

# HYPOTHESIS TESTING • MTH107

## Class R FAQ

by Derek H. Ogle, revised Dec-17

### 1-Sample Z-Test

$H_0: \mu = \mu_0$  (where  $\mu_0 =$  specific value)

**Statistic:**  $\bar{X}$     **Test Statistic:**  $Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$     **Conf. Region:**  $\bar{X} + Z^* \frac{\sigma}{\sqrt{n}}$

**Assumptions:** 1)  $\sigma$  is known  
2)  $n \geq 30$ ,  $n \geq 15$  and popn not strongly skewed, OR popn is normal

**R:** `z.test()`

### 1-Sample t-Test

$H_0: \mu = \mu_0$  (where  $\mu_0 =$  specific value)

**Statistic:**  $\bar{X}$     **Test Statistic:**  $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$     **Conf. Region:**  $\bar{X} + t^* \frac{s}{\sqrt{n}}$     **df:**  $n-1$

**Assumptions:** 1)  $\sigma$  is UNknown,  
2)  $n \geq 40$ ,  $n \geq 15$  & histogram not strongly skewed, OR histogram is normal

**R:** `t.test()`, `hist()`

### 2-Sample t-Test

$H_0: \mu_1 = \mu_2$     **Statistic:**  $\bar{X}_1 - \bar{X}_2$

**Test Statistic:**  $t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$  where  $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

**Conf. Region:**  $(\bar{X}_1 - \bar{X}_2) + t^* \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$     **df:**  $n_1 + n_2 - 2$

**Assumptions:**  
1) Individuals in groups are independent  
2) Group variances are equal (use Levene's Test)  
3)  $n_1 + n_2 \geq 40$ ,  $n_1 + n_2 \geq 15$  and both histograms are not strongly skewed, OR both histograms are normal

**R:** `t.test()`, `levenesTest()`, `hist()`

## Choosing a Hypothesis Test

1. If response variable is QUANTITATIVE, GOTO 2; if CATEGORICAL, GOTO 5.

### Quantitative Response

2. If 1 group, GOTO 3; if 2 or more groups, GOTO 4.
3. If  $\sigma$  is KNOWN, then **1-Sample Z**; If  $\sigma$  is UNKNOWN, then **1-Sample t**.
4. If individuals are INDEPENDENT between groups, then **2-Sample t**;  
otherwise, **Paired t**.

### Categorical Response

5. If 1 group, then **Goodness-of-Fit**; if 2 or more groups, then **Chi-Square**.

## Making a Decision about $H_0$

If p-value  $< \alpha$ , then **REJECT  $H_0$** , otherwise **DNR  $H_0$** .

## Chi-Square Test

$H_0$ : "Distribution of individuals in response levels is the same for all groups"

$H_A$ : "Distribution of individuals in response levels is NOT the same for all groups"

**Statistic:** Observed frequency table

**Test Statistic:**  $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$     **df:** (rows-1)(columns-1)

**Assumptions:**  $\geq 5$  in each cell of the EXPECTED table

**R:** `xtabs()`, `matrix()`, `chisq.test()`, `percTable()`

## Goodness-of-Fit Test

$H_0$ : "Distribution of individuals in response levels follows the theoretical distribution"

$H_A$ : "Distribution of individuals in response levels does NOT follow the theoretical distribution"

**Statistic:** Observed frequency table

**Test Statistic:**  $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$     **df:** cells-1

**Assumptions:**  $\geq 5$  in each cell of the EXPECTED table

**R:** `xtabs()`, `c()`, `chisq.test()`, `percTable()`, `chiGOF()`

# 11 STEPS FOR ANY HYPOTHESIS TEST

- 1) State the rejection criterion ( $\alpha$ )
- 2) State the null & alternative hypotheses and define the parameter(s)
- 3) Determine which test to perform – Explain!
- 4) Collect the data (address type of study and randomization)
- 5) Check all necessary assumption(s)
- 6) Calculate the appropriate statistic(s)
- 7) Calculate the appropriate test statistic
- 8) Calculate the p-value
- 9) State your rejection decision
- 10) Summarize your findings in terms of the problem
- 11) If rejected  $H_0$ , compute a **100(1- $\alpha$ )%** *confidence region* for parameter