Name:_____

Calculating a t-test by hand

This document includes information that is important for you to know about t-tests, the process of calculating a t-test statistic, and a problem that you will calculate by hand. The goals of this are to have you practice thinking about the components that go into calculating a t-test to help you better understand what they are and how they work, and to practice interpreting the results of a t-test. Note that there are many different types of t-test and in this case you are calculating a form of 2-sample t-statistic.

SUBMIT THIS COMPLETED DOCUMENT BY MOODLE TUES ~~MON~~ OCT. 25 BEFORE MIDNIGHT

Conceptually, a t-test compares the size of the difference between two means with the standard error of that difference. The numerator is the difference between the means, and the denominator is the difference in "standard error". "Standard error" is a measure of variance in descriptive statistics that combines the standard deviations (*s*) of the groups with the sample sizes of each group. (Note the

$$t = \frac{difference\ between\ means}{differences\ in\ standard\ error\ between\ groups}$$

$$\text{So, } t = \frac{Sample\ mean\ of\ pop\ 1 - Sample\ mean\ of\ pop\ 2}{Square\ root\ of\ (Sample\ std\ dev\ pop\ 1\ squared\ /\ N\ for\ pop\ 1) + (Sample\ std\ dev\ pop\ 2\ squared\ /\ N\ for\ pop\ 2)}$$

** Note that "pop 1" and "pop 2" will be two groups or categories in a categorical variable in a spreadsheet (data frame).*

To calculate a t-statistic t:

1. Calculate the mean of both groups ($\bar{x}_1$, $\bar{x}_2$)
2. Calculate the standard deviation for each group (s_1 , s_2)
3. Find the population size for each group (n_1 , n_2)
4. Add these values to the equation and calculate the t-statistic.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

To translate that t-statistic to a p-value

1. How many degrees of freedom in a two-sample t-test (df)?
    a. Df = (n_1 + n_2) - 2
2. Use the t-statistic table on the last page to find the p-value that is associated with your t-statistic and degrees of freedom. Use the 2-tailed option with a significance level of 0.05.

We are going to use a randomly-drawn subset from the penguins data set to compare the body mass of Adelie and Gentoo penguins.

species	island	body_masss_g	sex
Adelie	Biscoe	3950	male
Adelie	Dream	3425	male
Adelie	Torgersen	3350	female
Adelie	Biscoe	4300	male
Adelie	Biscoe	3600	female
Adelie	Biscoe	3825	female
Adelie	Biscoe	3950	male
Adelie	Torgersen	3450	female
Adelie	Torgersen	4300	male
Adelie	Biscoe	2925	female
Gentoo	Biscoe	4875	female
Gentoo	Biscoe	5600	male
Gentoo	Biscoe	5950	male
Gentoo	Biscoe	5050	male
Gentoo	Biscoe	4450	female
Gentoo	Biscoe	4400	female
Gentoo	Biscoe	4150	female
Gentoo	Biscoe	5000	female
Gentoo	Biscoe	4750	female
Gentoo	Biscoe	4400	female

	Adelie	Gentoo
Mean body mass		
Standard deviation of body mass		
Sample size		

How many degrees of freedom?

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300

What is the critical value for your t-statistic given your degrees of freedom and a significance level of 0.05? What does this value mean?

Write a sentence or two that accurately reports your t-statistic, the p-value associated with it the critical value, and what that means in terms of whether or not the two populations you are comparing are statistically significantly different.