Lecture 2

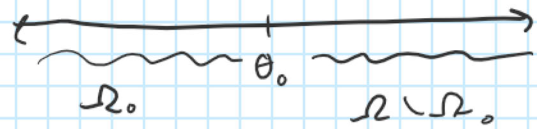
Thursday, January 9, 2020 2:00 PM

Recall

Decisions	Truth: H_0	H_A
Fail to reject H_0		Type II β
Reject H_0	Type I α	$(1-\beta)$ = Power

Type II $= P(\text{failing to reject } H_0 \mid H_A)$

Type I $= P(\text{rejecting } H_0 \mid H_0 \text{ true})$



Simple vs simple hypothesis:

$H_0: \theta = \theta_0$
$H_A: \theta = \theta_1 \quad (\theta_1 \neq \theta_0)$

α is called the "significance" of the test

Composite v.s. composite hypothesis:

$\begin{cases} H_0: \theta \leq \theta_0 \\ H_A: \theta > \theta_0 \end{cases}$ or $\begin{cases} H_0: \theta \geq \theta_0 \\ H_A: \theta < \theta_0 \end{cases}$ $\to$ α is the "size" of the test

(this is interpreted as the maximum probability of rejecting H_0 when H_0 is true; when maximizing the power function over the parameter space)

Returning to example on defection in batch #2

- Let θ be the proportion of defectives,

given $\begin{cases} H_0: \theta = 0.10 = \theta_0 \\ H_A: \theta = 0.15 = \theta_1 \end{cases}$

- If X_i is a random variable representing if part is defective, then $X_i \sim Ber(\theta)$ (Assume independence)
- Then, $\sum_{i=1}^{n} X_i \sim Bin(n, \theta)$
- Suppose $n = 20$, $\alpha = 0.05$

Q: Construct an exact test for this hypothesis.

- In the interest of being conservative, we would like to find a test which measures:

$$P(\text{Type I Error}) \leq \alpha$$

* Pgs 598-601 have binomial probability tables

- $P(\sum X_i \leq 3 \mid \theta = 0.1) \approx 0.8670 \Rightarrow P(\sum X_i \geq 4 \mid \theta = 0.1) \approx 0.1324$
- $P(\sum X_i \leq 4 \mid \theta = 0.1) \approx 0.9568 \Rightarrow P(\sum X_i \geq 5 \mid \theta = 0.1) \approx \boxed{0.04322}$

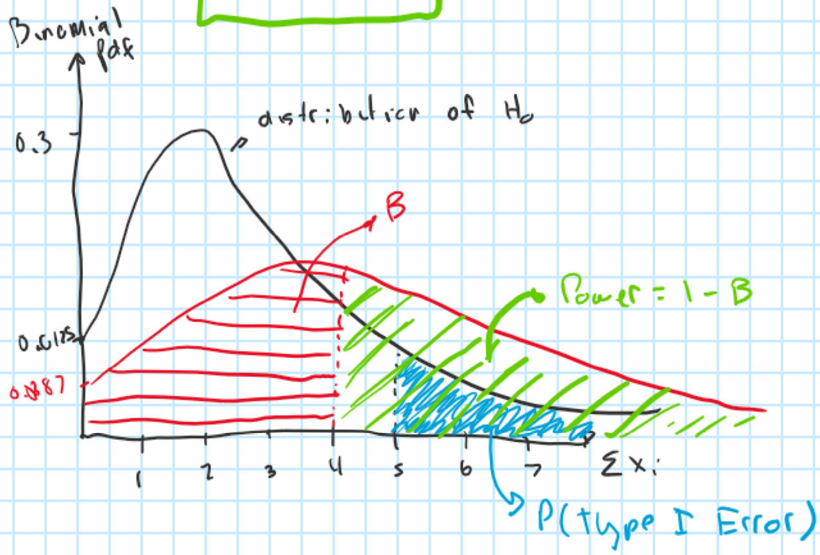
$\boxed{\text{* thus we choose } c = 5 \text{ for } P(\text{Type I}) \leq 0.05}$

$\boxed{\text{thus we choose } c=5 \text{ for } P(\text{Type I}) \le 0.05}$

- Now finding $P(\text{Type II})$ for a given rejection region.

$$C = \{ \underline{x} \mid \Sigma x_i \ge 5 \}$$

$$\beta = P(\text{Type II error})$$
$$= P(\text{fail to reject the null} \mid H_A)$$
$$= P(\Sigma x_i \le 4 \mid \theta = 0.15)$$
$$\boxed{\approx 0.8298}$$



- Let's make a table of c & β for different values n

n	c	β	α
20	5	0.8298	0.0432
40	8	0.7539	0.0419
100	16	0.5683	0.0398
10	4	0.9500	0.0128
10	3	0.8202	0.072

Take-away: Given α, we can choose n to attain our target β level.

- We use the normal approximation via the central Limit Theorem to find a method to calculate the n required as the exact test can only yield approximate n's through trial and error.

- Recall: $\hat{\theta}$ by the CLT is distributed approximately normal if:

$$n\theta \ge 10 \quad \text{and} \quad n(1-\hat{\theta}) \ge 10$$ thumb rules

- By CLT, $\hat{\theta} \sim N\left(\theta,\ SE(\hat{\theta}) = \sqrt{\frac{\theta(1-\theta)}{n}}\right)$

- Under H_0: $\hat{\theta} \sim N\left(\theta_0,\ SE(\hat{\theta}) = \sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right)$

- Under H_0: $\hat{\theta} \sim N\left(\theta_0, SE(\hat{\theta}) = \sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right)$

- let's use this approximation to find β for different n:

I ↳ Note: c should satisfy $P(\hat{\theta} \geq c \mid \theta_0 = 0.1) = 0.05$

OR

$$P\left[\frac{\hat{\theta}-\theta_0}{SE(\hat{\theta})} \geq \frac{c-0.1}{\sqrt{\frac{0.1(0.9)}{20}}}\right] = 0.05$$

II $$\beta = P\{\hat{\theta} < c \mid \theta = 0.15\}$$

$$= P\left\{\frac{\hat{\theta}-\theta}{SE(\hat{\theta})} < \frac{c-0.15}{\sqrt{\frac{0.15(0.85)}{20}}}\right\}$$

- let's make a table

n	c	B
10	0.26	0.83
20	0.21	0.78
40	0.18	0.69
100	0.15	0.49

- Determine sample size for α & β choices.

$H_0: \mu = \mu_0$

$H_A: \mu = \mu_1$ where $\mu_1 \geq \mu_0$

- Suppose α is fixed, then we can find β:

$$\beta = P(\text{Type II Error}) = P(\text{fail to reject } H_0 \mid H_A \text{ true})$$

$$= P(\bar{X} < c \mid \mu = \mu_1)$$

$$= P\left\{\frac{\bar{x}-\mu_1}{\sigma/\sqrt{n}} < \frac{c-\mu_1}{\sigma/\sqrt{n}}\right\}$$; we know $c = z_{1-\alpha} \cdot \sigma/\sqrt{n} + \mu_0$

$$= \beta = P\left\{\frac{\bar{x}-\mu_1}{\sigma/\sqrt{n}} < \frac{z_{1-\alpha} \cdot \sigma/\sqrt{n} + \mu_0 - \mu_1}{\sigma/\sqrt{n}}\right\}$$

because $z_{1-\alpha} = \frac{c-\mu_0}{\sigma/\sqrt{n}}$

$$= P\left\{\frac{\bar{x}-\mu_1}{\sigma/\sqrt{n}} < z_{1-\alpha} + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right\}$$

$$\Phi(z_\beta) = \Phi\left(z_{1-\alpha} + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$$

- We know $\beta = \Phi(z_\beta)$ since $\Phi()$ is CDF normal, monotonically increasing, $\Phi^{-1}()$ exists.

- $z_\beta = z_{1-\alpha} + \left(\frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$

- Given the symmetry of the standard normal about 0:

$$z_\beta = -z_{1-\beta}$$

$$\boxed{n = (z_{1-\alpha} + z_{1-\beta})^2 \sigma^2}$$

$$\boxed{n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \sigma^2}{(\mu_0 - \mu_1)^2}}$$

· Suppose $\alpha = 0.05$, $\beta = 0.1$ for $\sigma^2 = 16$,

and $\mu_0 = 10$, $\mu_1 = 11$

$$n = \frac{(1.645 + 1.28)^2 \cdot 16}{(10 - 11)^2}$$

$$\boxed{\therefore n \approx 137}$$

What if $\mu_0 > \mu_1$??

$$\begin{aligned}
\beta &= P(\text{Type II Error}) \\
&= P\left(\bar{X} > \mu_0 + z_\alpha \sigma/\sqrt{n} \,\middle|\, \mu = \mu_1\right) \\
&= 1 - \Phi\left(z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

$$\begin{aligned}
\alpha &= P(\bar{X} < c \mid \mu = \mu_0) \\
&= 1 - \Phi\left(z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right) \\
&= 1 - \Phi(z_\beta)
\end{aligned}$$

⋮

$$n = \frac{\sigma^2 (z_\beta + z_\alpha)^2}{(\mu_0 - \mu_1)^2}$$

→ HW
Problems up to 7