§6.3 Testing Binomial Data (for unknown proportion)

Suppose a sample of size n from $X \sim Bernoulli(p)$ with unknown p: $X_1 = k_1$, $X_2 = k_2$, ... , $X_n = k_n$, ($k_i \in \{0,1\}$).

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

Each X_i is Bernoulli with mean $E(X_i) = p$ and variance $\sigma^2 = p(1-p)$

By CLT, we know that $\overline{X} \sim \text{Normal}\left(p, \frac{p(1-p)}{n}\right)$ for large n. Then,

$$Z = \frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \text{Normal}(0,1)$$

To Test the Null Hypothesis $H_0 : p = p_0$ at the level of significance α, we define the **test statistic:**

$$z = \frac{\overline{k} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

1.(Right-sided) $H_0 : p = p_0$ versus $H_1 : p > p_0$.
Decision Rule: Reject H_0 if $z > z_\alpha$.

2. (Left-sided) $H_0 : p = p_0$ versus $H_1 : p < p_0$.

Decision Rule: Reject H_0 if $z < -z_\alpha$.

3. (Two-sided) $H_0 : p = p_0$ versus $H_1 : p \neq p_0$.

Decision Rule: Reject H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$.

Example 1. (Example 4 in §5.3) A poll was conducted to find out the percentage of people who will vote A or B for mayor of a city. Out of 500 people polled, 263 said A and the rest said B.

(1) Can we conclude that A, B are even with level of significance 0.05?

Let p be the probability that a person will vote A.
$H_0 : p = 0.5$ versus $H_1 : p \neq 0.5$.
Since $\alpha = 0.05$, then $z_{\alpha/2} = \textbf{invNorm}(0.975, 0, 1) \approx 1.96$.
The average $\overline{k} = 263/500 = 0.526$. The **test statistic** is

$$z = \frac{\overline{k} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.526 - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{500}}} = 1.16 < z_{\alpha/2}$$

So, we fail to reject H_0.

(2) Can we conclude that B is 2% ahead with level of significance 0.05?

Let p be the probability that a person will vote B.
$H_0 : p = 0.51$ versus $H_1 : p < 0.51$.
Since $\alpha = 0.05$, then $z_{\alpha} = \textbf{invNorm}(0.95, 0, 1) \approx 1.65$.
The average is $\overline{k} = 0.474$. The **test statistic** is

$$z = \frac{\overline{k} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{0.474 - 0.51}{\sqrt{\dfrac{0.51(1-0.51)}{500}}} = -1.61 > -z_{\alpha}$$

So, we fail to reject H_0.

(3) Can we conclude that A is 2% ahead with level of significance 0.05?

Let p be the probability that a person will vote A.
$H_0 : p = 0.51$ versus $H_1 : p < 0.51$.
$z = 0.71 > -z_{\alpha}$ implies that we fail to reject H_0.

We can also use the P-value to solve the problem.

If $H_1 : p > p_0$, then P-value=$P(\overline{X} \geq \overline{k} | H_0 \text{ is true})$

If $H_1 : p < p_0$, then P-value=$P(\overline{X} \leq \overline{k} | H_0 \text{ is true})$

If $H_1 : p \neq p_0$, then P-value=$2P(\overline{X} \geq \overline{k} | H_0 \text{ is true})$ or P-value= $2P(\overline{X} \leq \overline{k} | H_0 \text{ is true})$

Decision Rule: Reject H_0 if P-value $< \alpha$.

Example 2. Suppose 4 years ago in a voting campaign 40% of people supported a certain program. A recent poll of 420 people found that 151 people supported the program. Test at 5% significance level to see if the support for the program has changed.

Let p be the probability that a person will support the program.
$H_0 : p = 0.4$ versus $H_1 : p \neq 0.4$.
Since $\alpha = 0.05$, then $z_{\alpha/2} = \textbf{invNorm}(0.975, 0, 1) \approx 1.96$.
The average $\overline{k} = 151/420$. The **test statistic** is

$$z = \frac{\overline{k} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{151/420 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{420}}} = -1.693 > -z_{\alpha/2}$$

So, we fail to reject H_0.

Solution by P-Value: Suppose H_0 is true. By CLT,

$$\overline{X} \sim \text{Normal}(p_0, p_0(1-p_0)/n)$$

Since $\overline{k} < \mu_0$, the P-value is

$$P\text{-value} = (2)\textbf{normalcdf}(-\infty, 151/420, 0.4, \sqrt{\frac{0.4(1-0.4)}{420}}) \approx 0.0904 > \alpha$$

So, we fail to reject H_0.

Remark: If we use $\overline{k} = 151/420 \approx 0.36$, the error is big.

We can also use the calculator to find the test statistic and P-value:

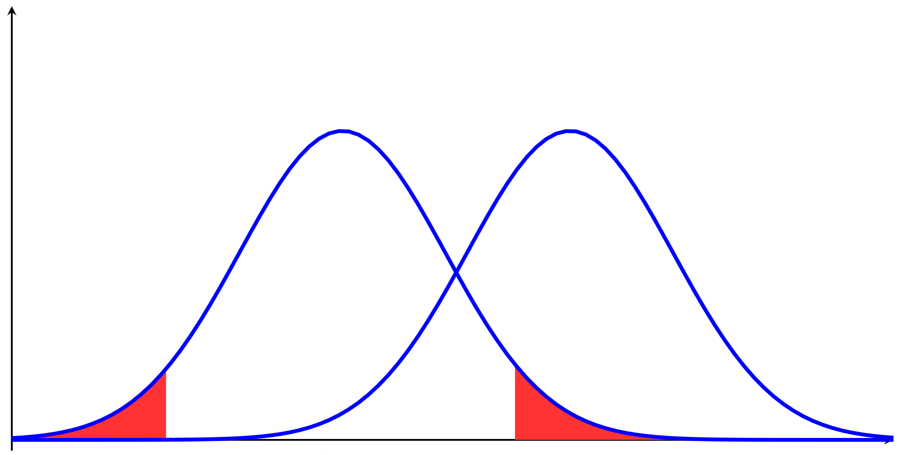
STAT $\rightarrow$ TESTS $\rightarrow$ 1-PropZTest

More calculator functions are summarized on the **summary file** on Blackboard/Course Materials.

Compare §5.3 with §6.2-6.3

Next, we compare the Confidence Interval and the Hypothesis Test.

We use Normal Distribution $X \sim \text{Normal}(\mu, \sigma^2)$ with known standard deviation σ as an example. Similar for the proportion.



Example 3. The 98% confidence interval for the mean is $[76.5, 83.5]$ with data from a normal distribution with known σ. Now a hypothesis test is run at the $\alpha = 0.02$ level of significance using $H_0 : \mu = 77$ against $H_1 : \mu \neq 77$.

(1) Find the sample mean $\bar{x}$.

$$\bar{x} = \frac{76.5 + 83.5}{2} = 80$$

(2) Find the test statistic value z for the test.

$z_{\alpha/2} \approx 2.326$. By margin of error, $z_{\alpha/2}(\sigma/\sqrt{n}) = \frac{83.5 - 76.5}{2} = 3.5$. So $\sigma/\sqrt{n} = \frac{3.5}{z_{\alpha/2}} \approx 1.505$.

The test statistic value $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{80 - 77}{1.505} \approx 1.99$.

(3) Would the null hypothesis be rejected at the $\alpha = 0.02$ level of significance?

(A) Reject H_0. (B) Fail to reject H_0. (C) Information is not enough.

(4) Would the null hypothesis be rejected at the $\alpha = 0.01$ level of significance?

(A) Reject H_0. (B) Fail to reject H_0. (C) Information is not enough.