

HYPOTHESIS FORMULATION AND TESTING

1. Hypothesis:

- A statement about one or more population
- Research question
 - Statement
 - Research hypothesis
- Postulating the existence of:
 - A difference between groups
 - An association among factors
- Usually derived from a hunch, an educated guess based on published results or preliminary observations.
- There are 2 type:
 - Null hypothesis (H_0)
 - Hypothesis of no difference
 - Hypothesis to tested
 - Alternative hypothesis (H_A)
 - The hypothesis that postulates that there is a treatment effect, an association factors or a difference between groups.
- Inferential statistics – estimating the probability that a given outcome is due to chance
- If the sample data provide sufficient evidence to discredit $H_0 \rightarrow$ reject H_0 in favor of H_A .

2. Hypothesis Testing:

- To aid the researcher in reaching a decision concerning a population y examining the sample.
- Observed differences or associations may have occurred by chance.

- H_0 : the proportion of patients with disease who die after treatment with the new drug is not different from the proportion of similar patients who die after treatment with placebo.
- H_A : the proportion of patients with disease who die after treatment with the new drug is lower than the proportion of similar patients who die after treatment with placebo.

Statistical Decision based on a sample	In the population	
	H_0 true	H_0 false
Do not reject H_0	Correct (Confidence Limit) $1 - \alpha$ (level of certainty of our statistical data hold true)	Type II error Beta (β) (probability of wrongly not rejecting H_0 when the H_0 is false)
	Type I error (level of significance) Alpha (α) (probability of wrongly rejecting the H_0 when the H_0 is true)	Correct (Power of study) $1 - \beta$ (probability that the H_0 correctly rejected)
Reject H_0		

- Type I error (α)
 - The probability of **wrongly rejecting** the null hypothesis when the null hypothesis is **true**.
- Type II error (β)
 - The probability **wrongly not rejecting** the null hypothesis when the null hypothesis is **false**.
- The test statistics
 - A value with a known distribution when the null hypothesis is true.
- Normal distribution (refer to descriptive statistics note)

- Level of significant (α)
 - The null hypothesis is rejected if the probability of obtaining a value as extreme or more extreme than that observed in the sample is small when the null hypothesis is true.
 - “Small” is usually taken to less than or equal to 5%
 - If the 2 tail test is taken then the α must be divided by 2
 - E.g. $\alpha = 0.05$, p value = 0.05, (2 tail test taken). Significant level $\alpha = 0.05/2 = 0.025$. Conclusion does not reject the H_0 .
- The p value
 - The probability of obtaining a value as extreme or more extreme than that observed in the sample given that the null hypothesis is true is called p value
 - The smallest value of α for which the null hypothesis can be rejected
 - The p value is compared to the predetermined significance level α (usually 0.05) to decide whether the null hypothesis should be rejected
 - If p value less than α , reject the H_0 .
 - If p value greater than α , do not reject the H_0 .

3. Steps in hypothesis testing

- Step 1
 - Generate the null hypothesis and alternative hypothesis
 - $H_0 : ??$
 - $H_A : ??$
 - What are the characteristics of interest?
 - E.g. mean, proportion
 - 1-tail (one sided) or 2-tail (both sided)
 - E.g. 1-tail research hypothesis

- The proportion of patients with disease after treatment with new drug is **lower** than the proportion of similar patients who die after treatment with placebo
- E.g. 2-tail research hypothesis
 - The mean blood pressure of patients in the new treatment group is **not different** from the mean blood pressure of patients in the old treatment
- E.g. research questions:
 - Effectiveness of a new antihypertensive drug
 - H_0 : the mean blood pressure of patients in the new treatment group is **not different** from the mean blood pressure of patients in the old treatment ($\mu_1 = \mu_2$)
 - H_A : the mean blood pressure of patients in the new treatment group is **different** from the mean blood pressure of patients in the old treatment ($\mu_1 \neq \mu_2$)

Notes:

1. in population

- μ = mean

- σ = standard deviation

Notes:

2. in sample

- \bar{x} = mean

- SD = standard deviation

- Step 2:
 - Set the significance level
 - Usually set at 0.05, 0.01, 0.1
- Step 3:
 - Decide which statistical test to use and check the assumption of the test
 - Population is approximately normally distributed
 - Data values are obtained by independent random sampling
 - Adequate sample size
 - To decide which statistical test should be used

- E.g. mean, proportion
- Assumption must be adequately met
- If not met alternative procedures can be used
 - E.g. non parametric test would be used when the data is seriously non-normal)
- Step 4:
 - Compute the test statistic and associated p value
 - Calculate appropriate test statistics
- Step 5:
 - Interpretation
 - Compare p value with the level of significance
 - Decide whether or not to reject the null hypothesis
 - $p \text{ value} < \alpha$ – reject the null hypothesis
 - $p \text{ value} > \alpha$ – do not reject the null hypothesis
- Step 6:
 - Draw conclusions
 - Conclude accordingly based on rejecting/not rejecting null hypothesis
 - Decision rule
 - Rejection region
 - To reject the null hypothesis if the value of the test statistic that computed from the sample is one of the values in the rejection region
 - Acceptance region
 - To accept the null hypothesis of the computed values in the acceptance region
 - E.g. conclusion:
 - The mean blood pressure of patients in new treatment group is different from the mean blood pressure of patient in old treatment