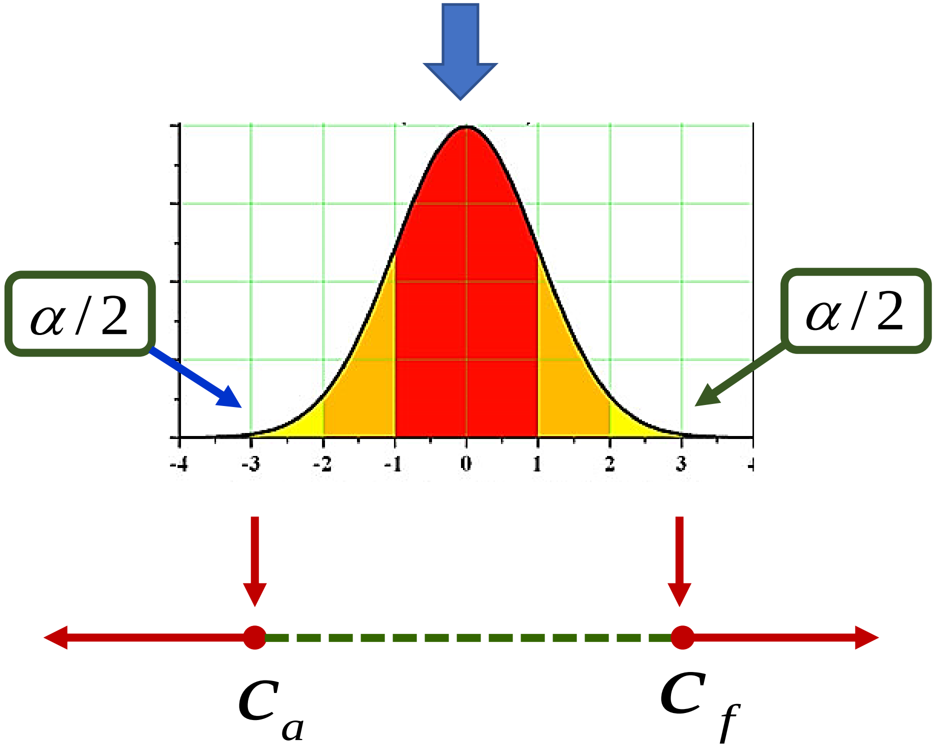
One sample expected value test

$$H_0^T: \mu = \mu_0$$

$$H_0: \mu = \mu_0 \qquad H_0: \mu \geq \mu_0 \qquad H_0: \mu \leq \mu_0$$

$$H_1: \mu \neq \mu_0 \qquad H_1^b: \mu < \mu_0 \qquad H_1^j: \mu > \mu_0$$

B: left side
J: right side



In these cases the whole alfa goes to the right or left side, position depends on the alternative hypotheses.

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

One sample proportion test

$$H_0^T: \quad P = P_0$$

Asymptotic assumption for sample size:

$$\min\{nP_0, n(1-P_0)\} \geq 10$$

$$z = \frac{p - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

p is the proportion in sample (favourable cases/all cases)
P0 is the theoretical proportion (what you state)

Two independent samples T test (expected value)

$$H_0^T:\ \mu_Y - \mu_X = \delta_0$$

$$H_1:\ \mu_Y - \mu_X \neq \delta_0$$
$$H_1:\ \mu_Y - \mu_X > \delta_0$$
$$H_1:\ \mu_Y - \mu_X < \delta_0$$

$$t = \frac{(\bar{y} - \bar{x}) - \delta_0}{s_c \sqrt{\frac{1}{n_Y} + \frac{1}{n_X}}}$$

$$\nu = n_Y + n_X - 2$$

$$s_c = \sqrt{\frac{(n_Y - 1)s_Y^2 + (n_X - 1)s_X^2}{n_Y + n_X - 2}} = \sqrt{\frac{\sum d_Y^2 + \sum d_X^2}{n_Y + n_X - 2}}$$

Assumption: two variances from the samples should be equal. Before the T test, we have to check it. If it is not true, then we should use the not equal variance version of T test in R.

Two samples variance test

$$H_0: \sigma_Y = \sigma_X \quad \text{or} \quad H_0: \sigma_Y^2 = \sigma_X^2 \qquad F = \frac{s_Y^2}{s_X^2} \qquad \nu_1 = n_Y - 1 \text{ és } \nu_2 = n_X - 1$$

$$H_1: \quad \sigma_Y \neq \sigma_X$$

$$H_1: \quad \sigma_Y > \sigma_X$$

$$H_1: \quad \sigma_Y < \sigma_X$$

$$H_1^j: \quad \sigma_Y > \sigma_X \qquad c_f = F_{1-\alpha}(\nu_1, \nu_2)$$

$$H_1^b: \quad \sigma_Y < \sigma_X \qquad c_a = F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)}$$

$$H_1: \quad \sigma_Y \neq \sigma_X \qquad c_a = F_{\alpha/2}(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha/2}(\nu_2, \nu_1)}$$

$$c_f = F_{1-\alpha/2}(\nu_1, \nu_2)$$

ANOVA

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ ("all *k* population means are equal")
H_1: At least one μ_i different ("at least one of the *k* population means is not equal to the others")

Source	*SS*	*df*	*MS*	*F*	*Sig.*
Between	SS_b	k-1	MS_b	MS_b/MS_w	*p* value
Within	SS_w	N-k	MS_w		
Total	$SS_b + SS_w$	N-1			

- Independence of observations: the data were collected using statistically-valid methods, and there are no hidden relationships among observations. If your data fail to meet this assumption because you have a confounding variable that you need to control for statistically, use an ANOVA with blocking variables.
- Normally-distributed response variable: The values of the dependent variable follow a normal distribution.
- Homogeneity of variance: The variation within each group being compared is similar for every group. If the variances are different among the groups, then ANOVA probably isn't the right fit for the data.

Independence testing for qualitative variables

$$H_0 : P_{ij} = P_{i.} \cdot P_{.j}$$

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \left[\frac{\left(n_{ij} - n_{ij}^*\right)^2}{n_{ij}^*} \right]$$

$$\nu = k - b - 1 = rc - (r-1) - (c-1) - 1 = (r-1)(c-1)$$

$$n_{ij}^* \geq 10(5)$$

$$c_f = \chi^2_{1-\alpha}$$

Crosstab:

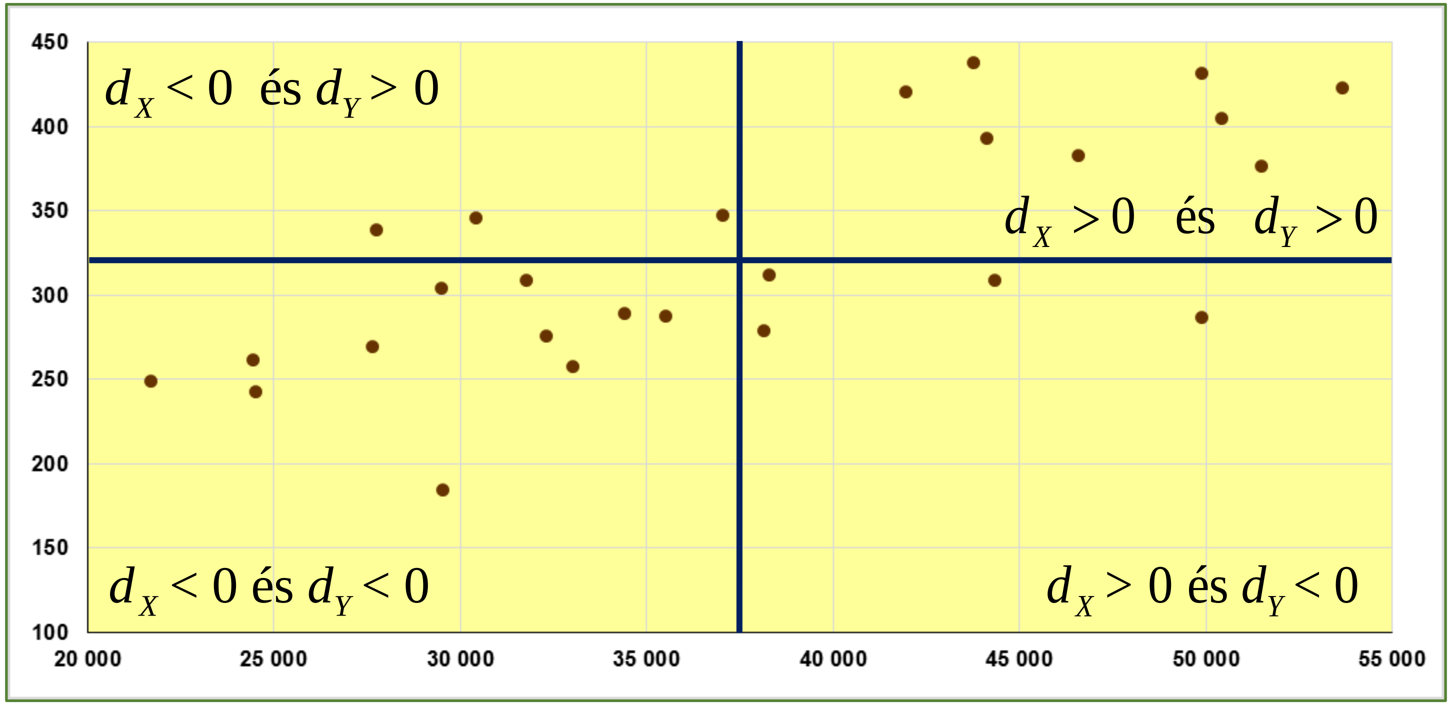
	Column 1	Column 2	Row totals
Row 1	a	b	$a+b$
Row 2	c	d	$c+d$
Column totals	$a+c$	$b+d$	$a+b+c+d$

Covariance

$$C(X,Y)=\frac{\sum X_i Y_i}{N}-\overline{X}\overline{Y}=\frac{\sum d_{X_i} d_{Y_i}}{N}$$

$$d_{X_i}=X_i-\overline{X} \qquad d_{Y_i}=Y_i-\overline{Y}$$

$\overline{Y}=324$



$$0\le\left|C(X,Y)\right|\le\sigma_X\sigma_Y$$

$\overline{X}=37378,9$

Correlation

$$r(X,Y) = \frac{C(X,Y)}{\sigma_X \sigma_Y} \qquad C(X,Y) = \frac{\sum d_{X_i} d_{Y_i}}{N}$$

$$r(X,Y) = \frac{\frac{\sum d_{X_i} d_{Y_i}}{N}}{\sqrt{\frac{\sum d_{X_i}^2}{N}}\sqrt{\frac{\sum d_{Y_i}^2}{N}}} = \frac{\sum d_{X_i} d_{Y_i}}{\sqrt{\sum d_{X_i}^2 \sum d_{Y_i}^2}}$$

$$-1 \le r \le 1$$