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 samples, respectively. To obtain our proportions and n values, we reference the outputs from our code:

```
> race_count
             race    n
1 African-American 5807
2           Asian   58
3       Caucasian 4077
```

```
> # Test data against model
> c_data <- data.frame("race"= "Caucasian", "age" = 25, "sex"="Male", "juv_fel_count" = 0,
 "juv_misd_count" = 0, "juv_other_count" = 0, "priors_count" = 0, "days_b_screening_arrest"
 = 90, "c_days_from_compas" = 30, "c_charge_degree" = "F", "c_time_in_jail" = 60)
> round(predict(model_fit,c_data,type = "p"), 3)
   Low Medium   High
 0.532  0.335  0.133
>
> aa_data <- data.frame("race"= "African-American", "age" = 25, "sex"="Male", "juv_fel_coun
t" = 0, "juv_misd_count" = 0, "juv_other_count" = 0, "priors_count" = 0, "days_b_screening_
arrest" = 90, "c_days_from_compas" = 30, "c_charge_degree" = "F", "c_time_in_jail" = 60)
> round(predict(model_fit,aa_data,type = "p"), 3)
   Low Medium   High
 0.415  0.388  0.197
```

Using the total n-values for each group and these proportions, we can derive our n-values:

	Caucasian	African American	Combined
Low/Medium risk	3535	4663	8198
High risk	542	1144	1686
Total	4077	5807	9884

Thus

$P_c = 0.133$

$P_a = 0.197$

$P_{co} = 1686/9884 = 0.1706$

$n_c = 4077$

$n_a = 5807$

We calculate $\sigma_{P_c - P_a} = \sqrt{\frac{.1706(.8294)}{4077} + \frac{.1706(.8294)}{5807}} = 0.0077$

And our Z-score is $$Z = \frac{.133 - .197 - 0}{.0077} = -8.312$$

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.