Test t à 1 échantillon

Exemple 1: Durée de traitement d'un emprunt

Problème

Une durée de traitement d'emprunt plus courte augmente la productivité et la satisfaction des clients. Un établissement financier souhaite mettre en place un processus de référence en évaluant sa durée de traitement moyenne. Il souhaite également savoir si sa durée moyenne diffère de celle d'un concurrent faisant état d'une durée de 6 heures.

Collecte des données

Un analyste financier sélectionne aléatoirement 7 demandes d'emprunt et calcule manuellement le temps écoulé entre la demande effectuée par le client et la réception par le client de la décision de l'établissement.

Outils

- **Statistiques descriptives**
- **Guide statistique StatGuide**
- **Test t à 1 échantillon**
- **Test de normalité**
- **Diagramme de série chronologique**
- **Diagramme des valeurs individuelles**

Fichier de données

Emprunt.MPJ

Variable	Description
Prêt	Numéro de demande d'emprunt
Heures	Nombre d'heures s'écoulant jusqu'à réception par le client de la notification

Affichage des statistiques descriptives

Les statistiques descriptives vous permettent de récapituler les caractéristiques importantes des données. En particulier, les statistiques descriptives fournissent des informations utiles sur l'emplacement et la variabilité des données.

Afficher les statistiques descriptives

1. Ouvrez Emprunt.MPJ.
2. Sélectionnez **Stat** > **Statistiques élémentaires** > **Afficher les statistiques descriptives**.
3. Dans la zone **Variables**, saisissez *Heures*.
4. Cliquez sur **OK**.

Interprétation des résultats

Les statistiques indiquent que la moyenne de l'échantillon est de 5,079 heures. Cette valeur est légèrement inférieure à la durée cible de 6 heures. Le test t à 1 échantillon permet de comparer cette différence (0,921) à la variation des données. Comme l'écart type des données échantillons est de 1,319 heure, le test t à 1 échantillon utilise l'ErT moyenne.

Et maintenant...

Utilisons le guide statistique StatGuide pour expliquer l'ErT moyenne.

Statistiques descriptives : Heures

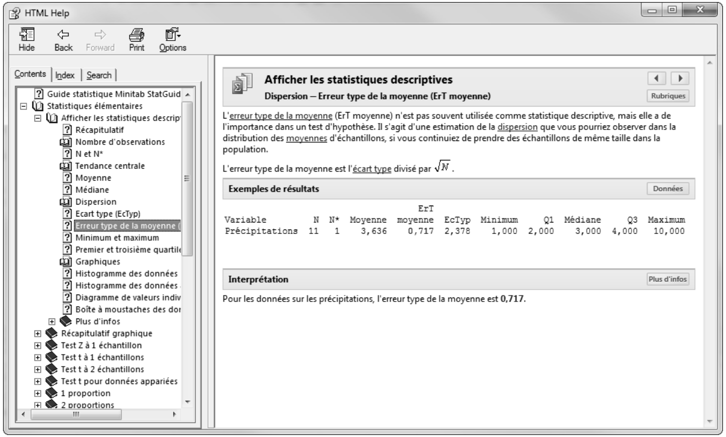
Variable	N	N*	Moyenne	ErT moyenne	EcTyp	Minimum	Q1	Médiane	Q3	Maximum
Heures	7	0	5,079	0,498	1,319	3,250	3,700	5,200	6,250	6,750

Utilisation du guide statistique StatGuide

Nous pouvons observer que la valeur ErT moyenne est simplement l'écart type divisé par la racine carrée du nombre de points de données et qu'elle représente la dispersion ou la variation dans la distribution des moyennes de l'échantillon. Le test t à 1 échantillon utilise la distribution de la moyenne de l'échantillon (et non celle des données) pour effectuer l'analyse. Ainsi, l'erreur type de la moyenne servira d'estimation de la variation pour le test t et l'intervalle de confiance.

Guide statistique StatGuide

1. Dans la fenêtre **Session**, placez votre curseur sur les résultats desstatistiques descriptives.
2. Cliquez sur l'icône .
3. Sélectionnez **Erreur type de la moyenne (ErT moyenne)** comme indiqué ci-dessous.



Test d'hypothèse

Qu'est-ce qu'un test d'hypothèse ?

Un test d'hypothèse utilise des données échantillons pour tester une hypothèse sur la population dont l'échantillon est prélevé. Le test t à 1 échantillon est l'une des nombreuses procédures disponibles dans Minitab pour les tests d'hypothèse.

Par exemple, pour savoir si la durée moyenne d'une transaction est égale à la cible souhaitée, mesurez la durée d'un échantillon de transactions et utilisez la moyenne de cet échantillon pour évaluer la moyenne de toutes les transactions. La conclusion tirée d'informations provenant d'un échantillon est appelée inférence statistique.

Quand utiliser un test d'hypothèse ?

Utilisez un test d'hypothèse pour faire des déductions sur une ou plusieurs populations lorsque des données échantillons sont disponibles.

Pourquoi utiliser un test d'hypothèse ?

Le test d'hypothèse permet de répondre aux types de questions suivants :

- Les temps de réponse correspondent-ils aux attentes des clients ou les dépassent-ils ?
- Le service dans une filiale est-il meilleur que celui d'une autre filiale ?

Par exemple :

- En moyenne, un centre d'appels respecte-t-il la durée cible pour répondre aux questions d'un client ?
- Le temps de facturation moyen est-il plus court dans l'agence disposant d'une nouvelle méthode de facturation ?

Test t à 1 échantillon

Qu'est-ce qu'un test t à 1 échantillon ?

Un test t à 1 échantillon permet de déterminer si μ (moyenne de la population) est égal à une valeur hypothétisée (moyenne du test).

Le test utilise l'écart type de l'échantillon pour estimer σ (écart type de la population). Si la différence entre la moyenne de l'échantillon et la moyenne du test est importante par rapport à la variabilité de la moyenne de l'échantillon, il est peu probable que μ soit égal à la moyenne du test.

Quand utiliser un test t à 1 échantillon ?

Utilisez un test t à un échantillon lorsque des données continues sont disponibles à partir d'un seul échantillon aléatoire.

Le test suppose que la population est distribuée normalement. Toutefois, le test est robuste à un non-respect de cette hypothèse pour des effectifs d'échantillons supérieurs ou égaux à 30, sous réserve que les observations soient collectées aléatoirement et que les données soient continues, unimodales et relativement symétriques (reportez-vous à [1]).

Pourquoi utiliser un test t à 1 échantillon ?

Un test t à un échantillon permet de répondre aux types de questions suivants :

- La durée moyenne de transaction est-elle satisfaisante ?
- Le service client satisfait-il les attentes des clients ?

Par exemple :

- En moyenne, un centre d'appels respecte-t-il la durée cible pour répondre aux questions d'un client ?
- La durée de facturation avec une nouvelle méthode est-elle plus courte que la durée actuelle de 20 jours ?

Test de l'hypothèse nulle

La société souhaite savoir si la durée moyenne d'un processus d'acceptation est statistiquement différente de celle des concurrents, qui avancent une durée de 6 heures. En termes statistiques, la moyenne du procédé correspond à la moyenne de la population, ou μ (mu).

Hypothèses statistiques

Soit μ est égal à 6 heures, soit il ne l'est pas. Vous pouvez affirmer ces alternatives sous la forme de deux hypothèses :

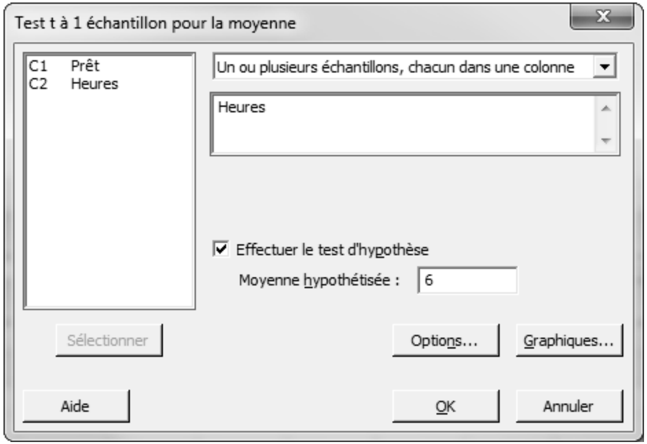
L'*hypothèse nulle* (H_0) : μ est égal à 6 heures.

L'*hypothèse alternative* (H_1) : μ n'est *pas* égal à 6 heures.

Comme les analystes ne vont pas mesurer toutes les demandes d'emprunt au sein de la population, ils ne connaîtront pas la valeur réelle de μ. Toutefois, un test d'hypothèse adapté leur permet de prendre une décision fondée. Pour ces données, le test approprié est un test t à 1 échantillon.

Test t à 1 échantillons

1. Sélectionnez **Stat** > **Statistiques élémentaires** > **Test t à 1 échantillon**.
2. Configurez la boîte de dialogue comme suit.



3. Cliquez sur **OK**.

Interprétation des résultats

Logique du test d'hypothèse

Tous les tests d'hypothèse reposent sur la même procédure :

1. Supposez que l'hypothèse H_0 est vraie.
2. Déterminez l'importance de la différence entre l'échantillon et ce que l'hypothèse ci-dessus laisserait prévoir.
3. Si la statistique issue de l'échantillon est suffisamment improbable dans le cas où l'hypothèse H_0 est vraie, rejetez H_0 en faveur de H_1.

Par exemple, les résultats du test t indiquent que la moyenne de l'échantillon est de 5,079 heures. Le test répond à la question "Si μ est égal à 6 heures, quelle est la probabilité d'obtenir une moyenne de l'échantillon qui présente au moins telle différence ?". La réponse est donnée sous la forme d'une valeur de probabilité (P), en l'occurrence 0,114 pour ce test.

Statistique de test

La statistique t (-1,85) se calcule comme suit :

t = (moyenne de l'échantillon - moyenne du test) / ErT moyenne

où ErT moyenne représente l'erreur type de la moyenne (mesure de variabilité). Au fur et à mesure que la valeur absolue de la statistique t augmente, la valeur de p devient plus petite.

Test T à 1 échantillon : Heures

```
Test de μ = 6 et ≠ 6


                                        ErT
Variable  N  Moyenne  EcTyp  moyenne     IC à 95 %         T      P
Heures    7    5,079  1,319    0,498  (3,859; 6,298)  -1,85  0,114
```

Interprétation des résultats

Prise d'une décision

Pour prendre une décision, choisissez le seuil de signification, α (alpha), avant le test :

- Si P est inférieur ou égal à α, rejetez H_0.
- Si P est supérieur à α, ne rejetez pas H_0. (En principe, vous n'*acceptez* jamais l'hypothèse H_0. Vous vous contentez de ne pas la rejeter.)

Une valeur typique d'α est 0,05, mais vous pouvez choisir des valeurs supérieures ou inférieures suivant la sensibilité nécessaire pour le test et les conséquences d'un rejet à tort de l'hypothèse nulle. Dans l'hypothèse d'un niveau d'α de 0,05 pour les données sur l'emprunt, il n'existe pas de preuve suffisante pour rejeter H_0 car P (0,114) est supérieur à α.

Et maintenant...

Vérifions l'hypothèse de normalité.

Test T à 1 échantillon : Heures

```
Test de µ = 6 et ≠ 6


                                       ErT
Variable  N  Moyenne  EcTyp  moyenne     IC à 95 %         T      P
Heures    7    5,079  1,319    0,498  (3,859; 6,298)  -1,85  0,114
```

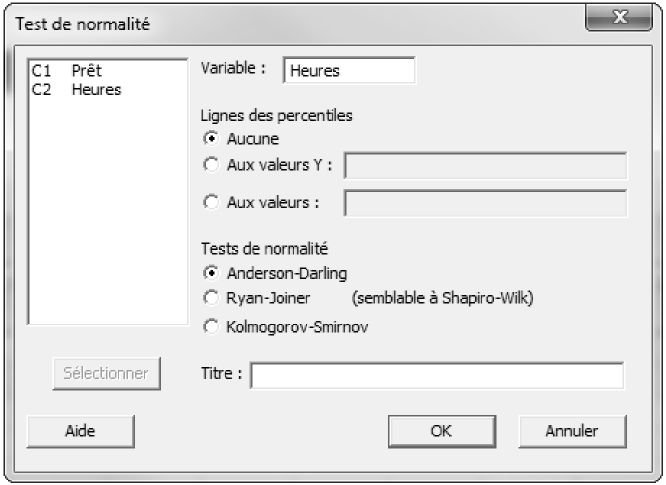
Test de l'hypothèse de normalité

Le test t à 1 échantillon suppose que les données sont échantillonnées à partir d'une population distribuée normalement.

Utilisez un test de normalité pour déterminer si l'hypothèse de normalité est valide pour ces données.

Test de normalité

1. Sélectionnez **Stat** > **Statistiques élémentaires** > **Test de normalité**.
2. Configurez la boîte de dialogue comme suit.



3. Cliquez sur **OK**.

Interprétation des résultats

Utilisez la droite de Henry pour vérifier que les données ne s'écartent pas sensiblement des résultats attendus lorsque l'échantillonnage est réalisé à partir d'une loi normale.

- Si les données proviennent d'une loi normale, les points doivent suivre plus ou moins la droite d'ajustement.
- Si les données ne proviennent pas d'une loi normale, les points ne suivent pas la droite.

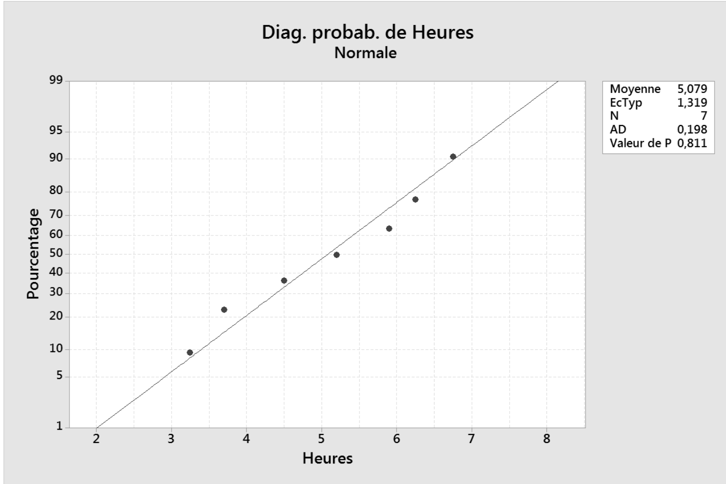
test de normalité d'Anderson-Darling

Les hypothèses du test de normalité d'Anderson-Darling sont les suivantes :

H_0 : les données proviennent d'une population distribuée normalement.

H_1 : les données ne proviennent pas d'une population distribuée normalement.

La valeur de p issue du test d'Anderson-Darling (0,811) évalue la probabilité que les données proviennent d'une population distribuée normalement. Avec un niveau d'α de 0,05, il n'existe pas de preuve suffisante permettant de supposer que les données ne proviennent pas d'une population normalement distribuée.



Et maintenant...

Trions les données chronologiquement pour vérifier s'il existe des schémas non aléatoires déterminés dans le temps.

Conclusion

D'après le diagramme et le test de normalité, vous pouvez supposer que les données proviennent d'une population normalement distribuée.

Remarque Lorsque les données ne sont pas distribuées normalement, il se peut que vous puissiez les transformer à l'aide d'une transformation de Box-Cox ou utiliser une procédure non paramétrique, telle qu'un test du signe à 1 échantillon.

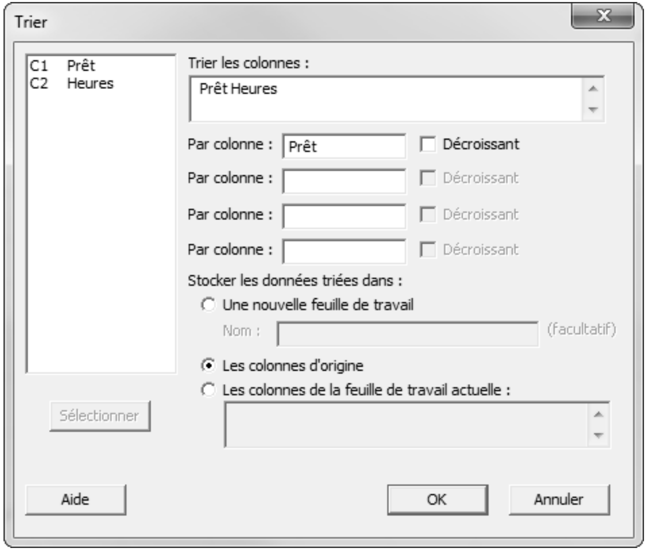
Tri des données

Vous pouvez utiliser la commande Trier de Minitab pour trier les données dans l'ordre croissant ou décroissant, en fonction de critères numériques, alphabétiques ou de dates. Dans cet exemple, triez les données dans l'ordre croissant des numéros de demande d'emprunt. Le tri des données par numéro de demande d'emprunt permettra de les classer par ordre chronologique car les numéros des demandes d'emprunt sont affectés par ordre croissant par le système. Le tri des données dans l'ordre chronologique facilite le traçage des données dans le temps et la détection de schémas ou de tendances dans celle-ci.

Remarque Incluez toutes les colonnes adaptées dans l'étape de tri, afin de conserver la relation entre les colonnes de données.

Trier

1. Sélectionnez **Données** > **Trier**.
2. Configurez la boîte de dialogue comme suit.



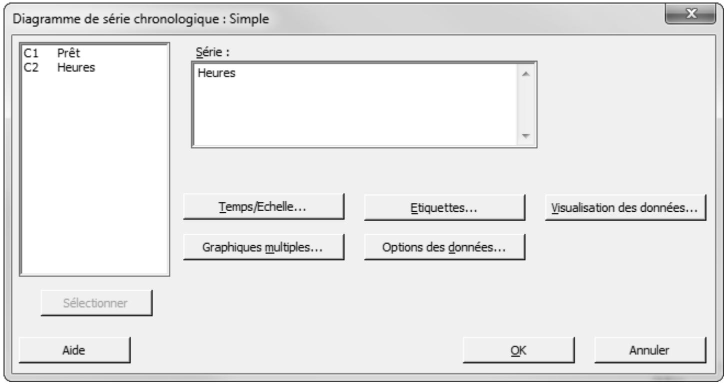
3. Cliquez sur **OK**.

Test d'hypothèse sur le caractère aléatoire

Utilisez un diagramme de série chronologique pour rechercher les tendances ou schémas dans vos données, ce qui peut indiquer que vos données ne sont pas aléatoires dans le temps.

Diagrammes de série chronologique

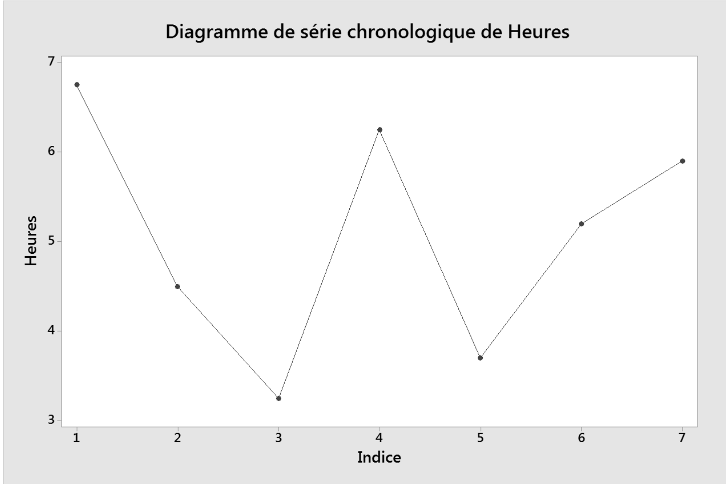
1. Sélectionnez **Graphique** > **Diagramme de série chronologique**.
2. Sélectionnez **Simple**, puis cliquez sur **OK**.
3. Configurez la boîte de dialogue comme suit.



4. Cliquez sur **OK**.

Interprétation des résultats

Si une tendance ou un schéma existe dans vos données, nous souhaitons en comprendre les raisons. Dans ce cas, les données ne présentent pas de tendances ou de schémas évidents.



Et maintenant...

Calculons un intervalle de confiance pour la moyenne réelle de la population.

Intervalles de confiance

Qu'est-ce qu'un intervalle de confiance ?

Un intervalle de confiance est une plage de valeurs probables pour un paramètre de population (tel que µ), basée sur des données échantillons. Par exemple, avec un intervalle de confiance à 95 % pour µ, vous pouvez être certain à 95 % que l'intervalle contient µ. En d'autres termes, 95 intervalles sur 100 vont contenir µ sur plusieurs échantillonnages.

Quand utiliser un intervalle de confiance ?

Utilisez un intervalle de confiance pour faire des déductions sur plusieurs populations à partir de données d'échantillons ou pour quantifier la précision de votre estimation d'un paramètre de population, comme µ.

Pourquoi utiliser un intervalle de confiance ?

Les intervalles de confiance peuvent permettre de répondre, en grande partie, aux mêmes questions que le test d'hypothèse :

- µ est-il centré sur la cible ?
- Dans quelle mesure existe-t-il des erreurs dans une estimation de µ ?
- Quelle valeur élevée ou faible µ peut-il présenter ?

Par exemple :

- La durée de transaction moyenne est-elle supérieure à 30 secondes ?
- Quelle est la plage de valeurs probables pour des chiffres d'affaires quotidiens moyens ?

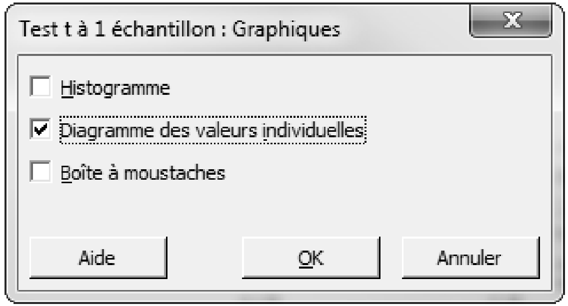
Utilisation de l'intervalle de confiance

Dans l'analyse précédente, vous avez utilisé un test d'hypothèse pour déterminer si la moyenne de la durée de traitement d'un emprunt différait de la valeur cible. Vous pouvez également utiliser un intervalle de confiance pour évaluer cette différence.

Les résultats de la fenêtre Session du test t à 1 échantillon comprennent les valeurs des bornes supérieure et inférieure de l'intervalle de confiance à 95 %. Pour obtenir une représentation graphique de l'intervalle, sélectionnez **Diagramme des valeurs individuelles** dans la sous-boîte de dialogue **Graphiques**.

Test t à 1 échantillon

1. Sélectionnez **Stat** > **Statistiques élémentaires** > **Test t à 1 échantillon**.
2. Cliquez sur **Graphiques**.
3. Configurez la boîte de dialogue comme suit.



4. Cliquez sur **OK** dans chaque boîte de dialogue.

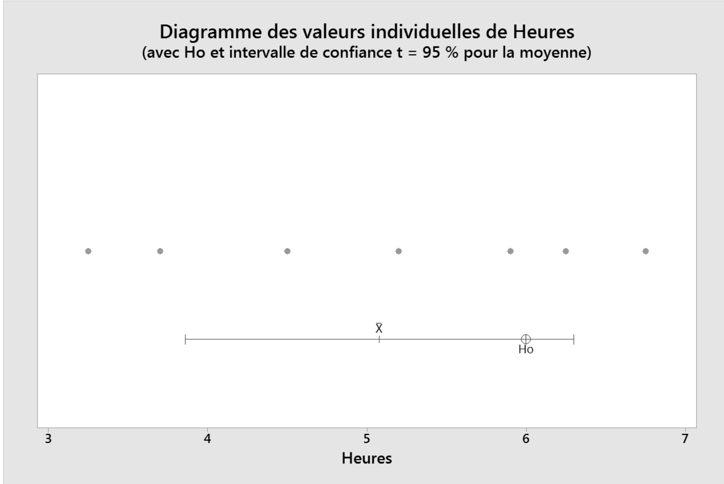
Interprétation des résultats

Intervalle de confiance

L'intervalle de confiance est une étendue de valeurs probables pour μ. Dans le diagramme de valeurs individuelles, Minitab représente l'intervalle sous la forme d'une ligne bleue.

Si vous effectuiez un échantillonnage plusieurs fois à partir de la même population, les intervalles de confiance d'environ 95 % des échantillons comprendraient μ. Par conséquent, pour n'importe quel échantillon, vous pouvez être certain à 95 % que μ se situe à l'intérieur de l'intervalle de confiance.

Remarque Un intervalle de confiance ne représente pas 95 % des données ; il s'agit d'une idée fausse courante.



Interprétation des résultats

Test d'hypothèse

Le marqueur de repère du milieu, étiqueté $\overline{X}$, représente la moyenne de l'échantillon, tandis que le cercle rouge, étiqueté H_0, représente la moyenne de la population supposée (6 heures). Vous pouvez être sûr à 95 % que la moyenne du procédé est comprise entre 3,859 heures et 6,298 heures.

Utilisez l'intervalle de confiance pour tester l'hypothèse nulle :

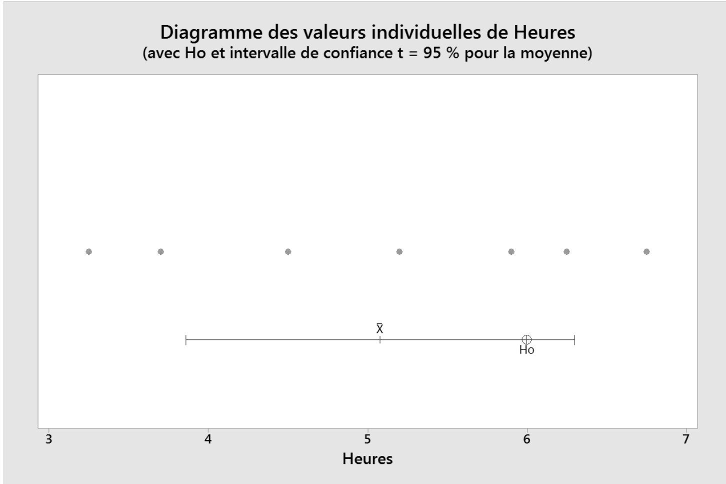
- Si H_0 se situe en dehors de l'intervalle, la valeur de p du test d'hypothèse est inférieure à 0,05. Vous pouvez rejeter l'hypothèse nulle au niveau d'α de 0,05.
- Si H_0 se situe dans l'intervalle, la valeur de p est supérieure à 0,05. Vous ne pouvez pas rejeter l'hypothèse nulle au niveau d'α de 0,05.

Etant donné que H_0 se situe dans l'intervalle de confiance, vous ne pouvez pas rejeter l'hypothèse nulle. Il n'existe pas de preuves suffisantes pour conclure que μ est différent de la valeur cible de 6 heures au seuil de signification de 0,05.

Test T à 1 échantillon : Heures

Test de μ = 6 et ≠ 6

Variable	N	Moyenne	EcTyp	ErT moyenne	IC à 95 %	T	P
Heures	7	5,079	1,319	0,498	(3,859; 6,298)	-1,85	0,114



Observations finales

Résumé et conclusions

D'après le test t et les données échantillons, vous ne pouvez pas rejeter l'hypothèse nulle au niveau d'α de 0,05. En d'autres termes, les données ne fournissent pas de preuve suffisante permettant de conclure que la durée moyenne de traitement est significativement différente de 6 heures.

Le test de normalité et le diagramme de série chronologique indiquent que les données confirment les hypothèses de normalité et de caractère aléatoire du test t.

L'intervalle de confiance à 95 % indique que la valeur réelle de la moyenne de la population est comprise entre 3,859 et 6,298 heures.

Observations finales

Hypothèses

Un test d'hypothèse commence toujours par deux hypothèses opposées.

L'hypothèse nulle (H_0) :

- Affirme généralement qu'une propriété d'une population (telle que la moyenne) n'est pas différente d'une valeur spécifiée ou de celle d'une référence.
- Est supposée vraie jusqu'à ce que suffisamment de preuves indiquent le contraire.
- N'est jamais démontrée comme étant vraie. Vous vous contentez de ne pas la rejeter.

L'hypothèse alternative (H_1) :

- Affirme que l'hypothèse nulle est fausse.
- Peut également spécifier le sens de la différence.

Seuil de signification

Choisissez le niveau d'α *avant* d'effectuer le test.

- L'augmentation du niveau d'α accroît les chances de détecter une différence, mais également les risques de rejeter H_0 à tort (erreur de 1ère espèce).

Hypothèses

Chaque test d'hypothèse repose sur une ou plusieurs hypothèses au sujet des données analysées. Si ces hypothèses ne sont pas satisfaites, les conclusions peuvent être incorrectes.

Les hypothèses pour un test t à 1 échantillon sont les suivantes :

- L'échantillon doit être aléatoire.
- Les données échantillons doivent être continues.
- Les données de l'échantillon doivent être normalement distribuées (bien que cette hypothèse soit moins importante lorsque l'effectif de l'échantillon est de 30 ou plus).

La procédure du test t est robuste à un non-respect de l'hypothèse de normalité, à condition que les observations soient collectées aléatoirement et que les données soient continues, unimodales et relativement symétriques (reportez-vous à [1]).

Intervalle de confiance

L'intervalle de confiance fournit une plage probable de valeurs pour μ (ou pour d'autres paramètres de population).

Vous pouvez effectuer un test d'hypothèse bilatéral (hypothèse alternative de ≠) à l'aide d'un intervalle de confiance. Par exemple, si la valeur du test ne se situe pas dans un intervalle de confiance à 95 %, vous pouvez rejeter H_0 au niveau d'α de 0,05. De même, si vous définissez un intervalle de confiance à 99 % et qu'il ne

- La diminution du niveau d'α réduit les risques de commettre une erreur de 1ère espèce, mais également les chances de détecter correctement une différence.

comprend pas la moyenne du test, vous pouvez rejeter H_0 au niveau d'α de 0,01.