Hypothesis testing

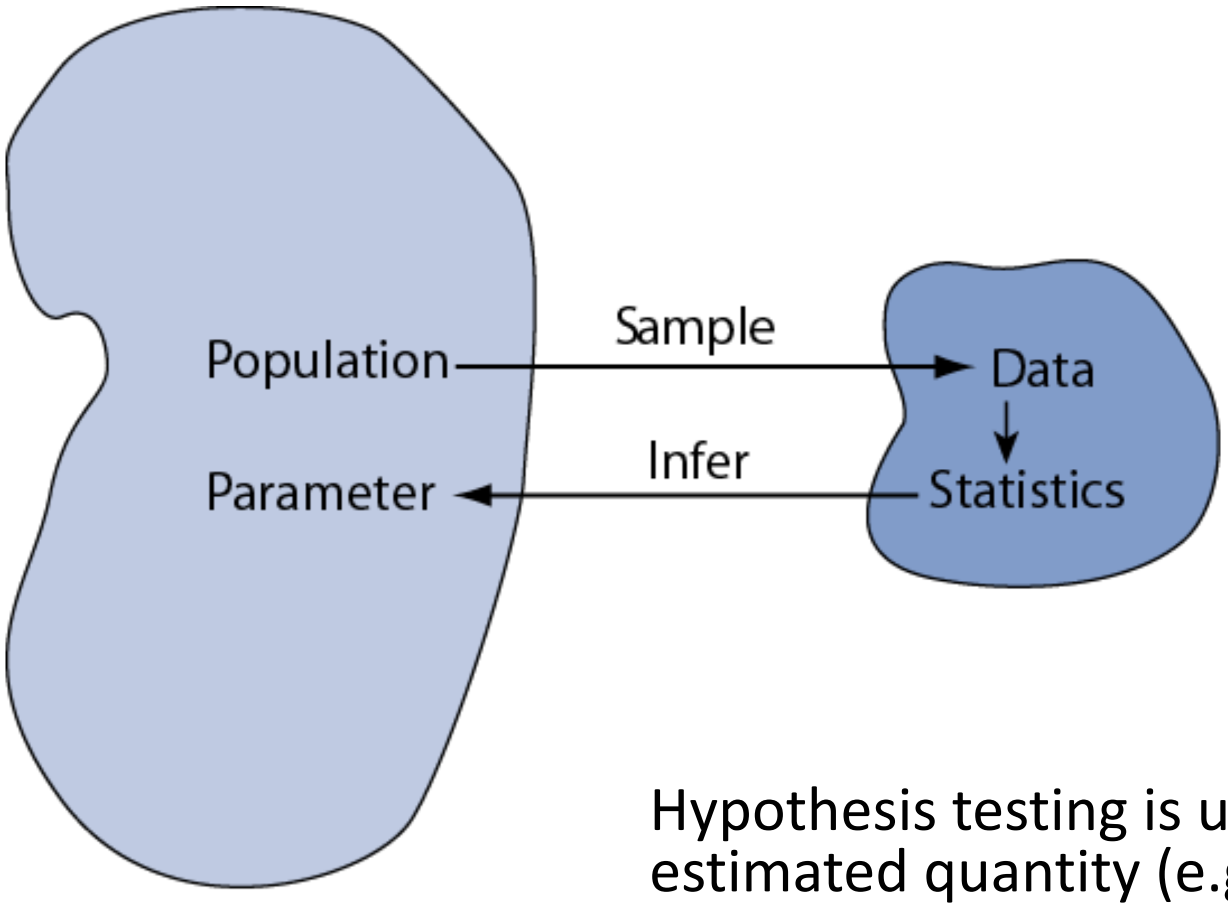
Hyeri You

Graduate student

Dep of Epi & Biostat

Contents

Why do we perform Hypothesis Testing?



Hypothesis testing is used to draw a conclusion about how an estimated quantity (e.g. mean) relates to the true population value (under certain conditions).

Hypothesis Testing Steps

Step 1. Define your hypotheses (null, alternative) and significant level(alpha)

Step 2. Specify your null distribution and calculate statistic

Step 3. Calculate the p-value of what you observed

Step 4. Interpretation (Reject or fail to reject (~accept) the null hypothesis)

Step1. Null/alternative hypothesis

- Convert the research question to null and alternative hypotheses

- EX1) hypothesis test of one mean.
- Two sided test:
  - Null hypothesis: H_0: $\mu=\mu_0$
  - Alternative hypothesis: H_A: $\mu\neq\mu_0$
- One sided test:
  – Null hypothesis: H_0: $\mu\geq\mu_0$
  – Alternative hypothesis: H_A: $\mu<\mu_0$

or

  – Null hypothesis: H_0: $\mu\leq\mu_0$
  – Alternative hypothesis: H_A: $\mu>\mu_0$

- EX2) hypothesis test of one proportion.
- Two sided test:
  - Null hypothesis: H_0: $p = p_0$
  - Alternative hypothesis: H_A: $p \neq p_0$
- One sided test:
  – Null hypothesis: H_0: $p \geq p_0$
  – Alternative hypothesis: H_A: $p < p_0$

or

  – Null hypothesis: H_0: H_0: $p \leq p_0$
  – Alternative hypothesis: H_A: $p > p_0$

Step 2. Null distribution and Test Statistic

- Ex1) Hypothesis test of a mean
- Goal: to test whether a mean is equal to some hypothesized value

✓If n is large enough, the distribution of the sample mean $\bar{x}$ is normal by the CLT.

$$z_{stat} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

✓If σ is not known, and n is not large enough, the distribution of the sample mean is student-T.

$$t_{stat} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Step 2. Null distribution and Test Statistic

- Ex2) hypothesis test of a proportion
- Goal: to test whether a sample proportion is equal to some hypothesized value $p_{0.}$

✓ If n is large enough, the distribution of estimated proportion p̂ is approximately a normal by the CLT. (mean: p, SD= √(p(1-p)/n))

$$z_{stat} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

❖ straightforwardly, the distribution of proportion p̂ is a binomial (exact test). (mean: np, SD= √(np(1-p))

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Step 3. P-value

- The *P*-value answer the question: What is the probability of the observed test statistic or more extreme **when H_0 *is true*?**
- This corresponds to the Area Under the Curve(AUC) in the tail of the Standard Normal distribution beyond the $z_{stat.}$
- Convert *z* statistics to *P*-value :

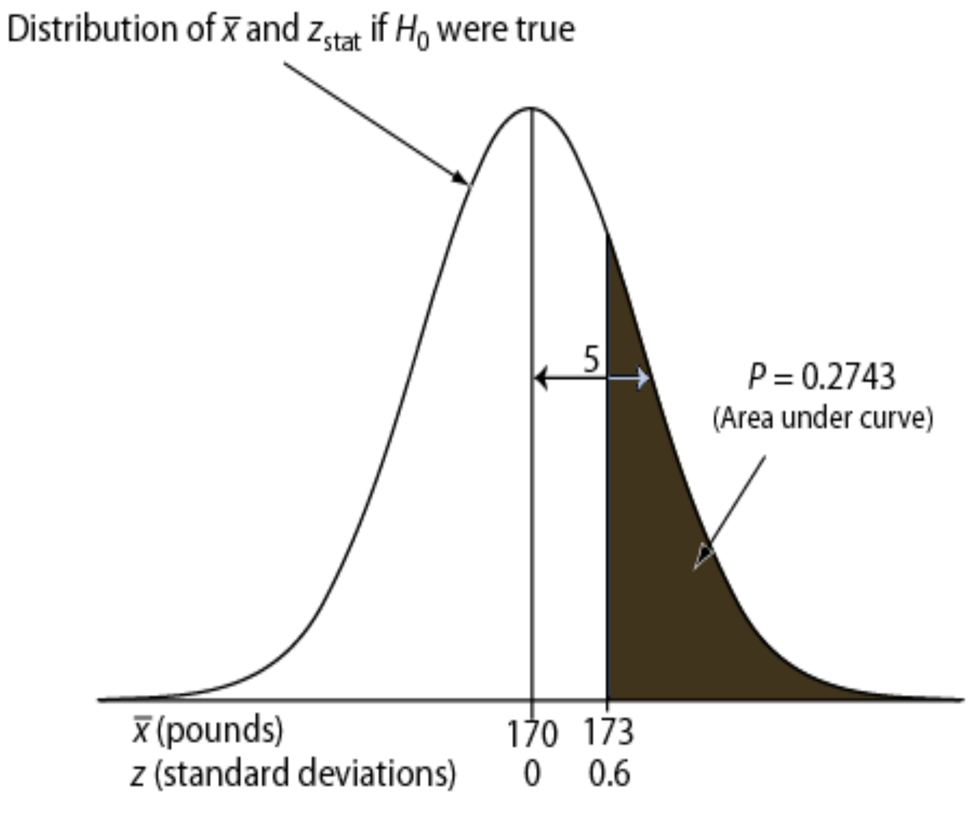
  For H_a: $\mu > \mu_0 \Rightarrow P = \Pr(Z > z_{stat})$ = right-tail beyond z_{stat}

  For H_a: $\mu < \mu_0 \Rightarrow P = \Pr(Z < z_{stat})$ = left tail beyond z_{stat}

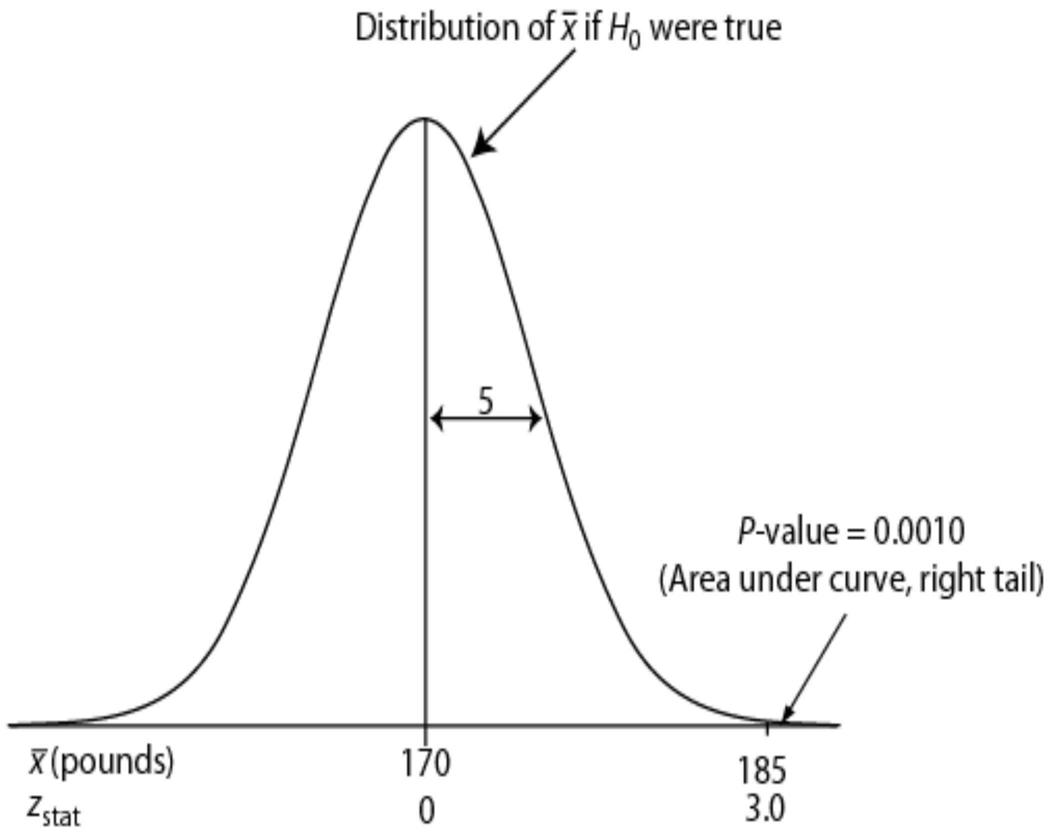
  For H_a: $\mu \neq \mu_0 \Rightarrow P = 2 \times$ one-tailed *P*-value
- Use p-value table or software to find these probabilities.

Step 3. P-value_One-sided *P*-value

For Z(stat) of 0.6



For Z(stat) of 3.0

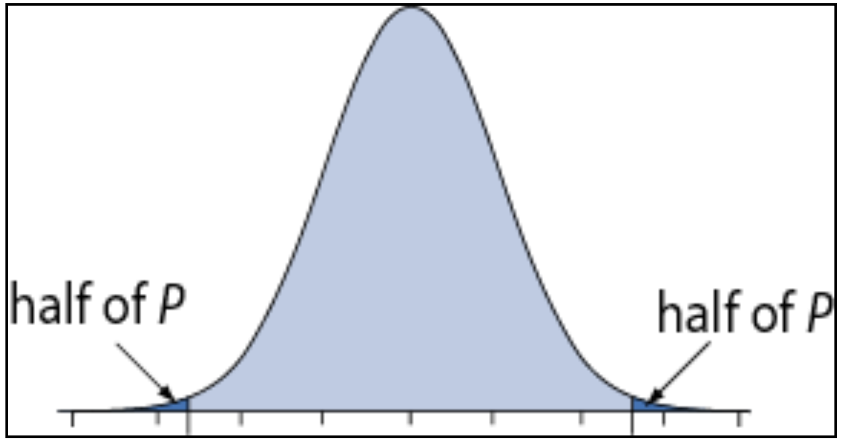


Step 3. P-value_Two sided p-value

- One-sided $H_a \Rightarrow$ Area under curve in tail beyond z_{stat}
- Two-sided $H_a \Rightarrow$ consider potential deviations in both directions $\Rightarrow$ double the one-sided P-value

Ex)
If one-sided $P = 0.0010$,
then two-sided $P = 2 \times 0.0010 = 0.0020$.

If one-sided $P = 0.2743$,
then two-sided $P = 2 \times 0.2743 = 0.5486$.

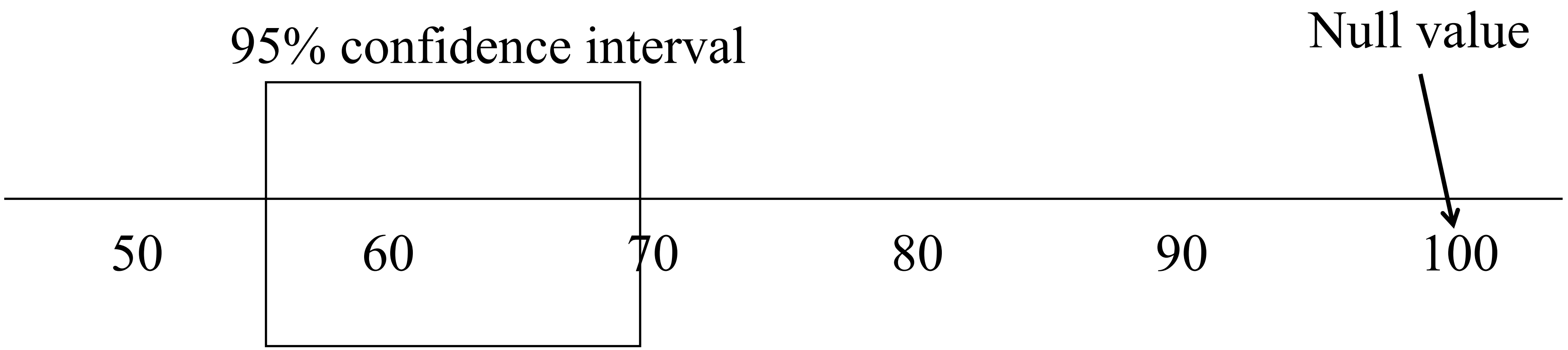


Step 4. Interpretation

- Let significance level (α) ≡ probability of erroneously rejecting H_0
- Set significance level threshold (e.g., let $\alpha = .10$, .05, *or whatever*)
- Reject H_0 when $P \leq \alpha$, since smaller P-values provide stronger evidence against H_0
- Retain H_0 when $P > \alpha$

Confidence Interval vs. P-value

- Confidence intervals give the same information (and more) than hypothesis tests



Null hypothesis: Average vitamin D is 100nmol/L
Alternative hypothesis: Average vitamin D is not 100nmol/L
P-value<0.05

Types of Error

Decision	Truth: H_0 true	Truth: H_0 false
Retain H_0	Correct retention(1-α)	Type II error(β)
Reject H_0	Type I error(α)	Correct rejection(1-β)

Type I error = erroneous rejection of true H_0

Type II error = erroneous retention of false H_0

Power

- $\beta \equiv$ probability of a Type II error

  β = Pr(retain H_0 | H_0 false)

- $1 - \beta =$ "Power" $\equiv$ probability of avoiding a Type II error

  $1 - \beta$ = Pr(reject H0 | H0 false)

- You can construct a power curve by plotting "power vs. different alternative hypotheses" or "power vs. different sample sizes"
  - ✓The curve with power vs. different sample sizes will allow you to see what gains you might have versus cost of adding participants

Power _Characteristics

- The power of a statistical test is lower for alternative values that are closer to the null value.
- The power of a statistical test can also be increased by increasing n
- For a fixed n, the power of a statistical test can be increased by increasing α, the probability of a Type I error
- The balance between type I error and type II error (or power) should depend on the context of the study
- When setting up a study, most investigators make sure that there will be at least 80% power.

Thank you!