Assignment 5 - CSC 262 - Fall 2017 - Due December 5

Q1: A coin is tossed $n = 2030$ times, and the outcome is heads $X = 1050$ times. Let p be the probability of heads for a single toss.

(a) Construct a confidence interval for p, using a 95% confidence level.

(b) What sample size is needed to guarantee a margin of error $E_o = 0.01$ using a 95% confidence level.

Q2: Suppose a binomial random variable $X \sim bin(n, p)$ is observed to be $X = 21$, with sample size $n = 56$.

(a) Test hypothesis $H_o : p \leq 0.25$ against $H_a : p > 0.25$. Report a P-value. Is the null hypothesis rejected at a significance level of $\alpha = 0.05$? Use the continuity correction.

(b) Verify your answer using the R `prop.test()` function. What is the P-value without the continuity correction (use the `prop.test()` with the appropriate option to answer this)?

Q3: A chain of grocery stores conducts a customer satisfaction survey at several of its properties. Customers were asked the question: 'Do you shop primarily at this site?'. The following table gives the number surveyed; and the number who answer 'yes' for two different sites.

	Site 1	Site 2
Answered Yes	65	100
Answered No	185	400
Total	250	500

Let p_1, p_2 represent the proportion of each population sampled who would answer 'yes'.

1. Construct a level 0.95 confidence interval for proportion difference $p_2 - p_1$.
2. Test hypothesis $H_o : p_1 = p_2$ against $H_a : p_1 \neq p_2$. Report a P-value. Is the null hypothesis rejected at a significance level of $\alpha = 0.05$?

Q4: Suppose in a clinical trial involving 91 liver cancer patients $n_1 = 34$ were treated with an experimental drug, and $n_2 = 57$ were given the conventional treatment. The following contingency table reports the number of each group that experienced a recurrence within 6 months.

	Experimental Treatment	Standard Treatment	
Cancer recurs	3	12	15
Cancer does not recur	31	45	76
Total	34	57	91

Construct a level 0.95 confidence interval for the log odds ratio of recurrence between groups Experimental Treatment and Standard Treatment. Can you reject the null hypothesis $H_o : OR = 1$ against $H_a : OR \neq 1$ at significance level of $\alpha = 0.05$?

Q5: We are given an *iid* sample from a normal distribution $N(\mu, \sigma^2)$:

$$9.43, 9.85, 10.12, 9.89, 9.81, 10.3,$$

of sample size $n = 6$.

(a) Calculate a level 0.9 upper confidence bound for the standard deviation σ.

(b) Use this upper bound to estimate the sample size needed to construct a 95% confidence interval for the mean μ with a margin of error $E_o = 0.1$ (use a normal approximation).

Q6: We are given samples of size $n_1 = 14$ and $n_2 = 18$ from independent normally distributed populations. Suppose we observe sample variances $S_1^2 = 645.16$ and $S_2^2 = 1413.76$. Do a hypothesis test of

$$H_o : \sigma_2^2 = \sigma_1^2$$
$$H_a : \sigma_2^2 \neq \sigma_1^2$$

using an $\alpha = 0.1$ significance level. Give explicitly the rejection regions, and also report a P-value.

Q7: [**For Graduate Students**] A company has developed a predictive model for the screening of applicants based on a questionnaire. The responses are converted to 3 components:

$$\begin{aligned} X_1 &= \text{Leadership skills} \\ X_2 &= \text{Communication skills} \\ X_3 &= \text{Level of expertise.} \end{aligned}$$

Each component is normally distributed, and has been standardized to have zero mean and standard deviation $\sigma_X = 25$. A composite score believed to be especially predictive of success is given by

$$T = \frac{1}{2}X_1 + \frac{1}{6}X_2 + \frac{1}{3}X_3.$$

The company wishes to use T for screening job applicants. If an applicant's score exceeds a threshold $T \geq t$ they are selected for further interviews. The company wishes to select 10% of applicants for further screening, so it sets the threshold at the value

$$t = \sigma_T \times z_{0.1}$$

where σ_T is the standard deviation of T, and $z_{0.1}$ is the 10% critical value of a standard normal distribution $N(0, 1)$. If σ_T is correctly calculated, and $E[T] = 0$ as expected, we would have

$$P(T > \sigma_T \times z_{0.1}) = 0.1.$$

It is then noted that in order to calculate σ_T, the correlations between X_1, X_2 and X_3 must be known (Sections 4.8 - 4.9 of the lecture notes). Following this, two points of view emerge, which we'll refer to as the null and alternative hypotheses.

- H_o Scales from psychometric questionnaires are designed to measure independent constructs. So, although we might expect, say, leaderships skills and communication skills to be positively correlated in everyday life, the scales X_1, X_2 and X_3 are designed to measure these qualities in a manner that is independent of the others. Therefore, we should expect zero correlation between X_1, X_2 and X_3.

- H_a A statistical analysis has estimated the following correlations, and these should therefore be used to calculate σ_T.

$$\begin{aligned} \rho_{X_1,X_2} &= 0.56 \\ \rho_{X_1,X_3} &= 0.18 \\ \rho_{X_2,X_3} &= 0.21. \end{aligned}$$

In order to test these hypotheses, a sample of n test scores T is to be collected. Design a size $\alpha = 0.05$ hypothesis test for null and alternative hypotheses H_o and H_a. Note that only the scores T will be be available, and not the underlying scores X_1, X_2 and X_3. Construct a plot of power against sample size n, for $n = 2, 3, \ldots, 199, 200$. Superimpose a horizontal line at *power* $= 90\%$. What is the minimum sample size needed to attain at least 90% power?