Statistics Project

Problem Statement 1:

You were recently hired as a business analyst in a top sports company. The senior management team has asked you to come up with metrics with which they can gauge which team will win the upcoming La Liga Cup (Football tournament). The given data set contains information on all the teams that have so far participated in all the past tournaments. It has data about how many goals each team scored, conceded; how many times they came within the first 6 positions, how many seasons they have qualified, their best position in the past, etc.

Before doing any analysis it would be a good idea to check for any hyphens or other symbols in the data set and make appropriate replacements to make sure you can perform arithmetic operations on the data. Prepare a short report to answer the following questions:

1. Which are the teams which started playing between 1930 and 1980?
2. Which are the top 5 teams in terms of points?
3. What is the distribution of the winning percentage for all teams? Which teams are in the top 5 in terms of winning percentage? (Winning percentage= (GamesWon / GamesPlayed)*100)
4. Is there a significant difference in the winning percentage for teams which have attained the best position between 1-3 and those teams which have had the best position between 4-7?

Problem Statement 2:

A study was done to measure the blood pressure of 60-year-old women with glaucoma. A random sample of 200 60-year-old women with glaucoma was chosen. The mean of the systolic blood pressure in the sample was 140 mm Hg and the standard deviation was 25 mm Hg. Prepare a short report to answer the following questions:

1. Calculate the estimated standard error of the sample mean. What does the standard error indicate?
2. Estimate a 95% confidence interval for the true mean blood pressure for all 60-year-old women with glaucoma.
3. Assume that instead of 200, a random sample of only 100 60-year-old women with glaucoma was chosen. The sample mean and standard deviation estimates are the same as those in the original study. What is the estimated 95% confidence interval for the true mean blood pressure?
4. Which of the two above intervals is wider?
5. Explain in non-technical terms why the estimated standard error of a sample mean tends to decrease with an increase in sample size.

Problem Statement 3:

Par Inc., is a major manufacturer of golf equipment. Management believes that Par's market share could be increased with the introduction of a cut-resistant, longer-lasting golf ball. Therefore, the research group at Par has been investigating a new golf ball coating designed to resist cuts and provide a more durable ball. The tests with the coating have been promising.

One of the researchers voiced concern about the effect of the new coating on driving distances. Par would like the new cut-resistant ball to offer driving distances comparable to those of the current-model golf ball. To compare the driving distances for the two balls, 40 balls of both the new and current models were subjected to distance tests. The testing was performed with a mechanical hitting machine so that any difference between the mean distances for the two models could be attributed to a difference in the design. The results of the tests, with distances measured to the nearest yard, are contained in the data set "Golf". Prepare a short report to answer the following questions:

1. Formulate and present the rationale for a hypothesis test that Par could use to compare the driving distances of the current and new golf balls.
2. Analyze the data to provide the hypothesis testing conclusion. What is the p-value for your test? What is your recommendation for Par Inc.?
3. What is the 95% confidence interval for the population mean of each model, and what is the 95% confidence interval for the difference between the means of the two populations?