Testing with Normal Distributions

ECE 341: Random Processes

Lecture #20

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

Testing with Normal Distributions

If a variable has a normal distribution, you can determine certain probabilities. This lecture covers three of these:

- Single-sided confidence interval
- Two-sided confidence interval
- Comparison of Two Distributions
  - False Positives
  - False Negatives

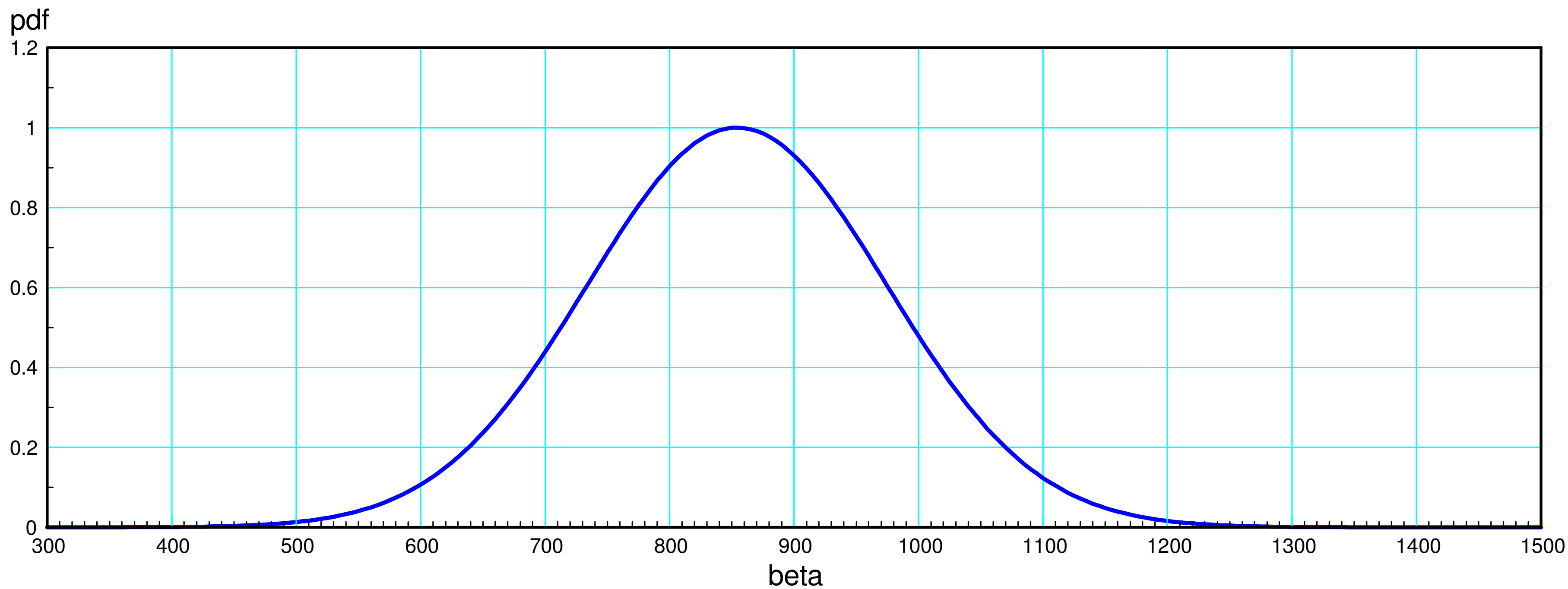
For illustration purposes, consider the gain of 62 Zetex1051a transistors:

```
915, 602, 963, 839, 815, 774, 881, 912, 720, 707, 800, 1050, 663, 1066, 1073,
802, 863, 845, 789, 964, 988, 781, 776, 869, 899, 1093, 1015, 751, 795, 776,
860, 990, 762, 975, 918, 1080 774, 932, 717, 1168, 912, 833, 697, 797, 818,
891, 725, 662, 718, 728, 835, 882, 783, 784, 737, 822, 918, 906, 1010, 819,
955, 762
```

From this data, the mean and standard deviation of these transistors can be found

```
x = mean(beta)          x =  854.1290

s = std(beta)           s =  120.2034
```

From the Central Limit Theorem, with a sample size of 62, these approach a normal distribution with the same mean and standard deviation.



Normalized pdf for the current gain, beta (hfe).

With this curve, we can answer several questions.

Single-Sided Test:

The data sheets state that the gain is at least 300.

- What is the probability that a transistor will have a gain less than 300?

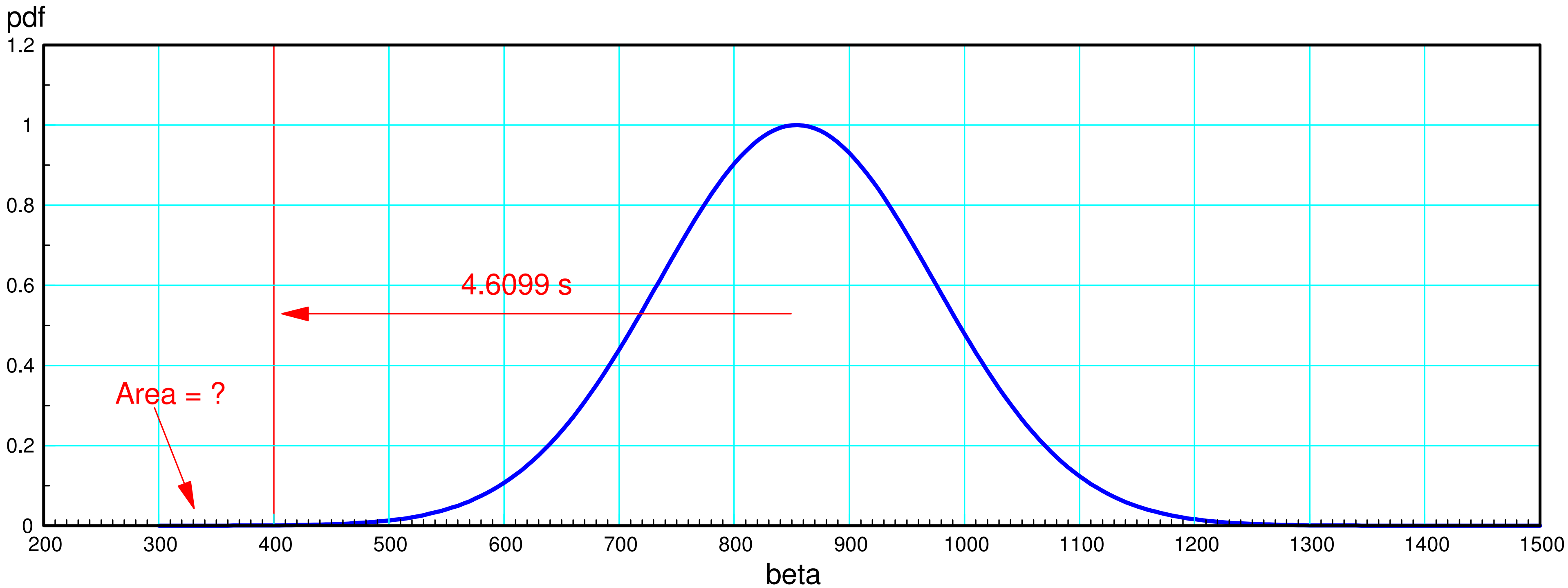
Find the z-score (distance to the mean in terms of standard deviations)

$$z = \left(\frac{\mu - 300}{\sigma}\right) = \left(\frac{854.129 - 300}{120.2013}\right) = 4.6099$$

```
>> z = (854.129-300)/120.2013
z =    4.6100

>> p = (1 - erf(z/sqrt(2)))/2
p =  2.0133e-006
```

You can also use StatTrek or a standard normal table

Deviations	+0	+1	+2	+3	+4	+5
Area of Tail	0.5	0.158655	0.022750	0.0013498	3.16712e-5	2.8665e-7

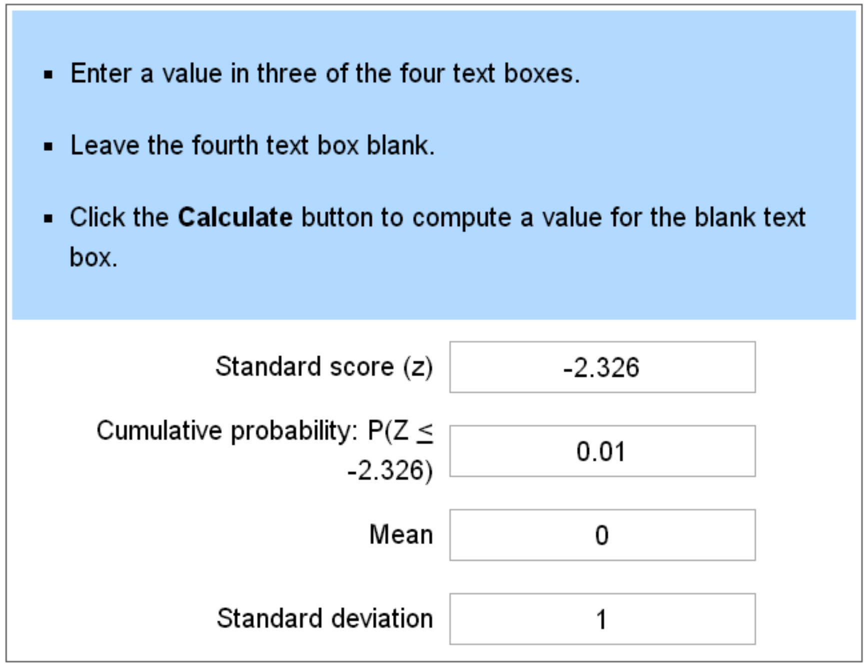


Probability that a transistor has a gain less than 300 is less than 2.0133e-006

Essentially, the manufacturer is cautions in the claims.

Example 2: Determine the gain that 99% of all transistor will meet or exceed.

Solution: Use StatTrek to determine how many standard deviations you have to go for the area to be 0.01 (1%)

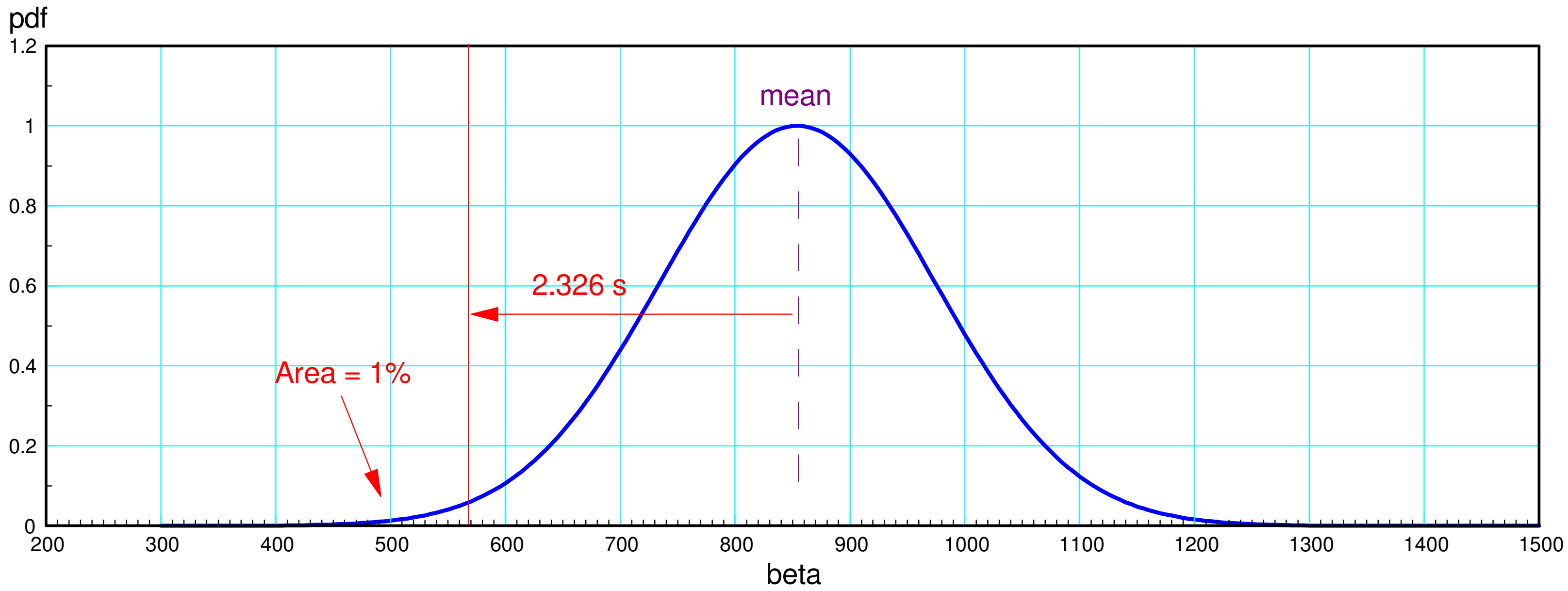


What this means is

- Go 2.326 standard deviations to the left of the mean
- The area under the curve (i.e. the probability) will be 1%

Translating to gain:

$$\beta > \mu - 2.326\sigma = 574.578 \qquad (p = 0.99)$$



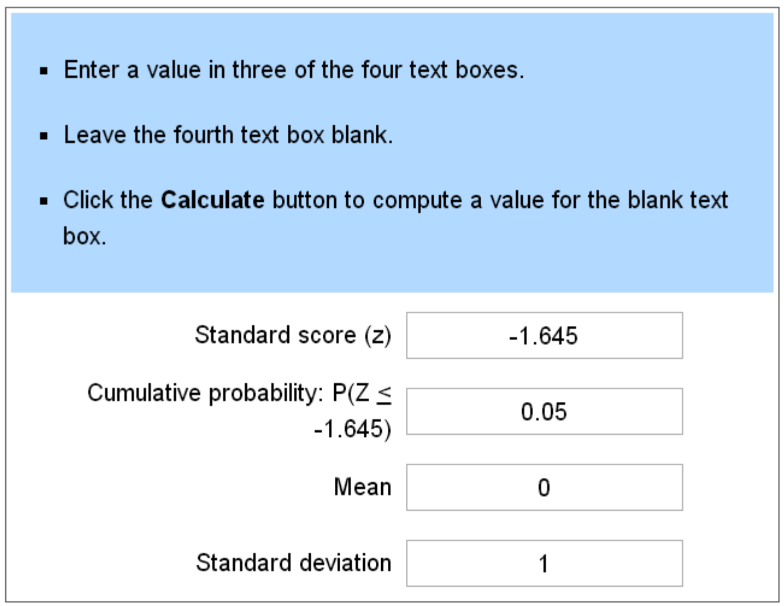
Single-Sided Test: Any given transistor will have a gain > 574 with a probability of 0.99

Two-Sided Tests (Confidence Intervals)

Example 3: Determine the 90% confidence interval for the gain of a given transistor.

Solution:

- Area in the middle = 90%
- Each tail = 5% (two tails)
- Find the z-score for 5% tails

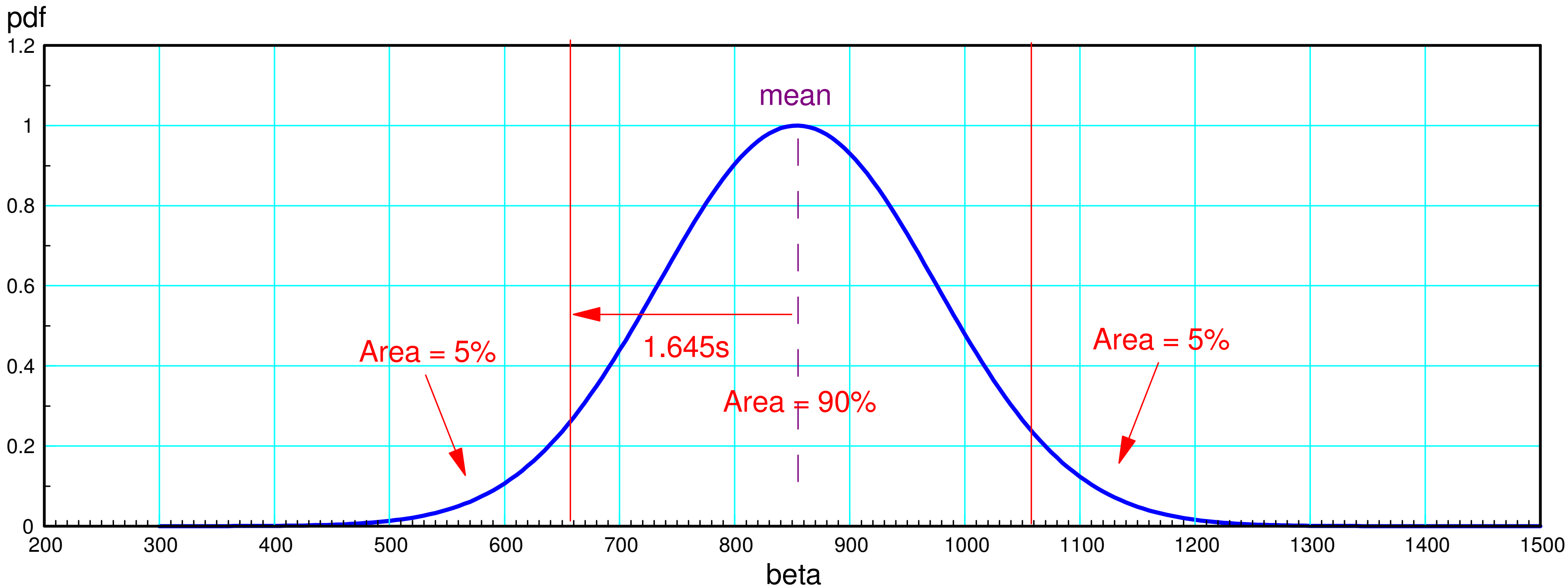


The 90% confidence interval is then

$$\mu - 1.645\sigma < \beta < \mu + 1.645\sigma \qquad\qquad p = 0.9$$

Plugging in numbers

$$656.39 < \beta < 1051.9 \qquad\qquad p = 0.9$$



90% Confidence Interval for Transistor Gain: Note that each tail has an area of 5%

Example 4: What is the 99% confidence interval?

- Repeat but with each tail being 0.5%:

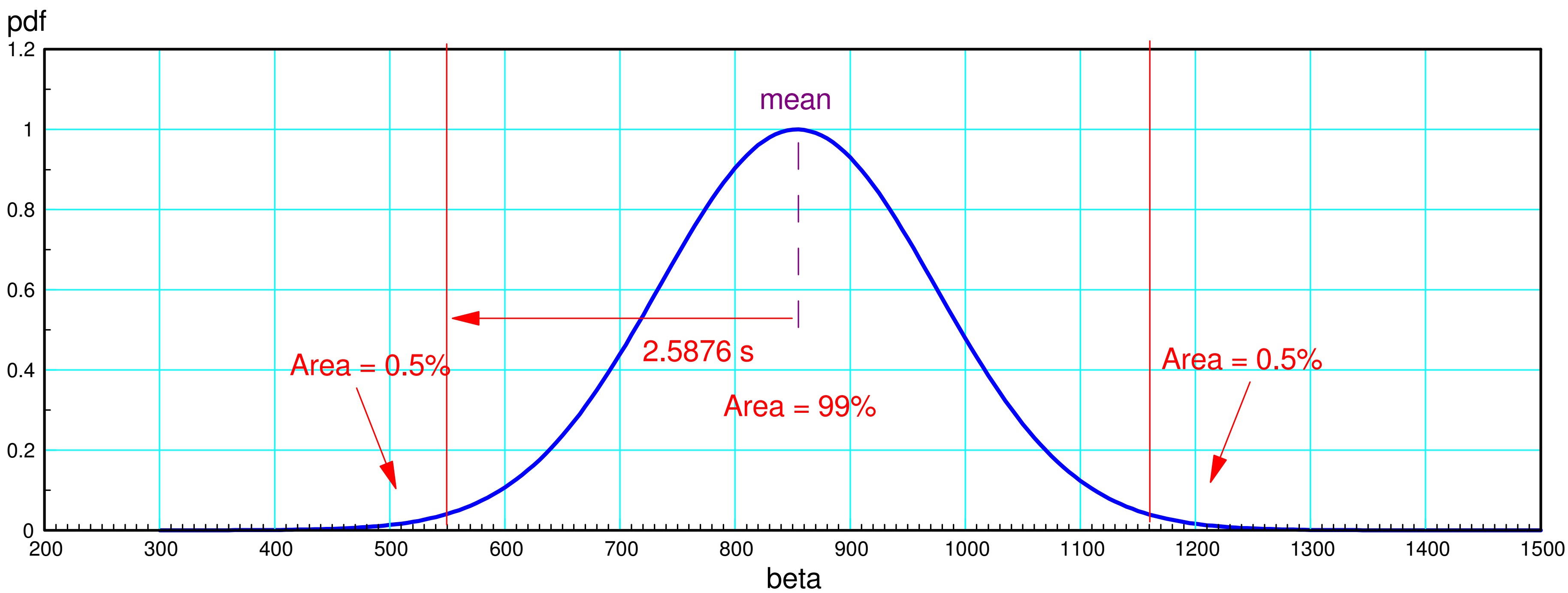
  - Enter a value in three of the four text boxes.
  - Leave the fourth text box blank.
  - Click the **Calculate** button to compute a value for the blank text box.

Standard score (z)	-2.576
Cumulative probability: P(Z ≤ -2.576)	0.005
Mean	0
Standard deviation	1

The 99% confidence interval is then

$$\bar{x} - 2.576s < \beta < \bar{x} + 2.576s$$

$$544.84 < \beta < 1163.8$$



99% confidence integral for the current gain

Example 5: What is the 100% confidence interval?

- Repeat but with each tail being 0%
- The z-score is infinity

$$-\infty < \beta < \infty \qquad\qquad p = 100\%$$

100% probability makes no sense.

- Nothing is 100% certain
- We aren't even 100% certain the world exists

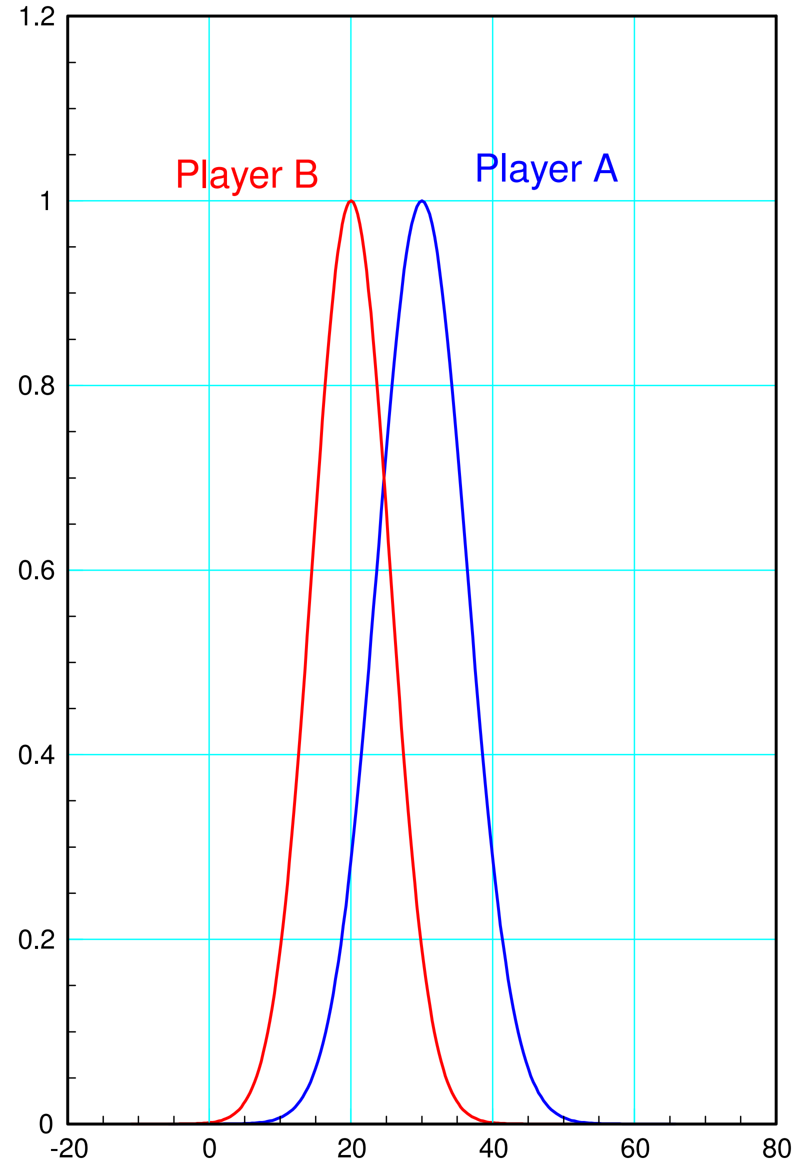
Normal Distributions: A vs B

Suppose you have two normal distributions

$$A \sim N(\mu_a, \sigma_a^2)$$

$$B \sim N\left(\mu_b, \sigma_b^2\right)$$

What is the probability that a random sample from A will be larger than a random sample from B?



New Distribution: W = A - B

Solution: Create a new distribution

$$W = A - B$$

The mean and variance of W are

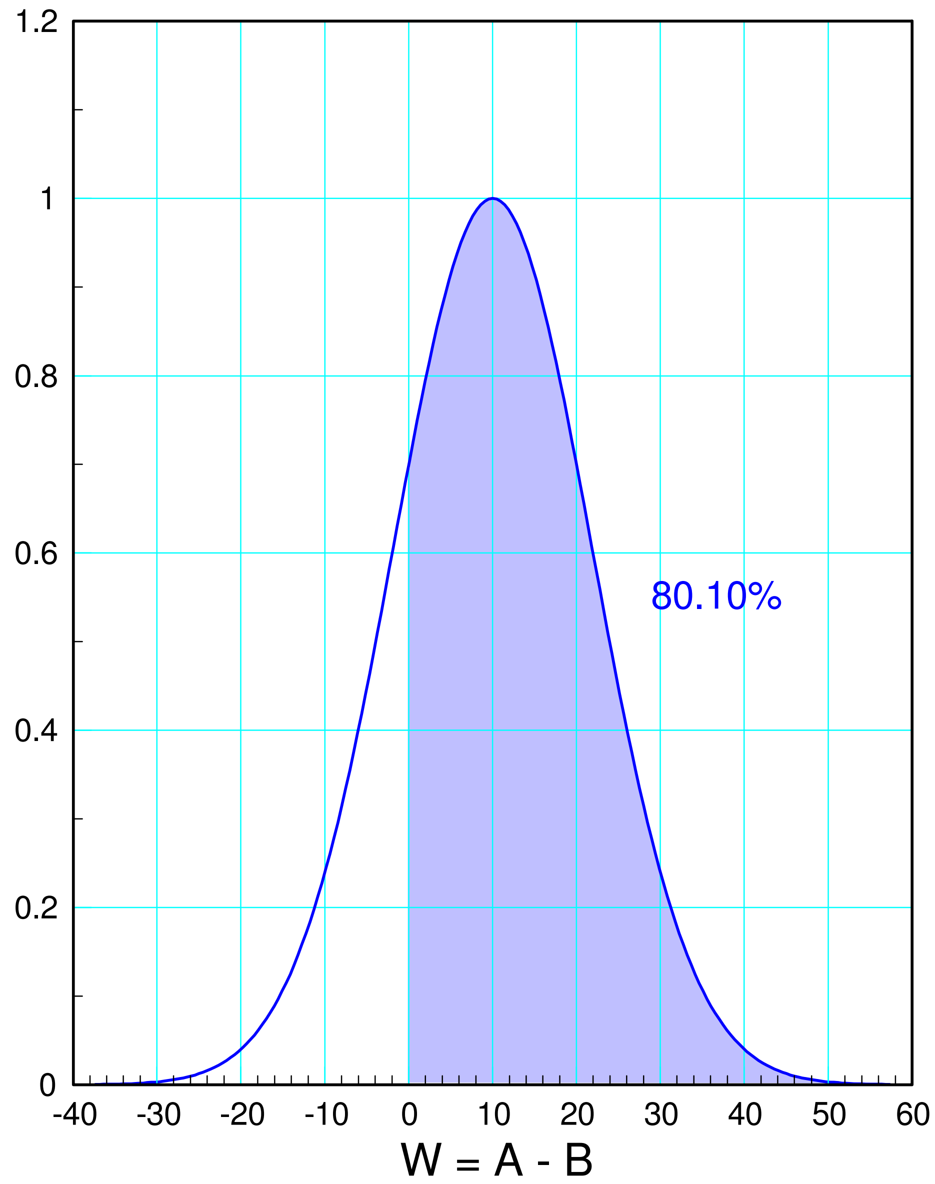
$$\mu_w = \mu_a - \mu_b$$

$$\sigma_w^2 = \sigma_a^2 + \sigma_b^2$$

Find the probability that W > 0

```
>> z = Xw / Sw
z =     0.8452

>> p = (1 - erf(-z/sqrt(2)))/2
p =     0.8010
```

A has a 80.10% chance of winning



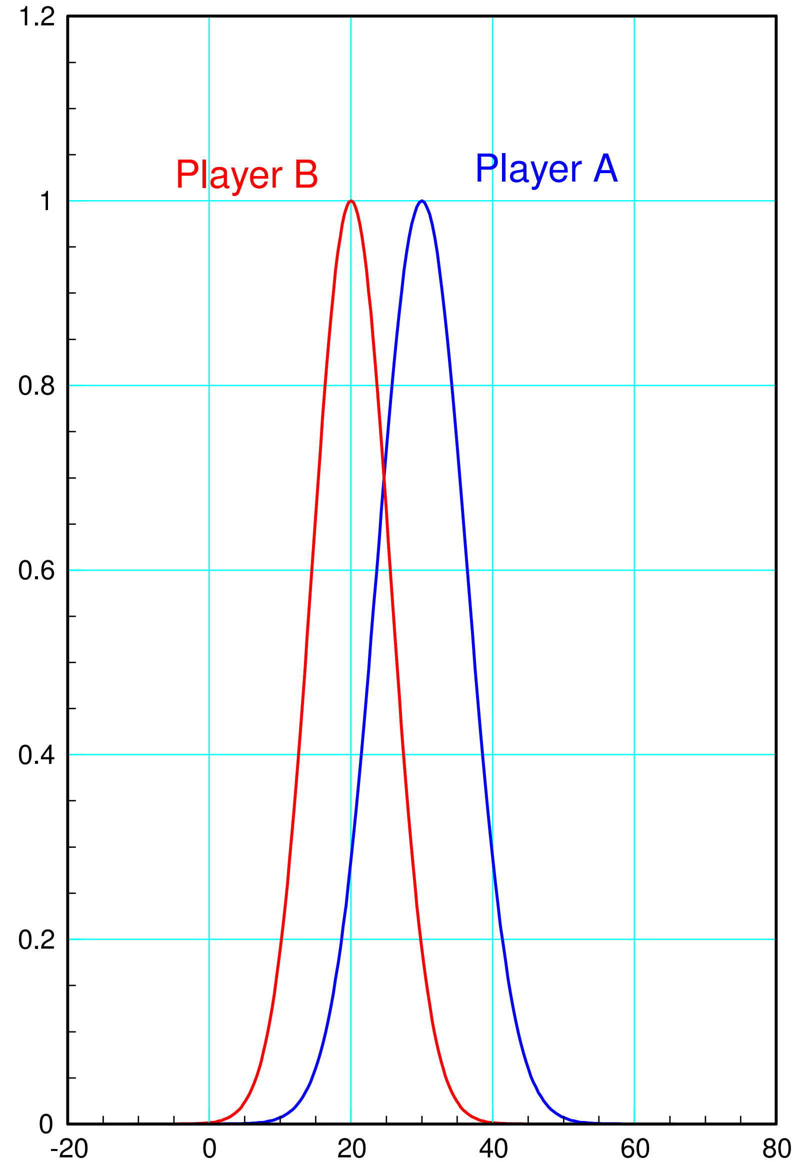
Example

Let A and B be playing a game

- A and B have normal distributions
- mean(A) = 30, var(A) = 80
- mean(B) = 20, var(B) = 60

What is the probability that

- A wins the next game?
- A wins by more than 15 points?



A wins:

Create a new variable, W = A-B

$$\mu_w = 30 - 20 = 10$$

$$\sigma_w^2 = 60 + 80 = 140$$

Find the z-score for W=0

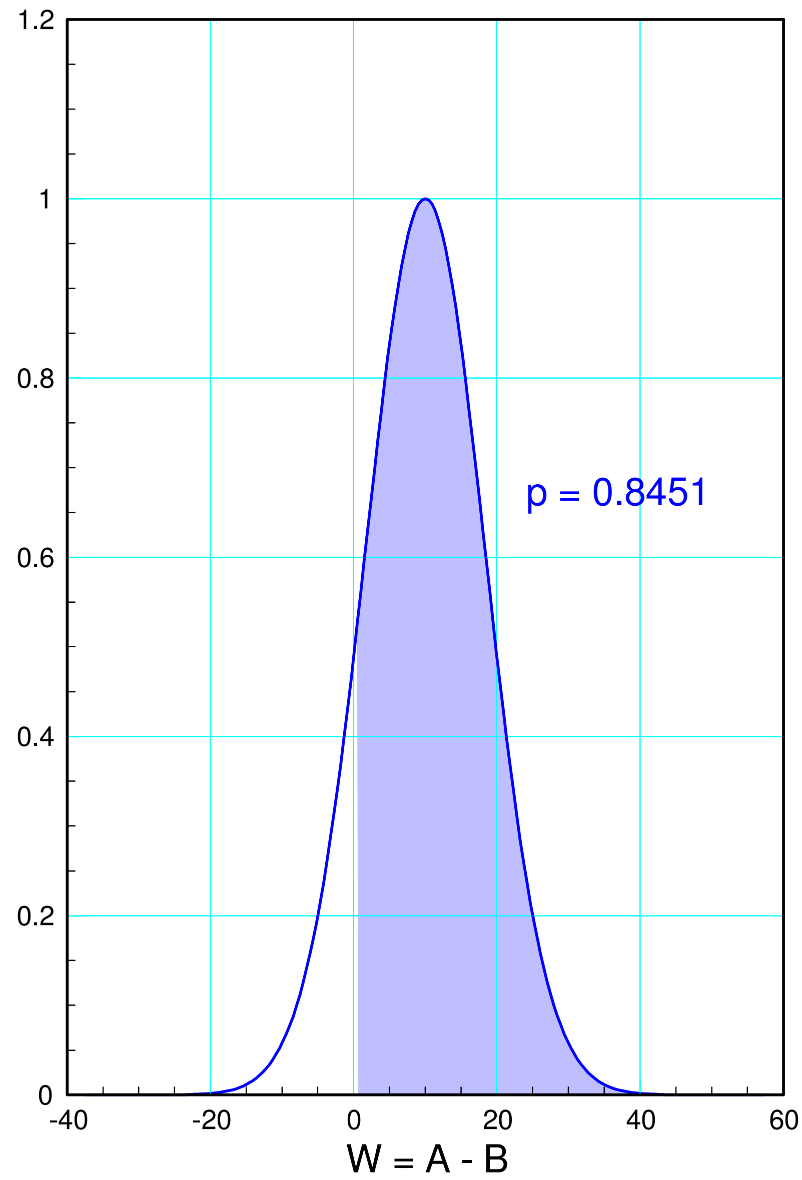
- A wins

$$z = \left(\frac{10-0}{\sqrt{140}}\right) = 0.8451$$

Convert to a probability

```
>> z = Xw / Sw
z =      0.8452

>> p = (1 - erf(-z/sqrt(2)))/2
p =      0.8010
```

A has an 80.10% chance of winning



A wins by 15

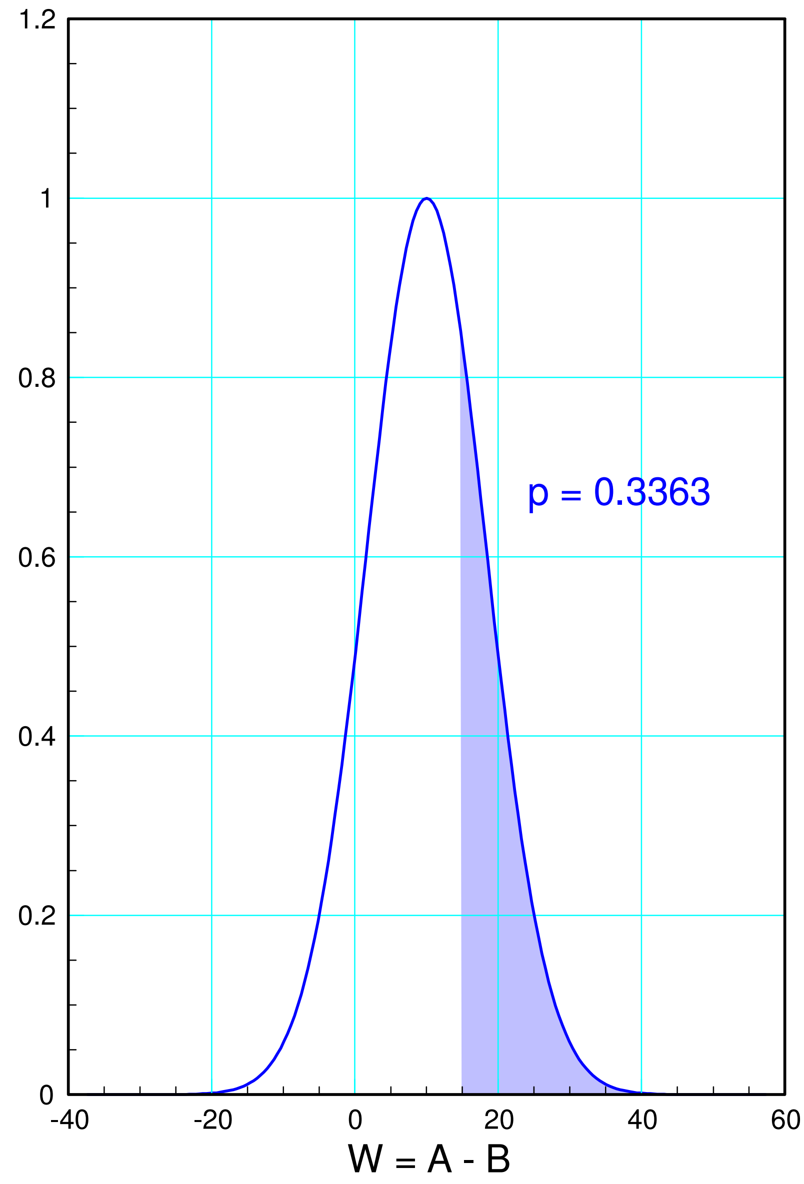
Find the z-score for W=15

$$z = \left(\frac{10-15}{\sqrt{140}}\right) = -0.422577$$

Convert to a probability

p = 0.3363

A has a 33.6% chance of winning by 15



Example: Zetex Transistors

Going back to the Zetex transistors, suppose you split these into two populations:

Population A (First 30)

```
915, 602, 963, 839, 815, 774, 881, 912, 720, 707,
800, 1050, 663, 1066, 1073, 802, 863, 845, 789, 964,
988, 781, 776, 869, 899, 1093, 1015, 751, 795, 776
```

Population B (Next 30)

```
860, 990, 762, 975, 918, 1080 774, 932, 717, 1168,
912, 833, 697, 797, 818, 891, 725, 662, 718, 728,
835, 882, 783, 784, 737, 822, 918, 906, 1010, 819
```

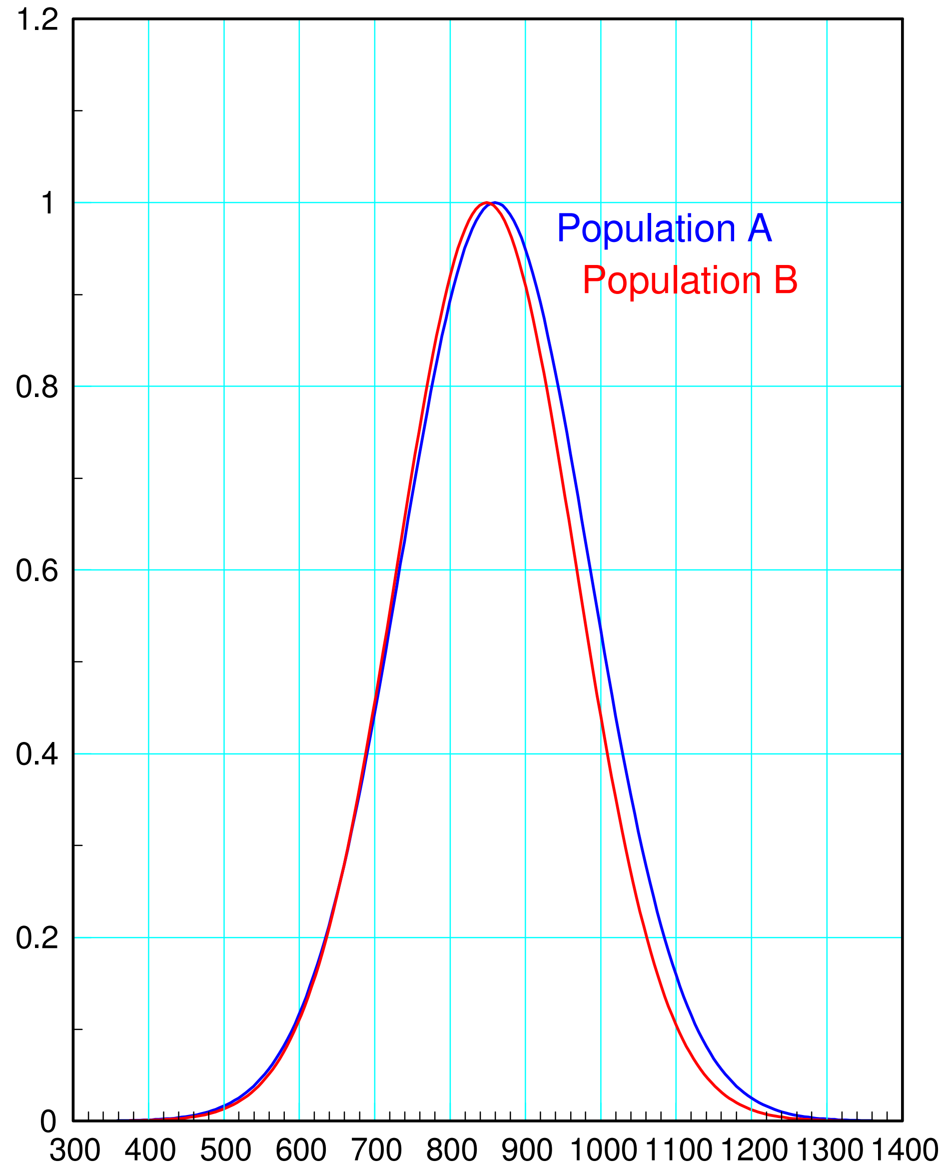
What is the probability that a transistor from Population A has a higher gain than a transistor from Population B?

Step 1:

Find the Mean and Variance:

	Mean	Variance
Population A	859.5333	1568.1
Populaion B	848.4333	1400.5

```
Xa = mean(A)
Sa = std(A)
Xb = mean(B)
Sb = std(B)

s = [-4:0.01:4]';
p = exp(-s.^2/2);
plot(s*Sa+Xa,p,'b',s*Sb+Xb,p,'r
');
```



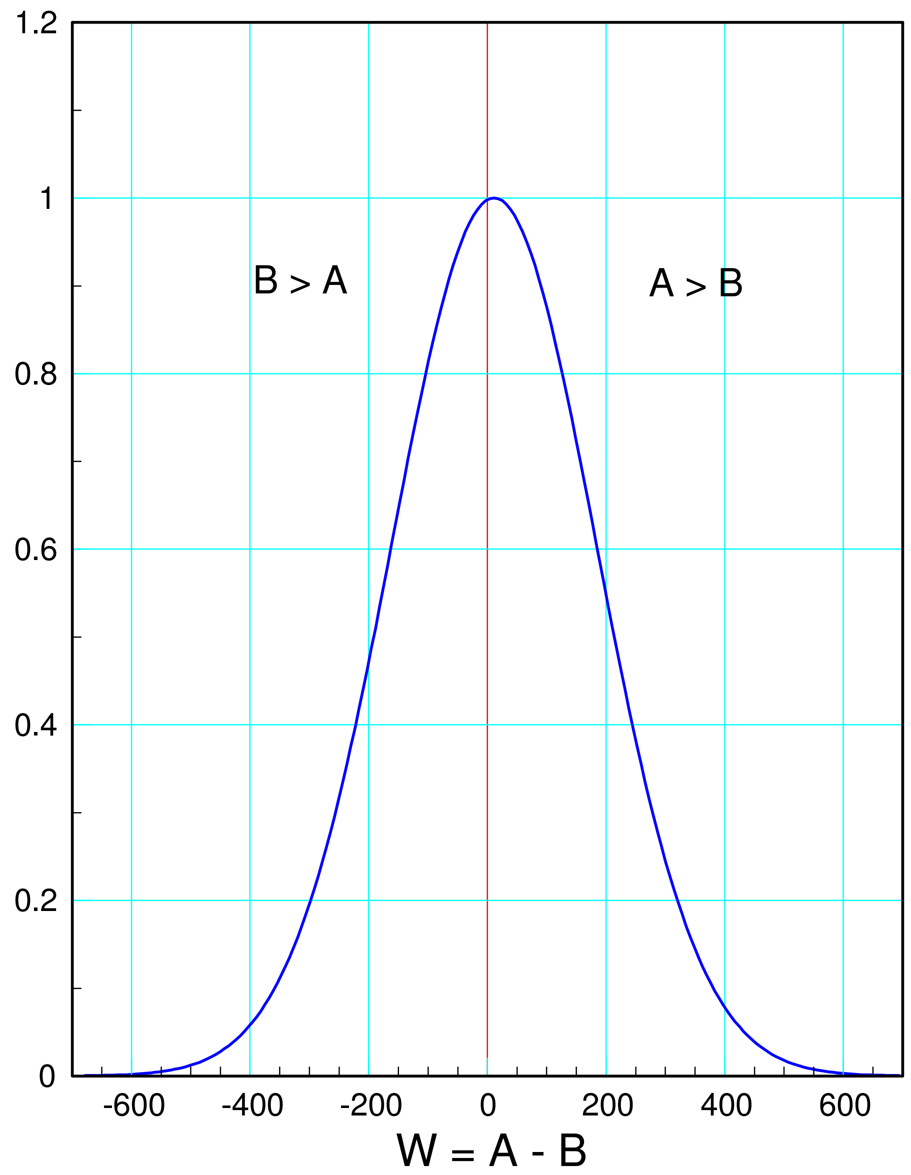
Step 2: W = A - B

Create a new variable, W = A - B

- $\mu_w = \mu_a - \mu_b$
- $\sigma_w^2 = \sigma_a^2 + \sigma_b^2$

In Matlab

```
>> Xw = mean(A) - mean(B)

Xw =  11.100

>> Vw = var(A) + var(B)

Vw =  29685

>> Sw = sqrt(Vw)

Sw =  172.29
```



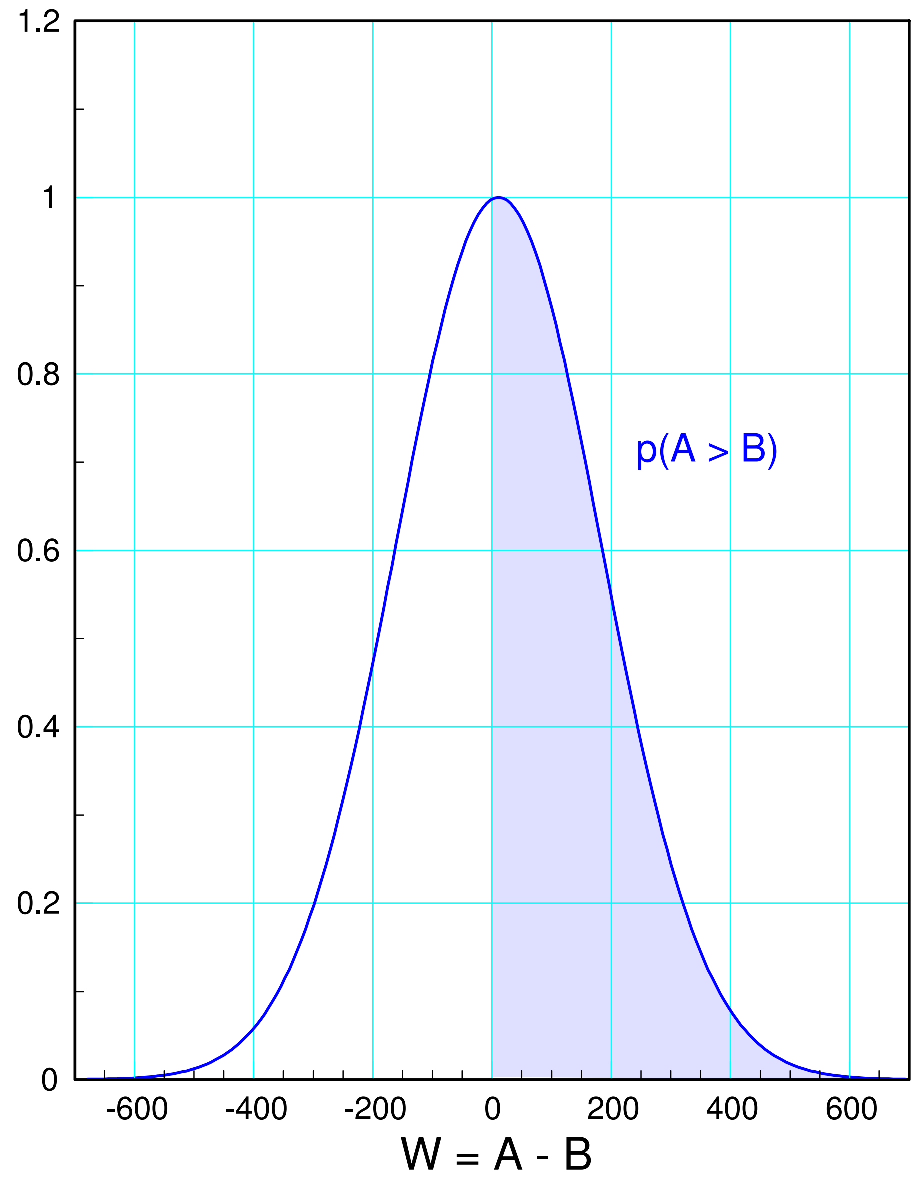
Step 3: Find the area A>B

Probability that A > B

```
>> z = Xw / Sw

z =  0.064425

>> p = (1 - erf(-z/sqrt(2)))/2

p =  0.52568
```

A has a 52.568% chance of having a higher gain

- Essentially a coin flip
- A & B are from the same populations



Testing: (False Positives, False Negatives)

Assume

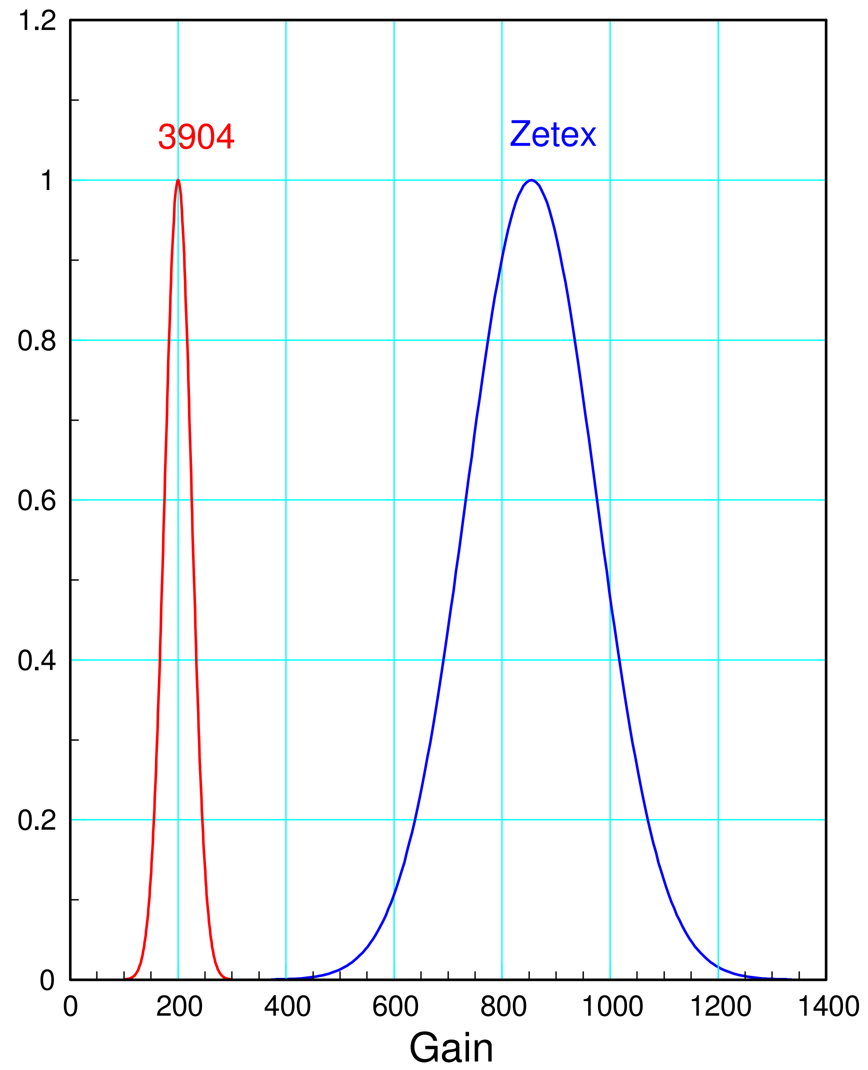
- A population falls into two groups, A and B.
- You run a test on individual X
  - The results differ for each group:

Based upon the test result, does X belong to

- Population A or
- Population B?

Example: The gain of a transistor

- Population A: Zetex 1051a transistors
  - $\mu_A = 854.1290$ mean
  - $\sigma_A = 120.2034$ standard deviation
- Population B: 3904 transistor
  - $\mu_B = 200$ mean
  - $\sigma_B = 25$ standard deviation



Test Procedure

Pick a threshold

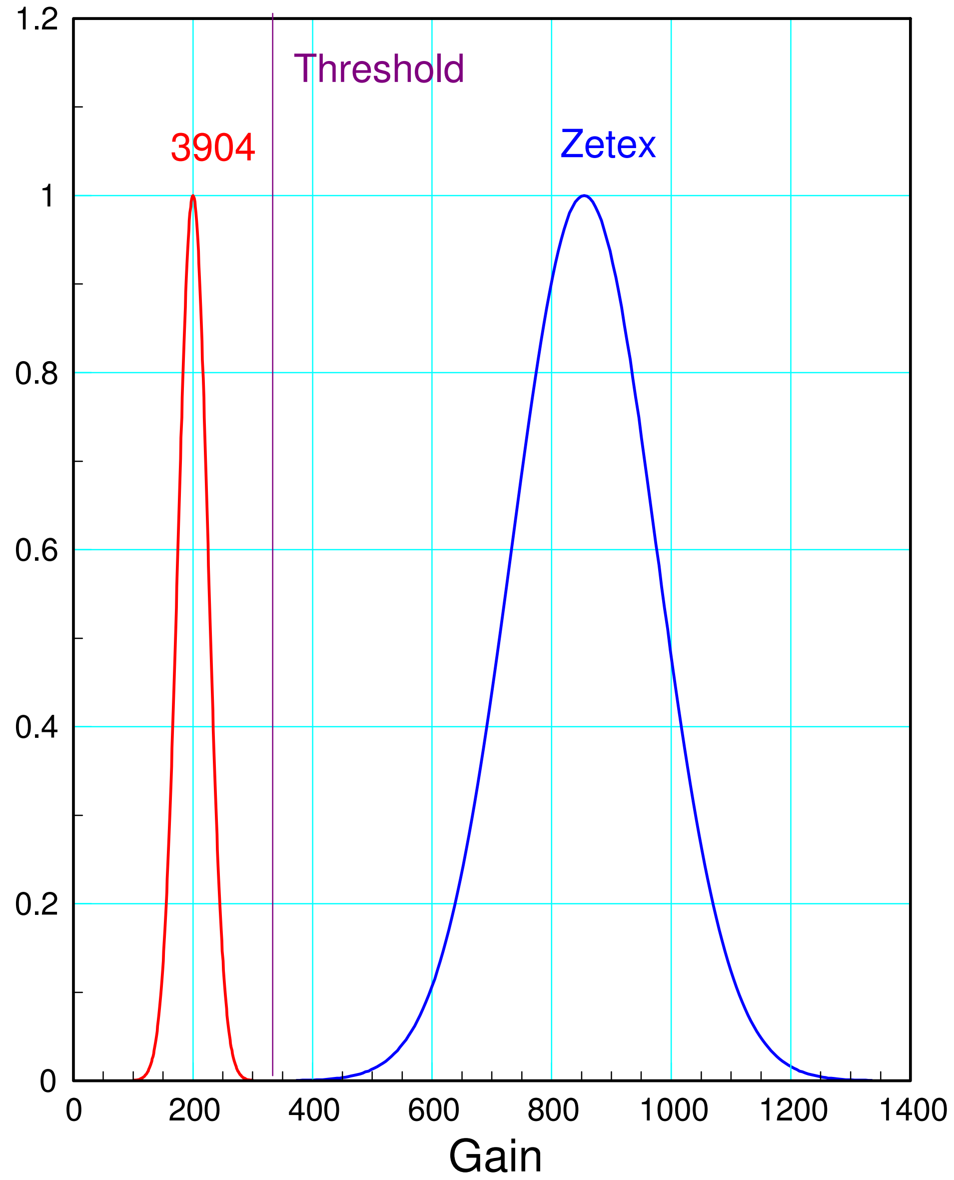
- 375 works here

Measure the gain of a transistor

- If it's more than 375, it's a Zetex transistor
- If it's less, it's a 3904 transistor

This is the ideal situation

- Both populations are distinct
- There is very little overlap



More Typical Situation

Let population A have (positive)

- mean = 150 standard deviation = 35

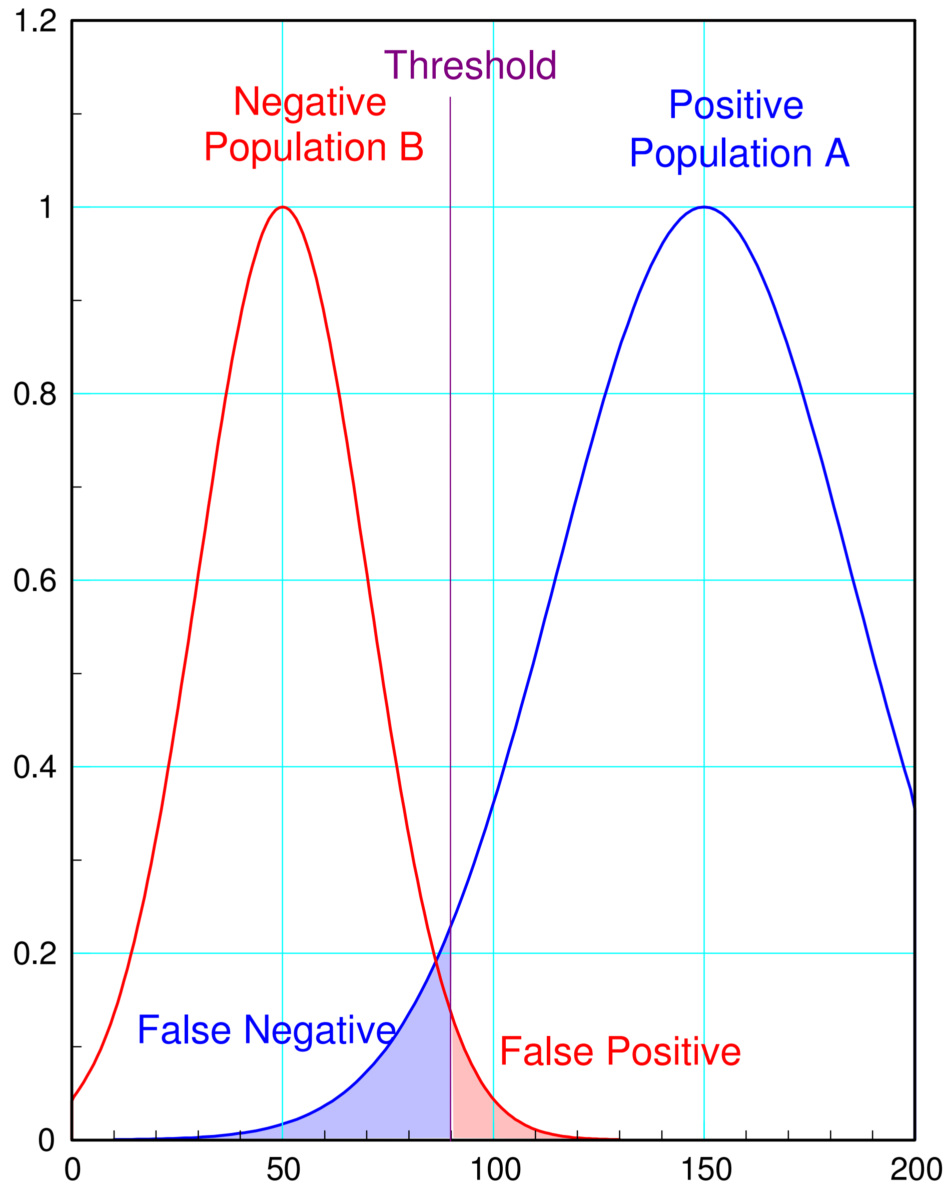
Let population B have

- mean = 50 standard deviation = 20

Since there's overlap

- There's a chance a sample from populaton B will be assigned to population A
  - False Positive
- There's a chance a sample from population A will be assigned to populaton B
  - False Negative

How do you choose the threshold?



Case 1: p(False Negative) = 1%

Assume false-negatives are bad

- Keep them at 1%

The z-score for 1% tails is 2.236

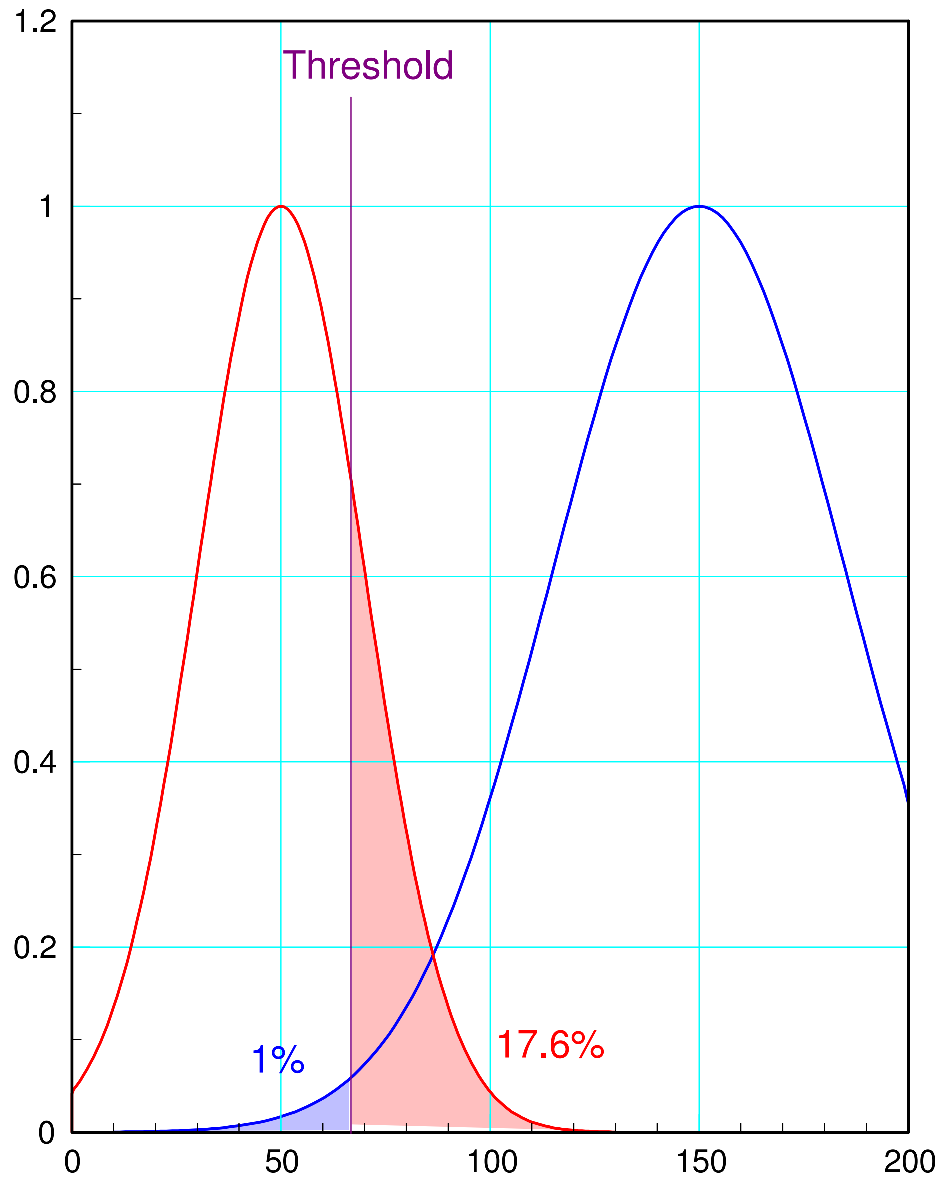
The threshold is

$$T = \mu_a - 2.236\sigma_a$$

$$T = 68.59$$

But.. p(False Positive) = 17.63%

```
>> z = (68.59-Xb)/Sb
z =     0.9295

>> p = (1 - erf(z/sqrt(2)))/2
p =     0.1763
```



Case 2: p(False Positive) = 1%

Assume false-positives are bad

- Keep them at 1%

The z-score for 1% tails is 2.236

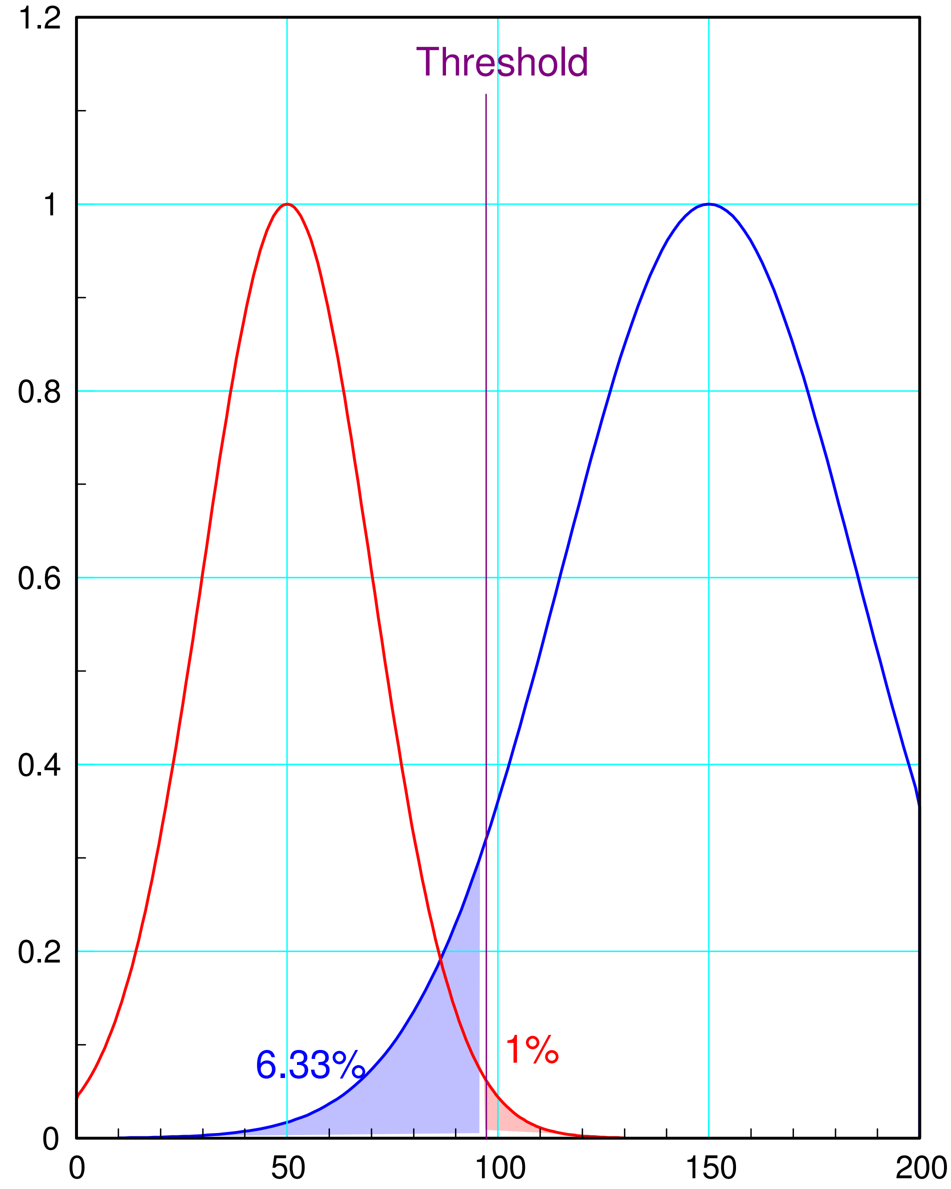
The threshold is

$$T = \mu_B + 2.326\ \sigma$$

$$T = 96.52$$

But.. p(False Negative) = 6.33%

```
>> z = (96.52-Xa)/Sa
z =   -1.5280

>> p = (1 - erf(-z/sqrt(2)))/2
p =    0.0633
```



Case 3: p(FP) = p(FN)

For equal probability, set the z-scores to match

$$z = \left(\frac{\mu_A - T}{\sigma_A}\right) = \left(\frac{T - \mu_B}{\sigma_B}\right)$$

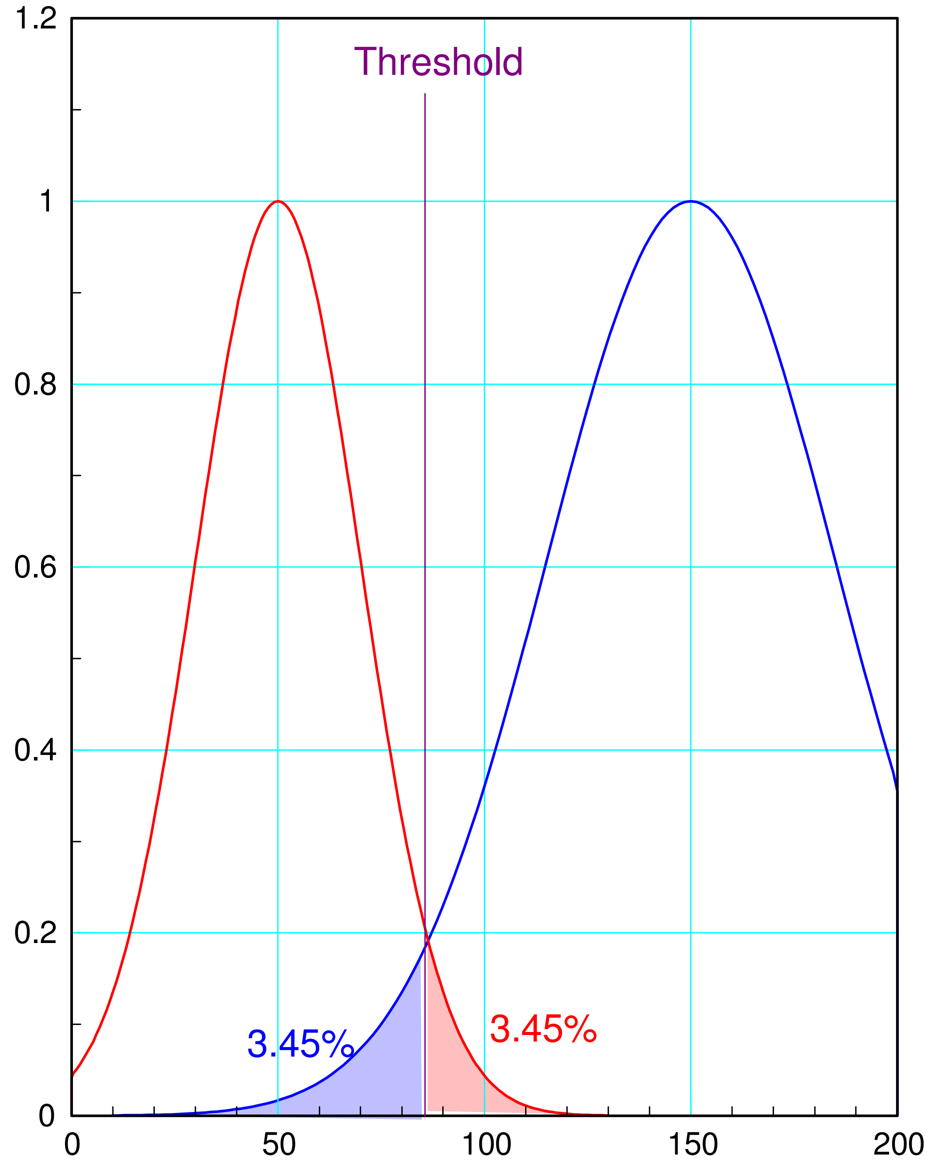
$$\left(\frac{150 - T}{35}\right) = \left(\frac{T - 50}{20}\right)$$

$$T = \left(\frac{20 \cdot 150 + 50 \cdot 35}{20 + 35}\right) = 86.36$$

The probability of false readings is 3.45%

$$z = \left(\frac{86.36 - 50}{20}\right) = \left(\frac{150 - 86.36}{35}\right) = 1.81812$$

$$p = 0.0345$$



Greater Certainty

If you want to be more certain, you could run a separate (independent) test.

If this test also has a 3.5% error rate, then the two tests will give

- Correct results 93.12% of the time
  - Both read positive (correctly)
  - Both read negative (correctly)
  - $p = (0.965)(0.965) = 0.93127$
- Incorrect results 0.12% of the time
  - $p = (0.035)(0.035) = 0.0012$
- Conflicting results 6.76% of the time
  - One says yes
  - One says no

With two tests, you can greatly reduce the probability of a false positive or false negative.

Summary

If a population has a normal distribution, you can determine

- The probability $a < x$
- The probability $a < x < b$

To do so, you use the z-score

If you want to compare two variables with normal distributions

- Create a new variable, $W = A - B$
- $\text{mean}(W) = \text{mean}(A) - \text{mean}(B)$
- $\text{var}(W) = \text{var}(A) + \text{var}(B)$

If you have two populations, you can assign a test result to Populaton A or Population B

- There *will* be incorrect results, however (false positives and false negatives)
- The probability of each depends upon where you place the threshold