Homework 11

In Problems 1-4, you will be analyzing a data set on the performance of mutual funds. The data is stored in the file "Funds.MTP". The funds in the data set are a selection of the largest stock mutual funds available to the general public. They are classified by type: (1) Balanced [aiming for a balance of stocks and bonds, to keep the risk relatively low], (2) Equity-Income [dividend-paying stocks of well-established companies] (3) Growth-and-Income [aiming for growth in share price and income from dividends], (4) Growth [stocks with potential for rapid growth], (5) Aggressive-Growth [Riskier version of Growth], (6) Small-Company [small young companies with potential for rapid growth], (7) International and Global [International funds invest in stocks outside of the U.S., while Global funds invest in both foreign and U.S. stocks].

Performance of the funds is measured in two ways. FiveYR gives the five year performance of the funds (from 1988-1992), measured as the amount an investor would have accumulated by investing 2000 Dollars in the fund at the beginning of each year. Any front-end sales charges and redemption fees have been deducted from the dollar figure. The variable Return92 gives the fund’s return for 1992, (in percent) rounded to the nearest integer. One might wonder whether there is a difference in performance based on the fund type. We will focus only on the five year performance.

1) Use Graph → Boxplot, Y: FiveYR, X: Type, to get multiple boxplots of the five year returns.

A) What differences in performance do you see from the boxplots? Which fund did best, and which did worst?

B) Based on the boxplots, and your knowledge of the stock market, do you think it is reasonable to assume that the population distributions for the groups are normal, that the population variances are equal, and that the observations are independent? Justify your answers. (Pay particular attention to the question about independence. Note that a given stock may be used by more than one fund. Furthermore, don’t forget that most stocks have a tendency to move with the overall market trend.)
C) Accordingly, do the assumptions needed for a one-way analysis of variance seem to be satisfied? (Even if the answer is no, we will proceed anyway, for illustrative purposes).

2) Run the one way ANOVA, using Stat → ANOVA → Oneway, Response: FiveYR, Factor: Type, Comparisons → Tukey’s (5), Fisher’s (5).

A) According to the $F$-test, is there a significant difference in the average performances of the funds, at level .05? At level .01? At level .001?

B) Do Bartlett’s and Levene’s tests indicate any problems with the assumptions needed for the ANOVA? To run these tests, use Stat → ANOVA → Homogeneity of Variance.

C) Get the value of $s^2$ from the ANOVA output. (You will find it somewhere in the ANOVA table, but it won’t be called $s^2$.)

D) Use the ANOVA output to calculate $s_y$, the sample standard deviation of the five year returns for all funds taken together. Which is larger, $s_y$ or $s$? Does the answer make sense?

E) What was the average performance of the best group of funds? What was the average for the worst?

3) Consider Fisher’s (LSD) pairwise comparisons. Note that for this data set, LSD finds that the mean for the international funds is significantly different from the mean for any other group, since none of the confidence intervals for Group 7 includes zero. Also, there is a significant difference between (Growth, Equity-Income), even though there is no significant difference between the pair (Growth, Growth & Income), or between the pair (Growth & Income, Equity-Income). The "Critical value" is $t_{w/2; n-g} = t_{.05; .178} = 1.973$, as is appropriate for LSD comparisons.

A) Find the 95% LSD confidence interval for the difference in means between the Balanced fund and the Growth fund.

B) Do the means in part A) differ significantly, according to LSD, at level .05?
4) Consider Tukey’s pairwise comparisons.

    A) Find the 95% Tukey confidence interval for the difference in means between the Balanced fund and the Growth fund.

    B) Do the means in part A) differ significantly, according to Tukey, at level .05?

    C) Explain briefly why the answers to 3B) and 4B) might be different.

5) In the one-way ANOVA model, prove that $z_{g+1}, \ldots, z_n$ are iid $N(0, \sigma^2)$, even if the group population means are unequal.

6) In the one-way ANOVA model with $g = 2$, use the formula on top of page 2, Handout 22, and the formula for $s^2$ on page 12 of Handout 21, to show that a level $1 - \alpha$ confidence interval for $\mu_1 - \mu_2$ is given by

   $$ ar{y}_1 - \bar{y}_2 \pm t_{\alpha/2, n-2} \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 n_2} / (n_1 n_2)} . $$

   (The corresponding hypothesis test for $\mu_1 - \mu_2$ is called the two-sample $t$-test.)