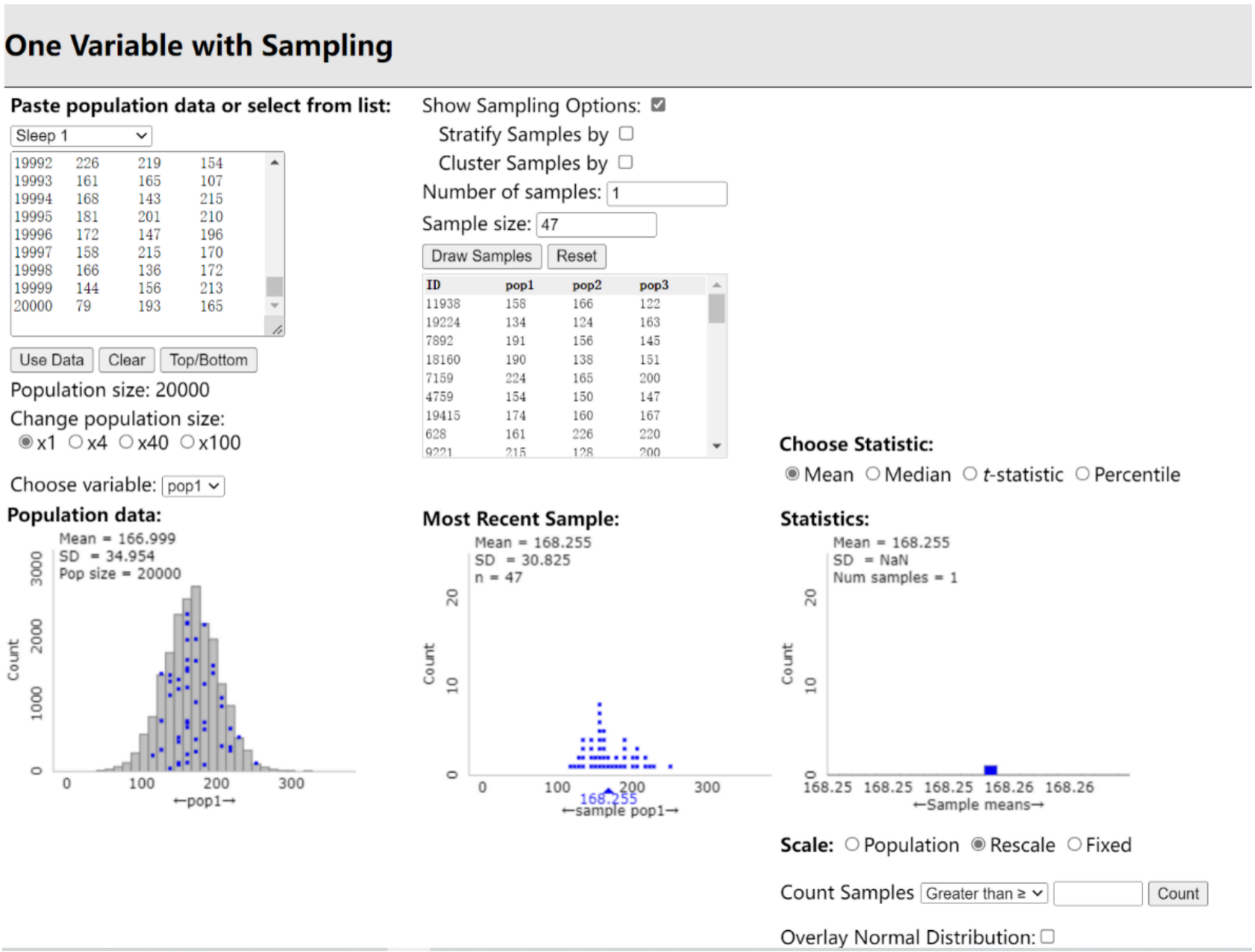
Chapter 2 Summary

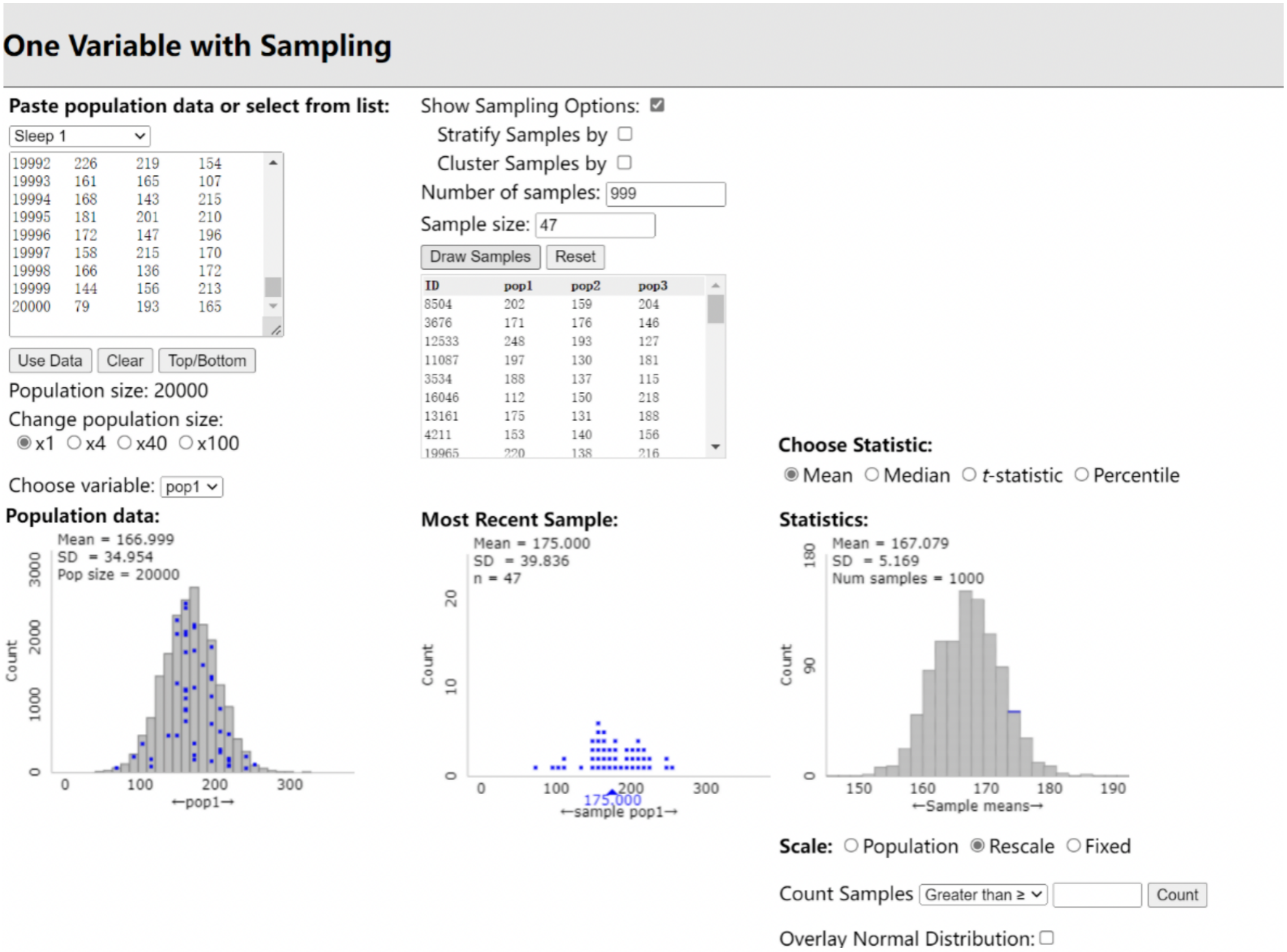
Tina Wang, Paul Luo, Jackson Cong

Investigation 2.4: Population Mean

- Sampling from Finite Population applet (Normal Population)

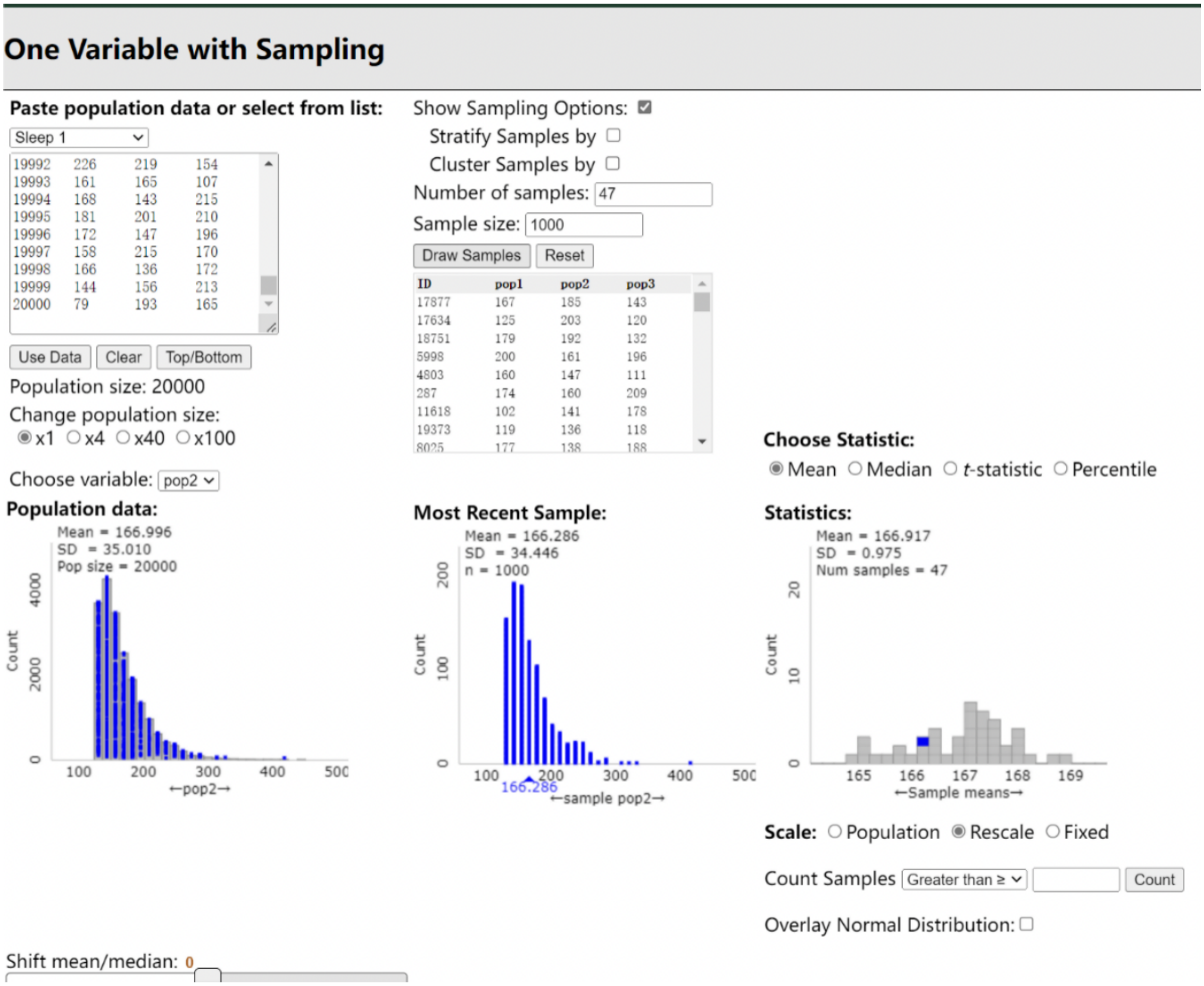


(1 Number of samples, 47 Sample size)

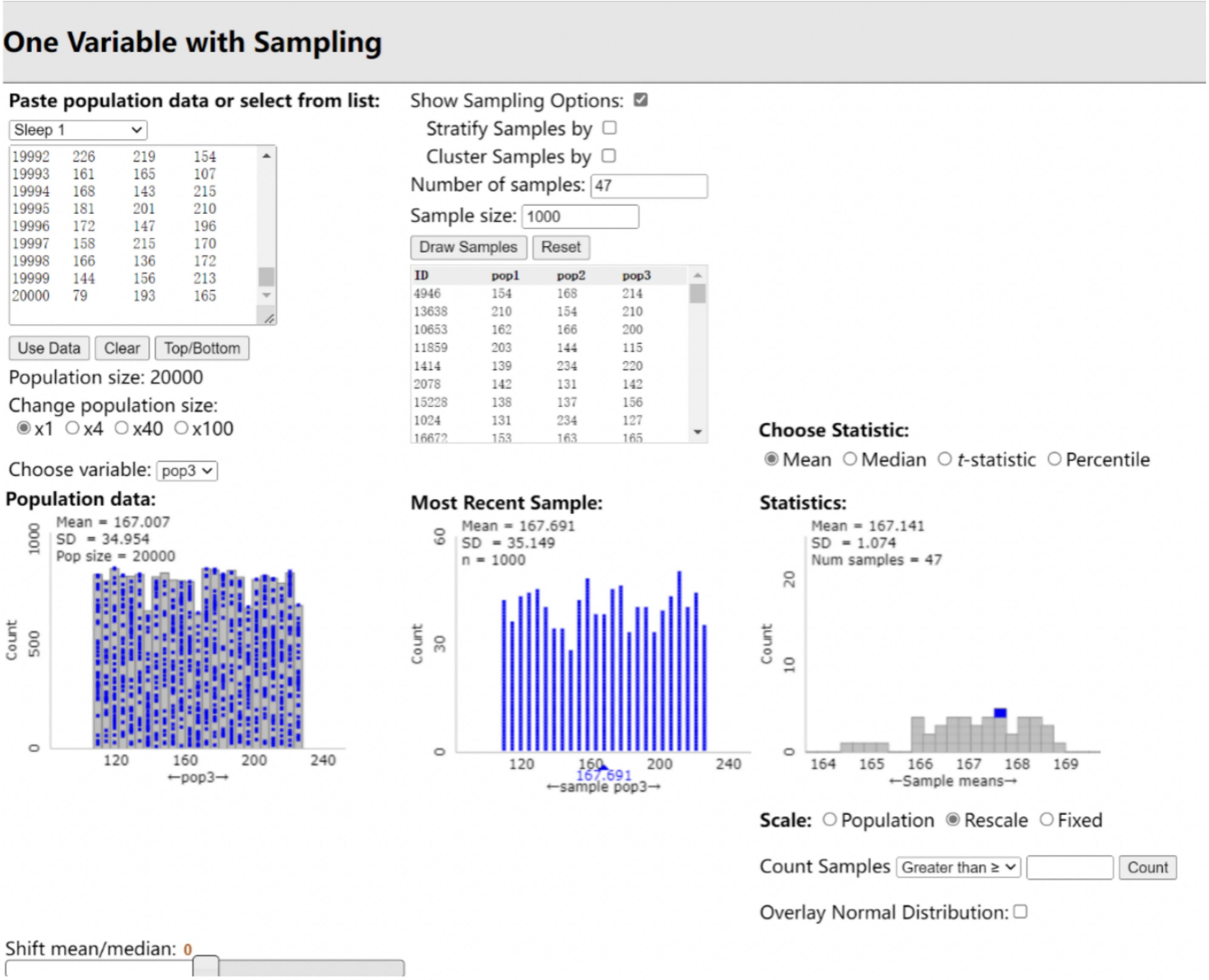


(1000 Number of samples, 47 Sample size)

- Non-normal Population



(Skewed to right)



(Uniform)

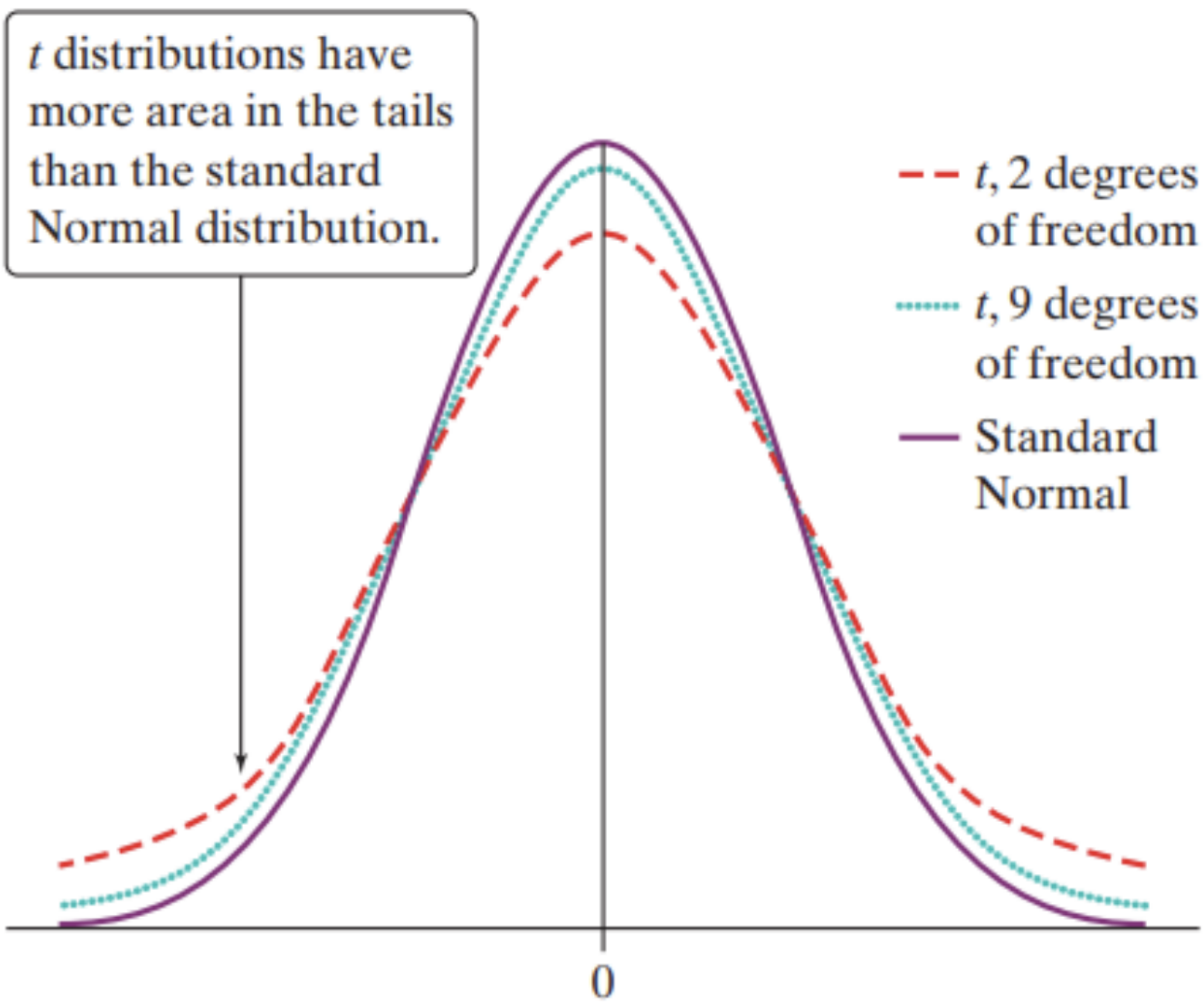
- The sample data come from a well-designed random population or randomized experiment of size n with mean μ and standard deviation, the sampling distribution of sample means has following characters:
    - **Shape**: If the population distribution is **normal**, then so is the sampling distribution of $\bar{x}$; If the population distribution **isn't normal**, the sampling distribution of $\bar{x}$ will be approximately **normal** if the sample size is large enough by the central limit theorem (CLT).
    - **Center**: $\mu_{\bar{x}} = \mu$ (The sampling mean equals to the population mean)
    - **Spread**: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (The sampling standard deviation equal to $\frac{\sigma}{\sqrt{n}}$ of the population SD)
- **Central Limit Theorem**: The population has a normal distribution or the sample size is large (n ≥ 30), the shape of sampling distribution will be normal. If the population distribution has unknown shape and n < 30, use a graph of the sample data to assess the Normality of the population. Do not use t procedures if the graph shows strong skewness or outliers.

Investigation 2.5: t-test

- Standard deviation of statistic: $SE(\bar{x}) = \frac{s}{\sqrt{n}}$
- To test the hypothesis $H_0 : \mu = \mu_0$, compute the one-sample t statistic.

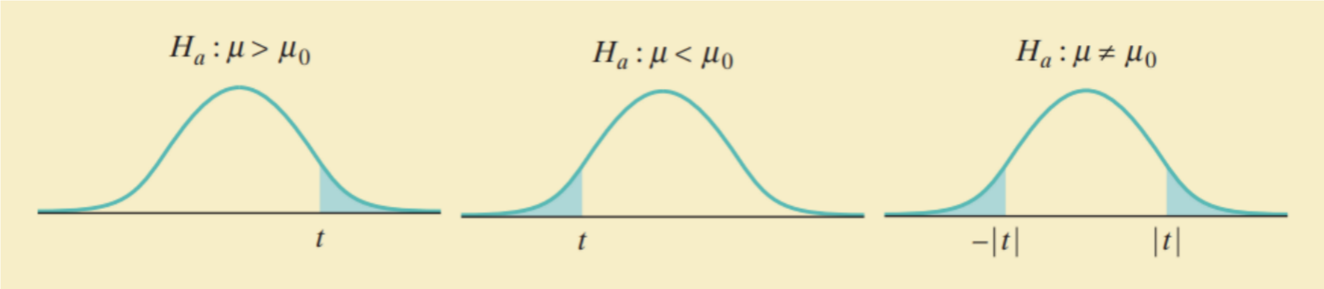
$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

- When we perform inference about a population mean m using a t distribution, the appropriate degrees of freedom are found by subtracting 1 from the sample size n, making df (degrees of freedom) = n – 1.



(t distribution with 2 and 9 degrees of freedom, and all are symmetric with center 0)

- Then find the P-value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_a in a t distribution with df= n – 1:

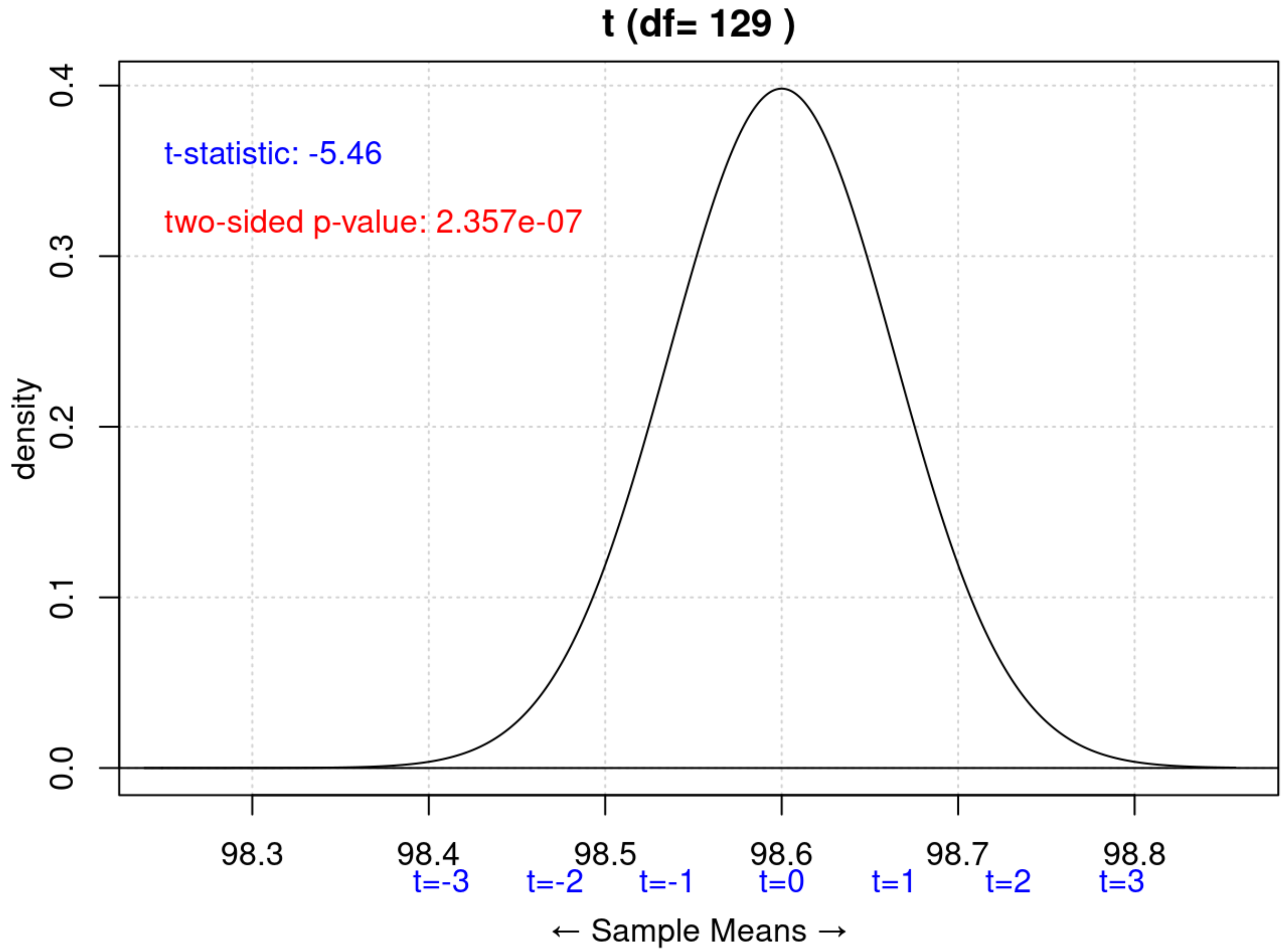


- **P-Value**: The P-value is based on a t-distribution with n 1 degrees of freedom. This value can be estimated in R.

- In R:
  - t.test(x, mu = μ_0, alternative = “two.sided”, conf.level = .95)
  - iscamonesamplet(xbar, sample standard deviation(s), n, hypothesized, alternative, conf.level)

```
iscamonesamplet(xbar = 98.249, sd = .733, n = 130, hypothesized = 98.6, alter
native = "two.sided", conf.level = .95)
```

```
##
## One Sample t test
##
## mean = 98.249, sd = 0.733,  sample size = 130
## Null hypothesis        : mu = 98.6
## Alternative hypothesis: mu <> 98.6
## t-statistic: -5.46
```




```
## 95 % Confidence interval for mu: ( 98.1218 ,  98.3762 )
## p-value: 2.35666e-07
```

- If $P < \alpha$, then Reject the H_0, otherwise Fail to Reject H_0.
- **Confidence Interval for μ.** (Check the percentage of confidence interval to corresponding t-value)
  - $\bar{x} \pm t^*_{n-1} \times (\frac{s}{\sqrt{n}})$
- In R: t.test(x, y = NULL, alternative = ("two.sided", "less", "greater"), mu = 0, conf.level = 0.95)

```
B3FD <- read.csv("~/math247/Project/B3FDAgeGender.csv")
age_diff <- abs(B3FD$difference)
t.test(age_diff, mu = 3, alt = "greater")
```

```
##
##  One Sample t-test
##
## data:  age_diff
## t = -3.3465, df = 498, p-value = 0.9996
## alternative hypothesis: true mean is greater than 3
## 95 percent confidence interval:
##  2.560345      Inf
## sample estimates:
## mean of x
##  2.705411
```

```
t.test(age_diff, mu = 3, conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  age_diff
## t = -3.3465, df = 498, p-value = 0.0008803
## alternative hypothesis: true mean is not equal to 3
## 95 percent confidence interval:
##  2.532456 2.878366
## sample estimates:
## mean of x
##  2.705411
```