

Chapter 2 Summary

Tina Wang, Paul Luo, Jackson Cong

Investigation 2.4: Population Mean

- Sampling from Finite Population applet (Normal Population)

One Variable with Sampling

Paste population data or select from list:

Sleep 1

19992	226	219	154
19993	161	165	107
19994	168	143	215
19995	181	201	210
19996	172	147	196
19997	158	215	170
19998	166	136	172
19999	144	156	213
20000	79	193	165

Use Data Clear Top/Bottom

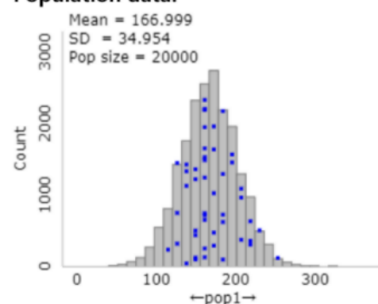
Population size: 20000

Change population size:

☒ x1 ☐ x4 ☐ x40 ☐ x100

Choose variable: pop1

Population data:



Show Sampling Options: ☒

Stratify Samples by ☐

Cluster Samples by ☐

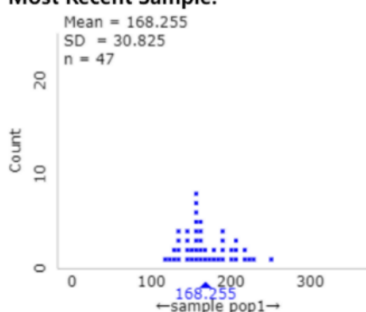
Number of samples: 1

Sample size: 47

Draw Samples Reset

ID	pop1	pop2	pop3
11938	158	166	122
19224	134	124	163
7892	191	156	145
18160	190	138	151
7159	224	165	200
4759	154	150	147
19415	174	160	167
628	161	226	220
9221	215	128	200

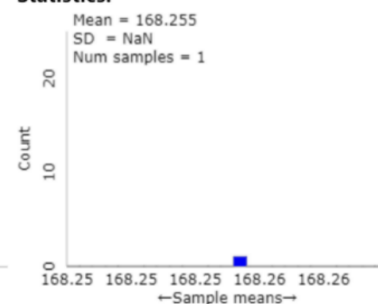
Most Recent Sample:



Choose Statistic:

☒ Mean ☐ Median ☐ t-statistic ☐ Percentile

Statistics:



Scale: ☐ Population ☒ Rescale ☐ Fixed

Count Samples Greater than \geq Count

Overlay Normal Distribution: ☐

(1 Number of samples, 47 Sample size)

One Variable with Sampling

Paste population data or select from list:

Sleep 1

19992	226	219	154
19993	161	165	107
19994	168	143	215
19995	181	201	210
19996	172	147	196
19997	158	215	170
19998	166	136	172
19999	144	156	213
20000	79	193	165

Use Data Clear Top/Bottom

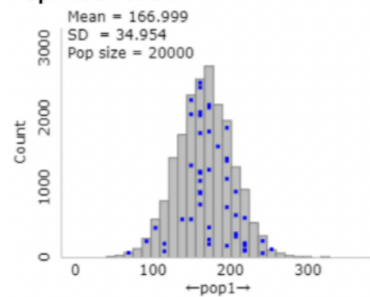
Population size: 20000

Change population size:

☒ x1 ☐ x4 ☐ x40 ☐ x100

Choose variable: pop1

Population data:



Show Sampling Options: ☒

Stratify Samples by ☐

Cluster Samples by ☐

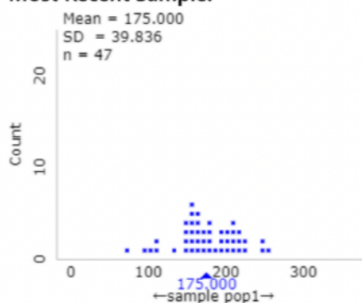
Number of samples: 999

Sample size: 47

Draw Samples Reset

ID	pop1	pop2	pop3
8504	202	159	204
3676	171	176	146
12533	248	193	127
11087	197	130	181
3534	188	137	115
16046	112	150	218
13161	175	131	188
4211	153	140	156
19965	220	138	216

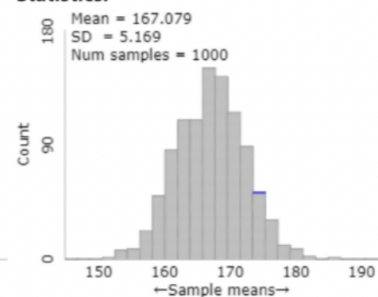
Most Recent Sample:



Choose Statistic:

☒ Mean ☐ Median ☐ t-statistic ☐ Percentile

Statistics:



Scale: ☐ Population ☒ Rescale ☐ Fixed

Count Samples Greater than \geq Count

Overlay Normal Distribution: ☐

(1000 Number of samples, 47 Sample size)

- Non-normal Population

One Variable with Sampling

Paste population data or select from list:

Sleep 1

19992	226	219	154
19993	161	165	107
19994	168	143	215
19995	181	201	210
19996	172	147	196
19997	158	215	170
19998	166	136	172
19999	144	156	213
20000	79	193	165

Use Data Clear Top/Bottom

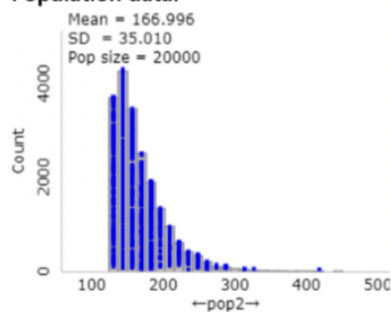
Population size: 20000

Change population size:

☒ x1 ☐ x4 ☐ x40 ☐ x100

Choose variable: pop2

Population data:



Show Sampling Options: ☒

Stratify Samples by ☐

Cluster Samples by ☐

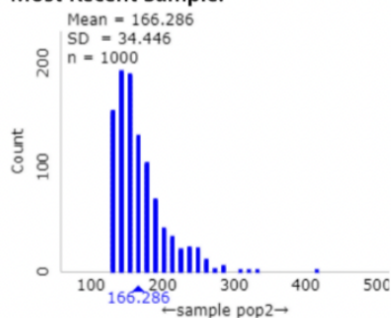
Number of samples: 47

Sample size: 1000

Draw Samples Reset

ID	pop1	pop2	pop3
17877	167	185	143
17634	125	203	120
18751	179	192	132
5998	200	161	196
4803	160	147	111
287	174	160	209
11618	102	141	178
19373	119	136	118
8025	177	138	188

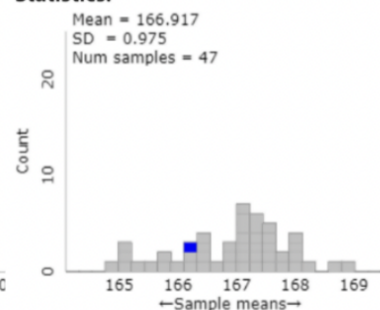
Most Recent Sample:



Choose Statistic:

☒ Mean ☐ Median ☐ t-statistic ☐ Percentile

Statistics:



Scale: ☐ Population ☒ Rescale ☐ Fixed

Count Samples Greater than \geq Count

Overlay Normal Distribution: ☐

Shift mean/median: 0

(Skewed to right)

One Variable with Sampling

Paste population data or select from list: Sleep 1

19992	226	219	154
19993	161	165	107
19994	168	143	215
19995	181	201	210
19996	172	147	196
19997	158	215	170
19998	166	136	172
19999	144	156	213
20000	79	193	165

Use Data Clear Top/Bottom

Population size: 20000
Change population size: ☒ x1 ☐ x4 ☐ x40 ☐ x100

Choose variable: pop3

Population data:

Mean = 167.007
SD = 34.954
Pop size = 20000

Show Sampling Options: ☒
Stratify Samples by ☐
Cluster Samples by ☐
Number of samples: 47
Sample size: 1000
Draw Samples Reset

ID	pop1	pop2	pop3
4946	154	168	214
13638	210	154	210
10653	162	166	200
11859	203	144	115
1414	139	234	220
2078	142	131	142
15228	138	137	156
1024	131	234	127
16672	153	163	165

Most Recent Sample:

Mean = 167.691
SD = 35.149
n = 1000

Choose Statistic: ☒ Mean ☐ Median ☐ t-statistic ☐ Percentile

Statistics:

Mean = 167.141
SD = 1.074
Num samples = 47

Scale: ☐ Population ☒ Rescale ☐ Fixed

Count Samples Greater than ≥ Count

Overlay Normal Distribution: ☐

Shift mean/median: 0

(Uniform)

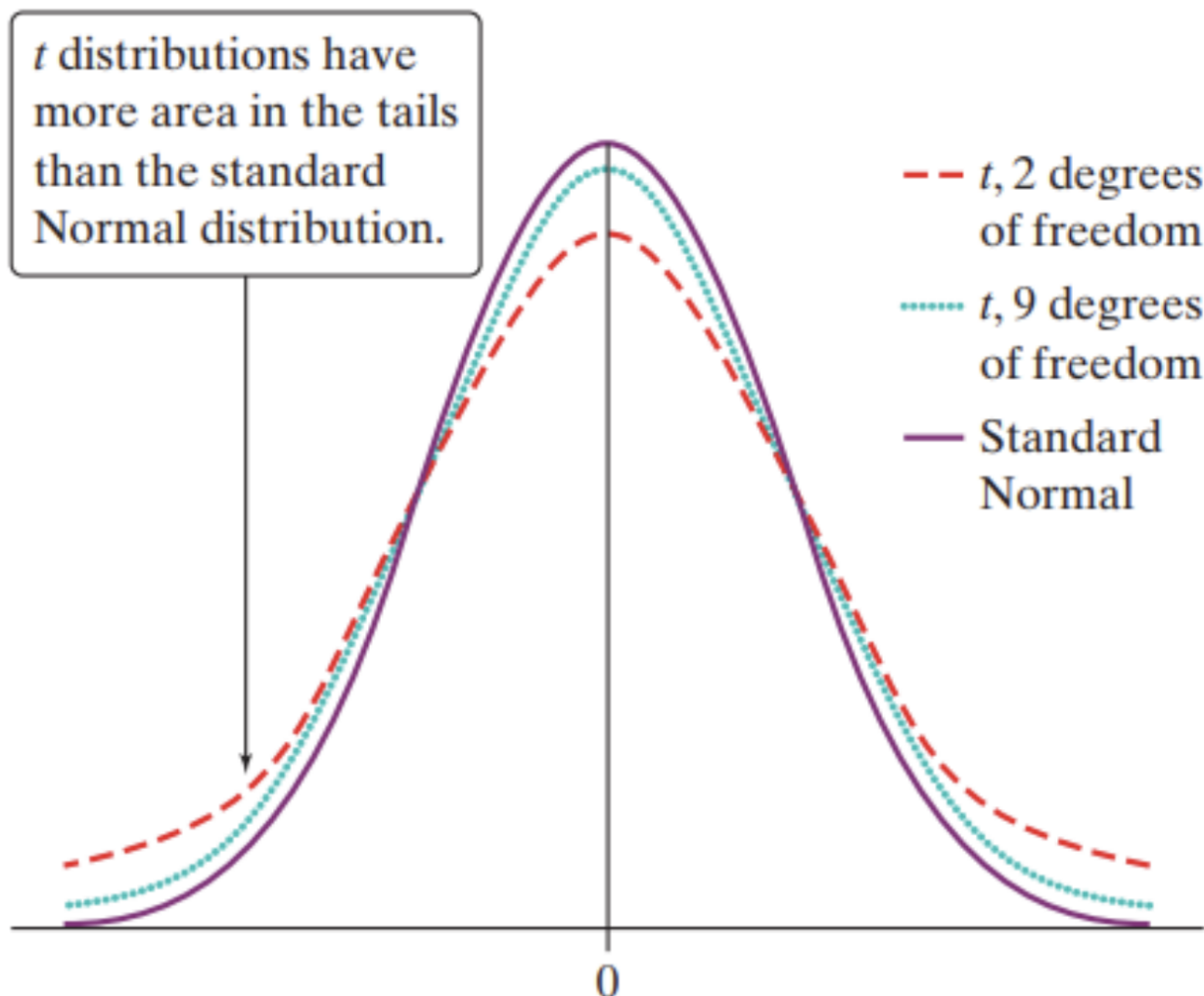
- The sample data come from a well-designed random population or randomized experiment of size n with mean μ and standard deviation, the sampling distribution of sample means has following characters:
 - Shape:** If the population distribution is **normal**, then so is the sampling distribution of \bar{x} ; 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 ($n \geq 30$), 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.

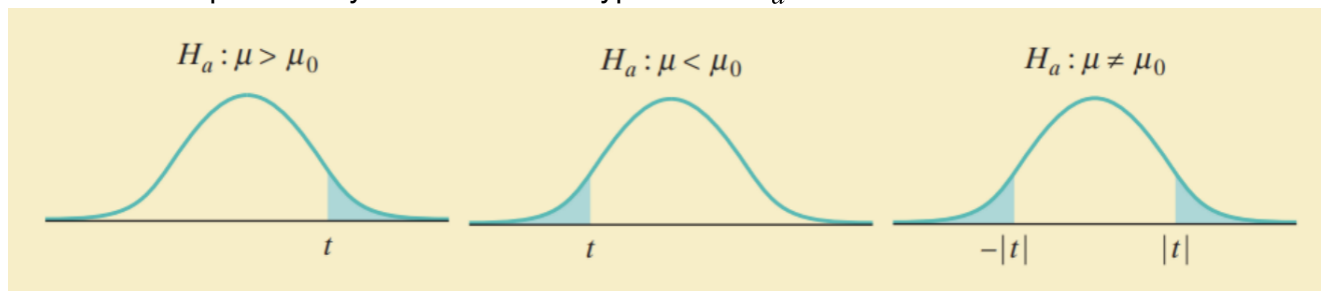
$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

- When we perform inference about a population mean μ 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)

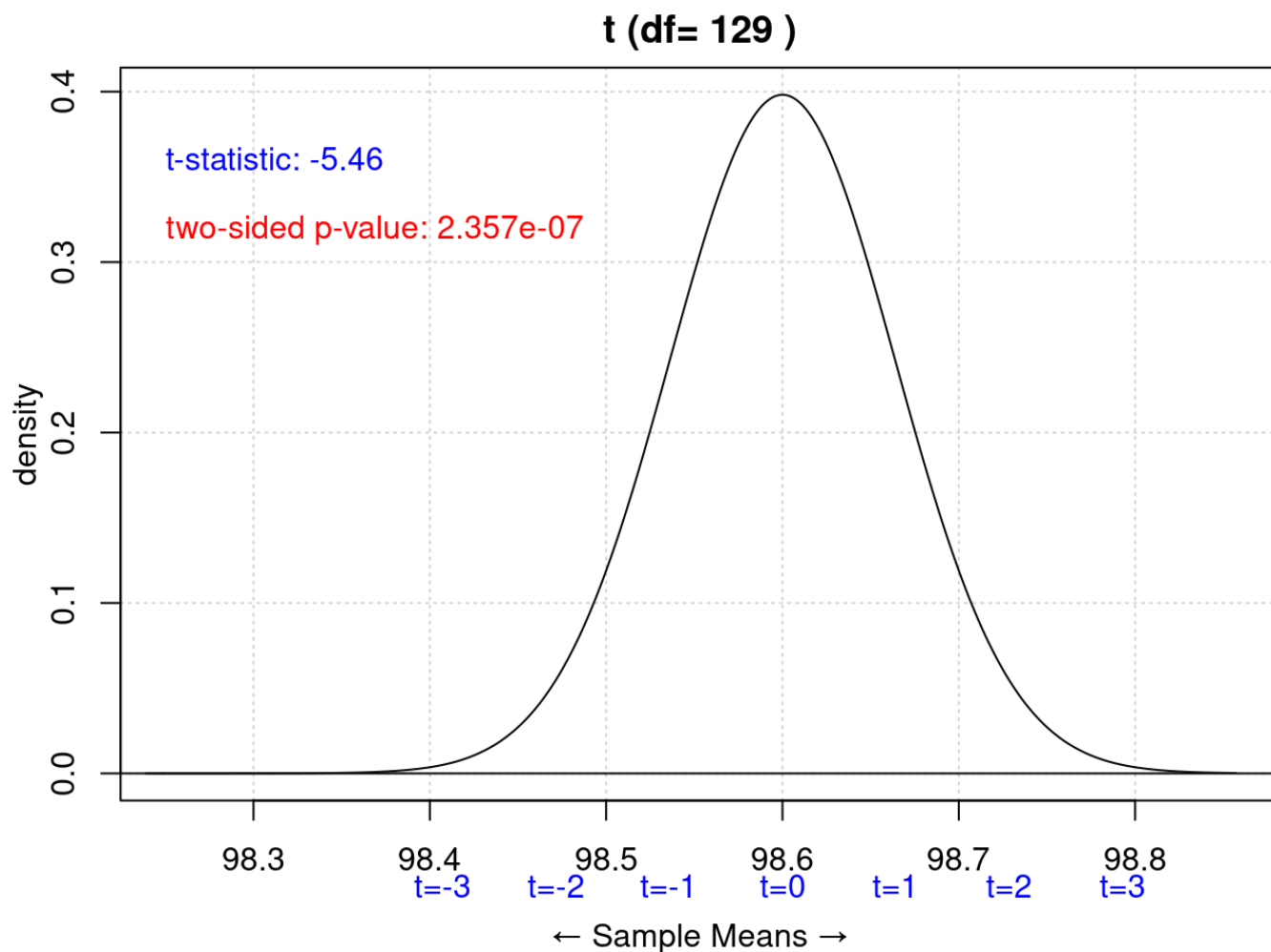
- Then find the P-value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_a in a t distribution with $df = n - 1$:



- **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, alternative = "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)
 - $\bar{x} \pm t_{n-1}^* \times \left(\frac{s}{\sqrt{n}} \right)$
- 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
```