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

$$H : p = 0.3 \qquad K : p = 0.8$$

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 $\mathcal{C}$. 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 $\mathcal{C}$ 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 $\mathcal{C}$.

$$\begin{aligned}\alpha = \alpha(\mathcal{C}) &= \text{Probability of making Type I error} \\ &= P_H(X \in \mathcal{C}) \\ &= \sum_{x \in \mathcal{C}} \binom{5}{x} (0.3)^x (0.7)^{5-x}\end{aligned}$$

and

$$\begin{aligned}\beta = \beta(\mathcal{C}) &= \text{Probability of making Type II error} \\ &= P_K(X \notin \mathcal{C}) \\ &= \sum_{x \notin \mathcal{C}} \binom{5}{x} (0.8)^x (0.2)^{5-x}\end{aligned}$$

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

	$\mathcal{C}$	$\alpha(\mathcal{C})$	$\beta(\mathcal{C})$
1	$\{1, 2, 3, 4, 5\}$	0.8319	0.0003
2	$\{2, 3, 4, 5\}$	0.4718	.0067
3	$\{3, 4, 5\}$	0.1630	0.0579
4	$\{4, 5\}$	0.0308	0.2627
5	$\{5\}$	0.0024	0.6723
6	$\{0, 1, 2, 3, 4, 5\}$	1	0
7	ϕ	0	1
8	$\{0, 1, 2\}$	0.8369	0.9421

Note the critcal region 8 is silly. In fact all of the critical regions from 1 trough 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 amoung 1 through 7. The answer depends on how strongly the subject wishes to avoid making the Type I error.