SOLUTIONS TO LAB ASSIGNMENT 4

Question 1

(a) The 11 subjects constitute a subset of a larger group of 16 individuals who believed they were addicted to caffeine and agreed to participate in the study. As the subjects are volunteers, they cannot be treated as a random sample from the population of all caffeine dependent individuals. As a consequence, no generalizations can be made to the population of all caffeine dependent individuals from the study.

(b) The study is double-blind because neither the subjects themselves nor the researchers who worked with them knew which treatment any subject had received in each of the two periods. Subjects were not told which pills were the placebos so that they would not develop physical symptoms just from the belief that they should experience withdrawal when they did not ingest caffeine. Because the subjects did not know which capsules were which, symptoms would be caused by their bodies physically reacting to the presence or absence of caffeine, and their symptoms, or lack thereof, would not be psychologically triggered. The interviewers were not told of the subjects' status so that they would not be influenced as they tried to record any evidence of functional impairment. For example, if a subject experienced unusual problems during one of the study periods, yet the interviewer knew that the subject had been taking the caffeine-free pills, the interviewer might try to elicit examples of impairment from the subject, when in fact none exist.

(c) The order of the two series of capsules was randomized so that any differences that might affect the psychological characteristics measured during the two periods would not consistently affect one period over the other. It was also necessary to ensure that the experiment was truly double-blind.

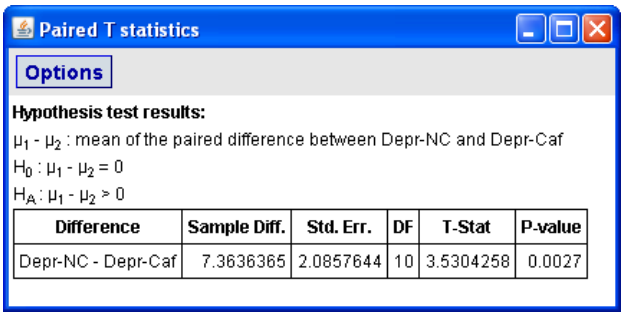
(d) The two study periods were not held consecutively to lessen the impact of carry-over effects. If a subject was denied caffeine in the first study period, withdrawal effects might linger during the next few days and confound the results from the second period if the periods are held to close together.

Question 2

(a) The null and alternative hypotheses should be defined as follows:

H_0: $\mu_D = 0$ versus H_A: $\mu_D > 0$, where μ_D = mean change in depression score (no-caffeine period minus caffeine period). The alternative hypothesis states that being deprived of caffeine raises depression scores. Given the study design, the paired-data t test should be used to answer the question of interest. The data are paired because each value is the change in the depression score for the same person.

The related StatCrunch output is shown below:



Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
Depr-NC - Depr-Caf	7.3636365	2.0857644	10	3.5304258	0.0027

According to the output, the value of the test statistic is 3.53. The test statistic follows a t-distribution with 10 degrees of freedom under the null hypothesis of no difference between the average depression scores for the no-caffeine and caffeine periods. The p-value of the one-sided test is 0.0027. The small p-value provides strong evidence that the mean of the differences is positive or equivalently that the

mean depression score for the no-caffeine period is higher than the mean depression score for the caffeine period.

(b) The related StatCrunch output is shown below:

Paired T statistics

Options

95% confidence interval results:

$\mu_1 - \mu_2$: mean of the paired difference between Depr-NC and Depr-Caf

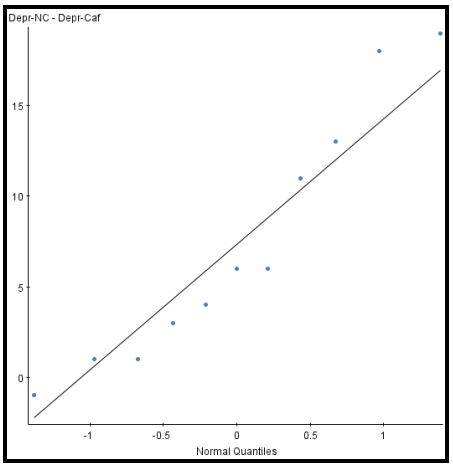
Difference	Sample Diff.	Std. Err.	DF	L. Lim	U. Lim
Depr-NC - Depr-Caf	7.3636365	2.0857644	10	2.7162635	12.011009

Based on the output, a 95% confidence interval for the mean increase in depression scores between no-caffeine and caffeine periods is approximately (2.72, 12.01).

Since the above interval has both endpoints positive, there is evidence that the mean depression score for the no-caffeine period is higher than the mean depression score for the caffeine period which is consistent with the outcome of the test in part (a).

(c) The t procedures used in parts (a) and (b) are valid only if the differences come from an approximately normal distribution (nearly normal condition) and the differences constitute a simple random sample.

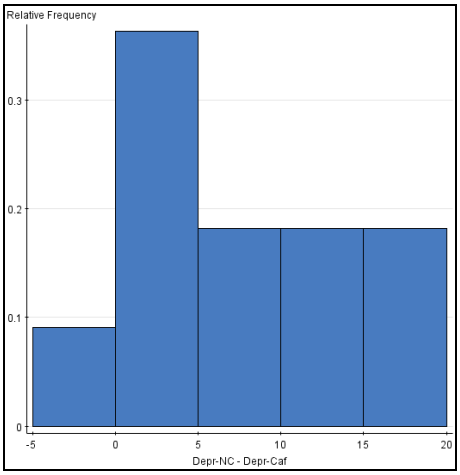
First we discuss the assumption of normality. The paired data t test in part (a) and a confidence interval in part (b) are quite robust against non-normality. Their robustness increases with the sample size. The most reliable graphical tool to verify the assumption of normality is the normal probability plot (Q-Q plot in StatCrunch). The Q-Q plot for the 11 differences is shown below:



The Q-Q plot for the 11 observations displays some weak concavity (several consecutive points below the line in the middle range, though relatively close to the line, and a few in the lower and upper range of the data). Some students also may claim that the concave pattern is not clear enough due to the relatively small number of observations and the plot does not exhibit any systematic deviations from a straight line pattern, so there is no sufficient evidence that the data are non-normal.

A histogram of the differences can also be obtained to check the assumption of normality. This graphical tool is clearly less reliable tool to check the assumption of normality than Q-Q plot. If a histogram is heavily skewed and has some outliers, there is evidence that data are non-normal.

A roughly symmetric histogram of the differences supports the assumption of normality. The histogram for the data displays some right skewness:



Due to robustness of the paired-data inferential tools, the weak right skewness in the data should not affect the validity of the paired-data inferences in parts (a) and (b).

Now we discuss the assumption that the data should constitute a simple random sample. The assumption is clearly not satisfied, because the data were obtained from 11 subjects addicted to caffeine who agreed to participate in the experiment. No generalizations to the populations can be made based on the data. The fact that the assumption is not satisfied is the chief threat to the validity of the results obtained in parts (a) and (b).

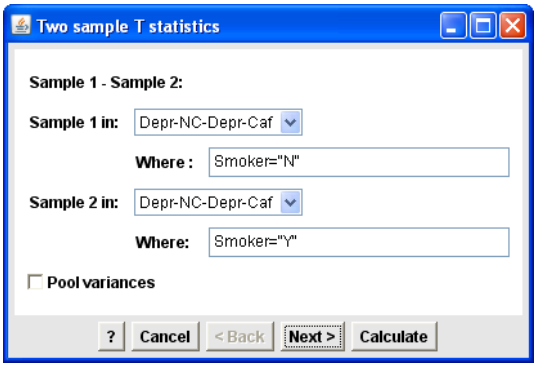
Question 3

(a) In order to compare the changes in the depression scores for smokers and non-smokers, the independent samples t-test will be used. As the sample sizes are small (3 and 8), the test without the assumption of equal variances should be the first choice to compare the two groups (small samples do not provide reliable information about variation in the corresponding populations).

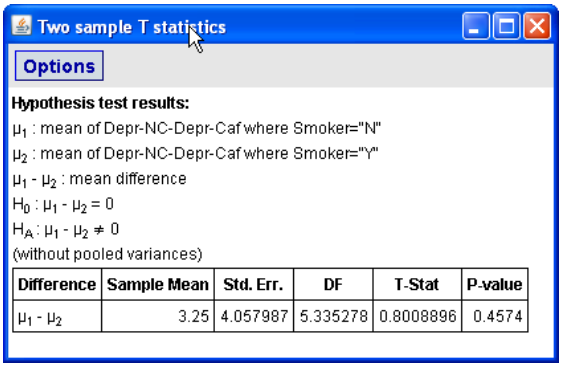
The null and alternative hypotheses should be defined as follows:

H_0: $\mu_N - \mu_S = 0$ versus H_A: $\mu_N - \mu_S \neq 0$,

where μ_N = mean change in the depression score for non-smokers and μ_S = mean change in the depression score for smokers. The related dialog box in StatCrunch should be filled as follows:



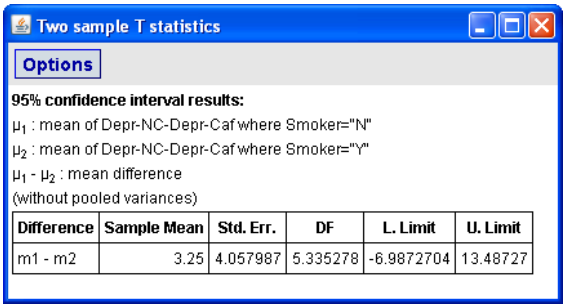
The corresponding StatCrunch output is shown on the next page. According to the output, the value of the test statistic is 0.8009. The t statistic follows a t distribution with DF=5.335 and the p-value of the two-sided test is 0.4574. The high p-value provides no evidence that the change in depression levels is different for smokers and non-smokers.



Difference	Sample Mean	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	3.25	4.057987	5.335278	0.8008896	0.4574

(b) The t test for independent samples is used in this case because the three observations for smokers and 8 observations for non-smokers are independent. The two-sample t test is valid under the assumption that the two samples are simple random samples from the appropriate populations, the samples are independent, and each population is normally distributed.

(c) The 95% confidence interval is (-6.9872704, 13.48727).



Difference	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
m1 - m2	3.25	4.057987	5.335278	-6.9872704	13.48727

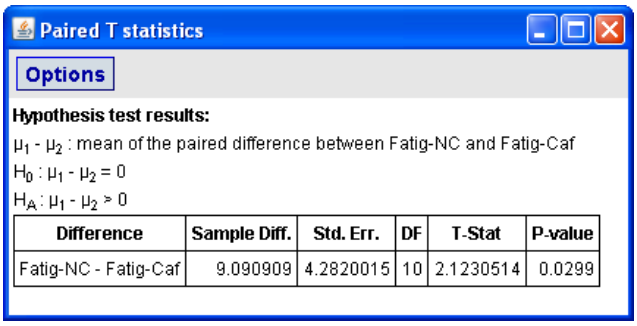
The difference in the mean depression for non-smokers and smokers with 95% confidence is between -6.9872 and 14.4873. Since the above interval contains zero, there is insufficient evidence that the mean depression score change for no-smokers is different than the mean depression score change for the smokers.

Remark: Some students noticed that the ratio of the larger and the smaller standard deviation 7.55/5.29<2 and used the test and confidence interval with the pooled estimate of the variance. In this case, the value of the test statistic is t=0.6748, DF=9, and p-value=0.5168. The corresponding 95% confidence interval is (-7.6452, 14.1452). The conclusions about the mean difference are the same as those reached above for the inferences without the assumption of equal variances.

Note that the assumption of equal variances is hard to check for the data due to the small sample sizes. The graders should also accept the above solution.

Question 4

(a) The null and alternative hypotheses should be defined as follows: H_0: $\mu_D = 0$ versus H_A: $\mu_D > 0$, where μ_D = mean change in fatigue score (no-caffeine period minus caffeine period). The alternative hypothesis states that being deprived of caffeine raises fatigue scores. The paired-data t test is used. The related StatCrunch output is shown below:

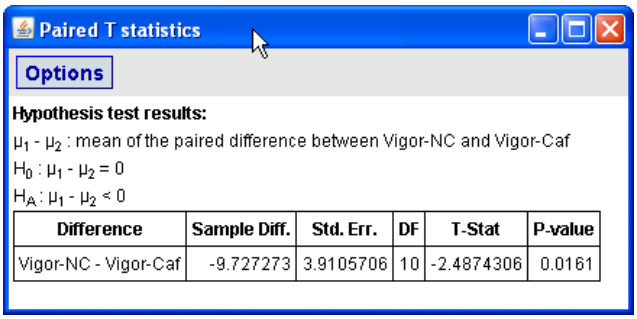


According to the output, the value of the test statistic is 2.12. The p-value of the one-sided test is approximately 0.0299. The small p-value provides strong evidence that the mean of the differences is positive or equivalently that the mean fatigue score for the no-caffeine period is higher than the mean fatigue score for the caffeine period.

(b) The null and alternative hypotheses should be defined as follows:

H_0: $\mu_D = 0$ versus H_A: $\mu_D < 0$,

where μ_D = mean change in vigor score (no-caffeine period minus caffeine period). The alternative hypothesis states that being deprived of caffeine lowers vigor scores. The paired-data t test is used. The related StatCrunch output is shown below:



According to the output, the value of the test statistic is -2.4874. The p-value of the one-sided test is 0.0161. The small p-value provides strong evidence that the mean of the differences is negative or equivalently that the mean vigor score for the no-caffeine period is smaller than the mean vigor score for the caffeine period.

Question 5

The above analysis shows that being deprived of caffeine raises depression scores, increases fatigue and reduces vigor scores. Higher caffeine daily intake may result in more depressive mood. There is no evidence of difference in depression scores between smokers and non-smokers.

The change in mood (more depressive) seems to be the most severe manifestation of the withdrawal symptoms (the largest absolute value of the test statistic and the smallest p-value of the test). As there is no random sample, the results cannot be generalized to the whole population of caffeine dependent individuals unless we assume that the volunteers are not different from the caffeine dependent individuals in the general population.

LAB 4 ASSIGNMENT: MARKING SCHEMA

Proper Header and appearance: 10 marks

Proper Header: include your name, course section (e.g. Stat 151 Q1) and lab assignment number on the cover page. Lab reports must be **typed**.

Question 1

(a) Random sample or not: 2 marks
Generalizations: 2 marks
(b) Double-blind study discussion: 2 marks
(c) Order of administration of pills: 2 marks
(d) Timing of the two study periods: 2 marks

Question 2

(a) Hypotheses: 3 marks
Value of the test statistic: 2 marks
Null Distribution: 2 marks
P-value: 2 marks
Conclusion: 2 marks

(b) 95% confidence interval: 4 marks
Comparison of the interval with the test: 2 marks

(c) Specifying assumptions (in general): 2 marks
Normality assumption for the data: 2 marks
Plot to verify the assumption of normality: 3 marks
SRS assumption for the data: 2 marks
Chief threat to validity: 2 marks

Question 3

(a) Hypotheses: 3 marks
Value of the test statistic: 2 point
Null distribution: 2 marks
P-value: 2 marks
Conclusions: 2 marks
(b) Choice of the test: 2 marks
Assumption of the two-sample t test: 2 marks
(c) Confidence interval: 4 marks
Consistency of the confidence interval with the test: 2 marks

Question 4

(a) Hypotheses: 3 marks
Value of the test statistic: 2 point
P-value: 2 marks
Conclusions: 2 marks

(b) Hypotheses: 3 marks
Value of the test statistic: 2 point
P-value: 2 marks
Conclusions: 2 marks

Question 5

Brief summary (including the answer to the question): 5 marks

TOTAL= 92