# CHAPTER 4

# HYPOTHESIS TESTING

If a new form of therapy is tested on 100 patients and gives good results in 64 of them, while this only holds for 50% of the patients with the old therapy, is the new therapy truly better than the old one, or were we just "lucky"?

This is a typical question that concerns a decision between two conflicting hypoteses.

$\boxed{\text{Set-up}}$:

$$\underline{X} \sim \text{Model } p_\theta(\underline{x}) \quad \theta \in \Theta$$

Null hypothesis $\theta \in \Theta_0$ or

Alternative $\theta \in \Theta_1 = \Theta \setminus \Theta_0$.

$\boxed{\text{Often the 2 hypotheses are not treated symmetrically}}$

- like in trials we want to know if $H_1$ is correct.
- If there is not enough evidence to reject $H_0$

$$\not\Rightarrow \begin{cases} H_1 \text{ incorrect} \\ \text{and } H_0 \text{ correct} \end{cases}$$

because it can also mean that there are not sufficient proofs for either of the hypotheses.

## Possible Conclusions

- Reject $H_0$ And accept $H_1$ as correct. strong
- Do not reject $H_0$ but do not accept $H_1$ as being correct. weak (more investigation needed)

## Possible errors

- type I : Reject $H_0$ when correct

$\Rightarrow$ accept strong conclusion what is wrong.

- type II : Mantain $H_0$ when incorrect.

$\Rightarrow$ accept the weak conclusion

ASYMMETRY

| | |
|---|---|
| type I | VERY UNDESIRABLE |
| type II | not as bad. |

Wisely choose the hypoteses :
In principle, choose the statement we want to "prove" as $H_1$.

**Example 4.1** $p = Pr$ (success) of a new therapy

$H_0 : p \leq 0.5 \quad H_1 : p > 0.5$

## Example 4.2 Playing dice

$\theta = (p_1 \dots p_6)$ $\quad H_0 : p_i = \frac{1}{6} \quad \forall i$

$H_1 : p_i \neq \frac{1}{6}$ at least on $i$

## Example 4.3 Two samples model

$(X_1 \dots X_m)$ $\quad (Y_1 \dots Y_n)$ all independent.

$\downarrow \qquad\qquad \downarrow$

$N(\mu, \sigma^2) \qquad N(\nu, \tau^2)$

See Notebook 4 Box-plot not helpful

$$\begin{cases} H_0 : \mu = \nu \\ H_1 : \mu \neq \nu \end{cases}$$ formal test.

$$\theta = (\mu, \nu, \sigma^2, \tau^2) \in \mathbb{R}^2 \times (0, \infty)^2$$

$\Theta_0$ for the null is

$$\Theta_0 = \{(\mu, \mu) : \mu \in \mathbb{R}\} \times (0, \infty)^2.$$

## $\boxed{\text{4.3 Sample size an critical region}}$

The decision on rejection of $H_0$ is based on the data $\underline{X} = (X_1, \dots X_n)$

A statistical test on the hypothesis $H_0$ is defined by the critical region $K$, a subset of the sample space

if $\underline{X} \in K \Rightarrow$ Reject $H_0$

if $\underline{X} \notin K \Rightarrow$ don't reject $H_0$

Typically $K$ is defined by a test statistic $T(\underline{X})$, real valued, that defines the distance between the data and the hypothesis.

$\boxed{\text{Example 4.5}}$ Binomial test.

- Experiment: give a new therapy to 100 patients.
- Let $X$ = # successes $\sim Bin(100, p)$
- Test if the new therapy has probability of success $> 0.5$.

- Test statistic $T(X) = X$
- Critical region $K = \{T(X) \geq c\}$

CRITICAL VALUE TO BE CHOSEN

So: $K = \{c, c+1, \dots, 100\}$.

A large value of $X$ gives an indication that $H_0 : p \leq 0.5$ is incorrect.

## Example 4.6 – Gauss test

$(X_1, \dots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ $\sigma^2$ known.
(unrealistic but see later)

Test of

$$H_0 : \mu \leq \mu_0 \text{ against } H_1 : \mu > \mu_0$$

Example : Quality control based on a sample $\underline{X}$ and a Teststatistic

$$T(\underline{X}) = \bar{X}$$

$$K = \{ \underline{x} : \bar{x} \geq c \}.$$

c chosen to get a sufficiently small probability of type I error.

**Notes:** – Two different test statistics may give the same K.

$$K = \left\{ \underline{x} : \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \geq z \right\}$$

- A different system of the hypotheses implies a different $K$.

$$H_0 : \mu \geq \mu_0 \qquad H_1 : \mu < \mu_0$$

$$K = \{ \underline{x} : \bar{x} \leq c \}.$$

## $\boxed{4.3.1}$ Size and power function

$\boxed{\text{what we want}}$

if $\theta \in \Theta_0$ $P(\underline{x} \in K)$ small

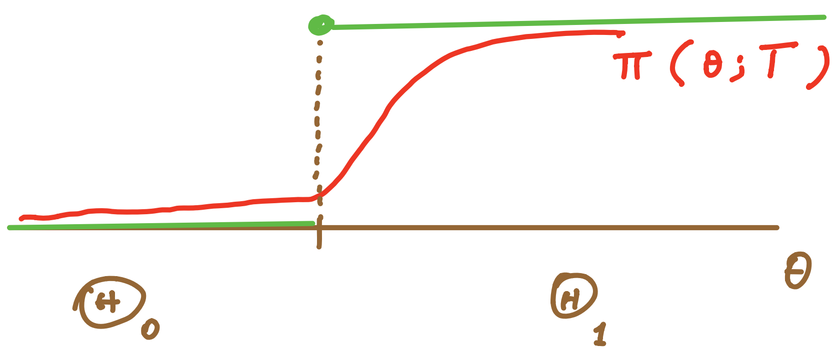
if $\theta \notin \Theta_0$ $P(\underline{x} \in K)$ high

A function that measures the quality (or sensitivity) of a test is

$$\theta \longmapsto P_\theta(\underline{X} \in K) = \pi$$

called $\boxed{\text{Power function}}$

Example of a real power function and ideal




Size of a test $T$

$$\alpha = \sup_{\theta \in \Theta_0} \pi(\theta; T)$$

A test has level $\alpha_0$ if $\alpha_0 \geq \alpha$.

Convention in testing (A)

Choose first a level $\alpha_0$ and then use only tests of size $\alpha \leq \alpha_0$.

- It's appealing to choose tests with a very small level
- But this implies that the probability

$$P(\mathrm{II}) = P_\theta(\underline{X} \notin K) \quad \theta \in \Theta_1$$

tends to become very large.
- We don't try to minimize the sum of the two errors $P(\mathrm{I}) + P(\mathrm{II})$
- The level $\alpha_0$ should be chosen depending on the consequences of the decision, but...
- in practice we choose only tests with a level 0.05
- Of the tests of level 0.05 we prefer those having the smallest $P(\mathrm{II})$.

## Convention in testing (B)

Given level $\alpha_0$, we prefer a test with level $\alpha_0$ that has the highest $\pi(\theta; K)$ for $\theta \in \Theta_1$

## Example 4.11 Binomial test.

## Example 4.12 - Gauss test

$\underset{\sim}{X} = (X_1, \dots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ ↳ known

$H_0: \mu \leq \mu_0 \quad H_1: \mu > \mu_0.$

Test statistic $T(\underset{\sim}{x}) = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

Equivalent to $T(\underset{\sim}{x}) = \bar{x}$.

Critical region $K = \left\{ \underset{\sim}{x} : \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq c \right\}$

Note that

$$\boxed{T(\underset{\sim}{x}) \sim N(0,1) \text{ under } \mu = \mu_0}$$

So we look for a test of level at most $\alpha_0$

$$\sup_{\mu \leq \mu_0} P_\mu\left( \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq c \right) \leq \alpha_0 \quad (*)$$

**Result**

The critical region of level at most $\alpha_0$ is obtained for $c \geq \xi_{1-\alpha_0}$ where $\xi_{1-\alpha}$ is the $1-\alpha_0$-quantile of the $N(0,1)$

Proof:

$$P_\mu\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \geq c\right) = P_\mu\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq c + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(c + \frac{\mu_0-\mu}{\sigma/\sqrt{n}}\right).$$

is an increasing function of $\mu$.


So the $\sup_{\mu \leq \mu_0}$ is attained for $\mu = \mu_0$.

Condition (*) $\iff P_{\mu_0}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq c\right) \leq \alpha_0$

$$\iff 1-\Phi(c) \leq \alpha_0$$

$$\iff 1-\alpha_0 \leq \Phi(c)$$

$$\iff \Phi^{-1}(1-\alpha_0) \leq c$$

and $c \geq \xi_{1-\alpha_0}$ is the solution. CHOOSE $c = \xi_{1-\alpha_0}$ //

Critical region for the gauss test

$H_0 \geq \mu_0 \quad H_1 < \mu_0$

$$K = \left\{ \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \leq -\xi_{1-\alpha_0} \right\}$$

critical region for the Gauss test

$H_0 = \mu_0 \quad H_1 \neq \mu_0$

combining two one-sided regions with level $\alpha_0/2$ each

$$K = \left\{ \left| \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \right| \geq \xi_{1-\alpha_0/2} \right\}$$

See Notebook 4