DSA 8301 - Statistical Inference in Big Data
CAT # 2 (due 11:59 PM June 25, 2025)

INSTRUCTIONS:

1. You must show *ALL* work to receive *ANY CREDIT*.

2. Submit your own work. Do not consult *ANYONE*. Violations will be heavily penalized.

1) Let X be the alpha particles emitted by barium-133 in one tenth of a second. An experimenter takes 50 observations of X with a Geiger counter in a fixed position and partitions the set of outcomes into sets:

$$A_1 = \{0, 1, 2, 3\}, \ A_2 = \{4\}, \ A_3 = \{5\}, \ A_4 = \{6\}, \ A_5 = \{7\}, \ A_6 = \{8, 9, 10, ...\}.$$

The observed values for the sets are 13 for A_1, 9 for A_2, 6 for A_3, 5 for A_4, 7 for A_5 and 10 for A_6. The sample mean number of particles is $\bar{x} = 5.4$. Test H_0 : X is distributed as Poisson(λ) versus H_a : X is not distributed as Poisson(λ), at $\alpha = 5\%$.

2) It is claimed that the median weight, m, of certain loads of candy, is 40000 pounds.

a) Use the following 13 observations and the Wilcoxon statistic to test the null hypothesis $H_0 : m = 40000$ vs. $H_a : m < 40000$ at $\alpha = 5\%$.

```
41195 39485 41229 36840 38050 40890 38345
34930 39245 31031 40780 38050 30906
```

b) What is the approximate p-value of this test?
c) Use the sign test to test the same hypothesis.
d) Calculate the p-value from the sign test and compare it with the p-value obtained from the Wilcoxon test.

3) Let X be distributed as $N(\theta, \sigma^2)$, where σ^2 is known. Assume Θ is distributed as $N(\theta_0, \sigma_0^2)$, where θ_0 and σ_0^2 are known.

a) Show that, under the squared error loss function, the Bayes estimator of θ is the posterior mean of Θ.

b) Find the 99% credible interval for θ.

4) Use the Hotelling's T^2 test and the data in the test score data set (scores on Math and Reading tests given to a sample of girls and a sample of boys) to test for a difference in the mean vector of the boys and the mean vector of the girls.

The following R code can be used to read in the data:

```
testdata <- read.table ("Your File Folder/testscoredata.txt", header = T)
attach(testdata)
testdata.noIDs <- testdata[,-1] # removes ID numbers
```