Hypothesis testing about the mean of a population

Sample of 2234 persons in Nicaragua, income per capita in 1998 in córdobas (C$) (1998 US$1~ C$10)

```
. sum i_income urban indigenous

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    i_income |      2234    5369.052    7533.916   5.409909   168552.9
       urban |      2234    .5022381    .5001069          0          1
  indigenous |      2234     .045658    .2087891          0          1
```

1. Testing about the mean in a population

The conjecture is that the average income per capita in the population is 5000 córdobas. Is there statistical evidence to the contrary in our sample? H0 is presumed correct until we find statistical evidence against it.

a. Set the hypotheses
Null hypothesis on the mean in the population: H0: $\mu = \mu_0 = 5000$ córdobas (with unknown variance $\text{var}(X)$)
Alternative hypothesis H1: $\mu > 5000$ córdobas (We don't even consider the values <5000)

Under H0, and for large sample, the sample mean $\bar{X} \sim N\left(\mu_0, \text{var}(\bar{X})\right)$,

or $\dfrac{\bar{X}-\mu_0}{\text{sd}(\bar{X})} \sim N(0,1)$, with $\text{sd}(\bar{X}) = \dfrac{\text{sd}(X)}{\sqrt{n}}$

or $\dfrac{\bar{X}-\mu_0}{\text{se}(\bar{X})} \sim t_{n-1}$, where $\text{se}(\bar{X}) = \dfrac{s}{\sqrt{n}}$ is an estimation of $\text{sd}(\bar{X})$

Observations on a large sample of n = 2234 persons gives: $\text{se}(\bar{X}) = \dfrac{s}{\sqrt{n}} = \dfrac{7534}{\sqrt{2234}} = 159.4$

b. Compute the *t-statistic*: $\dfrac{\bar{X}-\mu_0}{\text{se}(\bar{X})} = t_{2233} = \dfrac{5369-5000}{159.4} = 2.31$

c. Choose a *level of significance*
Lookup the *critical value c* for this level: 5% $c = 1.645$
1% $c = 2.326$

d. Reject H0 if $t_{2233} > c$

➔ *Reject* H0 at the 5% significance level, because $t = 2.31 > 1.645$
➔ *Fail to reject* H0 at the 1% significance level, because $t = 2.31 < 2.326$

e. Conclude: At the 5% significance level, there is statistical evidence that the income per capita is higher than 5000 córdobas.
At the 1%, there is no statistical evidence that income per capita is not equal to 5000 córdobas

Level of significance = probability to reject H0 when in fact it is true.

Two-side alternative hypothesis:

a. Null hypothesis: H0: $\mu = \mu_0 = 5000$ córdobas with unknown variance $\text{var}(X)$ for X
Alternative H1: $\mu \neq 5000$ córdobas

b. Compute the t-statistic: $\dfrac{\overline{X}-\mu_0}{\text{se}\left(\overline{X}\right)} = t_{2233} = \dfrac{5369-5000}{159.4} = 2.31$

c. Choose the significance level and the corresponding critical value:
significance level 5% $c = 1.96$

d. Reject H0 at the 5% significant level
e. There is statistical evidence that income per capita is not equal to 5000 Cordobas

2. Testing for the equality in means in two populations.
Is the income per capita in rural areas equal to that in urban areas?

Urban

```
. sum urban i_income if urban==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       urban |      1122           1           0          1          1
    i_income |      1122    7061.576    8973.852   128.6425   168552.9
```

Rural

```
. sum urban i_income if urban==0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       urban |      1112           0           0          0          0
    i_income |      1112    3661.307    5197.585   5.409909   79047.75
```

For large samples, $\overline{X}_u \sim N\left(\mu_u, \text{var}\left(\overline{X}_u\right)\right)$, $\overline{X}_r \sim N\left(\mu_r, \text{var}\left(\overline{X}_r\right)\right)$, and thus

$\overline{X}_u - \overline{X}_r \sim N\left(\mu_u - \mu_r, \text{var}\left(\overline{X}_u - \overline{X}_r\right)\right)$

$\overline{X}_u - \overline{X}_r = 7061.6 - 3661.3 = 3400.3$

$\widehat{\text{var}\left(\overline{X}_u - \overline{X}_r\right)} = \widehat{\text{var}\left(\overline{X}_u\right)} + \widehat{\text{var}\left(\overline{X}_r\right)} = \dfrac{s_u^2}{n_u} + \dfrac{s_r^2}{n_r} = \dfrac{5198^2}{1112} + \dfrac{8974^2}{1122} = 96074$ ➔ $\text{se}\left(\overline{X}_u - \overline{X}_r\right) = \sqrt{96074} = 310$

a. Null hypothesis: H0: $\Delta = \mu_u - \mu_r = 0$
Alternative H1: $\Delta > 0$

b. $t_{1122+1112-2} = \dfrac{3400.3-0}{310} = 10.97$ follows a t distribution with $n_u + n_r - 2$ degrees of freedom

c. Choose the significance level and the corresponding critical value:
significance level 1% $c = 2.33$

d. Reject H0 at the 1% significance level that urban and rural per capita incomes are equal in favor of urban income per capita being larger than rural income per capita.

e. There is statistical evidence that urban income per capita is larger than rural income per capita

Always use the 5 steps in hypothesis testing