# Hypothesis Testing for the Population Proportion $p$.

**We test:**

$$H_0: p = p_0 \qquad \text{vs.} \qquad H_a: \begin{cases} p < p_0 \\ p \neq p_0 \\ p > p_0 \end{cases}$$

**Test Statistic = ?**

Sample Proportion $\hat{P}$ of "successes";

we know its (approximate) sampling dist'n **under the null hypothesis.**

For a large enough sample, we know that:

- the count r.v. satisfies:

$$X \approx \underline{\text{Normal}} \left( \text{mean} = \underline{n \cdot p_0}, \; sd = \sqrt{n \, p_0 (1 - p_0)} \right)$$

- the Sample proportion r.v. satisfies:

$$\hat{P} = \frac{X}{n} \approx \text{Normal} \left( \text{mean} = p_0, \; sd = \sqrt{\frac{p_0(1 - p_0)}{n}} \right)$$

The observed value of the sample proportion is denoted by $\hat{p}$.
The corresponding value of the $z$-statistic under the null hypothesis:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

By def'n, the p·value is the probability of observing what
we've "observed" or something "more extreme" under the null

$$\begin{cases} \text{IF} \quad H_a: p < p_0, \text{ then } p\cdot value = \mathbb{P}[Z < z] = \text{pnorm}(z) \\ \text{IF} \quad H_a: p \neq p_0, \text{ then } p\cdot value = \mathbb{P}[Z < -|z|] + \mathbb{P}[Z > |z|] \\ \qquad\qquad\qquad\qquad\qquad\qquad = 2 * \text{pnorm}(-abs(z)) \\ \text{IF} \quad H_a: p > p_0, \text{ then } p\cdot value = \mathbb{P}[Z > z] = 1 - \text{pnorm}(z) \end{cases}$$

If a significance level $\alpha$ is given, then ...

IF p·value $\leq \alpha$, then REJECT THE NULL.

IF p·value $> \alpha$, then FAIL TO REJECT THE NULL.

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 16

Hypothesis testing: One-sample proportion.

**Problem 16.1.** *Source: Problem* **8.99** *from the Moore/McCabe/Craig.*
*Castaneda v. Partida* is an important court case in which statistical methods were used as part of a legal argument. When reviewing this case, the Supreme Court used the phrase "two or three standard deviations" as a criterion for statistical significance. This Supreme Court review has served as the basis for many subsequent applications of statistical methods in legal settings. (The two or three standard deviations referred to by the Court are values of the z statistic and correspond to $p$-values of approximately 0.05 and 0.0026.)

In Castaneda the plaintiffs alleged that the method for selecting juries in a county in Texas was biased against Mexican Americans. For the period of time at issue, there were $181,535$ persons eligible for jury duty, of whom $143,611$ were Mexican Americans. Of the 870 people selected for jury duty, 339 were Mexican Americans.

(i) (1 point) What proportion of eligible jurors were Mexican Americans?

$$p_0 = \frac{143,611}{181,535} = 0.7911$$

(ii) (2 points) Let $p$ denote the probability that a randomly selected juror is a Mexican American. Formulate the null and alternative hypotheses to be tested.

$$H_0: p = p_0 = 0.7911 \quad \text{against} \quad H_a: p < p_0 = 0.7911$$

(iii) (1 point) What is the sample proportion of jurors who were Mexican American?

$$\hat{p} = \frac{339}{870} = 0.3897$$

(iv) (4 points) Compute the $z-$statistic, and find the $p-$value.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.3897 - 0.7911}{\sqrt{\frac{0.7911(1-0.7911)}{870}}} = -29. \ldots$$

$\Rightarrow$ the p·value is virtually zero.

(v) (2 points) How would you summarize your conclusions? (A finding of statistical significance in this circumstance does not constitute proof of discrimination. It can be used, however, to establish a prima facie case. The burden of proof then shifts to the defense.)

There is evidence in favor of the prima facie case!

INSTRUCTOR: Milica Čudina

# Statistical Inference for Two Proportions.

Our parameters of interest:

$p_i$ , $i = 1, 2 \ldots$ the (sub)population proportion for the (sub)population $i = 1, 2$.

e.g., $p_1 \ldots$ corresponds to the (sub)population who get the treatment;

$p_2 \ldots$ corresponds to the (sub)population who get the placebo.

Sample sizes are denoted by $n_1$ and $n_2$.

Assume that the two samples are independent.