STAT505 Assessment #9

1. J&W Exercise 8.13. The data are given in "radiotherapy.dat".

2. The data in "track.dat" give the men's national track records for 55 countries in 1984 for the following distances: 100m, 200m, 400m, 800m, 1500m, 5000m, 10000m, and marathon. The first three variables are measured in seconds, while the remaining variables are measured in minutes. Express the values of the first three variables in minutes by dividing each number by 60. This can be accomplished using the following data step:

```
data track;
   infile "v:\track.dat";
   input d100 d200 d400 d800 d1500 d5000 d10000 marathon country $;
   d100=d100/60;
   d200=d200/60;
   d400=d400/60;
   run;
```

   (a) Perform a principal component analysis using the covariance matrix; that is, using the raw data expressed in minutes. Include a scatter plot of the first two principal components.

      i. How many principal components are required to explain 90% of the total variation for this data?

      ii. For the number of components in part i, give the formula for each component and a brief interpretation.

      iii. What countries have the highest and lowest values for each principal component (only include the number of components specified in part i). For each of those countries, give the principal component scores.

   (b) Perform a principal component analysis using the correlation matrix. Include a scatter plot of the first two principal components.

      i. How many principal components are required to explain 90% of the total variation for this data?

      ii. For the number of components in part i, give the formula for each component and a brief interpretation.

      iii. What countries have the highest and lowest values for each principal component (only include the number of components specified in part i). For each of those countries, give the principal component scores.

   (c) Compare the results from parts (a) and (b). Which gives the best interpretation of the data?