Variables:

One quantitative (V1)

One qualitative (V2)

Assumptions of ANOVA

1. Variable Type
2. Independence
3. Normality of residuals
4. Equality of variance

STEP 1 :

Check for independence

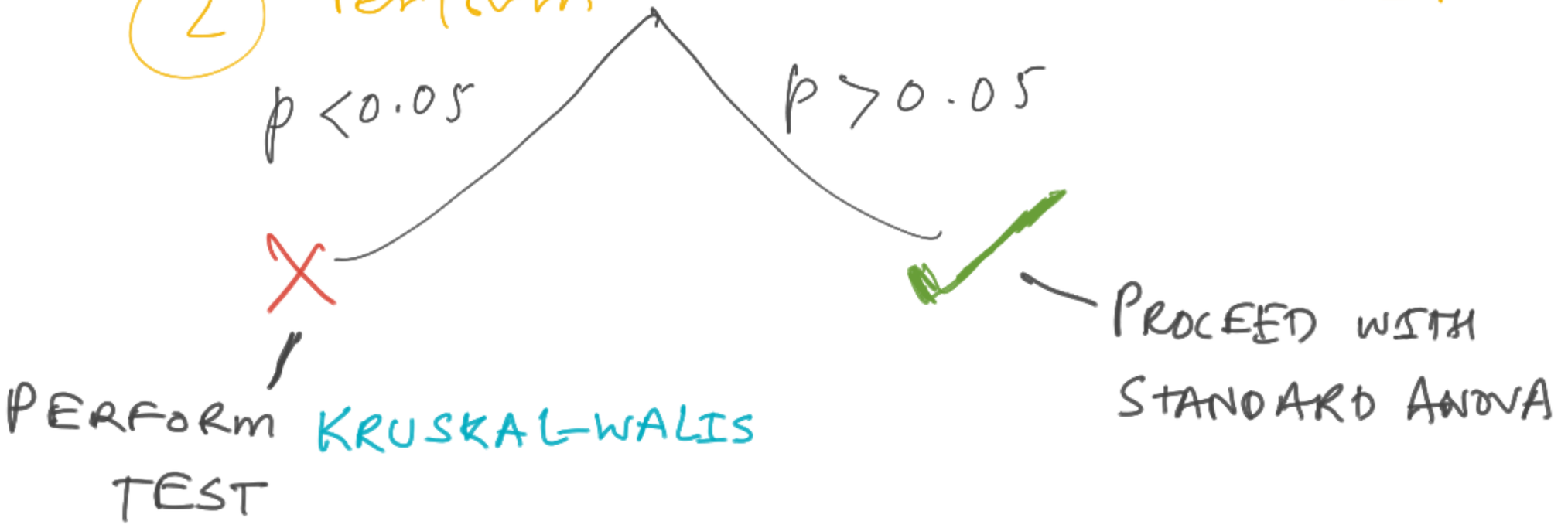
— data was randomly sampled

— there is no dependence b/w obs.

Step 2 :

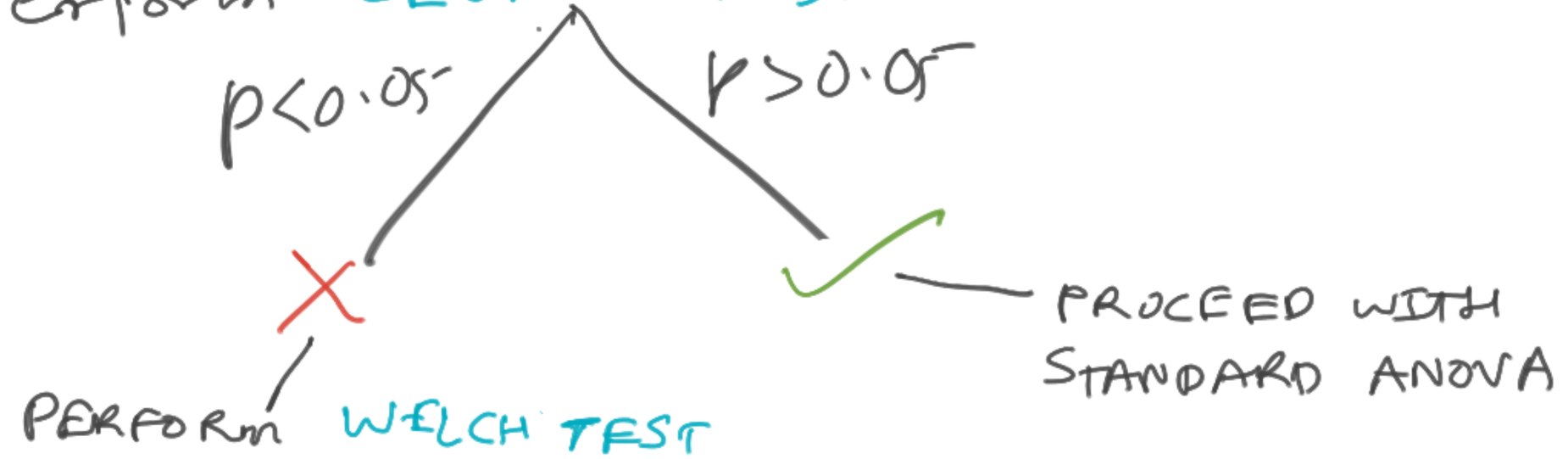
NORMALITY

1. Draw a histogram of residuals
2. Perform SHAPIRO-WILK TEST



STEP 3:
Equality of Variance

1. Draw a boxplot
2. Perform LEUENE TEST



In R:

1. Calculate degrees of freedom for groups (df_G)
2. Calculate SSG (sum of squares - groups)
3. Calculate MSG (mean square between groups)

---

4. Calculate degrees of freedom (df_E)
5. Calculate SSE (sum of squared error)
6. Calculate MSE (mean squared error)

DEGREES OF FREEDOM (GROUPS)

$$df_G = k - 1$$

where k: number of groups.

example: $k = 7$, $df_G = 6$

SUM OF SQUARES : GROUPS

$$SSG = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{\bar{x}})^2$$

where n_i : sample size in group i

$\bar{x}_i$: mean for group i

$\bar{\bar{x}}$: mean for the entire data.

MEAN SQUARE BETWEEN GROUPS (MSG)

$$MSG = \frac{SSG}{df_G}$$

MSG offers us a way to understand the average variation across groups.

Degrees of Freedom (Error)

$$df_E = n - k$$

where n : number of observations

k : number of groups

example : $n = 100$, $k = 6$

$$df_E = 94$$

SUM OF SQUARED ERRORS (SSE)

$$SSE = \sum (n_i - 1) s_i^2$$

n_i : sample size within each group

s_i^2 : sample variance

NOTE:
Why $(n-1)$ here?

because you don't want to underestimate the variance.

MEAN SQUARE ERROR (MSE)

$$MSE = \frac{SSE}{df_E}$$

This measures the variation within groups.

F-statistics and p-value

$$F = \frac{MSG}{MSE}$$

MSG → variation across groups

MSE → variation within groups

using the F-stat, you can compute p-value in R:

```
pf(F, df_G, df_E, lower.tail = FALSE)
```