Hypothesis Testing for Population Proportion

Requirements:
Random sample
$n*p >= 5$; $n*q <= 5$
n is sample size

Traditional Method (Non-Bayesian Testing):

Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic

$Zt = p\_hat - p \;/\; sqrt((p * q) / n)$
p_hat - sample proportion of success
p and q are hypothetical values (success and failure)
$q = 1 - p$

Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
<u>On the picture put Z-critical value with corresponding alpha from Z-table</u>

Step 5: Interpret results:
If Z-test statistic is in Rejection Region => Reject H_0 and accept H_1
If the Z-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

P-value method (Bayesian Testing):

Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic

$Zt = p\_hat - p \;/\; sqrt((p * q) / n)$
p_hat - sample proportion of success
p and q are hypothetical values (success and failure)

Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
<u>On the picture put Z-test statistic</u>
P-value is : 1 - Area (Fail to Rejection Region); <u>P-value is the area in the Tail</u>;
If we have two tails we need to multiply by 2

Step 5: Interpret results:
If P-value <= alpha => Reject H_0 and accept H_1
If P-value > alpha => Fail to Reject H_0; We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing with Two Proportions:
Traditional Method (Non-Bayesian Testing)

Step 1: Define Claim and Opposite, H_0 and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic:

$Zt = (p1\_hat - p2\_hat) - (p1 - p2) \;/\; sqrt(p\_hat\_total * (1 - p\_hat\_total) * (1/n1 + 1/n2))$

p1 - p2 is hypothetical difference
p_hat_total is (X1 + X2) / (n1 + n2)

Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
<u>On the picture put Z-critical value with corresponding alpha from Z-table</u>
Step 5: Interpret results:
If Z-test statistic is in Rejection Region => Reject H_0 and accept H_1
If the Z-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing for Population Mean. Population standard deviation is known

Requirements:
Random sample
Population standard deviation is known
n > 30 or Population is normally distributed
n is sample size
(If n <30 we use T distribution)

Traditional Method (Non-Bayesian Testing):

Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic:

Zt = X_bar - μ / σ / sqrt(n)

X_bar - sample mean
μ - population hypothetical mean
σ - population standard deviation
n - sample size
Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
<u>On the picture put Z-critical value with corresponding alpha from Z-table</u>
Step 5: Interpret results:
If Z-test statistic is in Rejection Region => Reject H_0 and accept H_1
If the Z-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

P-value method (Bayesian Testing):

Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic

Zt = X_bar - μ_0 / σ / sqrt(n)

X_bar - sample mean
μ_0 - population hypothetical mean
σ - population standard deviation
n - sample size
Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
On the picture put Z-test statistic
P-value is : 1 - Area (Fail to Rejection Region); P-value is the area in the Tail;
If we have two tails we need to multiply by 2
Step 5: Interpret results:
If P-value <= alpha => Reject H_0 and accept H_1
If P-value > alpha => Fail to Reject H_0; We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing with Two Means.
Populations are independent and standard deviations σ1, σ2 are known

Requirements:
Random sample
Population standard deviation is known
n > 30 or Population is normally distributed
n is sample size
(If n <30 we use T distribution)

Traditional Method (Non-Bayesian Testing):
Step 1: Define Claim and Opposite, H_0 and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic:

Zt = (X1_bar - X2_bar) - D_0 / sqrt(((σ1 * σ1) / n_1) + ((σ2 * σ2) /n2))

D_0 - Hypothetical difference
Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
On the picture put Z-critical value with corresponding alpha from Z-table
Step 5: Interpret results:
If Z-test statistic is in Rejection Region => Reject H_0 and accept H_1
If the Z-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing with Two Means.
Populations are independent and standard deviations σ1, σ2 are unknown or n <= 30

Traditional Method (Non-Bayesian Testing):
Step 1: Define Claim and Opposite, H_0 and H_1
Step 2: Define significance level (alpha)

Step 3: Calculate T-test statistic:

$Tt = (X1\_bar - X2\_bar) - D_0 / sqrt(((s_1 * s_1) / n_1) + ((s_2 * s_2) /n_2))$

D_0 - Hypothetical difference

s1 and s2 - sample standard deviation

Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail

On the picture put T-critical value with corresponding alpha from T-table and DF

Degree of Freedom = n1 + n2 - 2

Step 5: Interpret results:

If T-test statistic is in Rejection Region => Reject H_0 and accept H_1

If the T-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing with Two Means.
Paired data

Traditional Method (Non-Bayesian Testing):

Step 1: Define Claim and Opposite, H_0 and H_1

Step 2: Define significance level (alpha)

Step 3: Calculate T-test statistic:

$Tt = (X1\_bar - X2\_bar) - D_0 / s_d / sqrt(n)$

D_0 - Hypothetical difference

s_d - standard deviation difference; s_1-s_2 or **$\sigma 1 - \sigma 2$ (and then we have Z-test statistic)**

Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail

On the picture put T-critical value with corresponding alpha from T-table and DF

Degree of Freedom = n1 + n2 - 2

Step 5: Interpret results:

If T-test statistic is in Rejection Region => Reject H_0 and accept H_1

If the T-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing for Population Mean. Population Standard Deviation is unknown

Requirements:

Random sample

Population standard deviation is unknown

n > 30 or Population is normally distributed
n is sample size
(If n <30 we use T distribution too)

Traditional Method (Non-Bayesian Testing):
Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Z-test statistic:

Tt = X_bar - μ / s / sqrt(n)

X_bar - sample mean
μ - population hypothetical mean
s - sample standard deviation
n - sample size
Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail
On the picture put T-critical value with corresponding alpha from T-table
Step 5: Interpret results:
If T-test statistic is in Rejection Region => Reject H_0 and accept H_1
If the T-test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing for Variance and Standard Deviation
Chi-squared distribution

Traditional Method:
Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1
Step 2: Define significance level (alpha)
Step 3: Calculate Test statistic:

X^2 = (n - 1) (s * s) / (σ * σ)

Step 4*: Draw a picture: according to H_1 it is left-tail, right-tail or two-tail; Keep in mind that distribution starts from Zero and it's only Right-Skewed
On the picture put Chi-squared critical value with corresponding alpha and DF
Be careful if it is left-squaded (1 - alpha) and after that look in the table
Step 5: Interpret results:
If Test statistic is in Rejection Region => Reject H_0 and accept H_1
If the Test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing for Comparing Two Variances - Two independent Samples F- distribution

Step 1: Define Claim and Opposite, H_0 (contains equal sign) and H_1

Step 2: Define significance level (alpha)

Step 3: Calculate Test statistic:

$F = \text{variance}_{(larger)} / \text{variance}_{(smaller)}$

Step 4*: Draw a picture: it is always upper-tailed

On the picture put F-critical value which is based on DF of these two samples and alpha, which is calculated with calculator (or in Python)

Step 5: Interpret results:

If Test statistic is in Rejection Region => Reject H_0 and accept H_1

If the Test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Hypothesis Testing: Chi-Square Test

It helps us to understand the relationship between two categorical variables: grade level, sex, age group, year. Chi-Square test involve the frequency of events; the count; Expected Vs Observed categorical distribution

Example - determine if this die is a fair or not with 95% certainty; 600 trials for the next 6 days;

Example - a school principal expected that students will be absent equally during the 5-day school week;

Step 1: Construct two tables: Observed vs Expected value

Step 2: Define Claim and Opposite, H_0 (contains equal sign) and H_1

Step 3: Define significance level (alpha)

Step 4: Calculate Test statistic:

Chi-Squared = sum**(**(Observed - Expected)^2 / Expected**)**

Step 5: Draw a picture:

Keep in mind that distribution starts from Zero and it's only Right-Skewed

On the picture put Chi-Squared critical value which is based on DF(n - 1) and alpha, which is calculated with calculator (or with Python)

Step 6: Interpret results:

If Test statistic is in Rejection Region => Reject H_0 and accept H_1

if the Test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1

Test of Independence Using Chi-Square Distribution

(watch the video again in your playlist)

Example: Is the average number of studying hours depend on the type of student;

Step 1: Construct two tables: Observed vs Expected value ($E.V = Row_{total} * Col_{total} / N$)

Step 2: Define Claim and Opposite, H_0 (contains equal sign) and H_1

Step 3: Define significance level (alpha)
Step 4: Calculate Test statistic:

Chi-Squared = sum**(**(Observed - Expected)^2 / Expected**)**

Step 5: Draw a picture:

Keep in mind that distribution starts from Zero and it's only Right-Skewed

<u>On the picture put Chi-Squared critical value</u> which is based on DF(n - 1) and alpha, which is calculated with calculator (or with Python)

Step 6: Interpret results:

If Test statistic is in Rejection Region => Reject H_0 and accept H_1

if the Test statistic is in the Fail to Rejection Region => We know nothing! There is not enough evidence to accept H_1