Finding Confidence interval for the population proportion:

What we need:

Random probability sample

Conditions for Binomial:

- Fix number of trials
- Trials are independent
- Two outcomes: Success / Failure
- At least 10 'Yes' and 10 'No'

n - sample size; p^ - probability of success (or sample proportion of success); q^ - probability of failure (or sample proportion of failure)

n * p^ >= 10; n * q^ >= 10

Margin of Error (E) - the maximum possible difference between p (population proportion) and p^ (sample proportion for success; is a point estimate for p).

$E = Z_{\alpha/2} \cdot \sqrt{p\hat{} * q\hat{} / n}$ (the square root included p^ , q^, and n)

If p^ and q^ are not accurate, p^ * q^ = 0.5 * 0.5 and after that we use the above formula

$Z_{\alpha/2}$ - Critical value and we use Confidence level to find it:

for 90% Confidence level -> $Z_{\alpha/2}$ = 1,645;

for 95% Confidence level -> $Z_{\alpha/2}$ = 1,96;

for 98% Confidence level -> $Z_{\alpha/2}$ = 2,326;

for 99% Confidence level-> $Z_{\alpha/2}$ = 2,576;

Construct Confidence interval for the population proportion:

p^ - E < p < p^ + E

How to find required sample size for the Survey with given E:

a) If we know p^ and q^:

   n = $(Z_{\alpha/2} * Z_{\alpha/2})$ * p^ * q^ / (E * E)

b) If we don't know p^ and q^ :

   n = $(Z_{\alpha/2} * Z_{\alpha/2})$ * 0.25 / (E * E)

From the given Confidence interval find p^ and E:

p^ = (upper boundary + lower boundary) / 2

E = (upper boundary - lower boundary) / 2

How to estimate the difference between two populations proportion:

$p_1 - p_2 \pm Z_{\alpha/2} * \sqrt{(p\hat{}_1 * q\hat{}_1 / n_1) + (p\hat{}_2 * q\hat{}_2 / n_2)}$ (the square root end after n2)

Finding Confidence interval for the population mean, when population standard deviation is known (σ):

What we need:
Random probability sample
Population standard deviation is known (σ)
n > 30 **or** Population is normally distributed
$\overline{X}$ is a sample mean (point estimate) for population mean (μ)

Margin of Error (E) - the maximum possible difference between p (population proportion) and p^ (sample proportion for success; is a point estimate for p).

$E = Z_{\alpha/2} * \sigma / \sqrt{n}$

$\sigma / \sqrt{n}$ - standard error

$Z_{\alpha/2}$ - Critical value and we use Confidence level to find it:

for 90% Confidence level -> $Z_{\alpha/2} = 1,645$;

for 95% Confidence level -> $Z_{\alpha/2} = 1,96$;

for 98% Confidence level -> $Z_{\alpha/2} = 2,326$;

for 99% Confidence level-> $Z_{\alpha/2} = 2,576$;

Construct Confidence interval for the population mean:

$\overline{X} - E < \mu < \overline{X} + E$

From the given Confidence interval find $\overline{X}$ and E:

$\overline{X}$ = (upper boundary + lower boundary) / 2

E = (upper boundary - lower boundary) / 2

How to find required sample size for the Survey with given E and σ :

n = $(Z_{\alpha/2} * Z_{\alpha/2}) * (\sigma * \sigma) / (E * E)$

How to estimate the difference between two population means, if we have two independent groups, and σ is known for each population (example- BMI between men and
women Mexican-American):

$\mu_1 - \mu_2 \pm Z_{\alpha/2} * \sqrt{((\sigma_1 * \sigma_1) / n_1)) + ((\sigma_2 * \sigma_2) / n_2))}$ (the square root end after n2)

Finding Confidence interval for the population mean, when population standard deviation is unknown (more realistic case):

What we need:

Random probability sample

n > 30 **or** Population is normally distributed

$\overline{X}$ is a sample mean (point estimate) for population mean (μ)

If we don't know σ, we can't use $Z_{\alpha/2}$. Instead we use $T_{\alpha/2}$ (T - score).
Critical values are given by $T_{\alpha/2}$.

If the sample size is big enough $T_{\alpha/2} = Z_{\alpha/2}$

Steps to find T-score:

Calculate Degrees of Freedom (**D.F. = n - 1)**, n is a sample size;

Calculate **α** (Alpha); **α** is a Significance Level and is a complement of **Critical level**; For example if **Critical Level** is 95%, **α** = 5% or 0.05)

Use statistics table to find T - score;

Margin of Error (E) - the maximum possible difference between p (population proportion) and p^ (sample proportion for success; is a point estimate for p).

$E = T_{\alpha/2} * s / \sqrt{n}$

s - sample standard deviation

$s / \sqrt{n}$ - estimate standard error (sample standard error)

Construct Confidence interval for the population mean:

$\overline{X} - E < \mu < \overline{X} + E$

From the given Confidence interval find $\overline{X}$ and E:

$\overline{X}$ = (upper boundary + lower boundary) / 2

E = (upper boundary - lower boundary) / 2

How to estimate the difference between two population means, if we have two independent groups (for example- BMI between men and women Mexican-American) and standard deviation for these two populations is unknown:

a) First approach - assumption that $(\sigma_1 * \sigma_1)$ is not equal to $(\sigma_2 * \sigma_2)$

Degree of Freedom = $\min(n_1 - 1; n_2 - 1)$; take the minimum of these two

After that find $T_{\alpha/2}$ from the statistics table.

$\mu_1 - \mu_2 +- T_{\alpha/2} * \sqrt{((s_1 * s_1) / n_1)) + ((s_2 * s_2) / n_2))}$ (the square root end after n2)

b) Second approach - assumption that $(\sigma_1 * \sigma_1)$ is equal to $(\sigma_2 * \sigma_2)$

Degree of Freedom = $n_1 + n_2 - 2$

After that find $T_{\alpha/2}$ from the statistics table

$$\mu_1 - \mu_2 \pm T_{\alpha/2} * \sqrt{(((n_1 - 1)(s_1 * s_1)) + ((n_2 - 1)(s_2 * s_2))) / (n_1 + n_2 - 2)}$$ (the square root end after 2)

$$* \sqrt{(1 / n_1) + (1 / n_2)}$$ (the square root end after n_2)

How to estimate the difference between two population means, if we have paired data (for example older twin education vs younger twin education)

n - is the same for these two groups
Degree of Freedom = n - 1
After that find $T_{\alpha/2}$ from the statistics table.

$$\mu_1 - \mu_2 \pm T_{\alpha/2} * (s_d / \sqrt{n})$$

$S_d = S_1 - S_2$

Finding Confidence interval for Variance (σ^2) and Standard Deviation (σ). Chi-Squared Distribution. One population

Chi-Squared Distribution is not symmetrical, it's right-skewed.
Values are only positive, because the distribution has only one tail to the right.
If Degrees of Freedom goes up the distribution becomes more symmetrical.
Chi-Squared Distribution gives Critical value to the Right (X_R^2) for current Confidence Level (90%, 95%, 98%, 99%).

Find Left and Right critical values

n = 12, 95% Confidence Level, Degrees of Freedom = 11 right_critical = scipy.stats.chi2.ppf(1-.025, df = 11) is $\mathbf{X_R}^2$ = 21.92

left_critical = scipy.stats.chi2.ppf(.025, df=11) is $\mathbf{X_L}^2$ = 3,816

Construct Confidence interval for the population variance:

$(n - 1)(s*s) / \chi_R^2 < \sigma^2 < (n - 1)(s*s) / \chi_L^2$

Construct Confidence interval for the population standard deviation:

$$\sqrt{(n - 1)(s*s) / \chi_R^2} < \sigma < \sqrt{(n - 1)(s*s) / \chi_L^2}$$

(the square roots end after the two Chi-squared which we found from Statistics Table or with Python)

How to estimate the ratio between two population variances

Find F left critical value and F right critical value

F_right = scipy.stats.f.ppf (q = 1 - .025, df1, df2)

F_left = scipy.stats.f.ppf(q = .025, df1, df2)

first is the population with bigger size

Construct confidence interval for the difference

1 / F_right * (variance_1 / variance_2) < variance_1 / variance_2 > 1 / F_left *

(variance_1 / variance_2)