

# Licence de Psychologie - Semestre N° 5 - TD n° 2

## Test du khi-2, lois de distribution classiques et tests paramétriques avec Statistica

### 10 Organiser son espace de travail avec Statistica

#### 10.1 Modifier le comportement de Statistica

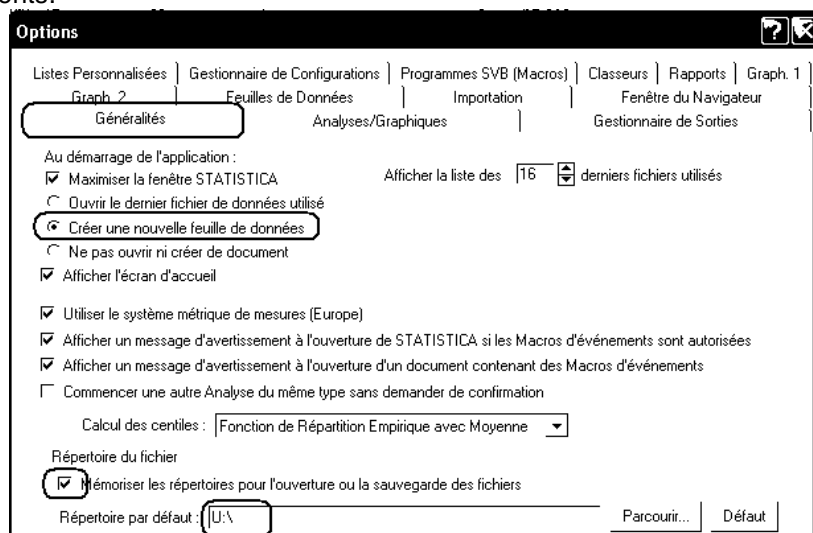
Le comportement de Statistica peut être modifié en intervenant dans la fenêtre de dialogue affichée par le menu Outils - Options.

Par exemple, nous souhaitons :

- que Statistica n'ouvre plus systématiquement la dernière feuille de données utilisée lors du chargement du logiciel ;
- que Statistica nous propose par défaut le volume U: pour enregistrer nos documents, au lieu du répertoire "Mes Documents".

Exécutez le menu Outils - Options. Sous l'onglet Généralités, activez le bouton radio "Créer une nouvelle feuille de données".

Désactivez la boîte à cocher "mémoire des répertoires pour l'ouverture ou la sauvegarde des fichiers". Complétez la zone d'édition "Répertoire par défaut" en indiquant U:\, puis réactivez la boîte à cocher (N.B. Bien que l'option soit en apparence désactivée, Statistica proposera par défaut le répertoire U:\ pour l'enregistrement de nouveaux documents).



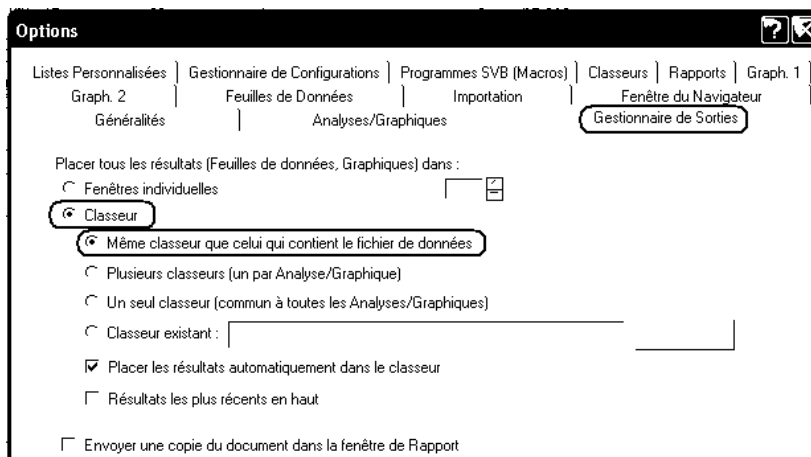
#### 10.2 Gérer les sorties

Lorsqu'on utilise Statistica sans se préoccuper des options de sortie des résultats, on se retrouve vite à la tête d'une quantité de fenêtres (classeurs, feuilles de données de résultats, fenêtres de graphiques...). Pour réaliser un travail que l'on souhaite conserver et reprendre au cours de plusieurs séances de travail, il paraît indispensable d'organiser correctement son espace de travail et ses sauvegardes.

##### 10.2.1 Enregistrer données et résultats dans un seul classeur

Cette méthode consiste à enregistrer les données, les résultats de traitements, et les commentaires éventuels comme objets d'un même classeur. Ainsi, un unique fichier du disque rassemble l'ensemble de notre travail sur un cas donné.

Ce comportement correspond aux réglages suivants dans le menu Outils - Options - Onglet Gestionnaire de Sorties :



**Remarque :** Le réglage ne sera actif que si la feuille de données se trouve effectivement dans un classeur. Or, ce ne sera pas le cas si la feuille de données a été ouverte à partir d'un fichier \*.sta, ou importée à partir d'une feuille Excel. Dans ce cas, vous devez insérer la feuille de données dans le classeur comme il a été indiqué au paragraphe précédent.

### 10.2.2 Indiquer quelle est la feuille de données active

Statistica utilise la feuille de données active comme source des données pour effectuer les différents traitements que nous demandons. Cette notion permet de résoudre plusieurs problèmes :

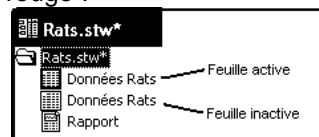
- Ouvrir plusieurs fichiers .sta et effectuer un travail sur l'un d'eux (pas nécessairement le dernier ouvert)
- Utiliser une feuille de résultats comme feuille de données pour des traitements ultérieurs.
- Lorsque l'on travaille avec une feuille de données insérée dans un classeur, il arrive couramment que Statistica ne retrouve pas la feuille à partir de laquelle les traitements doivent être effectués. Mais on peut éviter ce comportement en spécifiant la propriété "feuille de données active" pour l'objet du classeur qui contient nos données.

Pour spécifier comme feuille de données active une feuille d'un classeur :

- Cliquez avec le bouton droit de la souris sur l'icône de la feuille de données dans le volet gauche du classeur.
- Utilisez l'item Feuille de données active du menu local.

On peut également utiliser le menu Données - Feuille de données active.

Remarquez que le volet gauche d'un classeur indique si une feuille insérée dans le classeur est active ou non : l'icône d'une feuille active est encadrée en rouge :



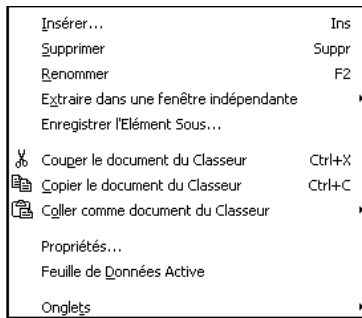
### 10.2.3 Enregistrer les données et l'ensemble des traitements réalisés dans un même classeur

Ouvrez un fichier de données (un fichier d'extension .sta) et réalisez un ou plusieurs traitements relatifs à ces données (par exemple, des statistiques descriptives et un graphique). Si vous avez gardé les options par défaut de Statistica, les résultats de tous ces traitements se trouvent dans un classeur.

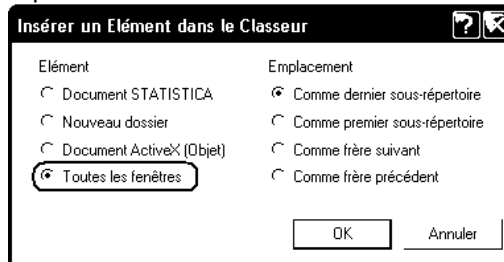
Pour enregistrer données, traitements et rapport dans un seul classeur :

Affichez la fenêtre du classeur contenant les résultats.

Cliquez avec le bouton droit de la souris dans le volet gauche de la fenêtre du classeur.



Sélectionnez l'item Insérer..., puis l'option "Toutes les fenêtres" :



N'oubliez pas, ensuite, de spécifier la feuille contenant les données de base comme feuille active du classeur.

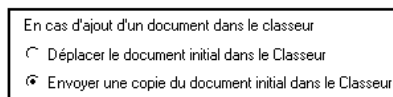
### 10.3 Manipuler les objets contenus dans un classeur

#### 10.3.1 Copier - coller entre classeurs, entre un classeur et un objet Statistica

Pour déplacer un objet d'un classeur à un autre, il suffit de déplacer son icône depuis le volet gauche du premier classeur dans le volet gauche du second. On peut également utiliser les menus locaux Copier et Coller obtenus à l'aide d'un clic droit dans le volet gauche de chaque classeur.

Le menu local "Insérer" du volet gauche d'un classeur permet également d'insérer dans ce classeur un document contenu dans une fenêtre indépendante. Il suffit de choisir les options : Document Statistica - Créer à partir d'une fenêtre.

L'opération faite par Statistica est soit une copie (l'original de l'objet est conservé) soit un déplacement (l'original de l'objet n'est pas conservé) selon le paramétrage choisi dans le menu Outils - Options - Onglet Classeurs - Item "En cas d'ajout d'un document dans le classeur".



#### 10.3.2 Supprimer un objet d'un classeur

Il est également possible de supprimer un objet d'un classeur, à l'aide d'un clic droit et de l'item de menu Supprimer. Cela permet notamment de ne garder, pour un traitement donné, que le résultat le plus abouti. Attention cependant : lorsque l'on supprime un objet qui n'est pas une feuille de la hiérarchie, on supprime en même temps tous les objets qui en dépendent.

## 11 Travail sur un tableau de contingence - Test du khi-2

### 11.1 Enoncé

Une étude a été menée en 1990-91 sur les facteurs pouvant influencer sur le port de la ceinture de sécurité par les conducteurs et les passagers de voitures de tourisme et de véhicules utilitaires. De nombreuses observations ont été effectuées (9434 au total), et ont donné lieu au relevé des éléments suivants :

- Nature du véhicule (voiture de tourisme / véhicule utilitaire)
- Age du conducteur (trois classes d'âge)
- Sexe (M / F)
- Port de la ceinture (port / non port)

- Présence d'un passager avant (oui / non)
- Le cas échéant, âge, sexe et port de la ceinture pour le passager
- Présence de passagers arrière (oui / non)

On s'intéresse tout d'abord à l'effet du type d'occupation du véhicule (conducteur seul, conducteur + passagers avant, conducteur + passagers arrière, conducteur + passagers avant et arrière) sur le port de la ceinture par le conducteur. On dispose de 8374 observations concernant cette partie de l'étude. Les données sont les suivantes :

	Port ceinture	non port de ceinture
Seul	2825	3468
Cond. + pass. avant	729	815
Cond. + pass. arrière	80	113
Cond. + pass. av. et arr.	168	176

On souhaite en particulier tester l'existence d'un lien entre les deux variables "Type d'occupation" et "Port de la ceinture" à l'aide d'un test du khi-2.

## 11.2 Mise en oeuvre du test du khi-2

### 11.2.1 Le test du khi-2 à partir d'un tableau protocole

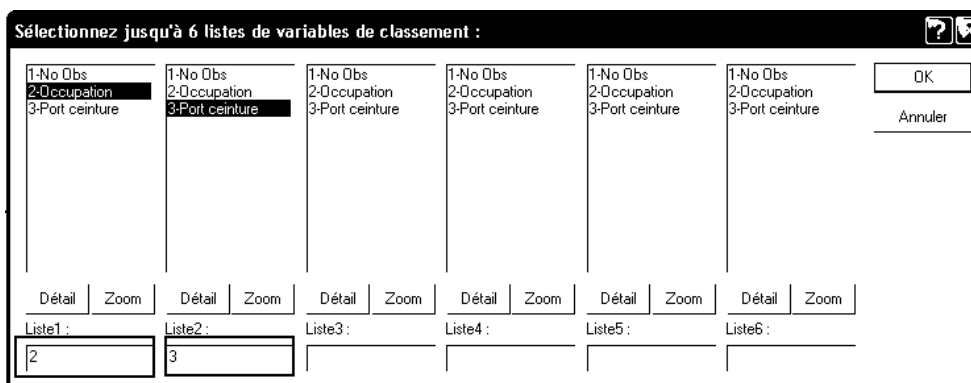
En général, une ligne d'une feuille de données Statistica correspond à **une** observation. Autrement dit, nous devions ici avoir 8374 lignes du type suivant :

N° obs	Type d'occupation	Port ceinture
1	Seul	Oui
2	Seul	Oui
...	...	...
8374	Cond. + pass. av. et arr.	Non

Ouvrez le classeur Ceinture.stw et affichez la feuille de données Ceinture-protocole. Observez la façon dont elle a été constituée.

#### 11.2.1.1 Vérifier la saisie des données

La feuille de données Ceinture-protocole correspond-elle à l'énoncé ci-dessus ? On peut le vérifier en utilisant le menu : Statistiques - Statistiques élémentaires - Tableaux et tris croisés, et en sélectionnant l'onglet Tableaux croisés ou l'onglet Tris croisés. Pour l'onglet Tableaux croisés, le bouton "Spécifier les tables (sélection des variables)" permet d'avoir accès au dialogue suivant :

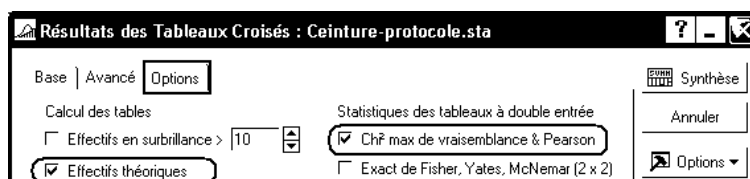


Avec l'onglet Tris croisés, le bouton "Spécifier les tables" affiche un dialogue analogue, mais beaucoup plus simple. On sélectionne alors des deux variables "Occupation" et "Port ceinture". Après avoir validé ce dialogue, on peut cliquer sur le bouton "Synthèse" pour obtenir un tableau de contingence du type suivant :

Occupation	Port ceinture Oui	Port ceinture Non	Totaux Ligne
Seul	2825	3468	6293
Avant	729	815	1544
Arrière	80	113	193
Avant et arrière	168	176	344
Ts Grpes	3802	4572	8374

### 11.2.1.2 Test du khi-2 sur un tableau protocole - Première méthode

Le test du khi-deux sur ce tableau de contingence peut être obtenu comme résultat supplémentaire de l'étude en cours. Pour cela, reprenez l'analyse en cours, activez l'onglet "Options" et cochez les cases "Chi<sup>2</sup> max de vraisemblance et Pearson" et un autre item de l'onglet, "Effectifs théoriques", par exemple).



Outre les effectifs théoriques, on obtient comme résultat la valeur de la statistique du khi-2 ( $\chi^2 = 5,56$ ) et son niveau de significativité ( $p=0,13$ ) :

Occupation	Port ceinture Oui	Port ceinture Non	Totaux Ligne
Seul	2857,175	3435,825	6293,000
Avant	701,014	842,986	1544,000
Arrière	87,627	105,373	193,000
Avant et arrière	156,184	187,816	344,000
Ts Grpes	3802,000	4572,000	8374,000

#### Remarque importante :

Statistica (comme tous les logiciels de Statistiques) indique le résultat du test en termes de niveau de significativité (valeur de  $p$ ) et non en termes de valeur observée / valeur critique pour un seuil donné. Il faut donc interpréter le résultat en appliquant la règle vue dans le cours de Statistiques :

- Si le niveau de significativité  $p$  est supérieur ou égal au seuil  $\alpha$  que l'on s'est fixé, c'est l'hypothèse nulle  $H_0$  qu'il convient de retenir ;
- Si le niveau de significativité  $p$  est strictement inférieur au seuil  $\alpha$  que l'on s'est fixé, c'est l'hypothèse alternative  $H_1$  qu'il convient de retenir.

### 11.2.1.3 Test du khi-2 sur un tableau protocole - variante

Une deuxième méthode pour obtenir le khi-2, avec une meilleure présentation, est de procéder comme suit :

- Utiliser le menu Statistiques - Statistiques élémentaires - Tableaux et tris croisés
- Sélectionner l'onglet "Tableaux croisés" ou l'onglet "Tris croisés".
- Cliquer sur le bouton "Spécifier les tables (sélection des variables)" et spécifier "Occupation" comme première variable et "Port ceinture" comme deuxième variable.
- Après avoir validé, afficher l'onglet "Options" et cocher la case "Chi<sup>2</sup> max de vraisemblance et Pearson"
- Enfin, afficher l'onglet "Avancé" et cliquer sur le bouton "Tableaux détaillés à double entrée".

On obtient alors parmi les résultats la feuille de données suivante :

Statistique	Stats : Occupation(4) x Port ceinture(2) (Ceinture-protocole.sta)		
	Chi?	dl	p
Chi? de Pearson	5,563066	dl=3	p=,13492
Chi? Max-Vr.	5,561035	dl=3	p=,13504

Rassemblez au besoin les différentes feuilles de résultats dans le classeur Ceinture.stw et enregistrez-le.

### 11.2.2 Le test du khi-2 à partir d'un tableau d'effectifs

Si les données que nous devons traiter sont des données que nous avons nous-mêmes recueillies, nous disposons sans doute d'un tableau protocole, et le traitement précédent convient. Cependant, si nous disposons au départ d'un tableau d'effectifs ou d'un tableau de contingence, il est évidemment très fastidieux de composer une feuille de données du type précédent.

Cependant, Statistica permet de travailler sur des données **pondérées**. Nous allons donc saisir nos données de la façon suivante :

	Type occupation	Ceinture	Effectif
1	seul	oui	2825
2	avant	oui	729
3	arrière	oui	80
4	avant et arrière	oui	168
5	seul	non	3468
6	avant	non	815
7	arrière	non	113
8	avant et arrière	non	176

Saisissez ces données dans une nouvelle feuille du classeur Ceinture.stw, nommez cette feuille Ceinture-effectifs. et rendez-la active.

Vérifiez bien que vous définissez exactement 4 modalités pour la variable "Type d'occupation" et 2 modalités pour la variable "ceinture". Pour cela, faites un double-clic sur la colonne correspondante, puis cliquez sur le bouton : Valeurs/Stats...

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 :



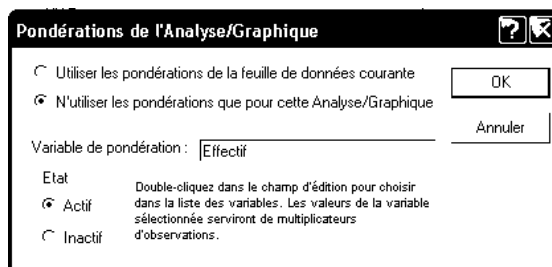
. Les pondérations s'appliquent alors à toutes les analyses utilisant cette feuille.

Dans le deuxième cas, on utilise le menu Statistiques Élémentaires et on clique sur le bouton "pondérations"



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

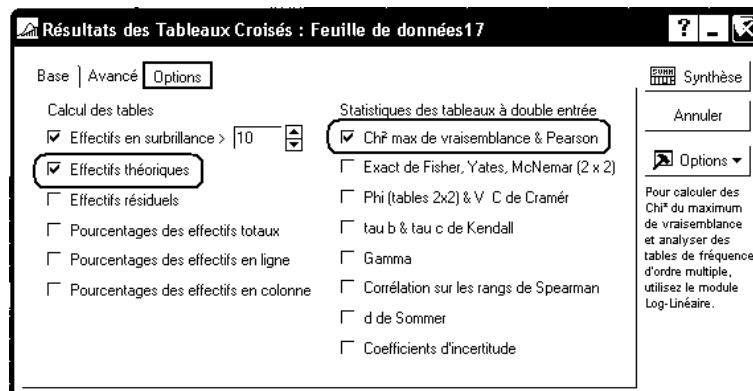
Indiquez que la variable 3 (Effectif) est la variable de pondération :



Le reste du traitement peut alors être réalisé de la même façon que pour un tableau-protocole.

Pour réaliser un test du khi-2, utilisez l'item Tableaux et tris croisés. Indiquez Type occupation comme variable dans la première liste, Ceinture dans la deuxième, puis cliquez sur le bouton OK. La fenêtre de dialogue "Résultats des tableaux croisés" s'affiche pour nous permettre de sélectionner les résultats dont nous souhaitons le calcul.

Affichez l'onglet Options et sélectionnez "Effectifs théoriques" et "Chi2 ...".



Statistica affiche alors un tableau de contingence classique, le tableau des effectifs théoriques et le résultat du test du khi-2 :

Synthèse : Effectifs Théoriques (Feuille de données17)			
Effectifs en surbrillance > 10			
Chi² de Pearson : 5,56307, dl=3, p=,134922			
Type occupation	Ceinture oui	Ceinture non	Totaux Ligne
seul	2857,175	3435,825	6293,000
avant	701,014	842,986	1544,000
arrière	87,627	105,373	193,000
avant et arrière	156,184	187,816	344,000
Ts Grpes	3802,000	4572,000	8374,000

Ici, le khi-2 observé vaut 5,56, et son niveau de significativité est de 13,5%. On retient donc l'hypothèse H0 d'absence de lien entre les variables.

Rassemblez au besoin les différentes feuilles de résultats dans le classeur Ceinture.stw, et enregistrez-le.

### 11.2.3 Le test du khi-2 à partir d'un tableau de contingence

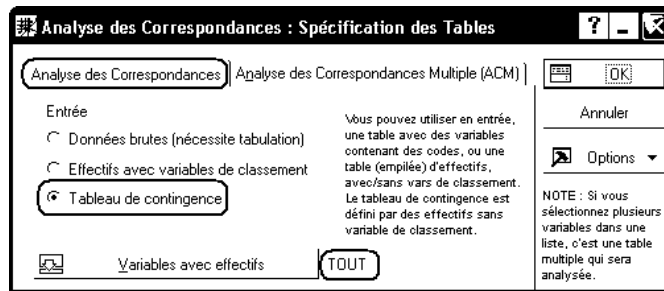
On peut aussi fournir à Statistica un tableau de contingence, sous la forme suivante :

Occupation	Port ceinture	Port ceinture
	Oui	Non
Seul	2825	3468
Avant	729	815
Arrière	80	113
Avant et arrière	168	176

Parmi les méthodes d'analyse qui sont présentes dans le menu Statistiques, l'une d'elles accepte des données structurées sous cette forme, et calcule un khi-deux de contingence comme résultat annexe. C'est la méthode "Techniques exploratoires multivariées - Analyse de correspondances".

On pourra procéder de la manière suivante :

- Saisir les données dans une nouvelle feuille de données du classeur Ceinture.stw, et la nommer Ceinture-contingence.
- Rendre active la feuille Ceinture-contingence.
- Utiliser le menu Statistiques - Techniques exploratoires multivariées - Analyse de correspondances
- Activer l'onglet "Analyse des correspondances" et compléter la fenêtre de dialogue comme suit :



- Cliquer sur le bouton OK
- La valeur du khi-2 et son niveau de significativité apparaissent alors dans l'en-tête de la fenêtre de dialogue suivante :

Nombre de variables (colonnes de la table) : 2  
 Nombre d'observations actives (lignes de la table) : 4

Valeurs propres : ,0007  
 Chi<sup>2</sup> Total = 5,56307 dl = 3 p = ,1349

- Pour obtenir des détails sur la manière dont ce khi-2 a été obtenu (effectifs théoriques, ou contributions au khi-2), on pourra activer l'onglet "Etude".

**Remarque** : Cette méthode permet également d'obtenir le calcul du khi-2 à partir du tableau protocole ou du tableau d'effectifs utilisé au paragraphe précédent, et en évitant de définir des pondérations.

Pour travailler à partir du tableau des effectifs, par exemple :

- Chargez la feuille de données Ceinture-effectifs.sta saisie au paragraphe précédent
- Utilisez le menu Statistiques - Techniques exploratoires multivariées - Analyse de correspondances
- Cochez "Effectifs avec variables de classement" comme type de données d'entrée.

Rassemblez au besoin les différentes feuilles de résultats dans le même classeur et enregistrez-le et refermez-le.

### 11.2.4 Exercice :

1) On s'intéresse, pour les conducteurs non accompagnés, au lien entre le sexe et le port de la ceinture. Les données sont les suivantes :

	Port ceinture	non port de ceinture
Homme	1981	2647
Femme	844	821

Saisissez ces données dans une feuille de données Statistica.

Représentez graphiquement le taux de port de la ceinture selon le sexe à l'aide d'un graphique à barres.

Réalisez, de même, un test du khi-2 pour déterminer si le port de la ceinture par le conducteur dépend ou non du sexe du conducteur.

2) On se limite ici aux véhicules dans lesquels se trouvaient des passagers. On s'intéresse d'une part au port de la ceinture par la paire conducteur/passager avant et d'autre part au type de véhicule. Les données sont les suivantes :

	Véh. de tourisme	Véh. utilitaire
Cond. sans ceint., pass avec ceint.	199	11
Cond. et pass. sans ceinture	596	111
Cond. et pass. avec ceinture	549	24
Cond. avec ceinture, pass. sans ceinture	161	10

Ouvrez un nouveau document Statistica et saisissez les données ci-dessus.



Représenter par un diagramme circulaire la répartition des paires d'occupants de véhicules de tourisme selon les 4 modalités de la variable "port de la ceinture".

Réalisez un test du khi-2 pour déterminer si les variables "comportement des occupants vis-à-vis du port de la ceinture" et "nature du véhicule" sont indépendantes ou non.

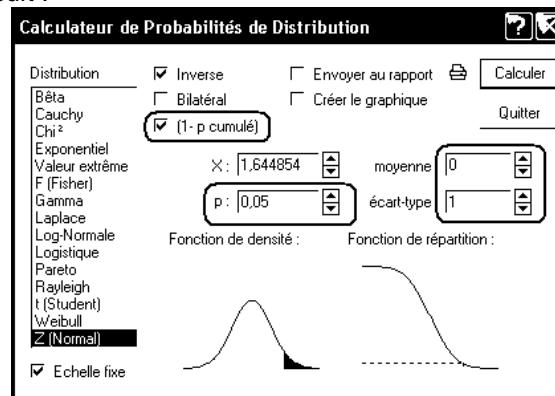
## 12 Lois statistiques classiques

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.

### 12.1 La loi normale centrée réduite

Quelle est la valeur de Zcritique pour un test unilatéral à 5% ?

Compléter le dialogue comme suit :

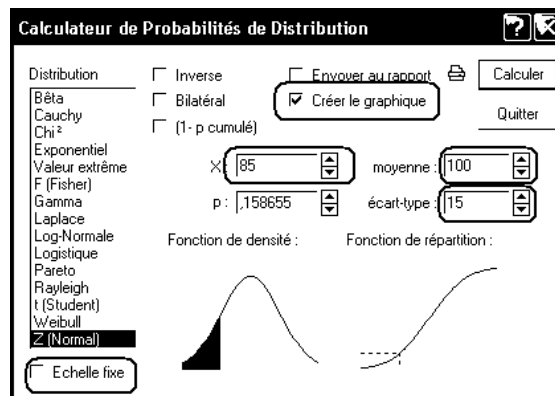


La réponse doit être lue dans la zone d'édition "X: \_\_\_\_\_". C'est ici :  $Z_c=1,644854$ .

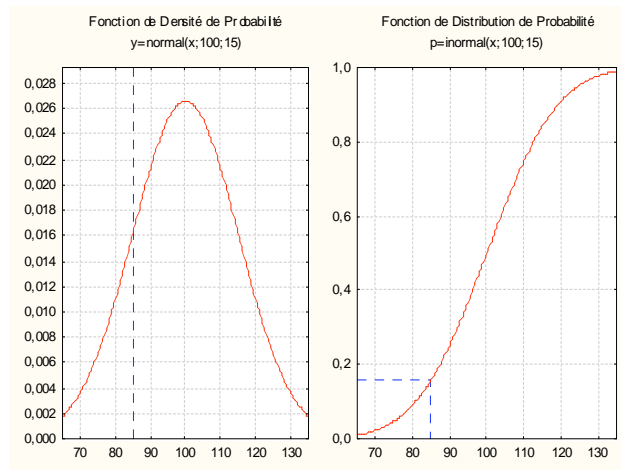
### 12.2 Représenter graphiquement la densité d'une loi normale quelconque

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 boîte "Echelle Fixe" ne soit pas cochée :

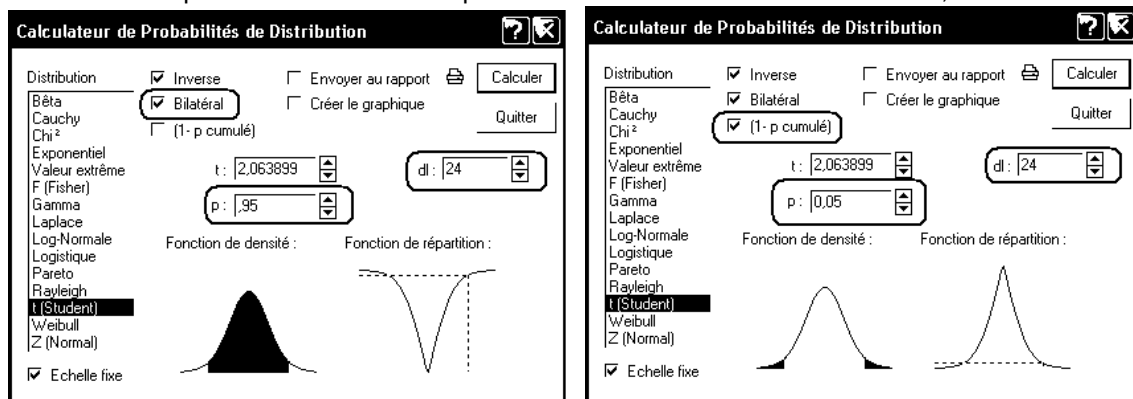


Remarquez que Statistica exige que l'un des deux champs "X:" ou "p:" soit complété, et repère la valeur correspondante sur le graphique. On obtient ainsi un graphique ayant l'allure suivante :



### 12.3 Loi de Student

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



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.

### 12.4 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 :  $\chi^2 = 12,59$ .

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).

## 13 Tests paramétriques classiques

### 13.1 Test de comparaison d'une moyenne à une norme

Ouvrez le classeur ADD.stw et consultez la présentation des données qui se trouve dans le rapport contenu dans ce classeur.

On se pose la question suivante : la population étudiée diffère-t-elle significativement de la population générale du point de vue du IQ ?

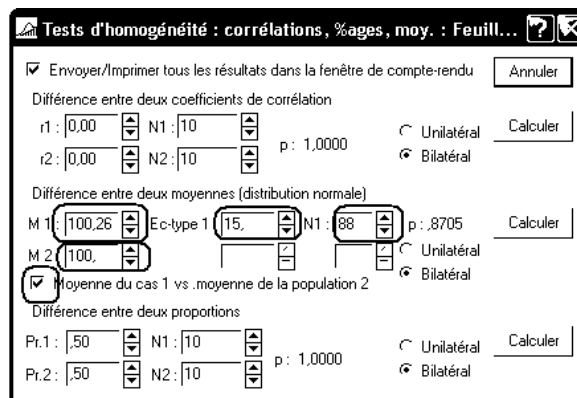
Utilisez le menu Statistiques - Statistiques Élémentaires, puis l'item Comparer une moyenne à un standard. Sélectionnez la variable IQ et indiquez 100 comme valeur de référence. Vous devriez obtenir le résultat suivant :

Variable	Comparaison de moyennes à un standard (constante) (ADD.sta)							
	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 IQ.

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



### 13.2 Test de comparaison de deux moyennes sur des groupes indépendants

#### 13.2.1 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".

	1	2
	Groupe	VD
1	1	5,7
2	1	4,8
3	2	7,6
4	2	6,5

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

**Exemple** : 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 :

Tests t ; Classmt : GENDER (ADD.sta)											
Groupe1: 1											
Groupe2: 2											
Variable	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 est  $t=1,66$ , 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.

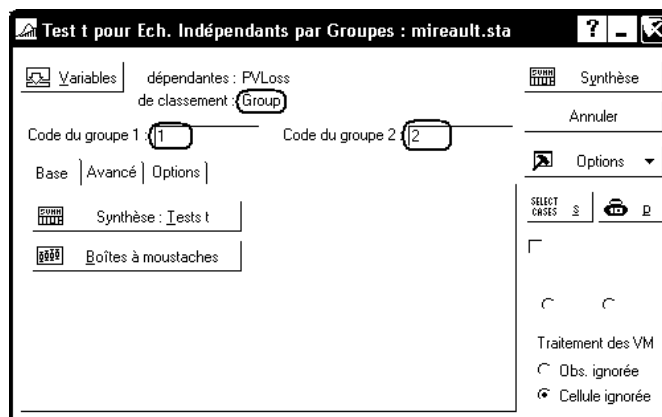
Insérez si besoin les fenêtres de résultats dans le classeur, enregistrez le classeur ADD.stw sur votre compte et refermez-le.

**13.2.2 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 /ariances	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".

Enregistrez le classeur sur votre compte et refermez-le.

### 13.2.3 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 Élémentaires - 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.

### 13.2.4 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.

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

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

N.B. iif est une fonction ("si immédiat") de Statistica, analogue à la fonction SI d'Excel. Réalisez ensuite le test sur la variable GPA pour les deux groupes ainsi définis.

## 14 Test de comparaison de deux moyennes sur des groupes appariés

On reprend l'exemple ADD.dat.

Les variables ENGG et GPA représentent les résultats obtenus en Anglais d'une part, et la moyenne des points obtenus en 9<sup>e</sup> année d'autre part. Ces résultats utilisent la même échelle de notation : le score varie de 0 à 4. Nous souhaiterions savoir si, dans la population étudiée, les scores en anglais sont égaux à la moyenne des scores.

Utilisez ensuite le menu Statistiques - Statistiques Élémentaires - Test t pour des échantillons appariés. Sélectionnez ENGG comme variable pour constituer la première liste, et GPA comme variable pour constituer la seconde. Vous devriez obtenir le résultat suivant :

Test t pour des Echantillons Appariés (ADD.sta)								
Différences significatives marquées à p < ,05000								
Variable	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
ENGG	2,6591	0,9455						
GPA	2,4562	0,8614	88	0,2028	0,5188	3,6677	87	0,0004

Enregistrez le classeur ADD.stw sur votre compte puis refermez-le.

## 15 Tests sur les proportions

### 15.1 Test de comparaison de deux proportions sur des groupes indépendants

#### 15.1.1 Variables dichotomiques données par des effectifs

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

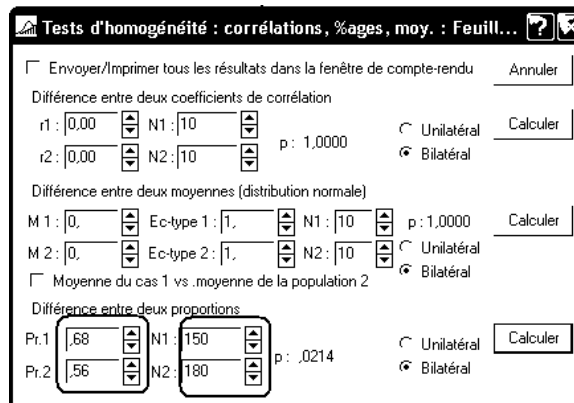
Pour faciliter l'interprétation des résultats, il sera commode de personnaliser le codage de la variable "Resultat" Succès sera codé 1 tandis que Echec sera codé 0.

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

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 peut aussi utiliser le menu Statistiques - Statistiques Élémentaires - 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% ?

### 15.1.2 Variables dichotomiques données par leur protocole

Dans l'exemple traité précédemment, le protocole est un tableau de données à 2 colonnes et 330 lignes. Il est en fait assez facile de constituer à l'aide de Statistica un tableau répétitif de ce type.

- Ouvrez une nouvelle feuille de données de 330 observations, pour 2 variables.
- Pour la première variable, spécifiez le type "texte".
- Pour la seconde variable, spécifiez le type "Numérique", avec, comme valeurs-texte, "Succès" pour la valeur numérique 1, et Echec pour la valeur numérique 0.
- Dans la première variable, saisissez l'identifiant du premier groupe ("A" par exemple) en ligne 1, puis utilisez ensuite le menu Edition - Remplir / Centrer-Réduire le bloc - Remplir/Copier vers le bas pour recopier cette valeur jusqu'en ligne 150.
- Saisissez l'identifiant du second groupe en ligne 151, puis recopiez de même jusqu'en ligne 330
- Dans la seconde variable, saisissez la valeur 1, ou le texte Succès en ligne 1 et recopiez jusqu'en ligne 102.
- De même pour le texte Echec de la ligne 103 à la ligne 150.
- Continuez en saisissant "Succès" de la ligne 151 à la ligne 250, puis "Echec" de la ligne 251 à la ligne 330.

Réalisez ensuite un test de comparaison de moyennes en utilisant ces deux variables. Vous devriez retrouver les résultats précédents.