MATH 140 Worksheet **Name:** Key **Date:**

Inference on Two Means

Inference on two means

- Parameter of interest: $\mu_1 - \mu_2$, difference of two population means.
  Point estimate: $\bar{x}_1 - \bar{x}_2$, difference of sample means.
- Conditions:
  - independence within groups (often verified by a random sample, and if sampling without replacement, $n <$ 10% of population) and between groups
  - no extreme skew in either group
- For hypothesis testing, use test statistic

$$t = \frac{\text{point estimate} - \text{null value}}{SE}, \quad = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

where $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, and assume t lives in a t-distribution with df $= \min(n_1 - 1, n_2 - 1)$.

- Confidence interval: point estimate $\pm\, t^\star_{df} \times SE$

1. Do students have older cars than faculty at Linfield? A Random sample of 14 student cars have ages with a mean of 8 years and a standard deviation of 3.6 years, while 28 randomly selected faculty cars have ages with a mean of 5.8 years and a standard deviation of 3.3 years. Use a $\alpha = 0.01$ significance level to test the claim that student cars are older than faculty cars.

   (a) What is the parameter of interest? State hypotheses for the appropriate hypothesis test.

   $\mu_1 - \mu_2$ where μ_1 = Linfield Student ~~car mean~~ mean car age, and μ_2 = Linfield Faculty mean car age.

   $H_0: \mu_1 - \mu_2 = 0$

   $H_a: \mu_1 - \mu_2 > 0$ — this corresponds to "Student cars older on average"

   (b) Determine the test statistic. What distribution do we assume this test statistic lives in?

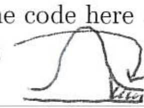
   Data Stats

	Student (1)	Faculty (2)
n	14	28
$\bar{x}$	8.0	5.8
s	3.6	3.3

   $$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{8.0 - 5.8}{\sqrt{\frac{(3.6)^2}{14} + \frac{(3.3)^2}{28}}} = \frac{2.2}{\sqrt{.9257 + .3889}} = 1.92$$

   df = 13 (smaller sample size minus 1)

   (c) Use RStudio to calculate the p-value. Record the code here as well as the p-value.

   Ha is "greater than" so p-value is  this area

   1.92

   $= 1 - pt(1.92, 13)$
   $= .0385$

   (d) Is there sufficient evidence to support the claim that student cars are older than faculty cars? Explain. Be sure to reference the p-value in your explanation.

   ~~Test~~ ~~[illegible]~~

   At the $\alpha = .01$ level, we do <u>not</u> have significant evidence to reject H_0 in favor of H_a, since p-value $> \alpha$.

2. Using the data from the previous question, construct a 99% confidence interval estimate of the difference $\mu_s - \mu_f$, where μ_s is the mean age of all student cars at Linfield and μ_f is the mean age of all faculty cars at Linfield.

   99% CI: $(\bar{X}_1 - \bar{X}_2) \pm t^* \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ $\quad (t^* = qt(.995, 13) = 3.01)$

   $2.2 \pm 3.01 \cdot (1.1466)$

   $\boxed{2.2 \pm 3.45}$ - OR - -1.25 to 5.65

   this is an interval that we believe captures the true value of $\mu_S - \mu_F$

3. A researcher wished to compare the average amount of time spent in extracurricular activities by high school students in a suburban school district with that in a school district of a large city. Is there a difference? The researcher obtained an SRS of 60 high school students in a large suburban school district and found the mean time spent in extracurricular activities per week to be $\bar{x}_1 = 6$ hours, with a standard deviation $s_1 = 3$ hours. The researcher also obtained an independent SRS of 40 high school students in a large city school district and found the mean time spent in extracurricular activities per week to be $\bar{x}_2 = 4$ hours, with a standard deviation $s_2 = 2$ hours. Let μ_1 and μ_2 represent the mean amount of time spent in extracurricular activities per week by the populations of all high school students in the suburban and city school districts, respectively. Assume the two-sample t procedures are safe to use.

(a) What is the parameter of interest? State hypotheses for the test.

Parameter: $\mu_1 - \mu_2$

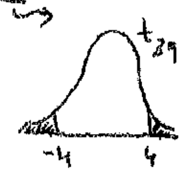
where μ_1 = mean hours for suburban district

and μ_2 = mean hours for big city district

$H_0: \mu_1 - \mu_2 = 0$ vs $H_a: \mu_1 - \mu_2 \neq 0$. (2-sided test!)

(b) Determine the test statistic and calculate the p-value using RStudio. (df = 40 − 1 = 39 (smaller n − 1))

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{6-4}{\sqrt{\frac{3^2}{60} + \frac{2^2}{40}}} = 4.0$$



2 sided test so we want both tail areas!

p-value = 2 * pt(-4, 39) = .0003

(c) Assuming a significance level of $\alpha = .05$, is there sufficient evidence to conclude that there is a difference in the average amount of time spent in extracurricular activities in these two groups? Explain. Be sure to reference the p-value in your explanation.

Yes p-value < α = .05 so we have sufficient evidence to reject Ho in favor of Ha and conclude there is a difference in the average hours students spend on extracurricular activities in these 2 districts.

4. Test the claim that the two samples described below come from populations with the same mean. Assume that the samples are independent simple random samples.

Sample 1: $n_1 = 14$, $x_1 = 11.0$, $s_1 = 2.5$

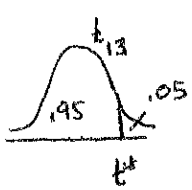
Sample 2: $n_2 = 23$, $x_2 = 9.0$, $s_2 = 5$.

(a) Determine the test statistic.

$$t = \frac{11 - 9}{\sqrt{\frac{2.5^2}{14} + \frac{5^2}{23}}} = 1.62$$

(we assume this is a t-score on a t_{13} distn.

(b) Find the t critical value t^* for a significance level of 0.05 for an alternative hypothesis that the first population has a larger mean (one-sided test). This value will be the t score in the appropriate t distribution that has area .05 above it.



t^* = qt(.95, 13) = 1.77

(c) Comparing our test statistic to the critical t^*, state your conclusion. Is there sufficient evidence to warrant rejection of the claim that the two populations have the same mean and accept that the first population has a larger mean?

We reject Ho in favor of H_a if our test statistic t is greater than t^*. Since it isn't, we fail to reject Ho here.