Hypothesis testing (part 2)

Lecture 15b (STAT 24400 F24)

Hypothesis testing - general framework

A general framework:

- **Test statistic** = the function of the data that we'll use to make our decision (i.e., choose H_0 or H_1)

  e.g., test statistic $\bar{X}$ is a function of the data $X_1, \cdots, X_n$

- **Null distribution** = distrib. of the test statistic, assuming H_0 is true
- **Rejection region** = values of the test statistic for which we reject H_0

Hypothesis testing (Poisson)

Example — $X \sim \text{Poisson}(\lambda)$:

- Testing $H_0 : \lambda = 100$ vs $H_1 : \lambda \in \{110, 120, 130\}$
- If X is in the range 84–117 then choose H_0, otherwise choose H_1

In the terminology of the general framework:

- Test statistic $= X$
- Null distribution $= \text{Poisson}(100)$
- Rejection region $= \{0, 1, \ldots, 83\} \cup \{118, 119, \ldots\}$

Hypothesis testing (Normal LRT)

Example — $X \sim \text{N}(\mu, \sigma^2)$:

- Testing $H_0 : \mu = 1,\ \sigma^2 = 1$ vs $H_1 : \mu = 2,\ \sigma^2 = 2$
- LR test: if LR > 1.3 then choose H_0, otherwise choose H_1

In the terminology of the general framework:

- Test statistic $= \text{LR} = \sqrt{2e} \cdot e^{-X^2/4}$
- Null distribution $=$ [distrib. of $\sqrt{2e} \cdot e^{-X^2/4}$ when $X \sim \text{N}(1, 1)$]
- Rejection region $= [0, 1.3]$

Hypothesis testing (Normal, non-LR test)

Example — $X \sim \mathrm{N}(\mu, \sigma^2)$:

- Testing $H_0 : \mu = 1,\ \sigma^2 = 1$ vs $H_1 : \mu = 2,\ \sigma^2 = 2$
- A different test: if $X \leq 1.5122$ then choose H_0, otherwise choose H_1

In the terminology of the general framework:

- Test statistic $= X$
- Null distribution $= \mathrm{N}(1,1)$
- Rejection region $= (1.5122, \infty)$

Test statistics & rejection regions

In general, we want to choose the test statistic to satisfy:

- It must be a function of the data (can't depend on unknown parameters).
- Its distribution under H_0 (i.e., the null distribution) should be appropriate or easy to calculate.
- Its typical values should be as different as possible under H_0 vs H_1, so that it can help us accurately distinguish between H_0 & H_1.

Common rejection regions

Common types of rejection regions: "one sided" and "two sided"

- (c, ∞) or $[c, \infty)$ for some threshold $c \in \mathbb{R}$
- $(-\infty, c)$ or $(-\infty, c]$ for some threshold $c \in \mathbb{R}$
- $(-\infty, -c) \cup (c, \infty)$ or $(-\infty, -c] \cup [c, \infty)$ for some threshold $c > 0$
- $(-\infty, c_1) \cup (c_2, \infty)$ or $(-\infty, c_1] \cup [c_2, \infty)$ for some thresholds $c_1 < c_2$

E.g., for i.i.d. data from a distrib. with mean μ, using test statistic $\bar{X}$:

- $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0 \rightsquigarrow$ use rej. region (c, ∞) for some $c > \mu_0$
- $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0 \rightsquigarrow (-\infty, c)$ for some $c < \mu_0$
- $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0 \rightsquigarrow (-\infty, -c) \cup (c, \infty)$ for some $c > 0$
- $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0 \rightsquigarrow (-\infty, c_1) \cup (c_2, \infty)$ for $c_1 < \mu_0 < c_2$

Test level and rejection regions

Defining the rejection region:
If we have a fixed level α, we choose the rejection region to satisfy

$$\mathbb{P}_{H_0}(\text{test statistic} \in \text{rejection region}) = \alpha$$

this means probability under the null distrib.

for a discrete distrib., may need to be $< \alpha$

E.g.,

- For rejection region of the form (c, ∞), find c to satisfy
$$\mathbb{P}_{H_0}(\text{test statistic} > c) = \alpha$$
- For rejection region of the form $(-\infty, -c) \cup (c, \infty)$, find c to satisfy
$$\mathbb{P}_{H_0}(\text{test statistic} > c \ \text{ or } \ < -c) = \mathbb{P}(|\text{test statistic}| > c) = \alpha$$
- For rejection region of the form $(-\infty, c_1] \cup [c_2, \infty)$, find c_1, c_2 to satisfy
$$\mathbb{P}_{H_0}(\text{test statistic} \leq c_1) = \frac{\alpha}{2} \ \text{ and } \ \mathbb{P}_{H_0}(\text{test statistic} \geq c_2) = \frac{\alpha}{2}$$

Hypothesis tests: overview

General recipe for a hypothesis test:

- Design the test:
  - Choose a test statistic (a function of the data) & rejection region
  - Compute $\alpha = \mathbb{P}_{H_0}(\text{test statistic} \in \text{rejection region})$
  - If applicable... compute $\beta = \mathbb{P}_{H_1}(\text{test statistic} \notin \text{rej. region})$
- Run the test:
  - Observe the data, compute the test statistic, & choose H_0 or H_1

$\rightsquigarrow$ By construction, the test has the property

$$\mathbb{P}_{H_0}(\text{reject } H_0) = \alpha$$

(and if applicable, the test may satisfy $\mathbb{P}_{H_1}(\text{fail to reject } H_0) = \beta$)

Example: Exponential (two-sided)

Suppose we have data $X \sim \text{Exponential}(\lambda)$,
and we want to test $H_0 : \lambda = 20$ vs. $H_1 : \lambda \neq 20$ at level $\alpha = 0.1$

- Choose the test statistic: X
- Choose the rejection region:

$$0.05 = \mathbb{P}_{H_0}(X < c_1) = 1 - e^{-20 \cdot c_1} \;\rightsquigarrow\; c_1 = 0.0026$$

$$0.05 = \mathbb{P}_{H_0}(X > c_2) = e^{-20 \cdot c_2} \;\rightsquigarrow\; c_2 = 0.1498$$

$\Rightarrow$ if $0.0026 \leq X \leq 0.1498$ then do not reject H_0, otherwise reject H_0

- We may want to calculate power against some alternatives, e.g.,
  - If $\lambda = 10$, $\mathbb{P}(0.0026 \leq X \leq 0.1498) = 0.751 \Rightarrow$ power $= 0.249$
  - If $\lambda = 1$, $\mathbb{P}(0.0026 \leq X \leq 0.1498) = 0.137 \Rightarrow$ power $= 0.863$

Example: Exponential (one-sided)

Suppose we have data $X \sim \text{Exponential}(\lambda)$,
and we want to test $H_0 : \lambda = 20$ vs. $H_1 : \lambda < 20$ at level $\alpha = 0.1$

- Choose the test statistic: X
- Choose the rejection region:

$$0.1 = \mathbb{P}_{H_0}(X > c) = e^{-20 \cdot c} \;\rightsquigarrow\; c = 0.1151$$

$\Rightarrow$ if $X \leq 0.1151$ then do not reject H_0, otherwise reject H_0

- We may want to calculate power against some alternatives, e.g.,
  - If $\lambda = 10$, $\mathbb{P}(X \leq 0.1151) = 0.684 \Rightarrow$ power $= 0.316$
  - If $\lambda = 1$, $\mathbb{P}(X \leq 0.1151) = 0.109 \Rightarrow$ power $= 0.891$

Remarks: the increase of the power for one-sided test and the trade-off.

Multiple testing concerns

Caution—the type of test (two sided / one sided / which side?)
must be chosen **before** seeing the data.

Suppose $X \sim \text{N}(\mu, 1)$, & we want to test if $\mu = 0$ at level $\alpha = 0.05$.
We observe X and then choose the test:

- If we observe a positive value,
  we'll choose to test $H_0 : \mu = 0$ vs $H_1 : \mu > 0$
  $\rightsquigarrow 1 - \Phi(1.65) = 0.05$, so we use rejection region $= (1.65, \infty)$
- If we observe a negative value,
  we'll choose to test $H_0 : \mu = 0$ vs $H_1 : \mu < 0$
  $\rightsquigarrow \Phi(-1.65) = 0.05$, so we use rejection region $= (-\infty, -1.65)$

Overall, we will reject $H_0 : \mu = 0$ if *either* $X > 1.65$ or $X < -1.65$
$\Rightarrow$ Type I error $= 0.1$ (twice as large as the desired level 0.05)
This is not valid!

Understanding p-values

"p-value" is a way to quantify how strong the evidence from data is against H_0.

Once we've picked the type of rejection region
(e.g., (c,∞) or $(-\infty,c)$ or $(-\infty,-c)\cup(c,\infty)$ or ...),
the rejection regions for different test levels α's are nested.

E.g., for rejection regions of the form (c,∞),
reduce $\alpha \Leftrightarrow$ increase c
$\rightsquigarrow$ (rejection region at a smaller α) $\subseteq$ (rejection region at a larger α)

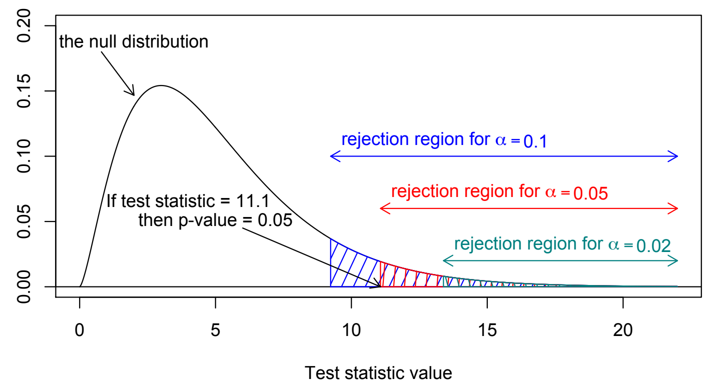
For any value t of the test statistic, can calculate

- For which values of α would t lie in the rejection region? larger α's
- For which values of α would t lie outside the rejection region? smaller α's

The **p-value** is the value of α where t switches
from being inside the rejection region to being outside the rejection region.

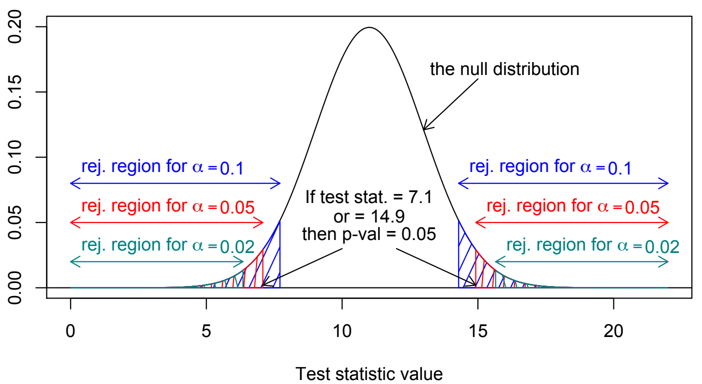
Illustrate p-value: one-sided H_a

For a one-sided alternative:



Illustrate p-value: two-sided H_a

For a two-sided alternative:



Pitfalls about p-values

Another common interpretation of the p-value:
quantifying the probability of the test statistic being
at least as extreme as what we observed, given that H_0 is true.

A common misinterpretation is to think the p-value
is the prob. that H_0 is true, given the observed data

- In a frequentist setting, H_0 being true is not random
- In a Bayesian setting, $\mathbb{P}(\text{data} \mid H_0) \neq \mathbb{P}(H_0 \mid \text{data})$.

Example of p-value: Exponential

Data $X \sim \text{Exponential}(\lambda)$

Testing $H_0 : \lambda = 20$ vs $H_1 : \lambda < 20$

- What's the value of α for which x is at the boundary of the rej. region?

$$\alpha = \mathbb{P}_{H_0}(X > x) = e^{-20x}$$

$\rightsquigarrow$ after observing data X, the p-value is e^{-20X}.

For example, if we observe $x = 0.2$, then the p-value is $e^{-20\cdot 0.2} = 0.018$.

- If (pre-specified) $\alpha > 0.018$ (e.g. $\alpha = 0.05$), then $x = 0.018 \in$ (rej. region).
- If (pre-specified) $\alpha < 0.018$ (e.g. $\alpha = 0.01$), then $x = 0.018 \notin$ (rej. region).
- If (pre-specified) $\alpha = 0.018$, then $x = 0.018$ is at the boundary of the rej. region.

Hypothesis testing vs. confidence intervals

The connection between hypothesis testing / p-values / conf. intervals...

For a two-sided test / confidence interval with a pre-specified test level α, these outcomes are equivalent:

- If we test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, we can reject at the level α
- If we test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, the p-value is $\leq \alpha$
- If we construct a $(1 - \alpha)$ C.I. for θ, then θ_0 lies outside the C.I.

(The same is true for one-sided tests / one-sided conf. int.)