PrivateTeacher

Cours Privés de Science

p-value Interprétation

Sommaire

Énoncé

Dans une grande entreprise américaine, le salaire annuel moyen des hommes possédant entre 3 et 5 années d'expérience est de 58'000$. Les salaires (en milliers de dollars) d'un échantillon aléatoire de 10 femmes possédant entre 3 et 5 ans d'expérience sont les suivants :

54 57 61 51 49 56 60 52 45 66

Voici les résultats d'un test :

```
        One Sample t-test

data: x
t = -1.4729, df = 9, p-value = 0.08744
alternative hypothesis : true mean is less than 58
95 percent confidence interval :
        -Inf 58.70926

sample estimates :
mean of x
        55.1
```

Y a-t-il suffisamment d'évidence pour attester que les femmes gagnent moins que les hommes en prenant un risque de 0.05 ?

Réponse

La valeur-p est supérieure au risque de première espèce fixé à 5%. Nous ne pouvons donc pas rejeter **l'hypothèse nulle** qui indique que la moyenne du salaire des femmes est **supérieure ou égale à 58**. La réponse correcte est donc : Non, ces données ne fournissent pas suffisamment d'évidence pour attester que les femmes gagnent moins que les hommes.

Statistique de test

Dans cet exercice, on cherche à comparer le salaire moyen des hommes μ_h avec le salaire de quelques femmes choisies au hasard $\hat{\mu}_f$

$\hat{\mu}_f$ est un <u>estimateur</u> car il provient d'un <u>échantillon</u> de femmes.

Ici, on cherche à <u>comparer</u> la moyenne d'une population μ_h avec la moyenne estimée à partir d'un échantillon $\hat{\mu}_f$

A cette fin, on utilise un nombre qui a été imaginé exprès pour ça.
Il s'agit de la statistique t qui mesure l'écart entre deux moyennes, l'une réelle μ_h, l'autre estimée $\hat{\mu}_f$:

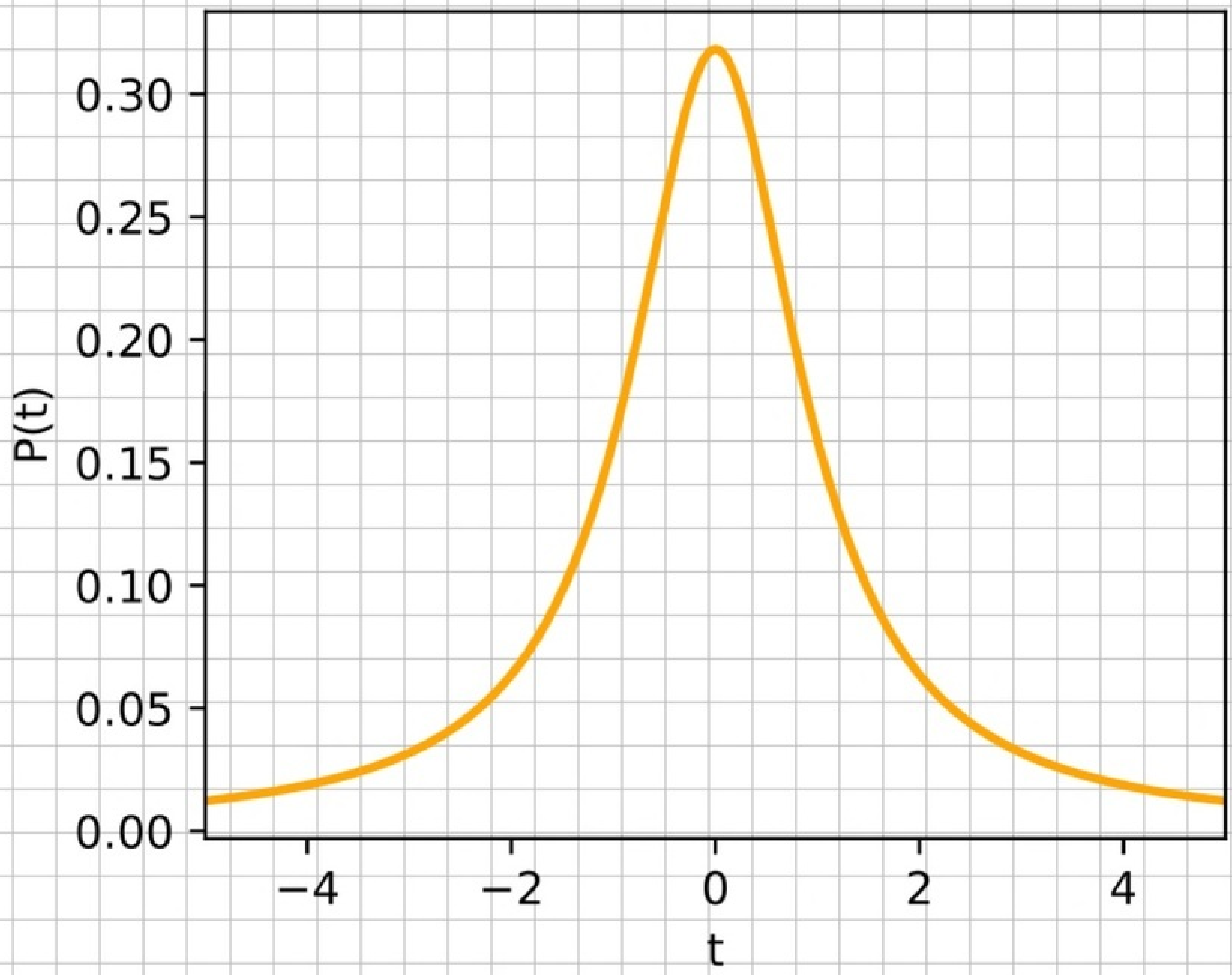
$$t = \frac{\hat{\mu}_f - \mu_h}{\sigma_f / \sqrt{n_f}}$$

<u>La statistique de test "t" mesure l'écart entre deux moyennes</u>

Loi de Student

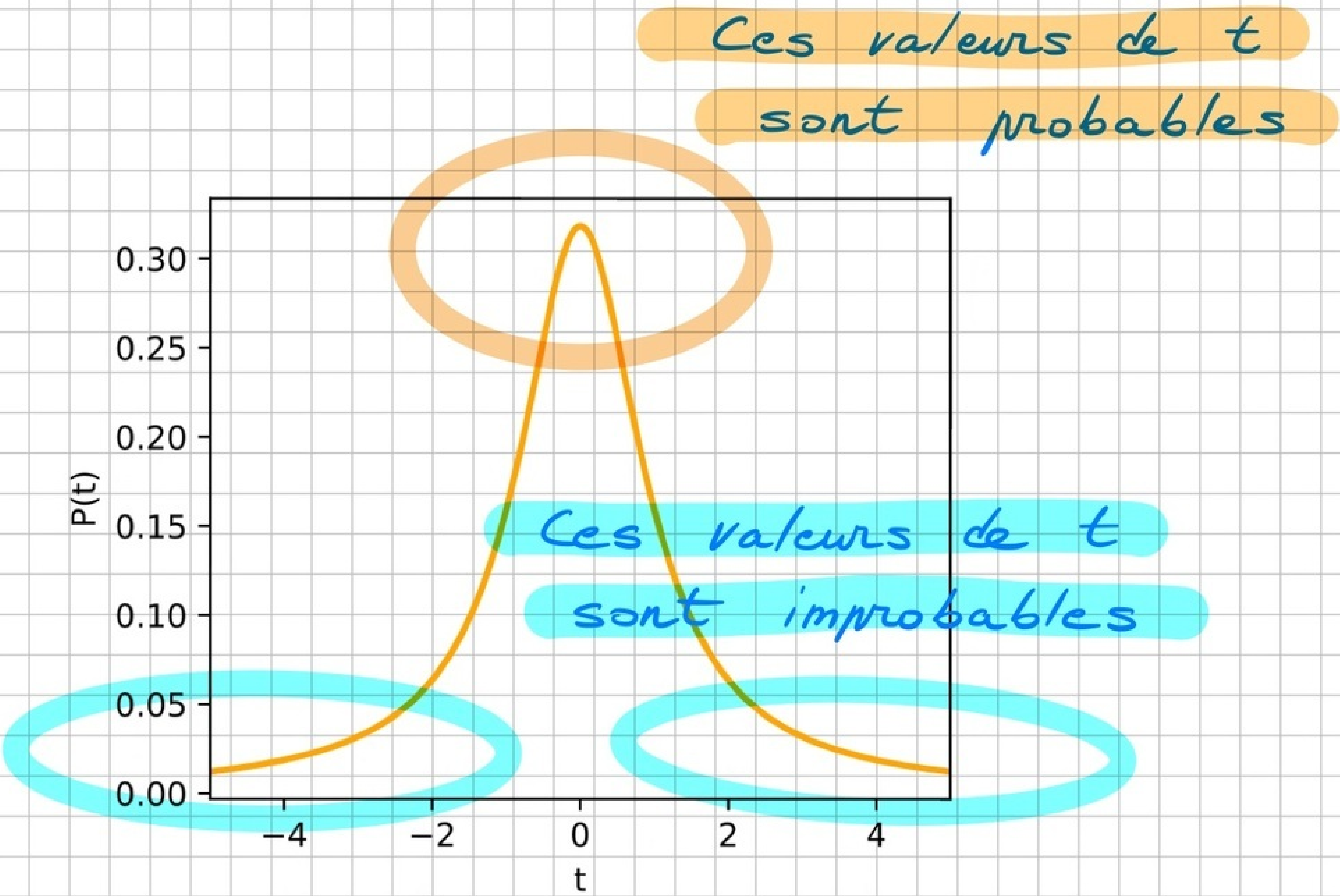
Il se trouve que la statistique t suit une distribution dite "loi de Student"

La loi de Student a la forme suivante :



Cette distribution n'est pas une loi normale (gaussienne) Bien qu'elle soit, elle aussi, symmétrique par rapport à l'axe $x = 0$, la loi de Student est plus épaisse aux extrémités.

L'intérêt de connaître la distribution de probabilité de la statistique de test "t" est de nous permettre de connaître la probabilité d'observer chacune des valeurs possible de t.



Tout le test d'hypothèse repose sur ce graphique

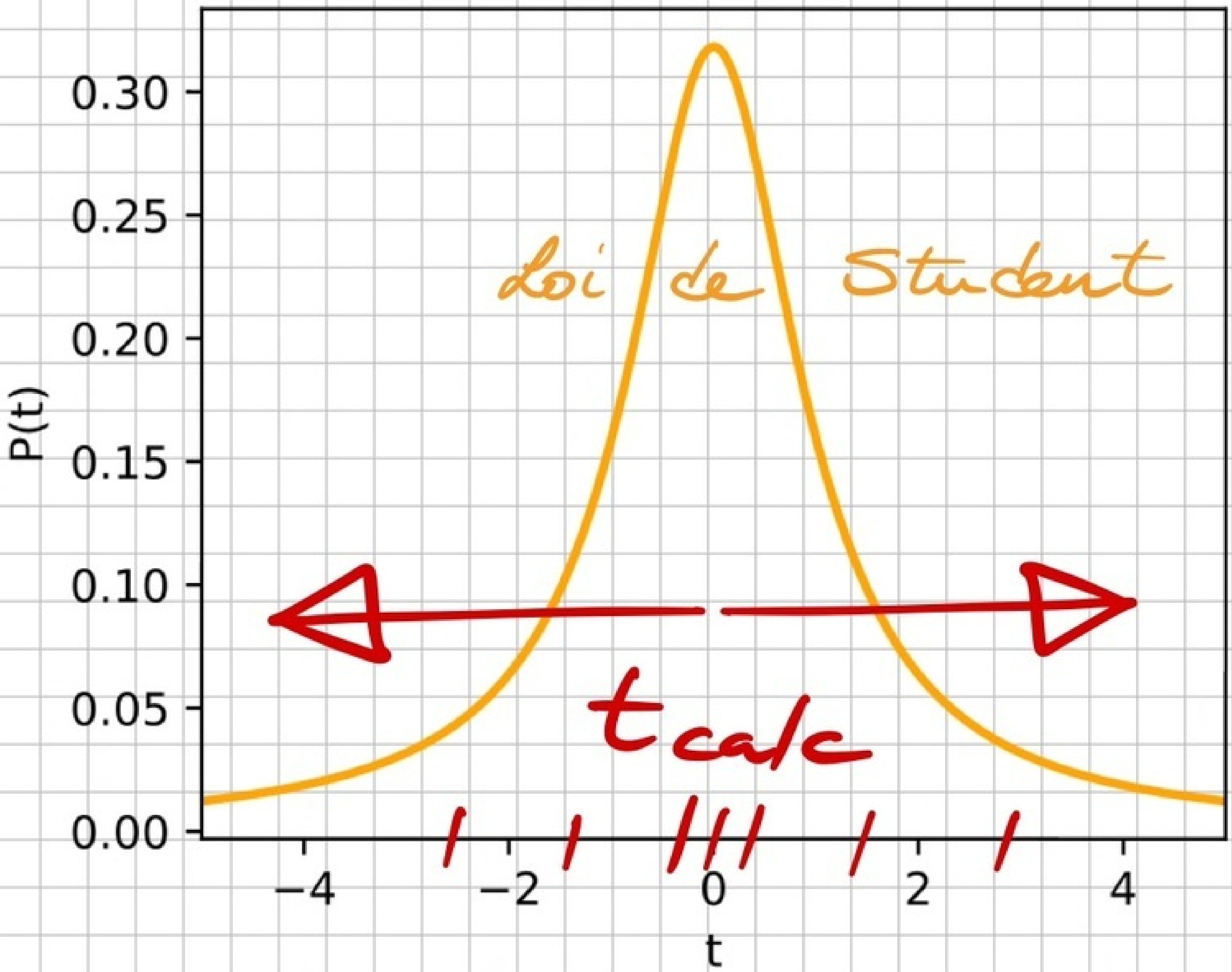
Voyons comment le logiciel l'utilise pour calculer le résultat...

Valeur calculée

Le logiciel de statistique commence par calculer la valeur de t dans notre cas particulier. On note cette valeur t_{calc}

la statistique de test calculée t_{calc} prendra place quelque part sur la distribution de Student :

Probabilité d'observer une valeur de t en particulier.



toutes les valeurs possibles de la statistique t

Test d'hypothèse

On le voit, la valeur de t la plus probable est 0. Cette valeur correspond à l'hypothèse nulle H_0

Elle signifie qu'il n'y a pas de différence entre les deux moyennes.

On note : $\boxed{H_0 : \mu_h = \hat{\mu}_f}$

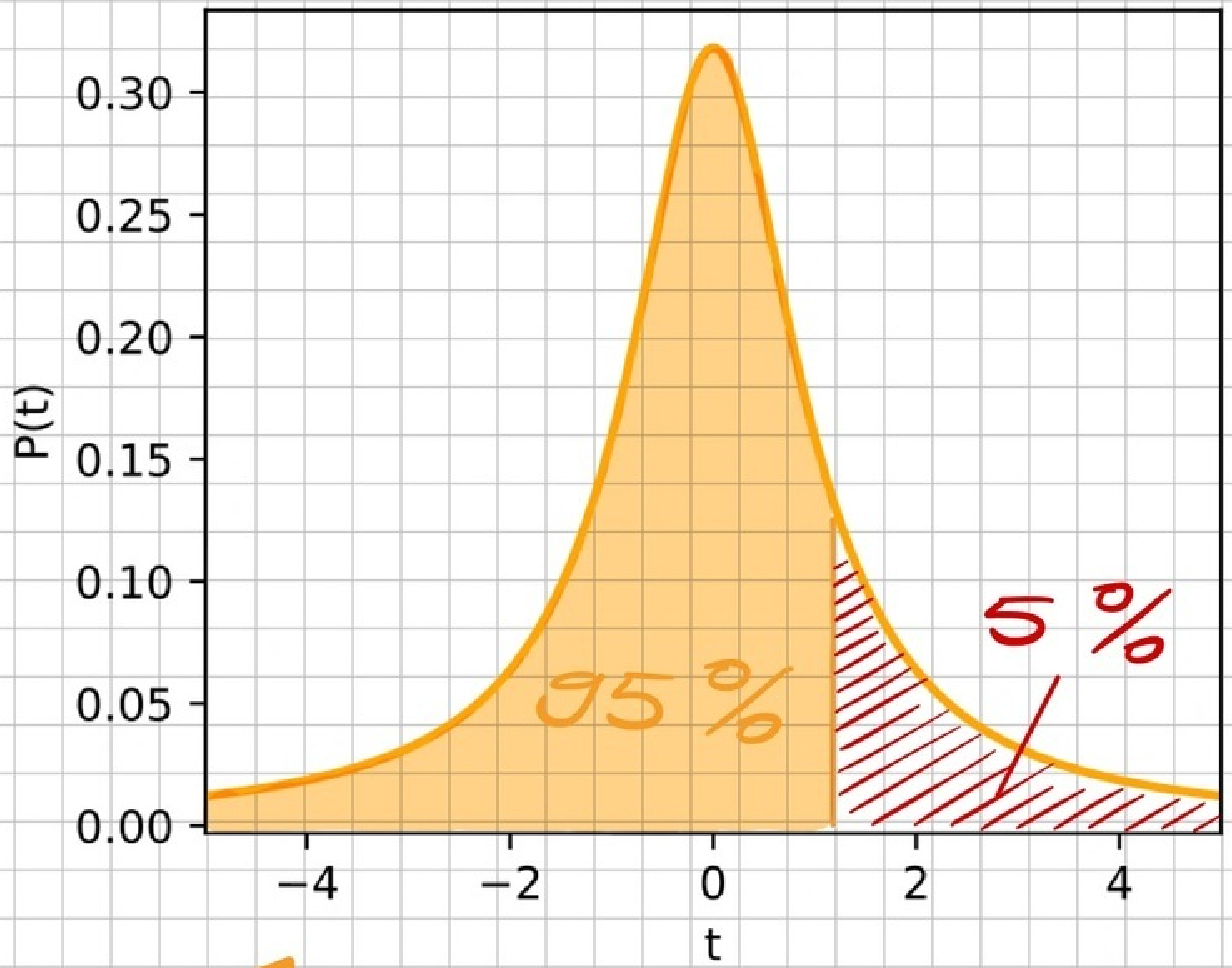
Un test d'hypothèse consiste à situer la valeur calculée t_{calc} au sein de la distribution de Student afin de déterminer si cette valeur est probable ou pas

Si la valeur calculée de la statistique de test t_{calc} est probable, alors on détient un argument en faveur de H_0

Si la valeur calculée de la statistique de test t_{calc} est improbable, alors on détient un argument en faveur de H_1

Seuil de significativité

C'est à nous de fixer le seuil de ce que l'on considère comme "probable". On choisit souvent 95 %



Ces valeurs de t sont probables

$\Leftrightarrow$

H_0 est raisonnable

Zone d'acceptation

Ces valeurs de t sont improbables

$\Leftrightarrow$

H_0 est déraisonnable

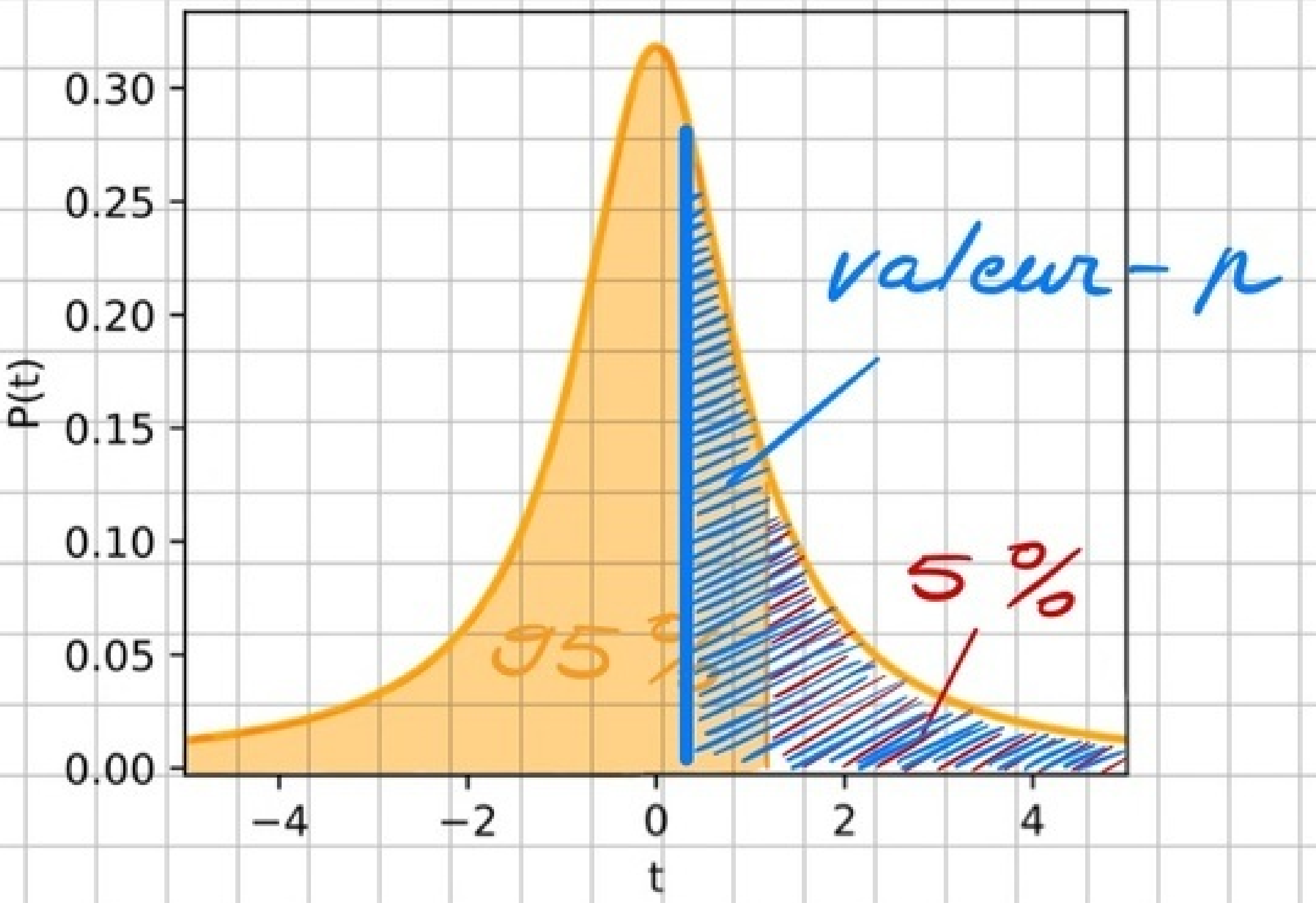
Zone de rejet

La valeur-p

Une fois le seuil fixé, le logiciel calcul la probabilité d'observer une valeur de t plus grande ou égale à la valeur de t calculée t_{calc}

Cette probabilité se nomme : valeur-p

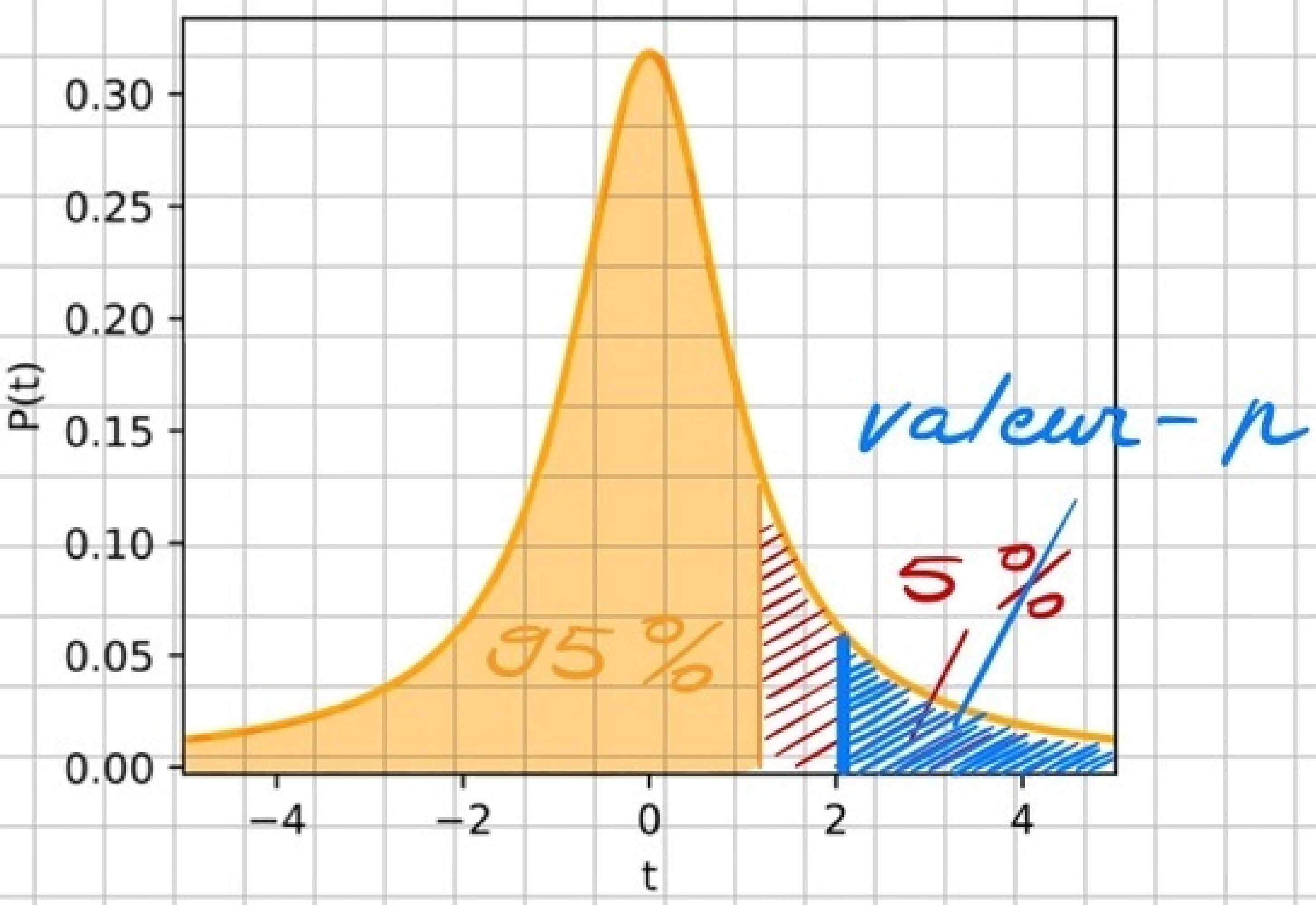
$$\boxed{\text{valeur-p} = P(t \geq t_{calc})}$$



Valeur-p > 5 %

=▷ t_{calc} probable

=▷ H_0 raisonnable



Valeur-p < 5 %

=▷ t_{calc} peu probable

=▷ H_0 déraisonnable

Interprétation

Ainsi lorsque le logiciel nous retourne
p-value = 0.087 (> 0.05)
cela signifie que la valeur calculée de t, $t_{calc} = -1,47$ est <u>probable</u>.

Il s'agit donc d'une indication <u>en faveur de</u> H_0

$$H_0 : \hat{\mu}_f = \mu_h$$

$$\Leftrightarrow$$

l'estimation du salaire moyen des femmes est égale au salaire moyen des hommes.

A la question : "y a-t-il suffisamment d'évidence pour attester que les femmes gagnent moins que les hommes"
on répondra donc : "Non, il n'y a pas suffisamment d'évidence"