Testing Statistical Hypothesis (Section 9.4)

Set up:

Population: $X \sim f_\theta(x)$

may be a vector $\rightarrow$ θ = unknown

$X_1, X_2, \ldots, X_n$ (Random sample).

Testing hypotheses: Verifying claims regarding unknown θ based upon the evidence provided by the data.

Hypotheses: Two *mutually exclusive* statements about θ.

- **Null hypothesis** H_0: Value of θ corresponding to "status quo", "common belief", "no change", etc. Often, $H_0 : \theta = \theta_0$ (a given value)
- **Alternative hypothesis** H_1: The claim the researcher is hoping to prove.

null value

Three possible hypotheses in this course:

known

- (**Two-sided**) $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$
- (**One-sided, right-tailed**)
  $H_0 : \theta = \theta_0$ (or $\theta \leq \theta_0$) against $H_1 : \theta > \theta_0$
- (**One-sided, left-tailed**)
  $H_0 : \theta = \theta_0$ (or $\theta \geq \theta_0$) against $H_1 : \theta < \theta_0$

EX:

$H_0 : \theta = 0.5$ vs.

$H_1 : \theta \neq 0.5$

Let us set up the hypotheses in the following examples.

Ex 1: A long-time authorized user of a computer account takes 0.2 seconds on average between keystrokes. One day, when a user typed in the username and password, 15 times between keystrokes were recorded. These data had mean of 0.3 seconds and standard deviation of 0.12 seconds. Do these data given evidence of an unauthorized login attempt?

X = time b/w keystrokes for the person who is trying to log in.

$$\boxed{\begin{array}{l} X \sim f_\theta(x) \\ \theta = E[X]. \end{array}}$$

$X_1, X_2, \ldots, X_n,\quad n = 15$

H_0: attempt is not unauthorized, i.e., $\theta = 0.2$

θ_0.

H_1: attempt is unauthorized, i.e., $\theta \neq 0.2$

Ex 2: The number of concurrent users for an ISP has historically averaged 5000. After a marketing campaign, the management would like to know if it has resulted in an increase in the number of concurrent users. To test this, data were collected by observing the number of concurrent users at 100 randomly selected moments of time. Suppose that the average and the standard deviation of the sample data are 5200 and 800, respectively. Is there evidence that the mean number of concurrent users has increased?

$X = $ # concurrent users.

$\theta = E[X]$.

$H_0: \theta = 5000$

$H_1: \theta > 5000$

Ex 3: A recent poll of 1,000 American people estimated that the approval rating of the current congress is 31%. Do these data give evidence that less than 30% of the American people approve the performance of the congress?

X = indicator of whether a randomly selected American approves the performance or not [i.e., $X=1$ or 0].

θ = proportion of American who approve the performance of the congress

$H_0: \theta = 0.30$

$H_1: \theta < 0.30$

Outcome of a hypotheses test: Accept H_0 or reject H_0 (i.e., accept H_1)

- We do not know the truth. (If we knew, there was no point in collecting data.)
- H_0 is rejected **only if** there is strong evidence against it, otherwise H_0 is accepted.
- Evidence is provided by the data.
- If H_0 is accepted, it doesn't mean that H_0 is true. It just means that there is not enough evidence in the data to reject it.
- If H_0 is rejected, it doesn't mean that H_1 is true. It simply means that the data strongly favors H_1.
- Analogous to a court case.

Two types of errors:

	Truth	
Test outcome	H_0 is true	H_1 is true
Accept H_0	NO error	type II error
Reject H_0 (accept H_1)	type I error	NO error

- Trade-off between the two error probabilities. As one decreases, the other increases. So, it may not be possible for a procedure with a given sample size n to have both probabilities to be small.
- Hypotheses are set up in a way that ensures *type I error is more serious than type II error.*

- Design a test procedure that guarantees that its type I error probability does not exceed a small prescribed value α, known as the **level of significance** or simply the α level of the test. , i.e., $P[\text{type I error}] = P[\text{Reject } H_0 \mid H_0 \text{ is true}] \le \alpha$.
- In practice, $\alpha = 0.01, 0.05$ (most popular), or 0.10. In practice, $P[\text{type I error}] = \alpha$.
- No guarantee of P(type II error). We try to keep it small by choosing a large enough n.
- Power of test $=$ 1−P(type II error).
- Typically, the error probabilities depend on the true θ.

Possible to do a sample size computation that guarantees a specified power at a specified alternative — not in the course.

Analogy with a court case

A suspect is brought to the court — "presumed innocent until proven guilty."

H_0: suspect is innocent

H_1: suspect is ~~guilty~~ guilty

H_0 is rejected (i.e., the suspect is convicted) only if there is strong evidence against his/her innocence. Otherwise, H_0 is accepted (i.e., the suspect is acquitted).

Type I error: $\underbrace{\text{convicting}}_{\text{Reject } H_0}$ an $\underbrace{\text{innocent suspect}}_{H_0 \text{ is true}}$.

Type II error: $\underbrace{\text{Acquitting (i.e., releasing)}}_{\text{Accept } H_0}$ a $\underbrace{\text{guilty suspect}}_{H_1 \text{ is true}}$.

Basic premise: Reject H_0 only if evidence against H_0 is strong; otherwise accept H_0.

Q: How can we make P(type I error) = 0? What happens to P(type II error)?

Release everyone $\Rightarrow P[\text{type I error}] = 0$

$\Rightarrow P[\text{type II error}] = 1.$

Q: How can we make P(type II error) = 0? What happens to P(type I error)?

Convict everyone $\Rightarrow P[\text{type II error}] = 0$

$\Rightarrow P[\text{type I error}] = 1.$

Note: Tradeoff b/w the two error prob. It is not possible to minimize both error prob. ~~fix~~ with a procedure based on fixed n.

A general approach ~~for~~ to get a level α test

- Estimate θ by its point estimator $\hat{\theta}$
- Compute s.e.$(\hat{\theta})$ assuming $\theta = \theta_0$. Estimate it if it's unknown.
- Compute a **test statistic** T that measures measures how consistent the data are with H_0. Often, T has the form:

$$T = \frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})}.$$

- Find the **null distribution** — the distribution of T assuming H_0 is true.
- Find the form of the **rejection region** $\mathcal{R}$ — the set of values of T for which H_0 is rejected.
- **Acceptance region** $\mathcal{A}$ = Complement of $\mathcal{R}$.
- Determine $\mathcal{R}$ by ensuring that the level of significance of the test is α, i.e., $P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$.