**Using regression based on real data to examine the distribution of rmse .**

The data set examsub2 is in the s5066/Nhanes3 subdirectory. **Use 54321 as the seed for all random number streams**.

1. Use a PROC REG step with the data set examsub2 to regress fvc (dependent variable) on bmpht (independent variable). Output the estimated coefficients to a data set, betas. Use a PROC PRINT step to examine the contents of the data set betas.

Correct Answer:



1. Use a DATA \_null\_ step to read the data set betas and create three macro variables: beta0 (the intercept), beta1 (the slope), and rmse (the root mean square error). Use a %let statement to examine the values stored in the macro variables.

Correct Answer (from log):



1. Use a PROC MEANS step to find the mean and standard deviation of the variable bmpht (height in centimeters). Use and output statement to create a data set, distnvalues. The data set distnvalues should contain the variables meanht (the average height) and sdht (the standard deviation of height). Use a PROC PRINT step to examine the dataset distnvalues.

Correct Answer:



1. Use a DATA \_null\_ step to read the data set distnvalues and create two macro variables: mu (the average height) and std (the standard deviation of height). Use a %let statement to examine the values of the macro variables mu and std.

Correct Answer (from log):



1. Use a DATA step to generate 10,000 samples of size 100. Each observation should contain three variables: rep (the sample number), ht (a random observation from a normal distribution with mean and standard deviation defined by the macro variables mu and std), fvc (a random observation, the linear part is defined by the variable ht and the macro variables beta0 and beta1. To this you should add random noise from a normal distribution with mean 0 and standard deviation defined by the macro variable rmse).
2. Use a PROC REG step with the noprint option and with a by variable to estimate the regression of fvc on ht separately for each of the 10,000 samples. Output the estimated coefficients and the root mean square error to a file, RegOut. Use a PROC PRINT step to display the first 10 observations on the data set RegOut.

Correct Answer:



1. Use a PROC UNIVARIATE step to display a histogram of the approximate sampling distribution of the root mean square error. The histogram should have a normal distribution overlayed. Use and ODS statement to display only the histogram and the goodness of fit tests.

Correct Answer (Partial)

