

This exercise uses your **latest** IVH SAS data set, from homework 3 (or use mine, g:\shared\bio113\ivh2.sas7bdat). There are two new raw (ASCII) data sets that you need to combine with your IVH SAS data set. Use LABEL, FORMAT, and TITLE to make the output readable. **WARNING: Throughout this homework, be very careful not to use unnecessary data steps and watch out for possible missing data values when you create new variables.** Hand in the programs, logs and output.

Create a **temporary** SAS data set by inputting the data in the **raw** data file, g:\shared\bio113\idob.dat. The file contains each baby's ID variable in columns 1-6 and its date of birth in columns 7-16 (you may use the mmddyy10. INFORMAT to read the dates).

Join each baby's date of birth in the temporary SAS data set you just created to the **end** of the matching record in your large permanent IVH data set and save this latest combined data set as a new **permanent** SAS data set named **ivh3**.

Calculate each baby's age on **October 20, 2004**, in years (a year has 365.25 days). Use PROC UNIVARATE with the PLOT option (see page 71 of the handout or LSB 7.1 for correct syntax) to display the distribution of the babies' ages.

In a **one single data step**, make three **temporary** data sets; one for babies born in 1991, one for babies born in 1992 and one for babies born in 1993. In each temporary data set **keep only the variables you need** to make the following tables.

Use PROC FREQ to display a table of HOSP vs RACE in the 1991 data set, a table of CS vs PIH in the 1992 data set and a table of SINGLE vs MEDU in the 1993 data set

Create another **temporary** SAS data set by inputting the data in the second **raw** data file, g:\shared\bio113\std.dat. This file contains national standards for birth weight and T4 by gestational age and sex. The variables in order are: gestational age (completed weeks), sex (0=female and 1=male), mean birth weight, standard deviation of birth weight, mean T4, and standard deviation of T4. The values are separated by spaces so you may use list-style input.

Join this new temporary data set to the permanent SAS data set you saved above (**ivh3**), matching on gestational age and sex. Take care when you combine the permanent SAS data set and the temporary one with the standards that the resulting data set has **the same number of observations** as the original permanent one. Make the data set that results from the join **permanent**, and live dangerously, give it the same name, **ivh3**.

Create a birth weight Z-score for each baby by subtracting the mean birth weight (from the standard) from the baby's birth weight and dividing the result by the standard deviation of birth weight (also from the standard) $[(bw - \text{mean } bw) / \text{stdev } bw]$. Also calculate a T4 Z-score for each baby $[(t4 - \text{mean } t4) / \text{stdev } t4]$.

Now, read ahead to the end to see if you need to create any other variables. If so, do it now so the variables get saved into the permanent data set.

Use this data set for the following analyses.

Use PROC PLOT or PROC GPLOT to plot each of the Z-scores (y axis) vs gestational age (x axis).

Use PROC TTEST (see page 73 of the handout for syntax) to see if the birth weight Z-scores of male babies are different from those of female babies.

Use PROC NPAR1WAY (see page 74 of the handout for syntax) to see if the duration of labor is related to maternal age in three categories, < 20, 20 to 30, > 30 years).

Use PROC REG (see page 75 of the handout or LSB 7.5 for syntax) to see if the birth weight adjusted crystalloid-colloid sum on day 2 (CC2 from HW2) (dependent variable) the mean arterial blood pressure on day 2 (MAP2) (independent variable).

Create an indicator variable for black race (a baby is black if race=2). Use PROC GLM (see page 79 of the handout) to examine the relationship between T4 Z-score (dependent variable) and black race, ACS (3 levels) and duration of labor categories (variable made in HW2). Fit a model with **all** possible interactions.

Use PROC LOGISTIC (page 80 of the handout) to see if IVH is predicted by birth hospital, gestational age < 26 weeks, gestational age 26-28 weeks, a complete ACS course, diagnosis of pregnancy-induced hypertension, duration of membrane rupture 1 or more hours, Caesarian delivery, and T4 (**not** the Z-score) ≤ 5 . Notice you need to **make several n/y (0/1) indicator variables**.

Make a variable that divides the MAP on day 1 into four categories: < 28, 28 to < 32, 32 to < 37, and ≥ 37 . Use PROC LIFETEST (see page 82 of the handout) to see if hospital length of stay is related to MAP category. Censor deaths.

This is the last homework!