HW 4: Two Sample Location Problems (Dependent and Independent)

Instructions: Work must be shown to receive full credit. You may work with others on the homework, but you must write and turn in your own copy. **This does not mean that you can simply copy someone else's work!!** Also, make sure your homework is neat, stapled, and all answers are written in complete sentences!! Come and see me if you have any questions.

On problems that require the use of R, PLEASE give me the RELEVANT R code and output to for each problem so I can assess partial credit. I may take off for including unnecessary R output. If one problem refers back to output from another problem, make sure to cite that output in your answer. Incorrect one-sentence answers will get little or no credit.

NOTE: If a problem asks you to perform a hypothesis test, make sure to give the hypotheses, test statistic, p-value, and a conclusion in the terms of the problem. Also, if the problem asks you to perform a confidence interval, make sure to interpret the confidence interval.

"By Hand" Problems: For hypothesis tests, you may use R to find the p-value. For confidence intervals, you may use R to find the multiplier.

1. An automotive researcher wanted to estimate the difference in distance required to come to a complete stop while traveling 40 miles per hour on wet versus dry pavement. Because car type plays a role, the researcher used 8 different cars with the same drive and tires. The braking distance (in feet) on both wet and dry pavement was recorded. The following table gives a summary of the results.

	n	Mean	Std. Dev.
Wet	8	107.188	3.442
Dry	8	74.488	3.531
Difference	8	32.7	3.767

   (a) Why would you consider these paired or dependent samples? Explain briefly.

   (b) Using a t tool, **construct and interpret** a 90% confidence interval for the mean difference in braking distance on wet versus dry pavement.

   (c) Based on the confidence interval, is there a significant difference in the braking distance on wet pavement as opposed to dry pavement? Why?

2. An experiment is conducted to compare the starting salaries of male and female college graduates who find jobs. Pairs are formed by choosing a male and a female with the same major and similar grade point averages (GPAs). Suppose a random sample of 10 pairs is formed in this manner and the starting annual salary (in $1000s) of each person is recorded. The results are shown in the table below

Pair	1	2	3	4	5	6	7	8	9	10
Male	29.3	41.5	40.4	38.5	43.5	37.8	69.5	41.2	38.4	59.2
Female	28.8	41.6	39.8	38.5	42.6	38.0	69.2	40.1	38.2	58.5

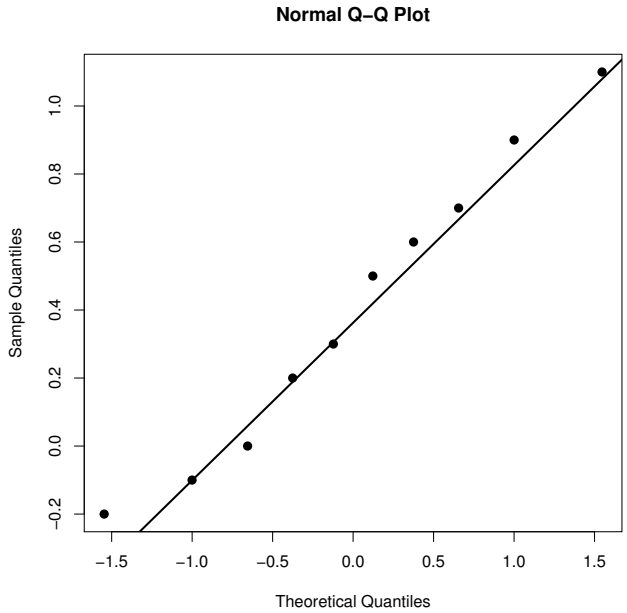
In addition, a table of summary statistics was also computed.

	n	Mean	Std. Dev.
Male	10	43.93	11.665
Female	10	43.53	11.617
Difference	10	0.4	0.4346

(a) Using the paired t test, is there enough evidence to conclude that males make more money than females on average? Be sure to state the hypotheses, calculate the test statistic by hand, find the p-value in `R`, interpret the p-value and state your conclusion in context.

(b) Using the paired Wilcoxon Signed Rank test, is there enough evidence to conclude that males typically make more money than females? Be sure to state the hypotheses, calculate the test statistic by hand, find the p-value in `R`, interpret the p-value and state your conclusion in context.

(c) Below is a Normal probability plot of the data difference between pairs of male and female salaries. Based on this plot, which of the above analyses do you feel is most appropriate?



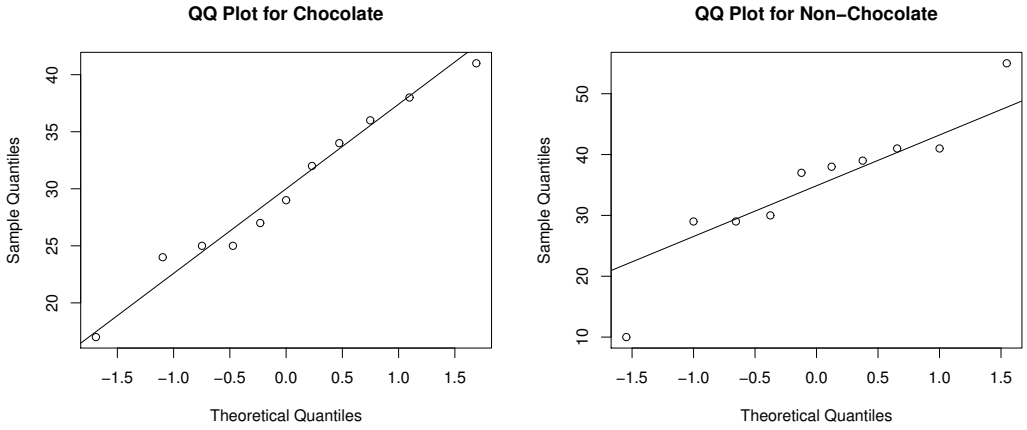
3. The number of grams of carbohydrates contained in 1-ounce servings of randomly selected chocolate and non-chocolate candy is listed below.

Chocolate	25	17	36	41	25	32	38	34	24	27	29	$\bar{x}_1 = 29.82$	$s_1 = 7.11$
Non-chocolate	41	41	37	29	30	38	39	10	29	55		$\bar{x}_2 = 34.9$	$s_2 = 11.66$

(a) Use the Wilcoxon Rank Sum Test to test the claim that a difference exists between the typical amount of carbohydrates for chocolate candy as opposed to non-chocolate candy. Be sure to show all steps expected of a hypothesis test write-up.

(b) Use the t-test to determine if a difference exists in the mean amount of carbohydrates in chocolate and non-chocolate candy. Be sure to show all steps expected of a hypothesis test write-up.

(c) Consider the plots provided and determine which analysis is most appropriate. Justify your answer.



4. A researcher was interested in examining the age of gamblers. A survey was conducted that examined the ages of 24 randomly selected slot machine players and 35 randomly selected roullette players. The following table gives a summary of the results.

Game	n_i	$\bar{x}_i$	s_i
Slots	24	48.7	6.8
Roulette	35	55.3	3.2

Construct and interpret a 95% confidence interval for the difference in mean age between slot machine players and roulette players.

"R" Problems:

5. In a statistics class, students took their pulses before and after being frightened. The frightening event was having the teacher scream and run from one side of the room to the other. The pulse rates (beats per minute) before and after the scream were obtained separately and given below.

Pulse Before	64	100	80	60	92	80	68	84	80	68	60	68	68
Pulse After	68	112	84	68	104	92	72	88	80	92	76	72	100

Treat this as though it were a random sample of students.

(a) Test the hypothesis that the pulse rate is higher after a fright using a t tool. Be sure to provide the `R` code for data entry and the appropriate test along with a full hypothesis test write-up of the results.

(b) Test the hypothesis that the typical pulse rate is higher after a fright using a nonparametric method.Be sure to provide the `R` code for data entry and the appropriate test along with a full hypothesis test write-up of the results.

(c) Construct a Normal probability plot of the differences and comment on which of the above methods is most appropriate in this setting.

6. A study was conducted to analyze whether the use of cell phones impairs reaction times while driving. A random sample of 32 subjects were selected and were asked to perform a specific diving task while using a cell phone and then without a cell phone. The reaction time (in milliseconds) was measures for each subject. The data can be found on *Moodle* (`reacttime.xlsx`) where `wocell` represents the reaction times without a cell phone and `withcell` represents the reaction times with a cell phone.

   (a) Using a t tool, **construct and interpret** a 95% confidence interval for the mean difference in reaction time while using a cell phone while driving as compared to without a cell phone.

   (b) Construct a Normal probability plot of the differences and comment on the appropriateness of the procedure.

7. A study was conducted to analyze whether a home with a pool is typically larger than than a home without a pool for a particular suburb. To answer this question, a random sample of 20 homes with a pool was selected along with a random sample of 20 homes without a pool. The size of each home (in square feet) was measured. The data can be found on Moodle (`poolhomesize.xlsx`).

   (a) Use the appropriate t tool, to test the claim that homes with a pool are larger than homes without a pool.

   (b) Use the appropriate nonparametric method to test the claim that homes with a pool are larger than homes without a pool.