# Random Processes

Probability and hypothesis testing

## Topics

- Rolling dice in Matlab
- Determining if a die is fair (chi-squared test)
- Normal Distribution
- Predicting the result of die rolls using a normal distribution
- Predicting the result of die rolls using a Student-t distribution

## Introduction

Every time you do an experiment, you will get slightly different results. For example,

- If you roll a 6-sided die, the results will (probably) be different each time you roll the dice.
- If you measure the gain of a transistor, each transistor will have a slightly different gain

This variation in the results means that each is a random process. Statistics is a branch of mathematics that lets you describe and predict the result of random processes.

This is an important area in ECE since any time you collect data in lab, the results will vary each time you run the experiment.

This lecture covers common statistical tests used to analyze such data.

- Chi-Squred Test: Is a die really fair?
- t-Test with One Population: What is the 90% confidence interval?
- t-Test with Two Population: Is the mean of A more than the mean of B?

## Chi-Squared Test

The chi-squared test is a test of a distribution: it is used to detect if a set of data is consistent with the assumed distribution. For example, with a Chi-Squared test, you can determine if a die is fair or not (all numbers have equal probability).

The procedure for using a chi-squared test is as follows:

- You collect a bunch of data
- Separate the data in to N bins (such as 6 numbers for testing a 6-sided die).
- Count the number of times the data wound up in each bin
- Compare it to the expected frequency using the metric

$$\chi^2 = \sum \left( \frac{(np_i - N_i)^2}{np_i} \right)$$

Use a chi-squared table to convert this to a probability. A large number means that the data is inconsistent with the assumed distribution.

- df is the degrees of freedom (number of bins minus 1)
- % is the probability level
- The number in the table is the chi-square value

Chi-Squared Table

Probability of rejecting the null hypothesis
http://people.richland.edu/james/lecture/m170/tbl-chi.html

| df | 99.5% | 99% | 97.5% | 95% | 90% | 10% | 5% | 2.5% | 1% | 0.5% |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7.88 | 6.64 | 5.02 | 3.84 | 2.71 | 0.02 | 0 | 0 | 0 | 0 |
| 2 | 10.6 | 9.21 | 7.38 | 5.99 | 4.61 | 0.21 | 0.1 | 0.05 | 0.02 | 0.01 |
| 3 | 12.84 | 11.35 | 9.35 | 7.82 | 6.25 | 0.58 | 0.35 | 0.22 | 0.12 | 0.07 |
| 4 | 14.86 | 13.28 | 11.14 | 9.49 | 7.78 | 1.06 | 0.71 | 0.48 | 0.3 | 0.21 |
| 5 | 16.75 | 15.09 | 12.83 | 11.07 | 9.24 | 1.61 | 1.15 | 0.83 | 0.55 | 0.41 |
| 6 | 18.55 | 16.81 | 14.45 | 12.59 | 10.65 | 2.2 | 1.64 | 1.24 | 0.87 | 0.68 |
| 7 | 20.28 | 18.48 | 16.01 | 14.07 | 12.02 | 2.83 | 2.17 | 1.69 | 1.24 | 0.99 |
| 8 | 21.96 | 20.09 | 17.54 | 15.51 | 13.36 | 3.49 | 2.73 | 2.18 | 1.65 | 1.34 |
| 9 | 23.59 | 21.67 | 19.02 | 16.92 | 14.68 | 4.17 | 3.33 | 2.7 | 2.09 | 1.74 |
| 10 | 25.19 | 23.21 | 20.48 | 18.31 | 15.99 | 4.87 | 3.94 | 3.25 | 2.56 | 2.16 |

## Example: Fair Die:

Matlab has a random number generator. Does is produce truly random numbers?

To determine this, set up an experiment where we roll a 6-sided die 120 times.

```
result = zeros(6,1);
for i=1:120
   D6 = ceil( 6 * rand );
   result(D6) = result(D6) + 1;
   end

result

chi2 = sum(  (result - 20).^2 / 20  )
```

Now set up a table where we compare the actual frequency of each number (N) vs. the expected results (np):

| Number | probability (p) | Expected Frequency (np) | Actual Frequency (N) | $\frac{(np-N)^2}{np}$ |
|---|---|---|---|---|
| 1 | 1/6 | 20 | 22 | 0.2000 |
| 2 | 1/6 | 20 | 17 | 0.4500 |
| 3 | 1/6 | 20 | 17 | 0.4500 |

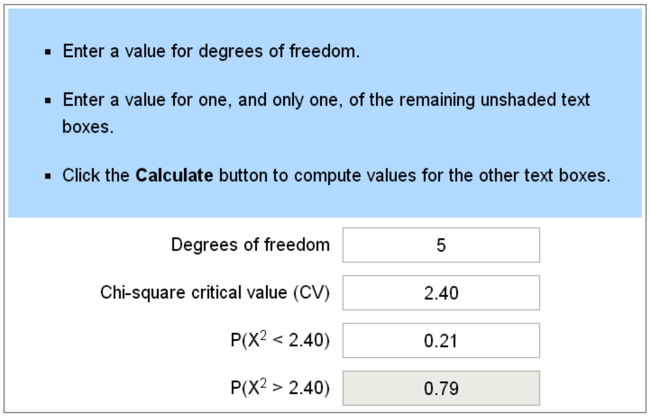
| | | | | |
|---|---|---|---|---|
| 4 | 1/6 | 20 | 19 | 0.0500 |
| 5 | 1/6 | 20 | 25 | 1.2500 |
| 6 | 1/6 | 20 | 20 | 0.0000 |
| | | | **Sum** | **2.4** |

The chi-squared table converts the chi-squared score (2.40) to a probability. Using the above table with 5 degrees of freedom (6 bins), 2.40 is more than 10% and less than 90%

- More than 10% means the data probably wasn't fudged. It the data is too perfect, be suspicious
- Less than 90% means there is no reason to claim that Matlab's rand function is biased.

You can also use StatTrek.com which tells you that a chi-squared value of 2.40 equates to a probability of 0.21

- You cannot say that the die is loaded with > 90% confidence



- Enter a value for degrees of freedom.
- Enter a value for one, and only one, of the remaining unshaded text boxes.
- Click the **Calculate** button to compute values for the other text boxes.

| | |
|---|---|
| Degrees of freedom | 5 |
| Chi-square critical value (CV) | 2.40 |
| P($X^2$ < 2.40) | 0.21 |
| P($X^2$ > 2.40) | 0.79 |

Chi-Sqared Result from StatTrek.com.

## Example 2: Fair Die, Non-Random Process

Instead of using the *rand* function, generate the die roll by going through the sequence {1, 2, 3, 4, 5, 6} and repeating. Is this a fair die?

```
result = zeros(6,1);
for i=1:120
   D6 = mod(i, 6) + 1;
   result(D6) = result(D6) + 1;
   end

result

chi2 = sum(  (result - 20).^2 / 20  )
```

Now set up a table where we compare the actual frequency of each number (N) vs. the expected results (np):

| Number | probability (p) | Expected Frequency (np) | Actual Frequency (N) | $\frac{(np-N)^2}{np}$ |
|---|---|---|---|---|
| 1 | 1/6 | 20 | 20 | 0 |
| 2 | 1/6 | 20 | 20 | 0 |
| 3 | 1/6 | 20 | 20 | 0 |
| 4 | 1/6 | 20 | 20 | 0 |
| 5 | 1/6 | 20 | 20 | 0 |
| 6 | 1/6 | 20 | 20 | 0 |
|  |  |  | **Sum** | **0.000** |

The chi-squared value for going through the sequence {1,,6} is 0.000

From a chi-squred table (or StatTrek), this corresponds to a probability of 0.0000. This tells you that

- The data is consistent with a fair die, and
- The data was probably fudged.

The latter tells you that this isn't the result of a random process: you won't get such perfect data (most likely) if it were.

## Example3: Loaded Die:

Suppose 10% of the time you cheat: the die is forced to be a six. Can you detect this?

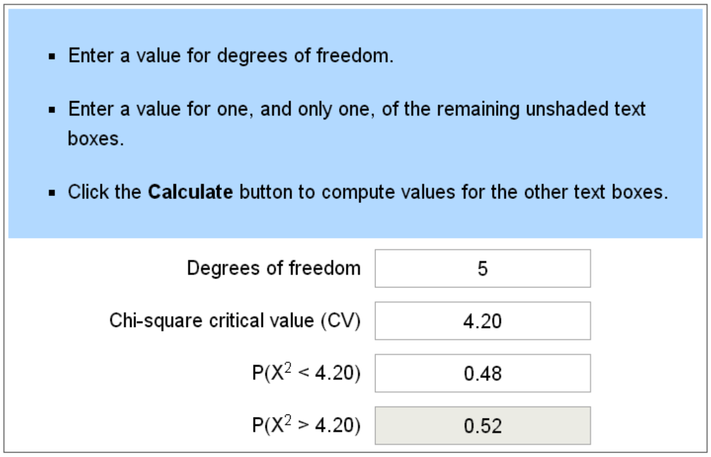
```
result = zeros(6,1);
for i=1:120
   D6 = ceil( 6 * rand );
   if (rand < 0.1)
      D6 = 6;
      end
   result(D6) = result(D6) + 1;
   end

result

chi = sum(  (result - 20).^2 / 20 )
```

Again, set up a chi-squared table:

| Number | probability (p) | Expected Frequency (np) | Actual Frequency (N) | $\frac{(np-N)^2}{np}$ |
|---|---|---|---|---|
| 1 | 1/6 | 20 | 17 | 0.4500 |
| 2 | 1/6 | 20 | 21 | 0.0500 |
| 3 | 1/6 | 20 | 23 | 0.4500 |
| 4 | 1/6 | 20 | 18 | 0.2000 |
| 5 | 1/6 | 20 | 15 | 1.2500 |
| 6 | 1/6 | 20 | 26 | 1.8000 |
|  |  |  | **Sum** | **4.2000** |

From StatTrek.com, a chi-squared value of 4.20 corresponds to a probability of 0.48

- Based upon this data, this is a 48% chance the die is loaded



- Enter a value for degrees of freedom.
- Enter a value for one, and only one, of the remaining unshaded text boxes.
- Click the **Calculate** button to compute values for the other text boxes.

| | |
|---|---|
| Degrees of freedom | 5 |
| Chi-square critical value (CV) | 4.20 |
| $P(X^2 < 4.20)$ | 0.48 |
| $P(X^2 > 4.20)$ | 0.52 |

From the Matlab code, we *know* the die is loaded. From the data, I can't tell with only 120 data points.

If you increase the sample size to 600

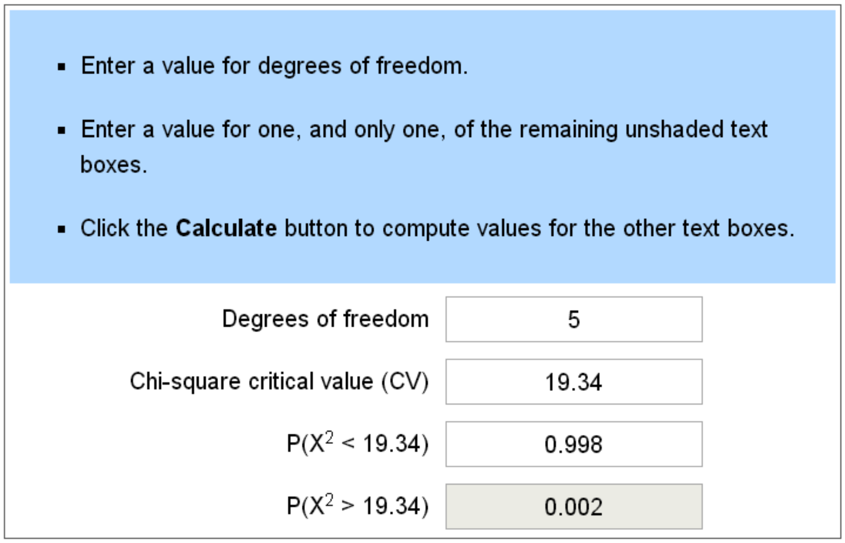
```
result = zeros(6,1);
for i=1:600
   D6 = ceil( 6 * rand );
   if (rand < 0.1)
      D6 = 6;
      end
   result(D6) = result(D6) + 1;
   end

result

chi = sum(  (result - 100).^2 / 100 )
```

| Number | probability (p) | Expected Frequency (np) | Actual Frequency (N) | $\frac{(np-N)^2}{np}$ |
|---|---|---|---|---|
| 1 | 1/6 | 100 | 95 | 0.2500 |
| 2 | 1/6 | 100 | 85 | 2.2500 |
| 3 | 1/6 | 100 | 91 | 0.8100 |
| 4 | 1/6 | 100 | 91 | 0.8100 |
| 5 | 1/6 | 100 | 99 | 0.0100 |
| 6 | 1/6 | 100 | 139 | 15.2100 |
| | | | **Sum** | **19.3400** |

Now you can tell that the die is loaded. From StatTrek, a chi-squared score of 19.34 equates to a probability of 0.998

- Based upon this data, I'm 99.8% certain that the die is loaded.

Given enough data, you can spot even slight loading. It might cost you a *lot* of money getting this data though...



- Enter a value for degrees of freedom.
- Enter a value for one, and only one, of the remaining unshaded text boxes.
- Click the **Calculate** button to compute values for the other text boxes.

| | |
|---|---|
| Degrees of freedom | 5 |
| Chi-square critical value (CV) | 19.34 |
| $P(X^2 < 19.34)$ | 0.998 |
| $P(X^2 > 19.34)$ | 0.002 |

### Example: Fudging the Data:

An interesting aspect of chi-squred tests is it can also detect fudged data: data that is *too* good. For example, instead of rolling a fair die 600 times, roll the die 120 times (like before) then add 80 to the result (making it *look* like I rolled the dice 600 times).

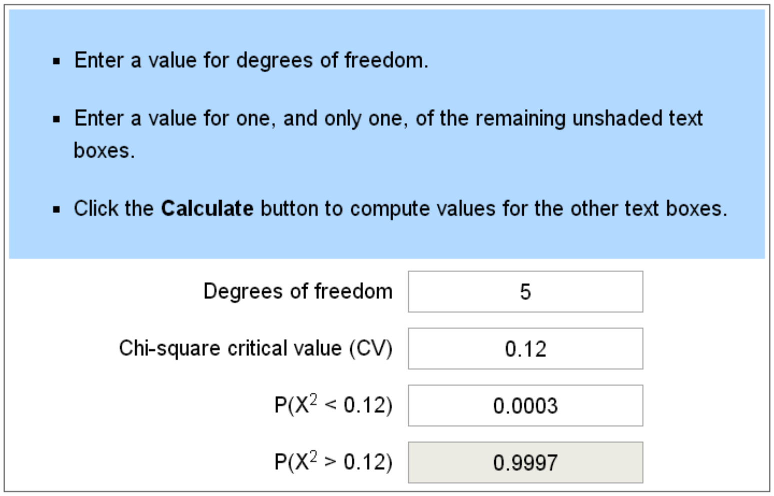
```
result = zeros(6,1);
for i=1:60
   D6 = ceil( 6*rand);
   result(D6) = result(D6) + 1;
   end

result = result + 90

chi2 = sum(  (result - 100).^2 / 100  )
```

| Number | probability (p) | Expected Frequency (np) | Actual Frequency (N) | $\frac{(np-N)^2}{np}$ |
|---|---|---|---|---|
| 1 | 1/6 | 100 | 99 | 0.01 |
| 2 | 1/6 | 100 | 101 | 0.01 |
| 3 | 1/6 | 100 | 99 | 0.01 |
| 4 | 1/6 | 100 | 101 | 0.01 |
| 5 | 1/6 | 100 | 102 | 0.04 |
| 6 | 1/6 | 100 | 98 | 0.04 |
| | | | **Sum** | **0.12** |

From StatTrek, a chi-squred value of 0.12 corresponds to a probability of 0.0003

- The data looks like a fair die
- The data is actually *too* good: the odds against getting such good data is 0.03% ( 3333 : 1 odds against)

The data was most likely fudged.



- Enter a value for degrees of freedom.
- Enter a value for one, and only one, of the remaining unshaded text boxes.
- Click the **Calculate** button to compute values for the other text boxes.

Degrees of freedom 5

Chi-square critical value (CV) 0.12

$P(X^2 < 0.12)$ 0.0003

$P(X^2 > 0.12)$ 0.9997

If the data is fudged, it shows up as a probability which is *too* close to zero

is *too* good, it show

# Student t-Test: One Population

Switching gears, another common statistical test is the Student t-Test. This is a test of the mean. With this test, you can determine

- The 90% confidence interval for where your data will lie,
- The 90% confidence interval for the population's mean, and
- The probability that a random measurement will be more than a threshold. For example, the probability that any given transistor has a gain more than 500.

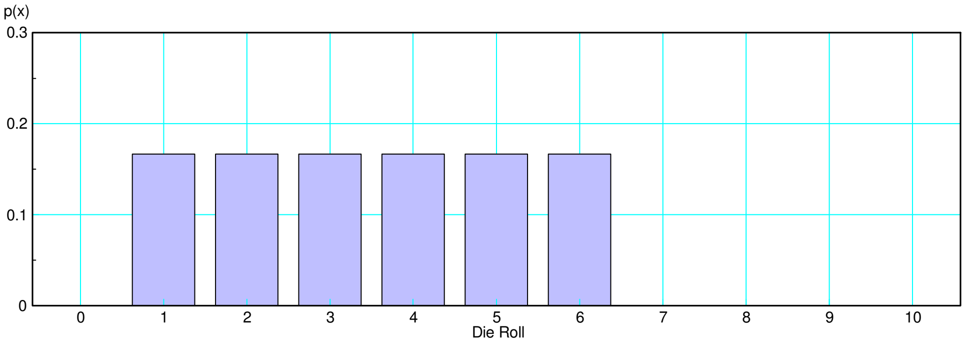
The heart of the Student t-Test is the assumption that your data from a normal distribution. The normal distribution is the bell-shaped curve you're probably familiar with that describes height, weight, resistor values, etc. This assumption that your data is normally distributed is usually a valid assumption due to the Central Limit Theorem.

## Central Limit Theorem

The Central Limit Theorem states that

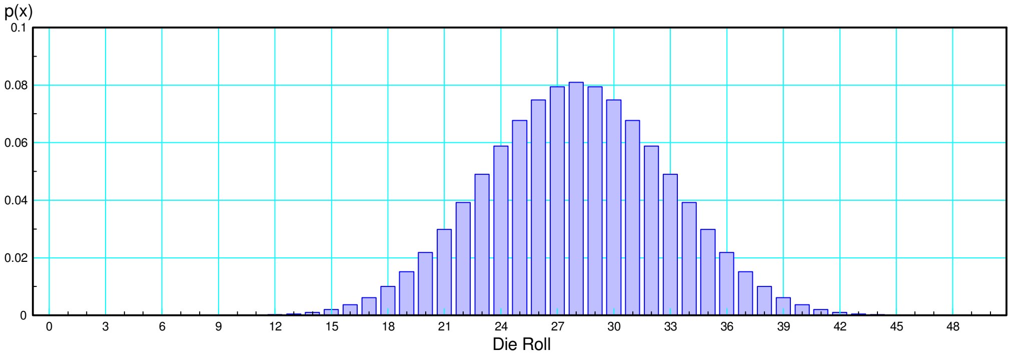
- All distributions converge to a normal distribution as the number of samples goes to infinity, and
- Once you have a normal distribution, you remain with a normal distribution.

For example, take a six sided die with each number having a probability of 1/6.



Probability of rolling each number with a fair 6-sided die

If you sum 8 dice, the result in a bell curve (it approaches a Normal distribution)



Probability distribution for rolling 8 6-sided dice

## Normal (Gaussian) Distributions:

The normal distribution is written as

$$N(\bar{x}, \sigma^2)$$

and has the probability density function of

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{-(x-\bar{x})^2}{2\sigma^2}\right)$$

where

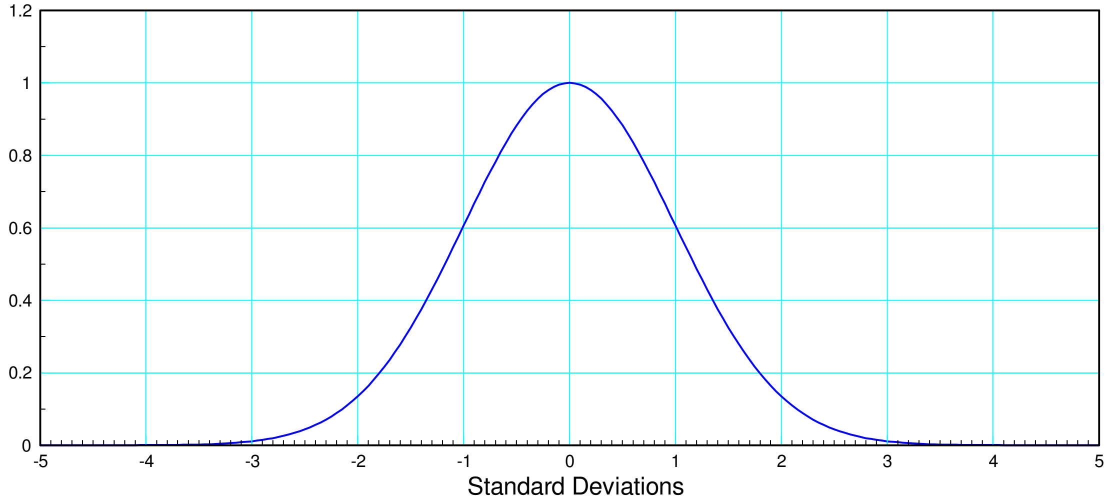
- $\bar{x}$ is the mean and
- $\sigma$ is the standard deviation (a measure of the spread)

N(0,1) is the standard-normal distribution with

- mean equal to zero, and
- standard deviation equal to one

It's probability density function (scaled so the peak is 1) is:

```
s = [-5:0.05:5]';
p = exp(-s.^2 / 2);
plot(s,p);
xlabel('deviations');
```



Standard Normal Distribution (normalized so the peak is 1.000)

The area under the curve is the probability of an event happening. For example, the area within X standard deviations of the mean is:

| +/- 1 deviations | +/- 2 deviations | +/- 3 deviation |
|---|---|---|
| 0.68 | 0.95 | 0.996 |

As a rough rule of thumb, 95% of the data should lie within +/- 2 standard deviations of the mean. (The mean tells you the average of the data, the standard deviation tells you the spread.)

**Example: Rolling Dice**

If you know the mean and standard deviation, use a normal approximation for the distribution.

The mean and standard deviation for a 6-sided die are

$$\bar{x} = \frac{1}{n} \Sigma (x_i) = 3.5$$

$$s^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2 = 1.7078$$

What is the distribution of summing 10 six-sided dice (10d6)?

$$y = x_1 + x_2 + \ldots + x_{10}$$

Answer: The mean and variance add

$$\bar{x}_y = 10 \cdot 3.5 = 35$$

$$s_y^2 = 10 \cdot 1.7078 = 17.078$$

$$s_y = \sqrt{s_y^2} = 4.1326$$

What is the probability of rolling 45 or mode with 10d6?

Answer: Find the distance from 45 to the mean in terms of standard deviations

$$z = \left( \frac{45 - \bar{x}}{s} \right)$$

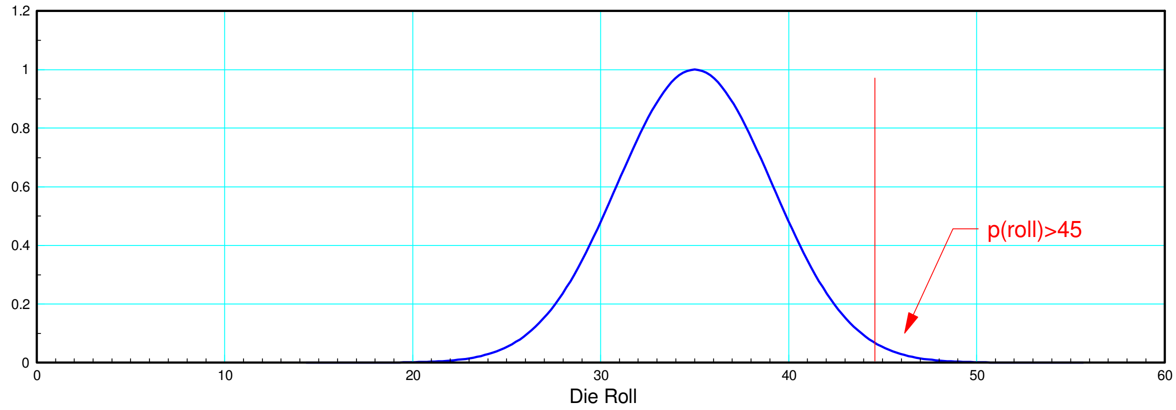
$$z = \left( \frac{45 - 35}{4.1326} \right) = 2.4198$$

Using a Normal distribution table, determine the area to the right of +2.4198 standard deviations

| Normal Distribution | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.29 |

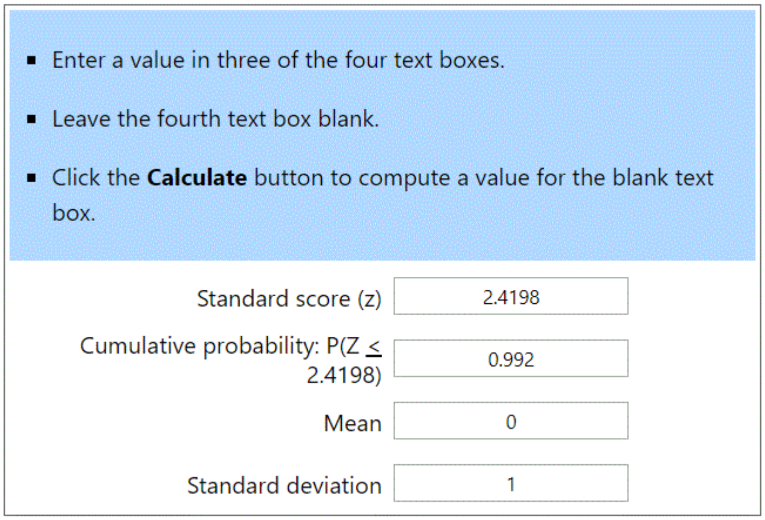
The area to the left is between 0.99 and 0.995

The chance of rolling 45 or more is between 0.01 and 0.005



Probability Distribution of Rolling 10d6: The Probability of Rolling Less than 45 is 0.992

You can also use StatTrek to compute this.



- Enter a value in three of the four text boxes.
- Leave the fourth text box blank.
- Click the **Calculate** button to compute a value for the blank text box.

| | |
|---|---|
| Standard score (z) | 2.4198 |
| Cumulative probability: P(Z ≤ 2.4198) | 0.992 |
| Mean | 0 |
| Standard deviation | 1 |

Normal Distribution from StatTrek

Problem: Determine the 90% confidence interval for rolling 10d6

Solution: For a 90% confidence interval, we need the tails to have an area of 5%. Use Normal distribution table to find out how far away from the mean you have to go for the tails to be 5%

| Normal Distribution | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.29 |

From a t-table

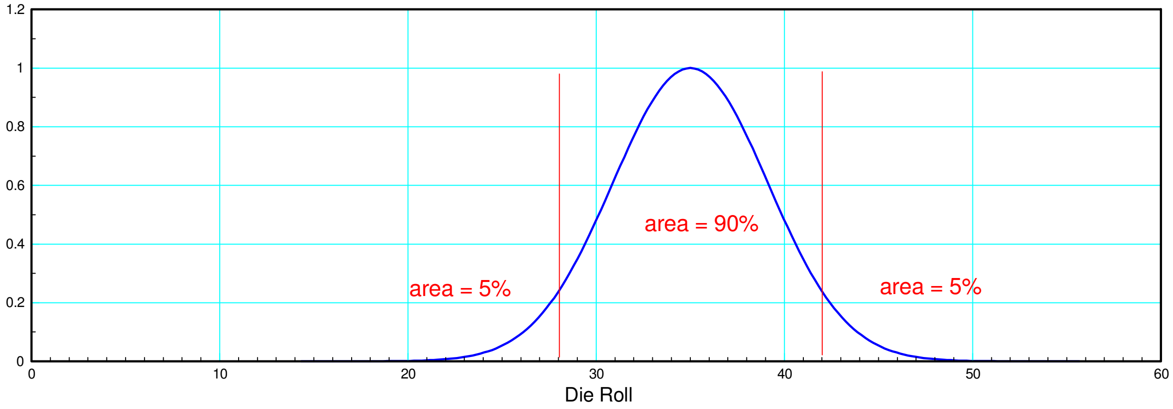
t = 1.645

The 90% confidence interval is then

$$\bar{x} - 1.645s < sum < \bar{x} + 1.645s$$

$$28.20 < sum < 41.79$$

90% of the time you should roll numbers in-between 28.20 and 41.79



90% Confidence Interval for Rolling 10d6. Note that each tail has an area of 5%

## Student t-distribution

If you don't know the mean and standard deviation of what you're sampling, you can estimate them from your data. If you do that, then you use a student-t distribution.

The t-distribution is like the normal distribution, but it takes the sample size into account. A t-table looks like the following:

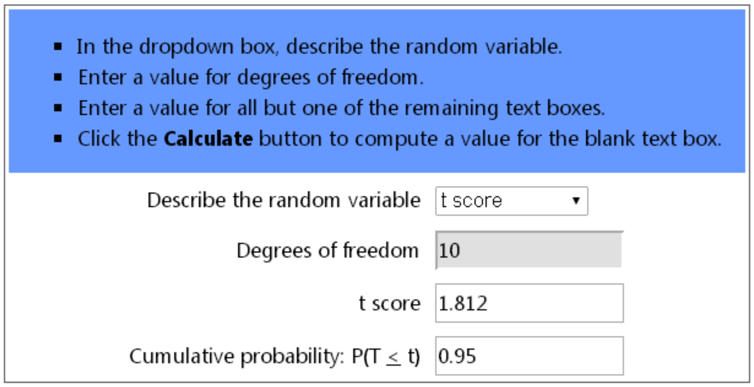
- The left column is the degrees of freedom. This is the sample size minus one.
- The top tells you the probability level (the area to the left in terms)
- The table entries tell you how many standard deviations away from the mean you have to go to capture that much area
- Infinite sample size is a Normal distribution (cental limit theorem)

Student t-Table

(http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf)

| p | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
|---|---|---|---|---|---|---|---|---|---|---|
| 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 |
| infinity | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.29 |

This is also available at StatTrek.com. For example, a probability of 0.95 with 10 degrees of freedom gives 1.81 - the same as the above table



- 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.

Describe the random variable: t score ▾

Degrees of freedom: 10

t score: 1.812

Cumulative probability: P(T $\leq$ t): 0.95

StatTrek.com t-distribution

## t-test and Dice

Suppose you don't know the mean and standard deviation for rolling 10d6. Find the probability of rolling a 45 or more.

Solution: Toss 10d6 N times. From this sample, estimate the mean and standard deviation. Since you're estimating these rather than using the actual mean and standard deviation, this is a student-t distribution.

Example: Roll 10d6 five times.

```
A = [];

for i=1:5
   A = [A, sum(ceil(6*rand(1,10)))];
   end

A  =  37.     32.     40.     41.     36.
```

From this, find the mean and standard deviation

```
x = mean(A)

 x   = 37.2

s = stdev(A)

 s   = 3.5637059
```

Note that these are close to what we found before but are a little bit off. The t-tables take the error due to a finite sample size into account. Now find the distance from 45 to the mean in terms of standard deviations

$$t = \left(\frac{45-37.2}{3.5637}\right) = 2.1887$$

Convert this to a probability using a t-table with 4 degrees of freedom (sample size minus 1)

| Student t-Table | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| p | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| 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 |
| infinity | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.29 |

From the t-table, a t-score of 2.1887 corresponds to a probability of 0.95

**There is a 95% chance of rolling less than 45 with 10d6.**

Note that this is a different result from before. If you increase the sample size, the two should eventually agree.

## t-tests and Weather Data:

The high for the month of April from 2010 - 2015 was

```
{77F, 70F, 77F, 73F, 79F, 82F}
```

From this data, determine the probability that it will break 85F this coming year.

Solution: This is similar to a Normal distribution, but

- We don't know the mean and standard deviation of the high for April, and
- We instead estimate these from the data

That makes this a student-t distribution.

The statistics for this data is

$$\bar{x} = \frac{1}{n} \Sigma x_i = 76.333$$

$$s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2 = 18.267$$

$$s = \sqrt{s^2} = 4.2740$$

(note that the variance is slightly different when you estimate the mean from the data)

The t-score is the distance 85F is from the mean in terms of standard deviations

$$t = \left( \frac{85-76.333}{4.2740} \right) = 2.0278$$

To convert this to a probability, use a t-table

- There are 5 degrees of freedom (6 data points)
- The t-score is 2.0278
- This corresponds to a probability of p = 0.95

There is a 95% chance the high for April will be less than 85F

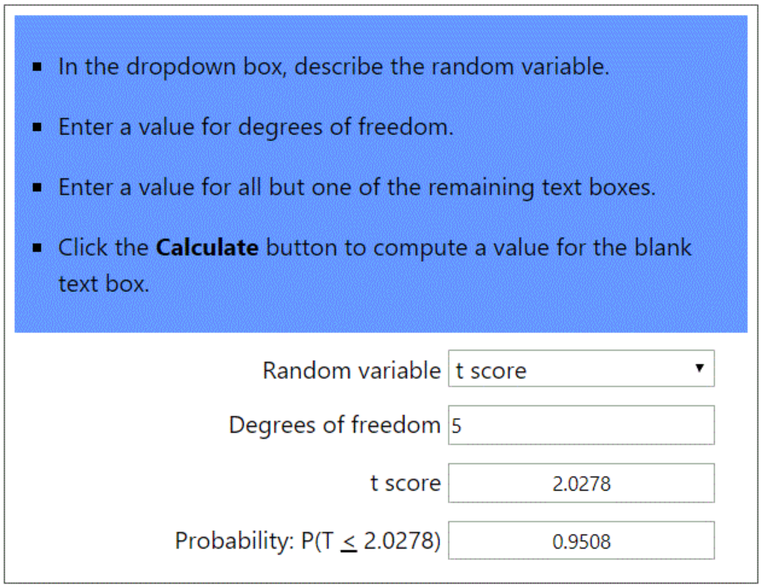
There is a 5% chance the high for April will be more than 85F

Student t-Table

(http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf)

| p | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
|---|---|---|---|---|---|---|---|---|---|---|
| 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 |

You can also use StatTrek to convert t-scores to probability



- 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.

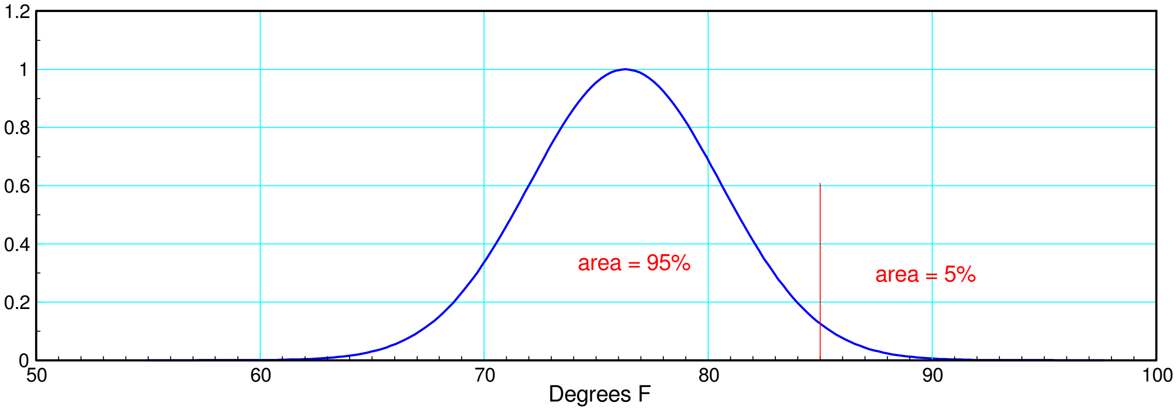
Random variable: t score

Degrees of freedom: 5

t score: 2.0278

Probability: P(T ≤ 2.0278): 0.9508

Converting a t-score to a probability using StatTrek



Distribution for the High in April: 95% of the time the high will be less than 85F

Problem: What is the 90% confidence interval for the high for April?

Solution: From a t-table with 5 degrees of freedom, find the t-score for tails of 0.05

| Student t-Table<br>(http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| p | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| 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 |
| infinity | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.29 |

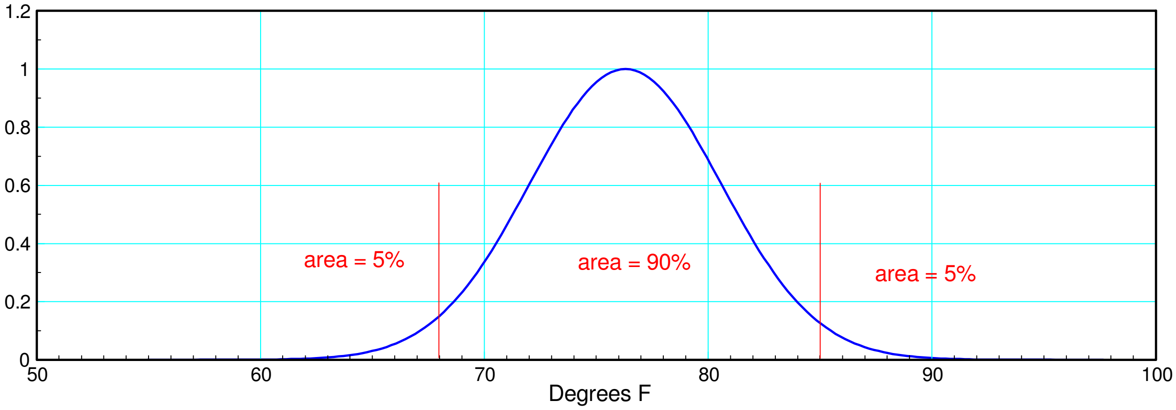
From this

$$t = 2.02$$

The 90% confidence interval is thus

$$\bar{x} - 2.02s < high < \bar{x} + 2.02s$$

$$67.98F < high < 84.68F$$



90% confidence interval for the high in April

# t Test with 2 Populations

## Comparison of Means

If you want to compare to populations (A vs. B), create a new variable, W, as

W = A - B

The mean and standard deviation of W is then

$$\bar{x}_w = \bar{x}_a - \bar{x}_b$$

$$s_w^2 = s_a^2 + s_b^2$$

Example: What is the chance that the sum of 10d6 will be more than the sum of 8d6?

Solution: The mean and standard deviations are

A: 10d6

$$\bar{x}_a = 35$$

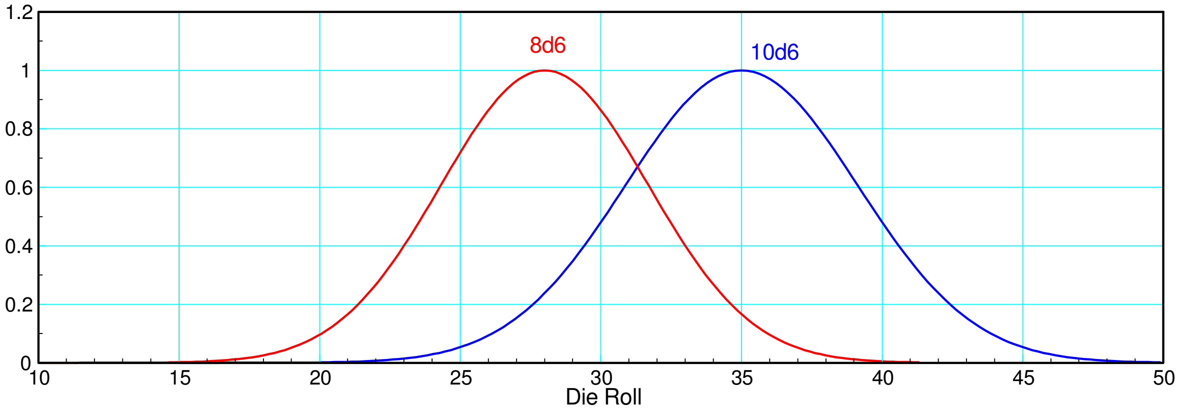
$$s_a^2 = 17.078$$

B: 8d6

$$\bar{x}_b = 28$$

$$s_b^2 = 13.66$$

The degrees of freedom is the smaller of the sample sizes minus one



Probability Distribution of Rolling 10d6 (blue) and 8d6 (red)

Note that there is overlap in the two probability distributions: it is possible for the red curve (8d6) to be to the right of the blue curve (10d6)

To compute the chance of 10d6 being more than 8d6, create a variable, W, which is the difference

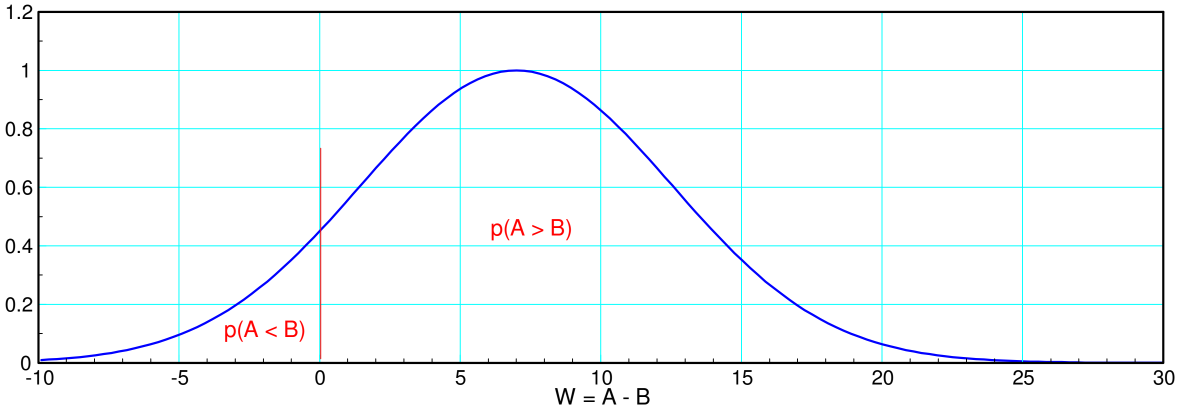
W = 10d6 - 8d6

$$\bar{x}_w = \bar{x}_a - \bar{x}_b = 7.00$$

$$s_w^2 = s_a^2 + s_b^2 = 30.74$$

$$s_w = \sqrt{s_w^2} = 5.54$$

The degrees of freedom are infinite since we know the distribution of a d6, 10d6, and 8d6



Probability Distribution of 10d6 minus 8d6 (variable W).
Area to the right of 0.00 is the probability that 10d6 > 8d6

The t-score is then the distance from the mean to zero in terms of standard deviations

$$t = \left(\frac{7.00}{5.54}\right) = 1.26$$

From a t-table with infinite degrees of freedom (also known as a Normal distribution), this corresponds to a probability of 0.9

| Student t-Table | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| p | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| 1 | 1 | 1.38 | 1.96 | 3.08 | 6.31 | 12.71 | 31.82 | 63.66 | 318.31 | 636.62 |
| 10 | 0.7 | 0.88 | 1.09 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 | 4.14 | 4.59 |
| infinity | 0.674 | 0.842 | 1.036 | **1.282** | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.29 |

p = 0.9

**10d6 should beat 8d6 90% of the time.**