Testing Statistical Hypothesis (Section 9.4)

Set up:

Ex:

$$\boxed{\begin{array}{l}\text{Population: } X \sim f_\theta(x) \\ \theta = \text{unknown}\end{array}}$$

$$\downarrow$$

$X_1, \ldots, X_n$ (RS)

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)

null value

- **Alternative hypothesis** H_1: The claim the researcher is hoping to prove.

Note: Hypotheses are statements about θ — Nothing from data should be used to formulate the hypotheses.

Three possible hypotheses in this course:

- (**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 $H_1 : \theta < \theta_0$

Let us set up the hypotheses in the following examples. $\Rightarrow$ take μ as parameter of interest

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?

Recall: X = (typical) time b/w two keystrokes for the user who is trying to log in

$\mu = E[X]$

$H_0 : \mu = 0.2$ [i.e., attempt is not unauthorized]

$H_1 : \mu \neq 0.2$ [i.e., " " unauthorized]

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?

Here! X = # concurrent users at a typical time.

$\mu = E[X]$ — of interest

$H_0: \mu = 5000$ (i.e., no change from before)

$H_1: \mu > 5000$ (i.e., campaign is successful)

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 approval for a typical American

$$= \begin{cases} 1, & \text{if "yes"} \\ 0, & \text{otherwise} \end{cases}$$

$X \sim \text{Bernoulli}(p)$, $p = P[X=1]$ — approval rating.

$H_0: p = 0.3$

$H_1: p < 0.3$

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.

Key idea: Accept H_0 if the data are consistent with what we expect when the H_0 is true. Otherwise, reject H_0.

Two types of errors:

based on data

	Truth	
Test outcome	H_0 is true	H_1 is true
Accept H_0	No error	type II error
Reject H_0	type I error	No error

⇓ Accept H_1

- 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.*

$\boxed{\text{level-}\alpha \text{ test}}$

$P[\text{Type I error}] = P[\text{Reject } H_0 \mid H_0 \text{ is true}] \le \alpha$

- 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.
- In practice, $\alpha = 0.01, 0.05$ (most popular), or 0.10.
- 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.

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

Court's decisions:

Rejects H_0 ⇒ suspect is pronounced guilty = convict the suspect

Accepts H_0 ⇒ suspect is pronounced innocent = release the suspect

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: Reject H_0 if H_0 is true ⇒ convicting an innocent suspect

Type II error: Accept H_0 if H_1 is true ⇒ releasing a guilty suspect

Basic premise: $\underbrace{\text{convicting an innocent}}_{\text{type I error}}$ is a more serious error than $\underbrace{\text{releasing a guilty}}_{\text{type II error}}$

Q: How can we make $P(\text{type I error}) = 0$? What happens to $P(\text{type II error})$?

never convict anyone $\Rightarrow P[\text{type I error}] = 0$

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

Q: How can we make $P(\text{type II error}) = 0$? What happens to $P(\text{type I error})$?

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

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

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

Issue: How to construct a level-α test? i.e., $P[\text{Type I error}] \le \alpha$

- 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:

Expect T to be small (i.e., close to zero) if H_0 is true. $\leftarrow$ $T = \dfrac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$ $\theta_0 \to$ null value

[Recall: $T = \dfrac{\hat{\theta} - \theta}{se(\hat{\theta})}$] pivot

- Find the **null distribution** — the distribution of T assuming H_0 is true.

Test statistic $T \to$ a random quantity.
$T_{obs} \to$ obs. value of T — a real #.

- 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., $P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$.

$P[\text{Type I error}] = \alpha \Rightarrow$ level-α test.

Some common rejection regions

Suppose $T = \dfrac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$. When H_0 is true, we expect T to be close to zero.

In this case, it is often easy to guess $\mathcal{R}$.

Case 1: $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$

Reject H_0 if: $|T|$ large $\Rightarrow$ $|T| > c$, where c is some positive cutoff.

Case 2: $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$

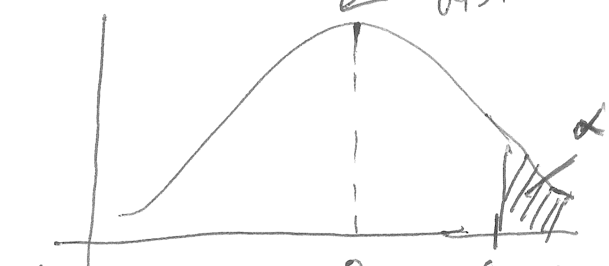
Case 3: $H_0 : \theta = \theta_0$ against $H_1 : \theta < \theta_0$

Compute the critical point in a way that ensures that the level of the test equals the prescribed α.

The corresponding level α tests:

Suppose c_α is such that $P(T > c_\alpha | \theta = \theta_0) = \alpha$.

Suppose:



Recall: $T = \dfrac{\hat{\theta} - \theta_0}{\hat{SE}(\hat{\theta})}$ $\Rightarrow$ Intuitively: reject H_0 if $|T| >$ cutoff. positive

Case 1: $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$

$\mathcal{R} = \{|T| > c_{\alpha/2}\}$, i.e., $\boxed{\text{reject } H_0 \text{ when } |T| > c_{\alpha/2},}$ otherwise accept it.

$P[\,|T| > c_{\alpha/2} \mid H_0 \text{ is true}\,] = P[\text{TYPE I error}] = \alpha$

— see next page

Case 2: $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$

$\mathcal{R} = \{T > c_\alpha\}$, i.e., reject H_0 when $T > c_\alpha$, otherwise accept it.

Case 3: $H_0 : \theta = \theta_0$ against $H_1 : \theta < \theta_0$

$\mathcal{R} = \{T < -c_\alpha\}$, i.e., reject H_0 when $T < -c_\alpha$, otherwise accept it.

Note: $P\left[\,|T| > c_{\alpha/2} \mid H_0 \text{ is true}\right]$

$= P\left[\,T > c_{\alpha/2} \text{ or } T < -c_{\alpha/2} \mid H_0 \text{ is true}\right]$

$= P\left[\,T > c_{\alpha/2} \mid H_0 \text{ is true}\right] + P\left[\,T < -c_{\alpha/2} \mid H_0 \text{ is true}\right]$

$= \alpha/2 + \alpha/2 = \alpha.$