Talis Biomedical Statistics Course - Homework 9

Due: 12 February 2020 9:00 AM

Name: [your first and last name]
Collaborators: [list all the people you worked with]
Date: [date of submission]

By turning in this assignment, I agree by the **Stanford honor code** and declare that all of this is my own work.

The exercises can be found in the DeGroot & Schervish's textbook, available here.

Problem 1

Section 10.1. Exercise 3.

```
import numpy as np
import pandas as pd
from scipy.stats.distributions import chi2

def draw_data(n, seed=1):
  np.random.seed(seed)
  x = np.random.uniform(size=n)
  return x

def chi_square_test(x):
  bins = np.histogram(x, np.arange(0, 1.1, .1))
  expected_count = x.size*0.1
  Q = ((bins[0] - expected_count)**2 / expected_count).sum()
  df = bins[0].size - 1
  p = chi2.sf(Q, df)
  return (Q, df, p)

x = draw_data(200)
chi_square_test(x)
```

The result shows a test statistic of 8.90 and p-value of 0.45, leading to a failure of rejecting the null hypothesis.

Problem 2

Section 10.1. Exercise 4.

	Wears moustache	No moustache
18-30	8	32
Over 30	12	48

We have that

$$Q = \frac{(10 - 24 * 0.25)^2}{24 * 0.25} + \frac{(10 - 24 * 0.5)^2}{24 * 0.5} + \frac{(4 - 24 * 0.25)^2}{24 * 0.25} \quad (1)$$
$$= \frac{11}{3} \quad (2)$$

Looking up the p-value under the $\chi^2(2)$ distribution, we find that it is 0.16. We consequently fail to reject the null.

Problem 3

Section 10.3. Exercise 4.

We can calculate out the marginal probabilities for age group and whether the subject has a moustache. This gives us $P(\text{Over } 30) = 0.6$ and $P(\text{moustache}) = 0.2$. Using this, we get the following expectations

Applying the χ^2-test, we get $Q = \frac{25}{6}$. With $(R-1)(C-1) = 1$ degree of freedom, we have a p-value of 0.04 allowing us to reject the null.

Problem 4

Section 10.5. Exercise 4.

Define the event A to be if the subject is a man and A^C to be the event that they are a woman. Similarly, define B to be the event that the subject receives treatment I and B^C to be the even that they receive treatment II. Then we can use the identity provided in Exercise 2 to prove our result. Note that the only way that the equality in Exercise 2 can hold is if A and B are independent events.

Problem 5

Section 10.6. Exercise 4.

To apply the KS test, we order the observations and compare $F_n(x)$ to $F(x)$ under a uniform distribution. Taking the maximum value of $|F_n(x) - F(x)|$ gives us $D_n^* = 0.60 - 0.42 = 0.18$. Consequently, our test statistic is $n^{1/2} D_n^* = 0.90$. Looking up the corresponding cdf in Table 10.32, we get $H(0.90) = 0.6073$. Therefore, the corresponding p-value is $1 - H(0.90) = 0.3927$. We fail to reject.

Problem 6

Section 10.8. Exercise 4.

We apply the test by ordering all the observations and summing up the ranks of $x_1, ..., x_{25}$, which gives us $S = 399$. Since $m = 25$ and $n = 15$, it follows from Equations 10.8.3 and 10.8.4 that $\mathbb{E}[S] = 512.5$ and $\mathrm{var}\,(S) = 1281.25$. Thus, $Z = \frac{399-512.5}{(1281.25)^{1/2}} = -3.17$. The p-value (from a standard normal distribution) corresponding to a two sided test is 0.0015, allowing us to reject the null.