*Multiple Regression using SPSS for Windows 11.0*

- Open the **bpdat.sav** data set.

- We're going to work with the following variables. The criterion variable (dependent variable) will be **digspan1** (digit span scores at time 1). The potential predictor variables we'll be examining are **age**, **gender**, **traitan1**, **diabp1**, and **sysbp1**

Using the Enter method for specifying a set of predictor variables

- Click **Analyze**
- Click **Regression**
- Click **Linear**
- Let's say that we want to maximize the predictive power of the regression equation by entering all of the predictor variables. Move the variable **digspan1** over to the **Dependent** box. Move the variables **age**, **gender**, **traitan1**, **sysbp1**, and **diabp1** into the **Independent(s)** box.
- Make sure that the option just to the right of **Method** says **Enter**. If it says anything other than that, click on the down arrow, find the term **Enter**, and click on it. **Enter** is the term that SPSS uses for the method where the researcher specifies the variables that will go into the regression equation and the stage at which they go in.
- Under **Statistics** click on the options for **Descriptives** and **R Squared Change**. Click **Continue** to return to the main regression menu.
- Click **OK** to generate the regression output.

- The output is presented in the same format as the one that you got when there was only one predictor variable.

  - First, SPSS shows you the predictor variables that were entered or removed from the model at each step. In this case there was just one step in which all five predictors were entered at once.

  - Second, SPSS gives you a **Model Summary** in which you get a sense of how good a job the regression equation is doing at each step. In this case, the R Square is .178, which indicates the predictor variables account for 17.8% of the variability in the criterion. You get the standard error of estimate at this point and you also get the statistic **R Square Change** and an F-test to go with it. This part of the output is telling you about the contribution of the most recent set of predictors added to the model. Because there was only one step here the **R Square Change** represents the increases in the proportion of variability accounted for in going from no predictors to the five predictors we entered in that one step. The **F Change** is testing whether that most recent contribution represents a significant improvement is the predictive power of the regression equation.

- The ANOVA section of the output contains an F-test that tells you whether the regression equation accounts for a significant amount of variability in the criterion variable.  In this case, the significance level is .054, which tells us that the regression equation was not quite good enough for us to be able to say that this set of five predictors accounts for a significant amount of variability in the digit span scores.  In other words, the regression equation is not quite significantly better than using no regression equation at all.

- Now let's see if SPSS can generate a regression model using one of it's automated strategies (Forward, Backward, Stepwise) that we can say is statistically significant.

Algorithms for generating regression equations

SPSS offers a number of strategies for helping the researcher to generate an efficient regression equation.  By efficient, we could say that it's the regression equation that accounts for greatest proportion of variability using the fewest number of predictor variables.

- To use one of these strategies, specify the predictor variables you want to start with.  In this case, use the same five predictor variables that we used above.  Now, for example, change the method to **Stepwise**.  Make sure the R Square Change option is still selected.  Click **OK**.

- The first thing you see in the output is a list of the steps that SPSS went through in generating the final version of it's regression equation.  It selected **sysbp1** to enter the model first and then, in a second step, it added **gender** to the model.  Apparently, at no step in the process did SPSS remove a variable from the model.  If it had, it would have listed this in the **Variables Removed** column.
- Moving down to the next part of the output, there are two rows in the **Model Summary** table.  This is because variables were added to the regression equation in two separate steps.  The values in the two rows show you how the regression equation was doing at each of these two steps.  For example, when just the one predictor variable **sysbp1** was in the equation the **R Square** was .08.  When there were two predictor variables in the equation the R Square increased to .146.
- The ANOVA section of the output also contains two separate ANOVA tables.  The ANOVA table for model 1 tests whether the one predictor variable **sysbp1** accounts for a significant amount of variability all by itself (it does, p=.029).  The ANOVA table for model 2 is testing whether the regression equation with two predictor variables is accounting for a significant amount of variability (it does, p=.011).
- The **Coefficients** section of the output shows you the values for the regression coefficients at each step in generating the  regression equation.  Look in the

column for **Unstandardized Coefficients**. In model 1 the regression equation is **Predicted digspan1 = 4.391 * (.01765)(sysbp1)**. The E-02 part of the value for the slope means that you're supposed to move the decimal point over to the left by two numbers. In model 2 the regression equation is **Predicted digspan1 = 5.284 + (.01719)(sysbp1) – (.549)(gender)**. These are the regression equations using unstandardized coefficients (the variables are in raw score units, rather than in standard score units).

- The last section of the output (**Excluded Variables**) gives you information about how SPSS selected the next variable to add to the model. After **sysbp1** was selected first there were four remaining variables to consider. The one that would contribute the most is **gender**. The significance level for its unique contribution is .041, which is better than the other three. After **gender** has been added none of the remaining three variables contribute a significant amount of variability to the model. All of the significance levels are greater than .05. So that's why SPSS stopped after entering **gender**.

<u>Determining the unique contribution of a particular predictor variable</u>

You can use **hierarchical regression** to determine whether a predictor variable contributes a significant amount of variability to a regression equation. To do this you must be in a position to control the order in which variables are entered into the regression. The **Enter** method in SPSS gives you the flexibility to say which variable(s) you want to enter in a first block (set) of variables and which variable you want to enter in a second block.

Let's say that you want to determine if **sysbp1** contributes a significant amount to a regression equation predicting scores for **digspan1** after **gender** has already been entered.

- Clear any predictor variables that are there out of the **Independent(s)** list in the regression window.
- To answer this kind of question you can't just enter both predictor variables at the same time. You've got to enter them one at a time. You've got to take the variable that you're testing in terms of its unique contribution and enter it all by itself <u>after</u> you've entered the other predictor variable(s) being considered. So it this case we've got to enter **gender** first and then **sysbp1** second. This makes sense because we're trying to see if **sysbp1** accounts for a significant amount of variability above and beyond what **gender** can.
- Make sure that **digspan1** is selected as the dependent variable.
- Select gender first and move it over into the **Independent(s)** box. Make sure it's the only variable listed at this step. You'll notice that it says **Block 1 of 1** just above the **Independent(s)** box.
- Now click on the **Next** button. Now it says **Block 2 of 2**.
- Move the variable **sysbp1** into the **Independent(s)** box.
- Make sure that the **R Square Change** option is selected.
- Click **OK**

- The output looks about the same as what we got for the Stepwise example above.
- The **Variables Entered/Removed** box tells you which variable was entered first and which variable was entered second.
- The **Model Summary** table contains the results that we need for this type of question. Notice that the **R Square** is .070 when the only predictor variable in the model is **gender**. In the bottom row, notice that the **R Square** goes up to .146 when both **gender** and **sysbp1** are entered. Over on the right hand side of the table, look in the **R Square Change** column. It's .070 for the first equation, which means that **gender** accounts for 7% of the variability when it's entered all by itself. When you add **sysbp1** in the second step you have an R Square Change of .076. This means that **sysbp1** accounts for an additional 7.6% of the variability, above and beyond what **gender** can do all by itself. The **F Change** is significant, which indicates that this unique contribution of sysbp1 is significant.
- To determine the unique contribution of **gender**, you just reverse the order of entry into the regression equation.

<u>Exercises</u>

1. Enter the variables **traitan1**, **sysbp1**, **diabp1**, **age**, and **gender** into a multiple regression model predicting scores for **digsym1**. What proportion of variability is accounted for? What is the regression equation using unstandardized coefficients? Does the model account for a significant amount of variability? Why do you think so?
2. Use the Stepwise method to determine regression equation when starting with the same predictor variables listed in Part 1. Please describe the steps SPSS went through in generating its regression equation. At each stage of the process list (a) the variable that was entered or removed from the equation (b) that variable's unique contribution, and (c) the R Square for the regression equation up to that point. Report the final version of the regression equation. What proportion of variability is accounted for by the final version of the regression equation.
3. Repeat Part 2, except use the Backward method. Is the solution different from the one you got using the Stepwise method?
4. Does **age** account for a significant amount of variability above and beyond that of **diabp1**? What is the unique contribution of **age**? Is this unique contribution significant? Why do you say this?
5. Generate the Venn diagram with **digsym1** as the criterion and **age** and **diabp1** as predictors.

***Please hand in a printout of your output window*** for this assignment. Please make sure that the **Notes** section of each section of output is visible. Also hand in a separate piece of paper with your Venn diagram. The assignment is due the Thursday after we get back from Thanksgiving Break.