Math 150 - Methods in Biostatistics - Homework 2

your name here

Due: Wednesday, February 6, 2019, in class

```
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=3, fig.width=5,
                      fig.align = "center")
library(tidyverse)
library(broom)

tumor <- readr::read_csv("http://pages.pomona.edu/~jsh04747/courses/math150/vonHippelLindau.csv")
```

Note: there are two places to check for hints on R code. One is the class notes (http://st47s.com/Math150/Notes/, see R Examples) and the other is the R manual associated with the textbook which is on Sakai.

1. Eisenhofer et al. (1999) investigated the use of plasma normetanephrine and metanephrine for detecting pheochromocytoma in patients with von Hippel-Lindau disease and multiple endocrine neoplasia type 2. The data set (vonHippelLindau.csv) contains data from this study on 26 patients with von Hippel-Lindau disease and nine patients with multiple endocrineneoplasia. The variables in the data set are (problem from Dupont, chp 2.22, PubMed article at http://www.ncbi.nlm.nih.gov/pubmed/10369850):

disease = 0: patient has von Hippel-Lindau disease 1: patient has multiple endocrine neoplasia type 2

p_ne = plasma norepinephrine (pg/ml)

tumorvol = tumor volume (ml)

   (a) Regress plasma norepinephrine against tumor volume. Draw a scatter plot of norepinephrine against tumor volume together with the estimated linear regression curve. What is the slope estimate for this regression? What proportion of the total variation in norepinephrine levels is explained by the regression?

   (b) Experiment with different transformations of norepinephrine and tumor volume. Find transformations that provide a good fit to a linear model. Report your new linear model. What is your new R^2? Does the R^2 matter in choosing your transformation? Explain.

   (c) Using the transformed model, what is the predicted average plasma norepinephrine concentration for a patient with a tumor volume of 100 ml? What is the 95% confidence interval for your prediction? Interpret. What is the 95% prediction interval for a new patient with a 100 ml tumor?

2. Which of the following assumptions are required to test hypotheses using simple linear regression. If you think it isn't valid, explain why not.

   (a) The random variable Y (not conditional on X) is normally distributed.

   (b) The variance of Y depends on X.

   (c) The random variable Y is normally distributed at each value of X.

   (d) The mean of Y (given X) is a linear function of X.

   (e) The random variable X is randomly distributed on some scale.

3. Chp 6, E1 Cancer and Smoking: Fisher's Exact Test and Simulations Studies (simulation R code is given in the supplement to the text which is on Sakai.)

4. Chp 6, E5 Donner Party (simulation R code is given in the supplement to the text which is on Sakai.)