Hypothesis Testing

```
key words: types of errors, H0 and H1, p-value, difference in means,
differences in variances, mean differences, population means, population
variance, correlation, independent, Adjusted Significance
```

1. Hypothesis Tests and Types of Errors

- Hypothesis is stated with population parameters, testing whether a hypothesis is reasonable statement and shouldn't be rejected or not.

- Procedure:

    1. state the hypothesis(one/two-tailed, H0 & H1)
    2. select the test statistics
    3. specify the significant level
    4. state the decision rule
    5. collect samples ,calculate the sample statistics and critical value
    6. Comparison and make a decision

- Null and Alternative Hypothesis

    - $H0$:
        - researchers want to **reject**
        - actually tested as basis for selecting the test statistics
        - 通常叙述中带等号
    - $H1$:
        - if there is sufficient evidence to reject $H0$

        - The hypothesis the researcher want to **access**
        - 不能带=

- One-tailed VS Two-tailed Test
    - Mainly depend on the testing purpose deviation on one or two side of hypothesized value
    - Critical value(or Rejection area):
        - Reject $H0$ if (*Decision Rule*)
            - test statistics > upper critical value OR
            - test statistics < lower critical value
                - critical value=$\pm z_{\alpha/2}$, for two-tailed test with z-score
                - critical value=$\pm z_{\alpha}$, for one-tailed test with z-score with interests in upper-tail(+) or lower-tail(-) tests respectively

- Test Statistics, Type I and Type II Error, Significant Level, Power
    - Test Statistics:
        - calculated from <u>sampled data</u>, while critical value gained from test statistics distribution and significant level
        - The comparison between **test statistics and critical value** is important step of decision making.
        - 🌟Calculation: $test\ statistics = \frac{sample\ statistic - hypothesized\ value}{standard\ error\ of\ the\ sample\ statistics}$
            - sample statistics: point estimation of population parameter
            - standard error: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt[2]{n}}$(known) or $s_{\bar{x}} = \frac{s}{\sqrt[2]{n}}$(unknown)

- TypeI and TypeII Error
    - TypeI Error: True H_0 is rejected, α, false positive
    - TypeII Error: Not reject the false H_0, β
    - Relationship:
        - sample size(n) and Significant level(α) together determine the TypeII Error.
        - $\alpha\uparrow, \beta\downarrow, 1-\beta\uparrow, n\uparrow$

- Power of Test
    - Definition: correctly rejecting false H_0, $1-\beta$
    - Usage: when multiple test statistics, power decides which to use. Test statistic with the **highest** power is selected.

- Decision Rule Explanation

- Decision Rule: whether reject H_0 or fail to reject H_0 based on tbe test statistic distribution and significant level.
- Confidence Intervals & Hypothesis Test:🌟
  - are linked with ***critical value***
  - {[sample statistic- critical value * standard error]} ≤ population parameter ≤ {[sample statistic+ critical value * standard error]}
  - -critical value≤test statistics≤critical value

- Statistical Significant
  - Not means always economical significance.

2. P-Values and Tests of Means

- p-value:
  - Definition:
    - the probability of getting a test statistic that will reject H_0
    - the **smallest** level of significance for H_0 can be rejected.
    - 🌟Position: the **area** from test statistic to (negative) infinity tail, which can be compared with significant level α. If < α, reject H_0.
      - OR compare the **critical value** with **test statistic** to make decision. If |critical value|<|test statistic|, reject H_0.

- Significant of Test among Multiple Tests
  - Normally, we reject H_0 when there are more than 5 false positives with 100 tests, with $\alpha = 0.5$. For multiple tests:
  - Process
    - rank p-values in ascending order which are < pre-set α
    - calculate adj. significance for each test
      - Formula: Adjusted Significance= α * (Rank of p-value/Numbers of Tests)
    - compare the calculated results and reported p-values
    - if ***adjusted significance ≥ reported p-value***(the ranking values), treat those tests as <u>actual rejections</u>.

- Tests of Means

- t-Test
  - test statistics: $t_{n-1} = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt[2]{n}}}$
  - Usage Condition:
    - Unknown population variance
    - $n < 30$, normal distribution
    - $n \geq 30$, with any types of distribution(CLT)
- z-Test
  - test statistics:
    - $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt[2]{n}}}$
      - Usage Condition: normally distributed with known variance
    - $z = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt[2]{n}}}$
      - **very large** sample size, unknown population variance
      - in this case, use t-test is better.

3. Mean Differences and Difference in Means

- Two Population Means Test

Difference in Mean

- Condition
  - respectively in normally distributed populations **iid**
  - unknown but equal variance
  - test whether these two means are equal
- T-statistic:
  - $t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt[2]{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}}$
    - $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
    - total degree of freedom= $n_1 + n_2 - 2$
  - Hypothesis:
    - Normally state as $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 = 0$, and we need to declare the specific representations of μ_1 and μ_2

Mean Difference(Paired Comparison)

- test the hypothesis that the **mean of differences in the pairs** are 0.($\mu_{dz} = 0$)
- Usage condition:
  - samples are dependent, depending on some other factor.
  - normally distributed samples
- Hypothesis:
  - $H_0 : \mu_d = \mu_{dz}, H_1 : \mu_d = \mu_{dz}$
    - μ_d: mean of population of paired differences
    - μ_{dz}: hypothesized mean of paired differences, commonly 0.
- Test Statistic:
  - $t = \frac{\bar{d}-\mu_{dz}}{s_{\bar{d}}}$
    - degree of freedom= $n-1$
    - $\bar{d}$: sample mean difference $= \frac{1}{n}\sum_{i=1}^{n} d_i$
    - $s_{\bar{d}}$: standard error of mean difference= $\frac{s_d}{\sqrt[2]{n}}$
    - s_d: sample standard deviation= $\sqrt[2]{\frac{\sum_{i=1}^{n}(d_i-\bar{d})^2}{n-1}}$

4. Tests of Variance, Correlation, and Independence

- Tests of Single Population Variance

  - Hypothesis:
    - $H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 = \sigma_0^2$
  - Test Statistics:
    - use Chi-square distribution(χ^2), asymmetrical and ~normal distribution with df↑
      - range: $[0, \infty]$
    - test statistics: $\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$
      - degree of freedom: $n-1$

- Tests of Two Population Variances Equality

  - Condition Usage:
    - iid
    - two normally distributed populations

- Hypothesis:
    - $H_0 : \sigma_1^2 = \sigma_2^2, H_1 : \sigma_1^2 = \sigma_2^2$
    - OR $H_0 : \sigma_1^2 \le \sigma_2^2, H_1 : \sigma_1^2 > \sigma_2^2; H_0 : \sigma_1^2 \ge \sigma_2^2, H_1 : \sigma_1^2 < \sigma_2^2$
- F-test:
    - F-distribution: right-skewed,the shape depends on dfs, range: $[0, \infty]$
    - Test Statistics: $F = \frac{s_1^2}{s_2^2}$
        - The ratio of sample variances
        - s_1^2 or s_2^2: variance of sample of n_1 or n_2 observations drawn from population 1,2
        - Degree of freedom: $n_1 - 1(df_1$, numerator) and $n_2 - 1(df_2$, denominator)
    - Put the **larger variance** on numerator, for having **right tail** graph and getting a convenient critical value
    - sample variance equal, $F = 1$; upper critical value>1; lower critical value<1.
    - lower critical value is **the reciprocal** of the upper critical value
    - While, practically, we **only** have the **upper** critical value, by putting larger sample variance in numerator.

* Parametric and Nonparametric Tests

    - Parametric Tests
        - assumptions depend on <u>population distribution</u> and <u>population parameters</u>
    - Nonparametric Tests
        - used concern about ***quantity*** without carrying about any parameters or the parametric tests cannot be applied for improper assumption or unfitted dataset
            - data are ranked other than values
            - hypothesis not involve parameters of population
            - the ***distribution*** of random variable requirement is not met.

* Test of Correlation

    - measure the strength of linear relationship between two variables
    - Condition: two normal distributed populations
    - Test Statistic: $t_{(n-2)} = \frac{r\sqrt[2]{n-2}}{\sqrt[2]{1-r^2}}$
        - r: sample correlation
        - n: sample size

        - degree of freedom: $n-2$
    - Hypothesis: population correlation(ρ)=0
    - Other correlation coefficient describes the strength of **non-linear** relationship:
        - Spearman rank correlation test:
            - non-parametric test
            - test whether two sets of ranks are correlated
            - if 2nd and 3rd value is the equal, take the average $(2+3)/2 = 2.5$
            - $r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$
                - r_s: rank correlation
                - d_i: difference between two ranks

- Test of Independence with Contingent Table Data

    - Contingent Table: combination of two characteristics. We can test whether these two characteristics are independent with each other.
    - Chi-square test statistic: $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}$
        - r: numbers of row
        - c: numbers of column
        - $O_{i,j}$: the number of observations in cell i,j, observed frequency
        - $E_{i,j}$: the expected number of observations for cell i,j
            - $E_{i,j}$= (total for row i * total column for j)/total for all
        - degree of freedom: $(r-1)(c-1)$
-