Class Example 1. Two Sample t-test - equal variances assumption

Execute the following code and make sure that you understand the output.

```
# Create data
using Distributions
srand(2017) #Keep same seed for reproduceability
dBoys = Normal(19.7,2.8)
dGirls = Normal(21.1,3.1)

nBoys = 12
nGirls = 17

dataBoys = [rand(dBoys) for _ in 1: nBoys]
dataGirls =  [rand(dGirls) for _ in 1: nGirls];

(mean(dataBoys),std(dataBoys)),(mean(dataGirls),std(dataGirls))
```

Note that the statistics are: ((19.35,2.66),(21.87,3.18)).

```
# Perform two-sided t-test of the null hypothesis that the two
# populations from which the data come from have the same mean,
# and equal variance, against the null that the population means
# are different, but the variances are equal.
using HypothesisTests
EqualVarianceTTest(dataBoys,dataGirls)
```

```
# We now verify the StandardError, tStatistic, and p-value ourselves
# from the formulas.
sPooled = sqrt(((nBoys-1)*std(dataBoys)^2 + (nGirls-1)*std(dataGirls)^2)
          /(nBoys+nGirls-2));
sErr = sPooled*sqrt(1/nBoys + 1/nGirls);
tStatistic = (mean(dataBoys)-mean(dataGirls))/sErr;
pVal = 2*(1-cdf(TDist(nBoys + nGirls -2),abs(tStatistic)));
(sPooled,sErr,tStatistic,pVal)
```

Class Example 2. Two Sample t-test - un-equal variances

Use the same data (same seed) as in example 1. Again execute the following code and make sure that you understand the output.

```
# Perform two-sided t-test of the null hypothesis that the two
# populations from which the data come from have the same mean,
# and unequal variance, against the null that the population means
# and variances are different.
UnequalVarianceTTest(dataBoys,dataGirls)
```

```
# We now verify the StandardError, tStatistic, and p-value ourselves
# from the formulas.
sErr = sqrt(std(dataBoys)^2/nBoys + std(dataGirls)^2/nGirls);
tStatistic = (mean(dataBoys)-mean(dataGirls))/sErr;

# Weighted degrees of freedom
nu = sErr^4/(((std(dataBoys)^2/nBoys)^2)/(nBoys-1) +
    (std(dataGirls)^2/nGirls)^2/(nGirls-1));

pVal = 2*(1-cdf(TDist(nu),abs(tStatistic)));
(sErr, tStatistic,nu,pVal)
```

Class Example 3. A question from the 2016 sample exam

Question 3

You are investigating the wear of a bicycle chain using two types of lubricants. You obtain data for the wear using lubricant X on 12 bicycles and the wear using lubricant Y on 10 bicycles. The sample means are 8mm and 5mm respectively and the sample standard deviations are 2.3mm and 2.1mm .

(a) Assume you have reason to believe that the population variances of the wear are the same. Calculate the pooled sample standard deviation.

s_P= ____________

(b) Construct and interpret a 95% confidence interval for the difference in mean wear between the two lubricants.

Confidence Interval=____________________

(c) You wish to test whether the mean wear is the same for both lubricants. Carry out an hypothesis test for: $H_0 : \mu_X = \mu_Y$ vs. $H_A : \mu_X \neq \mu_Y$, assuming $\alpha = 0.01$, and state your conclusion.

p-value=____________________

conclusion: ____

____

____

Question 1. P-values with t

For the hypothesis test $H_0 : \mu = 7$ against $H_1 : \mu > 7$ with variance unknown and $n = 20$, approximate the P-value for each of the following test statistics (use the t-distribution table to get an approximation). Then compare your result to an exact p-value obtained from Julia.

(a) $t_0 = 2.05$

(b) $t_0 = -1.84$

(c) $t_0 = 0.4$

Question 2. Simple Hypothesis

Consider a normal population with mean μ and variance 4. Assume you are not sure if $\mu = 8$ or $\mu = 10$ so you devise a simple hypothesis test with $H_0 : \mu = 8$ and $H_1 : \mu = 10$. This is based on the sample mean, $\overline{X}$ taken over n observations. You reject if $\overline{X} > \tau$ and otherwise accept. Calculate the probabilities of type-I and type-II errors, denoted byα and β respectively. Do this for each of the following cases:

(a) $n = 1$ and $\tau = 9$.

(b) $n = 16$ and $\tau = 9.5$.

(c) $n = 25$ and $\tau = 9.5$.

(d) $n = 36$ and $\tau = 9.5$.

Question 3. Steel Rods

The diameter of steel rods manufactured on two different extrusion machines is being investigated. Two random samples of sizes $n_1 = 15$ and $n_1 = 17$ are selected, and the sample means and sample variances are $\overline{x_1} = 8.73, s_1^2 = 0.35, \overline{x_2} = 8.68$ and $s_2^2 = 0.40$, respectively. Assume that $\sigma_1^2 = \sigma_2^2$ and that the data are drawn from a normal distribution.

(a) Is there evidence to support the claim that the two machines produce rods with different mean diameters? Use $\alpha = 0.05$ in arriving at this conclusion. Find the P-value.

(b) Construct a 95% confidence interval for the difference in mean rod diameter. Interpret this interval.

Question 4. Wet Chemical Etching

In semiconductor manufacturing, wet chemical etching is often used to remove silicon from the backs of wafers prior to metallization. The etch rate is an important characteristic in this process and known to follow a normal distribution. Two different etching solutions have been compared using two random samples of 10 wafers for each solution. The observed etch rates are as follows (in mils per minute):

Solultion 1		**Solution 2**	
9.9	10.63	10.2	10.0
9.4	10.3	10.6	10.2
9.3	10.0	10.7	10.7
9.6	10.3	10.4	10.4
10.2	10.1	10.5	10.3

(a) Construct normal probability plots for the two samples. Do these plots provide support for the assumptions of normality and equal variances? Write a practical interpretation for these plots.

(b) Do the data support the claim that the mean etch rate is the same for both solutions? In reaching your conclusions, use $\alpha = 0.05$ and assume that both population variances are equal. Calculate a P-value.

(c) Find a 95% confidence interval on the difference in mean etch rates

Question 5. Gold Ball Distance

The overall distance travelled by a golf ball is tested by hitting the ball with Iron Byron, a mechanical golfer with a swing that is said to emulate the distance hit by the legendary champion Byron Nelson. Ten randomly selected balls of two different brands are tested and the overall distance measured. The data is as follows:

Brand 1: 275, 286, 287, 271, 283, 271, 279, 275, 263, 267
Brand 2: 258, 244, 260, 265, 273, 281, 271, 270, 263, 268

(a) Is there evidence that overall distance is approximately normally distributed?

(b) Test the hypothesis that both brands of ball have equal mean overall distance. Use $\alpha = 0.05$. What is the P-value?

(c) Construct a 95% two-sided CI on the mean difference in overall distance for the two brands of golf balls.

Question 6. Computer Output

Consider the following computer output:

Two-Sample T-Test and CI

Sample	N	Mean	StDev	Se Mean
1	12	16	1.26	0.36
2	16	12.15	1.99	0.50

Difference = mu(1) - mu(2)
Estimate for difference: -1.210
95% CI for difference: (-2.560, 0.140)
T-test of difference = 0 (vs not =):
T-value = ?
P-value = ?
DF = ?
Both used Pooled StDev = ?

(a) Fill in the missing values. Is this a one-sided or a two-sided test? Use lower and upper bounds for the P-value.

(b) What are your conclusions if $\alpha = 0.05$ What if $\alpha = 0.01$?

(c) This test was done assuming that the two population variances were equal. Does this seem reasonable?

(d) Suppose that the hypothesis had been $H_0 : \mu_1 = \mu_2$ vs. $H_0 : \mu_1 < \mu_2$. What would your conclusions be if $\alpha = 0.05$?

Question 7. Melting Point

The melting points of two alloys used in formulating a solder were investigated by melting 16 samples of each material. The sample mean and standard deviation for alloy 1 was $\overline{x_1} = 420\,\deg F$ and $s_1 = 3.5\,\deg F$. For alloy 2, they were $\overline{x_2} = 416\,\deg F$ and $s_2 = 3.2\,\deg F$.

(a) Does the sample data support the claim that both alloys have the same melting point? Use $\alpha = 0.05$ and assume that both populations are normally distributed and have the same standard deviation. Find the P-value for the test.

(b) Suppose that the true mean difference in melting points is 3 $degF$. how large a sample would be required to detect this difference using an $\alpha = 0.05$ level test with probability at least 0.9? Use $\sigma_1 = \sigma_2 = 4$ as an initial estimate of the common standard deviation.

Question 8. More Simple Hypothesis

Let X be exponentially distributed with parameter λ. Assume you are sampling a single observation, X, and wish to carry out a simple hypothesis test with $H_0 : \lambda = 5.0$ and $H_1 : \lambda = 2.0$. Your test rejects H_0 if $X > \tau$.

(a) Plot the pdfs for the two distributions (two values of λ), above each other. Argue why it is sensible to reject when $X > \tau$.

(b) Assume $\tau = 0.75$. Calculate the α and β (the probabilities of type-I and type-II errors respectively).

(c) What would you use for the value of τ if you wish for α to be 0.05. In this case what is β?

(d) What would you use for the value of τ if you wish to have an equal value of α and β?