Math 310 Winter 2021 HW 2 Due Sunday Feb. 21st at midnight.

Your work should be typed and submitted through Canvas.

1. Hexavalent chromiium has been identified as an inhalation carcinogen and an air toxin of concern in a number of different locales. The article "Airborne Hexavalent Chromium in Southwestern Ontario" gave the accompanying data on both indoor and outdoor concentration (nanograms/m^3) for a sample of houses selected from a certain region.

   (a) Calculate a confidence interval for the population mean difference between indoor and outdoor concentrations using a confidence level of 95%, and interpret the resulting interval

   (b) If a 34th house were to be randomly selected from the population, between what values would you predict the difference in concentrations to lie?

2. Cushing's disease is characterized by muscular weakness due to adrenal or pituitary dysfunction. To provide effective treatment, it is important to detect childhood Cushing's disease as early as possible. Age at onset of symptoms and age at diagnosis (months) for 15 children suffering from the disease were given in the article "Treatment of Cushing's Disease in Childhood and Adolescence by Transphenoidal Microadenomectomy". Here are the values of the differences between age at onset of symptoms and age at diagnosis:

   -24 -12 -55 -15 -30 -60 -14 -21 -48 -12 -25 -53 -61 -69 -80

   (a) Construct a normal probability plot of the differences and comment on normality.

   (b) Calculate a lower 95% confidence bound for the population mean difference and interpret the resulting bound.

   (c) Suppose the (age at diagnosis)-(age at onset) differences had been calculated. What would be a 95% upper confidence bound for the corresponding population mean difference?

   (d) By far the most frequently tested null hypothesis when data is paired is $H_0 : \mu_D = 0$. Is that a sensible hypothesis in this context? Explain.

   (e) Carry out a test of hypothesis to decide whether there is compelling evidence for concluding that on average diagnosis occurs more than 25 months after the onset of symptoms.

3. Sometimes experiments involving success or failure responses are run in a paired or before/after manner. Suppose that before a major policy speech by a political candidate, n individuals are selected and asked whether (S) or not (F) they favor the candidate. Then after the speech the same n people are asked the same question. The responses can be entered in a table as follows:

		After	
		S	F
Before	S	x_1	x_2
	F	x_3	x_4

where $x_1 + x_2 + x_3 + x_4 = n$. Let p_1, p_2, p_3, and p_4 denote the four cell probabilities, so that $p_1 = P(S \text{ before and } S \text{ after})$, and so on. We wish to test the hypothesis that the true proportion of supporters(S) after the speech has not increased against the alternative that it has increased.

(a) State the hypothesis of interest in terms of p_1, p_2, p_3, and p_4.

(b) Construct an estimator for the after/before difference in success probabilities.

(c) When n is large, it can be shown that the random variable $\frac{X_i - X_j}{n}$ has approximately a normal distribution with variance given by $[p_i + p_j - (p_i - p_j)^2]/n$. Use this to construct a test statistic with approximately a standard normal distribution H_0 is true (the result is called McNemar's test).

(d) If $x_1 = 350$, $x_2 = 150$, $x_3 = 200$, and $x_4 = 300$ what do you conclude?

(e) McNemar's test can also be used when individuals are matched to create n pairs and then one member of each pair is given treatment 1 and the other is given treatment 2. Then X_1 is the number of pairs in which both treatments were successful, and similarly for X_2, X_3, and X_4. The test statistic for testing equal efficacy of the two treatments is given by $\frac{(X_2 - X_3)}{\sqrt{X_2 + X_3}}$ which has approximately a standard normal distribution when H_0 is true. Use this to test whether the drug ergotamine is effective in the treatment of migraine headaches.

		Ergotamine	
		S	F
Placebo	S	44	34
	F	46	30

4. The attached data gives the observations on time (h) for a AA 1.5-volt alkaline battery to reach a 0.8 voltage.
   (a) Construct a normal probability plot and assess normality
   (b) Does the data suggest that the variance of the Energizer population distribution differs from that of the Ultracell population distribution. Test the relevant hypothesis at the 0.05 significance level.
   (c) The Energizer batteries are much more expensive than the Ultracell batteries. Would you pay the extra money?