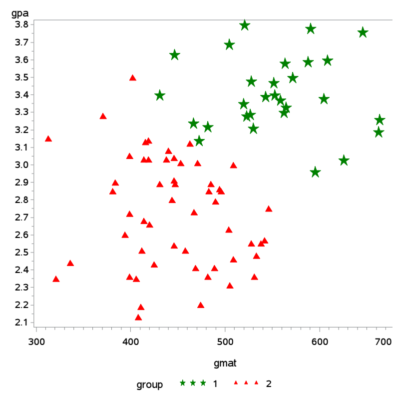
(d) What assumption does the MANOVA model make about $Cov(Y_{ijk})$? Is it reasonable here? Justify your answer from the output.

The MANOVA model assumes that $\Sigma = Cov([Y_{ij1}, \ldots, Y_{ij4}]')$ for all i and j. The results of this test are given on p.2 of the output. The p-value of 0.8988 do not suggest strongly against this assumption.

3. Information on past students who have applied to a graduate program is used to study future acceptances. Variables are college GPA, GMAT score, and whether the student was accepted (1 = yes, 2 = no). Output for the following SAS code is available in the file "ex2_grad.pdf".

```
proc discrim data=grad pool=yes;
  class group;
  var gpa gmat;
  priors '1'=1 '2'=1;
  run;
```

(a) The plot below is of GPA versus GMAT, using separate symbols for the acceptance status. Comment on the appropriateness of these variables as discriminators of admittance.



These two variables appear to distinguish the groups well in that knowing the values of GPA and GMAT tend to indicate which group the student belongs to (whether admitted or not). Moreover, they appear to be roughly elliptical in shape with similar variation, which are consistent with the assumptions of discriminant analysis.

(b) Use the discriminant analysis for classifying an applicant based on GPA and GMAT score, and apply the results to the existing observations to classify them as either accepted or rejected. What fraction of them are classified correctly?

The results of this (resubstitution) method are on p.3. Only one of each group was incorrectly classified, so the correct fractions are 30/31 and 53/54, respectively.

(c) From this analysis, what is the posterior probability of accepting a student with GPA and GMAT score of 2.9 and 550, respectively?

Applying the linear discriminate function coefficients from p.2 to these variable values gives

$$d_1 = -134.68959 + 48.53955(2.9) + .18312(550) = 106.7911$$
$$d_2 = -86.30887 + 38.88024(2.9) + 0.14588(550) = 106.6778$$

So, the posterior probability of admittance is

$$\frac{\exp(106.7911)}{\exp(106.7911) + \exp(106.6778)} = 0.5283$$

(d) Before consideration of a student's GPA and GMAT scores, what assumption is made about a student's probability of acceptance for this analysis?

Before consideration of the data, any assumption about group classification is specified in the prior probabilities. From the SAS code, these are equal, which means a student is equally assumed to be admitted as to be rejected.