Student t Distribution

ECE 341: Random Processes

Lecture #23

note: All lecture notes, homework sets, and solutions are posted on www.BisonAcademy.com

Student t Distribution

t-Test: Comparison of Means.

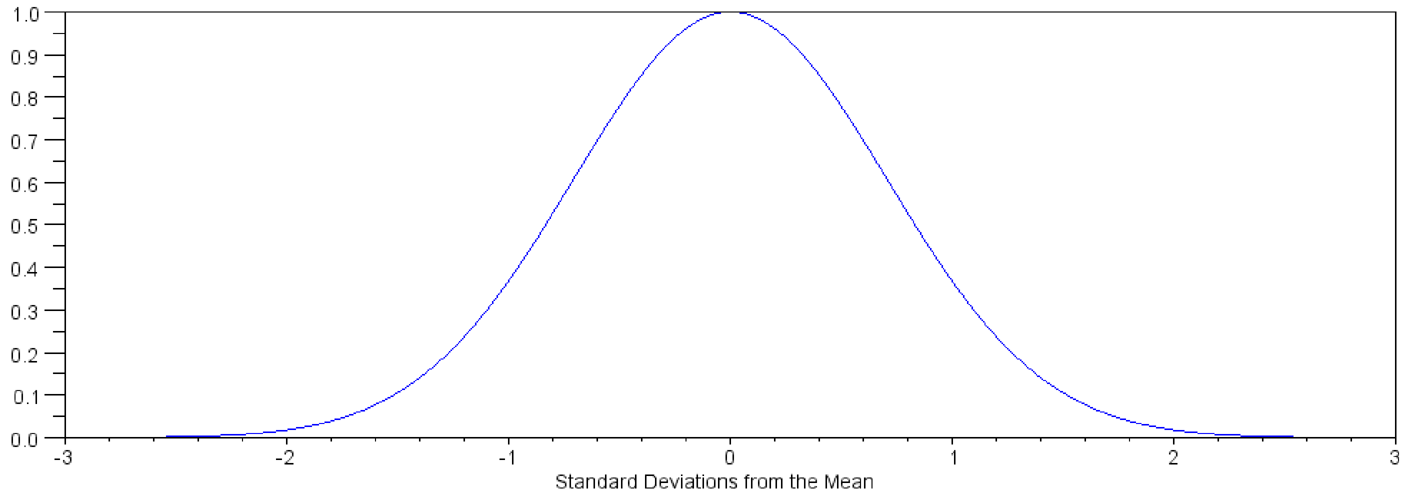
- What is the mean of rolling 10d6?
- What is the probability that 10d6 > 44.5?
- What is the probability of drawing 3-of-a-kind in 5-card stud?
- What is the probability that April will break 90F this year?
- What is the 90% confidence interval for the hottest it will get this coming April?
- How much energy is in a AA battery?
- What is the chance that a given AA battery has more than 500mAh?

Central Limit Theorem

- All distributions converge to a Normal distribution
- Normal + Normal = Normal

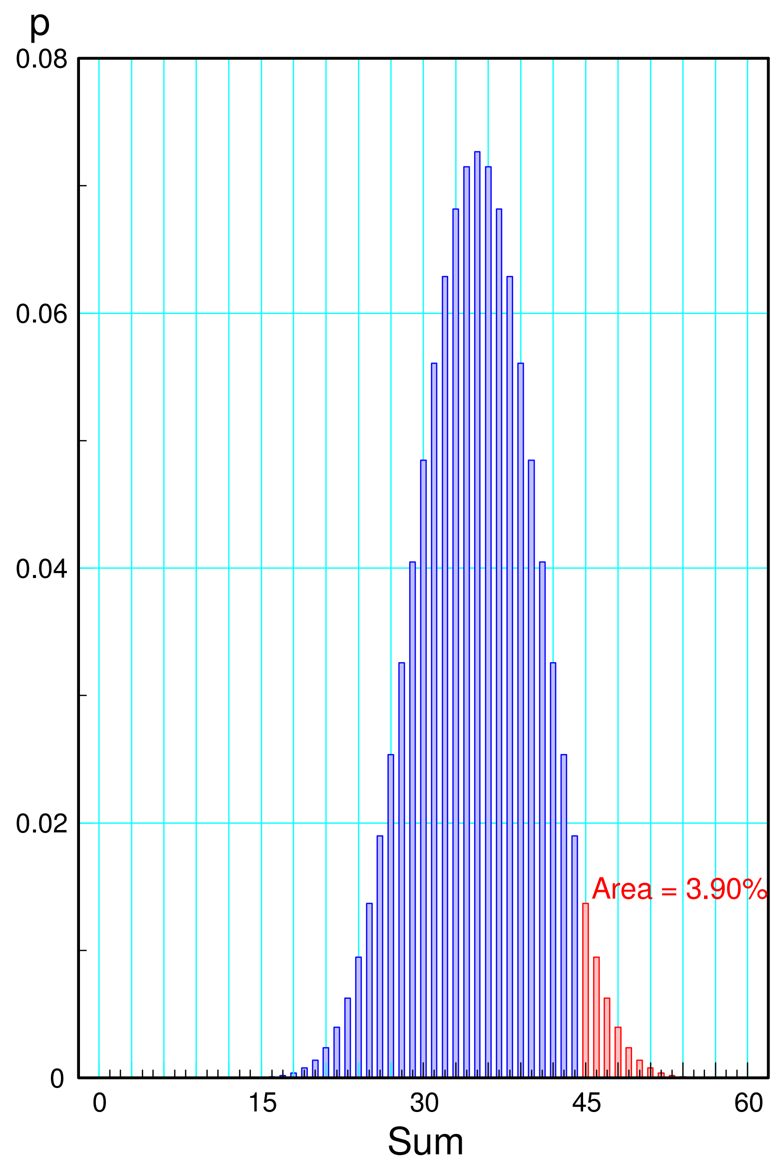
A Normal distribution is defined by two parameters: $N(\mu, \sigma^2)$

- μ: The mean (average) value
- σ: The standard deviation (a measure of the spread)
- Standard Normal: Mean = 0, Standard Deviation = 1



Example: Dice

- Roll 10 6-sided dice (10d6)
- mean = ?
- p(rolling 45 or higher) = ?



Exact Solution: Solve using convolution

```
>> d6 = [0,1,1,1,1,1,1]' / 6;
>> d6x2 = conv(d6,d6);
>> d6x4 = conv(d6x2,d6x2);
>> d6x8 = conv(d6x4,d6x4);
>> d6x10 = conv(d6x8,d6x2);

>> sum(N .* d6x10)
ans =    35.0000

>> sum(d6x10(46:61))
ans =     0.0390
```

Solution: Use the Central Limit Theorem

- d6: $\mu = 3.5$ $\sigma^2 = 2.917$
- 10d6: $\mu = 35$ $\sigma^2 = 29.17$

z-score:

$$z = \left(\frac{44.5-\mu}{\sigma}\right) = \left(\frac{44.5-35}{5.401}\right) = 1.7589$$

Normal Distributiont able converts z-score to a probability

Normal Distribution (area of tail)									
0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

StatTrek also works

- p = 3.9%

Student t-Distribution (t-test)

Similar to a Normal distribution & z-score

$$z = \left(\frac{x-\mu}{\sigma}\right)$$

Replace the mean and variance with their estimates

- $\mu \rightarrow \bar{x} = \frac{1}{n}\sum x_i$ *normal distribution*
- $\sigma \rightarrow s = \frac{1}{n-1}\sum (x_i - \bar{x})^2$ *gamma distribution*
- $t = \left(\frac{x-\bar{x}}{s}\right)$ *t-score: student t distribution*
- n-1 *degrees of freedom*

Note: Individual vs. Population

- $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ *divide the variance by n when talking about populations*
- $s \rightarrow \left(\frac{s}{\sqrt{n}}\right)$ *when talking about the population (average of data)*

A Student-t Table looks like the following:

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	1	1.38	1.96	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	0.82	1.06	1.39	1.89	2.92	4.3	6.97	9.93	22.33	31.6
3	0.77	0.98	1.25	1.64	2.35	3.18	4.54	5.84	10.22	12.92
4	0.74	0.94	1.19	1.53	2.13	2.78	3.75	4.6	7.17	8.61
5	0.73	0.92	1.16	1.48	2.02	2.57	3.37	4.03	5.89	6.87
10	0.7	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Student t-Table. Note that as the sample size goes to infinity (p = infinity), you converge to a Normal distribution.

Example: Rolling Dice

Roll 10d6

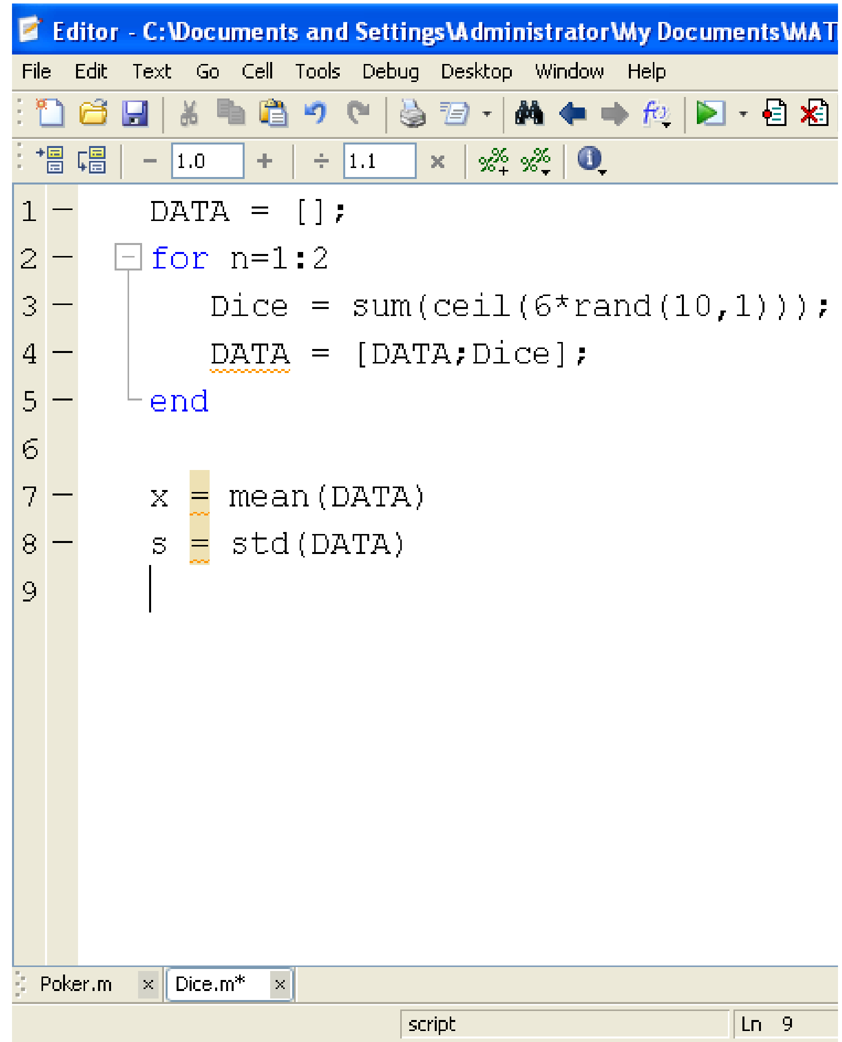
```
>> sum(ceil(6*rand(10,1)))

ans =     42
```

Result: meaningless

- You need at least 2 measurements

Roll 10d6 two times

```
>> sum(ceil(6*rand(10,1)))

ans =     29

>> sum(ceil(6*rand(10,1)))

ans =     38

>> DATA = [29,38];
>> x = mean(DATA)

x =   33.5000

>> s = std(DATA)

s =    6.364
```



Individual Tests (2 die rolls)

What is the 90% confidence interval for my next roll?

- Spread does not depend upon sample size
- 1 degree of freedom, 5% tails >> t-score = 6.31

$$\bar{x} - 6.31s < roll < \bar{x} + 6.31s$$

$$-6.0258 < roll < 73.0258$$

What is the chance my next roll will be more than 44.5?

$$t = \left(\frac{44.5 - \bar{x}}{s}\right) = 1.7285$$

p = 16.69%

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	1	1.38	1.96	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	0.82	1.06	1.39	1.89	2.92	4.3	6.97	9.93	22.33	31.6
3	0.77	0.98	1.25	1.64	2.35	3.18	4.54	5.84	10.22	12.92

Population Tests (2 die rolls)

What is the 90% confidence interval for the population's mean?

- 1 degree of freedom, 5% tails >> t-score = 6.31

$$\bar{x} - 6.31\left(\frac{s}{\sqrt{2}}\right) < \mu < \bar{x} + 6.31\left(\frac{s}{\sqrt{2}}\right)$$

$$5.10 < \mu < 61.89$$

What is the chance that the population's mean is more than 39.5?

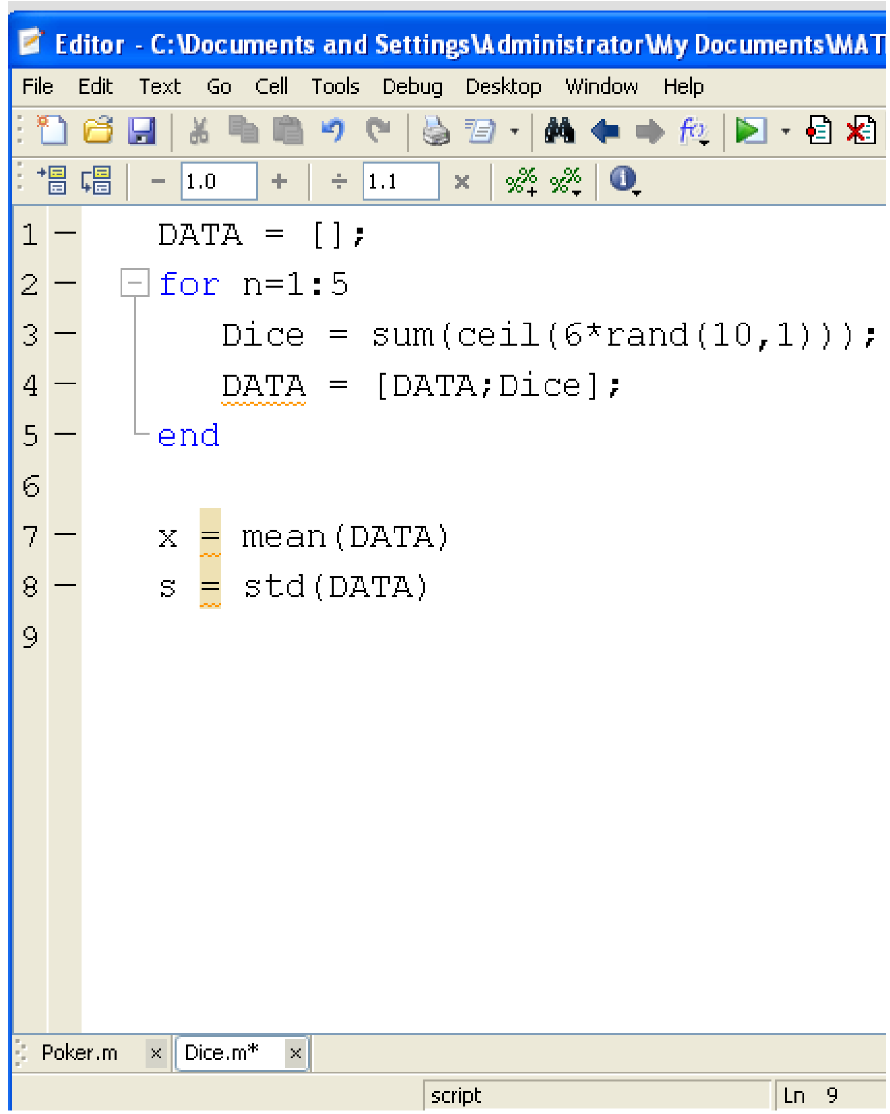
$$t = \left(\frac{39.5 - \bar{x}}{\left(\frac{s}{\sqrt{2}}\right)}\right) = 1.3333$$

p = 20.48%

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	1	1.38	1.96	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	0.82	1.06	1.39	1.89	2.92	4.3	6.97	9.93	22.33	31.6
3	0.77	0.98	1.25	1.64	2.35	3.18	4.54	5.84	10.22	12.92

Roll the dice 5 times

```
DATA = [];
for n=1:5
    Dice = sum(ceil(6*rand(10,1)));
    DATA = [DATA;Dice];
    end
x = mean(DATA)
s = std(DATA)

x =   34.4000

s =    4.6690
```



Individual Tests (5 die rolls)

What is the 90% confidence interval for my next roll?

- 4 degrees of freedom, 5% tails >> t-score = 2.13

$$\bar{x} - 2.13s < roll < \bar{x} + 2.13s$$

$$24.455 < roll < 44.345$$

What is the chance my next roll will be more than 44.5?

$$t = \left(\frac{44.5 - \bar{x}}{s}\right) = 2.1632$$

p = 4.83%

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	1	1.38	1.96	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	0.82	1.06	1.39	1.89	2.92	4.3	6.97	9.93	22.33	31.6
3	0.77	0.98	1.25	1.64	2.35	3.18	4.54	5.84	10.22	12.92
4	0.74	0.94	1.19	1.53	2.13	2.78	3.75	4.6	7.17	8.61
5	0.73	0.92	1.16	1.48	2.02	2.57	3.37	4.03	5.89	6.87

Population Tests (5 die rolls)

What is the 90% confidence interval for the population's mean?

- 4 degrees of freedom, 5% tails → t-score = 2.13

$$\bar{x} - 2.13\left(\frac{s}{\sqrt{5}}\right) < \mu < \bar{x} + 2.13\left(\frac{s}{\sqrt{5}}\right)$$

$$29.953 < \mu < 38.747$$

What is the chance that the population's mean is more than 39.5?

$$t = \left(\frac{39.5 - \bar{x}}{\left(\frac{s}{\sqrt{5}}\right)}\right) = 2.4425$$

p = 3.55%

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
1	1	1.38	1.96	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	0.82	1.06	1.39	1.89	2.92	4.3	6.97	9.93	22.33	31.6
3	0.77	0.98	1.25	1.64	2.35	3.18	4.54	5.84	10.22	12.92
4	0.74	0.94	1.19	1.53	2.13	2.78	3.75	4.6	7.17	8.61
5	0.73	0.92	1.16	1.48	2.02	2.57	3.37	4.03	5.89	6.87

Roll the dice 21 times

```
DATA = [];
for n=1:1001
   Dice = sum(ceil(6*rand(10,1)));
   DATA = [DATA;Dice];
   end
x = mean(DATA)
s = std(DATA)

x =   34.0476
s =    4.0924
```

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
10	0.7	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Individual Tests (21 die rolls)

What is the 90% confidence interval for my next roll?

- 20 degrees of freedom, 5% tails >> t-score = 1.73

$$\bar{x} - 1.73s < roll < \bar{x} + 1.73s$$

$$26.968 < roll < 41.127$$

What is the chance my next roll will be more than 44.5?

$$t = \left(\frac{44.5 - \bar{x}}{s}\right) = 2.5541$$

p = 0.95%

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
10	0.7	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Population Tests (21 die rolls)

What is the 90% confidence interval for the population's mean?

- 20 degrees of freedom, 5% tails >> t-score = 1.73

$$\bar{x} - 1.73\left(\frac{s}{\sqrt{21}}\right) < \mu < \bar{x} + 1.73\left(\frac{s}{\sqrt{21}}\right)$$

$$32.503 < \mu < 35.593$$

What is the chance that the population's mean is more than 39.5?

$$t = \left(\frac{39.5 - \bar{x}}{\left(\frac{s}{\sqrt{21}}\right)}\right) = 6.1055$$

$p \approx 0\%$ *(off scale)*

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
10	0.7	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Roll the dice 1001 times

```
DATA = [];
for n=1:1001
   Dice = sum(ceil(6*rand(10,1)));
   DATA = [DATA;Dice];
   end
x = mean(DATA)
s = std(DATA)

x =   34.9710
s =    5.4608
```

At this point, the t-distribution converges to the Normal distribution

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
1,000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Individual Tests (1001 die rolls)

What is the 90% confidence interval for my next roll?

- 1000 degrees of freedom, 5% tails >> t-score = 1.646 (1.645 for a Normal dist)

$$\bar{x} - 1.646s < roll < \bar{x} + 1.646s$$

$$25.982 < roll < 43.959$$

What is the chance my next roll will be more than 44.5?

$$t = \left(\frac{44.5 - \bar{x}}{s}\right) = 1.7450$$

p = 4.06%

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
1,000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Population Tests (1001 die rolls)

What is the 90% confidence interval for the population's mean?

- 1000 degrees of freedom, 5% tails >> t-score = 1.646 (1.645 for a Normal dist)

$$\bar{x} - 1.646\left(\frac{s}{\sqrt{1001}}\right) < \mu < \bar{x} + 1.646\left(\frac{s}{\sqrt{1001}}\right)$$

$$34.687 < \mu < 35.255$$

What is the chance that the population's mean is more than 39.5?

$$t = \left(\frac{39.5 - \bar{x}}{\left(\frac{s}{\sqrt{21}}\right)}\right) = 26.24$$

$p \approx 0$ *(off scale)*

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
20	0.69	0.86	1.06	1.33	1.73	2.09	2.53	2.85	3.55	3.85
1,000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
infinity	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.29

Summary:

# Rolls	Individual		Population	
	next roll	p(roll > 44.5)	mean	p(mean > 39.5)
1	-	-	-	-
2	(-6.02, 73.05)	16.69%	(5.10, 61.89)	20.48%
5	(24.45, 44.34)	4.83%	(29.95, 38.74)	3.55%
21	(26.96, 41.12)	0.95%	(32.50, 35.59)	0%
1,001	(25.98, 43.95)	4.06%	(34.69, 35.26)	0%

Example #2: Poker & 3-of-a-kind odds

What is the probablity of being dealt 3-of-a-kind in 5-card stud?

Solution 1: Enumeration

$$p = \left(\frac{54{,}912}{2{,}598{,}960}\right) = 2.11\%$$

Solution 2: Combinatorics:

$$p = \left(\frac{54{,}912}{2{,}598{,}960}\right) = 2.11\%$$

Solution 3: Monte-Carlo Simulation

- Count the number of times you get a 3-of-a-kind
- Answer should be 2112.84 for 100,000 hands

Monte-Carlo Simulation

- 100,000 hands
- mean = 2111.1
- std = 53.08
- sample size = 11

4-of-kind	full-house	**3-of-kind**	2-pair	pair
31	136	**2230**	4835	42365
27	132	**2037**	4681	42404
23	170	**2126**	4701	42381
21	138	**2153**	4807	42245
21	125	**2081**	4879	42378
24	142	**2069**	4672	42552
25	143	**2077**	4773	42506
27	158	**2111**	4700	42318
27	147	**2101**	4855	42230
24	135	**2154**	4794	42209
15	134	**2083**	4776	42017

Individual Tests

What is the 90% confidence interval for the next time I run the simulation?

- Individual (not a population)
- Spread does not depend upon sample size

t-score for 5% tails and 10 degrees of freedom = 1.812

$$\bar{x} - 1.812s < hands < \bar{x} + 1.812$$

$$2014.9 < hands < 2207.3$$

What is the chance that I'll draw more than 2199.5 3-of-a-kind hands?

$$t = \left(\frac{2199.5 - \bar{x}}{s}\right) = 1.6655$$

$$p = 6.34\%$$

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
10	0.7	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59

Population Tests

What is the 90% confidence interval for the odds of drawing a 3-of-a-kind?

- This is a population question
- The larger the sample size, the more you know

$$\bar{x} - 1.812\left(\frac{s}{\sqrt{11}}\right) < N < \bar{x} + 1.812\left(\frac{s}{\sqrt{11}}\right)$$

$$2082.1 < N < 2140.1 \qquad (\ actual\ N = 2112.84\ )$$

What is the chance that the odds are more than 2150 in 100,000 hands?

$$t = \left(\frac{2150 - \bar{x}}{\left(\frac{s}{\sqrt{11}}\right)}\right) = 2.4310$$

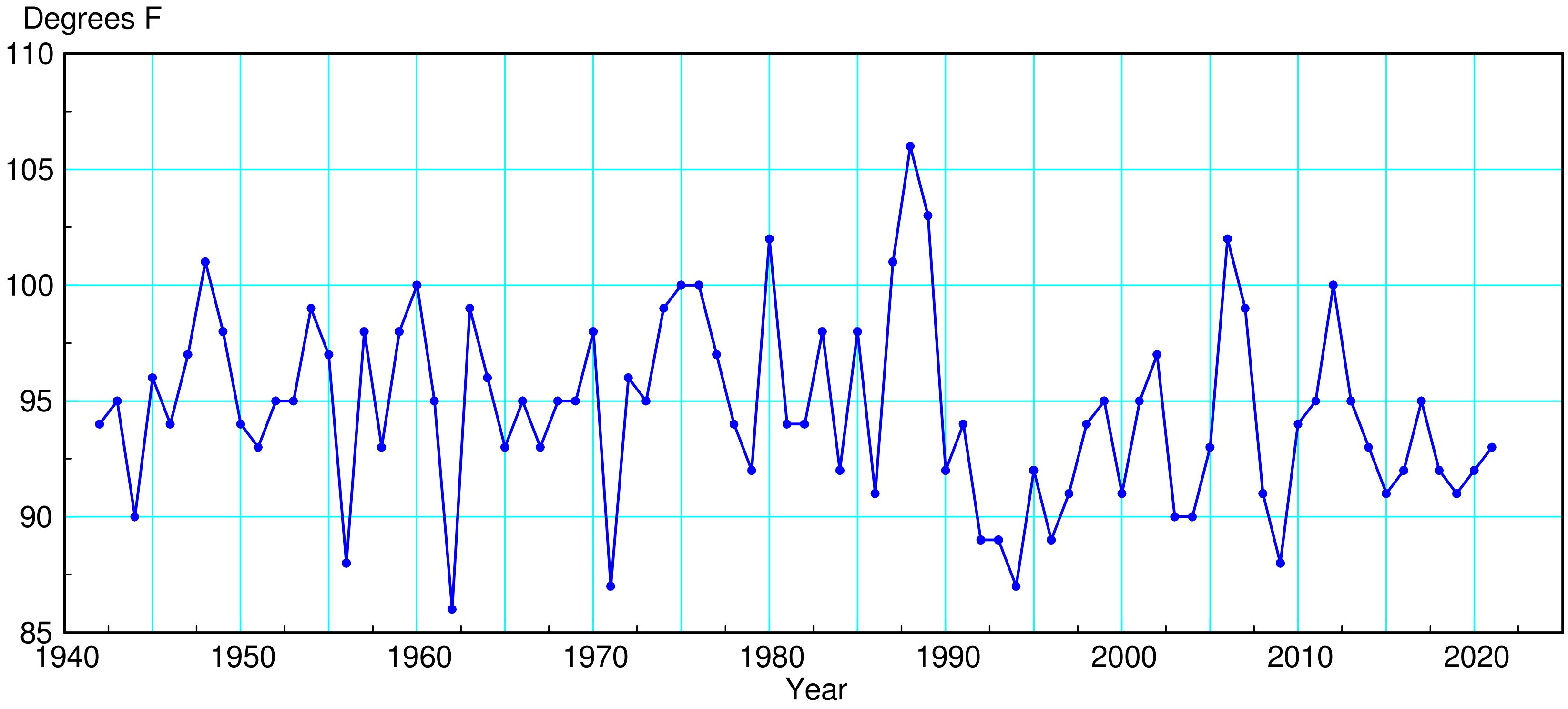
$$p = 1.77\%$$

Student t-Table (area of tail)										
p	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
10	0.7	0.88	1.09	1.37	1.81	2.23	2.76	3.17	4.14	4.59

More Fun with t-Tests

What is the probability that it will break 100F this coming July?

- Use 80 years of data



High for the Month of July since 1942 (National Weather Service)

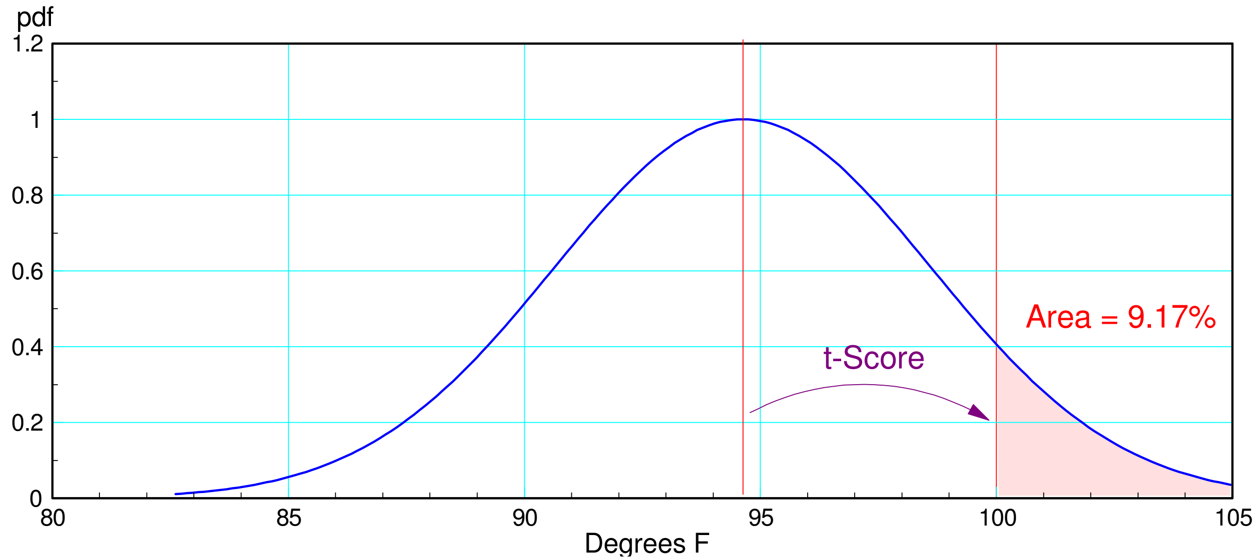
To determine this probability, compute the mean and standard deviation:

- $\bar{x} = 94.6250F$
- s = 4.0030F
- p = 79 *80 data points = 79 degrees of freedom*

$$t = \left(\frac{100-94.6250}{4.0030}\right) = 1.3423$$

Use a t-table to convert this to a probability: p = 0.0917

- There is 9.17% chance it will break 100F this coming July



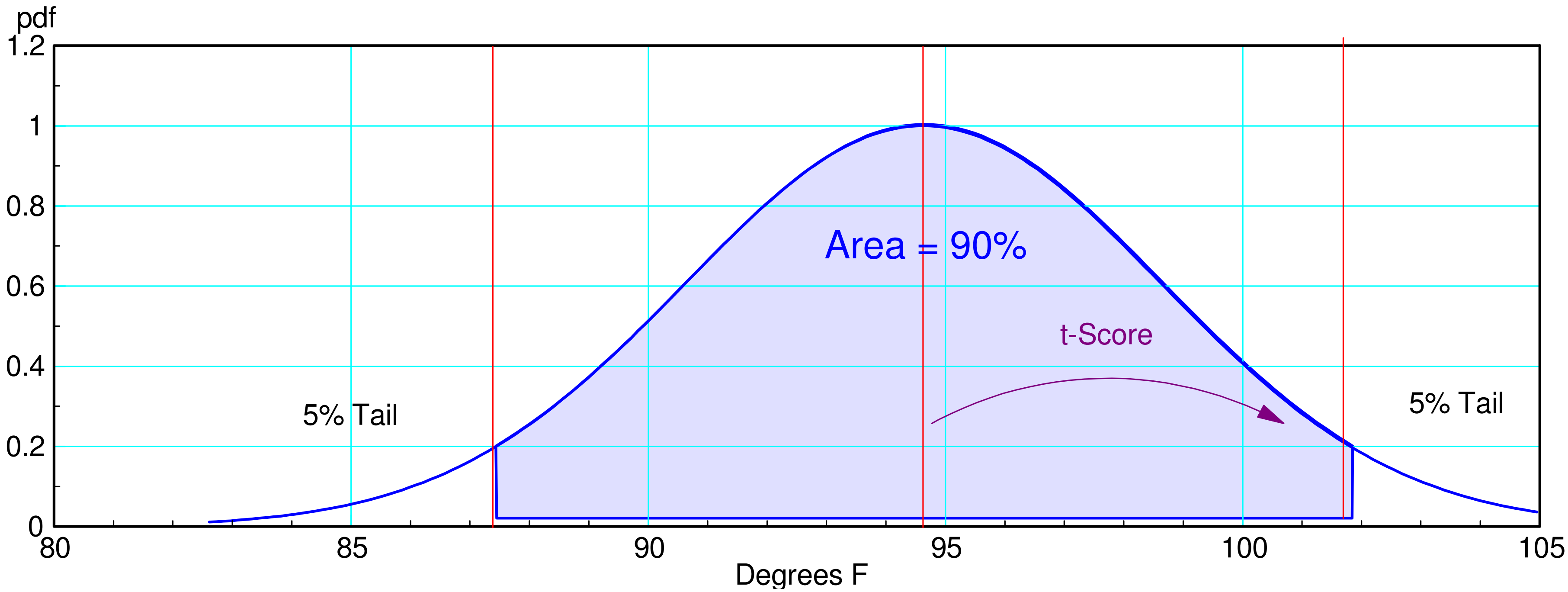
What is the 90% confidence interval for the high in July?

Use a Student t-Table to determine how far away from the mean you have to go for each tail to have an area of 5% (leaving 90% in the middle)

The 90% confidence interval is

$$\bar{x} - 1.665s < high < \bar{x} + 1.665s$$

$$87.96F < high < 101.29F \qquad p = 0.9$$



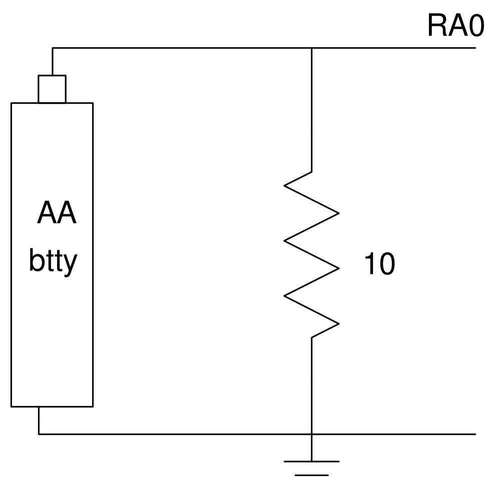
- In the dropdown box, describe the random variable.
- Enter a value for degrees of freedom.
- Enter a value for all but one of the remaining text boxes.
- Click the **Calculate** button to compute a value for the blank text box.

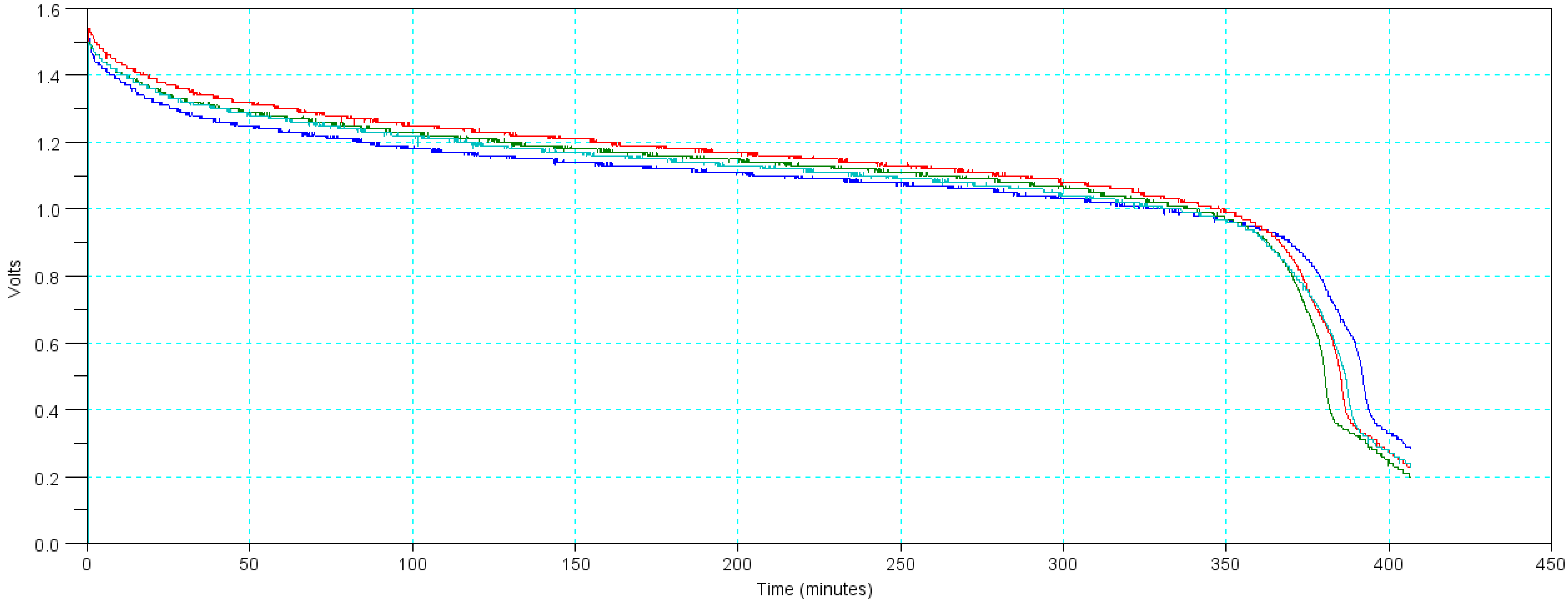
Random variable	t score
Degrees of freedom	76
t score	-1.665
Probability: $P(T \leq t)$	0.05

How much energy is in a AA battery?

Design of Experiment:

- Discharge a AA battery across a 10 Ohm resistor.
- Measure the voltage every 6 seconds for 7 hours.
- Compute the power at any time as $P = \frac{V^2}{R}$
- Integrate to get the energy dissipated across the 10 Ohm resistor





1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Volts
0
50
100
150
200
250
300
350
400
450
Time (minutes)

Statistical Analysis:

- Convert the data to a number
- Time to 1.00V
- Total energy in Joules
- Other Possibilities

Energy in Joules:

```
Joules = sum(Watts) * 6

 Joules  =   2937.5546    3063.8801    3204.829    3019.9991
```

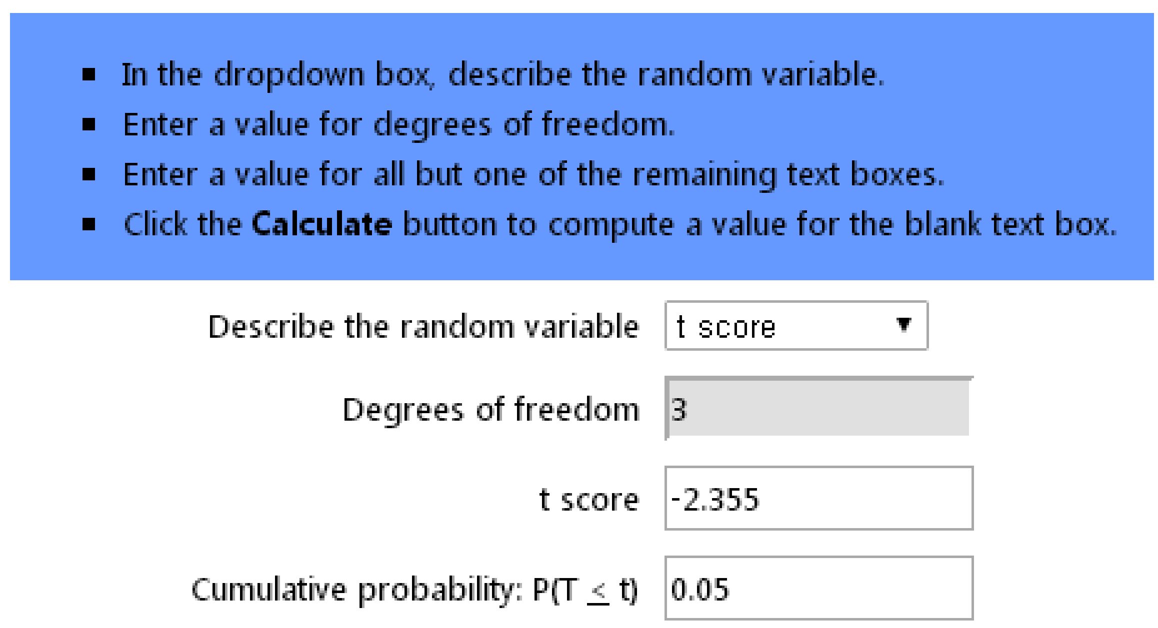
Find the mean and the standard deviation:

```
x = mean(Joules)

    3056.5657

s = stdev(Joules)

    111.85742
```

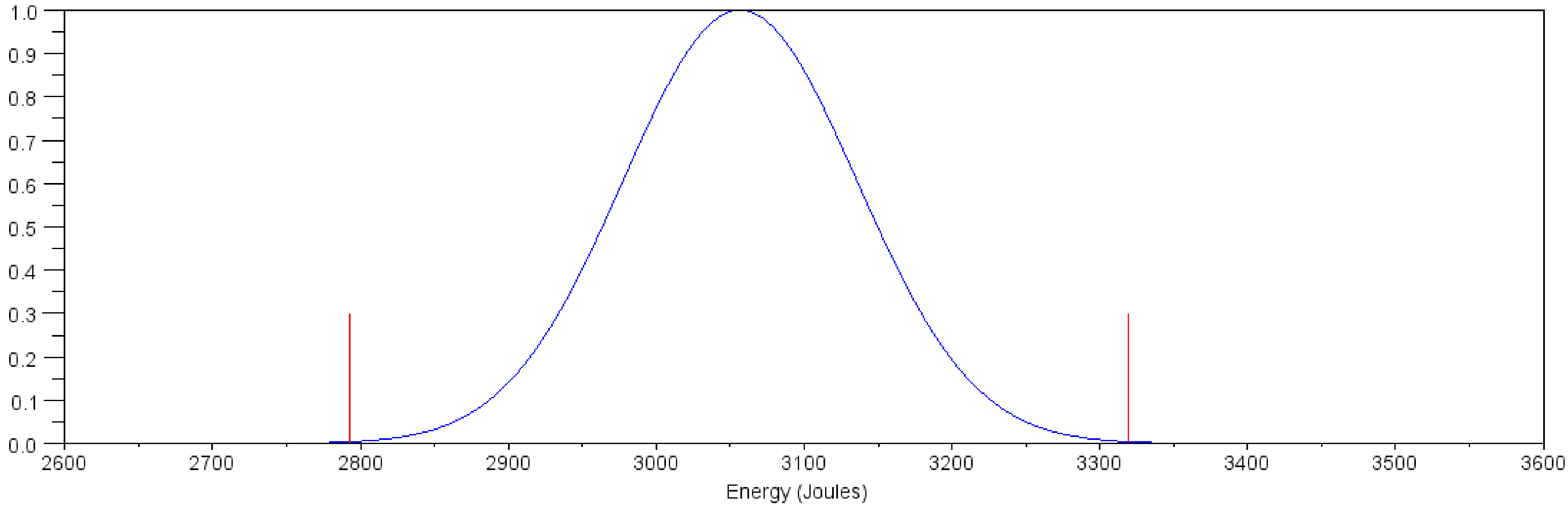
Determine how far you need to go (t-score) for 5% tails

- 2.355 deviations

  2793 < Joules < 3319 $\qquad\qquad p = 0.9$



StatTrek.com



90% Confidence Interval for the Total Energy in a type-D AA battery

What percentage of batteries have 500mAh?

- Rated energy

Convert 500mAh to Joules. Assuming 1.5V (the rated voltage):

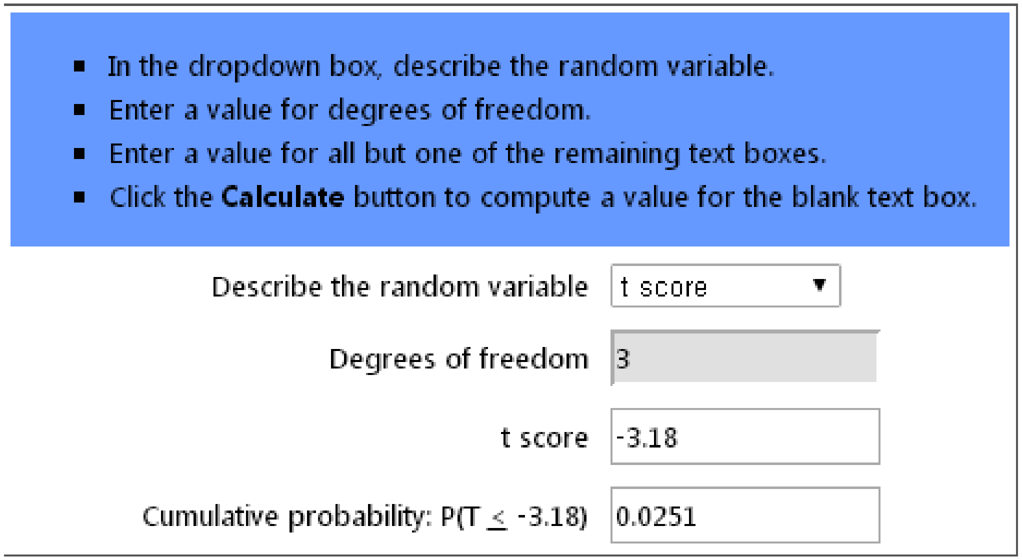
$$1.5V \cdot 500mAh \cdot \left(\frac{60\min}{hr}\right)\left(\frac{60s}{\min}\right) = 2700 \text{ Joules}$$

Find the t-score

$$t = \left(\frac{\bar{x} - 2700}{s}\right) = 3.187$$

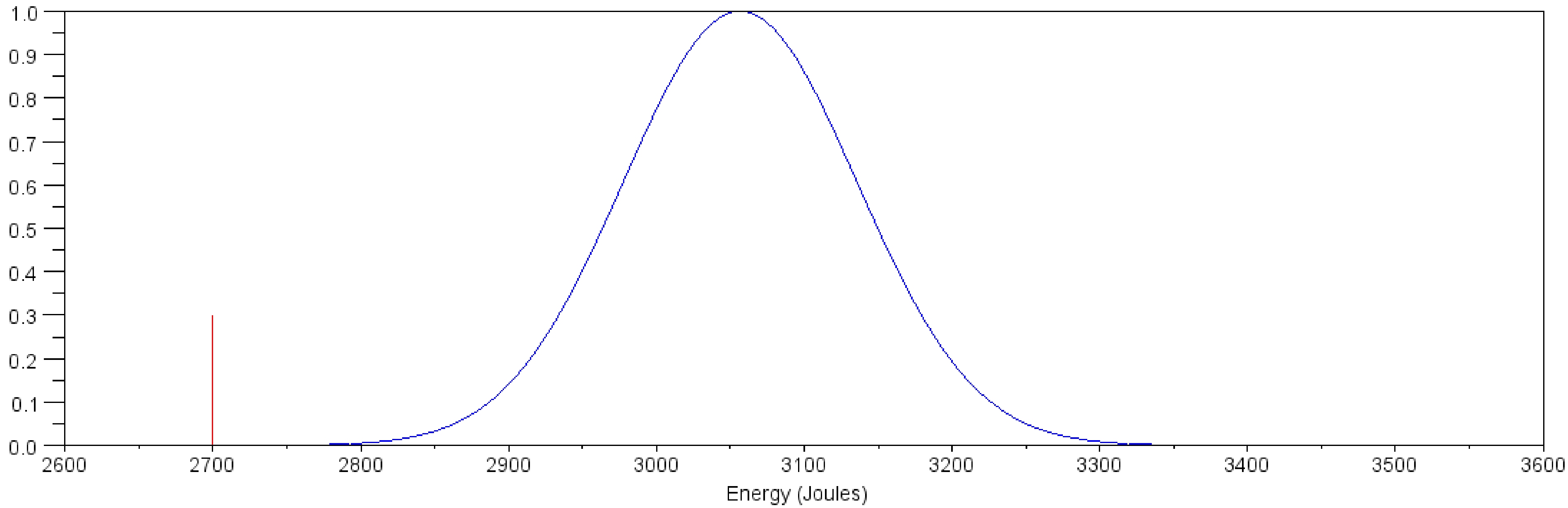
Use a t-table to convert the t-score to a probability

- p = 2.51%



Result:

- 97.49% of the batteries should meet the specifications.



97.49% of the batteries should have a total energy more than 500mAh (2700 Joules)

Summary

If you know the mean and the standard deviation, use a Normal distribution

If you have to estimate the mean and standard deviation using data, use a student-t distribution.

Notes:

- A sample size of 1 is meaningless
- With only 2 measurements, you *can* do something
- More measurements help, but you get diminishing returns
- If estimating something related to the population, divide the variance by the sample size