ST 516 - Homework 6
student Paul ReFalo 10/31/17

1. (6 points) Why is it important to correctly distinguish between the 2-sample t-test setup, and the paired t-test setup? Follow the steps to explore the difference between these two tests.

(a) Marquis de Laplace proved the Central Limit Theorem in 1810, so begin with set.seed(1810). Run the following code to generate three samples A, B and C

```
> set.seed.(1810)
> A <- rnorm(10)
> B <- rnorm(10)
> C <- 0.5 + (0.8 * A) + (sqrt(1 - 0.8^2) * B)
> B <- 0.5 + B
> A
 [1]  0.6213137  0.6345510  1.6268240  1.5267466 -2.5737812  0.3891985 -0.6036475  1.5232994
0.0438107 -1.4153427
> B
 [1]  0.167628  0.057247 -0.405866  1.180586  1.210035  1.053957  2.300863  1.098171 -0.423522
2.827727
> C
 [1]  0.7976279  0.7419890  1.2579397  2.1297487 -1.1330038  1.1437331  1.0976000  2.0775420
-0.0190648  0.7643623
```

(b) Conduct a two sample t-test of $H_0 : \mu_A - \mu_B = 0$ vs. $H_A : \mu_A - \mu_B /= 0$, assuming unequal group variances. Then calculate the t-statistic and p-value manually, using the formula for the test statistic and pt(). It should match the output of t.test(). Hint: you can use the degrees of freedom reported by t.test().

```
> t.test(A, B, mu = 0, paired = FALSE, conf.level = 0.95) # two-sided by
default

	Welch Two Sample t-test

data:  A and B
t = -1.319, df = 17.11, p-value = 0.205
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.89587  0.43710
sample estimates:
mean of x mean of y
 0.177297  0.906683
```

```
> # manually calculate the t statistic and pvalue
> t <- (Abar - Bbar) / sqrt((Avar/10) + (Bvar/10))
> t
[1] -1.31862
> df <- 17.11 # from t-test
> Pvalue <- 2 * pt(t, df) # multiply by two for two-sided test
> Pvalue
[1] 0.204676
```

(c) Now conduct a paired t-test of $H_0 : \mu_A - \mu_B = 0$ vs. $H_A : \mu_A - \mu_B /= 0$, using t.test(). Then calculate the t-statistic and p-value manually by constructing the test statistic using the sample statistics of the differences (i.e. using diffs where diffs <- A - B). It should match the output of t.test().

```
> t.test(A, B, mu = 0, paired = TRUE, conf.level = 0.95) # two-sided by
default

      Paired t-test

data:  A and B
t = -1.08, df = 9, p-value = 0.308
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.257507  0.798736
sample estimates:
mean of the differences
              -0.729385


> diffs <- A - B
> SEdiff <- sd(diffs) / sqrt(10)
>
> tPaired <- (mean(diffs) - 0) / (SEdiff)
> tPaired
[1] -1.07975
> dfDiff <- 9 # from t-test
> PvalueDiff <- 2 * pt(tPaired, dfDiff) # multiply by two for two-sided test
> PvalueDiff
[1] 0.308339
```

(d) Compare the test statistics, confidence intervals and p-values from part (b) and part (c). Do the two procedures reach the same conclusion?

	Test Statistic	Confidence Interval	p-value
Welch Two Sample t-test	-1.309	-1.90 to 0.44	0.205
Paired t-test	-1.08	-2.26 to 0.80	0.308

The Welch t-test results in a smaller CI and p-value. Both procedures reach the same conclusion. That is, at the 95% confidence interval there is not evidence that the sample means of A and B are different. These two tests fail to reject the Null Hypothesis that the sample means are the same.

(e) Repeat parts (b), (c) and (d), now comparing samples A and C (just use t.test(), you don't need to verify the results manually).

```
> t.test(A, C, mu = 0, paired = FALSE, conf.level = 0.95) # two-sided by default

	Welch Two Sample t-test

data:  A and C
t = -1.342, df = 16.04, p-value = 0.198
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.82723  0.41013
sample estimates:
mean of x mean of y
 0.177297  0.885847

> t.test(A, C, mu = 0, paired = TRUE, conf.level = 0.95) # two-sided by default

	Paired t-test

data:  A and C
t = -2.716, df = 9, p-value = 0.0237
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.298619 -0.118481
sample estimates:
mean of the differences
               -0.70855
```

	Test Statistic	Confidence Interval	p-value
Welch Two Sample t-test	-1.342	-1.83 to 0.41	0.198
Paired t-test	-2.716	-1.30 to -0.12	0.02

Here the Welch t-test results in a larger CI and p-value. Moreover, the Welch t-test failed to reject the Null Hypothesis while the Paired t-test rejects the Null Hypothesis in favor of the Alternative Hypothesis that the same means are different at a 95% confidence interval.

(f) You should find that the two procedures reach roughly the same conclusion when comparing samples A & B, but different conclusions when comparing samples A & C, despite the true differences in mean in both cases being 0.5. Explain why.
You may find it helpful to either examine how the samples were generated and/or examine the following plots of the the three samples.

The conclusions are different because when using the Paired t-test *the differences* are analyzed and because of the way this data was generated, the differences in A and C are large enough to reject the Null Hypothesis (Paired t-test, t = -2.72, p-value = 0.02). Because of the way the data is generated, when comparing A and B we get nearly the same number of paired samples having a positive slope (6) as we do getting a negative slope (4). This contrasts to the situation comparing A and C where the number of paired samples showing a positive slope is 8 with one have a negative slope and another with a slope nearly zero (slightly negative). This imbalance causes the Paired t-test for data sets A and C to diverge. While the mean of B (0.91) and C (0.89) are quite close, the way they are paired results in a different conclusion for the Welch v. Paired t-test at a 95% confidence interval.

2. (2 points) A language transcriptionist translates a random sample of seven speeches from Spanish to English and then from Spanish to French. The times it takes to transcribe them (in minutes) are recorded in the table below:

   The transciptionist would like to know if there is a difference in times it takes her to transcribe from Spanish to English and from Spanish to French for the same speech.

(a) Are these data paired? Why or why not? If so, what are the sources of variation?

Yes, the data here are paired because the two data sets are a result of the work of one person translating the same speech. Some sources of variation could be the vocabulary in the speeches may lend themselves more readily translated to one language or the other. Also, the paired translations might have occurred at different times (i.e. morning v afternoon) or under different circumstances that would induce a bias.

(b) Perform a hypothesis at the 5% level using R. Report the t-statistic, degrees of freedom, p-value, and a confidence interval.

```
> SE <- c(15, 19, 45, 35, 67, 13, 33)
> SF <- c(16, 18, 60, 54, 70, 11, 34)
> tDiffs <- SF - SE
> tDiffs
[1]  1 -1 15 19  3 -2  1
> t.test(SE, SF, mu = 0, paired = TRUE, conf.level = 0.95) # two-sided by default

	Paired t-test

data:  SE and SF
t = -1.632, df = 6, p-value = 0.154
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.85167   2.56595
sample estimates:
mean of the differences
               -5.14286
```

	t-statistic	df	confidence interval	p-value
Paired t-test	-1.632	6	-12.85 to 2.57	0.154

The CI contains the value zero, so we fail to reject the Null Hypothesis.

(c) State a conclusion in the context of the problem.

There is not strong evidence that the two populations, the rate of transition Spanish to English and French respectively, have means that differ (paired t-test, CI 95%, p-value = 0.154, df = 6).

(d) Might there be practical significance here? Why or why not?

Possibly there is a practical significance. Notice that the only two speeches with a relatively large difference in their translation time were near to one-hour in translation time while those with low differences were typically 30 minutes or less. There is one other speech of 70 minutes that has a low difference but all speeches of near the one-hour marks how Spanish to French being longer than Spanish to English. Practically, this might indicate that taking a break at 30 minutes would be helpful in reducing the differences or that taking an French class would reduce those times making the results closer to those of the Spanish to English times.

3. (2 points) A chemist wishes to compare the amount of residue left behind for chemical reaction A and chemical reaction B, given a certain environment for each reaction. If they are different, then she will adjust the starting amounts of each chemical involved until the amounts of residues are the same for each of the two reactions. Assume each run of each reaction is independent, and that the she tries her best to make the environment of each run the same within each chemical reaction.

. (a) Are these data paired? Why or why not? If so, what are the sources of variation?

The data are not paired. Though the reactions were conducted by the same Chemist, the resulting residue has a lot to do with the reaction itself (stoichiometry, efficiency, product m.w. and phase at room temp, etc.). Of course, there may be some variation performing these tests from chemist to chemist, but the main drivers are the nature of the reactions.

. (b) Perform a hypothesis test at the 5% level using R. Report the t-statistic, the degrees of freedom, the p-value, and a confidence interval.

```
> RxnA <- c(456, 222, 567, 344, 222, 334, 543, 447)
> RxnB <- c(343, 242, 990, 222, 344, 455, 600, 323)
> t.test(RxnA, RxnB, mu = 0, paired = FALSE, conf.level = 0.95)
```

```
        Welch Two Sample t-test

data:  RxnA and RxnB
t = -0.4752, df = 10.6, p-value = 0.644
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -271.341  175.341
sample estimates:
mean of x mean of y 
  391.875   439.875 
```

	t-statistic	df	confidence interval	p-value
Welch Two Sample t-test	-0.4752	10.6	-271.341 to 175.341	0.644

. (c) State a conclusion in the context of the problem.

There is no evidence that the two populations, micrograms of residue left from reaction A and B respectively, have means that differ (Welch Two Sample t-test, CI 95%, p-value = 0.644, df = 10.6). With 95% confidence the mean residue weight in of Reaction A is between 271.3 micrograms lighter and 175.3 micrograms heavier than Reaction B.