§9. Hypothesis Testing; Binomial Random Variable

Estimating distribution parameters (e.g. p in Binom $\sim (n, p)$) is part of statistical inference. Two major branches of statistical inference are:

- Estimation
- Hypothesis Testing

> **Example 1**
>
> A manufacturing process, historically has 20% of its production resulting in defective items. A recent sample of 20 items shows that 6 items must be reworked. Has the manufacturing process changed?

Assume that a production process is binomial with probability for a defective item, p. Our null hypothesis will be that the process has not changed and the alternative that the proportion of defective items has increased.

$$H_0: \quad p = 0.20 \qquad \text{(null hypothesis)}$$

$$H_1: \quad p > 0.20 \qquad \text{(alternative hypothesis)}$$

We must decide what sampling results lead to fail to reject H_0 and what results will lead to its rejection and hence the acceptance of H_1.

The sampling is subject to variability $\Rightarrow$ our conclusions cannot be reached without running the risk of error.

	H_0 **True**	H_0 **False**
H_0 **Rejected**	Type I Error (α)	Correct
H_0 **Not Rejected**	Correct	Type II Error (β)

Both α and β are conditional probabilities that depend on our selection of the critical region; the set of sample values that will lead to rejection of H_0. So how should This critical region be chosen? Let's look at an example.

Example: Cont'd

Historical failure rate $p = 0.2$. Took a sample of $n = 20$

Let's decide to reject $H_0 : p = 0.2$ if 9 or more items are defective out of 20. So the critical region is $\{x \mid x \geq 9\}$. With this choice we have:

$$\begin{aligned}\alpha = P[x \geq 9 \mid H_o \text{ is True }] &= P[x \geq 9 \mid p = 0.2]] \\ &= \sum_{x=9}^{20} {}_{20}C_x (0.2)^x (0.8)^{20-x} \\ &= 0.00998179 \\ &\approx 0.01\end{aligned}$$

So, with this choice of critical region we will reject H_0 even if H_0 is true ~ 1% of the time. α is called **significance level** of the test. It indicates what Type I error we are willing to tolerate.

To compute β we need specific values for p which invalidates H_0 say $p = 0.3$. We have:

$$\beta = P[x < 9 \mid p = 0.30] = \sum_{x=0}^{8} {}_{20}C_x (0.3)^x (0.7)^{20-x} = 0.886669$$

This is very high risk (and why we say-fail to reject H_0).

What if our rejection region is $\{x \mid x \geq 8\}$? then

$$\beta = P[x < 8 \mid p = 0.30] = \sum_{x=0}^{7} {}_{20}C_x \, (0.3)^x (0.7)^{20-x} = 0.772272$$

$$\alpha = P[x \geq 8 \mid p = 0.20] = \sum_{x=8}^{20} {}_{20}C_x (0.2)^{\times} (0.8)^{20-x} = 0.032147$$

$1 - \beta$ is called **the power of the test**.

Example: Complete

Historical rate $p = 0.20$.
Took a sample $n = 20$
Found $x = 6$ defective.

a. $H_0 : p = 0.2, H_1 : p > 0.2$ Rejection region $\{x \mid x \geq 9\}$.
Sample $x = 6$

Fail to reject H_0. $\underbrace{\alpha = 0.01}_{\text{did not commit}}$ $\beta(p_t = 0.3) = 0.88$

b. $H_0 : p = 0.2, H_1 : p > 0.2$ Rejection region $\{x \mid x \geq 8\}$. Sample $x = 6$

Fail to reject H_0. $\underbrace{\alpha = 0.03}_{\text{did not commit}}$ $\beta(p_t = 0.3) = 0.77$

c. $H_0 : p = 0.2, H_1 : p > 0.2$ Rejection region $\{x \mid x \geq 6\}$. Sample $x = 6$

Reject H_0. Accept $H_1 : p > 0.2$

$$\alpha = P[x \geq 6 \mid p = 0.2] = \sum_{x=6}^{20} {}_{20}C_x (0.2)^x (0.8)^{20-x} = 0.195792$$

Type II error - did not commit.

Example 2

Say we would like to test $H_0 : p = 0.2$ vs $H_1 : p > 0.2$ with $\alpha \approx 0.05$ and $\beta\,(p_t = 0.3) \approx 0.10$.

What sample size and rejection region do we need?
$n = ?;$ $\{x \mid x \geq k\}, k = ?$

$$\alpha = \sum_{x=k}^{n} {}_nC_x (0.20)^x (0.80)^{n-x} = 0.05$$

$$\beta = \sum_{x=0}^{k-1} {}_nC_x (0.30)^x (0.70)^{n-x} = 0.10$$

Search with CAS yields $n \approx 156$ and $k \approx 40$.