# Phase Analyser

# Vérification d'hypothèse:

## Les variables continues

Fethi Derbeli . 2021

# La dynamique d'exécution

- L'effet d'entonnoir

Schéma du processus

Modèle C&E et AMDEC

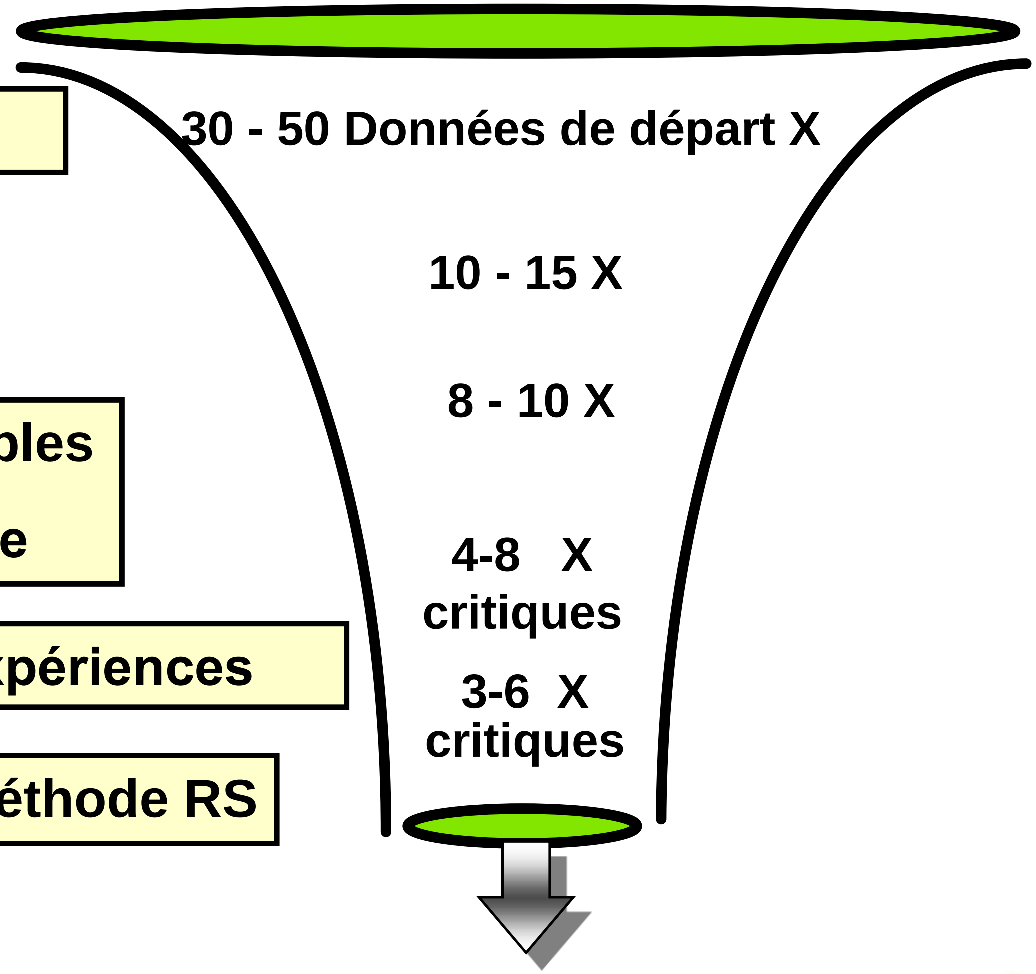
Etudes de variables multiples

Vérification d'hypothèse

Triage Conception D'Expériences

CDE / Opérations Evol. /Méthode RS

Plan de contrôle

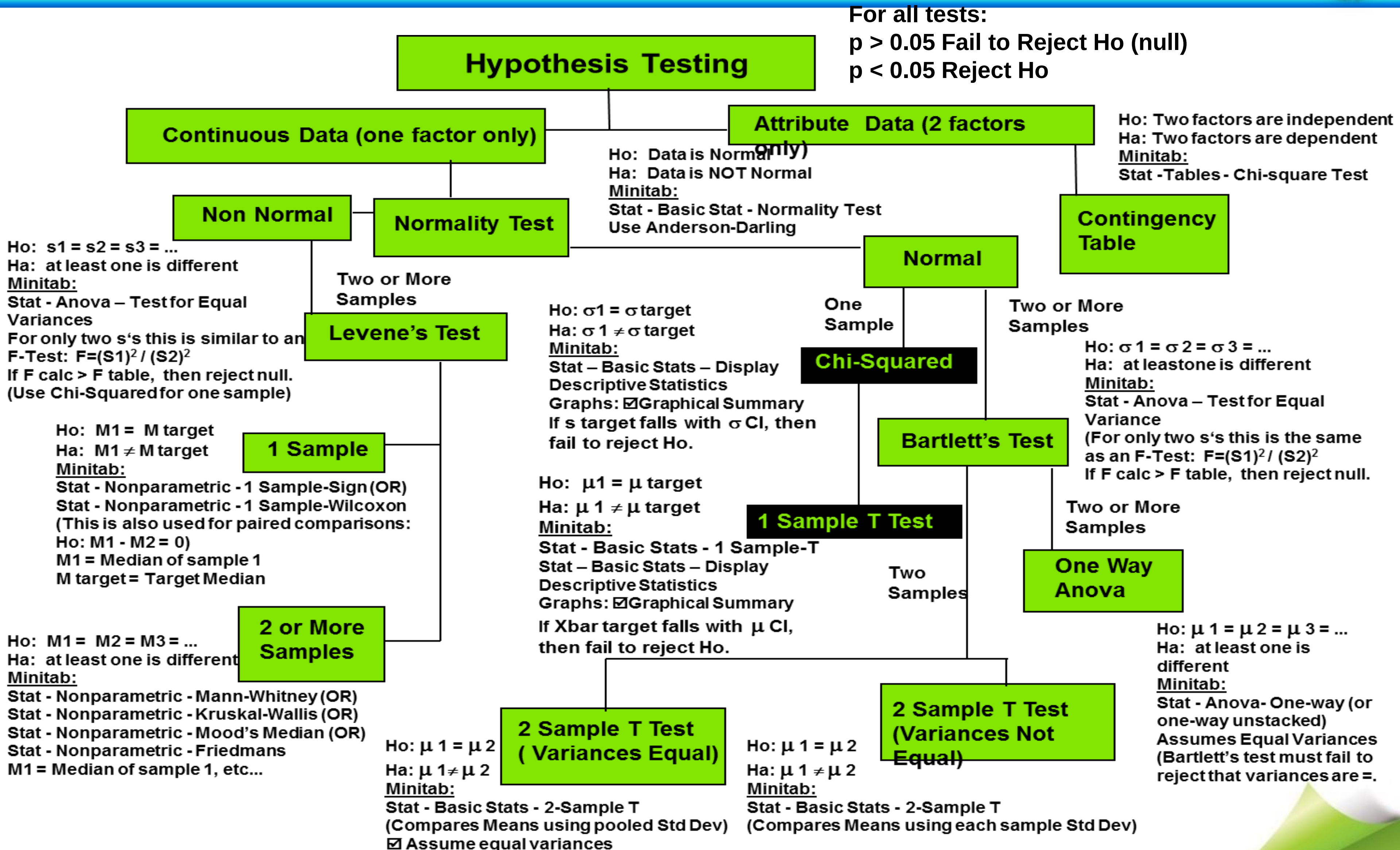



**Processus optimisé**

# Agenda

- **Test d’hypothèse pour 01 Echantillon**
  - Test pour la moyenne
    - Distribution Standard normale (Z) pour grande taille d’échantillons
    - t-distribution pour une taille réduite d’échantillons
  - Test pour la variance
  - Distribution Chi-carré
- **Test t pour données appariées**
- **Test d’hypothèse pour 02 Echantillons**
  - Test pour la variance
    - Test d’égailité des variances
    - F-distribution
  - Test pour la moyenne
    - Egalité des variances
    - Inégalité des variances

# Hypothesis Testing Roadmap




Fethi Derbeli . 2021

# Les étapes de la vérification d 'hypothèse:

1. Caractériser le problème et définir les objectifs
2. Elaborer les hypothèses
   - Enoncer l'hypothèse nulle (Ho)
   - Enoncer l'hypothèse alternative (Ha).
3. Décider du test statistique approprié (distribution de probabilité supposée, Z, t, $\chi^2$, F)
4. Indiquer le niveau Alpha (en général 5%)
5. Définir la taille de l'échantillon
6. Développer le plan d'échantillonnage
7. Effectuer la vérification et collecter les données
8. Calculer la statistique du test (z, t, ou F) à partir des données.
9. Déterminer la probabilité que se produise par hasard cette probabilité de test calculée = valeur-P .
   – Si p-value < $\alpha$, rejeter $H_o$
   – Si p-value > $\alpha$, accepter $H_o$
10. Reproduire les résultats et transposer la conclusion statistique en une solution pratique

Problème Pratique → Problème Statistique → Solution Statistique → Solution Practique

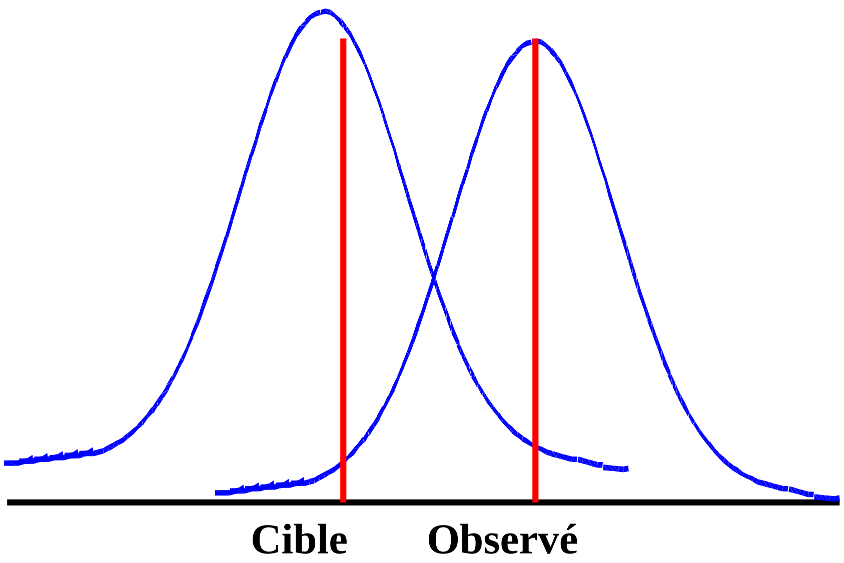
# Moyenne Population vs. valeur cible

- **Tester la différence entre les valeurs des moyennes demande des data variables et une distribution de référence**

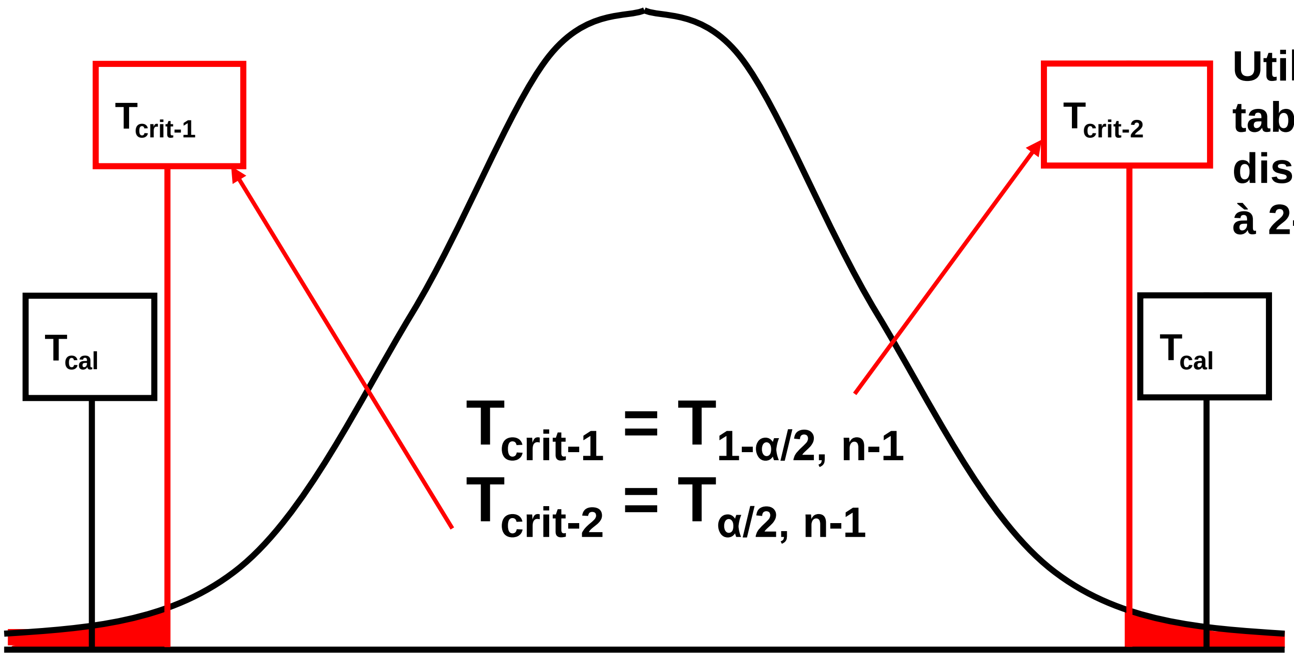
- **La différence entre la moyenne des échantillons et la cible est faite par le test statistique :“Test t pour 01 Echantillon”**

$$T_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

**Ou:** **n = taille échantillon**
**s = Ecart type**
**μ = Moyenne cible**



# $H_a$:  $\mu_{pop}$ # Cible

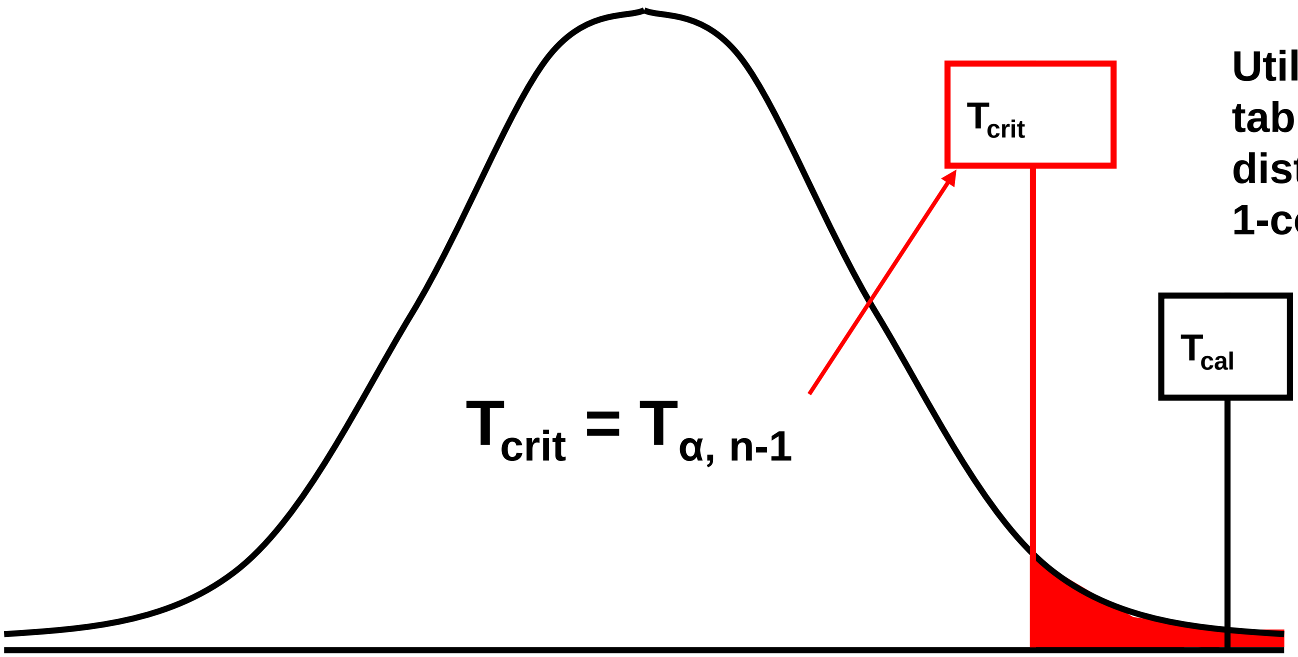



Utiliser la table de distribution à 2-cotés

**Rejet si $T_{cal} < T_{crit-1}$ ou $T_{cal} > T_{crit-2}$ …**

**… Puisque T-distribution est symétrique, on peut utiliser la valeur absolue  de $T_{cal}$ et comparer seulement à $T_{crit-2}$, rejetant $H_0$ si $T_{cal} > T_{crit-2}$**

# $H_a$: $\mu_{pop}$ > Cible

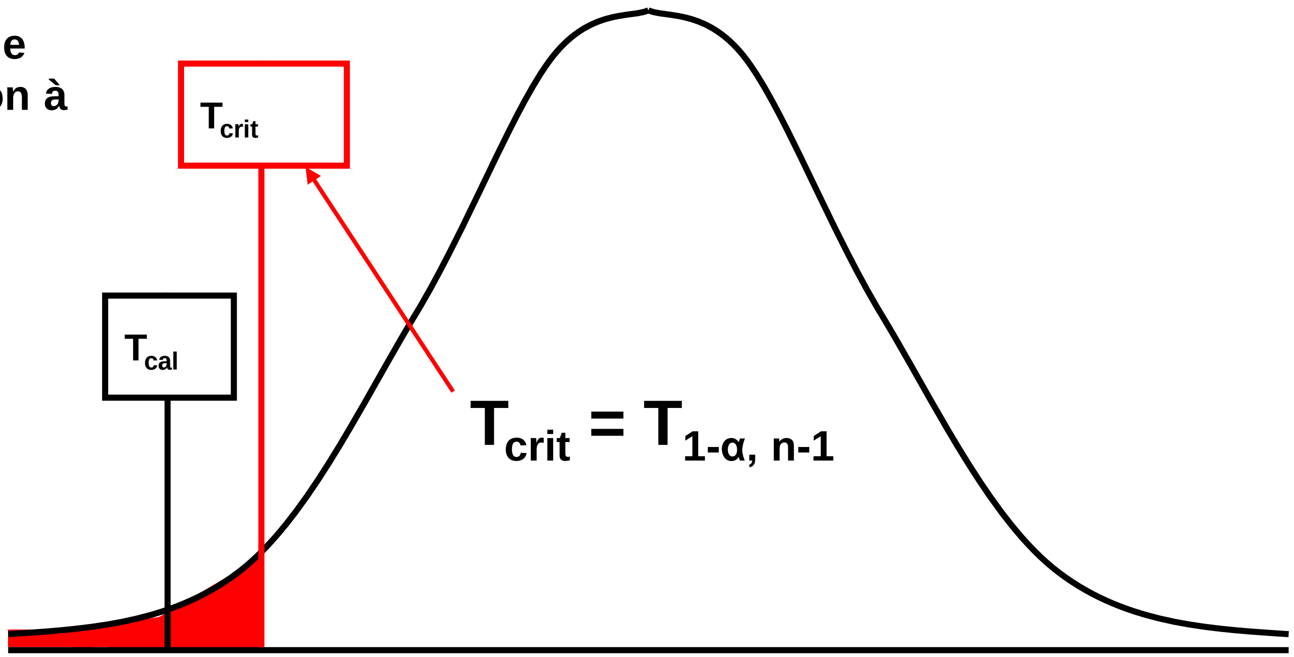



**Utiliser la table de distribution à 1-coté**

**Rejet $H_0$ si $T_{cal} > T_{crit}$**

# $H_a$: $\mu_{pop}$ < Cible

Utiliser la table de distribution à 1-cotés




**Rejet si $T_{cal} < T_{crit}$ …**

**Puisque t-distribution est symétrique on peut utiliser la valeur absolue de $T_{cal}$ et comparer à $T_{\alpha,\, n-1}$**

# Exemple

**Step 1:** Problème Pratique – Les impulsions totales optimale d'un moteur après le temps d'allumage (T0) de 0.5 seconds est 230 tr/sec. Un échantillon de 30 allumages prises donne Xbar = 226.95 et s = 8.63

**Step 2:** Statuer l'hypothèse nulle et l'hypothèse alternative

Pour $\mu$: $H_o: \mu_{population} = 230$

$H_a: \mu_{population} \neq 230$

**Step 3:** Le test statistique est une variable aléatoire normalement distribuée. Puisque la taille des échantillons est petite ($\leq 30$), la distribution –t est la référence appropriée.

**Step 4:** Le niveau de risque assumé (alpha) est 0.05 pour un test à deux cotés.

**Step 5:** Taille d'échantillon est 30.

**Step 6:** Développer un plan d'échantillonnage (aléatoire & représentatif).

**Step 7:** Collecter les data.

# Exemple (cont')

**Step 8:** **Calculer le test statistique approprié:**

**Xbar = 226.95 et s = 8.63**

$t_{calc} = \{(226.95-230)/(8.63 \div 5.5)\} = |-1.94|$

**Step 9:** **Trouver la valeur critique de la distribution approprié**

$t_{critical} = t_{\alpha/2,\ 29df} = 2.045$; **pour un à deux cotés**

**Puisque I-1.94I < 2.045, on échoue de rejeter $H_o$ parce que $t_{calculé} < t_{critique}$**

**p= 0,0634 > α =0,05**

**Step 10:** **Il n' y a pas de différence significative entre $\mu_{population}$ et $\mu_{cible}$, alors, on doit conclure que nous réalisons la valeur optimale.**

# Exemple (Intervalle de confiance )

**$Xbar - t_{\alpha/2} * \sigma/\sqrt{n} < \mu < Xbar + t_{\alpha/2} * \sigma/\sqrt{n}$**

**$t_{\alpha/2} = 2{,}045$ $n = 30$ $\sigma = 8{,}63$ $Xbar = 230$**

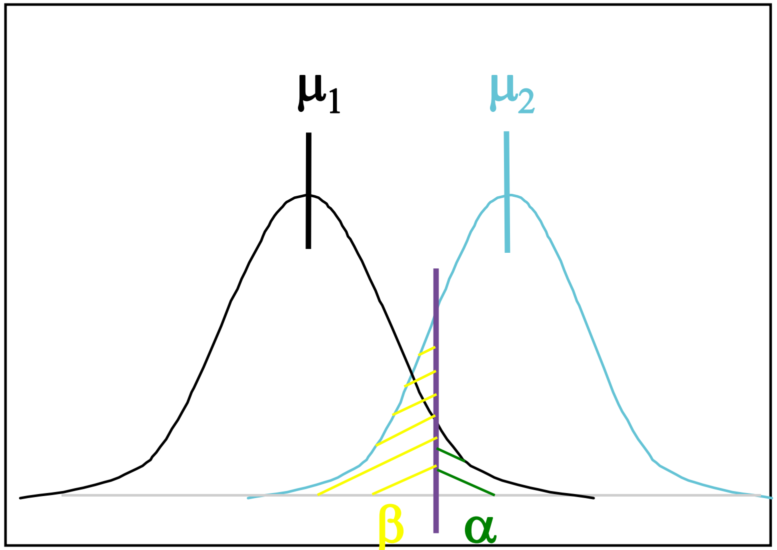
**$230 - 2{,}045*8{,}63/\sqrt{30} < \mu < 230 - 2{,}045*8{,}63/\sqrt{30}$**

**$226{,}78 < \mu < 233{,}22$**

**La nouvelle moyenne tombe dans l'intervalle de confiance, on peut conclure que le résultat observé est une conséquence du hasard et on échoue de rejeter l'hypothèse nul**

# Taille d'échantillon :1-Sample T test

- Elaborer les hypothèses
- Elaborer le test
- Déterminer alpha
- Déterminer beta
- Déterminer la sensibilité
- Définir la taille de l'échantillon




**α : est le risque de trouver une différence alors qu'il n'y en a pas.**

**β : est le risque de *ne pas* trouver de différence alors qu'il y en a une.**

**δ/s : est l'ampleur ou la taille de la différence qui est testée. Ceci s 'appelle parfois la sensibilité du test.**

# Taille d’échantillon 1-Sample T test

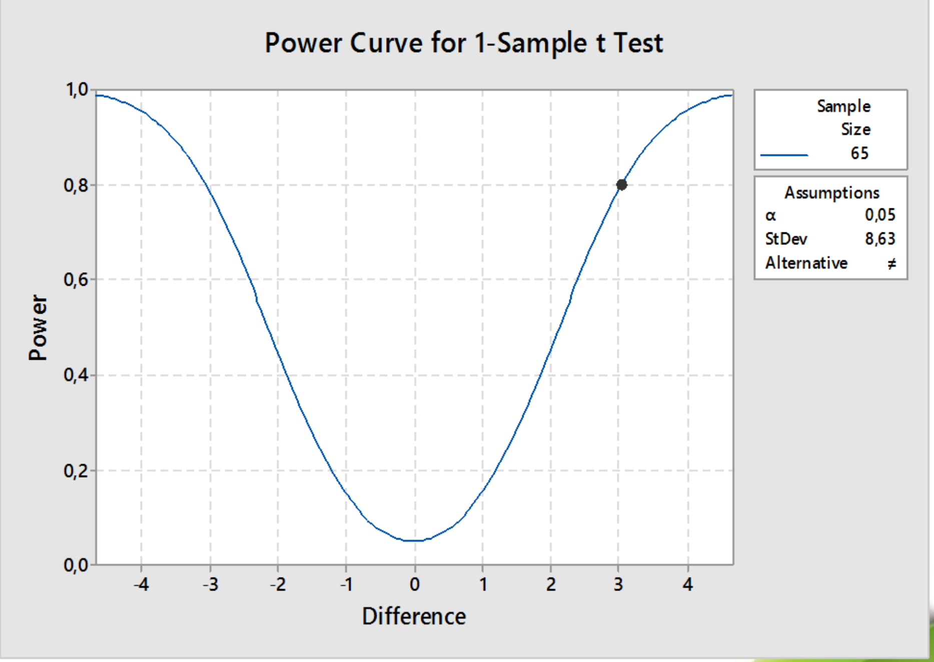
- **La formule:**

$$n = (t_{\alpha/2} + t_{\beta})^2 \frac{s^2}{\delta^2}$$

$$n = (Z_{\alpha/2} + Z_{\beta})^2 \frac{\sigma^2}{\delta^2}$$

- Choisir l’intervalle de confiance (alpha = 5%; β=20%)
- Choisir la différence minimum détectable = μ - xbar= 3,05
- Estimer la variabilité = s = 8,63
- nt=((2,045+0,854)*8,63/3,05))^2 = 67
- Nz= ((1,96+0,84)*8,63/3,05)^2 = 63

| Sample Difference | Target Size | Power | Actual Power |
|---|---|---|---|
| 3,05 | 65 | 0,8 | 0,801333 |



# Tests de la Variance

- **Rappelons la formule de l'intervalle de confiance pour σ**
  - Un échantillon statistique est une estimation d'une paramètre de la population
  - En prenant des échantillons de taille égale et on construisant des histogrammes des variances des échantillons crée une forme prédictible
  - Cette forme est appelée chi-carré et dépend de la taille des échantillons (n) et la valeur réelle de la variance, $\sigma^2$ de la population
- **La variable aléatoire chi-carré est définit:**

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

- **Cette statistique est utilisé pour définir le CI pour les variances.**

# Population $\sigma$ vs. cible $\sigma$

- **On a besoin d'une distribution $\chi^2$ et des data variables**
- **Cette statistique peut être utilisée pour tester si la variation d'une population est significativement différente d'une valeur cible.**
- **En d'autre mots:**

**"Y a t-il une différence significative entre $\sigma_{population}$ et $\sigma_{cible}$?"**

- **$\chi^2(calc) = (n – 1)s^2/ \sigma^2$**
  - Ou: n = Taille de l'échantillon
    $s^2$ = variance de l'échantillon
    $\sigma^2$ = variance cible

# Exercice

- **Rappelons l'équation:**
  - Cp = (USL – LSL) / 6s
- **Si nous sommes interressés de créer un processus six sigma, Cp = 2.0.**
- **On dévise les spécifications en 12 longueurs égaux (chaque s égale une longueur)**
- **Si notre objectif est différent de 6$\sigma$ process?**
  - Si on veut avoir une Cp de 1.67?
  - Si notre s objectif pour un 4$\sigma$ process?

# Exercice

**Step 1:** Problème pratique – L'Ingénieur qualité veut savoir si la variabilité actuelle du processus est au niveau six sigma.

- le USL du processus est 270, LSL est 190 et s = 8,63

**Step 1a:** Test des data.

- p >0,05 distribution normale

**Step 2:** Statuer l'hypothèse nulle et alternative.

- On doit résoudre l'équation Cp pour s.
- Cp >= 2 s = 6,67

Pour $\sigma$: $H_o$: $\sigma_{population} \leq 6.67$

$H_a$ $\sigma_{population} > 6.67$

**Step 3:** La distribution des échantillon est chi-carré

**Step 4:** On assume alpha est 0.05

**Step 5:** La taille de l'échantillon est 30

# Exercice (cont')

**Step 6:** On collecte aléatoirement 30 pièces de la production

**Step 7:** Mesurer et enregistrer les data

**Step 8:** Calculer le test statistique:

$$\chi_{calc}^{2} = \frac{(n-1)s^2}{\sigma^2} = \frac{(30-1)(8.63)^2}{(6.67)^2} = 48.6$$

**Step 9:** Trouver la valeur critique du table

- $\chi^2_{(critical)} = \chi^2_{0.05,\ 29df} = 42.56$
- Puisque $\chi^2_{(critical)} < \chi^2_{(calc)}$, on rejette l'hypothèse nulle

**Step 10:** Il y a une différence significative entre $\sigma_{population}$ et $\sigma_{cible}$; Le processus n'est capable à un niveau qualité six sigma dû à des variation excessive variation.

# Utilisation de l'intervalle de confiance

- On peut calculer le CI pour la variance et vérifier si la valeur cible se situe dans l'intervalle

- L'intervalle de confiance est généré via :

$$s\sqrt{\frac{n-1}{\chi^2_{\alpha/2}}} \leq \sigma \leq s\sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}}}$$

Ou :

$\alpha = 1 - \%$ confiance

$n =$ *taille'échantillon*

$8,63*\sqrt{(30-1)/45,72} \leq \sigma \leq 8,63*\sqrt{(30-1)/16,05}$

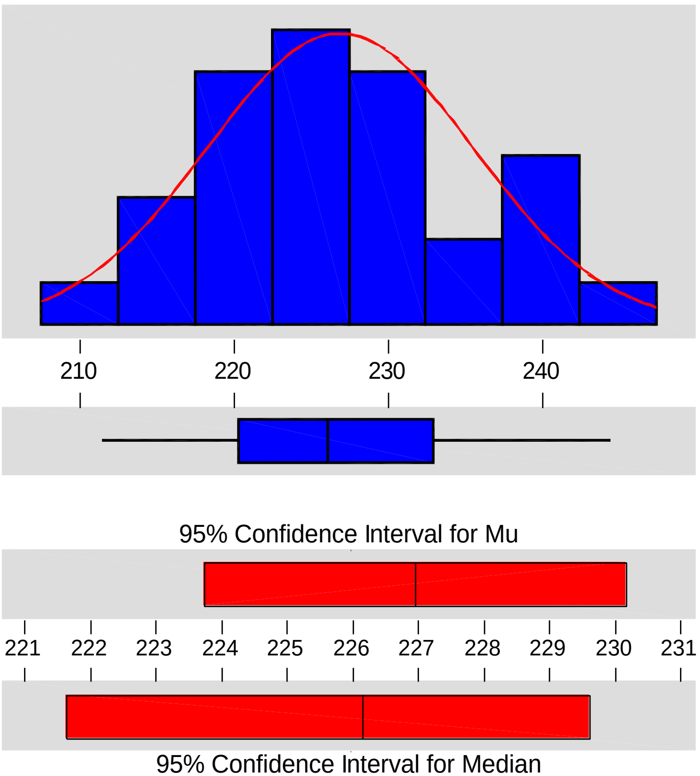
$6,87 \leq \sigma \leq 11,6$

$s_{cible} = 6.67$

- Ne se situe pas dans le CI 95% pour sigma
- Rejet hypothèse nulle

# Minitab Resume Graphique

## Descriptive Statistics




Variable: Total Impuls

Anderson-Darling Normality Test

| | |
|---|---|
| A-Squared: | 0.260 |
| P-Value: | 0.688 |
| Mean | 226.953 |
| StDev | 8.630 |
| Variance | 74.4798 |
| Skewness | 0.249783 |
| Kurtosis | -6.5E-01 |
| N | 30 |
| Minimum | 211.500 |
| 1st Quartile | 220.325 |
| Median | 226.150 |
| 3rd Quartile | 232.950 |
| Maximum | 244.500 |

95% Confidence Interval for Mu

| | |
|---|---|
| 223.731 | 230.176 |

95% Confidence Interval for Sigma

| | |
|---|---|
| 6.873 | 11.602 |

95% Confidence Interval for Median

| | |
|---|---|
| 221.640 | 229.600 |

**$s_{cible}$ = 6.67**

- **Ne se situe pas dans le CI 95% pour sigma**
- **Rejet hypothèse nulle**

# Test 02 échantillons

- **Pourquoi on vérifie des paramètres de 02 populations?**
- **Méthode Générale**
- **Variance population1 vs. Variance population2**
  - Manuellement: F-test
- **Moyenne Population1 vs. Moyenne Poulation2**
  - Manuellement: 2 sample t test
  - Options de l'hypothèse Alternative
  - Test de proportion

# Hypothesis Testing Roadmap

## Hypothesis Testing

### Continuous Data (one factor only)

### Attribute Data (2 factors only)

Ho: Two factors are independent
Ha: Two factors are dependent
Minitab:
Stat -Tables - Chi-square Test

**Contingency Table**

**Normality Test**

Ho: Data is Normal
Ha: Data is NOT Normal
Minitab:
Stat - Basic Stat - Normality Test
Use Anderson-Darling

**Non Normal**

Ho: s1 = s2 = s3 = ...
Ha: at least one is different
Minitab:
Stat - Anova – Test for Equal Variances
For only two s's this is similar to an F-Test: $F=(S1)^2 / (S2)^2$
If F calc > F table, then reject null.
(Use Chi-Squared for one sample)

Two or More Samples

**Levene's Test**

**Normal**

One Sample

**Chi-Squared**

Ho: $\sigma 1 = \sigma$ target
Ha: $\sigma 1 \neq \sigma$ target
Minitab:
Stat – Basic Stats – Display Descriptive Statistics
Graphs: ☑Graphical Summary
If s target falls with $\sigma$ CI, then fail to reject Ho.

**1 Sample T Test**

Two or More Samples

**Bartlett's Test**

Ho: $\sigma 1 = \sigma 2 = \sigma 3 = ...$
Ha: at least one is different
Minitab:
Stat - Anova – Test for Equal Variance
(For only two s's this is the same as an F-Test: $F=(S1)^2 / (S2)^2$
If F calc > F table, then reject null.

Two or More Samples

**1 Sample**

Ho: M1 = M target
Ha: M1 ≠ M target
Minitab:
Stat - Nonparametric - 1 Sample-Sign (OR)
Stat - Nonparametric - 1 Sample-Wilcoxon
(This is also used for paired comparisons:
Ho: M1 - M2 = 0)
M1 = Median of sample 1
M target = Target Median

**2 or More Samples**

Ho: M1 = M2 = M3 = ...
Ha: at least one is different
Minitab:
Stat - Nonparametric - Mann-Whitney (OR)
Stat - Nonparametric - Kruskal-Wallis (OR)
Stat - Nonparametric - Mood's Median (OR)
Stat - Nonparametric - Friedmans
M1 = Median of sample 1, etc...

**2 Sample T Test (Variances Equal)**

Ho: $\mu 1 = \mu 2$
Ha: $\mu 1 \neq \mu 2$
Minitab:
Stat - Basic Stats - 2-Sample T
(Compares Means using pooled Std Dev)
☑ Assume equal variances

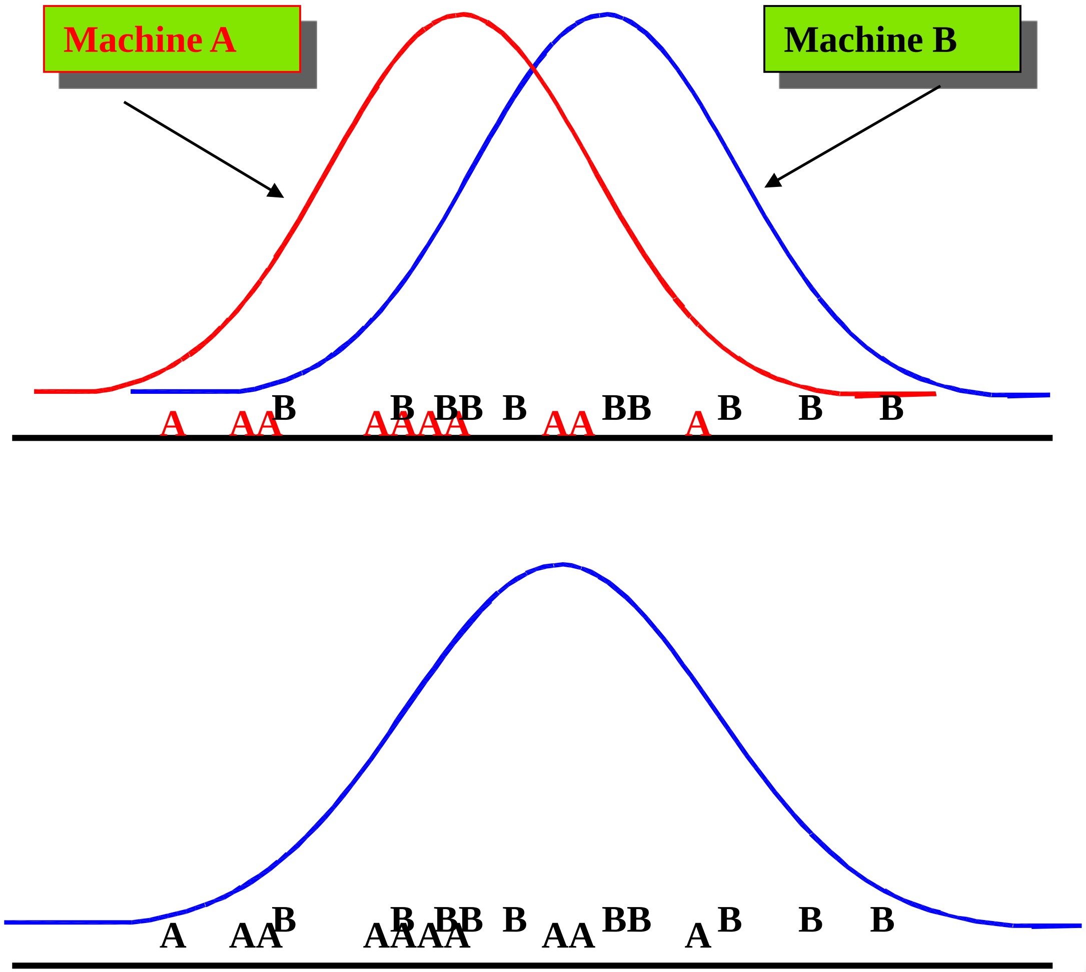
**2 Sample T Test (Variances Not Equal)**

Ho: $\mu 1 = \mu 2$
Ha: $\mu 1 \neq \mu 2$
Minitab:
Stat - Basic Stats - 2-Sample T
(Compares Means using each sample Std Dev)

Ho: $\mu 1 = \mu 2 = \mu 3 = ...$
Ha: at least one is different
Minitab:
Stat - Anova- One-way (or one-way unstacked)
Assumes Equal Variances
(Bartlett's test must fail to reject that variances are =.

# Pourquoi le Test d'hypothèse 02 échantillons?

- **Il y a plusieurs raisons de connaitre s'il y a une différence significative entre 02 populations.**
  - Evaluer le potentiel qualité d'un nouveau fournisseur
  - Savoir si une modification a généré des améliorations des performances
  - Quantifier la différences entre deux machines
  - Savoir si une modification du design d'un produit a apporté les améliorations attendues.




**"Y a t il ine différence significative entre A et B?"**

# Les étapes de la vérification d ’hypothèse:

1. statuer le problème pratique et définir les objectifs de l’amélioration

   a) Vérifier la normalité

1. Elaborer les hypothèses-

- Enoncer l’hypothèse nulle (Ho)
- Enoncer l’hypothèse alternative (Ha).

| For $\sigma$: | For $\mu$: |
|---|---|
| $H_o$: $\sigma_{pop1} = \sigma_{pop2}$ | $H_o$: $\mu_{pop1} = \mu_{pop2}$ (normal data) |
| $H_o$: M1 = M2 | (nonnormal data) |
| $H_a$: $\sigma_{pop1} \neq \sigma_{pop2}$ | Ha: $\mu_{pop1} \neq \mu_{pop2}$ (normal data) |
| $H_o$: M1 $\neq$ M2 | (nonnormal data) |

1. Décider du test statistique approprié (distribution de probabilité supposée, Z, t, $\chi^2$, F)

4. Indiquer le niveau Alpha (en général 5%)
5. Définir la taille de l’échantillon
6. Développer le plan d’échantillonnage
7. Effectuer la vérification et collecter les données
8. Calculer la statistique du test
   4. (z, t, ou F) à partir des données.
9. Déterminer la probabilité que se produise par hasard cette probabilité de test calculée = valeur-P .
   4. Si p-value < $\alpha$, rejeter $H_o$
   5. Si p-value > $\alpha$, accepter $H_o$
10. Reproduire les résultats et transposer la conclusion statistique en une solution pratique

# Population $\sigma_1$ vs Population $\sigma_2$

- **Lorsque on compare les moyennes de deux populations, utilisant des data variables on doit premièrement vérifier s'il y a une différence statistiques entre les variances (test d'égalité des variances). Ce test est important puisqu'il affecte la formule utilisée dans le test des moyennes.**

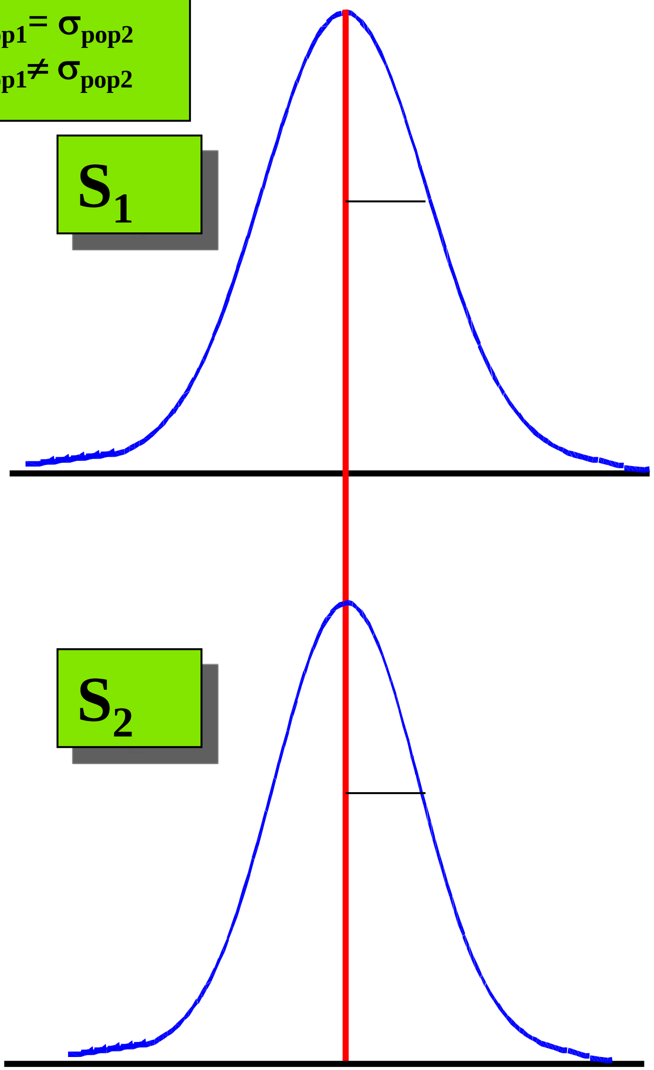
- **On a besoin aussi de savoir si la populations est normalement distribuée (test de normalité) pour déterminer le type de test de variance à effectuer.**

- **Les résultats de test de normalité et le test de variance vont déterminer la supposition adéquate pour l'analyse des tendances centrales de la population (moyennes vs. médianes avec ou sans égalité des variances)**

- **Le test exige l'utilisation du test statistique -F**
  - Le test statistique-F peut être utilisé pour comparer 02 ou plusieurs variances. (Le test F est utilisé pour 02 échantillons.) Ce test assume que les data sont normalement distribués.
  - Test statistique du Levene est utilisé pour comparer 2 ou plusieurs variances et il est approprié pour les données continues mais peuvent être non normalement distribuées.

# Test d’égalité des Variances

$H_o$: $\sigma_{pop1} = \sigma_{pop2}$
$H_a$: $\sigma_{pop1} \neq \sigma_{pop2}$

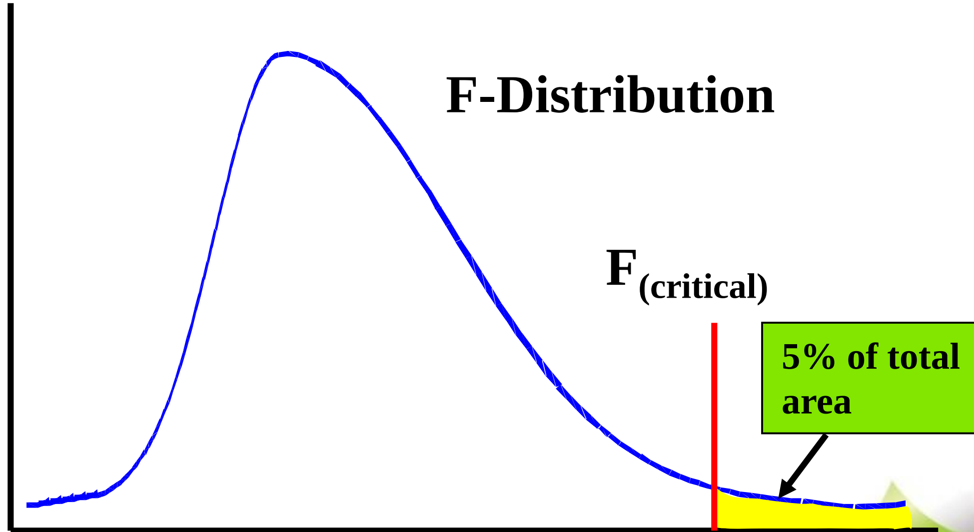



a) Calculer $F_{calc}$ sachant que $F_{calc} > 1$

$F_{calc} = s_2^2 / s_1^2$ ou $F_{calc} = s_1^2 / s_2^2$

b) Comparer $F_{crit}$ (pour $\alpha = 0.05$)

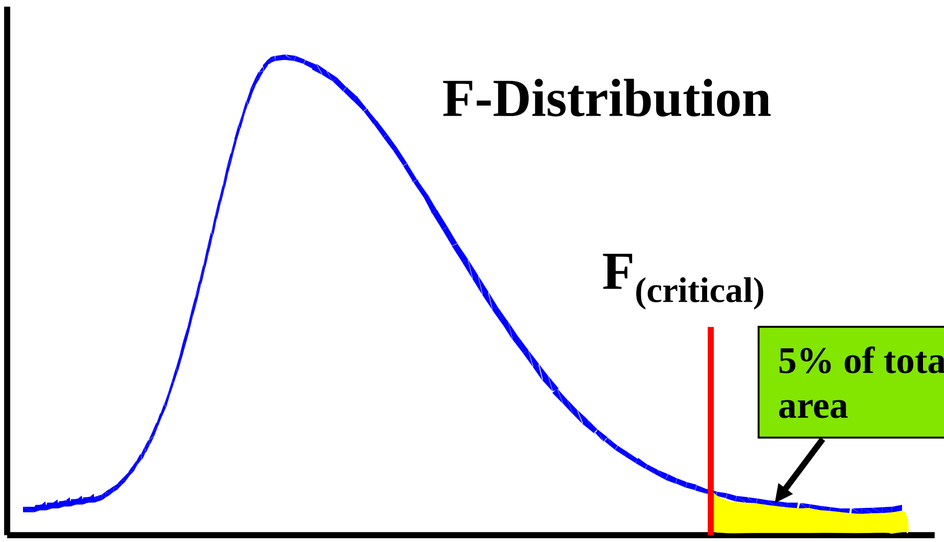
c) Si $F_{calc} > F_{crit}$, on rejete $H_o$



# C’est quoi $F_{(critique)}$?

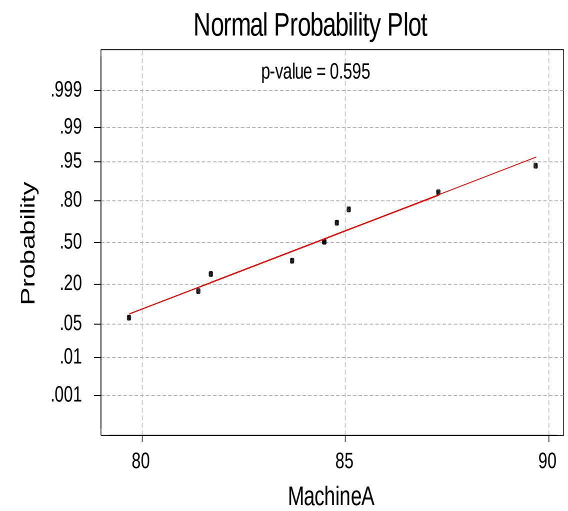
- Si on veut savoir la valeur critique, on peut l’avoir à partir du table F. la valeur dépend de la valeur alpha, et le degré de liberté des deux facteurs (numérateur and dénominateur).

- Quand la valeur F calculée dépasse la valeur F critique ( $\alpha = 0.05$), la valeur p va être plus petit que 0.05. implique que les deux variances sont différentes (p is low $H_0$ must go).



# Exercice

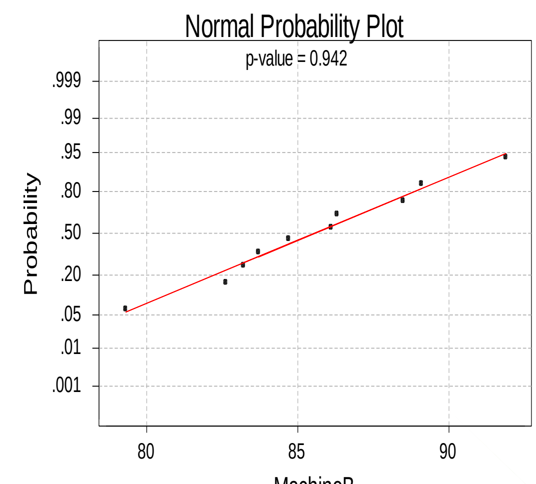
**Step 1: Problème Pratique – On a modifié une de deux machines. On veut savoir s'il y a une amélioration significative du rendement de la machine modifiée avant de la généraliser sur les autres machines. Après avoir collecté des échantillons sur les rendements des 2 machines. On veut déterminer s'il y a une vraie différence entre les deux?**




**Premièrement on vérifie si les data sont normalement distribuées.**

**Descriptive Statistics: MachineA, MachineB**

| Variable | N | Mean | Median | StDev | SE Mean |
|---|---|---|---|---|---|
| MachineA | 10 | 84.240 | 84.500 | 2.902 | 0.918 |
| MachineB | 10 | 85.54 | 85.40 | 3.65 | 1.15 |



# Exercice (cont')

**Step 2:** Vérifier l'égalité des variances. Statuer l'hypotheses nul et alternatif.

Ho: $\sigma_{MachineA} = \sigma_{MachineB}$

Ha: $\sigma_{MachineA} \neq \sigma_{MachineB}$

**Step 3:** Les variables aléatoires du test de l'égalité des variances suivent une distribution-F

**Step 4:** On assume le niveau de risque (alpha) de 0.05

**Step 5:** La taille d'échantillon est 10

**Step 6:** Collecter 10 échantillons aléatoires pour chaque machine

**Step 7:** Les data

# Exercice (cont')

- **Step 8: Calculer les paramètres F**
  - $F_{calc} = s_B^2 / s_A^2 = (3.65)^2 / (2.902)^2 = 1.5819$
- **Step 9: Déterminer la valeur critique du table F**
  - $F_{crit} = F_{.05,9,9} = 3.18$
- **Step 10: Puisque 3.18 > 1.5819, on échoue de rejeter $H_o$**

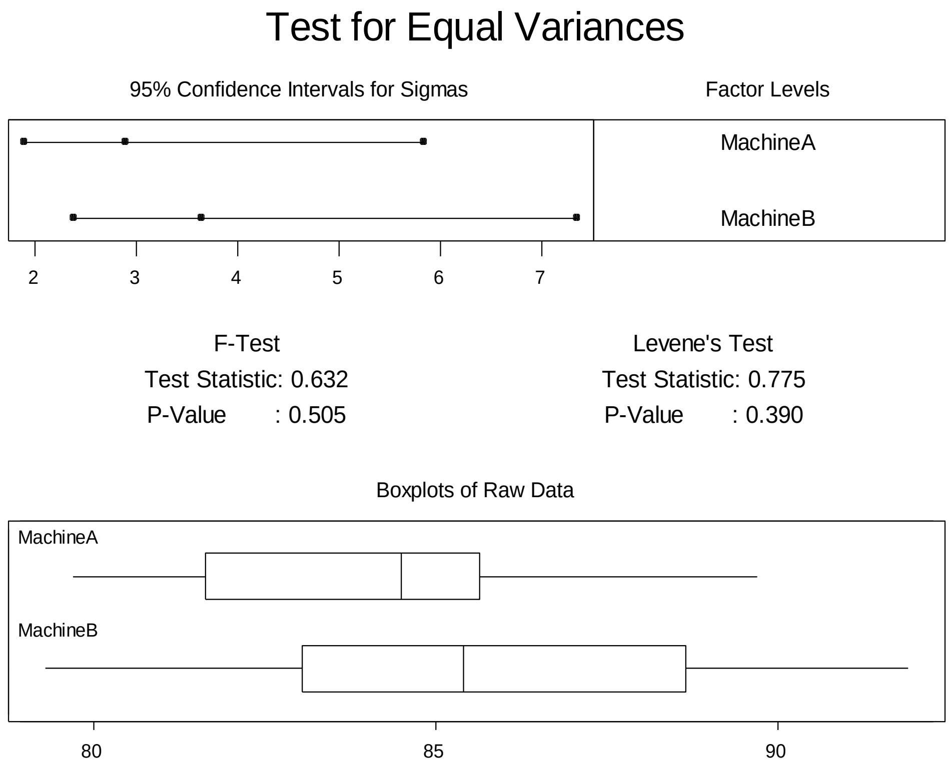
**"Il n'y a pas de différence significative entre $s^2_{machineA}$ et $s^2_{machineB}$, les variances sont supposées égaux."**

**Solution pratique: Eviter de modifier le reste des machines**

# Minitab: Test d'égalité des Variances

- Si la normalité est confirmée on utilise le test statistique de Bartlett (Minitab)

- Si la distribution n'et pas normale, on utilise test statistique de Levene (Minitab)




**Minitab:**
- Stat > ANOVA > Test for Equal Variances (stacked data)
- Stat > Basic Statistics > 2 Variances (stacked OR unstacked data)

Ho: $\sigma_{pop1} = \sigma_{pop2}$

Ha: $\sigma_{pop1} \neq \sigma_{pop2}$

# Moyenne Population-1 vs Moyenne Population-2

- **On a besoin d'une t-distribution et data variables**

- **On veut vérifier si les échantillons viennent de la même ou de deux distinctes distributions.**

- **D'autre mots:**

**"Y a t'il une différence significative entre $\mu_{pop1}$ et $\mu_{pop2}$?"**

- **Pour le test des deux échantillons "t test" est:**
  - Egalité des variances (degrees of freedom = n1 + n2 – 2):

$$t_{calc} = \frac{\bar{x}_{pop1} - \bar{x}_{pop2}}{\sqrt{s^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$s_p^2 = \frac{(n_1 - 1)\ s_1^2 + (n_2 - 1)\ s_2^2}{n_1 + n_2 - 2}$$

  - Inégalité des variances (degrees of freedom = n1 + n2 – 2):

$$t_{calc} = \frac{\bar{x}_{pop1} - \bar{x}_{pop2}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

# Exemple

**Step 1:** **Problème Pratique – On a modifié une de deux machines. On veut savoir s'il y a une amélioration significative du rendement de la machine modifiée avant de la généraliser sur les autres machines. Après avoir collecté des échantillons sur les rendements des 2 machines. On veut déterminer s'il y a une vraie différence entre les deux?**

| | N | Mean | StDev | SE Mean | |
|---|---|---|---|---|---|
| Mach A | 10 | 83,70 | 2,95 | 0,93 | Pooled σ= 3,29 |
| Mach B | 10 | 85,10 | 3,60 | 1,1 | |

- **Rappel: Les data sont normales et les variances sont égales.**

**Step 2:** **Déterminer l'hypothèse nul and alternatif pour l'amélioration des rendements.**

$$H_o: \mu_{machineA} = \mu_{machineB}$$

$$H_a: \mu_{machineA} < \mu_{machineB}$$

**Step 3:** **Les variables aléatoires pour le test des moyennes suivent la distribution-t.**

# Exemple (cont')

**Step 4-7: Alpha est 0.05, taille des échantillons est de 10, Les échantillons aléatoires sont  collectés de chaque machine.**

**Steps 8-9: test-t de 02 échantillons**

$T_{calc}= (85,1-83,7)/(\sqrt{3,29^2*(1/10+1/10)})$

$T_{calc}= 1,4/1,47 \ =0,95$

$T_{critic}= 1,833$

- $T_{calc} < T_{critic}$ ⟶ $P>\alpha$   on échoue à rejeter $H_0$

**Step 10: la différence des rendements des deux machines n'est pas statistiquement significative et dû au hasard. Ne peux généraliser la solution et chercher d'autres alternatives**

# Test t pour données appariées

- **Définition comparaison par paire:**

- **Chaque données est mesurée 02 fois (01 pour chaque machine dans les mêmes conditions).**

- **“Blocking” est utilisé pour minimiser l’excès de variabilité causé par la dépendance des échantillons. Règle: Bloquer ce qui tu peux, et échantillonner en mode aléatoire ce qui tu ne peut pas.**

$$t = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$$

**Intervalle de confiance**

$$\bar{d} \pm t_{\alpha/2}\left(s_d / \sqrt{n}\right)$$

**$\mu_0$** moyenne hypothétisée de la population pour les différences

**sd** écart type de l'échantillon des différentes d'échantillons appariés

**n** effectif d'échantillon

**$d$ $x_1 - x_2$ et $x_1$ et $x_2$ sont des observations appariées des populations 1 et 2, respectivement**

$$sd = \sqrt{\frac{\sum (d_i - \bar{d})^2}{(n-1)}}$$

$$\bar{d} = \sum d / n$$

# Test t pour données appariées: Exemple

- **Problème: Deux machines (A & B) produisant le même produit avec des performances similaires, une modification est apportée sur la machine B on veut savoir si cette modification a généré une différence significative par rapport à la machine A**

  **Une collecte des données est effectué dans les mêmes conditions (Lot matière, horaires du travail, équipe du travail etc..)**

| | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| **Machine A** | 24 | 96,875% | 0.02252 | 0.00460 |
| **Machine B** | 24 | 95,667% | 0.02160 | 0.00441 |
| **Difference** | 24 | 1,208% | 0.01956 | 0.00399 |

# Test t pour données appariées: Exemple (cont')

**Step 1:** **Problème Pratique – Est-ce que la modification sur la machine b a généré une amélioration significative du rendement comparé à la machine A**

**Test de Normalité des deux situations avec $p > 0.05$**

**Step 2:** **Statuer l'hypothèses:**

**Ho: $\delta = 0$ ($\delta$ - delta or change; $\mu A - \mu B = 0$)**

**Ha: $\delta \neq 0$**

**Steps 3-9:** **Déterminer le test statistique, déterminer la valeur critique, et rejeter ou accepter l'hypothèse nul.**

Tcal = (0,01208-0)/(0,01956/($\sqrt{24}$) =3,02

Tcrit = 2,07 P<α

95% CI for mean difference: (0,0038; 0,0203)

**Step 10:** **La modification a généré une amélioration significative**

# Exemple – L'Analyse erroné

- On va voir la même analyse des données avec une méthode différente test 02 échantillons-t:

| | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| Machine A | 24 | 96,875% | 0.02252 | 0.00460 |
| Machine B | 24 | 95,667% | 0.02160 | 0.00441 |
| Difference | 24 | 1,208% | 0.01956 | 0.00399 |

```
Sample    N    Mean    StDev   SE Mean
1        24  0,9688  0,0226   0,0046
2        24  0,9567  0,0216   0,0044


Difference = µ (1) - µ (2)
Estimate for difference:   0,01208
95% CI for difference:   (-0,00077; 0,02493)
T-Test of difference = 0 (vs ≠): T-Value = 1,89  P-Value = 0,065  DF = 45
```

**Un test est significative, l'autre n'est pas. pourquoi?**

# Considérations finales

## Echantillons par paires contre échantillons indépendants

- Les échantillons par paires contre des échantillons indépendants peuvent être difficiles à juger. La distinction est néanmoins cruciale puisque des procédures statistiques différentes sont applicables dans chaque cas.

  Par paires: en général dans le même sujet.
  Exemple: usure des pneus de deux marques sur la même voiture en même temps.

  Indépendant: nécessite des échantillons indépendants prélevés au hasard.
  Exemple: mesurer l'usure de deux marques de pneus, une marque sur la voiture A et une autre marque sur la voiture B.

## Variances égales

- Dans le cas du test-t à deux échantillons. En général, si l'égalité est supposée à tort, on peut se tromper de beaucoup lorsqu'on estime la différence des moyennes.
- Si l'on suppose l'inégalité alors que les variances sont en fait égales, on obtient une approche légèrement conservatrice et une faible partie de la précision est perdue lors de l'estimation.

## Echantillonnage aléatoire

- En général, pour les tests-t, on suppose que nous collectons des données à partir d'échantillons au hasard dans une distribution normale.
- Même si la distribution n'est pas normale, la distribution t donne de bonnes approximations, tant que l'échantillon est « prélevé de façon aléatoire ».
- Obtenir un échantillon qui ne soit pas aléatoire est plus problématique que les données non distribuées normalement.

# Résumé

**T 1-Echantillon**

- **But: Analyser la différence entre la moyenne obtenue et une valeur cible ou une moyenne passée.**
- **Caractéristiques clés:A utiliser lorsqu'il n'y a qu'un seul échantillon pour tirer des conclusions sur la moyenne de l'échantillon.**

**T 2- Echantillons**

- **But: Pour analyser la différence entre les moyennes obtenues de deux échantillons indépendants.**
- **Caractéristiques clés : à utiliser lorsqu'on a deux échantillons indépendants. Les variances peuvent être égales ou inégales.**

  **Les données peuvent être dans une seule colonne avec groupement variable ou dans deux colonnes différentes.**

**T par paires**

- **But: Analyser la différence entre les moyennes obtenues de deux échantillons liés.**
- **Caractéristiques clés :Utilisé avec les données par paires (souvent deux mesures du même sujet ou du même objet testé) l'ordre des observations est important. Chaque observation doit être dans la même rangée que celle qui lui correspond.**

# Résumé

**Test à 1 variance**

**Utiliser pour estimer la variance ou l'écart type d'une population, et la/le comparer à une valeur cible ou de référence. Vous pouvez ainsi :**

- **Déterminer si la variance ou l'écart type d'un groupe diffère d'une valeur spécifiée.**
- **Calculer une étendue de valeurs ayant de bonnes chances de contenir la variance ou l'écart type de la population.**

**Test à 2 variances**

- **Déterminer si les variances ou les écarts types de deux groupes diffèrent.**
- **Calculer une étendue de valeurs ayant de bonnes chances de contenir le rapport des variances ou des écarts types de la population des deux groupes.**