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 6 demandes de crédits effectuées les deux dernières semaines 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

- **Test t à 1 échantillon**
- **Test de normalité**
- **Diagramme de série chronologique**

Fichier de données

Emprunt.MPJ

Variable	Description
Date	Date de la notification du client
Heures	Nombre d'heures s'écoulant jusqu'à réception par le client de la notification

Test d'hypothèse

Qu'est-ce qu'un test d'hypothèse ?

Un test d'hypothèse utilise des données d'échantillon pour tester une hypothèse sur la population dont l'échantillon est prélevé. Le test t à 1 échantillon constitue 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. Il s'agit d'un exemple d'*inférence statistique*, qui utilise des informations sur un échantillon pour faire des déductions relatives à une population.

Quand utiliser un test d'hypothèse ?

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

Pourquoi utiliser un test d'hypothèse ?

Le test d'hypothèse permet de répondre aux questions suivantes :

- 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 questions suivantes :

- 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 processus 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.

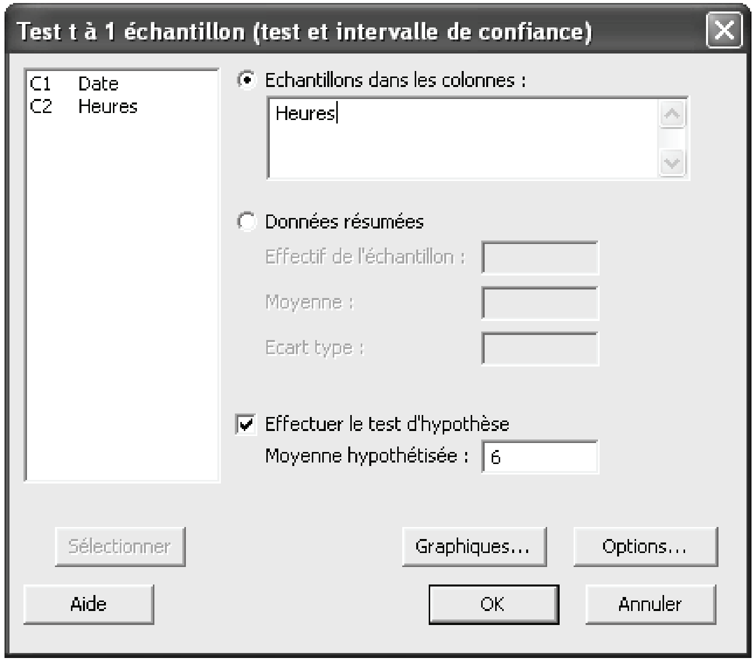
Comme les analystes ne vont pas mesurer toutes les demandes de prêt au sein de la population, ils ne connaîtront pas la valeur réelle de μ. Toutefois, un test d'hypothèse adapté leur permet de prendre une décision fondée. Pour ces données, le test approprié est un test t à 1 échantillon.

Test t à 1 échantillon

1 Ouvrez le fichier Emprunt.MPJ.

2 Sélectionnez **Stat ➤ Statistiques élémentaires ➤ Test t à 1 échantillons**.

3 Configurez la boîte de dialogue comme suit.



4 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 4,792 heures. Le test répond à la question suivante : "Si µ est égal à 6 heures, quelle est la probabilité d'obtenir une moyenne de l'échantillon aussi différente de 6 heures que celle que vous avez observée ?" La réponse est donnée sous la forme d'une valeur de probabilité (P), en l'occurrence 0,081.

Statistique de test

La statistique t (-2,18) 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 un échantillon : Heures

```
Test de mu = 6 en fonction de la différence 6

                                    ErT
Variable  N  Moyenne  EcTyp  moyenne      IC à 95 %        T      P
Heures    6    4,792  1,355    0,553  (3,370; 6,213)  -2,18  0,081
```

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 l'hypothèse H_0. P (0,081) est supérieur à α.

Et maintenant...

Vérifions l'hypothèse de normalité.

Test T à un échantillon : Heures

```
Test de mu = 6 en fonction de la différence 6

                                     ErT
Variable  N  Moyenne  EcTyp  moyenne     IC à 95 %         T      P
Heures    6    4,792  1,355    0,553  (3,370; 6,213)  -2,18  0,081
```

Test de l'hypothèse de normalité

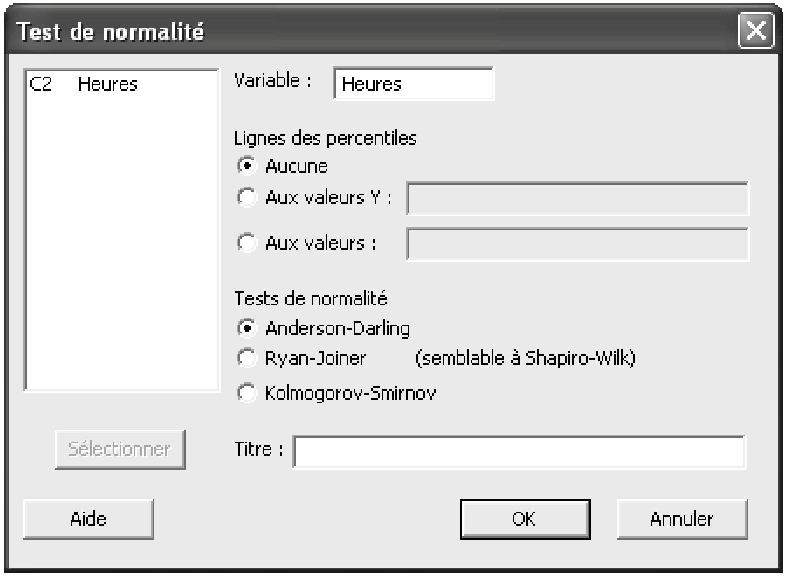
Le test t à 1 échantillon suppose que les données sont échantillonnées à partir d'une population distribuée normalement.

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.

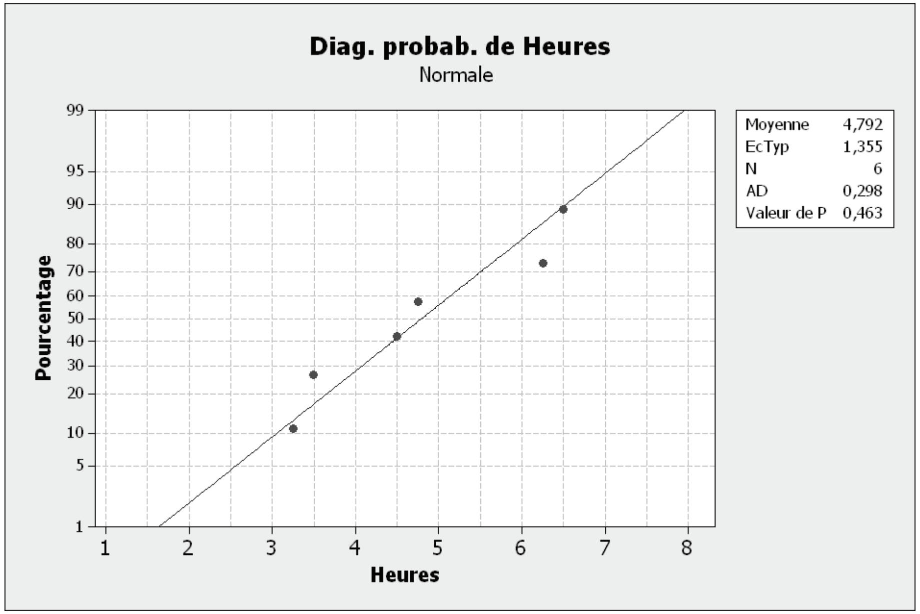


3 Cliquez sur **OK**.

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 (0,463) 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.

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.

Et maintenant...

Vérifions les données pour des schémas non aléatoires déterminés dans le temps.

Test d'hypothèse sur le caractère aléatoire

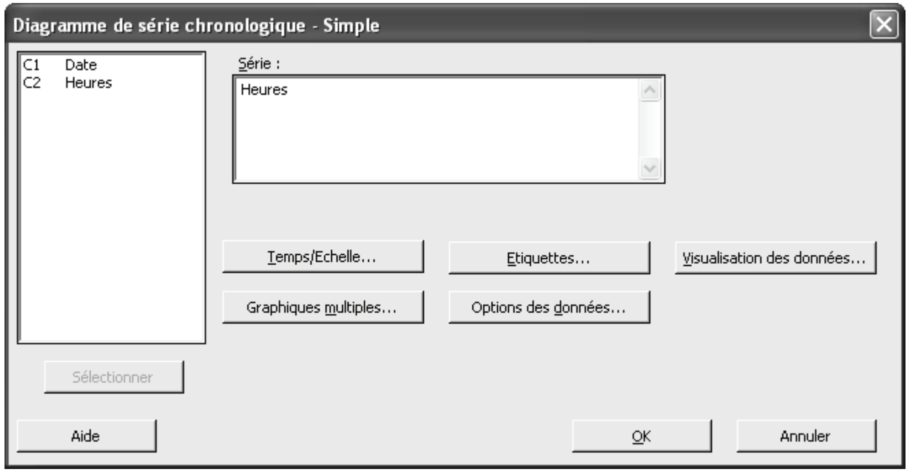
Utilisez un diagramme de série chronologique pour rechercher les tendances ou schémas dans vos données, ce qui peut indiquer que vos données ne sont pas aléatoires dans le temps.

Diagramme de série chronologique

1 Sélectionnez **Graphique ➤ Diagramme de série chronologique**.

2 Sélectionnez **Simple**, puis cliquez sur **OK**.

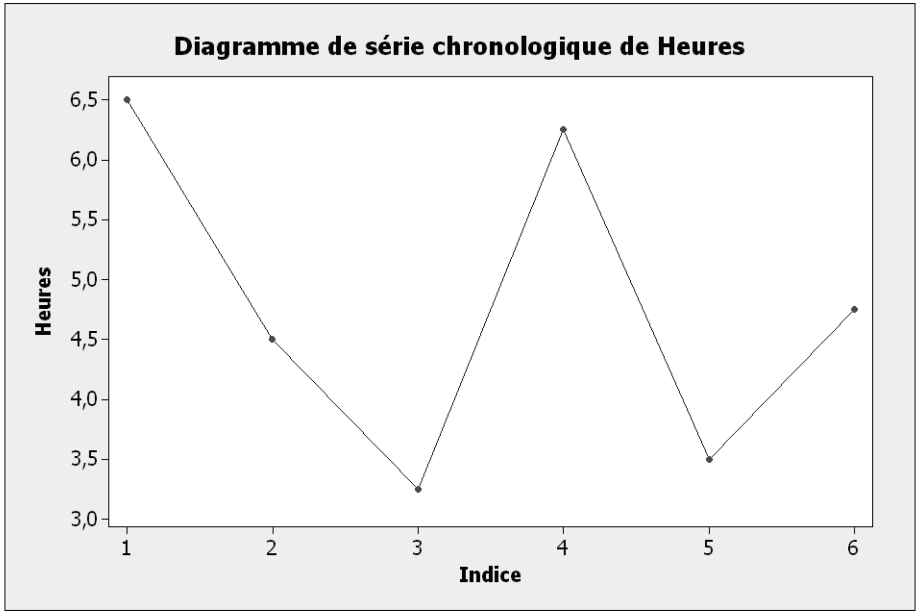
3 Configurez la boîte de dialogue comme suit.



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 d'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 ?

Interprétation des résultats

Intervalle de confiance

L'intervalle de confiance de 95 % pour la durée moyenne de traitement est compris entre 3,370 et 6,213 heures. L'intervalle de confiance de 95 % comprend la valeur de comparaison de 6. Cela revient à ne pas rejeter l'hypothèse nulle ($H_0 : \mu = 6$) pour ce test t avec $\alpha = 0,05$.

Test T à un échantillon : Heures

```
Test de mu = 6 en fonction de la différence 6

                                      ErT
Variable  N  Moyenne  EcTyp  moyenne     IC à 95 %        T      P
Heures    6    4,792  1,355    0,553  (3,370; 6,213)  -2,18  0,081
```

Observations finales

Résumé et conclusions

D'après le test t et les données de l'échantillon, 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 de 95 % indique que la valeur réelle de la moyenne de la population est comprise entre 3,37 et 6,213 heures.

Observations finales

Résumé et conclusions

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).
- 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.

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 un échantillon sont les suivantes :

- L'échantillon doit être aléatoire.
- Les données d'échantillons doivent être des variables 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 1/4) à l'aide d'un intervalle de confiance. Par exemple, si la valeur du test ne se situe pas à l'intérieur d'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 comprend pas la moyenne du test, vous pouvez rejeter H_0 au niveau d'α de 0,01.