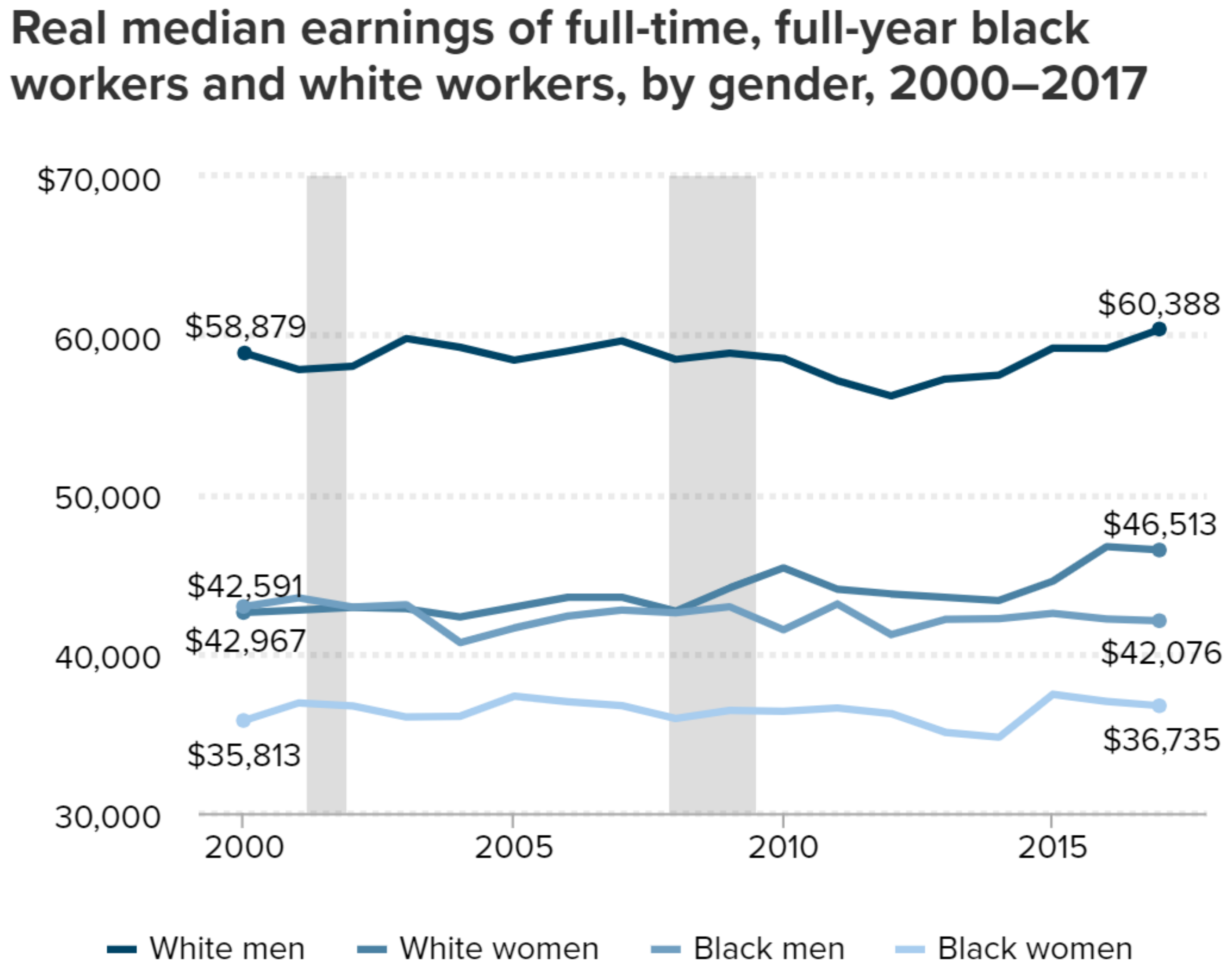
Hypothesis Test for Two Proportions

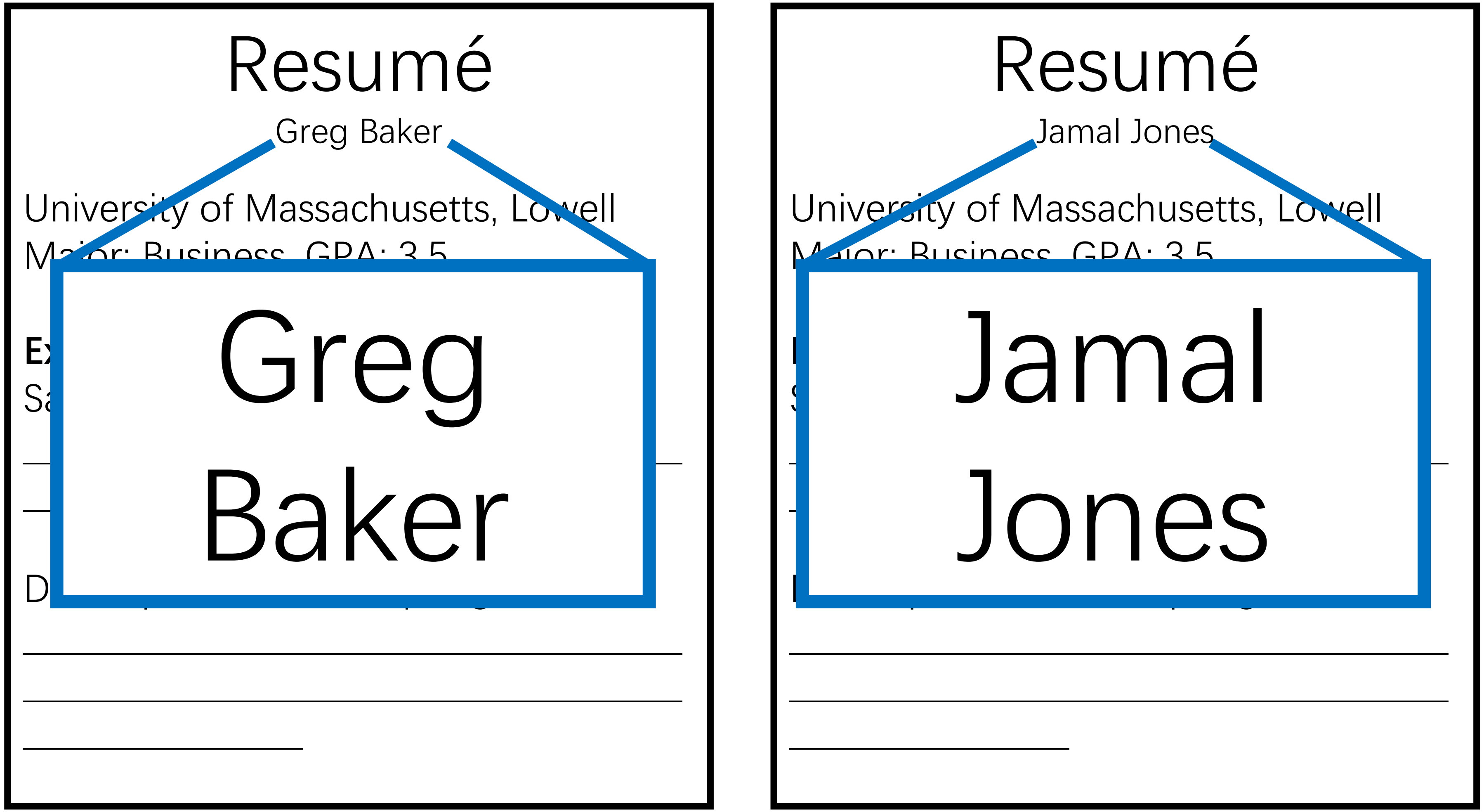


Economic Policy Institute, 2018: https://www.epi.org/blog/black-workers-have-made-no-progress-in-closing-earnings-gaps-with-white-men-since-2000/

Hiring discrimination

Researchers wanted to test if hiring discrimination was a factor in labor markets

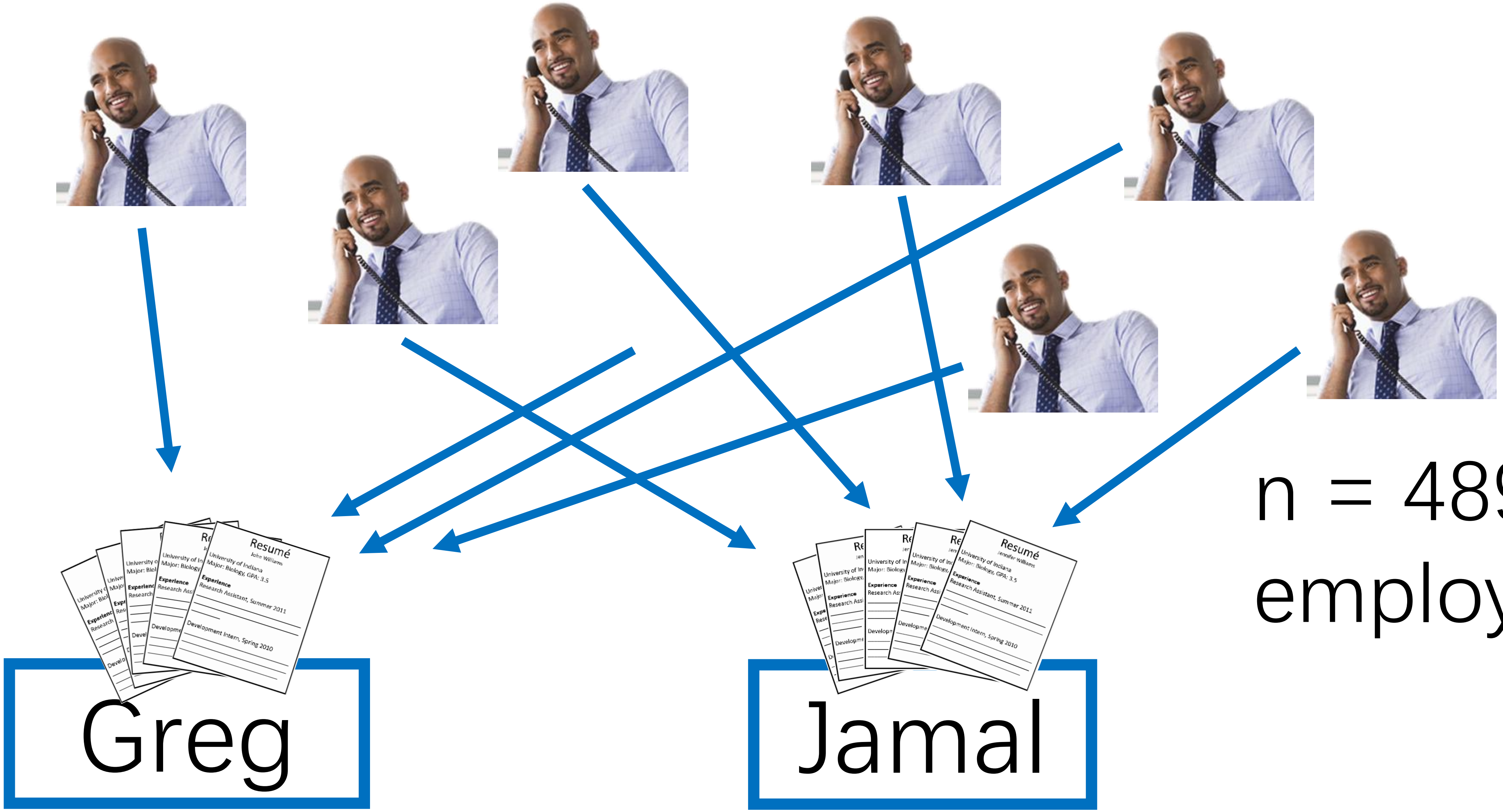
The Race/Resumé Study



The jobs

- Wide swath of jobs in the following **industries**: sales, administrative support, clerical services, and customer services
- Large range of **positions**, from “cashier work at retail establishments and clerical work in a mailroom to office and sales management positions.”

The Race/Resumé Study



n = 4890 employers

The Race/Resumé Study

Measured which group got more callbacks from potential employers

Greg $n_1 = 2445$

Jamal $n_2 = 2445$

The results

	Treatment 1 Commonly-White Names	Treatment 2 Commonly-Black Names	Total
Called back	246	164	410
Not called back	2199	2281	4480
Total	2445	2445	4890

$n_1 = 2445$ $\quad$ $n_2 = 2445$

$$\hat{p}_1 = \frac{246}{2445} = 0.101 \qquad \hat{p}_2 = \frac{164}{2445} = 0.067$$

Comparing the proportion who received callbacks from both treatments.

Two-Sample Situation

If there's hiring discrimination, $\hat{\boldsymbol{p}}_{\mathbf{1}} > \hat{\boldsymbol{p}}_{\mathbf{2}}$

Group 1: White

$\hat{p}_1$ = proportion of commonly-white name apps that got callback.

$$\hat{p}_1 = \frac{246}{2445} = \mathbf{0.101}$$

$>$

Group 2: Black

$\hat{p}_2$ = proportion of commonly-black name apps that got callback.

$$\hat{p}_2 = \frac{164}{2445} = \mathbf{0.067}$$

Are these proportions **different enough** to show discrimination, or could this difference have been a result of **chance alone**?

Hypotheses

$$H_0\colon p_1 = p_2$$

There is no discrimination, so the callback rate is the **same in both groups**. You're seeing if there's evidence to reject this default claim.

$$H_A\colon p_1 > p_2$$

There is discrimination, in which case the commonly-white named applications received a **higher rate** of callbacks.

Where:

p_1 is the proportion of **all** applicants with commonly-**white** names who'd receive callbacks when applying to jobs like the ones in this study.

p_2 is the proportion of **all** applicants with commonly-**black** names who'd receive callbacks when applying to jobs like the ones in this study.

Setting up the Hypotheses

$$H_0: p_1 = p_2$$
$$H_A: p_1 > p_2$$

OR

$$H_0: p_1 - p_2 = 0$$
$$H_A: p_1 - p_2 > 0$$

❤Preferred❤

Where:

p_1 is the proportion of **all** applicants with commonly-**white** names who'd receive callbacks when applying to jobs like the ones in this study.

p_2 is the proportion of **all** applicants with commonly-**black** names who'd receive callbacks when applying to jobs like the ones in this study.

Calculations

Since null assumes $p_1 = p_2$, so we can **combine** the proportion who got callbacks into one estimate: $\hat{p}_c$

Under certain conditions:

$$\hat{p}_1 - \hat{p}_2 \sim \mathrm{N}(\mu = 0, \sigma = \sqrt{\frac{\hat{p}_c\,(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c\,(1-\hat{p}_c)}{n_2}})$$

Centered at zero (since null assumes **no difference** between callback rates)

$$\hat{p}_1 = \frac{246}{2445} = 0.101$$

$$\hat{p}_2 = \frac{164}{2445} = 0.067$$

Combined proportion $\hat{p}_c = \frac{246+164}{2445+2445} = 0.084$

Calculations

Since null assumes $p_1 = p_2$, so we can **combine** the proportion who got callbacks into one estimate: $\hat{p}_c$

Under certain conditions:

$\hat{p}_1 - \hat{p}_2 \sim N(\mu = 0, \sigma = 0.0079)$

Centered at zero (since null assumes **no difference** between callback rates)

The Data:
The actual difference in callback rates from the experiment $\hat{p}_1 - \hat{p}_2 = \mathbf{0.034}$

How unlikely was our data?

Check the p-value!

$\hat{p}_1 = \frac{246}{2445} = 0.101$

$\hat{p}_2 = \frac{164}{2445} = 0.067$

Combined proportion $\hat{p}_c = \frac{246+164}{2445+2445} = 0.084$

Conclusion

Under my assumption that there is no difference in callback rates, the actually observed data (a 3.4% difference in callback rates among 4890 employers) is highly unlikely (**p-value = 0.00001 < alpha level of 0.05**). So, **I reject my earlier assumption**. There's convincing evidence that commonly-white named resumés receive a **higher callback rate**.

State-Plan-Do-Conclude

State: State the hypotheses, significance level, and define your parameters

$$H_0: p_1 - p_2 = 0$$
$$H_A: p_1 - p_2 > 0$$

$$\alpha = 0.05$$

Where:

p_1 is the proportion of **all** applicants with commonly-white names who'd receive callbacks when applying to jobs like the ones in this study.

p_2 is the proportion of **all** applicants with commonly-black names who'd receive callbacks when applying to jobs like the ones in this study.

State-**Plan**-Do-Conclude

Plan: Name your inference method and check conditions

We will conduct a **two-sample z-test** for $p_1 - p_2$, if all conditions are met.

Conditions

Recall: Why we check conditions

$$\hat{p} \sim \text{Normal}\left(\mu = 0, \sigma = \sqrt{\frac{\hat{p}_c\,(1-\hat{p}_c)}{n_1} + \frac{\hat{p}_c\,(1-\hat{p}_c)}{n_2}}\right)$$

1) Random condition
→ unbiased **center**

2) 10% condition
→ calculable **spread**

3) Large counts
→ approx. normal **shape**

State-Plan-Do-Conclude

Plan: Name your inference method and check conditions

We will conduct a **two-sample z-test** for $p_1 - p_2$, if all conditions are met.

Conditions

1. **Random:**
Employers were randomly **assigned** either a commonly-white or commonly-black named resumé ✓

Only have to do **10%** when sampling. However, this is an experiment. We don't have to check this condition!

3. **Large Counts:**

$n_1\hat{p}_c \geq 10$ $\quad$ $n_2\hat{p}_c \geq 10$

$n_1(1-\hat{p}_c) \geq 10$ $\quad$ $n_2(1-\hat{p}_c) \geq 10$

State-**Plan**-Do-Conclude

Plan: Name your inference method and check conditions

We will conduct a **two-sample z-test** for $p_1 - p_2$, if all conditions are met.

Conditions

1. **Random:** Employers were randomly **assigned** either a commonly-white or commonly-black named resumé

2. **Large Counts:**

$$n_1\hat{p}_c \geq 10$$

$$(2445)(.084) \geq 10$$

$$n_1(1-\hat{p}_c) \geq 10$$

$$(2445)(1-.084) \geq 10$$

$$n_2\hat{p}_c \geq 10$$

$$(2445)(.084) \geq 10$$

$$n_2(1-\hat{p}_c) \geq 10$$

$$(2445)(1-.084) \geq 10$$

State-**Plan**-Do-Conclude

Plan: Name your inference method and check conditions

We will conduct a **two-sample z-test** for $p_1 - p_2$, if all conditions are met.

Conditions

1. **Random:** Employers were randomly **assigned** either a commonly-white or commonly-black named resumé

2. **Large Counts:**

$$n_1\hat{p}_c \geq 10$$
$$205.4 \geq 10$$

$$n_1(1-\hat{p}_c) \geq 10$$
$$2239.6 \geq 10$$

$$n_2\hat{p}_c \geq 10$$
$$205.4 \geq 10$$

$$n_2(1-\hat{p}_c) \geq 10$$
$$2239.6 \geq 10$$

State-Plan-**Do**-Conclude

Do: Perform calculations (if conditions met), report the test statistic and the p-value

z = 4.231

p-value = 0.00001

State-Plan-Do-**Conclude**

Conclude: Reject or fail to reject H_0 and justify

$$H_0: p_1 - p_2 = 0$$
$$H_A: p_1 - p_2 > 0$$

$\alpha = 0.05$

z = 4.231

p-value = 0.00001

Conclusions template: Because our p-value (____) is **less/greater** than our alpha level (___), we **reject/fail to reject** H_0. We **do/don't** have convincing evidence that (H_A in context).

State-Plan-Do-**Conclude**

Conclude: Reject or fail to reject H_0 and justify

$$H_0: p_1 - p_2 = 0$$
$$H_A: p_1 - p_2 > 0$$

$\alpha = 0.05$

z = 4.231
p-value = 0.00001

Because our p-value (0.00001) is **less** than our alpha level (0.05), we **reject** H_0. We **do** have convincing evidence that commonly-white name resumés get a higher callback rate for jobs similar to the ones in this study.