SAS Training 2019
Exercise 3

1. The datafile **hosp.csv** is a comma delimited file containing made up data on 4 hospital stays. The variables on the file are:

Variable Name	Description
id	An id number linked to the patient
dob	Date of birth of patient
admit	Admission date for patient
dischrg	Discharge date for patient
fee	Total cost for hospitalization

   (a) Use **proc import** to bring in the data and create a SAS dataset called hospstay. Then in a datastep add the following variables to the dataset:

   (b) Compute a variable called **staydays** which is the number of days spent in the hospital. Include both the admission and discharge data as full days.

   (c) Compute a variable called **age** that is the age of the patient at time of admission.

   (d) Create a variable called **costperday** that is the average cost per day.

   (e) Run **proc contents** on the dataset. *What is the format for the date variables?*

   (f) Display all variables using **proc print**. Change the date formats to date9. formats.

2. The SAS dataset called **valung** contains survival data for patients diagnosed with lung cancer and entered in a study. The following variables are on the SAS dataset:

Variable Name	Description
SurvTime	Follow-up time in days after enrollment
Death	Indicator variable for death (1=died, 0=alive) at end of study
CellType	Type of Cancer: 1=squamous cell, 2=large cell, 3=adenocarcinoma, 4=small cell
Age	Age in years of patient at enrollment
DiagTime	Time in months from diagnosis to entry into the study.
Treatment	Type of therapy, test=1, standard=0.
Prior	Prior therapy, 1=yes, 0=no.
KPS	Karnofsky score (0-100); higher scores indicate more functionality.

(a) Download the dataset from the class website.

(b) Run a **proc contents** on the dataset. You will need to use a libname statement to tell SAS where the dataset is located. *How many observations are on the dataset?*
Using a data-step create a new dataset from valung that includes only patients with squamous cell cancer. Add a new variable the divides KPS in to 3 levels: (1) Less than 50, (2) 50-69, and (3) 70 and higher.

(c) Use a procedure to compare the proportion who died and the survival time for patients with squamous cell cancer by the three levels of KPS.

(d) Using proc copy make the new dataset a permanent SAS dataset stored in the same folder as **valung**.