Licence de Psychologie - Semestre N° 5 - TD n° 2
Intervalles de confiance, lois de distribution classiques et tests paramétriques avec Statistica

9 Travail sur des données recensées : statistiques descriptives sur un tableau d'effectifs

9.1 Introduire une pondération dans la feuille de données

On considère l'exemple suivant :
Dans le cadre d'une analyse médicale, deux méthodes de dosage peuvent être utilisées. A partir d'un même prélèvement, on répète 25 fois la méthode A et 30 fois avec la méthode B. Les résultats sont rassemblés dans les tableaux rassemblés dans le classeur Dosages.stw.

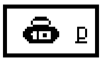
Ouvrez le classeur Dosages.stw et observez la façon dont les données ont été saisies dans les feuilles Méthode A, Méthode B et Ensemble.

Contrairement aux exemples traités précédemment, les données sont ici présentées sous la forme de tableaux recensés : les observations ont fait l'objet d'un tri à plat. Par exemple, la valeur 42 a été obtenue 7 fois comme résultat de mesure par la méthode A.

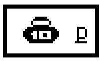
Comment effectuer les traitements de Statistiques descriptives sur des données structurées sous cette forme, par exemple dans la feuille "Méthode A" ?

Il faut indiquer à Statistica que la colonne "Nombre de dosages" est une colonne d'effectifs ou **pondérations** des données.

Les pondérations peuvent aussi bien être définies comme propriété de la feuille elle-même que comme propriété de l'une des analyses.

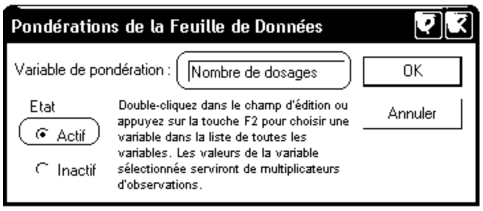
Dans le premier cas, on affiche la feuille de données et on utilise le bouton "pondérations" de la barre d'outils :


. Les pondérations s'appliquent alors à toutes les analyses utilisant cette feuille.

Dans le deuxième cas, on utilise l'un des items du menu Statistiques et on clique sur le bouton "pondérations"


de la fenêtre de dialogue. Les pondérations ne concerneront alors que l'analyse en cours.

Ici, rendez active la feuille "Méthode A" et indiquez que la variable 2 (Nombre de dosages) est la variable de pondération :

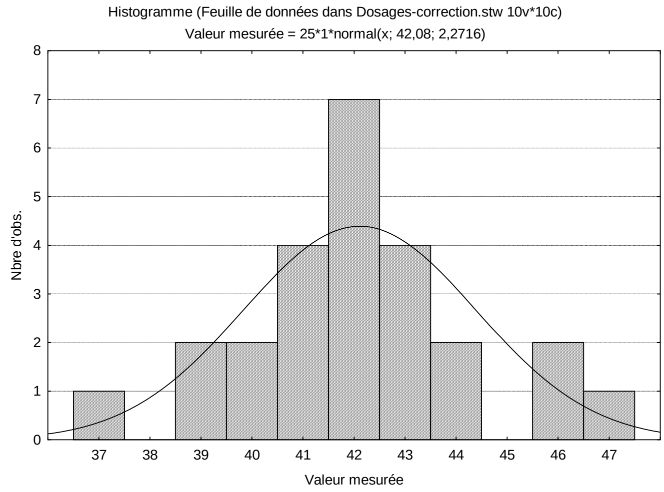


Le reste du traitement peut alors être réalisé de la même façon que pour un tableau-protocole. On obtient par exemple:

	Statistiques Descriptives (Méthode A dans Dosages.stw				
Variable	N Actifs	Moyenne	Minimum	Maximum	Ecart-type
Valeur mesurée	25	42,08000	37,00000	47,00000	2,271563

Remarquez que le nombre d'observations pris en compte n'est pas égal à 9 (nombre de lignes du fichier) mais à 25 (somme des effectifs contenus dans la deuxième colonne).

De même, réalisez à l'aide du menu Graphiques, l'histogramme suivant :



9.2 Calculer des paramètres de statistiques descriptives pour des données structurées "par groupe"

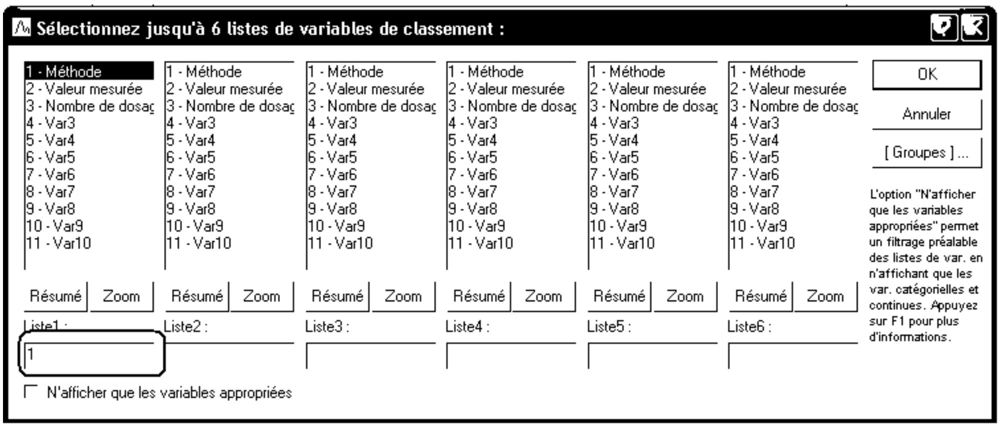
Nous souhaitons calculer la moyenne, la variance et l'écart type de la variable "Valeur mesurée" pour chacune des méthodes, et obtenir les résultats dans une même feuille de résultats.

- Rendez active la feuille "Ensemble" et définissez la variable "Nombre de dosages" comme variable de pondération.
- Utilisez ensuite le menu Statistiques - Statistiques Elémentaires et la méthode "Décompositions ; tableaux non factoriels".
- Sous l'onglet "Base", indiquez "Valeur mesurée" comme variable dépendante et "Méthode" comme variable de classement.
- Sous l'onglet "Statistiques descriptives", sélectionnez les paramètres à calculer : écart type, variance et N.
- Cliquez sur l'un des boutons "Synthèse".

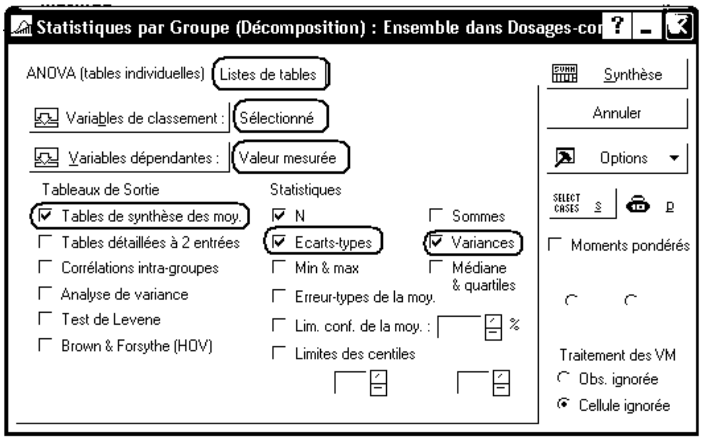
Vous devriez obtenir :

Statistiques Descriptives par Groupes (Ensemble dans Dosages-correction.stw) N=55(pas de VM dans les vars dépendantes)					
	Méthode	Valeur mesurée Moyennes	Valeur mesurée N	Valeur mesurée Ec-Types	Valeur mesurée Variance
	A	42,08000	25	2,271563	5,160000
	B	42,10000	30	1,398275	1,955172
	TsGrpes	42,09091	55	1,828506	3,343434

Remarque. Dans le cas de situations plus complexes, on pourra obtenir des résultats du même type à l'aide de la méthode "Décompositions & ANOVA à un facteur" du même menu Statistiques - Statistiques Elémentaires. Les résultats de statistiques descriptives sont disponibles sous l'onglet "Listes de tables". Mais l'interface de la fenêtre de dialogue est alors plus compliquée, puisque l'on peut indiquer jusqu'à 6 listes de variables de classement.



Choisissez ensuite la variable "Valeur mesurée" comme variable dépendante :



Les résultats obtenus sont identiques aux précédents.

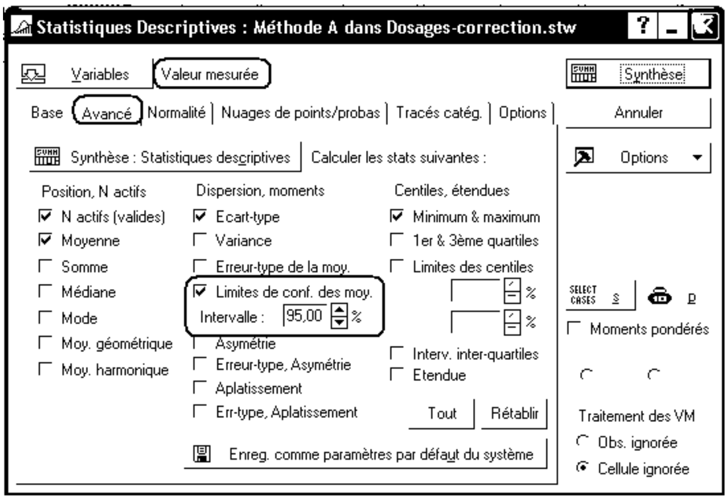
10 Intervalles de confiance

10.1 Intervalle de confiance pour une moyenne

Les méthodes "Statistiques descriptives", "Décompositions ; tableaux non factoriels" et "Décompositions & ANOVA à un facteur" du menu Statistiques - Statistiques élémentaires permettent également d'obtenir une estimation de la moyenne d'une variable numérique par un intervalle de confiance avec un degré de confiance donné.

Ainsi, à partir de l'échantillon de mesures réalisées par la méthode A, quel intervalle estimant la "vraie valeur" de la substance dosée peut-on donner avec un degré de confiance de 95% ?

Rendez active la feuille "Méthode A" et utilisez le menu Statistiques - Statistiques Elémentaires puis la méthode "Statistiques descriptives" en complétant la fenêtre de dialogue comme ci-dessous :

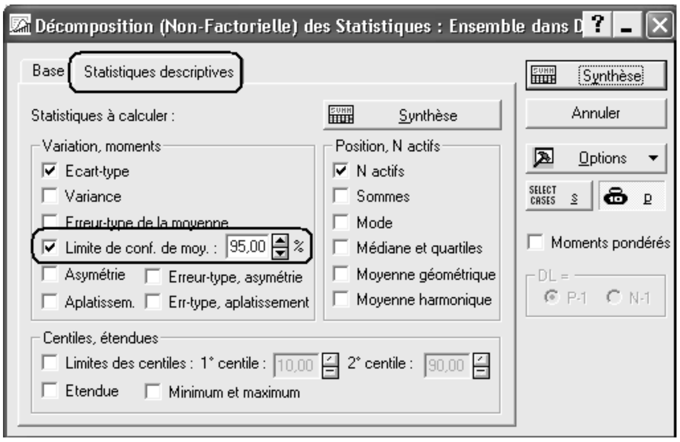


Vous devriez obtenir les résultats suivants :

Variable	Statistiques Descriptives (Méthode A dans Dosages-correction.stw) N Actifs	Moyenne	Confiance -95,000%	Confiance +95,000%	Minimum	Maximum	Ecart-type
Valeur mesurée	25	42,08	41,14	43,02	37,00	47,00	2,27

Autrement dit, on estime, avec un degré de confiance de 95%, que la vraie valeur de la quantité à doser est comprise entre 41,14 et 43,02.

- Rendez active la feuille "Ensemble" et utilisez le menu Statistiques - Statistiques Elémentaires puis la méthode "Décompositions ; tableaux non factoriels" en indiquant comme précédemment "Valeur mesurée" comme variable dépendante et "Méthode" comme variable de classement.
- Sous l'onglet "Statistiques descriptives", remplissez la fenêtre de dialogue comme suit :



Vous devriez obtenir :

Statistiques Descriptives par Groupes (Ensemble dans Dosages-correction.stw) N=55(pas de VM dans les vars dépendantes) Méthode	Valeur mesurée Moyennes	Valeur mesurée N	Valeur mesurée Ec-Types	Valeur mesurée Confiance -95.000%	Valeur mesurée Confiance +95.000%
A	42,08000	25	2,271563	41,14234	43,01766
B	42,10000	30	1,398275	41,57788	42,62212
TsGrpes	42,09091	55	1,828506	41,59659	42,58522

On voit que les intervalles de confiance obtenus pour chacune des méthodes, [41,14; 43,02] et [41,58; 42,62] se recouvrent largement. Un test de comparaison de ces moyennes devrait donc conduire à accepter leur égalité dans les populations parentes.

Exercice : De même, comparez les intervalles de confiance obtenus par les deux méthodes en utilisant la feuille "Ensemble" et le menu Statistiques - Statistiques Elémentaires - Décompositions & ANOVA à un facteur.

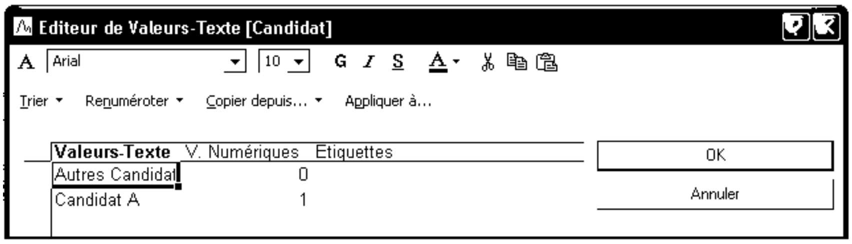
10.2 Intervalle de confiance pour une proportion

Lors d'un sondage électoral, on interroge au hasard 1000 personnes. 535 personnes déclarent vouloir voter pour le candidat A pendant que 465 déclarent vouloir voter pour un autre candidat. Quel intervalle de confiance, avec un degré de confiance de 95%, peut-on donner concernant le score du candidat A ?

Pour répondre à cette question, on peut :
- Créer la feuille de données suivante :

	1 Candidat	2 Suffrages
1	1	535
2	0	465

- Au besoin définir des étiquettes de texte pour la variable "Candidat", de façon que la feuille affiche "Candidat A" et "Autres Candidats". Pour cela, faire un double-clic sur la tête de la colonne "Candidat", puis utiliser le bouton "Valeurs-Texte" :



	1 Candidat	2 Suffrages
1	Candidat A	535
2	Autres Candidats	465

- Définir la variable "Suffrages" comme variable de pondération.

- Déterminer l'intervalle de confiance comme précédemment, à l'aide de la méthode "Statistiques descriptives".
 :

Variable	Statistiques Descriptives N Actifs	Moyenne	Confiance -95,000%	Confiance +95,000%
Candidat	1000	0,5350	0,5040	0,5660

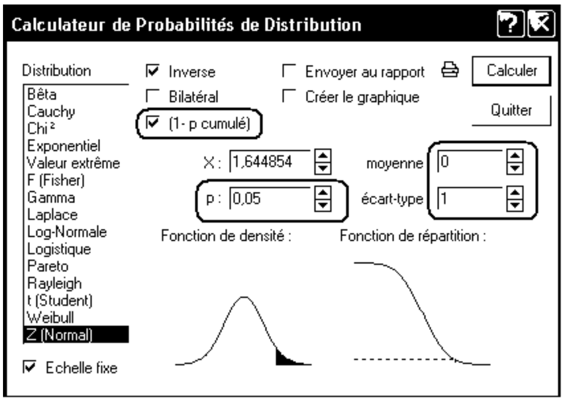
Au vu de l'intervalle trouvé, il semblerait que l'on puisse affirmer, avec un degré de confiance de 95%, que le candidat A sera élu...

11 Lois statistiques classiques

Le menu Statistiques - Calculateur de Probabilités - Distributions permet d'une part de trouver une valeur critique ou un niveau de significativité pour les lois statistiques continues usuelles, soit de réaliser des représentations graphiques de la densité ou de la fonction de répartition de ces lois.

11.1 La loi normale centrée réduite

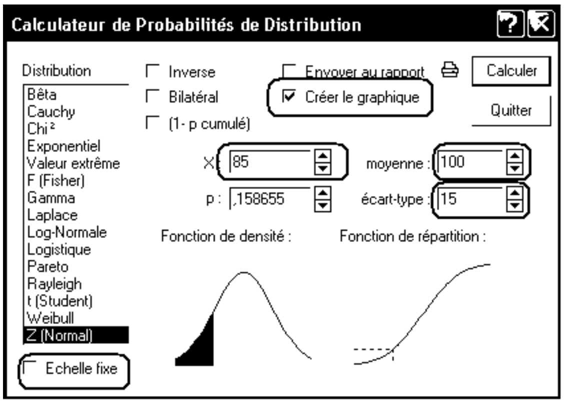
Quelle est la valeur de Zcritique pour un test unilatéral à 5% ?
Compléter le dialogue comme suit :



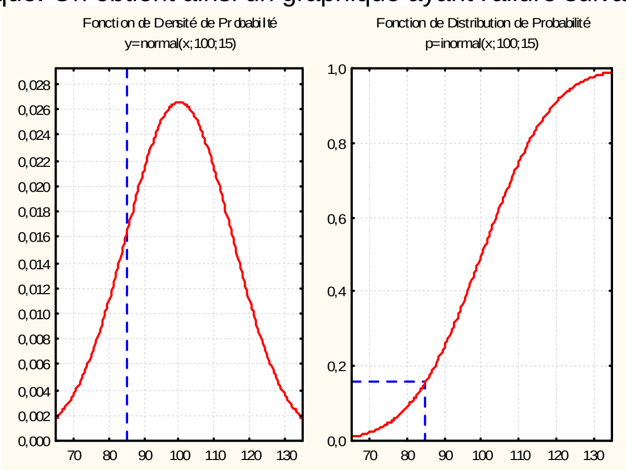
La réponse doit être lue dans la zone d'édition "X: ________". C'est ici : Zc=1,644854.

11.2 Représenter graphiquement la densité d'une loi normale quelconque

On veut représenter la densité de la loi normale de paramètres m = 100 et s=15 pour X compris entre 70 et 130
Complétez la fenêtre de dialogue comme suit, en veillant à ce que la boite "Echelle Fixe" ne soit pas cochée :

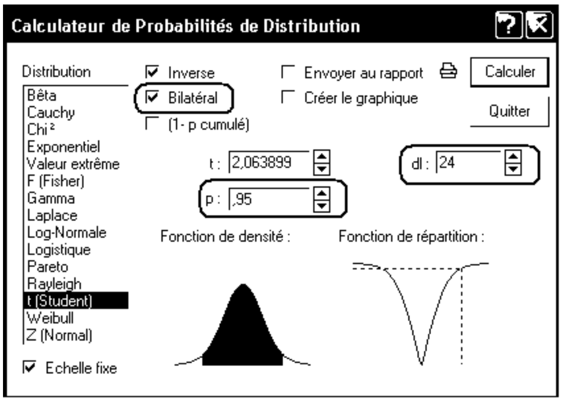


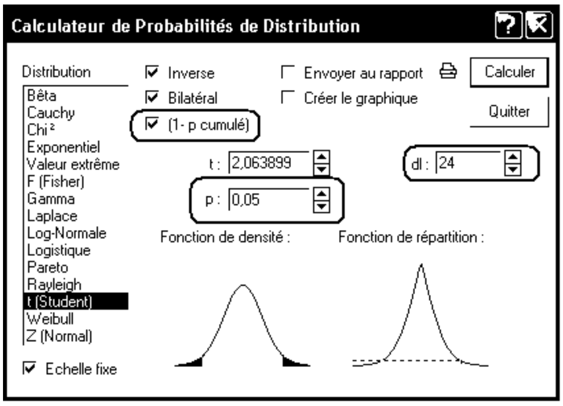
Remarquez que Statistica exige que l'un des deux champs "X:" ou "p:" soit complété, et repère la valeur correspondante sur le graphique. On obtient ainsi un graphique ayant l'allure suivante :



11.3 Loi de Student

Calculez la valeur critique de la loi de Student pour un test bilatéral avec ddl=24 et α=0,01 :





Remarquez que les paramètres peuvent être entrés de deux façons différentes :

- p = 0,95 si l'on ne coche pas la boîte "(1-p cumulé)
- p = 0,05 si l'on coche cette boîte.

Lecture du résultat : dans les deux cas, on obtient t_{crit}=2,063899.

On a réalisé un test de Student, avec un nombre de degrés de liberté égal à 58. La valeur observée de la statistique est t_{obs}=2,54. Quel est le niveau de significativité du résultat obtenu pour un test unilatéral ? pour un test bilatéral ?
Réponses : p=1,38% pour un test bilatéral et p=0,69% pour un test unilatéral.

11.4 Loi du khi-2

Selon les habitudes américaines, cette loi est désignée par Chi^2.
Déterminez la valeur critique du khi-2 pour un seuil de 5% et 6 ddl.
Vous devriez trouver : Khi-2 = 12,59.
Réaliser un graphique de la densité de la loi du khi-2 à 1 degré de liberté, en ajustant les échelles de manière à faire apparaître clairement la valeur critique correspondant à un seuil de 5% : 3,94.
Là, le choix d'échelle fait par Statistica est plus que discutable (que l'on ait, ou non, coché la case "Echelle Fixe"). Pour imposer notre choix d'échelle :

- Cliquez sur le graphique avec le bouton droit de la souris
- Sélectionnez l'item de menu Propriétés du Graphique (toutes options)
- Affichez l'onglet Axe -Echelle
- Indiquez pour l'axe X une plage de variation de 0,00015 à 5
- Indiquez pour l'axe Y une plage de variation de 0 à 1,5 (par exemple).

12 Tests paramétriques classiques

12.1 Test de comparaison d'une moyenne à une norme

Les données de la feuille de données du classeur Add.stw proviennent d'une étude de Howell et Huessy (1985). Ces auteurs ont rendu compte d'une étude portant sur 386 enfants qui avaient ou non manifesté, durant l'enfance, des symptômes liés à des troubles de l'attention. Ces données sont décrites dans l'appendice du livre.

Col 1 = Numéro d'identification du sujet
Col 2 = Moyenne des scores de troubles de l'attention sur 3 ans
Col 3 = Sexe. 1 = masculin; 2 = féminin
Col 4 = Nombre d'années scolaires que l'élève a doublées
Col 5 = QI calculé sur la base d'un test de QI administré au groupe
Col 6 = Niveau d'anglais: 1 = niveau préparatoire à l'université; 2 = niveau moyen; 3 = rattrapage
Col 7 = Résultats obtenus en anglais: 4 = très bon; 3 = bon; etc.
Col 8 = Moyenne des points obtenus en neuvième année
Col 9 = Problèmes sociaux: 0 = non; 1 = oui
Col 10 = 1 = l'élève a quitté l'école avant son terme; 0 = l'élève a terminé l'école

Ouvrez le classeur ADD.stw.

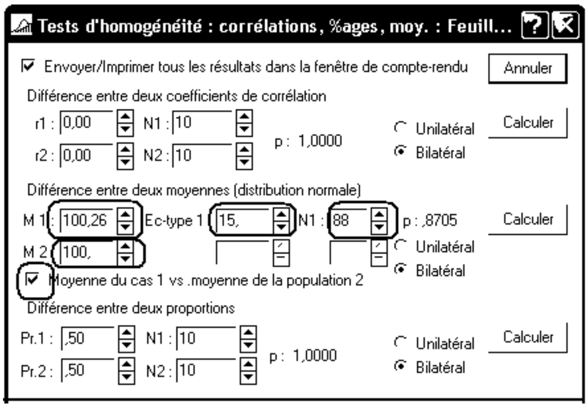
On se pose la question suivante : la population étudiée diffère-t-elle significativement de la population générale (moyenne du QI égale 100) du point de vue du QI ?

- Utilisez le menu Statistiques - Statistiques Elémentaires, puis la méthode Comparer une moyenne à un standard.
- Sélectionnez la variable IQ et indiquez 100 comme valeur de référence. Vous devriez obtenir le résultat suivant :

	Comparaison de moyennes à un standard (constante) (ADD.sta)							
Variable	Moyenne	Ec-Type	N	Erreur-T	Valeur de Référence	Valeur t	dl	p
IQ	100,2614	12,98496	88	1,384201	100,0000	0,188819	87	0,850674

Lecture du résultat :
La moyenne observée sur l'échantillon de 88 sujets est de 100,16. Statistica compare cette valeur à la valeur de référence (100), à l'aide d'un test de Student, avec 87 degrés de liberté. La statistique de test vaut t=0,189, ce qui correspond à un niveau de significativité de 85%. Autrement dit, rien n'indique une différence entre la population étudiée et la population générale du point de vue du QI.

Remarque : Dans le test ci-dessus, Statistica utilise l'estimation de l'écart type des QI faite à partir de l'échantillon, et non l'écart type de la distribution des QI dans la population (15). Il est possible d'introduire l'écart type de la population, à condition d'utiliser le menu Statistiques - Statistiques Elémentaires - Tests d'homogénéité et de remplir la fenêtre de dialogue comme suit :



12.2 Test de comparaison de deux moyennes sur des groupes indépendants

12.2.1 Données saisies "par sujet"

Lorsque la saisie a été faite correctement, la feuille de données Statistica rassemble sur une même ligne les observations relative à un même individu statistique. Ainsi, la saisie des observations relatives à un plan S<A> comportera au moins 2 colonnes :

- Une colonne "Groupe" ou "Condition expérimentale", avec, comme variable nominale, les différents niveaux du facteur A
- Une colonne "Variable dépendante".

Dans ce cas, on utilisera le menu Statistiques - Statistiques élémentaires - Test t pour échantillons indépendants - par groupes.

Exemple : On reprend le classeur Add.stw.
Le score ADDSC est-il significativement différent pour les garçons et les filles dans la population étudiée ?

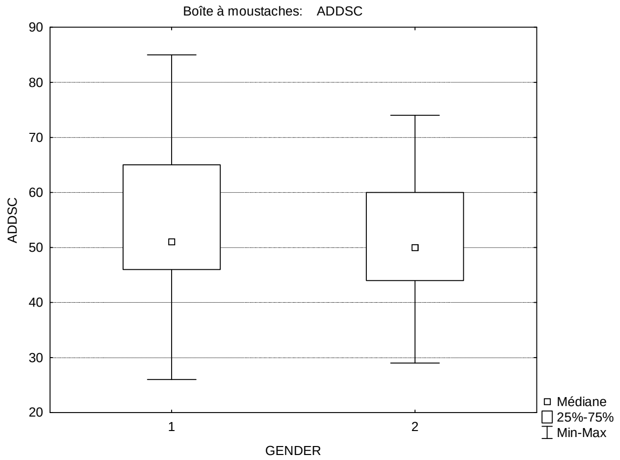
Utilisez le menu Statistiques - Statistiques élémentaires - Test t pour échantillons indépendants - par groupes. Indiquez ADDSC comme variable dépendante et GENDER comme variable de classement.
Statistica devrait produire le résultat suivant :

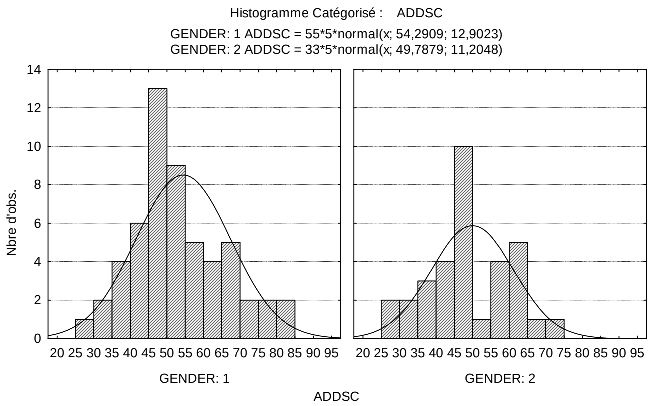
Variable	Tests t ; Classmt : GENDER (ADD.sta) Groupe1: 1 Groupe2: 2 Moyenne 1	Moyenne 2	Valeur t	dl	p	N Actifs 1	N Actifs 2	Ecart-Type 1	Ecart-Type 2	Ratio F Variances	p Variances
ADDSC	54,29091	49,78788	1,662895	86	,099974	55	33	12,90230	11,20479	1,325949	0,395292

Lecture des résultats :
Statistica fait un test t de Student. La valeur observée de la statistique t se lit dans la colonne "Valeur t". Elle est égale à 1,66. Le résultat du test, exprimé en termes de niveau de significativité, est lu dans la colonne "p". On a ici p=0,099974, ce qui correspond à un niveau de significativité de presque 10%. Bien que Statistica ne le mentionne pas, il fait ici un test bilatéral. Autrement dit, les différences ne sont pas significatives au seuil de 5% bilatéral.

Remarques

1. La même méthode permet facilement d'obtenir un graphique de type "boîtes à moustaches" ou "histogrammes catégorisés" comparant les deux situations. Il suffit pour cela d'utiliser les deux boutons ayant ces intitulés disponibles dans l'onglet "Avancé" de la méthode :





2. Il est tout à fait possible d'indiquer à Statistica plusieurs variables dépendantes. On obtient alors dans une seule feuille de résultats, les conclusions des tests sur chacune de ces variables.
Par exemple, reprenez le classeur Add.stw et comparez les moyennes selon le genre, pour l'ensemble des variables numériques, c'est-à-dire : ADDSC, IQ, ENGL, ENGG, GPA. Pour lesquelles d'entre elles observe-t-on une différence significative selon le genre ?

Vous devriez obtenir :

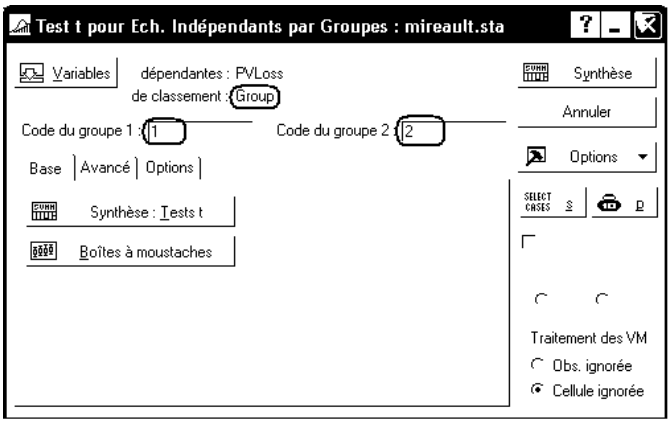
Tests t ; Classmt : GENDER (ADD dans Add.stw) Groupe1: 1 Groupe2: 2	Moyenne 1	Moyenne 2	Valeur t	dl	p
ADDSC	54,29	49,79	1,66	86	0,1000
IQ	99,33	101,82	-0,87	86	0,3867
ENGL	2,05	1,79	2,38	86	0,0197
ENGG	2,44	3,03	-2,98	86	0,0038
GPA	2,24	2,82	-3,27	86	0,0016

12.2.2 Données saisies "par sujet" : test sur une partie des observations

Lorsque le facteur A comporte plus de deux niveaux, il peut être utile de comparer les observations faites pour deux niveaux particuliers.

Reprenons, par exemple, les données Mireault. La variable Group comporte 3 modalités (codées 1, 2 et 3). Nous souhaitons comparer les scores de la variable PVLoss sur les groupes 1 et 2.

Chargez le classeur Mireault.stw, puis utilisez le menu Statistiques - Statistiques élémentaires - Test t pour échantillons indépendants - par groupes. Indiquez PVLoss comme variable dépendante, et Group comme variable de classement. Indiquez ensuite les deux niveaux choisis pour la variable groupe, comme ci-dessous :



Vous devriez obtenir le résultat suivant :

	Tests t ; Classmt : Group (mireault.sta) Groupe1: 2 Groupe2: 1										
Variable	Moyenne 2	Moyenne 1	Valeur t	dl	p	N Actifs 2	N Actifs 1	Ecart-Type 2	Ecart-Type 1	Ratio F Variances	p Variances
PVLoss	17,79	22,21	-6,44	320	0,000000	182	140	5,64	6,68	1,40	0,03

Remarque : Le test de Fisher sur l'égalité des variances conclut ici sur une différence significative des deux variances. On pourra donc recommencer le test en utilisant la boîte à cocher "Estimation séparée des variances", disponible dans le menu Options.

Enregistrez le classeur sur votre compte et refermez-le.

12.2.3 Données saisies par variable

Statistica permet également de réaliser un test de comparaison de moyennes sur deux groupes indépendants lorsque les observations de la VD ont été saisies dans deux colonnes différentes.

Exemple : On reprend l'énoncé suivant :
Lors d'une expérience pédagogique, on s'intéresse à l'effet comparé de deux pédagogies des mathématiques chez deux groupes de 10 sujets:

- pédagogie traditionnelle : Gr1
- pédagogie moderne : Gr2.

On note la performance à une épreuve de combinatoire.
Ces données expérimentales permettent-elles d'affirmer que la pédagogie a un effet sur les résultats à l'épreuve de combinatoire?

Définissez un nouveau classeur contenant une feuille de calcul et saisissez dans cette feuille les données suivantes :

	1 VD-Gr1	2 VD-Gr2
1	5	4
2	4	5,5
3	1,5	4,5
4	6	6,5
5	3	4,5
6	3,5	5,5
7	3	1
8	2,5	2
9	1,5	4,5
10	2,5	4,5

Utilisez ensuite le menu Statistiques - Statistiques Elémentaires - Test t pour échantillons indépendants, par variables. Vous devriez obtenir le résultat suivant :

	Test t pour des Echantillons Indépendants (Feuille de données35) Note : Variables traitées comme des échantillons indépendants								
Groupe1 vs. Groupe2	Moyenne Groupe 1	Moyenne Groupe 2	valeur t	dl	p	N Actifs Groupe 1	N Actifs Groupe 2	Ec-Type Groupe 1	Ec-Type Groupe 2
VD-Gr1 vs. VD-Gr2	3,25	4,25	-1,45	18	0,16	10	10	1,438556	1,637240

Enregistrez le classeur contenant la feuille de données et les résultats du traitement.

12.2.4 Construire une variable calculée pour définir les deux groupes

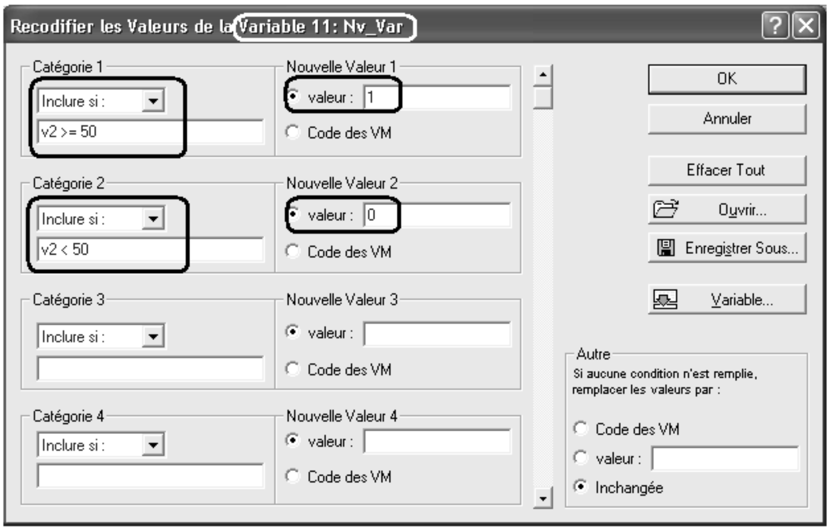
On reprend le classeur ADD.stw

La médiane de la variable ADDSC est égale à 50. On souhaiterait définir deux groupes en utilisant la position de l'observation par rapport à la médiane, et comparer ces deux groupes du point de vue de la variable GPA.

Pour créer une variable définissant les deux groupes selon ce critère, deux méthodes sont possibles : utiliser le menu Données - Recodifier... ou définir une variable calculée.

12.2.4.1 Utilisation du menu Données - Recodifier...

- Ajoutez une colonne supplémentaire au tableau de données (v11).
- Placez la sélection dans cette nouvelle colonne.
- Utilisez le menu Données - Recodifier... et complétez la fenêtre de dialogue comme suit :



12.2.4.2 Utilisation d'une variable calculée

- Ajoutez une colonne supplémentaire (v12) au tableau de données et définissez une variable calculée à l'aide de la formule :

```
=iif(v2 >=50;1 ; 0)
```

ou, de manière équivalente, mais supportant mieux d'éventuelles modifications ultérieures :

```
=iif('ADDSC' >= 50; 1 ; 0)
```

N.B. iif est une fonction ("si immédiat") de Statistica, analogue à la fonction SI d'Excel.

12.2.4.3 Réalisation du test

En utilisant l'une ou l'autre des deux variables définies dans les deux paragraphes précédents, on obtient :

Tests t ; Classmt : Nv_Var (ADD dans Add.stw)							
Groupe1: 0 Groupe2: 1							
	Moyenne 0	Moyenne 1	Valeur t	dl	p	N Actifs 0	N Actifs 1
GPA	2,9987	2,0440	**6,1406**	86	**0,000000**	*38*	*50*

Lecture du résultat du test : les moyennes dans le groupe 1 et le groupe 0 sont respectivement de 2,9987 et 2,0440. Le test de Student conduit à une statistique t égale à 6,1406. Le niveau de significativité correspondant est inférieur à 1 sur 1 000 000 ; on conclut donc sur H_1 : il existe une différence très significative entre les scores au GPA dans les deux populations parentes.
3
On remarque également que les groupes sont assez mal équilibrés (échantillons de tailles respectives 38 et 50) bien que la division en deux groupes ait été faite sur la base de la médiane. Ce déséquilibre est dû au grand nombre d'ex aequo pour lesquels ADDSC est égal à la médiane.

13 Test de comparaison de deux moyennes sur des groupes appareillés

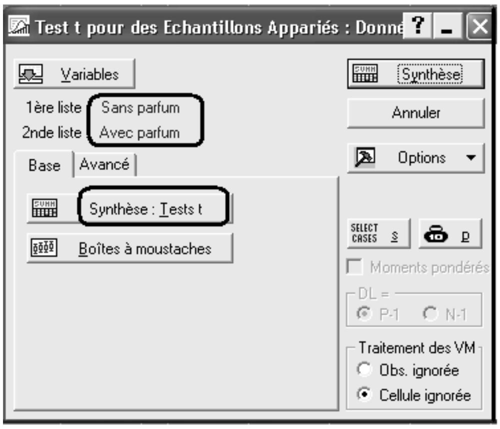
13.1 Comparer les scores observés dans deux conditions

On reprend le classeur Parfums1.stw (données présentées dans le polycopié N° 1) et on souhaite étudier si les temps de parcours du labyrinthe sont significativement différents dans la condition "Avec Parfum" et dans la condition "Sans Parfum".

Il s'agit là de données appariés, puisqu'il s'agit des mêmes sujets, évalués dans les deux conditions, sur deux épreuves présentant le même niveau de difficulté.

- Utilisez ensuite le menu Statistiques - Statistiques Elémentaires puis la méthode "Test t pour des échantillons appariés".
- Sélectionnez "Sans parfum" comme variable pour constituer la première liste, et "Avec parfum" comme variable pour constituer la seconde.

- Cliquez sur le bouton "Synthèse" ou le bouton "Synthèse : Tests t"

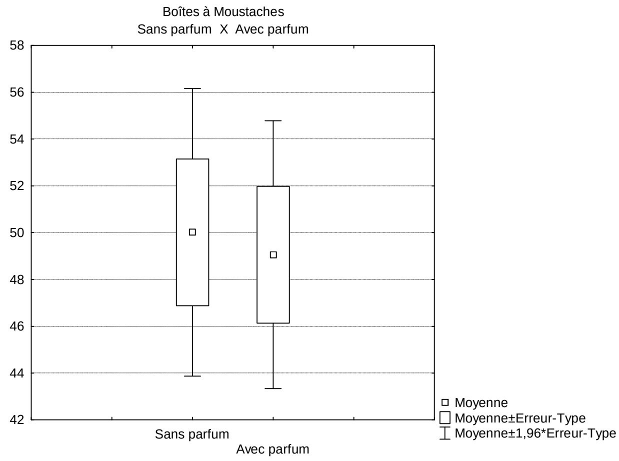


Vous devriez obtenir le résultat suivant :

Test t pour des Echantillons Appariés (Données Parfums dans Parfums1.stw) Différences significatives marquées à p < ,05000								
	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
Sans parfum	50,01	14,36						
Avec parfum	49,06	13,39	21	**0,96**	12,55	**0,35**	20	**0,73**

Lecture du résultat du test. La moyenne de la série des différences individuelles est de 0,96. La statistique de test suit une loi de Student et vaut 0,35 (colonne "t"). Le niveau de significativité correspondant est lu dans la colonne "p" : p-value = 0,73 = 73%. On conclut donc sur H0 aux seuils traditionnels : on n'a pas mis en évidence de différence significative entre les scores dans les deux conditions.

Remarque. Le bouton "Boîtes à moustaches" de la fenêtre de dialogue "Tests t pour des échantillons appariés" permet de construire un graphique comparant les deux conditions :



13.2 Comparer deux à deux les scores observés dans plusieurs conditions

Cette méthode permet d'indiquer plusieurs variables dans chacune des deux listes. Chacune des variables de la première liste est alors comparée à chaque variable de la deuxième liste.

Exemple :

Ouvrez le classeur "Performance-Cognitive.stw".

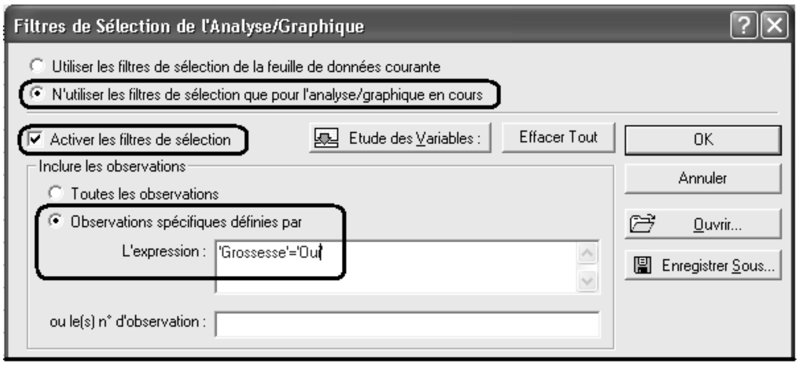
Dans une étude publiée en 2002, des chercheurs se sont intéressés à l'existence éventuelle d'un affaiblissement de la performance cognitive durant la grossesse. Dans cette étude, deux groupes de femmes ont été observés : 13 d'entre elles étaient enceintes, les 13 autres ne l'étaient pas et constituent le groupe "Contrôle" (variable : Grossesse - modalités : Oui / Non).

La performance cognitive est évaluée à l'aide du score au CFQ (Cognitive Failure Questionnaire). Il s'agit d'un questionnaire à items multiples demandant une auto-évaluation de la fréquence des troubles de la mémoire ou de la concentration, conduisant à une échelle de scores de 0 à 100. Chaque sujet est évalué à 3 reprises correspondant, pour les femmes enceintes au second et au troisième trimestre de la grossesse et au suivi 5 mois après l'accouchement (variables : CFQ-1 à CFQ-3).

On souhaite comparer deux à deux les scores CFQ-1 à CFQ-3 d'une part chez les femmes enceintes et d'autre part dans le groupe contrôle.

Comparaison des scores chez les femmes enceintes

On utilise comme précédemment la méthode "Test t pour des échantillons apparés", en indiquant les trois variables CFQ-1 à CFQ-3 dans chacune des deux listes de variables.
Par ailleurs, on séléctionne le groupe des femmes enceintes en réalisant un filtre, valable pour cette seule analyse :



Vous devriez aboutir au résultat suivant : chez les femmes encientes, les 3 scores ne semblent pas différents.

Test t pour des Echantillons Appariés (Donnees dans Performances-Cognitives.stw) Filtres de Sélection - 'Inclure' : 'Grossesse'='Oui'								
	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
CFQ-1	49,08	19,05						
CFQ-1	49,08	19,05	13	0,00	0,00	0,00	12	1,00
CFQ-1	49,08	19,05						
CFQ-2	53,23	18,87	13	-4,15	18,56	-0,81	12	0,44
CFQ-1	49,08	19,05						
CFQ-3	48,38	15,23	13	0,69	28,66	0,09	12	0,93
CFQ-2	53,23	18,87						
CFQ-1	49,08	19,05	13	4,15	18,56	0,81	12	0,44
CFQ-2	53,23	18,87						
CFQ-2	53,23	18,87	13	0,00	0,00	0,00	12	1,00
CFQ-2	53,23	18,87						
CFQ-3	48,38	15,23	13	4,85	23,60	0,74	12	0,47
CFQ-3	48,38	15,23						
CFQ-1	49,08	19,05	13	-0,69	28,66	-0,09	12	0,93
CFQ-3	48,38	15,23						
CFQ-2	53,23	18,87	13	-4,85	23,60	-0,74	12	0,47
CFQ-3	48,38	15,23						
CFQ-3	48,38	15,23	13	0,00	0,00	0,00	12	1,00

On reprend le même test, mais en sélectionnant maintenant le groupe-contrôle (condition 'Grossesse'='Non').
On obtient alors :

Test t pour des Echantillons Appariés (Donnees dans Performances-Cognitives.stw) Filtres de Sélection - 'Inclure' : 'Grossesse'='Non'

	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
CFQ-1	42,62	10,05						
CFQ-1	42,62	10,05	13	0,00	0,00	0,00	12	1,00
CFQ-1	42,62	10,05						
CFQ-2	41,54	8,80	13	1,08	13,74	0,28	12	0,78
CFQ-1	42,62	10,05						
CFQ-3	33,23	9,53	13	9,38	9,16	**3,69**	12	**0,00**
CFQ-2	41,54	8,80						
CFQ-1	42,62	10,05	13	-1,08	13,74	-0,28	12	0,78
CFQ-2	41,54	8,80						
CFQ-2	41,54	8,80	13	0,00	0,00	0,00	12	1,00
CFQ-2	41,54	8,80						
CFQ-3	33,23	9,53	13	8,31	9,90	**3,02**	12	**0,01**
CFQ-3	33,23	9,53						
CFQ-1	42,62	10,05	13	-9,38	9,16	**-3,69**	12	**0,00**
CFQ-3	33,23	9,53						
CFQ-2	41,54	8,80	13	-8,31	9,90	**-3,02**	12	**0,01**
CFQ-3	33,23	9,53						
CFQ-3	33,23	9,53	13	0,00	0,00	0,00	12	1,00

On constate que dans le groupe-contrôle, le score au premier test est significativement différent de ceux observés aux deuxième et troisième test.

En combinant les résultats précédents, on en déduit que le comportement des femmes enceintes semble significativement différent de celui des femmes du groupe-contrôle.

14 Test de comparaison de deux proportions sur des groupes indépendants

Deux échantillons provenant de deux populations différentes ont passé un test commun.
Dans le premier groupe, d'effectif 150, le taux de succès a atteint 68%. Autrement dit, 102 sujets ont passé le test avec succès, et 48 ont échoué.
Dans le deuxième groupe, d'effectif 180, le taux de succès a atteint 55,5%. Autrement dit, 100 sujets ont passé le test avec succès et 80 ont échoué.

Peut-on dire que la seconde population réussit l'épreuve moins facilement que la première ?

Statistica ne comporte pas de module spécifiquement destiné à traiter ce genre de situation. Cependant, la comparaison de deux proportions n'est qu'un cas particulier de la comparaison de deux moyennes.
Nous allons donc saisir nos données dans une feuille de calcul Statistica, de la façon suivante :

	1 Groupe	2 Resultat	3 Effectifs	V
1	A	Succès	102	
2	A	Echec	48	
3	B	Succès	100	
4	B	Echec	80	
5				

Pour faciliter l'interprétation des résultats, il sera commode de personnaliser le codage de la variable "Resultat"
Succès sera codé 1 tandis que Echec sera codé 0. Vous allez donc définir "Succès" comme valeur-texte correspondant à la valeur numérique 1 et "Echec" comme valeur-texte correspondant à la valeur numérique 0.

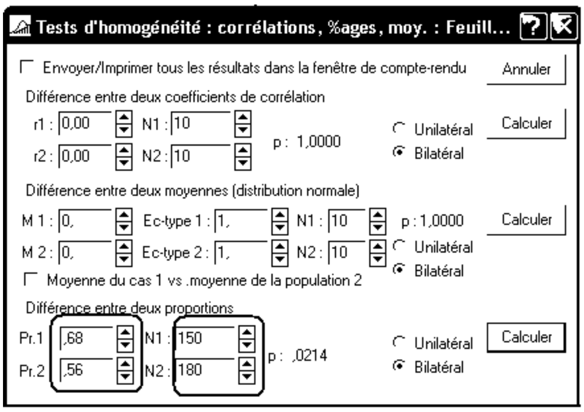
N'oubliez pas d'indiquer à Statistica que la colonne Effectifs est une colonne de pondérations.

Statistica fournit les résultats suivants :

	Tests t ; Classmt : Groupe (Feuille de données1) Groupe1: A Groupe2: B				
Variable	Moyenne A	Moyenne B	Valeur t	dl	p
Resultat	0,680000	0,555556	2,321956	328	0,020847

Lecture du résultat. On retrouve ainsi les valeurs des proportions observées (68% et 55,55%). La valeur observée de la statistique de test est 2,32. La formule de calcul n'est pas exactement la même que celle donnée en cours, mais la différence est très faible (2,3219 ici, alors que la formule du cours donnerait 2,3201). Statistica utilise une loi de Student avec 328 ddl. Vu le nombre de degrés de liberté, cette loi est très proche d'une loi normale, et le résultat produit est tout à fait correct. On obtient ici une p-value (colonne "p") égale à 0,020, c'est-à-dire à 2%. Au seuil de 5%, on conclut donc à une différence significative entre les deux groupes.

On peut aussi utiliser le menu Statistiques - Statistiques Elémentaires - Tests d'homogénéité. On complète alors la fenêtre comme suit :



Rappel des formules du cours :

$$p = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

$$Z = \frac{f_1 - f_2}{E} \text{ avec } E^2 = p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Exercice. 1) Faites varier les taux de succès dans les deux groupes. Que devient le résultat du test lorsque les taux varient ?

2) Faites varier les effectifs dans les deux groupes. Avec les taux de succès indiqués, quelles sont les tailles minimales des échantillons permettant d'obtenir un résultat significatif à 5% ?