Test about a Population Proportion

Let $p$ denote the proportion of individuals or objects in a population who possess a specified property (labeled as “$S$”). Let $X$ be the number of $S$’s in the sample. Then $\hat{p} = \frac{X}{n}$ is the sample proportion. $X$ is a binomial random variable with parameters $p$ and $n$, i.e. $X \sim Bin(n, p)$.

Furthermore, when the sample size $n$ itself is large, both $X$ and $\hat{p}$ are approximately normally distributed, i.e. $X \sim N(np, np(1 - p))$ and $\hat{p} \sim N(p, \frac{p(1-p)}{n})$.

Test about population proportion $p$ will depend on the sample size.
Large-Sample Tests
When the sample size is large \((n \geq 30)\), \(\hat{p}\) is approximately normally distributed with mean \(p\) and variance \(p(1 - p)/n\). In particular, under the null hypothesis \(H_0 : p = p_0\), \(\hat{p}\) is approximately normally distributed with mean \(p_0\) and variance \(p_0(1 - p_0)/n\), i.e. \(\hat{p} \sim N(p_0, p_0(1 - p_0)/n)\).
Therefore the test statistic

\[
Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}
\]

has approximately a standard normal distribution.
Example: Problem 35
State DMV records indicate that of all vehicles undergoing emissions testing during the previous year, 70% passed on the first try. A random sample of 200 cars tested in a particular county during the current year yields 124 that passed on the initial test. Does this suggest that the true proportion for this county during the current year differs from the previous statewide proportion? Test the relevant hypotheses using significance level $\alpha = 0.5$.

In this case, $\hat{p} = 124/200 = 0.62$.
The null hypothesis is $H_0 : p = 0.70$ and the alternative hypothesis is $H_a : p \neq 0.70$. 
2. Null hypothesis: $H_0 : p = 0.70$.
3. Alternative hypothesis: $H_a : p \neq 0.70$.
4. Test statistic value:

$$ z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.62 - 0.70}{\sqrt{0.70(1 - 0.70)/200}} = -2.469 $$

5. Rejection region: $z \geq z_{0.025}$ or $z \leq -z_{0.025}$, where $z_{0.025} = 1.96$.
6. Conclusion: since $-2.469 < -1.96$, we reject $H_0$. 

Test about a Population Proportion
Test about a Population Proportion

Summary for large-sample tests for population proportion $p$

Null hypothesis: $H_0 : p = p_0$

Test statistic value: $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

<table>
<thead>
<tr>
<th>Alternative Hypothesis</th>
<th>Rejection Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a : p &gt; p_0$</td>
<td>$z \geq z_\alpha$ (upper-tailed)</td>
</tr>
<tr>
<td>$H_a : p &lt; p_0$</td>
<td>$z \leq -z_\alpha$ (lower-tailed)</td>
</tr>
<tr>
<td>$H_a : p \neq p_0$</td>
<td>either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed)</td>
</tr>
</tbody>
</table>

Remark: These test procedures are valid provided that $np_0 \geq 10$ and $n(1 - p_0) \geq 10$. 
Test about a Population Proportion

Type II Error $\beta$ for large-sample tests

Alternative Hypothesis $\beta(p')$

\[
H_a : p > p_0 \quad \Phi \left[ \frac{p_0 - p' + z_\alpha \sqrt{p_0 (1-p_0)/n}}{\sqrt{p' (1-p')/n}} \right]
\]

\[
H_a : p < p_0 \quad 1 - \Phi \left[ \frac{p_0 - p' - z_\alpha \sqrt{p_0 (1-p_0)/n}}{\sqrt{p' (1-p')/n}} \right]
\]

\[
H_a : p > p_0 \quad \Phi \left[ \frac{p_0 - p' + z_{\alpha/2} \sqrt{p_0 (1-p_0)/n}}{\sqrt{p' (1-p')/n}} \right] - \Phi \left[ \frac{p_0 - p' - z_{\alpha/2} \sqrt{p_0 (1-p_0)/n}}{\sqrt{p' (1-p')/n}} \right]
\]
Test about a Population Proportion

**Small-Sample Tests**

When the sample size $n$ is small ($n \leq 30$), we test the hypotheses based directly on the binomial distribution.

For example, if the null hypothesis is $H_0 : p = p_0$ and the alternative hypothesis is $H_a : p > p_0$, then the rejection region is of the form $X \geq c$, where $X \sim Bin(n, p)$.

\[
P(\text{type I error}) = P(\text{reject } H_0 \mid H_0) = P(X \geq c \mid p = p_0) \\
= 1 - P(X < c \mid p = p_0) = 1 - P(X \leq c - 1 \mid p = p_0) \\
= 1 - B(c - 1; n, p_0)
\]

And

\[
P(\text{type II error}) = P(\text{fail to reject } H_0 \mid p = p') = P(X < c \mid p = p') \\
= P(X \leq c - 1 \mid p = p') = B(c - 1; n, p')
\]
Test about a Population Proportion

Small-Sample Tests
Remark: in the small-sample case, it is usually not possible to find a value $c$ for which $P(\text{type I error})$ is exactly the desired significance level $\alpha$. Therefore we choose the largest rejection region which satisfying

$$P(\text{type I error}) < \alpha.$$
Example:
A coin is tossed 10 times and $\times = 6$ heads are observed. Let $p = P(\text{head})$. Do you believe the coin prefers head with significance level 0.10?
Test about a Population Proportion

2. Null hypothesis: $H_0 : p = 0.5$.
3. Alternative hypothesis: $H_a : p > 0.5$.
4. Test statistic: $X = \text{number of heads}$.
5. Rejection region: $X \geq c$ for some $c$.
6. Significance level:

\[
0.10 \geq P(\text{type I error}) = P(\text{reject } H_0 \mid H_0) = P(X \geq c \mid p = 0.5) = 1 - P(X < c \mid p = 0.5) = 1 - B(c - 1; 10, 0.5)
\]

Therefore $c - 1 = 7$ and $c = 8$.
7. Conclusion: since $6 < 7$, we fail to reject $H_0$. 