Introduction to hypotheses testing

A sadistic monarch hands one of his subjects a coin and states that the probability of tossing a head, say \( p \), is either 0.3 or 0.8. The subject is required to toss the coin 5 times and then to state whether she believes \( p \) to be either 0.3 or 0.8. If the subject makes the correct decision she wins 100 pieces of gold. If she says \( p = 0.8 \) when \( p = 0.3 \) she will go to jail for 7 years. If she says \( p = 0.3 \) when \( p = 0.8 \) she will go to jail for 1 month.

The dilemma facing the subject can be formulated as a statistical testing problem.

Let \( X \sim \text{Binomial}(5, p) \). After observing \( X \) we must decide between the two hypotheses

\[
H : p = 0.3 \quad K : p = 0.8
\]

\( H \) is call the Null hypothesis.

\( K \) is called the Alternative hypothesis.

Before actually tossing the coin 5 times the subject can perform a thought experiment where she just imagines tossing the coin. For each possible outcome of \( X = x \) she can decide whether or not she would reject \( H \), that is decide \( p = 0.8 \). (For us rejecting \( H \) is the same as accepting \( K \). Similarly accepting \( H \) is the same as rejecting \( K \).)

A strategy for the subject is to determine for what points in \( \{0, 1, 2, 3, 4, 5\} \), the sample space of \( X \), she wishes to reject \( H \). We will call such a set a critical region and denote it by \( C \). How should the subject evaluate a possible critical region? Our answer depends on the lack of symmetry in the consequences of making the two types of error.

The two types of error are:

1. Type I error: Rejecting \( H \) when in fact \( H \) is true.
2. Type II error: Accepting \( H \) when in fact \( K \) is true.

For the subject the Type I error is deciding \( p = 0.8 \) when \( p = 0.3 \) is true and the Type II error is deciding \( p = 0.3 \) when \( p = 0.8 \). Note that for the subject the Type I error is the more serious of the two. The theory is based on this assumption and a testing problem needs to be set up to reflect this fact.

To evaluate a critical region \( C \) we must find the probability of making the Type I error when \( H \) is true and the probability of making the Type II error when \( K \) is true and we are using \( C \).

\[
\alpha = \alpha(C) = \text{Probability of making Type I error}
\]

\[
= P_H(X \in C)
\]

\[
= \sum_{x \in C} \binom{5}{x} (0.3)^x (0.7)^{5-x}
\]
and

\[ \beta = \beta(C) = \text{Probability of making Type II error} \]
\[ = P_K(X \notin C) \]
\[ = \sum_{x \notin C} \binom{5}{x} (0.8)^x (0.2)^{5-x} \]

We evaluate the critical region \( C \) by considering its two error probabilities \( \alpha = \alpha(C) \) and \( \beta = \beta(C) \).

<table>
<thead>
<tr>
<th>( C )</th>
<th>( \alpha(C) )</th>
<th>( \beta(C) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>{1, 2, 3, 4, 5}</td>
<td>0.8319</td>
</tr>
<tr>
<td>2</td>
<td>{2, 3, 4, 5}</td>
<td>0.4718</td>
</tr>
<tr>
<td>3</td>
<td>{3, 4, 5}</td>
<td>0.1630</td>
</tr>
<tr>
<td>4</td>
<td>{4, 5}</td>
<td>0.0308</td>
</tr>
<tr>
<td>5</td>
<td>{5}</td>
<td>0.0024</td>
</tr>
<tr>
<td>6</td>
<td>{0, 1, 2, 3, 4, 5}</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>\emptyset</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>{0, 1, 2}</td>
<td>0.8369</td>
</tr>
</tbody>
</table>

Note the critical region 8 is silly. In fact all of the critical regions from 1 through 5 are better than it.

One can prove that the critical regions 1 through 7 are the only sensible ones for this testing problem.

No best choice among 1 through 7. The answer depends on how strongly the subject wishes to avoid making the Type I error.