PRACTICAL TIPS & TRICKS

1. HYPOTHESIS AND SIGNIFICANCE TESTING FOR MEANS & VARIANCES:

- MEAN → average of the given set of data (vector)
- VARIANCE → sum of squares of differences between all numbers and the mean over the nb. of elements in the given set of data
- STANDARD DEVIATION → the square root of variance

- mean(X) → computes the mean
- var(X) → computes the variance
- std(X) → computes the standard deviation

- H_0 → null hypothesis
- H_1 → alternative hypothesis

- H_0 : $\theta = \theta_0$ → we always take the null hypothesis to be the equality

- H_1 :
  - $\theta < \theta_0 \Rightarrow$ left-tailed test
  - $\theta > \theta_0 \Rightarrow$ right-tailed test
  - $\theta \neq \theta_0 \Rightarrow$ two-tailed test

- alpha : $\alpha \in (0,1)$ → the significance level

- TS → statistics test
- $TS_0 = TS(\theta = \theta_0)$ → observed value of the statistic test

- RR → rejection region

- P-value → probability value

- sigma σ is the VARIANCE

1.1. ZTEST :

- [H, P, CI, ZVAL] = ztest(X, M, SIGMA, ALPHA, TAIL]

  - X ⟶ the vector
  - M ⟶ testing value (θ_0)
  - SIGMA ⟶ can be given in the statement or computed (std(X))
  - ALPHA ⟶ the significance level (usually 5% = 0.05)
  - TAIL = { -1 , for left tailed test ; 0 , for bilateral test ; 1 , for right-tailed test }

  - H ⟶ tells us if we reject the hypothesis or not (0 → no , 1 → reject)
  - P ⟶ the P-value of the test
  - CI ⟶ the confidence interval, containing the lower & upper boundaries of $100(1-\alpha)\%$
  - ZVAL = TS_0 ⟶ the value of the statistics test

- rejection region (RR) :

  - for ztest we will use norminv

  a) left - tailed test $\theta < \theta_0$:

  - $RR = (-\infty, tt_\alpha)$
  - tt_α = norminv(alpha)

  b) right - tailed test $\theta > \theta_0$:

  - $RR = (tt_{1-\alpha}, \infty)$
  - $tt_{1-\alpha}$ = norminv(1 - alpha)

  c) two - tailed test $\theta \neq \theta_0$:

  - $RR = (-\infty, tt_{\frac{\alpha}{2}}) \cup (tt_{1-\frac{\alpha}{2}}, \infty)$
  - $tt_{\frac{\alpha}{2}}$ = norminv(alpha /2)
  - $tt_{1-\frac{\alpha}{2}}$ = norminv(1- alpha /2)

- when we use it :
  - when we have to test the mean of a population and sigma σ is known

1.2. TTEST :

- $[H, P, CI, STATS] = ttest(X, M, ALPHA, TAIL)$

  - X ⟶ the vector
  - M ⟶ the testing value (θ_0)
  - ALPHA ⟶ the significance level (usually 5% = 0.05)
  - TAIL $= \begin{cases} -1, & \text{for left-tailed test} \\ 0, & \text{for bilateral test} \\ 1, & \text{for right-tailed test} \end{cases}$

  - H ⟶ tells us if we reject the null hypothesis or not

    $H = \begin{cases} 0, & \text{do not reject } H_0 \\ 1, & \text{reject } H_0 \end{cases}$
  - P ⟶ the P-value of the test
  - CI ⟶ the confidence interval
  - STATS ⟶ test statistics

    STATS:
    - .tstat $= TS_0$ ⟶ value of the statistics test
    - .df ⟶ degrees of freedom
    - .sd ⟶ estimated population standard deviation

- rejection region (RR) :

  - for ttest we will use the student distribution : $T(n-1)$, tinv

  a) left-tailed test $\theta < \theta_0$:

  - $RR = (-\infty, tt_\alpha)$
  - $tt_\alpha = tinv(alpha, n-1)$, where n is the length of the vector X

  b) right-tailed test $\theta > \theta_0$:

  - $RR = (tt_{1-\alpha}, \infty)$
  - $tt_{\alpha-1} = tinv(1-alpha, n-1)$

  b) two-tailed test $\theta \neq \theta_0$:

  - $RR = (-\infty, tt_{\frac{\alpha}{2}}) \cup (tt_{1-\frac{\alpha}{2}}, \infty)$
  - $tt_{\frac{\alpha}{2}} = tinv(alpha/2, n-1)$
  - $tt_{1-\frac{\alpha}{2}} = tinv(1-alpha/2, n-1)$

- when do we use it :
  - when we have to test the mean of a population and sigma σ is unknown

1.3. VARTEST2 :

- [H, P, Ci, STATS] = vartest2(X1, X2, ALPHA, TAIL)

  - X_1 → the first vector
  - X_2 → the second vector
  - ALPHA → the significance level (usually 5% = 0.05)
  - TAIL = $\begin{cases} -1, & \text{for left-tailed test} \\ 0, & \text{for bilateral test} \\ 1, & \text{for right-tailed test} \end{cases}$

  - H → tells us if we reject the null hypothesis or not

    H = $\begin{cases} 0, & \text{do not reject } H_0 \\ 1, & \text{reject } H_0 \end{cases}$
  - P → the P-value of the test
  - Ci → the confidence interval
  - STATS
    - fstat = TS_0 → the value of statistic test
    - df_1 → numerator degrees of freedom of the test
    - df_2 → denominator degrees of freedom of the test

- rejection region (RR) :

  - for vartest2 we will use finv

  a) left - tailed test $\theta < \theta_0$:

  - $RR = (-\infty, tt_{\alpha})$
  - tt_{α} = finv(alpha, n1-1, n2-1)
    - n1 = length of vector X_1
    - n2 = length of vector X_2

  b) right - tailed test $\theta > \theta_0$:

  - $RR = (tt_{1-\alpha}, \infty)$
  - $tt_{1-\alpha}$ = finv(1-alpha, n1-1, n2-1)

  c) two - tailed test $\theta \neq \theta_0$:

  - $RR = (-\infty, tt_{\frac{\alpha}{2}}) \cup (tt_{1-\frac{\alpha}{2}}, \infty)$
  - $tt_{\frac{\alpha}{2}}$ = finv(alpha/2, n1-1, n2-1)
  - $tt_{1-\frac{\alpha}{2}}$ = finv(1-alpha/2, n1-1, n2-1)

- when we use it :
  - when we have to test the variances of two populations

1.h. TTEST 2 :

omit it

- [H, P, Ci, STATS] = ttest2(X_1, X_2, ALPHA, TAIL, VARTYPE)

- → X_1 → the first vector
- → X_2 → the second vector
- → ALPHA → the significance level (usually 5 % = 0.05)
- → TAIL = { -1 , for left - tailed test ; 0 , for two - tailed test ; 1 , for right - tailed test }

- → H → tells us if we reject the null hypothesis or not
  - H: 0 , do not reject H_0 ; 1 , reject H_0
- → P → the P - value of the statistic test
- → STATS
  - . tstat = TS_0 → the value of the statistic test
  - . df → degrees of freedom of the test
  - . sd → estimation of the standard deviation

- rejection region (RR) :

  - → for ttest2 we will use tinv , but we have more cases :

    A. if the populations variances are equal (aka H from vartest2 is 0) :

  A. a) left- tailed test $\theta < \theta_0$:

  - $RR = (-\infty, tt_\alpha)$
  - $n = n_1 + n_2 - 2$
  - tt_α = tinv (alpha , n)

  n_1, n_2 → lengths of the vectors

  A. b) right- tailed test $\theta > \theta_0$:

  - $RR = (tt_{1-\alpha}, \infty)$
  - $n = n_1 + n_2 - 2$
  - $tt_{1-\alpha}$ = tinv (1- alpha , n)

  A. c) two - tailed test $\theta \neq \theta_0$:

  - $RR = (-\infty, tt_{\frac{\alpha}{2}}) \cup (tt_{1-\frac{\alpha}{2}}, \infty)$
  - $n = n_1 + n_2 - 2$
  - $tt_{\frac{\alpha}{2}}$ = tinv (alpha /2 , n)
  - $tt_{1-\frac{\alpha}{2}}$ = tinv (1 - alpha /2 , n)

B. if the populations variances are different (aka H from vartest2 is 1):

- $c = \dfrac{\frac{v_1}{n_1}}{\frac{v_1}{n_1} + \frac{v_2}{n_2}}$ - $n = \dfrac{c^2}{n_1 - 1} + \dfrac{(1-c)^2}{n_2 - 1}$

- $n = 1/n$

B. a) left - tailed test $\theta < \theta_0$:

- $RR = (-\infty, tt_\alpha)$
- tt_α = tinv (alpha, n)

B. b) right tailed test $\theta > \theta_0$:

- $RR = (tt_{1-\alpha}, \infty)$
- $tt_{1-\alpha}$ = tinv (1 - alpha, n)

B. c) two - tailed test $\theta \neq \theta_0$:

- $RR = (-\infty, tt_{\frac{\alpha}{2}}) \cup (tt_{1-\frac{\alpha}{2}}, \infty)$
- $tt_{\frac{\alpha}{2}}$ = tinv (alpha / 2, n)
- $tt_{1-\frac{\alpha}{2}}$ = tinv (1 - alpha / 2, n)

- when we use it:
  - when we have to test the difference of means of two populations
  - ON AVERAGE is mentioned

2. CONFIDENCE INTERVALS :

2.1. FOR A POPULATION MEAN :

2.1.1. SIGMA KNOWN :

```
n = length(X);
q = norminv(1 - alpha/2);
m = mean(X);
```

$$left = m - \frac{sigma}{\sqrt{n}} \cdot q \qquad right = m + \frac{sigma}{\sqrt{n}} \cdot q$$

2.1.2. SIGMA UNKNOWN :

```
n = length(X);
q = tinv(1 - alpha/2, n-1);
m = mean(X);
s = std(X);
```

$$left = m - \frac{s}{\sqrt{n}} * q \qquad right = m + \frac{s}{\sqrt{n}} * q$$

2.2. FOR A POPULATION VARIANCE :

```
n = length(X);
q1 = chi2inv(1 - alpha/2, n-1);
q2 = chi2inv(alpha/2, n-1);
v = var(X);
```

$$left = \frac{(n-1)v}{q_1} \qquad right = \frac{(n-1)v}{q_2}$$

2.3. FOR THE DIFFERENCE OF TWO POPULATION MEANS :

2.3.1. BOTH SIGMA KNOWN :

```
n1 = length(X1);
n2 = length(X2);
q = norminv(1 - alpha/2);
m1 = mean(X1);
m2 = mean(X2);
```

$$left = m_1 - m_2 - q \cdot \sqrt{\frac{sigma1^2}{n_1} + \frac{sigma2^2}{n_2}}$$

$$right = m_1 - m_2 + q \cdot \sqrt{\frac{sigma1^2}{n_1} + \frac{sigma2^2}{n_2}}$$

2.3.2. SIGMA UNKNOWN AND EQUAL :

$n_1 = \text{length}(X_1);$
$n_2 = \text{length}(X_2);$
$n = n_1 + n_2 - 2;$
$m_1 = \text{mean}(X_1);$
$m_2 = \text{mean}(X_2);$
$v_1 = \text{var}(X_1);$
$v_2 = \text{var}(X_2);$
$q = \text{tinv}(1 - \text{alpha}/2, n);$

$$sp = \sqrt{\frac{(n_1-1)v_1 + (n_2-1)v_2}{n}}$$

$$left = m_1 - m_2 - q \cdot sp \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$right = m_1 - m_2 + q \cdot sp \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

2.3.3. SIGMA UNKNOWN AND DIFFERENT:

$n_1 = \text{length}(X_1);$
$n_2 = \text{length}(X_2);$
$m_1 = \text{mean}(X_1);$
$m_2 = \text{mean}(X_2);$
$v_1 = \text{var}(X_1);$
$v_2 = \text{var}(X_2);$

$$c = \frac{\frac{v_1}{n_1}}{\frac{v_1}{n_1} + \frac{v_2}{n_2}} \qquad n = \frac{c^2}{n_1-1} + \frac{(1-c)^2}{n_2-1} \qquad n = 1/n$$

$q = \text{tinv}(1 - \text{alpha}/2, n);$

$$left = m_1 - m_2 - q \cdot \sqrt{\frac{v_1}{n_1} + \frac{v_2}{n_2}}$$

$$right = m_1 - m_2 + q \cdot \sqrt{\frac{v_1}{n_1} + \frac{v_2}{n_2}}$$

2.4. FOR THE RATIO OF TWO POPULATION VARIANCES:

```
n1 = length(X1);
n2 = length(X2);
v1 = var(X1);
v2 = var(X2);

q1 = finv(1- alpha/2, n1-1, n2-1)
q2 = finv(alpha/2, n1-1, n2-1)
```

$$left = \frac{1}{q_1} \cdot \frac{v_1}{v_2} \qquad right = \frac{1}{q_2} \cdot \frac{v_1}{v_2}$$