Z-test for difference of proportions

H_0: $P_c - P_a = 0$
H_a: $P_c - P_a \neq 0$
$\alpha = 0.01$

Where P_c is the proportion of Caucasians that are categorized as “high risk” and P_a is the proportion of African Americans categorized as “high risk”, and P_{co} will be the combined rate.

Our Z-score is given by $$Z = \frac{P_c - P_a - H_0}{\sigma_{P_c - P_a}}$$

And $$\sigma_{P_c - P_a} = \sqrt{\frac{P_{co}(1 - P_{co})}{n_c} + \frac{P_{co}(1 - P_{co})}{n_a}}$$

Where n_c and n_a are the total number of people in our Caucasian and African American test sample, respectively. We reference our code output to get these values as well as the proportion of the populations that were categorized high risk

```
> nrow(aadata[aadata$model =="High", ])
[1] 196

> nrow(aadata)
[1] 930

> nrow(aadata[aadata$model =="High", ]) / nrow(aadata)
[1] 0.2107527

> nrow(cdata[cdata$model =="High", ])
[1] 25

> nrow(cdata)
[1] 648

> nrow(cdata[cdata$model =="High", ]) / nrow(cdata)
[1] 0.03858025
>
```

	Caucasian	African American	Combined
Low/Medium risk	623	734	1357
High risk	25	196	221
Total	648	930	1578

Thus
$P_c = 0.0386$
$P_a = 0.2108$
$P_{co} = 0.1401$
$n_c = 648$
$n_a = 930$

We calculate $$\sigma_{P_c-P_a} = \sqrt{\frac{.14(1-.14)}{648}+\frac{.14(1-.14)}{930}} = 0.0178$$

And our Z-score is $$Z = \frac{.0386-.2108-0}{.0178} = -9.67$$

With a P-value ≈ 0, we reject H_0

Therefore we can conclude that there is a statistically significant difference in the rates at which Caucasians and African Americans are being categorized as high-risk.