No statistician left behind; or how (not) to measure the achievement gap

Thomas Pietraho

Before we begin

Federico Ardila enunciated the following axioms in his *Todos Cuendan*. They form a lens through with I view both teaching and doing mathematics.

- Mathematical potential is distributed equally among different groups, irrespective of geographic, demographic, and economic boundaries.
- Everyone can have joyful, meaningful, and empowering mathematical experiences.
- Mathematics is a powerful, malleable tool that can be shaped and used differently by various communities to serve their needs.
- Every student deserves to be treated with dignity and respect.

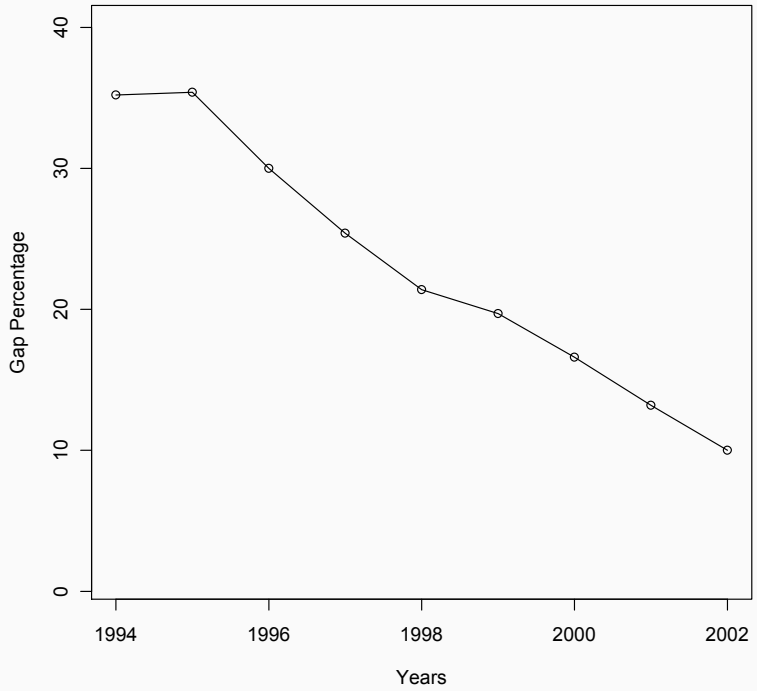
The Texas Success Story

G.W. Bush, address to Urban League, July 2003

"In my state, 73 percent of the White students passed the math test in 1994, while only 38 percent of Black students passed it. So we made that the point of reference. We had people focused on the results for the first time – not process, but results. And because teachers rose to the challenge, because the problem became clear, that gap has now closed to 10 points. Because every child can learn, you've just got to focus the attention and the resources when necessary."

The Texas Success Story

Table 1: White-Black achievement gap; percentage of White students minus percentage of Black students passing state math exam. Grades 3 through 8 and 10.



Year	Point Gap
1994	35.2
1995	35.4
1996	30.0
1997	25.4
1998	21.4
1999	19.7
2000	16.6
2001	13.2
2002	10.0

Sources:

- Texas Education Agency
- Charles Murray, "Acid Tests," *Wall Street Journal*
- *The Lion's Paw*: an anonymous blog

But I'm a skeptic

Suppose that we would like to close the **achievement gap** between groups A and B, that is, **the difference between passing rates on an exam.**

Solution

Let everyone from group A and group B pass the exam. The achievement gap is completely eliminated.

But I'm a skeptic

Suppose that we would like to close the **achievement gap** between groups A and B, that is, **the difference between passing rates on an exam.**

Solution

Let everyone from group A and group B pass the exam. The achievement gap is completely eliminated.

Alternate Solution

Fail everyone from group A and group B. No one passes. The achievement gap is completely eliminated.

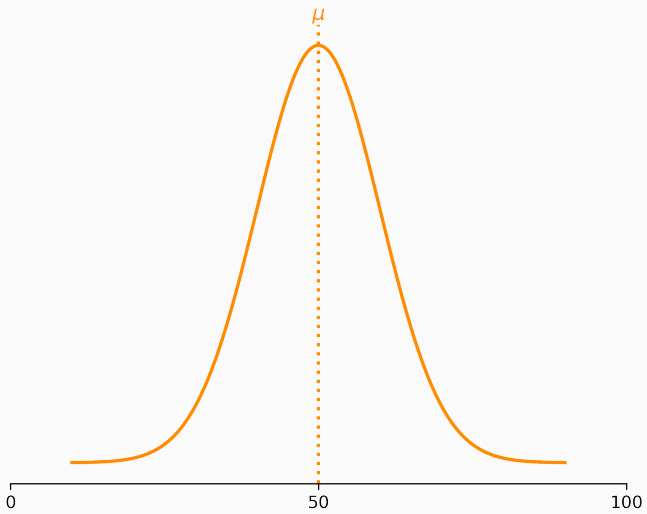
The Statistics of Exams

Observation

Exam distributions tend to be bell-shaped, or Gaussian.

The "bell curve" for the population with mean score μ and standard deviation σ is

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



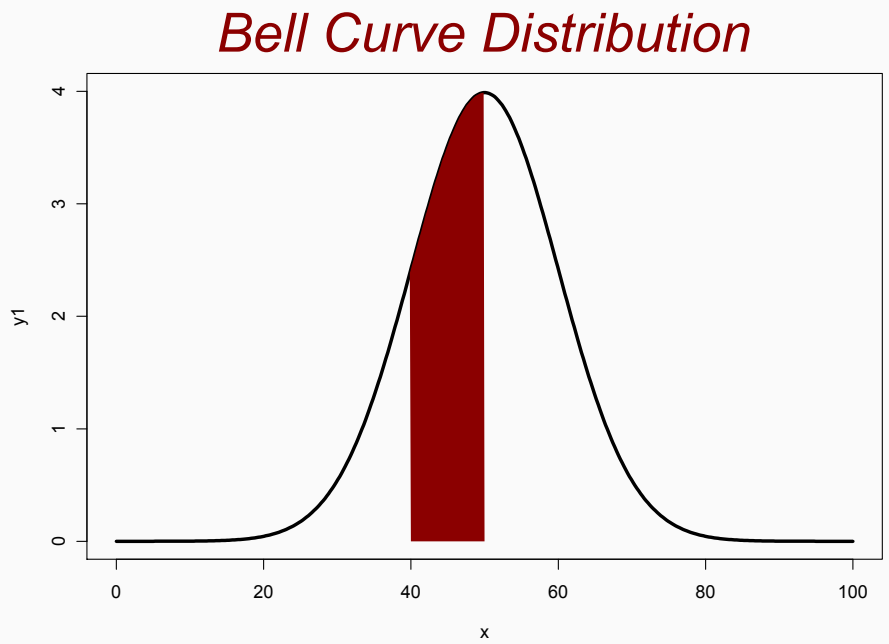
The Statistics of Exams

Observation

Exam distributions tend to be bell-shaped, or Gaussian.

The "bell curve" for the population with mean score μ and standard deviation σ is

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



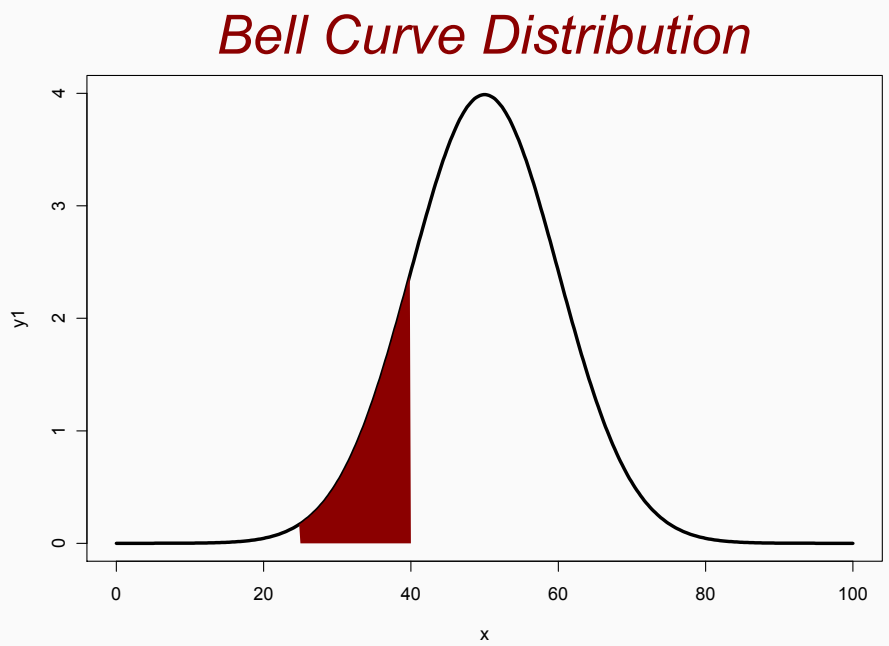
The Statistics of Exams

Observation

Exam distributions tend to be bell-shaped, or Gaussian.

The "bell curve" for the population with mean score μ and standard deviation σ is

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



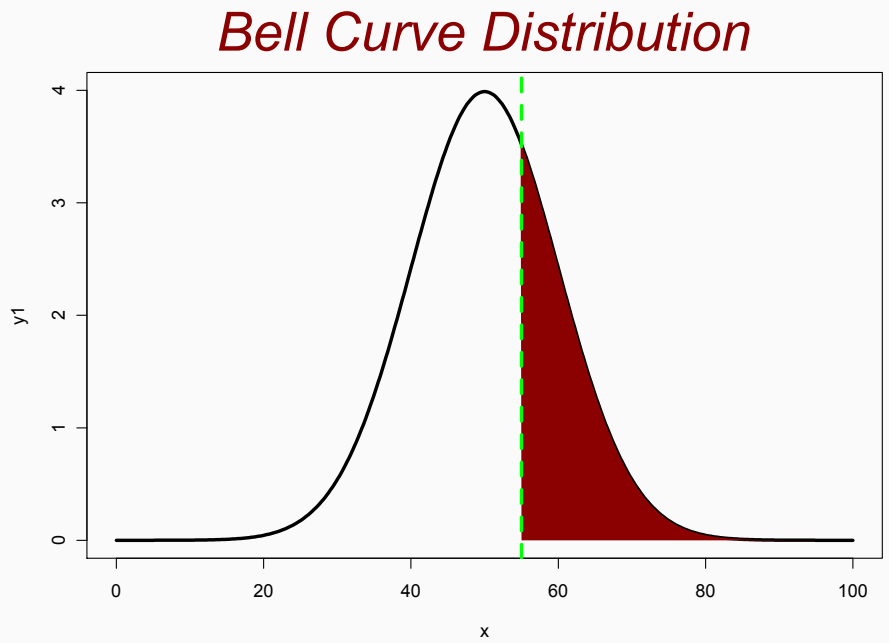
The Statistics of Exams

Observation

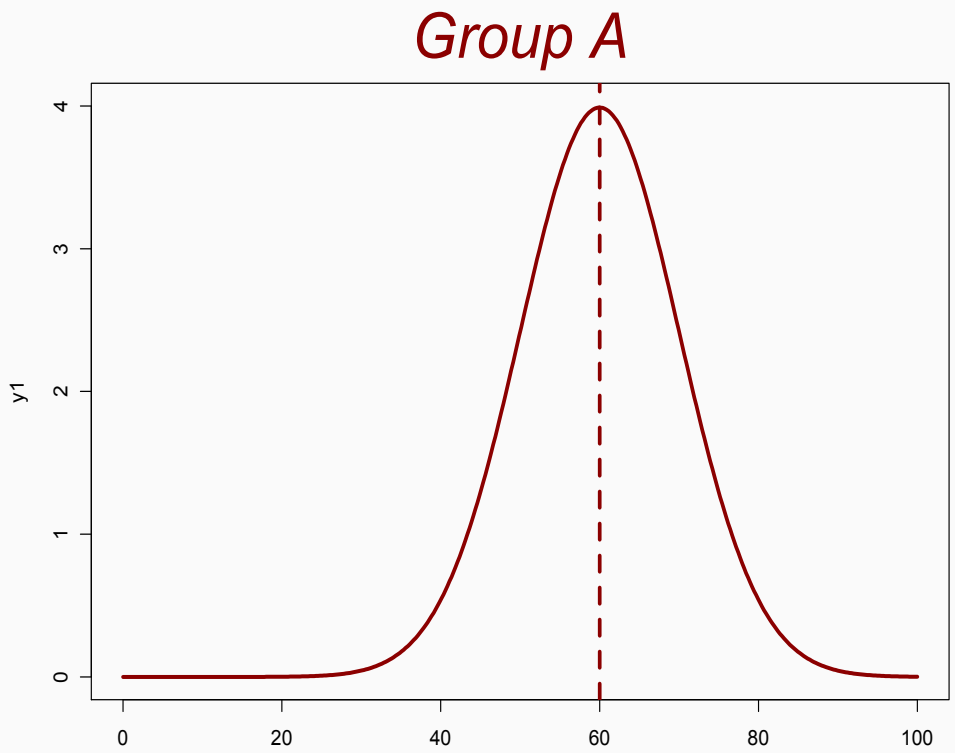
Exam distributions tend to be bell-shaped, or Gaussian.

The "bell curve" for the population with mean score μ and standard deviation σ is

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



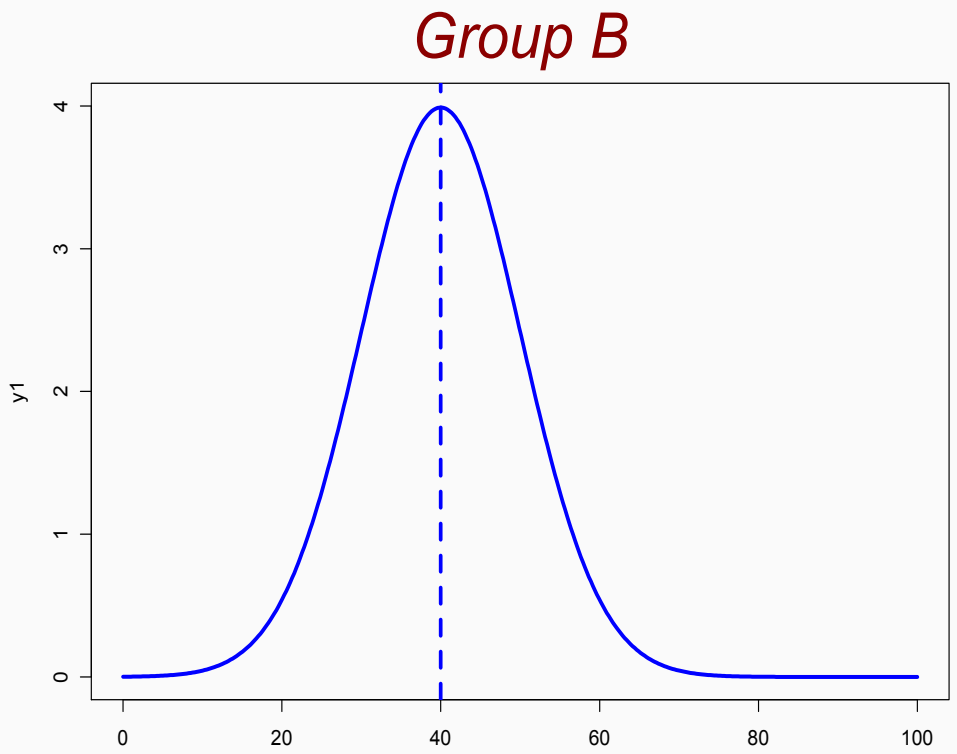
A Tale of Two Groups



Group A

Average test score: 60

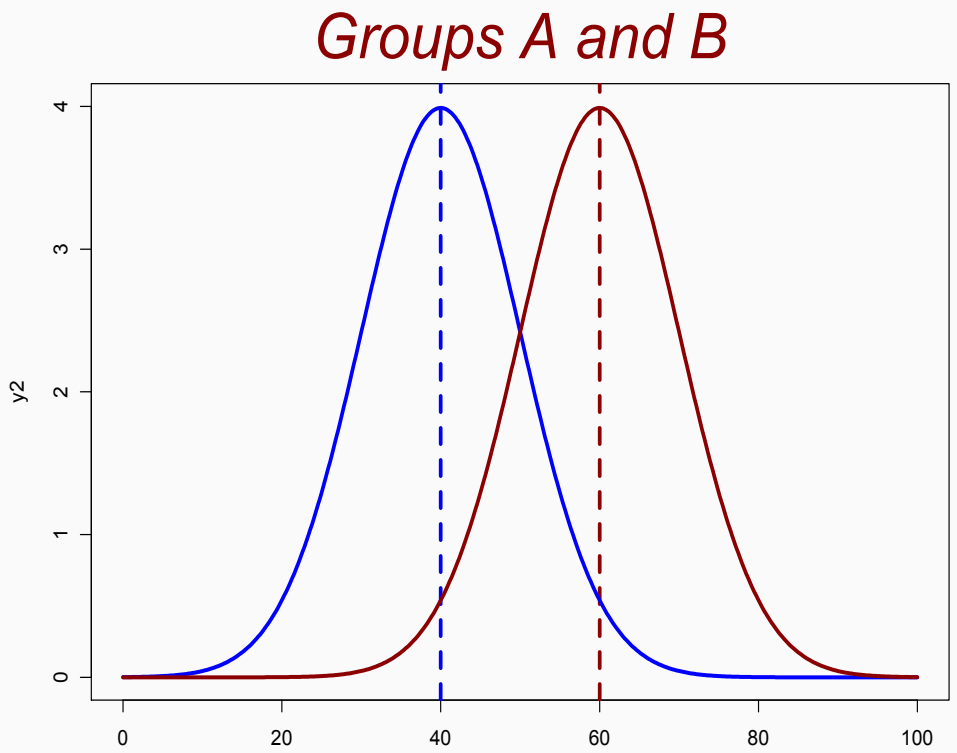
A Tale of Two Groups



Group B

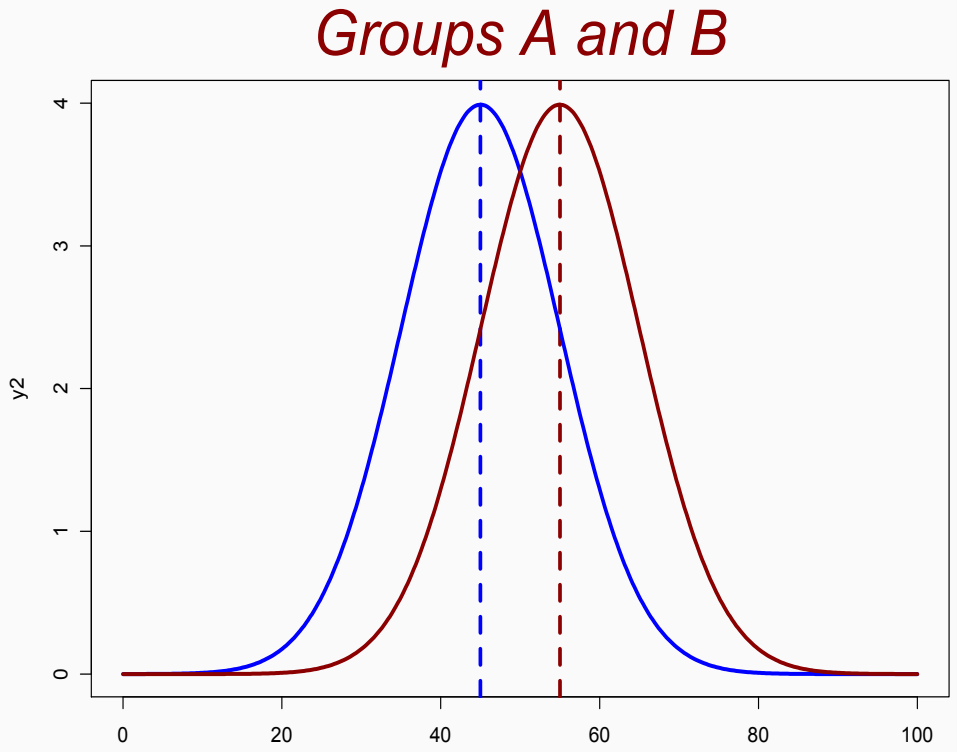
Average test score: 40

A Tale of Two Groups



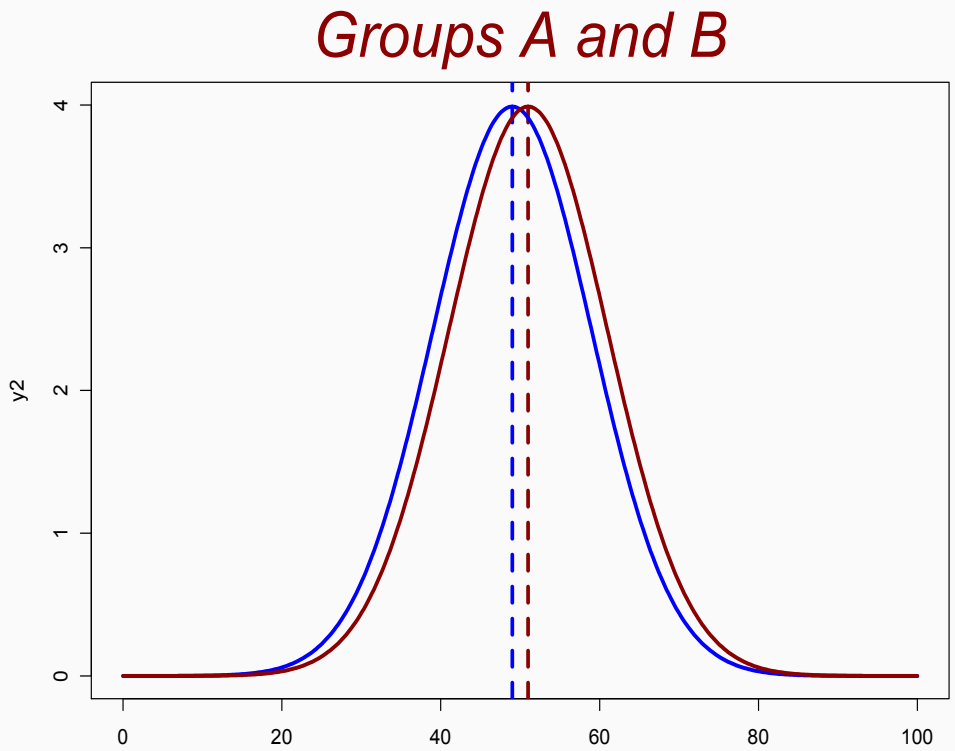
The usual way to measure the difference in performance between these groups is to measure the difference in the two averages, in this case $60 - 40$, or a 20 point difference.

A Tale of Two Groups



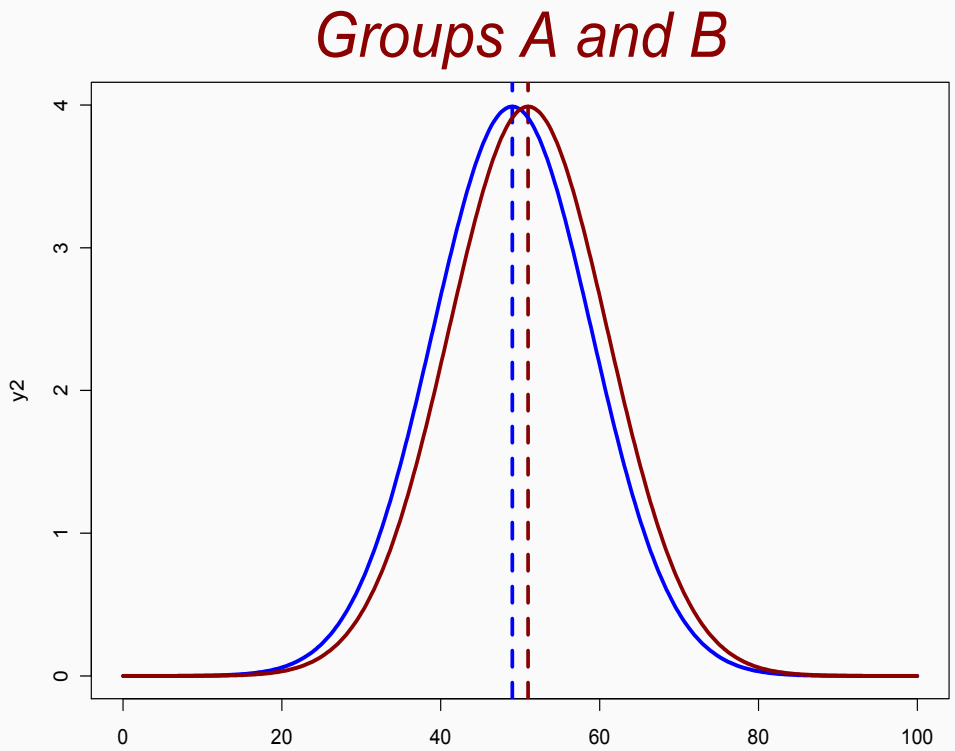
The differences between these two groups begin to disappear as the differences between the average scores decrease.

A Tale of Two Groups



The differences between these two groups begin to disappear as the differences between the average scores decrease.

A Tale of Two Groups



The differences between these two groups begin to disappear as the differences between the average scores decrease.

Question

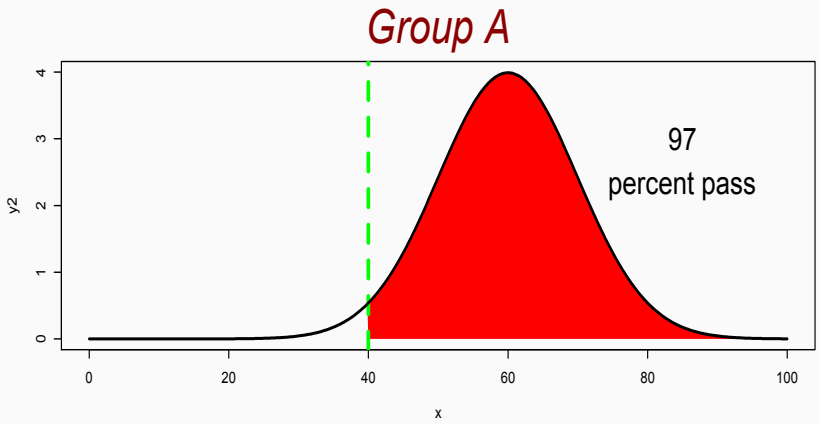
Is this what is going on in Texas?

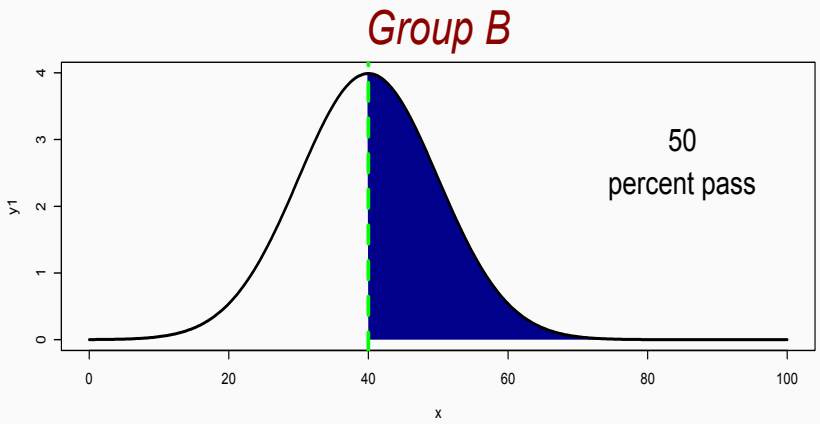
The Texas Approach

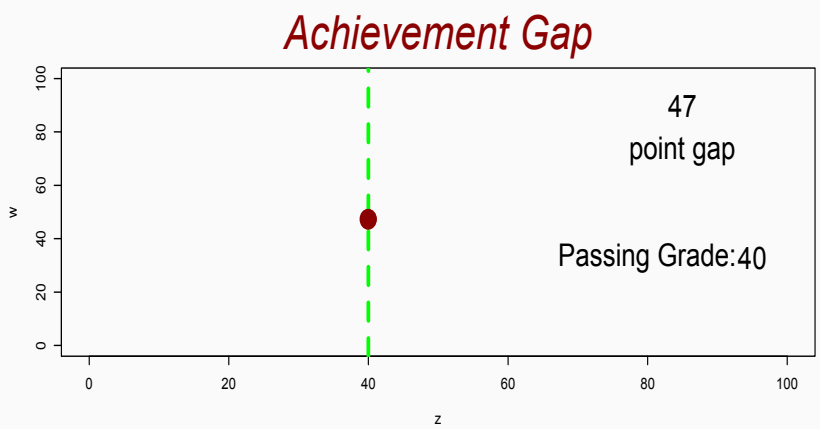
Consider the same two imaginary populations, Group A and Group B.

- Set a passing score on the exam.
- Compute what percentage of students in Group A passed.
- Compute what percentage of students in Group B passed.
- Compute the difference between these percentages.

The difference in percentages of students passing the exam will be called the **achievement gap**.





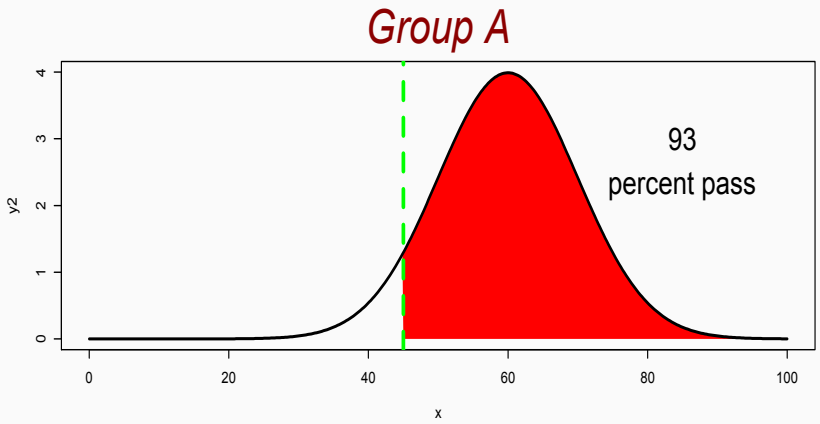


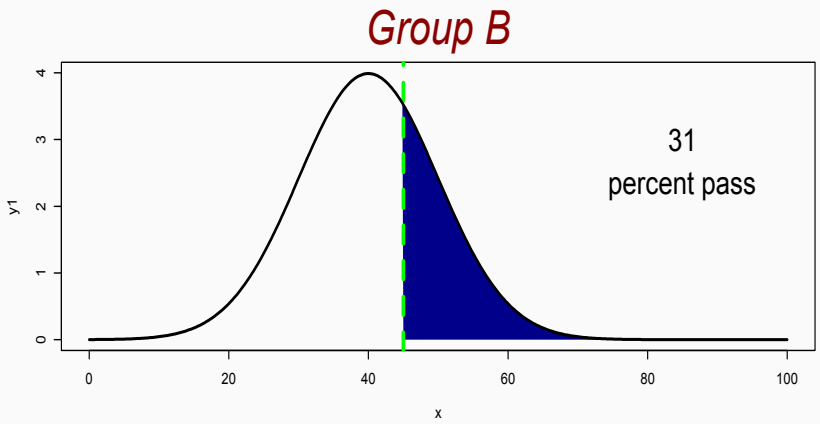
Suppose that the passing grade was set at **40 points**

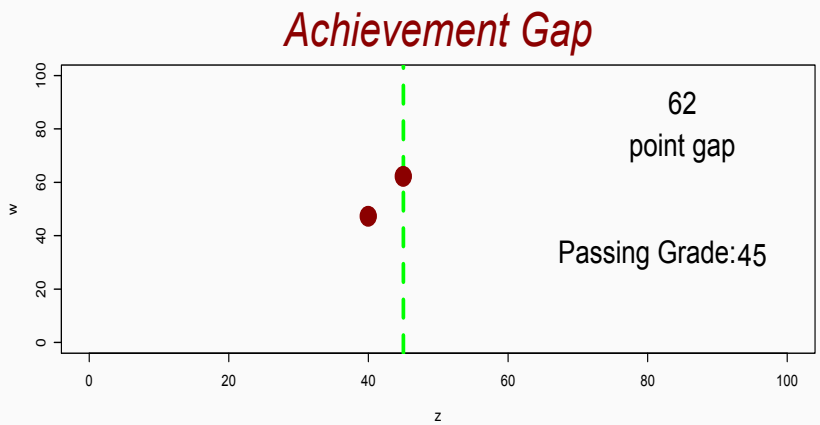
- **97 percent of Group A passes**
- **50 percent of Group B passes**

The achievement gap is:

47 points.





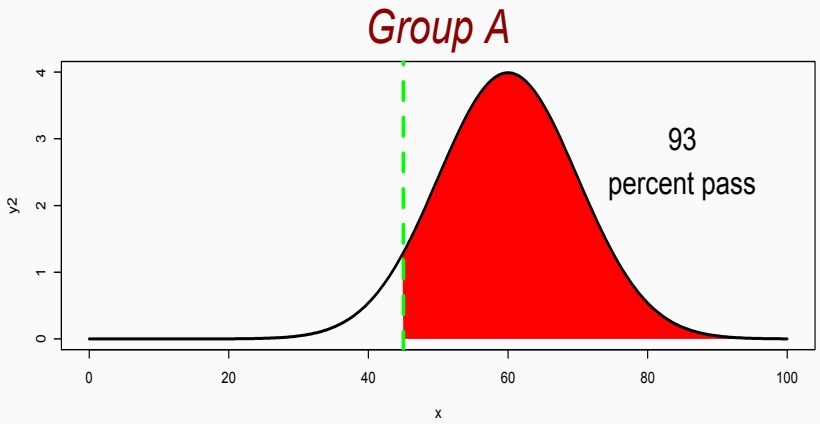


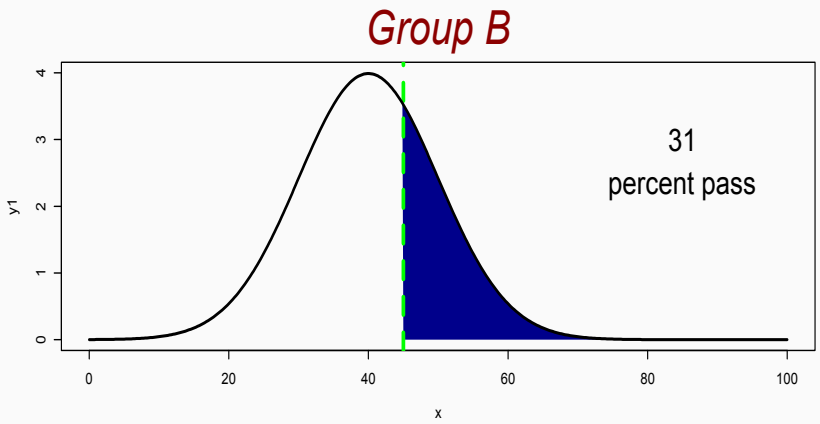
Suppose that the passing grade was set at **45 points**

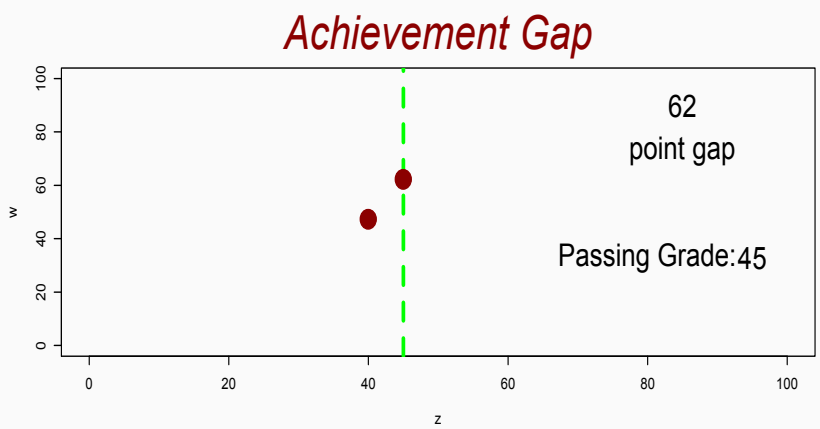
- **93 percent of Group A passes**
- **31 percent of Group B passes**

The achievement gap is:

62 points.

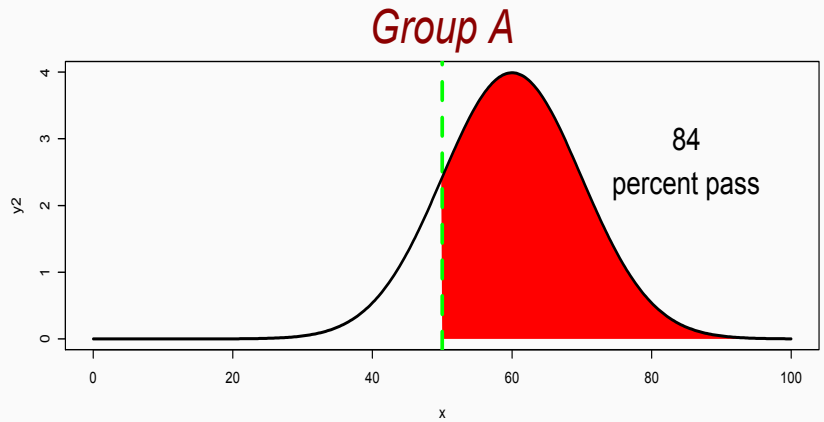


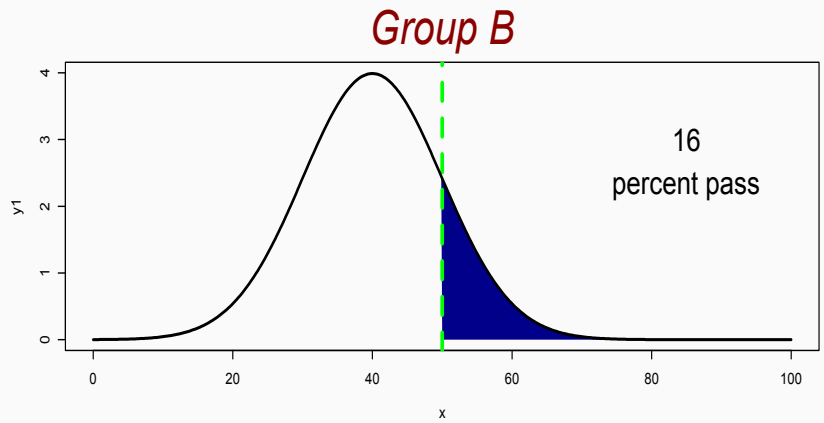


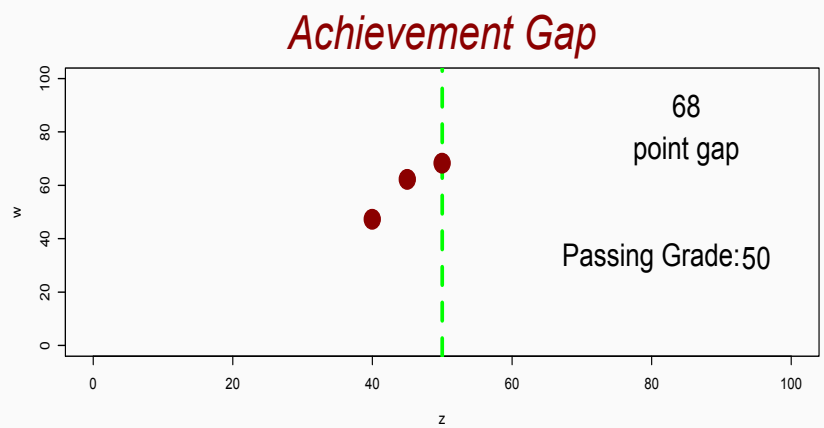


Already we've done something. We've changed the achievement gap from **47** to **62** points.

This was done without changing the populations at all, but merely by changing what we considered a passing score on a test they already took!





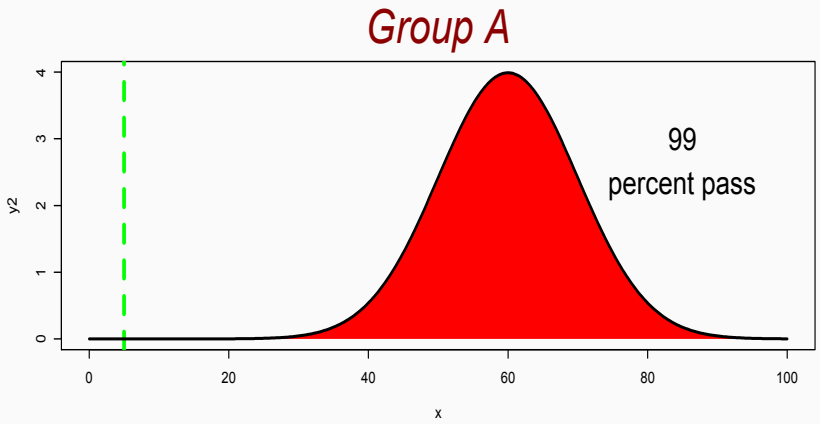


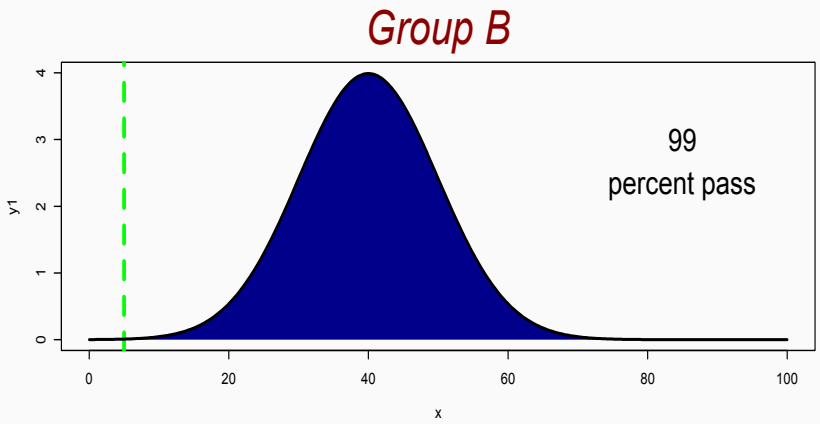
Suppose that the passing grade was set at **50 points**

- **84 percent of Group A passes**
- **16 percent of Group B passes**

The achievement gap is:

68 points.



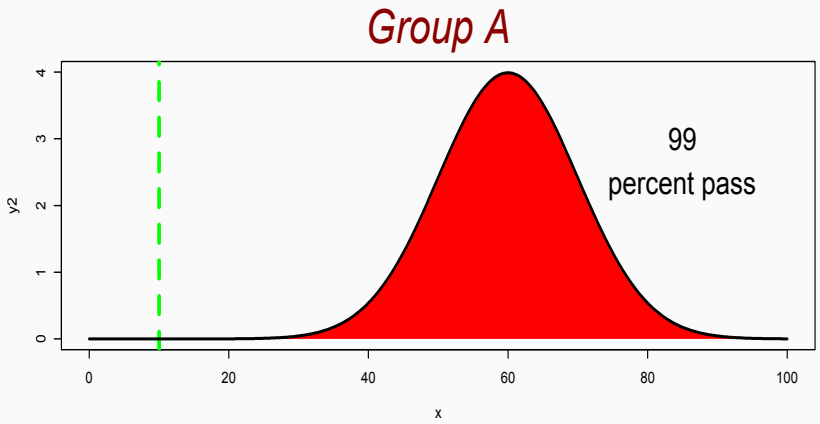


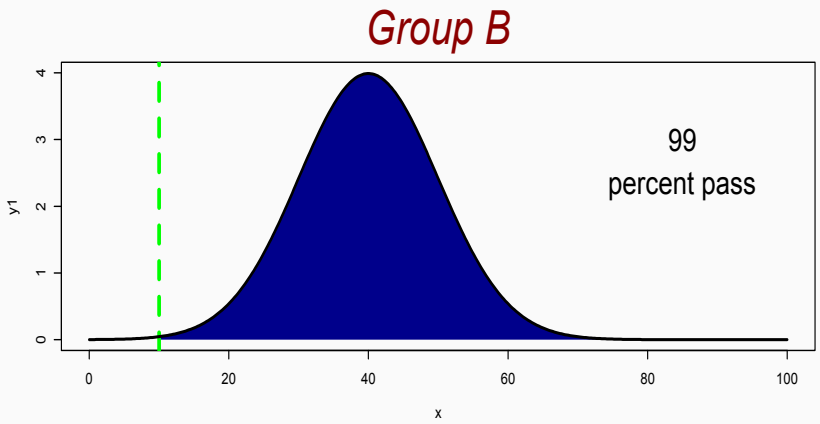
Suppose that the passing grade was set at **5 points**

- **99 percent of Group A passes**
- **99 percent of Group B passes**

The achievement gap is:

0 points.



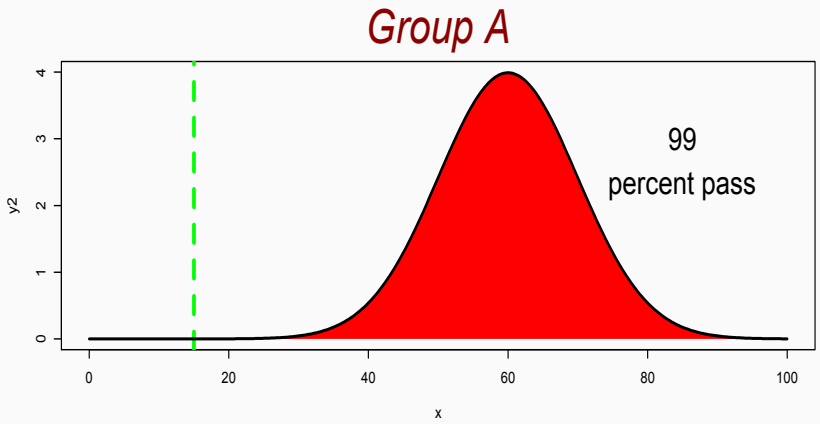


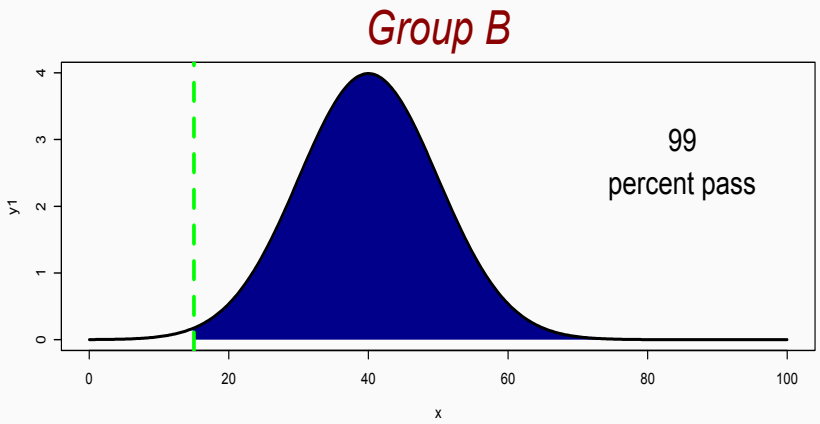
Suppose that the passing grade was set at **10 points**

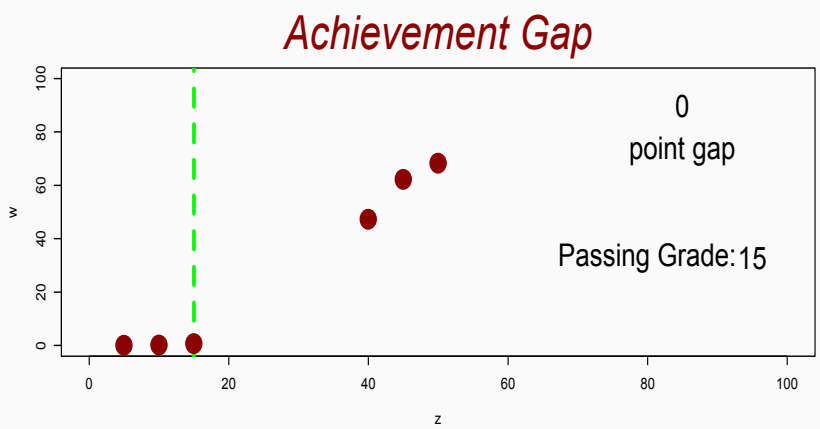
- **99 percent of Group A passes**
- **99 percent of Group B passes**

The achievement gap is:

0 points.





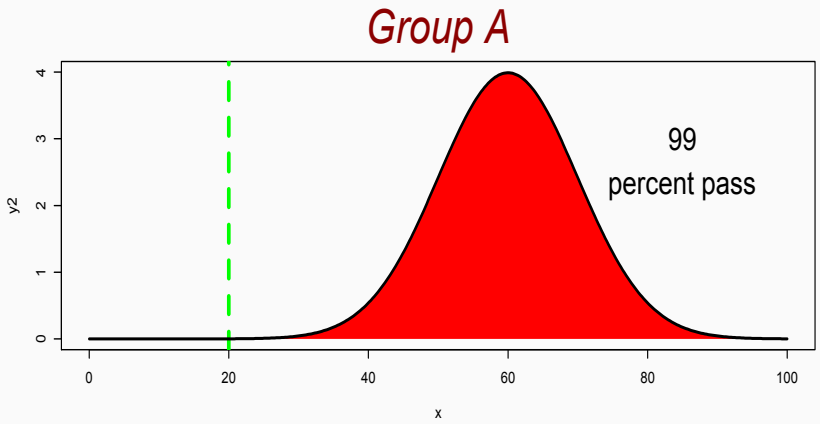


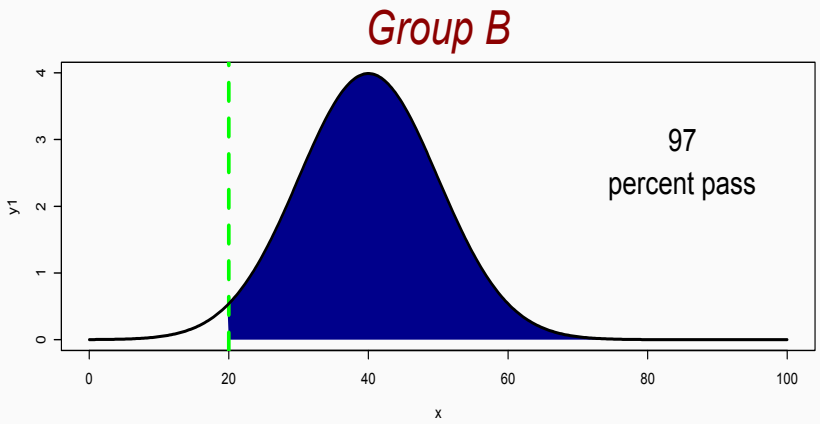
Suppose that the passing grade was set at **15 points**

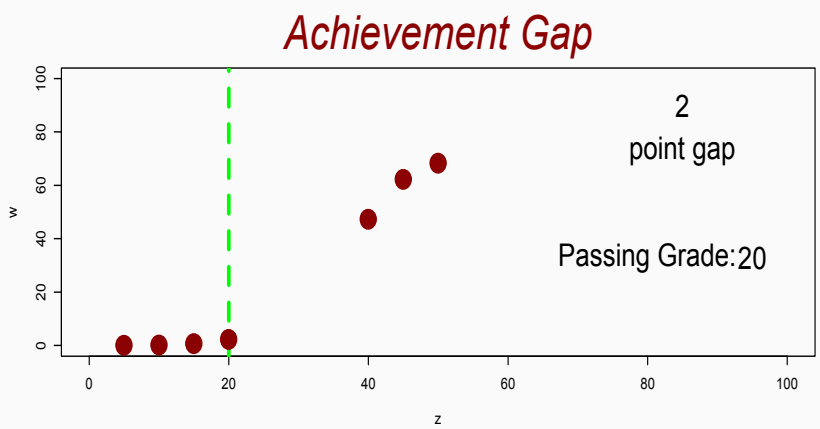
- **99 percent of Group A passes**
- **99 percent of Group B passes**

The achievement gap is:

0 points.





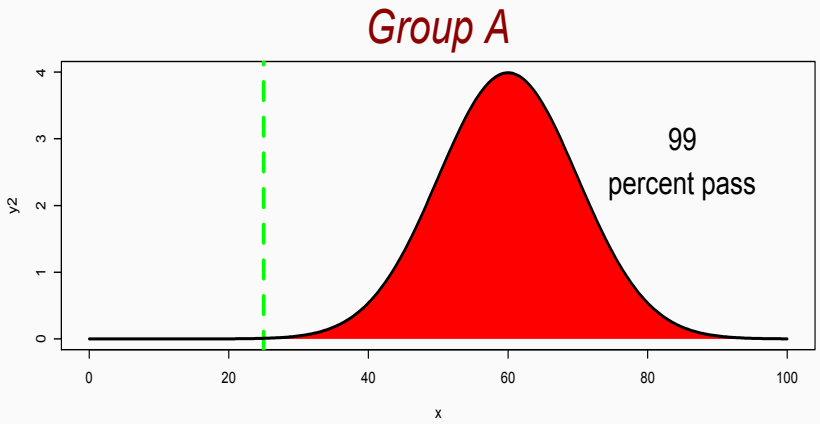


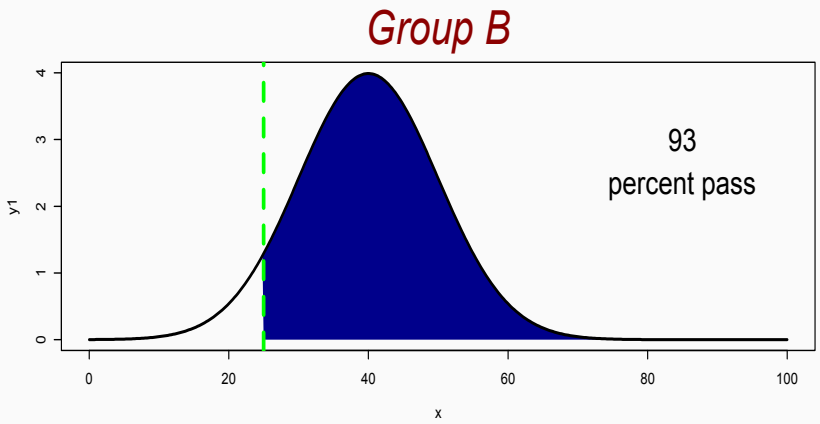
Suppose that the passing grade was set at **20 points**

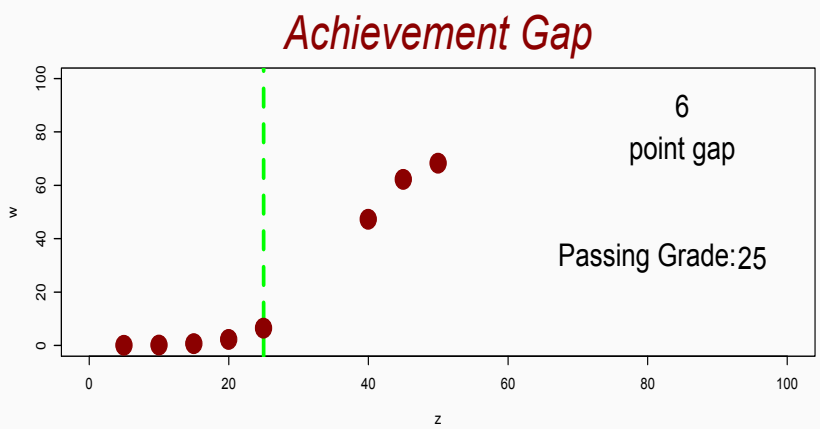
- **99 percent of Group A passes**
- **97 percent of Group B passes**

The achievement gap is:

2 points.





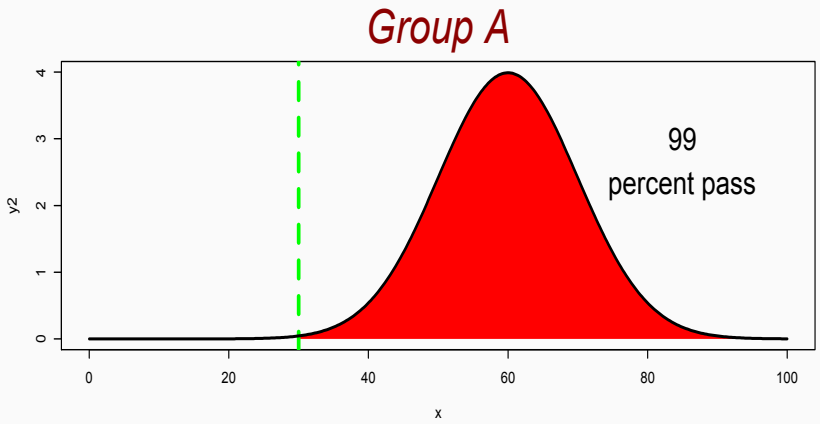


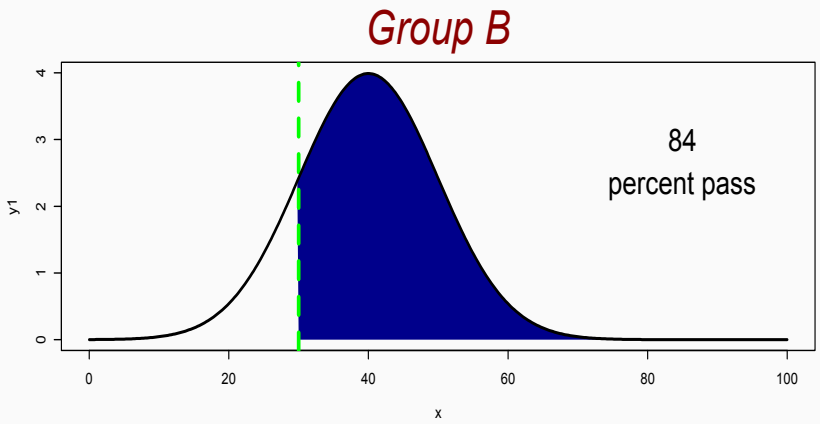
Suppose that the passing grade was set at **25 points**

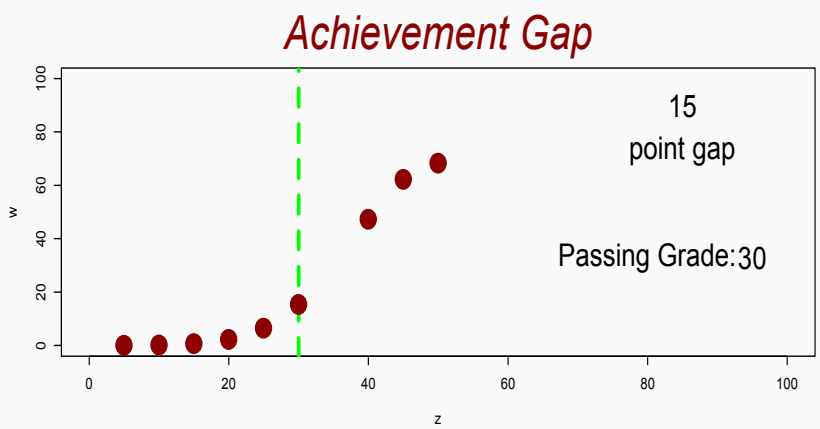
- **99 percent of Group A passes**
- **93 percent of Group B passes**

The achievement gap is:

6 points.





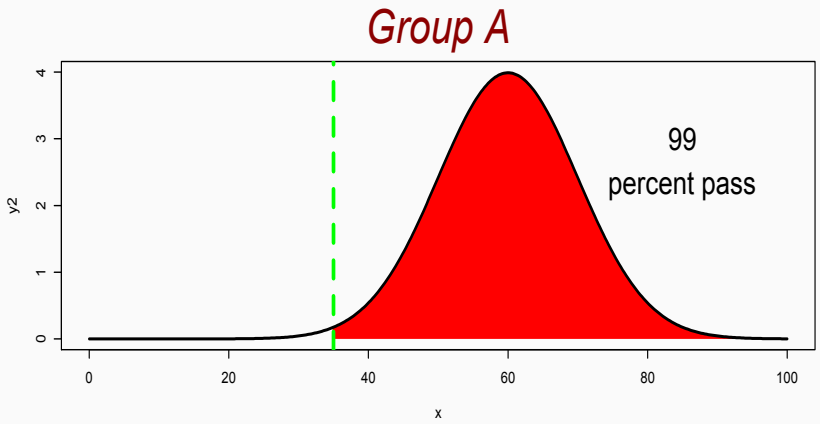


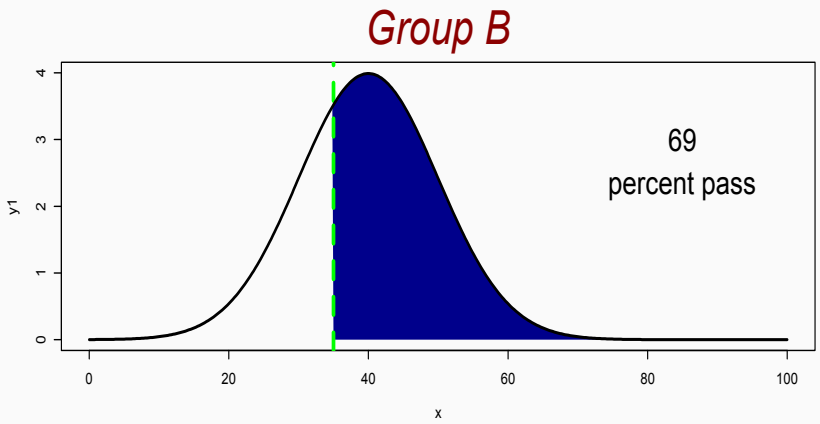
Suppose that the passing grade was set at **30 points**

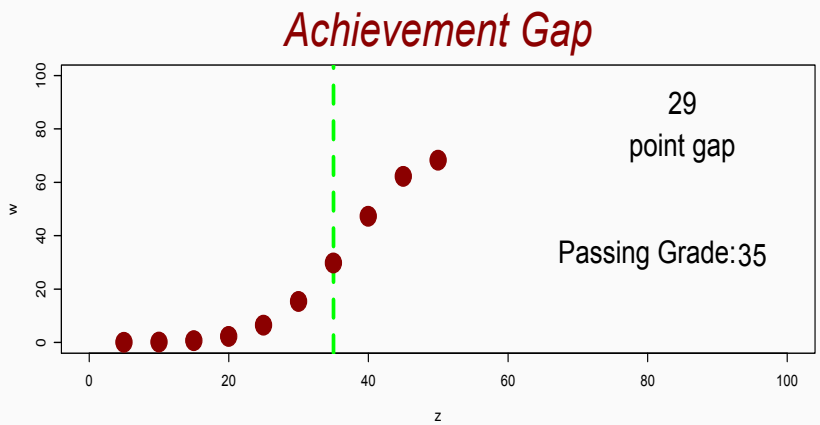
- **99 percent of Group A passes**
- **84 percent of Group B passes**

The achievement gap is:

15 points.





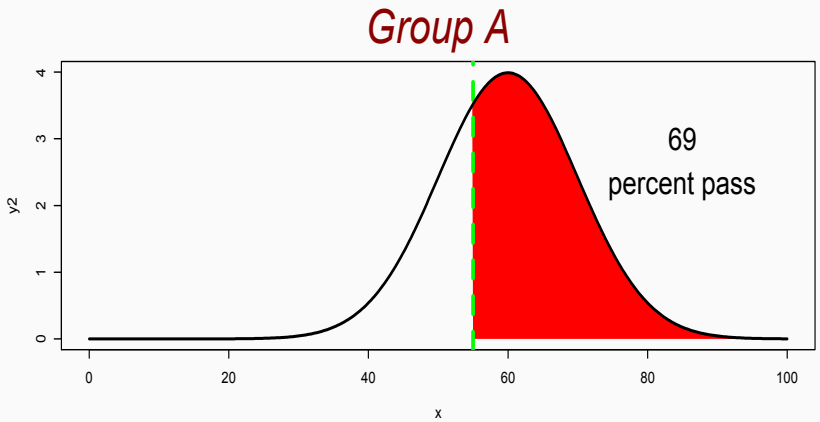


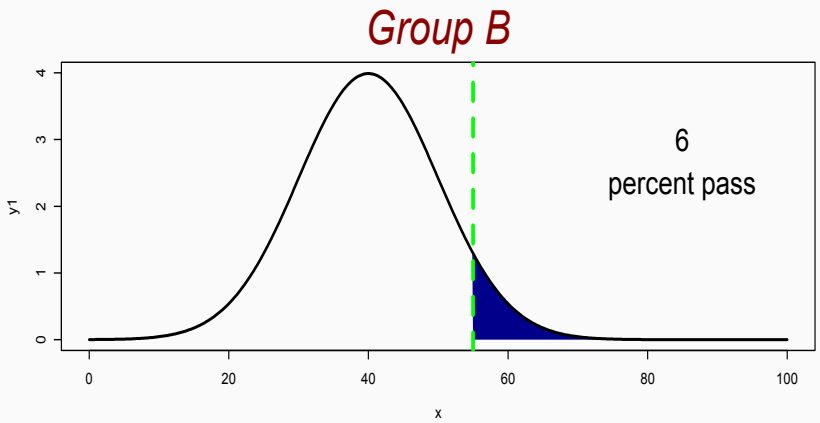
Suppose that the passing grade was set at **35 points**

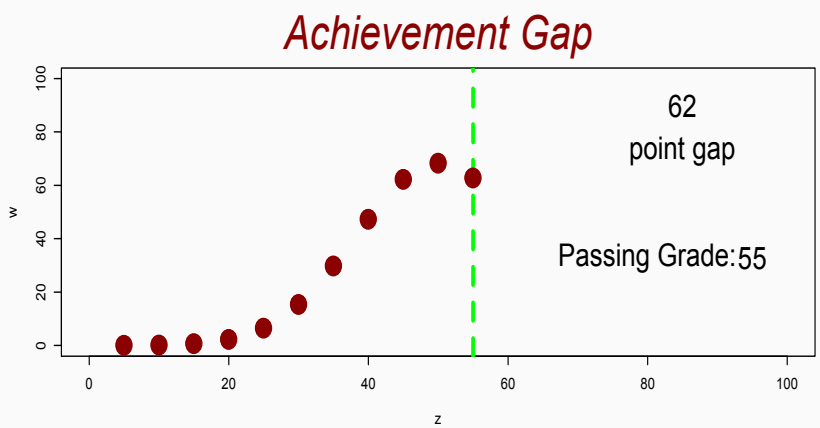
- **99 percent of Group A passes**
- **69 percent of Group B passes**

The achievement gap is:

29 points.





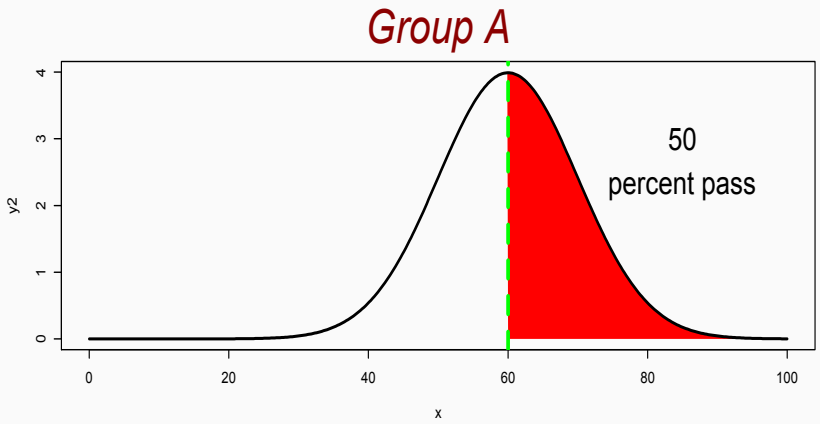


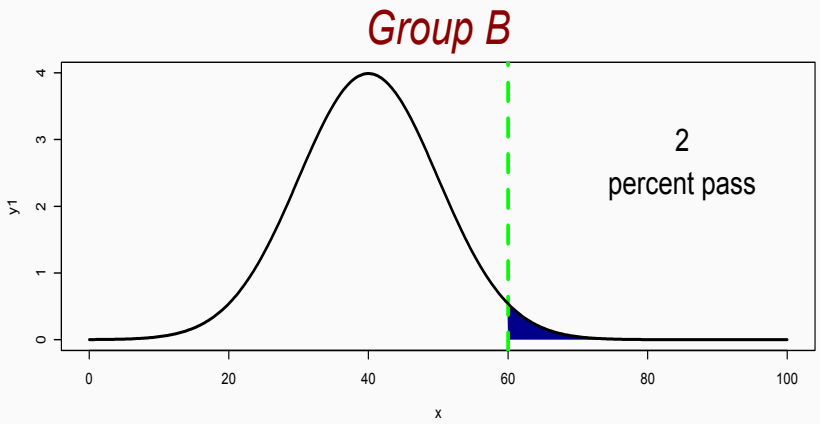
Suppose that the passing grade was set at **55 points**

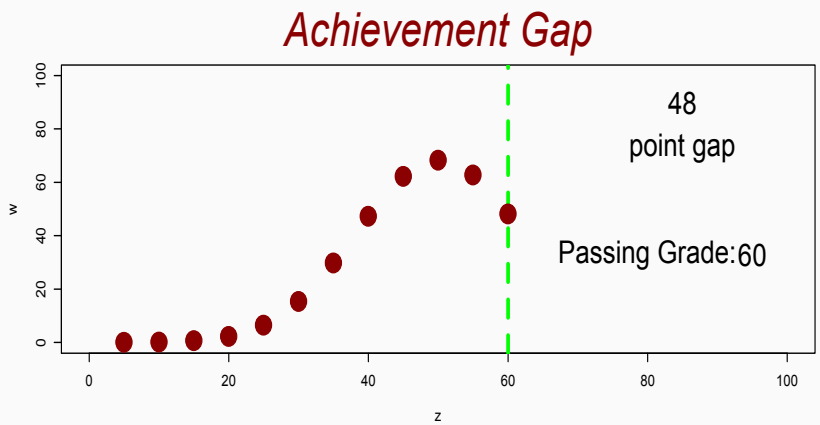
- **69 percent of Group A passes**
- **6 percent of Group B passes**

The achievement gap is:

62 points.





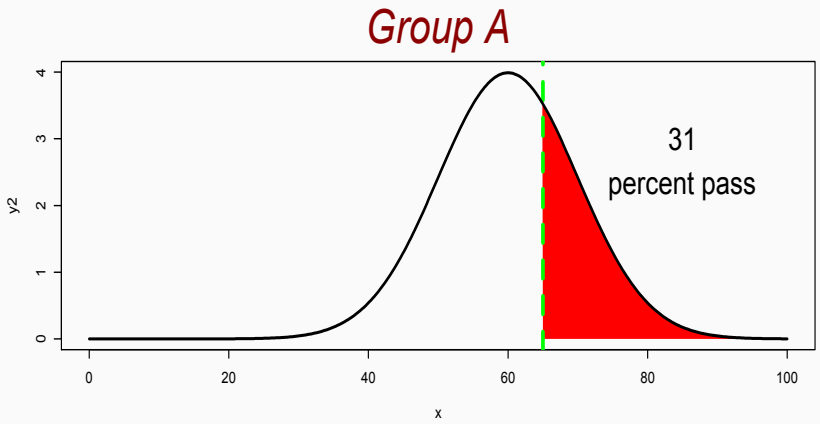


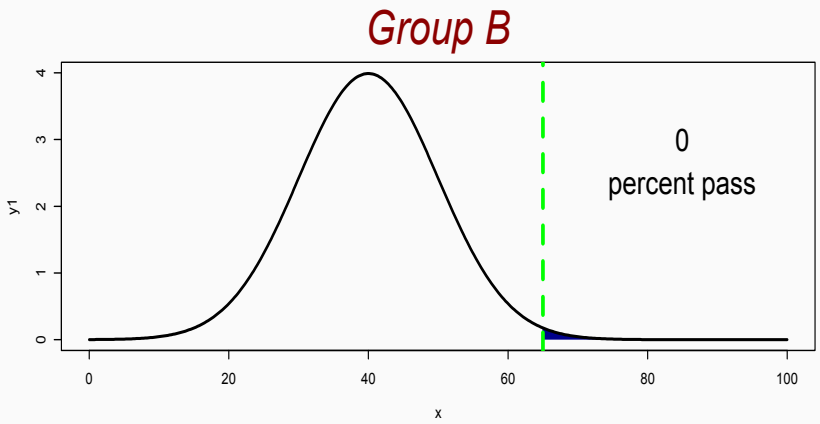
Suppose that the passing grade was set at **60 points**

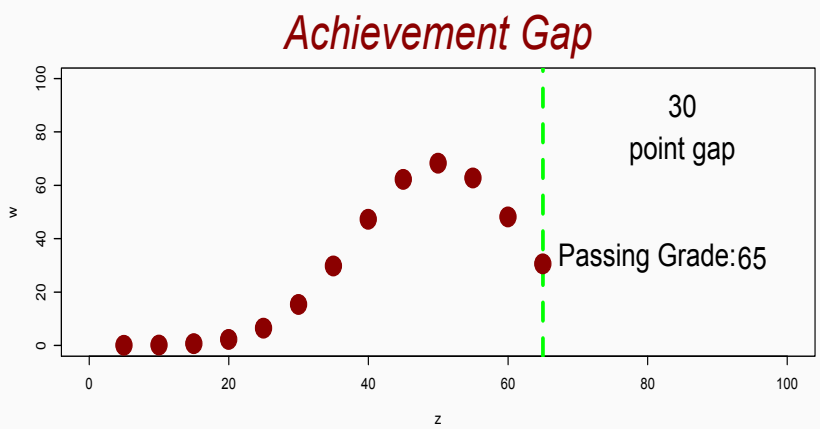
- **50 percent of Group A passes**
- **2 percent of Group B passes**

The achievement gap is:

48 points.





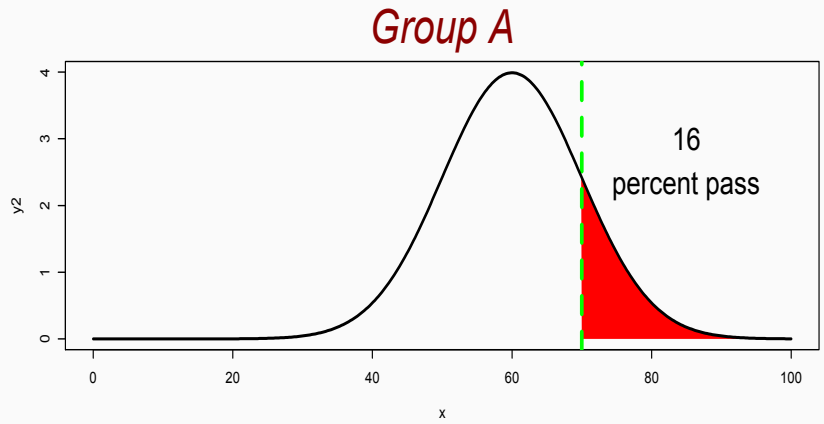


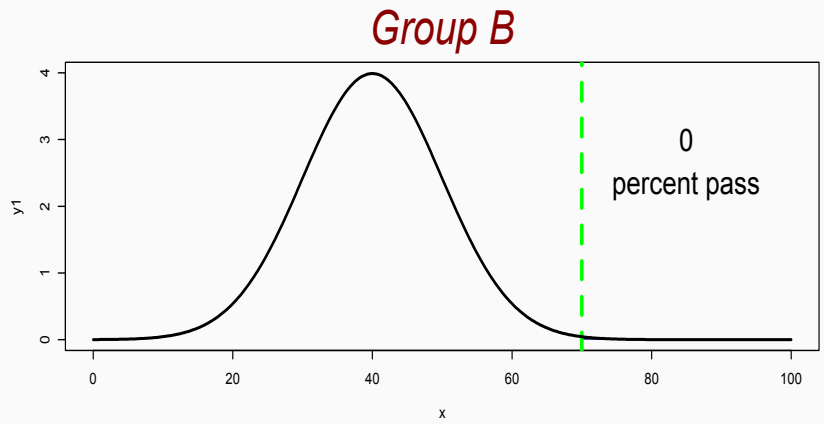
Suppose that the passing grade was set at **65 points**

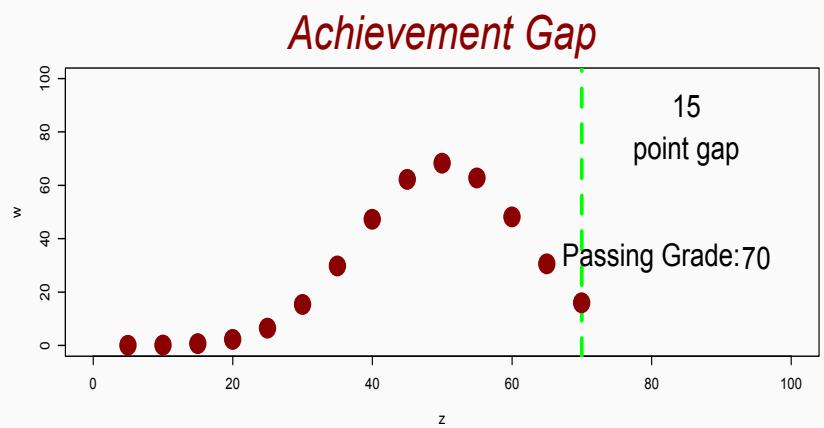
- **31 percent of Group A passes**
- **0 percent of Group B passes**

The achievement gap is:

30 points.





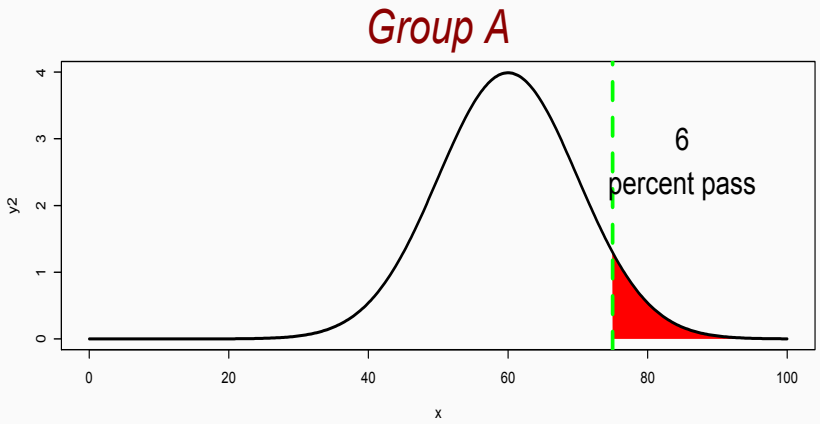


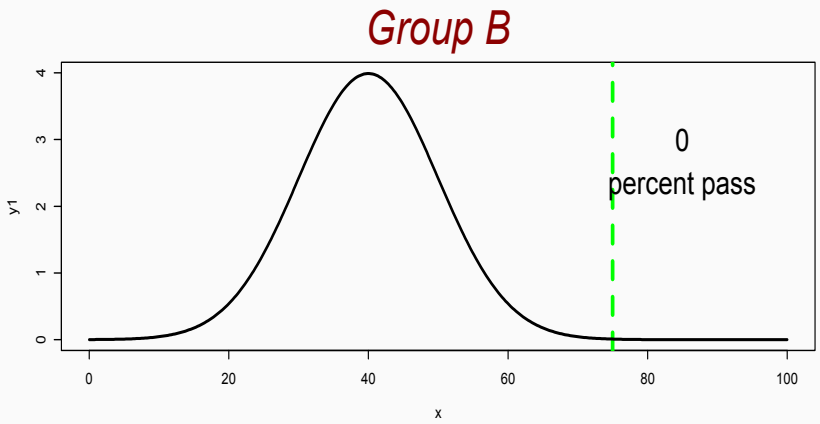
Suppose that the passing grade was set at **70 points**

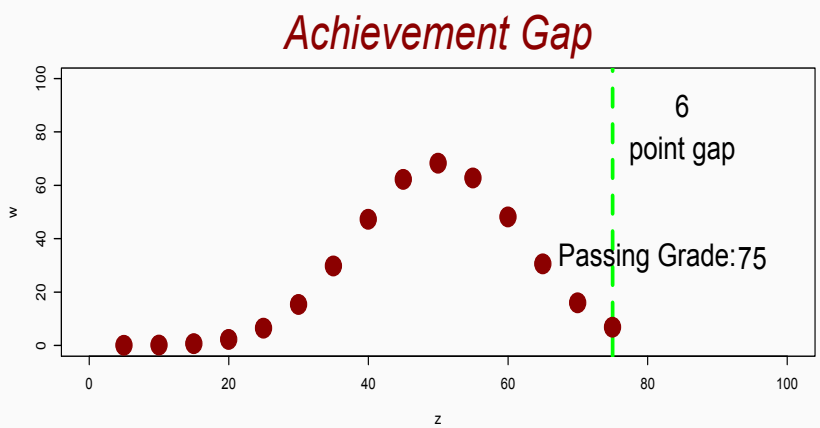
- **16 percent of Group A passes**
- **0 percent of Group B passes**

The achievement gap is:

15 points.





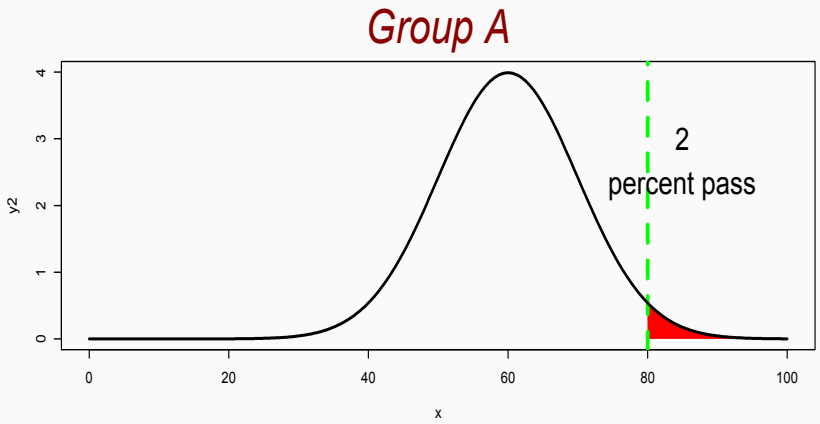


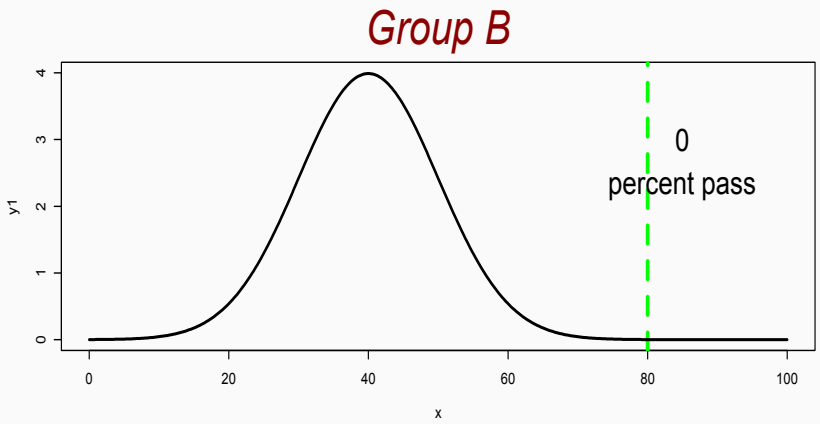
Suppose that the passing grade was set at **75 points**

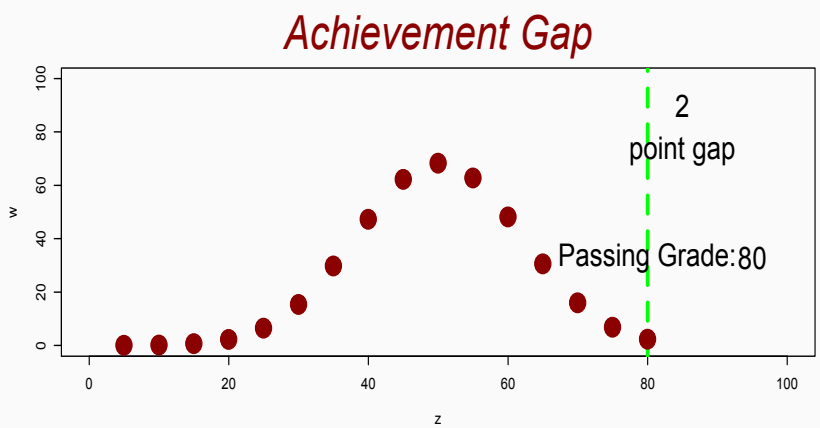
- **6 percent of Group A passes**
- **0 percent of Group B passes**

The achievement gap is:

6 points.





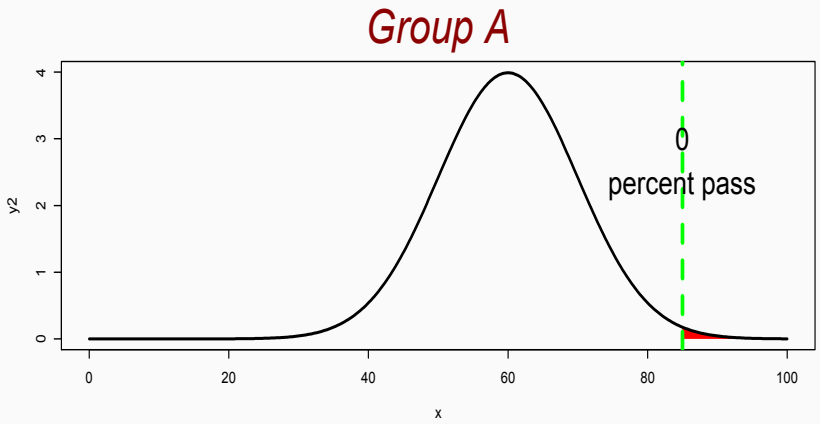


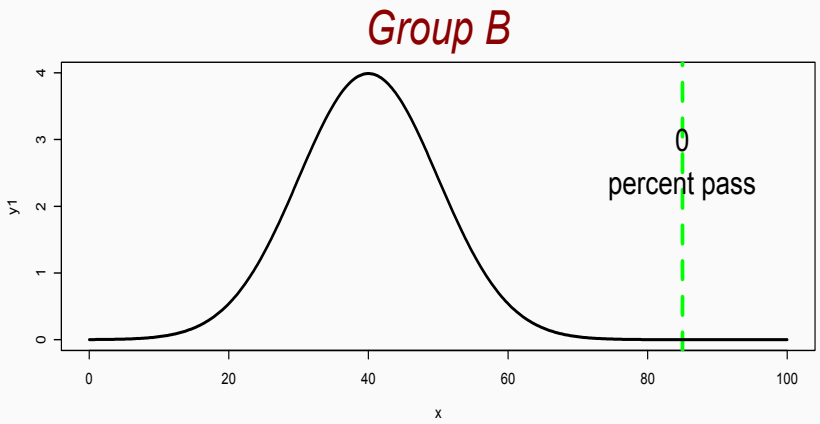
Suppose that the passing grade was set at **80 points**

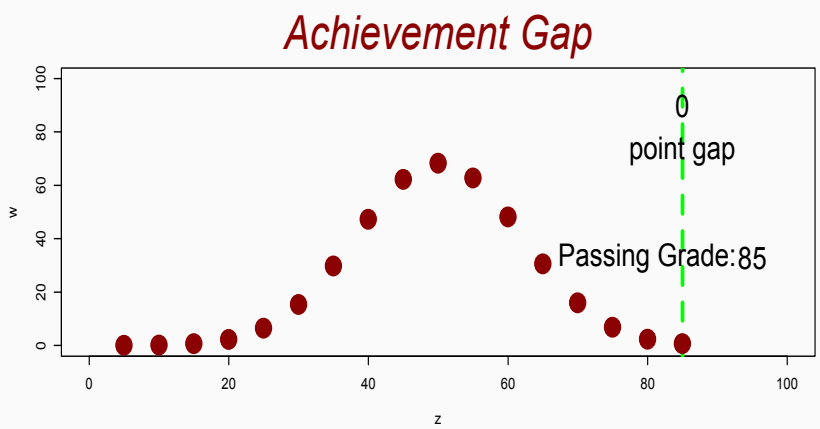
- **2 percent of Group A passes**
- **0 percent of Group B passes**

The achievement gap is:

2 points.





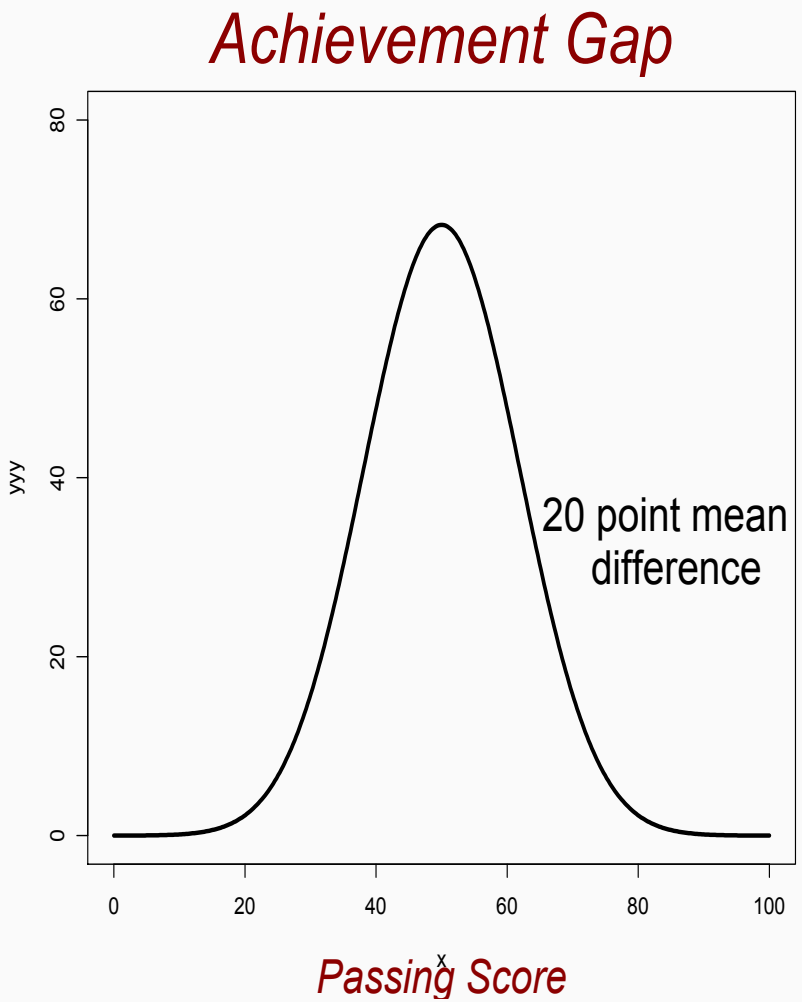


Suppose that the passing grade was set at **85 points**

- **0 percent of Group A passes**
- **0 percent of Group B passes**

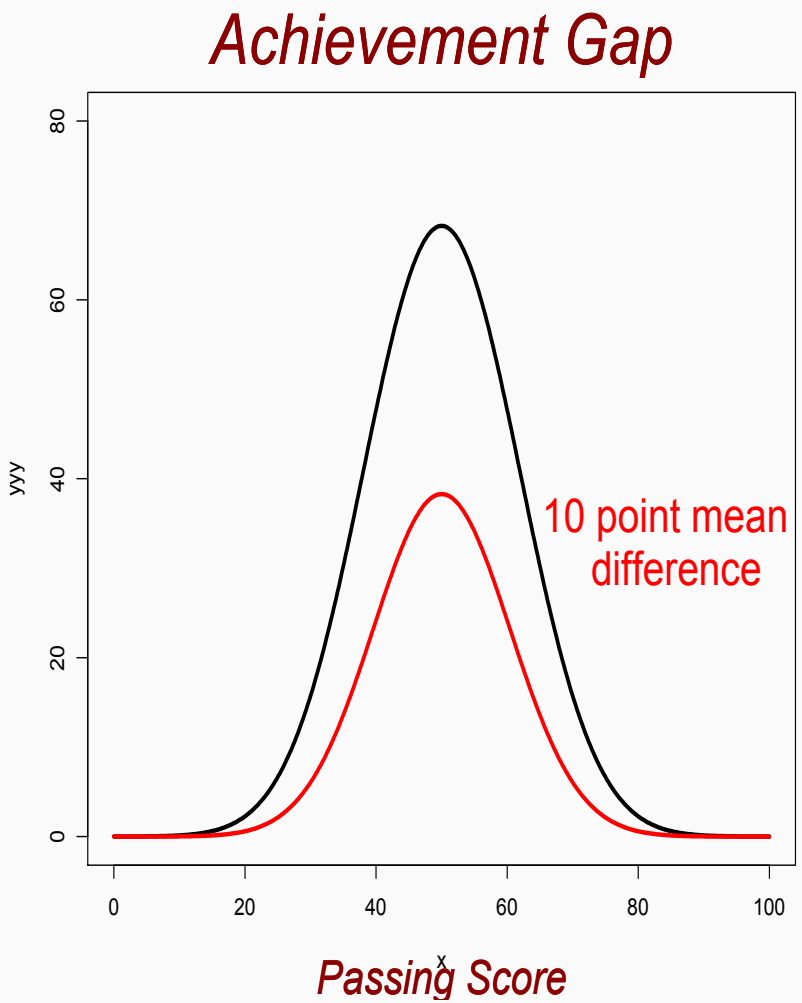
The achievement gap is:

0 points.

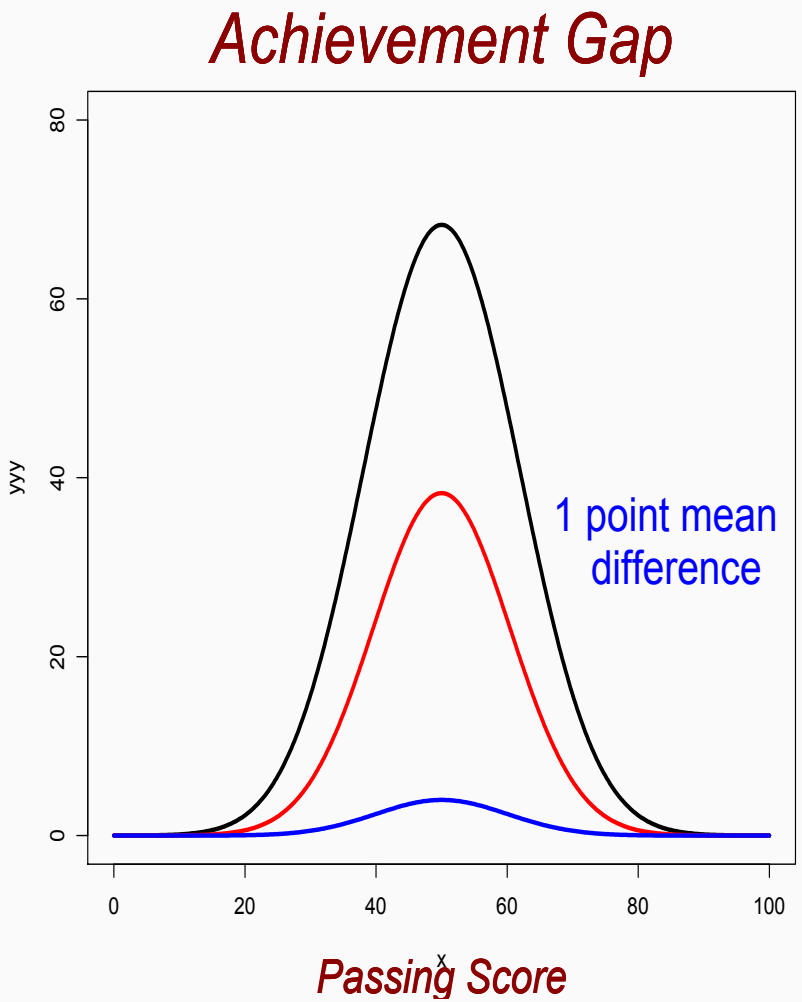


Conclusion: For Groups A and B, the achievement gap can be manipulated to be anywhere between 0 and 68 points.

This can be accomplished merely by changing what we consider to be a passing score.



If the two groups became more similar, the achievement gap still could be manipulated, but the achievement gaps would trace out a different curve.



If the two groups became more similar, the achievement gap still could be manipulated, but the achievement gaps would trace out a different curve.

Evaluating the Texas Success Story

While we have shown that the achievement gap could be manipulated, this does not mean it was. We will pursue this further. A problem is the lack of data. Essentially, the only data available to us is

- the year of the exam,
- the achievement gap,
- the percent of White students who passed, and
- the percent of Black students who passed.

Most importantly, we don't have access to either what the passing score was each year, or what the mean scores were for the two populations.

Evaluating the Texas Success Story

It turns out that we do not need to know all this data to evaluate Texas's results.

Probability Density Function

The "bell curve" describing the population with mean score μ and standard deviation σ is

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Assume that

- $f(x|0,1)$ describes population A, and
- $f(x|-D,r)$ describes population B.

Evaluating the Texas Success Story

Then we can express the percentage of students who pass the exam if the passing score is P as an integral:

$$p_A = \int_P^\infty f(x|0,1)\ dx$$

$$p_B = \int_P^\infty f(x|-D,r)\ dx$$

In terms of these integrals, the achievement gap is then:

$$G = p_A - p_B$$

Unfortunately, expression depends on P, which we don't know. However, we can eliminate it via some mathematical magic.

Evaluating the Texas Success Story

Since $p_A = \int_P^\infty f(x|0,1)\ dx$ we can solve for P:

$$P = \sqrt{2}\ Erf^{-1}(-2p_A)$$

Since $p_B = \int_P^\infty f(x|-D,r)\ dx$ we can again solve for P:

$$P = -D + r\sqrt{2}\ Erf^{-1}(-2p_B)$$

From which we can write p_B in terms of p_A by setting the two expressions for P equal to each other.

$$p_B = -\frac{1}{2}\ Erf\Big(\frac{D + \sqrt{2}Erf^{-1}(-2p_A)}{r\sqrt{2}}\Big)$$

Evaluating the Texas Success Story

From which

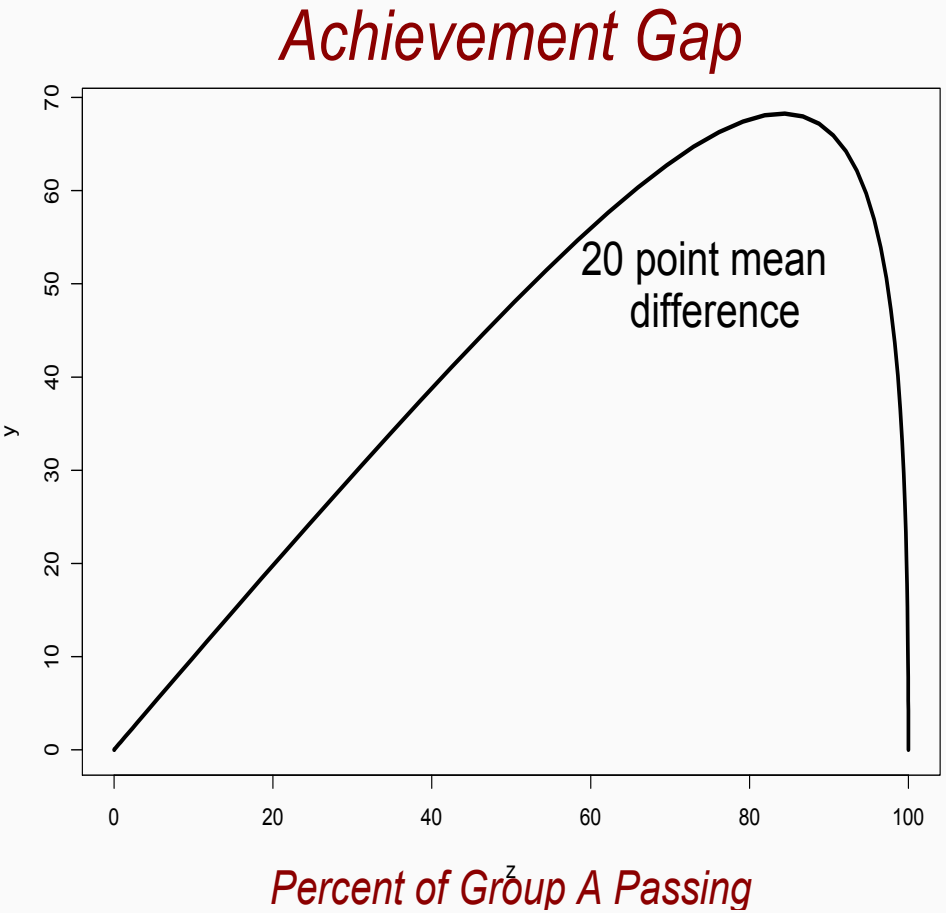
$$G = p_A - p_B = p_A + \frac{1}{2}\ Erf\Big(\frac{D + \sqrt{2}Erf^{-1}(-2p_A)}{r\sqrt{2}}\Big)$$

Moral

We do not need to know the passing score to predict the achievement gap. The achievement gap can be expressed in terms of only the percentage of students in group A who passed the exam.

Evaluating the Texas Success Story

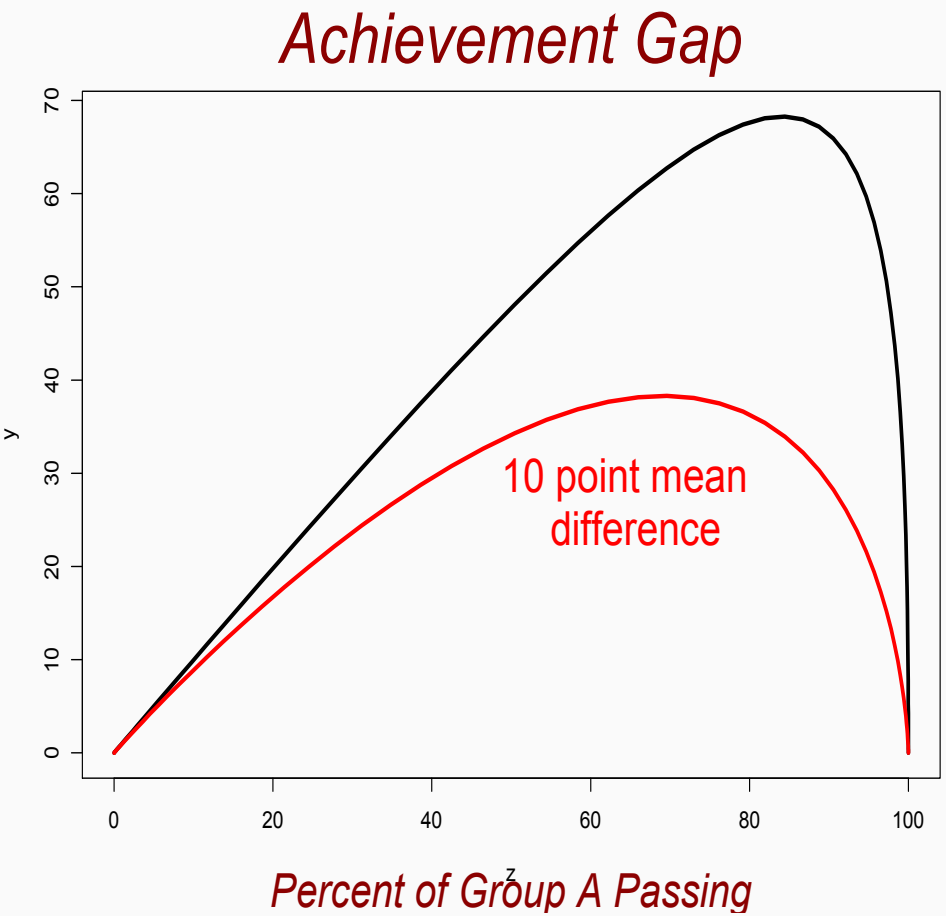
After some mathematical manipulations, one can eliminate the dependence of our results on the passing score.



The achievement gap can be predicted just from the pass rate of one of the groups. The passing score itself is not necessary.

Evaluating the Texas Success Story

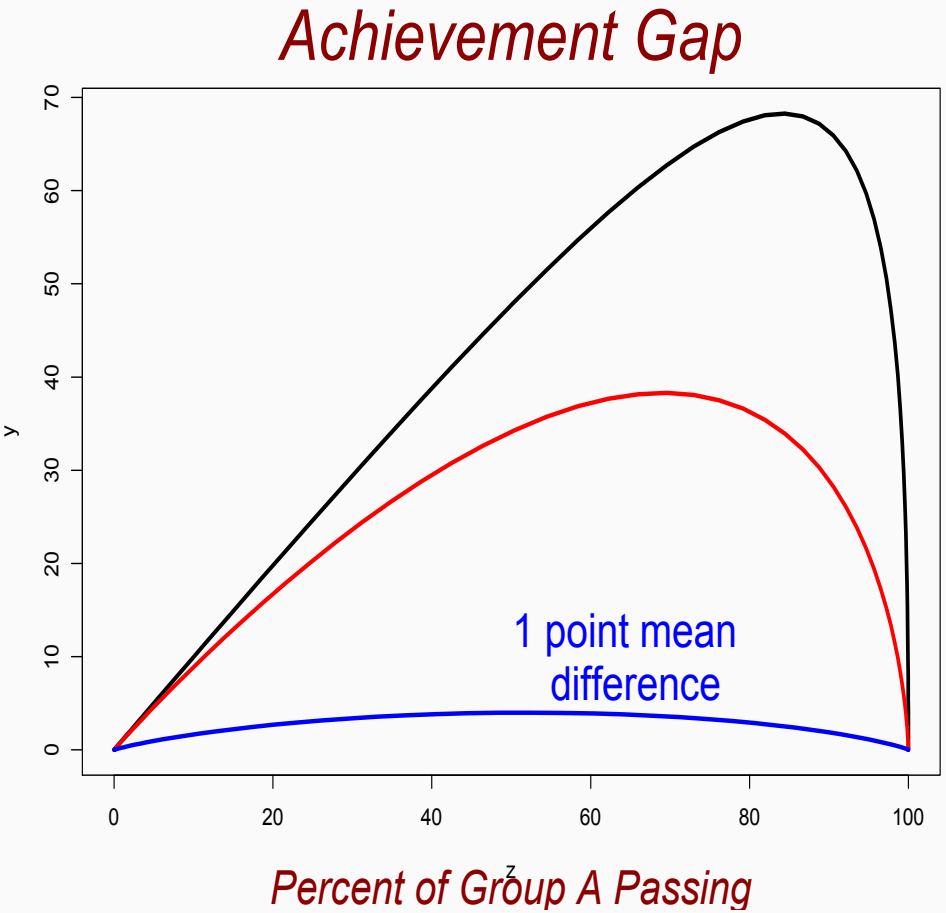
After some mathematical manipulations, one can eliminate the dependence of our results on the passing score.



The achievement gap can be predicted just from the pass rate of one of the groups. The passing score itself is not necessary.

Evaluating the Texas Success Story

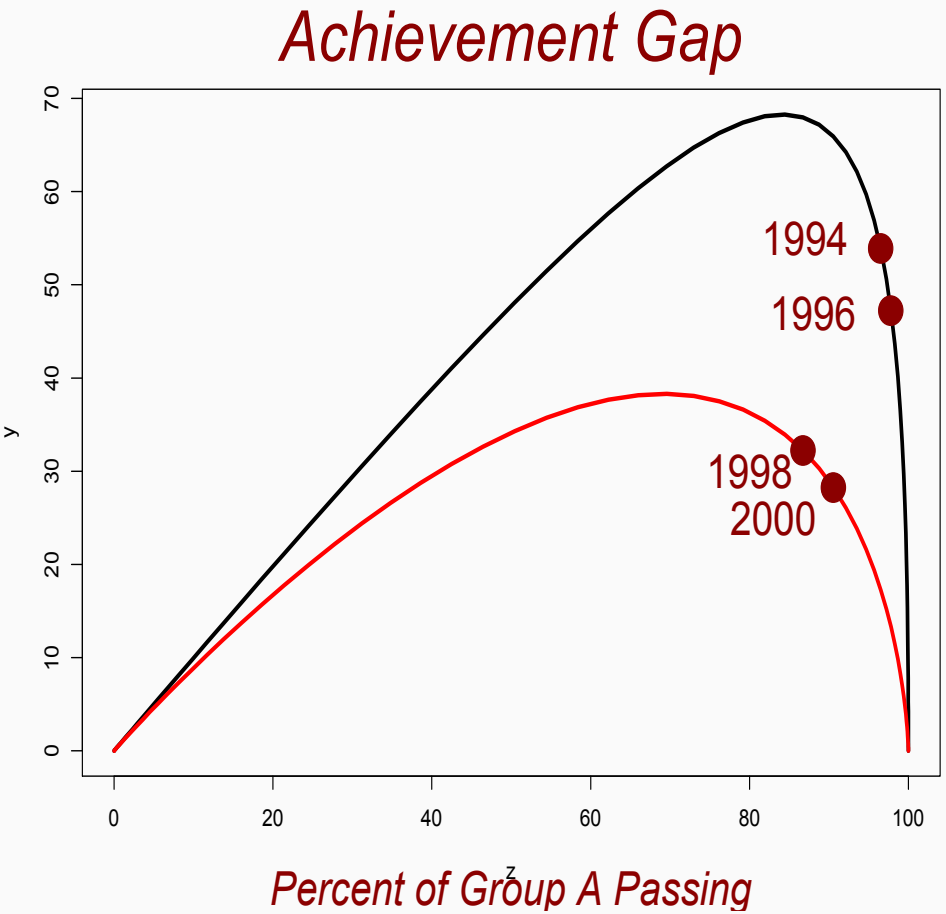
After some mathematical manipulations, one can eliminate the dependence of our results on the passing score.



The achievement gap can be predicted just from the pass rate of one of the groups. The passing score itself is not necessary.

Evaluating the Texas Success Story

How would you know if the mean difference between the populations got smaller?

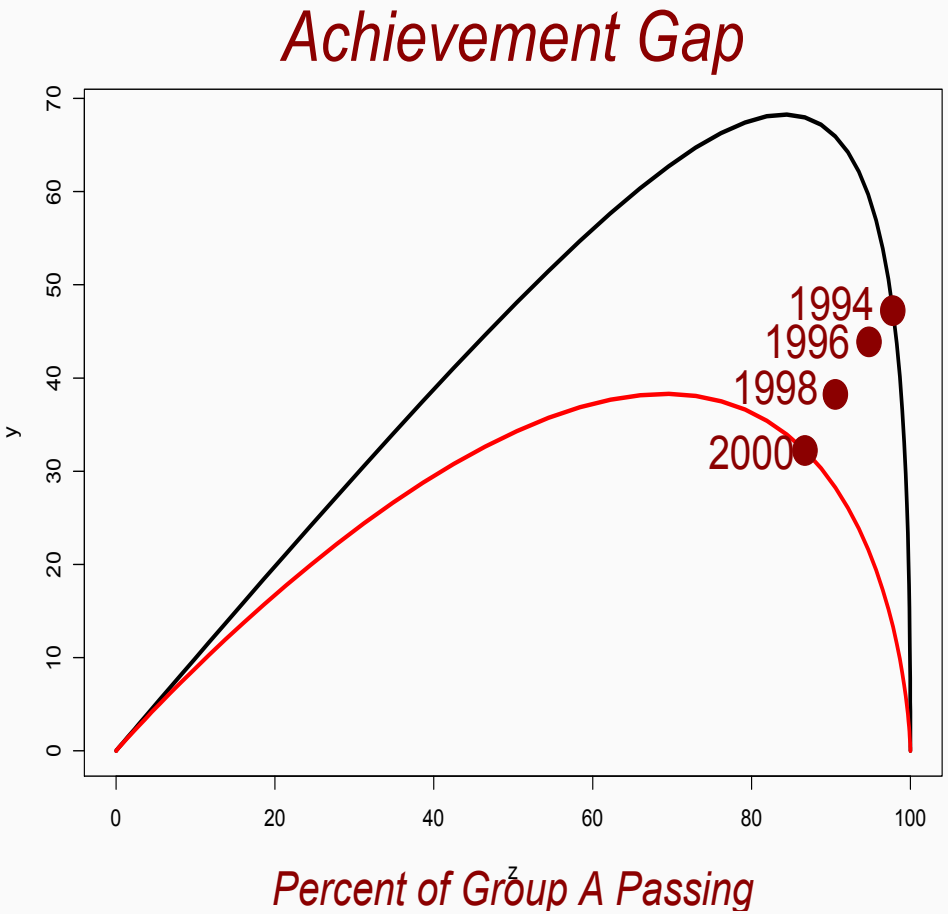


If the mean difference decreased from 20 to 10 points, this is what might be observed.

Punctuated improvement

Evaluating the Texas Success Story

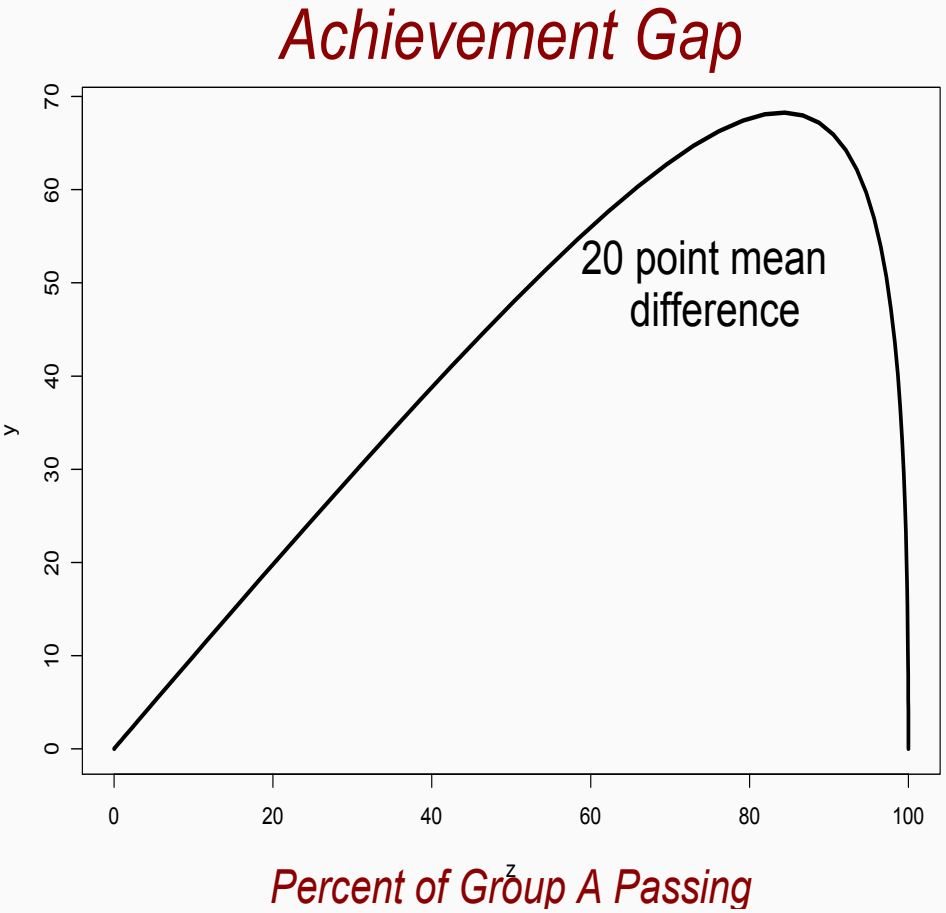
How would you know if the mean difference between the populations got smaller?



Or perhaps this.

Incremental improvement

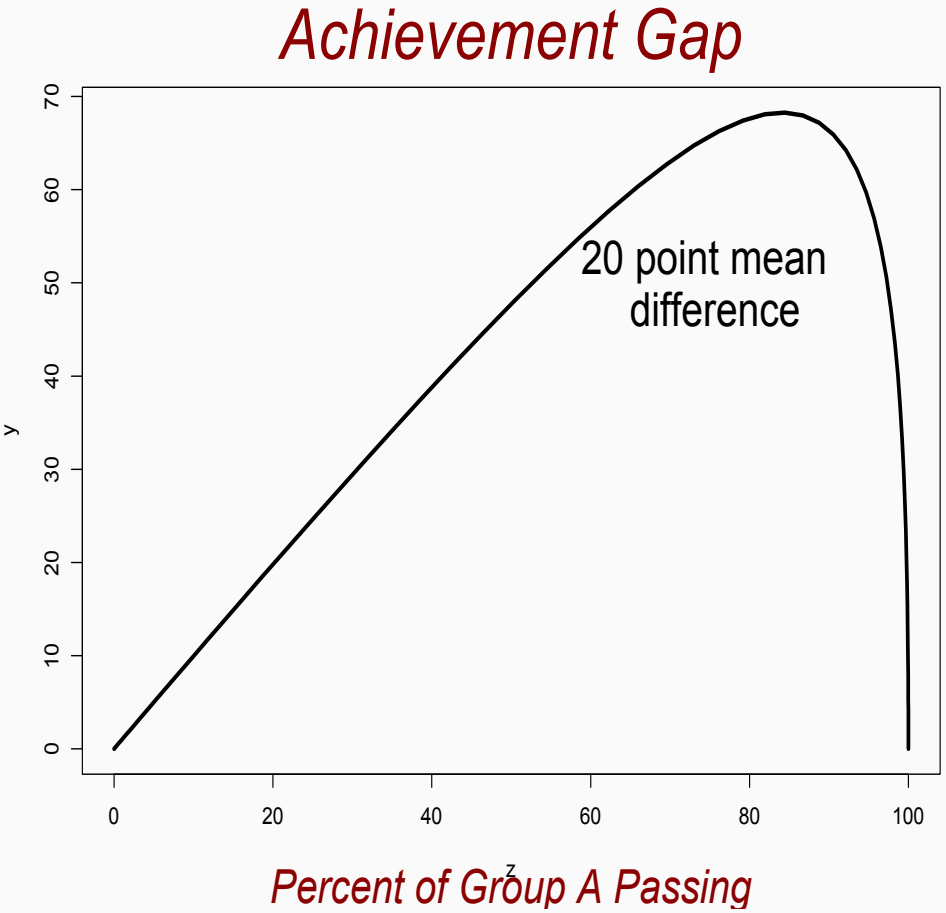
Evaluating the Texas Success Story



To evaluate the Texas Success story, we could compare Texas's data with the Achievement gap graphs. **But**, to make such a graph, we need to know

- r, the ratio of standard deviations,
- D, the difference between mean scores.

Evaluating the Texas Success Story



Ratio: We can glean r from other tests taken by Texas students for which the data is available:

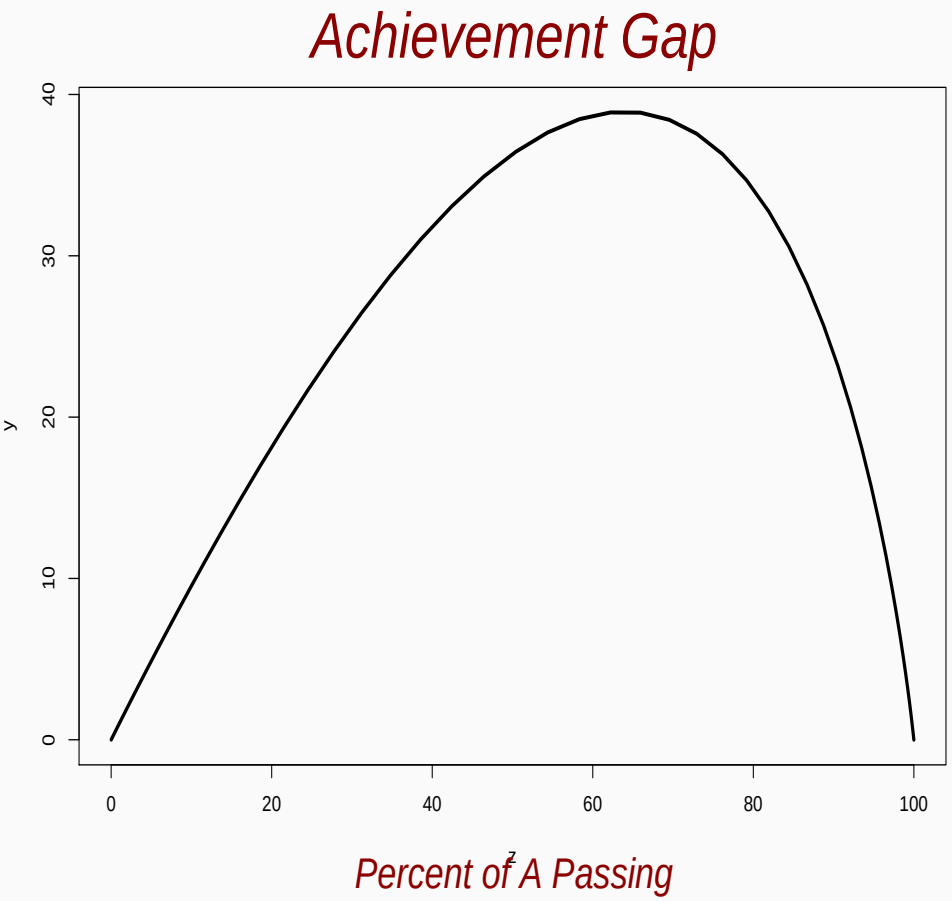
- Take $r = 0.85$.

Difference: This is harder. We will

- use the value of D which best fits the Texas data.

However, note that the curve should not fit the data perfectly, as *presumably*, this difference has been changing.

Evaluating the Texas Success Story



Ratio: We can glean r from other tests taken by Texas students for which the data is available:

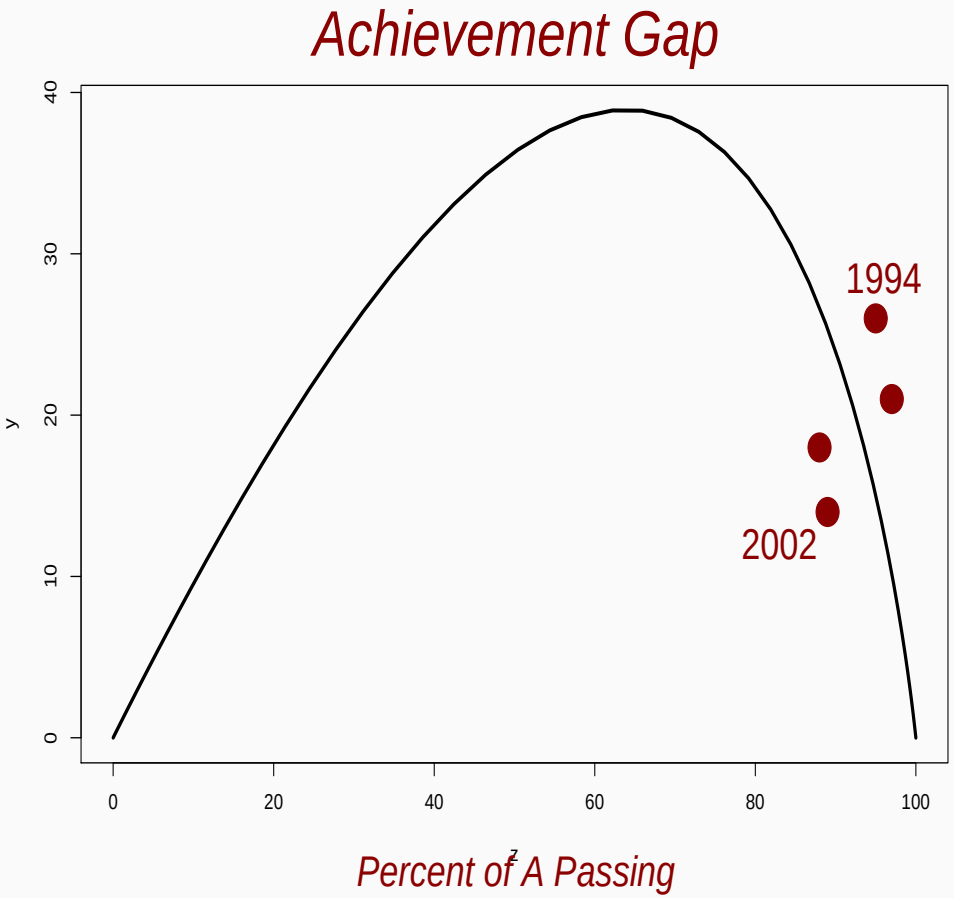
- Take $r = 0.85$.

Difference: This is harder. We will

- use the value of D which best fits the Texas data.

However, note that the curve should not fit the data perfectly, as *presumably*, this difference has been changing.

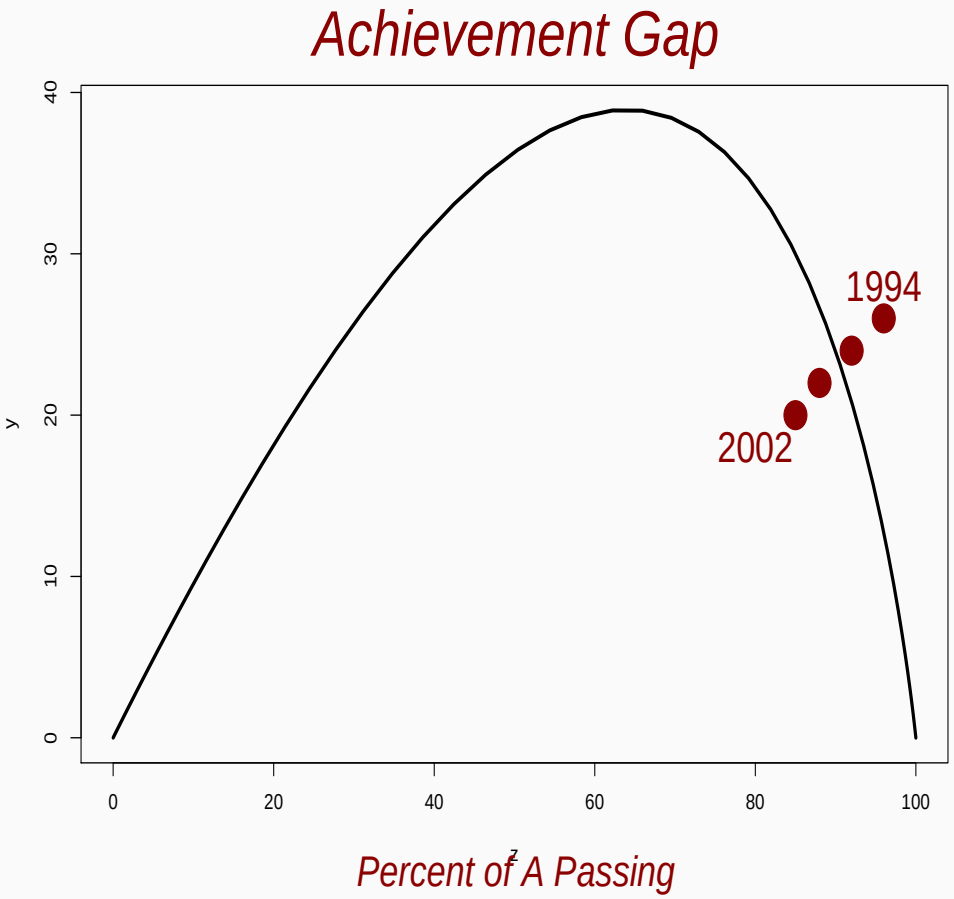
Evaluating the Texas Success Story



When we compare the Texas data, we would expect to see a graph similar to this one if there was

- punctuated improvement
-

Evaluating the Texas Success Story

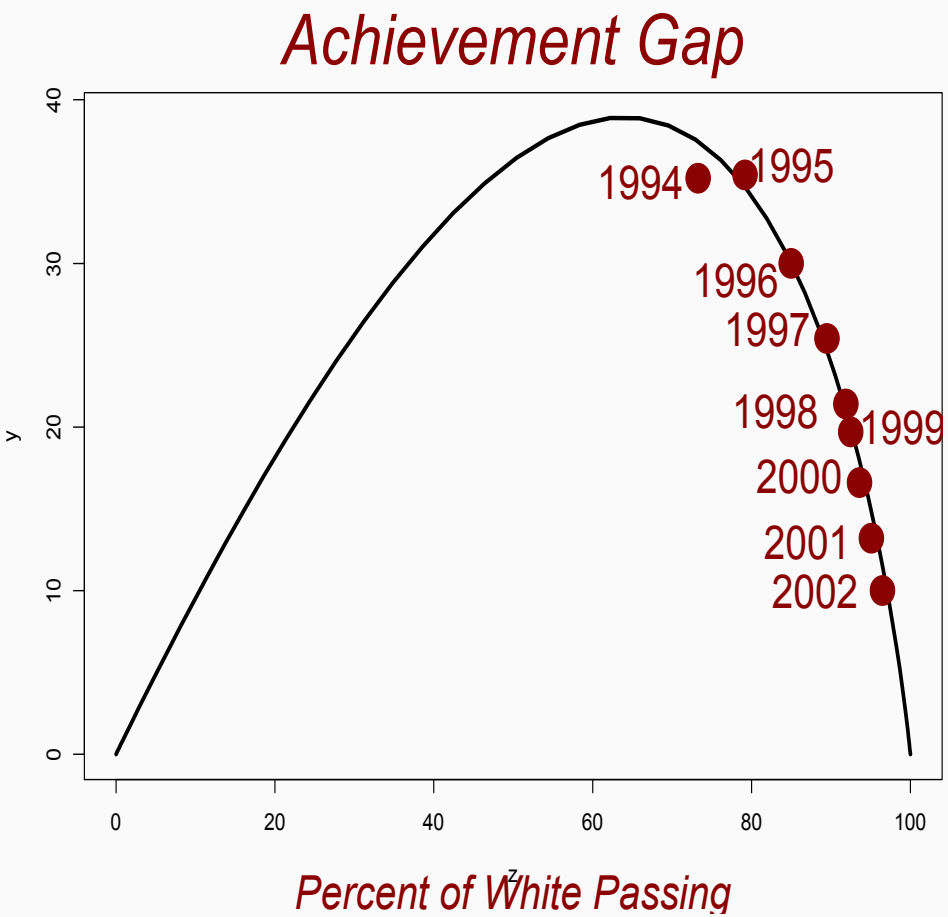


When we compare the Texas data, we would expect to see a graph similar to this one if there was

- punctuated improvement
- incremental improvement.

The Texas ~~Success~~ Story

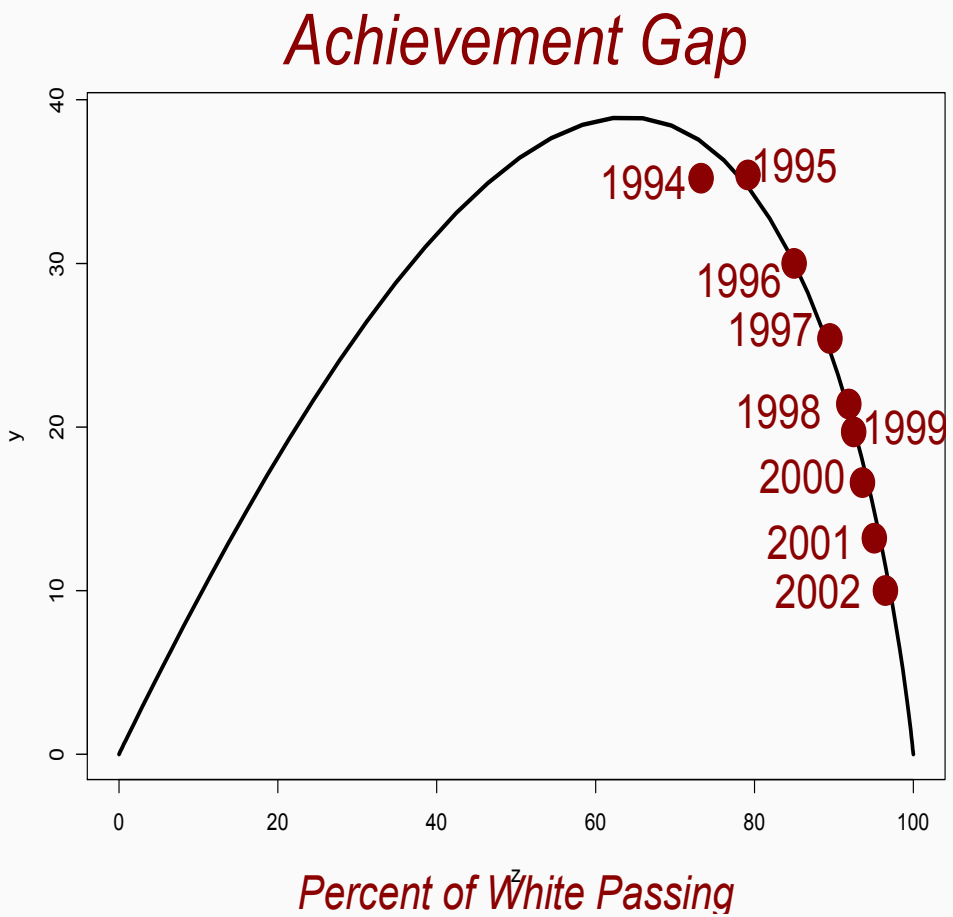
But this is what the data from Texas looks like:



Year	Point Gap	W pass	B pass
1994	35.2	73.3	38.1
1995	35.4	79.2	43.8
1996	30.0	85.0	55.0
1997	25.4	89.5	64.1
1998	21.4	91.1	70.5
1999	19.7	92.5	72.8
2000	16.6	93.6	77.0
2001	13.2	95.1	81.9
2002	10.0	96.5	86.5

The Texas ~~Success~~ Story

But this is what the data from Texas looks like:



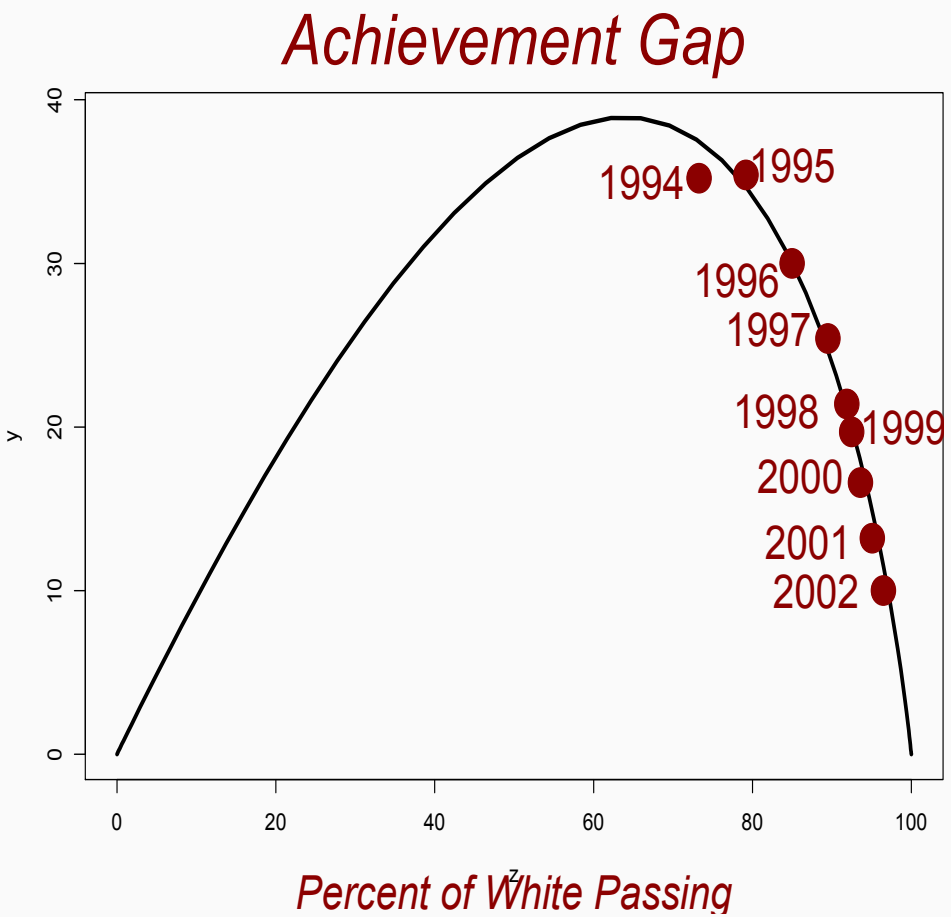
The data points lie on the curve predicted by

- a mean difference D of **0.97** standard deviation, and
- standard deviation ratio r of **0.85**,

which are consistent with other tests taken by Texas students.

The Texas ~~Success~~ Story

But this is what the data from Texas looks like:



Possible Conclusion:

No *meaningful* gap reduction occurred in Texas from 1994 to 2002.

Conclusions

If you believe that the above is

- a mathematical curio, or
- statistical nuance,

then you are right. **However,** schools are rewarded and penalized on the basis of such statistics.

Unfortunately, this method of measuring population differences is:

- uninformative, and
- potentially deceptive.

Headlines

Not just Texas.

Columbus Dispatch

"The state has set a goal of cutting the gap in passing rates in half"

Headlines

Not just Texas.

Columbus Dispatch

"The state has set a goal of cutting the gap in passing rates in half"

Detroit Free Press

"The State Board of Education and the Michigan Department of Education recently set closing the achievement gap – the difference in percentage of students passing the exam – as a key goal."

Headlines

Not just Texas.

Columbus Dispatch

"The state has set a goal of cutting the gap in passing rates in half"

Detroit Free Press

"The State Board of Education and the Michigan Department of Education recently set closing the achievement gap – the difference in percentage of students passing the exam – as a key goal."

Cringe...