

7

ÉCHANTILLONNAGE ET DISTRIBUTIONS D'ÉCHANTILLONNAGE

7.1	Le problème d'échantillonnage de la société Electronics Associates	386
7.2	Sélectionner un échantillon	387
7.3	Estimation ponctuelle	394
7.4	Introduction aux distributions d'échantillonnage	399
7.5	Distribution d'échantillonnage de \bar{x}	402
7.6	Distribution d'échantillonnage de \bar{p}	415
7.7	Autres méthodes d'échantillonnage	422

STATISTIQUES APPLIQUÉES

*La société MeadWestvaco** *Stamford, Connecticut*

La société MeadWestvaco, producteur majeur de papiers d'emballage, de papiers spéciaux, de produits pour professionnels et particuliers, emploie plus de 17 000 personnes. Elle est présente dans 30 pays à travers le monde et approvisionne des clients situés dans près de 100 pays. Les experts de l'entreprise utilisent des échantillons pour fournir une variété d'informations permettant à la société d'obtenir des gains de productivité significatifs et de rester compétitive.

Par exemple, MeadWestvaco possède une grande plantation forestière d'où proviennent les arbres qui constituent la matière première de nombreux produits fabriqués par l'entreprise. Les responsables ont besoin d'informations fiables et précises sur les régions d'abattage et les forêts, afin d'évaluer les capacités de l'entreprise à satisfaire ses besoins futurs en matière première. Quel est le volume actuel de bois dans les forêts ? Quelle était la croissance des forêts par le passé ? Quelles sont les prévisions de croissance des forêts ? Grâce aux réponses à ces questions, les responsables de la société MeadWestvaco peuvent développer les projets futurs, y compris le planning à long terme de plantation et d'abattage des arbres.

Comment MeadWestvaco obtient-elle les informations qu'elle souhaite sur ses réserves forestières ? Les données collectées à partir d'échantillons de parcelles, réparties à travers l'ensemble des propriétés de la société, sont à l'origine des informations sur la population des arbres que possède l'entreprise. Pour identifier les parcelles d'un échantillon, les propriétés forestières sont réparties en trois sections, selon leur situation géographique et le type d'arbres qu'elles contiennent. Sur la base de cartes et de nombres aléatoires, les statisticiens de la société identifient des échantillons aléatoires de parcelles de 1/5 à 1/7 acre (demi-hectare) dans chaque section de la forêt. Les gardes forestiers de la société collectent ensuite les données souhaitées dans ces échantillons de parcelles, à partir desquels sont obtenues les informations sur la population forestière entière.

Les gardes forestiers participent au processus de collecte des données sur le terrain. Périodiquement, des équipes de deux personnes rassemblent des informations sur chaque arbre de chaque échantillon de parcelles. Les données sont enregistrées dans le système informatique de gestion des forêts. Les rapports faits à partir de ce système informatique contiennent des résumés sous forme de distributions de fréquence, regroupant des statistiques sur les types d'arbre, le volume forestier actuel, les taux de croissance passés de la forêt, et les prévisions concernant la croissance et le volume forestier dans le futur. L'échantillonnage et les résumés statistiques des données fournissent les informations nécessaires à la gestion du parc forestier de la société MeadWestvaco.

Dans ce chapitre, vous vous familiariserez avec l'échantillonnage aléatoire simple et le processus de sélection d'un échantillon. De plus, vous apprendrez comment des statistiques comme la moyenne ou la proportion d'échantillon peuvent être utilisées pour estimer la moyenne ou une proportion de la population. Le concept de distribution d'échantillonnage est également introduit.

* Les auteurs remercient Dr. Edward P. Winkofsky de leur avoir fourni ce Statistiques appliquées.

Dans le chapitre 1, nous avons défini ce que sont un *élément*, une *population* et un *échantillon* :

- Un élément est une entité pour laquelle des données sont collectées.
- Une population est l'ensemble de tous les éléments auxquels on s'intéresse.
- Un échantillon est un sous-ensemble de la population.

La constitution d'un échantillon permet de collecter des données pour répondre à une question concernant une population.

Citons deux exemples dans lesquels un échantillon est utilisé pour répondre à une question concernant une population.

1. Les membres d'un parti politique au Texas sont supposés soutenir un candidat particulier aux élections du Sénat américain, et les leaders du parti voudraient estimer la proportion d'électeurs favorables à leur candidat. Un échantillon de 400 électeurs texans a été sélectionné et 160 de ces 400 électeurs ont indiqué être favorables au candidat. Une estimation de la proportion d'électeurs favorables au candidat est donc $160 / 400 = 0,40$.
2. Un fabricant de pneus a conçu un nouveau type de pneu permettant d'accroître le kilométrage effectué, comparativement au nombre de kilomètres effectués avec les pneus actuellement fabriqués par l'entreprise. Pour estimer le nombre moyen de kilomètres effectués avec les nouveaux pneus, le fabricant a sélectionné un échantillon de 120 nouveaux pneus, dans le but de les tester. D'après les résultats du test, la moyenne de l'échantillon est égale à 36 500 kilomètres. Par conséquent, une estimation du kilométrage moyen pour la population des nouveaux pneus est de 36 500 kilomètres.

Il est important de comprendre que les résultats d'un échantillon fournissent seulement des *estimations* de la valeur des caractéristiques de la population considérée. On ne s'attend pas à ce qu'exactly 40 % de la population des électeurs soit favorable au candidat considéré ; de même, on ne s'attend pas à ce que la moyenne d'échantillon de 36 500 kilomètres soit exactement égale au kilométrage moyen de tous les pneus de la population. Ceci tient au fait que l'échantillon ne contient qu'une partie de la population. Une certaine erreur d'échantillonnage est attendue. Avec des méthodes d'échantillonnage adéquates, les résultats de l'échantillon fournissent toutefois de « bonnes » estimations des paramètres de la population. Mais quelle justesse des résultats peut-on espérer ? Des procédures statistiques permettent de répondre à cette question.

Une moyenne d'échantillon fournit une estimation de la moyenne de la population et une proportion d'échantillon fournit une estimation de la proportion de la population. Avec de telles estimations, on doit s'attendre à des erreurs d'estimation. Ce chapitre fournit les bases pour déterminer l'importance de l'erreur d'estimation.

Définissons certains termes utilisés en échantillonnage. La **population échantillonnée** est la population à partir de laquelle l'échantillon est sélectionné et le **cadre d'analyse** est la liste des éléments d'où l'échantillon est issu. Dans le premier exemple, la population échantillonnée est l'ensemble des électeurs du Texas et le cadre d'analyse est

une liste de tous les électeurs. Puisque le nombre d'électeurs au Texas est fini, le premier exemple est un exemple d'échantillonnage à partir d'une population finie. Dans la section 7.2 nous discuterons de la manière de sélectionner un échantillon aléatoire simple lorsque l'échantillonnage se fait à partir d'une population finie.

La population échantillonnée dans l'exemple du fabricant de pneus est plus difficile à définir parce que l'échantillon de 120 pneus est obtenu à partir d'un processus de production à un moment particulier dans le temps. Nous pouvons penser à la population échantillonnée comme à la population conceptuelle de tous les pneus qui auraient pu être produits à partir de ce processus de production à ce moment particulier dans le temps. En ce sens, la population échantillonnée est considérée comme infinie, rendant impossible l'énumération des éléments de la population. Dans la section 7.2 nous discuterons de la manière de sélectionner un échantillon aléatoire dans une telle situation.

Dans ce chapitre, nous verrons comment sélectionner un échantillon à partir d'une population finie grâce à la méthode d'échantillonnage aléatoire simple et comment un échantillon aléatoire peut être issu d'une population infinie générée par un processus. Nous verrons ensuite comment utiliser les données obtenues à partir de l'échantillon pour estimer la moyenne, l'écart type ou une proportion de la population. De plus, nous introduirons le concept de distribution d'échantillonnage. Comme nous le montrerons, la connaissance de la distribution d'échantillonnage appropriée est ce qui nous permet de conclure quant à la justesse des résultats de l'échantillon. La dernière section traite des méthodes d'échantillonnage aléatoire alternatives à l'échantillonnage aléatoire simple, qui sont souvent employées dans la pratique.

7.1 LE PROBLÈME D'ÉCHANTILLONNAGE DE LA SOCIÉTÉ ELECTRONICS ASSOCIATES

Le directeur du personnel de la société Electronics Associates (EAI) a été chargé d'identifier le profil des 2 500 employés de la société. Les caractéristiques pertinentes à identifier comprennent le salaire annuel moyen des employés et la proportion d'employés ayant suivi le programme de formation au management, mis en place par la société.



En considérant les 2 500 employés comme la population de cette étude, on peut déterminer le salaire annuel de chaque individu et savoir s'il a suivi le programme de formation au management, en consultant les dossiers du personnel de l'entreprise. Vous trouverez la base de données contenant ces informations pour l'ensemble de la population dans le fichier en ligne intitulé EAI.

En utilisant l'ensemble de données EAI et les formules présentées au chapitre 3, nous pouvons calculer la moyenne et l'écart type du salaire annuel pour la population.

Moyenne de la population : $\mu = 51\,800$ dollars

Écart type de la population : $\sigma = 4\,000$ dollars

Les données concernant le programme de formation montrent que 1 500 des 2 500 employés l'ont effectivement suivi.

Les caractéristiques numériques d'une population sont appelées **paramètres**. Soit p la proportion de la population ayant suivi le programme de formation. Nous avons donc : $p = 1500 / 2500 = 0,60$. Le salaire annuel moyen de la population (dollars), l'écart type du salaire annuel de la population ($\sigma = 4\,000$ dollars) et la proportion de la population ayant suivi le programme de formation ($p = 0,60$) sont des paramètres de la population des employés de la société EAI.

Maintenant, supposez que les informations nécessaires sur les employés de la société EAI ne sont pas disponibles dans les bases de données de la société. La question qui se pose maintenant, est de savoir comment le directeur du personnel de la société peut obtenir des estimations des paramètres de la population, en utilisant un échantillon d'employés à la place de la population constituée de 2 500 employés. Supposez que l'on utilise un échantillon de 30 employés. Clairement, le temps et le coût nécessaire pour établir le profil de 30 employés sont moindres que ceux nécessaires pour établir le profil de l'ensemble de la population des employés de l'entreprise. Si le directeur du personnel est sûr qu'un échantillon de 30 employés fournira des informations correctes sur la population des 2 500 employés, travailler avec un échantillon, plutôt qu'avec la population entière, est préférable. Explorons la possibilité d'utiliser un échantillon pour l'étude de la société EAI en commençant par identifier un échantillon de 30 employés.

Souvent le coût de la collecte d'informations à partir d'un échantillon est largement inférieur à celui généré par la collecte d'informations à partir de la population entière, en particulier lorsque l'obtention de ces informations nécessitent des entretiens avec le personnel.

7.2 SÉLECTIONNER UN ÉCHANTILLON

Dans cette section, nous décrivons comment sélectionner un échantillon. Nous considérons tout d'abord comment sélectionner un échantillon à partir d'une population finie et décrivons ensuite comment sélectionner un échantillon à partir d'une population infinie.

7.2.1 Échantillonnage à partir d'une population finie

Les statisticiens recommandent de sélectionner un échantillon probabiliste lorsque l'on sélectionne un échantillon à partir d'une population finie parce qu'un échantillon probabiliste permet de faire de l'inférence statistique sur la population. Le type le plus simple d'échantillons probabilistes est celui dans lequel chaque échantillon de taille n a la même probabilité d'être sélectionné. On parle d'échantillon aléatoire simple. Un échantillon aléatoire simple de taille n , issu d'une population finie de taille N , est défini de la manière suivante.

D'autres méthodes d'échantillonnage probabilistes sont décrites dans la section 7.7.

- ▶ **Échantillon aléatoire simple (population finie)**
Un **échantillon aléatoire simple** de taille n , issu d'une population finie de taille N , est un échantillon sélectionné de manière à ce que chaque échantillon possible de taille n ait la même probabilité d'être sélectionné.

Une procédure de sélection d'un échantillon aléatoire simple, à partir d'une population finie, consiste à choisir les éléments de l'échantillon un par un, de façon à ce que les éléments restants dans la population aient la même probabilité d'être sélectionnés. Choisir n éléments de cette façon respecte la définition d'un échantillon aléatoire simple issu d'une population finie.

Nous décrivons comment utiliser Excel, Minitab et StatTools pour générer un échantillon aléatoire simple dans les annexes de ce chapitre.

Pour constituer un échantillon aléatoire simple à partir de la population finie des employés de la société EAI, nous assignons tout d'abord un numéro à chaque employé. Par exemple, on peut numéroter les employés de 1 à 2 500, en fonction de leur ordre d'apparition dans les fichiers du personnel de la société EAI. Ensuite, nous nous référons à la table des nombres aléatoires reproduite dans le tableau 7.1. Chaque chiffre de la première ligne, 6, 3, 2, ..., correspond à un chiffre aléatoire qui a une probabilité égale de survenir.

Tableau 7.1 Nombres aléatoires

63 271	59 986	71 744	51 102	15 141	80 714	58 683	93 108	13 554	79 945
88 547	09 896	95 436	79 115	08 303	01 041	20 030	63 754	08 459	28 364
55 957	57 243	83 865	09 911	19 761	66 535	40 102	26 646	60 147	15 702
46 276	87 453	44 790	67 122	45 573	84 358	21 625	16 999	13 385	22 782
55 363	07 449	34 835	15 290	76 616	67 191	12 777	21 861	68 689	03 263
69 393	92 785	49 902	58 447	42 048	30 378	87 618	26 933	40 640	16 281
13 186	29 431	88 190	04 588	38 733	81 290	89 541	70 290	40 113	08 243
17 726	28 652	56 836	78 351	47 327	18 518	92 222	55 201	27 340	10 493
36 520	64 465	05 550	30 157	82 242	29 520	69 753	72 602	23 756	54 935
81 628	36 100	39 254	56 835	37 636	02 421	98 063	89 641	64 953	99 337
84 649	48 968	75 215	75 498	49 539	74 240	03 466	49 292	36 401	45 525
63 291	11 618	12 613	75 055	43 915	26 488	41 116	64 531	56 827	30 825
70 502	53 225	03 655	05 915	37 140	57 051	48 393	91 322	25 653	06 543
06 426	24 771	59 935	49 801	11 082	66 762	94 477	02 494	88 215	27 191
20 711	55 609	29 430	70 165	45 406	78 484	31 639	52 009	18 873	96 927
41 990	70 538	77 191	25 860	55 204	73 417	83 920	69 468	74 972	38 712
72 452	36 618	76 298	26 678	89 334	33 938	95 567	29 380	75 906	91 807
37 042	40 318	57 099	10 528	09 925	89 773	41 335	96 244	29 002	46 453
53 766	52 875	15 987	46 962	67 342	77 592	57 651	95 508	80 033	69 828
90 585	58 955	53 122	16 025	84 299	53 310	67 380	84 249	25 348	04 332
32 001	96 293	37 203	64 516	51 530	37 069	40 261	61 374	05 815	06 714
62 606	64 324	46 354	72 157	67 248	20 135	49 804	09 226	64 419	29 457
10 078	28 073	85 389	50 324	14 500	15 562	64 165	06 125	71 353	77 669
91 561	46 145	24 177	15 294	10 061	98 124	75 732	00 815	83 452	97 355
13 091	98 112	53 959	79 607	52 244	63 303	10 413	63 839	74 762	50 289

Puisque le nombre le plus grand dans la population des employés de la société EAI, 2 500, a quatre chiffres, nous sélectionnons les nombres aléatoires de la table, formés de quatre chiffres. Bien que nous puissions débiter la sélection de nombres aléatoires n'importe où dans la table et nous déplacer dans n'importe quelle direction, nous utilisons la première ligne du tableau 7.1 et nous nous déplaçons de gauche à droite. Les sept premiers nombres aléatoires à quatre chiffres sont :

6 327 1 599 8 671 7 445 1 102 1 514 1 807

Puisque les nombres de la table sont aléatoires, ces nombres à quatre chiffres sont équiprobables.

Dans la table, les nombres aléatoires sont regroupés par groupe de cinq chiffres pour des raisons de commodité de lecture.

Nous pouvons maintenant utiliser ces nombres aléatoires à quatre chiffres pour donner à chaque employé de la population une probabilité identique d'être inclus dans l'échantillon aléatoire. Le premier nombre, 6 327, est supérieur à 2 500. Il n'est associé à aucun des employés numérotés dans la population ; par conséquent, il est écarté. Le second nombre, 1 599, est compris entre 1 et 2 500. Ainsi, le premier employé sélectionné dans l'échantillon aléatoire est celui qui porte le numéro 1 599 dans la liste des employés de la société. En poursuivant ce procédé, nous ignorons les nombres 8 671 et 7 445 avant d'inclure dans l'échantillon aléatoire les employés numérotés 1 102, 1 514 et 1 807. On poursuit ce procédé jusqu'à ce que 30 employés aient été sélectionnés.

En procédant à la sélection de cet échantillon aléatoire simple, il est possible qu'un nombre aléatoire déjà sélectionné réapparaisse dans la table, avant d'avoir constitué l'échantillon des 30 employés. Dans la mesure où nous ne voulons pas sélectionner un individu plus d'une fois, tous les nombres aléatoires déjà sélectionnés sont ignorés, puisque l'employé associé à ce nombre fait déjà partie de l'échantillon. Cette manière de sélectionner un échantillon correspond à une **procédure d'échantillonnage sans remise**. Si nous avons constitué l'échantillon en acceptant les nombres aléatoires déjà choisis et donc en incluant dans l'échantillon les individus plus d'une fois, nous aurions alors utilisé une **procédure d'échantillonnage avec remise**. L'échantillonnage avec remise est une façon correcte de constituer un échantillon aléatoire simple. Cependant, **l'échantillonnage sans remise est la procédure d'échantillonnage la plus utilisée**. Lorsque l'on se réfère à un échantillonnage aléatoire simple, il est sous-entendu que l'échantillonnage est sans remise.

7.2.2 Échantillonnage à partir d'une population infinie

Parfois, nous souhaitons sélectionner un échantillon à partir d'une population qui est infiniment grande ou dont les éléments sont générés par un processus pour lequel il n'y a pas de limite quant au nombre d'éléments qui peuvent être générés. Ainsi, il n'est pas possible de développer une liste de tous les éléments de cette population. C'est ce qu'on appelle le cas d'une **population infinie**. Dans un tel cas, **on ne peut pas sélectionner un échantillon aléatoire simple car on ne peut pas définir un cadre d'analyse contenant tous les éléments**.

Dans le cas d'une population infinie, les statisticiens recommandent de sélectionner ce qui est appelé un échantillon aléatoire.

► **Échantillon aléatoire (population infinie)**

Un **échantillon aléatoire** de taille n issu d'une population infinie est un échantillon sélectionné qui satisfait les conditions suivantes.

1. Chaque élément sélectionné est issu de la même population.
2. Chaque élément est sélectionné indépendamment des autres.

Précaution et bon sens doivent guider le processus de sélection d'un échantillon aléatoire à partir d'une population infinie. Chaque cas peut nécessiter une procédure de sélection différente. Considérons deux exemples pour illustrer les conditions (1) « chaque élément sélectionné est issu de la même population » et (2) « chaque élément est sélectionné indépendamment des autres ».

Une application courante en matière de **contrôle de la qualité** implique un processus de production dans lequel il n'y a pas de limite quant au nombre d'éléments qui peuvent être produits. La population conceptuelle d'où est issu l'échantillon, correspond à tous les éléments qui peuvent être produits (pas simplement ceux qui ont déjà été produits). Puisque nous ne pouvons pas constituer une liste de tous les éléments qui peuvent être produits, la population est considérée être infinie. Pour être plus précis, considérons **une chaîne de production conçue pour remplir des boîtes de céréale d'un poids moyen de 24 onces.** Des échantillons de 12 boîtes remplies via ce processus sont périodiquement sélectionnés par un inspecteur de la qualité pour déterminer si le processus fonctionne correctement ou si, par exemple, un dysfonctionnement a entraîné un sur- ou un sous-remplissage des boîtes.

Avec une opération de production de ce type, la principale difficulté dans la sélection d'un échantillon aléatoire est d'être sûr que la condition 1 est satisfaite, c'est-à-dire que **les éléments échantillonnés sont issus de la même population.** Pour s'assurer que cette condition est satisfaite, **les boîtes doivent être sélectionnées à peu près au même moment dans le temps.** De cette façon, l'inspecteur évite de sélectionner certaines boîtes lorsque la chaîne de production fonctionne correctement et d'autres boîtes lorsque le processus n'est plus sous contrôle et que **les boîtes sont sur- ou sous-remplies.** Avec un processus de production de ce type, la seconde condition, chaque élément est sélectionné indépendamment, est satisfaite en définissant le processus de production de façon à ce que chaque boîte de céréale soit remplie indépendamment. Avec cette hypothèse, l'inspecteur de la qualité n'a qu'à se soucier de la première condition.

Considérons un autre exemple de sélection d'un échantillon aléatoire à partir d'une population infinie, à savoir **la population des clients arrivant à un fast-food.** Supposez que l'on ait demandé à un employé de **sélectionner et d'interviewer un échantillon de clients afin de déterminer le profil des clients du restaurant.** Le processus d'arrivée des clients est permanent et il n'y a aucun moyen d'obtenir une liste de tous les clients formant la population. Aussi, pour des raisons pratiques, **la population** pour ce processus est considérée être **infinie.** Tant que la procédure d'échantillonnage est conçue de façon à ce que les éléments de l'échantillon soient les clients du restaurant et qu'ils sont sélectionnés de façon indépendante, un échantillon aléatoire sera obtenu. Dans ce cas, l'employé chargé

de collecter l'échantillon, doit sélectionner l'échantillon à partir des personnes qui entrent dans le restaurant et font un achat pour garantir que la condition de même population soit satisfaite. Si, par exemple, l'employé a sélectionné une personne qui est entrée dans le restaurant juste pour aller aux toilettes, cette personne n'est pas un client et la condition d'une même population est violée. Aussi, tant que l'employé sélectionne l'échantillon à partir des personnes effectuant un achat dans le restaurant, la condition 1 est satisfaite. S'assurer que les clients sont sélectionnés aléatoirement peut s'avérer plus difficile.

L'objectif de la seconde condition de la procédure de sélection d'un échantillon aléatoire (chaque élément est sélectionné indépendamment des autres) est d'éviter un biais de sélection. Dans ce cas, un biais de sélection survient si l'employé est libre de sélectionner les clients composant l'échantillon de façon arbitraire. L'employé pourrait se sentir plus à l'aise en sélectionnant des clients d'une tranche d'âge particulière et pourrait éviter de sélectionner les clients appartenant à d'autres tranches d'âge. Un biais de sélection surviendrait si l'employé sélectionnait un groupe de cinq clients qui entreraient ensemble dans le restaurant et leur demandait à tous de participer à l'enquête. Un tel groupe de clients auraient vraisemblablement des caractéristiques similaires, qui pourraient fournir des informations erronées sur la population des clients. Un biais de sélection de ce type peut être évité en s'assurant que la sélection d'un client particulier n'influence pas la sélection d'un autre client. En d'autres termes, les éléments (clients) sont sélectionnés indépendamment les uns des autres.

McDonald's, le leader de la restauration rapide, a mis en place une procédure d'échantillonnage aléatoire pour cette situation. La procédure d'échantillonnage était basée sur le fait que certains clients présentent des bons de réduction. Lorsqu'un client présentait un bon de réduction, on demandait au client suivant de remplir un questionnaire sur son profil. Puisque les clients présentant des bons de réduction arrivaient de façon aléatoire et indépendante des autres clients, cette procédure d'échantillonnage garantissait que les clients étaient sélectionnés indépendamment les uns des autres. En conséquence, l'échantillon satisfaisait les conditions d'un échantillon aléatoire issu d'une population infinie.

Des situations impliquant un échantillonnage à partir d'une population infinie, sont généralement associées à un processus durable. On peut citer à titre d'exemples les pièces fabriquées sur une chaîne de production, les essais expérimentaux répétés dans un laboratoire, les transactions bancaires, les appels téléphoniques reçus dans un centre de soutien technique, et les clients entrant dans un magasin. Dans chaque cas, la situation peut être vue comme un processus qui génère des éléments à partir d'une population infinie. Tant que les éléments échantillonnés sont sélectionnés à partir d'une même population et de façon indépendante, l'échantillon est considéré être un échantillon aléatoire provenant d'une population infinie.

REMARQUES

1. Dans cette section, nous avons défini avec précaution deux types d'échantillon : un échantillon aléatoire simple issu d'une population finie et un échantillon aléatoire issu d'une population infinie. Dans le reste de l'ouvrage, nous nous référerons généralement à ces deux types d'échantillons en parlant d'un échantillon aléatoire ou

simplement d'un échantillon. Nous ne distinguerons pas les échantillons aléatoires « simples » à moins que ce ne soit nécessaire pour l'exercice ou la discussion.

- Les statisticiens spécialisés dans les enquêtes d'échantillonnage à partir de populations finies, utilisent les méthodes d'échantillonnage qui fournissent des échantillons probabilistes. L'échantillonnage aléatoire simple est une de ces méthodes. Dans la section 7.7, nous décrirons d'autres méthodes d'échantillonnage probabilistes : l'échantillonnage aléatoire stratifié, l'échantillonnage par grappes et l'échantillonnage systématique. Nous utilisons le terme simple dans l'expression échantillonnage aléatoire simple pour indiquer qu'il s'agit d'une méthode d'échantillonnage probabiliste qui assure que chaque échantillon de taille n a la même probabilité d'être sélectionné.
- Le nombre d'échantillons aléatoires simples différents de taille n qui peuvent être sélectionnés à partir d'une population de taille N est

$$\frac{N!}{n!(N-n)!}$$

- Dans cette formule, $N!$ et $n!$ sont les factorielles dont nous avons parlé au chapitre 4. Pour le problème de la société EAI, avec $N = 2\,500$ et $n = 30$, selon cette expression, approximativement $2,75 \times 10^{69}$ échantillons aléatoires simples différents de 30 employés de la société EAI peuvent être constitués.

EXERCICES

Méthode



- Considérer une population finie composée de cinq éléments notés A, B, C, D et E. Dix échantillons aléatoires simples de taille égale à deux peuvent être sélectionnés.
 - Énumérer les dix échantillons en commençant par AB, AC, etc.
 - En utilisant la procédure d'échantillonnage aléatoire simple, quelle est la probabilité pour chaque échantillon de taille deux d'être sélectionné ?
 - Supposez que le nombre aléatoire 1 corresponde à A, le nombre aléatoire 2 corresponde à B, etc. Définir l'échantillon aléatoire de taille deux qui sera sélectionné en utilisant les chiffres 8 0 5 7 5 3 2.
- Supposez qu'une population finie soit composée de 350 éléments. En utilisant les trois derniers chiffres de chacun des nombres aléatoires suivants à cinq chiffres (601, 022, 448, ...), déterminer les quatre premiers éléments qui seront sélectionnés pour constituer l'échantillon aléatoire simple.

98601 73022 83448 02147 34229 27553 84147 93289 14209

Applications



- Fortune* publie des données sur les ventes, les profits, le capital, les capitaux des actionnaires, la valeur marchande et les bénéfices par action des 500 plus importantes sociétés industrielles

américaines (*Fortune 500*, 2006). Supposez que vous vouliez constituer un échantillon aléatoire simple de 10 sociétés parmi la liste des 500 sociétés établie par *Fortune*. Utilisez les trois derniers chiffres de la colonne 9 du tableau 7.1, en commençant par 554. Lire les chiffres en descendant dans la colonne et identifier les numéros des 10 sociétés qui seront sélectionnées.

4. L'association américaine de golf s'interroge sur l'opportunité d'interdire les clubs de golf longs et bombés. Cela a généré des débats parmi les golfeurs amateurs mais également les membres de l'Association professionnelle de golf (PGA) (*Golfweek*, 26 octobre 2012). Ci-dessous figurent les noms des 10 finalistes d'un tournoi récent de golf professionnel, le PGA Tour Mc Gladrey Classic.

- | | |
|---------------------|-----------------------|
| 1. Tommy Gainey | 6. David Love III |
| 2. David Toms | 7. Chad Campbell |
| 3. Jim Furyk | 8. Greg Owens |
| 4. Brendon de Jonge | 9. Charles Howell III |
| 5. D.J. Trahan | 10. Arjun Atwal |

- a) Sélectionnez un échantillon aléatoire simple de trois de ces joueurs pour connaître leur opinion concernant l'usage des clubs de golf longs et bombés. Utilisez les nombres aléatoires de la colonne 2 du tableau 7.1 pour effectuer votre sélection. Commencez avec 59986 et utiliser le dernier chiffre, 6, pour le premier joueur sélectionné (David Love III). Continuez en descendant dans la colonne pour sélectionner deux autres joueurs.
- b) Selon l'information contenue dans la remarque 3, combien d'échantillons aléatoires simples différents de taille 3 peuvent être constitués dans la liste des dix joueurs ?
5. Une organisation gouvernementale étudiante s'intéresse à l'estimation de la proportion des étudiants partisans de la politique d'évaluation « succès-échec » pour les cours facultatifs. Une liste des noms et adresses de 645 étudiants inscrits au cours du trimestre est disponible auprès du bureau des inscriptions. En utilisant les nombres aléatoires à trois chiffres de la ligne 10 du tableau 7.1 et en lisant de gauche à droite, identifiez les 10 premiers étudiants qui seront sélectionnés en utilisant la procédure d'échantillonnage aléatoire simple. Les nombres aléatoires à trois chiffres commencent par 816, 283 et 610.
6. Le *County and City Data Book*, publié par le bureau des recensements, fournit des informations sur 3 139 comtés américains. Supposez qu'une étude nationale collecte des données sur 30 comtés sélectionnés aléatoirement. Utilisez les nombres aléatoires à quatre chiffres à partir de la dernière colonne du tableau 7.1 pour identifier les nombres correspondant aux cinq premiers comtés sélectionnés pour constituer l'échantillon. Ignorez les premiers chiffres et commencez par les nombres aléatoires à quatre chiffres 9945, 8364, 5702, etc.
7. Supposez que nous voulions identifier un échantillon aléatoire simple de 12 des 372 médecins exerçant dans une ville particulière. Les noms des médecins sont disponibles auprès d'une organisation médicale locale. Utilisez la huitième colonne de nombres aléatoires à cinq chiffres du tableau 7.1 pour identifier les 12 médecins de l'échantillon. Ignorez les deux premiers chiffres aléatoires dans chaque ensemble de nombres aléatoires à cinq chiffres. Ce processus commence avec le nombre aléatoire 108 et se poursuit en descendant dans la colonne des nombres aléatoires.

8. Les actions suivantes composent l'indice Dow Jones Industriel (*Barron's*, 30 juillet 2012).

1. 3M	11. Disney	21. McDonald's
2. AT&T	12. DuPont	22. Merck
3. Alcoa	13. ExxonMobil	23. Microsoft
4. American Express	14. General Electric	24. J.P. Morgan
5. Bank of America	15. Hewlett-Packard	25. Pfizer
6. Boeing	16. Home Depot	26. Procter & Gamble
7. Caterpillar	17. IBM	27. Travelers
8. Chevron	18. Intel	28. United Technologies
9. Cisco Systems	19. Johnson & Johnson	29. Verizon
10. Coca-Cola	20. Kraft Foods	30. Wal-Mart

Supposez que vous vouliez sélectionner un échantillon de six de ces sociétés pour mener une étude approfondie sur les pratiques managériales. Utiliser les deux premiers chiffres de chaque ligne de la 9^e colonne du tableau 7.1 pour sélectionner un échantillon aléatoire simple de six sociétés.

9. L'indice Forbes 400 est un classement des 400 personnes les plus riches aux États-Unis (site Internet Forbes, 4 mars 2013). Supposez que vous vouliez sélectionner un échantillon aléatoire simple de 10 personnes parmi ces 400 pour effectuer une étude sur leur niveau d'études. Utilisez la quatrième colonne des nombres aléatoires du tableau 7.1, en commençant par 51102, pour sélectionner l'échantillon aléatoire simple de dix personnes. Commencez avec le numéro 102 et utilisez les trois derniers chiffres dans chaque ligne de la quatrième colonne pour effectuer votre sélection. Quels sont les numéros des 10 personnes sélectionnées dans l'échantillon ?
10. Indiquer lesquelles des situations suivantes impliquent un échantillonnage à partir d'une population finie et lesquelles impliquent un échantillonnage à partir d'une population infinie. Dans les cas où la population échantillonnée est finie, décrire la procédure d'échantillonnage.
- Obtenir un échantillon des conducteurs de l'État de New York.
 - Obtenir un échantillon des boîtes de céréale produites par la société Breakfast Choice.
 - Obtenir un échantillon des voitures passant sur le pont Golden Gate un jour de semaine ordinaire.
 - Obtenir un échantillon des étudiants en statistiques de l'Université d'Indiana.
 - Obtenir un échantillon des commandes gérées par une entreprise de vente par correspondance.

7.3 ESTIMATION PONCTUELLE

Maintenant que nous avons décrit comment constituer un échantillon aléatoire simple, revenons au problème de la société EAI. Supposez qu'un échantillon aléatoire simple de 30 employés ait été constitué et que les données correspondantes sur le salaire annuel et la participation au programme de formation au management soient celles présentées dans

le tableau 7.2. La notation x_1, x_2, \dots , est utilisée pour noter le salaire annuel du premier employé de l'échantillon, le salaire annuel du deuxième employé, etc. La participation au programme de formation est indiquée par un « oui » dans la colonne « programme de formation au management ».

Pour estimer la valeur d'un paramètre de la population, nous calculons la valeur d'une caractéristique correspondante de l'échantillon, dite **statistique d'échantillon**. Par exemple, pour estimer la moyenne μ et l'écart type σ du salaire annuel de la population des employés de la société EAI, nous utilisons les données du tableau 7.2 pour calculer les statistiques d'échantillon correspondantes : la moyenne de l'échantillon \bar{x} et l'écart type de l'échantillon s . En utilisant les formules présentées dans le chapitre 3, la moyenne de l'échantillon est égale à

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1\,554\,420}{30} = 51\,814 \text{ dollars}$$

et l'écart type de l'échantillon à

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{325\,009\,260}{29}} = 3\,348 \text{ dollars}$$

Tableau 7.2 Salaire annuel et participation au programme de formation pour un échantillon aléatoire simple de 30 employés de la société EAI

Salaire annuel (\$)	Programme de formation au management	Salaire annuel (\$)	Programme de formation au management
$x_1 = 49\,094,30$	Oui	$x_{16} = 51\,766,00$	Oui
$x_2 = 53\,263,90$	Oui	$x_{17} = 52\,541,30$	Non
$x_3 = 49\,643,50$	Oui	$x_{18} = 44\,980,00$	Oui
$x_4 = 49\,894,90$	Oui	$x_{19} = 51\,932,60$	Oui
$x_5 = 47\,621,60$	Non	$x_{20} = 52\,973,00$	Oui
$x_6 = 55\,924,00$	Oui	$x_{21} = 45\,120,90$	Oui
$x_7 = 49\,092,30$	Oui	$x_{22} = 51\,753,00$	Oui
$x_8 = 51\,404,40$	Oui	$x_{23} = 54\,391,80$	Non
$x_9 = 50\,957,70$	Oui	$x_{24} = 50\,164,20$	Non
$x_{10} = 55\,109,70$	Oui	$x_{25} = 52\,973,60$	Non
$x_{11} = 45\,922,60$	Oui	$x_{26} = 50\,241,30$	Non
$x_{12} = 57\,268,40$	Non	$x_{27} = 52\,793,90$	Non
$x_{13} = 55\,688,80$	Oui	$x_{28} = 50\,979,40$	Oui
$x_{14} = 51\,564,70$	Non	$x_{29} = 55\,860,90$	Oui
$x_{15} = 56\,188,20$	Non	$x_{30} = 57\,309,10$	Non

Pour estimer p , la proportion des employés de la population qui ont suivi le programme de formation au management, nous utilisons la proportion de l'échantillon \bar{p} . Soit x le nombre d'employés dans l'échantillon qui ont suivi le programme de formation au management. Les données du tableau 7.2 indiquent que $x = 19$. Ainsi, avec un échantillon de taille $n = 30$, la proportion d'échantillon est égale à

$$\bar{p} = \frac{x}{n} = \frac{19}{30} = 0,63$$

En faisant les calculs précédents, nous avons procédé à une *estimation ponctuelle*. En utilisant la terminologie de l'estimation ponctuelle, la moyenne d'échantillon \bar{x} correspond à l'**estimateur ponctuel** de la moyenne de la population μ , l'écart type d'échantillon s à l'**estimateur ponctuel** de l'écart type de la population σ et la proportion d'échantillon \bar{p} à l'**estimateur ponctuel** de la proportion de la population p . La valeur numérique obtenue pour \bar{x} , s ou \bar{p} est appelée **estimation ponctuelle**. Ainsi, pour l'échantillon aléatoire simple des 30 employés de la société EAI, présenté dans le tableau 7.2, 51 814 dollars est l'estimation ponctuelle de μ , 3 348 dollars est l'estimation ponctuelle de σ et 0,63 est l'estimation ponctuelle de p . Le tableau 7.3 résume les résultats d'échantillon et compare les estimations ponctuelles aux valeurs effectives des paramètres de la population.

Comme le montre le tableau 7.3, les estimations ponctuelles diffèrent quelque peu de la valeur du paramètre de la population qui lui est associé. Cet écart est prévisible puisque seul un échantillon et non un recensement de la population entière est utilisé pour effectuer les estimations ponctuelles. Dans le prochain chapitre, nous verrons comment obtenir des informations sur l'écart entre l'estimation ponctuelle et le paramètre de la population.

7.3.1 Conseil pratique

Le principal sujet traité dans le reste de l'ouvrage concerne l'inférence statistique. L'estimation ponctuelle est une forme d'inférence statistique. Nous utilisons une statistique d'échantillon pour faire de l'inférence à propos d'un paramètre d'une population. Lorsque l'on fait de l'inférence sur une population en se basant sur un échantillon, il est important d'avoir des liens forts entre la population échantillonnée et la population cible. La **population cible** est la population sur laquelle vous voulez faire de l'inférence, alors que la population

Tableau 7.3 Résumé des estimations ponctuelles obtenues à partir d'un échantillon aléatoire simple de 30 employés de la société EAI

Paramètre de la population	Valeur du paramètre	Estimateur ponctuel	Estimation ponctuelle
μ = Salaire annuel moyen de la population	51 800 \$	\bar{x} = Moyenne d'échantillon du salaire annuel	51 814 \$
σ = Écart type du salaire annuel de la population	4 000 \$	s = Écart type d'échantillon du salaire annuel	3 348 \$
p = Proportion de la population ayant suivi le programme de formation au management	0,60	\bar{p} = Proportion des employés de l'échantillon ayant suivi le programme de formation au management	0,63

échantillonnée est la population à partir de laquelle l'échantillon est sélectionné. Dans cette partie, nous avons décrit le processus de sélection d'un échantillon aléatoire simple à partir de la population des employés de la société EAI et réalisé des estimations ponctuelles des caractéristiques de cette même population. Aussi, la population échantillonnée et la population cible sont identiques, ce qui est la situation idéale. Mais dans d'autres cas, un soin particulier doit être pris pour faire correspondre population échantillonnée et population cible.

Considérez le cas d'un parc d'attraction sélectionnant un échantillon de ses clients pour déterminer leurs caractéristiques telles que l'âge et le temps passé dans le parc. Supposez que tous les éléments d'échantillon aient été sélectionnés un jour où l'entrée au parc était réservée aux employés d'une grande entreprise. Ainsi la population échantillonnée sera composée des employés de cette entreprise et des membres de leurs familles. Si la population cible sur laquelle on souhaite faire de l'inférence est la population des clients ordinaires du parc au cours d'un été ordinaire, alors on peut faire face à une différence significative entre la population échantillonnée et la population cible. Dans un tel cas, on peut douter de la validité des estimations ponctuelles faites. Les responsables du parc devraient être en mesure de déterminer si un échantillon constitué un jour donné est représentatif ou non de la population cible.

En résumé, lorsqu'un échantillon est utilisé pour faire de l'inférence sur une population, nous devons être sûrs que l'étude est menée de façon à ce que la population échantillonnée et la population cible soient proches. La question n'est pas mathématique mais exige du bon sens.

EXERCICES

Méthode

11. Les données suivantes sont issues d'un échantillon aléatoire simple.

5 8 10 7 10 14

- a) Quelle est l'estimation ponctuelle de la moyenne de la population ?
- b) Quelle est l'estimation ponctuelle de l'écart type de la population ?

12. Une question posée lors d'une enquête à un échantillon de 150 individus a fourni 75 réponses oui, 55 réponses non et 20 sans opinion.


- a) Quelle est l'estimation ponctuelle de la proportion d'individus dans la population qui ont répondu oui ?
- b) Quelle est l'estimation ponctuelle de la proportion d'individus dans la population qui ont répondu non ?

Applications

13. Un échantillon aléatoire simple des données sur les ventes au cours de cinq mois a fourni les informations suivantes :

<i>Mois :</i>	1	2	3	4	5
<i>Unités vendues :</i>	94	100	85	94	92

- a) Développer une estimation ponctuelle du nombre moyen d'unités vendues par mois pour la population entière.
- b) Développer une estimation ponctuelle de l'écart type de la population.

 **14.** Morningstar publie les évaluations de 1 208 actions émises par des sociétés (site Internet de Morningstar, 24 octobre 2012). Un échantillon de 40 de ces actions est contenu dans le fichier en ligne Morningstar. Utiliser ce fichier pour répondre aux questions suivantes.

- a) Développer une estimation ponctuelle de la proportion d'actions qui sont notées 5 étoiles par Morningstar.
- b) Développer une estimation ponctuelle de la proportion d'actions qui sont notées « au-dessus de la moyenne » au regard de leur risque.
- c) Développer une estimation ponctuelle de la proportion d'actions qui sont notées au plus 2 étoiles.

15. La ligue nationale de football (NFL) a mené une enquête auprès des supporters pour évaluer les matchs (site Internet de la NFL, 24 octobre 2012). Chaque match est évalué sur une échelle allant de 0 (sans intérêt) à 100 (mémorable). Les évaluations des supporters pour un échantillon aléatoire de 12 matchs sont indiquées ci-dessous.

57	61	86	74	72	73
20	57	80	79	83	74

- a) Développer une estimation ponctuelle de la note moyenne attribuée par les supporters pour la population des matchs de la NFL.
- b) Développer une estimation ponctuelle de l'écart type pour la population des matchs de la NFL.

16. On a demandé à un échantillon de 426 adultes américains âgés de 50 ans et plus quelle était l'importance de différents thèmes dans leur choix d'un candidat lors des élections présidentielles de 2012 (*AARP Bulletin*, mars 2012).

- a) Quelle est la population échantillonnée dans cette étude ?
- b) La sécurité sociale et Medicare ont été cités comme « très importants » par 350 personnes. Estimer la proportion de la population des adultes américains âgés de 50 et plus qui pensent que cette question est très importante.
- c) L'éducation a été citée comme « très importante » par 74 % des personnes interrogées. Estimer le nombre de personnes interrogées qui pensent que cette question est très importante.
- d) La croissance de l'emploi a été citée comme « très importante » par 354 personnes interrogées. Estimer la proportion d'adultes américains de 50 ans et plus qui pensent que la croissance de l'emploi est très importante.
- e) Quelle est la population cible des inférences faites aux questions (b) et (d) ? Est-ce la même que la population échantillonnée que vous avez identifiée à la question (a) ? Supposez que vous appreniez plus tard que l'échantillon était restreint aux membres de l'association américaine des personnes retraitées (AARP). Pensez-vous encore que les inférences faites aux questions (b) et (d) sont valides ? Pourquoi ?

17. L'une des questions posées aux adultes dans le cadre de l'enquête Pew « Internet & American Life Project » était : « Utilisez-vous Internet, au moins occasionnellement ? » (site Internet de Pew, 23 octobre 2012). Les résultats ont révélé que 454 des 478 adultes âgés de 18 à 29 ans ont répondu oui ; 741 des 833 adultes âgés de 30 à 49 ans ont répondu oui ; et 1 058 des 1 644 adultes âgés de 50 ans et plus ont répondu oui.
- Développer une estimation ponctuelle de la proportion d'adultes âgés de 18 à 29 ans qui utilisent Internet.
 - Développer une estimation ponctuelle de la proportion d'adultes âgés de 30 à 49 ans qui utilisent Internet.
 - Développer une estimation ponctuelle de la proportion d'adultes âgés de 50 ans et plus qui utilisent Internet.
 - Commenter toute relation entre l'âge et l'usage d'Internet qui semble apparente.
 - Supposez que votre population cible soit celle de tous les adultes (âgés de 18 ans et plus). Développer une estimation de la proportion de cette population qui utilise Internet.

7.4 INTRODUCTION AUX DISTRIBUTIONS D'ÉCHANTILLONNAGE

Dans la section précédente, nous avons défini la moyenne d'échantillon \bar{x} comme l'estimateur ponctuel de la moyenne de la population μ et la proportion d'échantillon \bar{p} comme l'estimateur ponctuel de la proportion de la population p . Dans le cadre de l'échantillon aléatoire simple des 30 employés de la société EAI, présenté dans le tableau 7.2, l'estimation ponctuelle de μ est $\bar{x} = 51\,814$ dollars et l'estimation ponctuelle de p est $\bar{p} = 0,63$. Supposez que nous sélectionnions un autre échantillon aléatoire simple de 30 employés de la société EAI, et que nous obtenions les estimations ponctuelles suivantes :

Moyenne d'échantillon $\bar{x} = 52\,670$ dollars
 Proportion de l'échantillon $\bar{p} = 0,70$

Tableau 7.4 Valeurs de \bar{x} et \bar{p} obtenues à partir de 500 échantillons aléatoires simples de 30 employés de la société EAI

Numéro de l'échantillon	Moyenne de l'échantillon (\bar{x})	Proportion de l'échantillon (\bar{p})
1	51 814	0,63
2	52 670	0,70
4	51 780	0,67
5	51 588	0,53
...
500	51 752	0,50

Tableau 7.5 Distribution de fréquence de \bar{x} obtenue à partir de 500 échantillons aléatoires simples de 30 employés de la société EAI

Salaire annuel moyen (\$)	Fréquence	Fréquence relative
49 500,00-49 999,99	2	0,004
50 000,00-50 499,99	16	0,032
50 500,00-50 999,99	52	0,104
51 000,00-51 499,99	101	0,202
51 500,00-51 999,99	133	0,266
52 000,00-52 499,99	110	0,220
52 500,00-52 999,99	54	0,108
53 000,00-53 499,99	26	0,052
53 500,00-53 999,99	6	0,012
Total	500	1,000

Ces résultats fournissent des valeurs de \bar{x} et \bar{p} différentes de celles obtenues avec le premier échantillon. De manière générale, un second échantillon aléatoire simple n'est pas censé fournir les mêmes estimations ponctuelles que le premier.

Supposez maintenant que nous répétons maintes et maintes fois le processus de sélection d'un échantillon aléatoire simple de 30 employés de la société EAI, calculant à chaque fois les valeurs de \bar{x} et \bar{p} . Le tableau 7.4 contient une partie des résultats obtenus pour 500 échantillons aléatoires simples et le tableau 7.5 présente les distributions de fréquence absolue et relative des 500 valeurs de \bar{x} . La figure 7.1 représente l'histogramme des fréquences relatives des valeurs de \bar{x} .

Dans le chapitre 5, nous avons défini une variable aléatoire comme étant une description numérique du résultat d'une expérience. Si nous considérons le processus de sélection d'un échantillon aléatoire simple comme une expérience, la moyenne d'échantillon \bar{x} correspond à la description numérique du résultat de l'expérience. Ainsi, la moyenne d'échantillon \bar{x} est une variable aléatoire. Par conséquent, comme pour toute autre variable aléatoire, \bar{x} a une espérance mathématique, une variance et une distribution de probabilité. Puisque les différentes valeurs possibles de \bar{x} résultent d'échantillons aléatoires simples différents, la distribution de probabilité de \bar{x} est appelée **distribution d'échantillonnage de \bar{x}** . La connaissance de cette distribution d'échantillonnage et de ses propriétés nous permet de tirer des conclusions en termes de probabilités quant à l'écart entre la moyenne d'échantillon \bar{x} et la moyenne de la population μ .

La bonne compréhension des chapitres suivants repose sur la capacité de compréhension et d'utilisation des distributions d'échantillonnage présentées dans ce chapitre.

Revenons au graphique 7.1. Pour déterminer de façon précise la distribution d'échantillonnage de \bar{x} , il faudrait énumérer tous les échantillons possibles de 30 employés

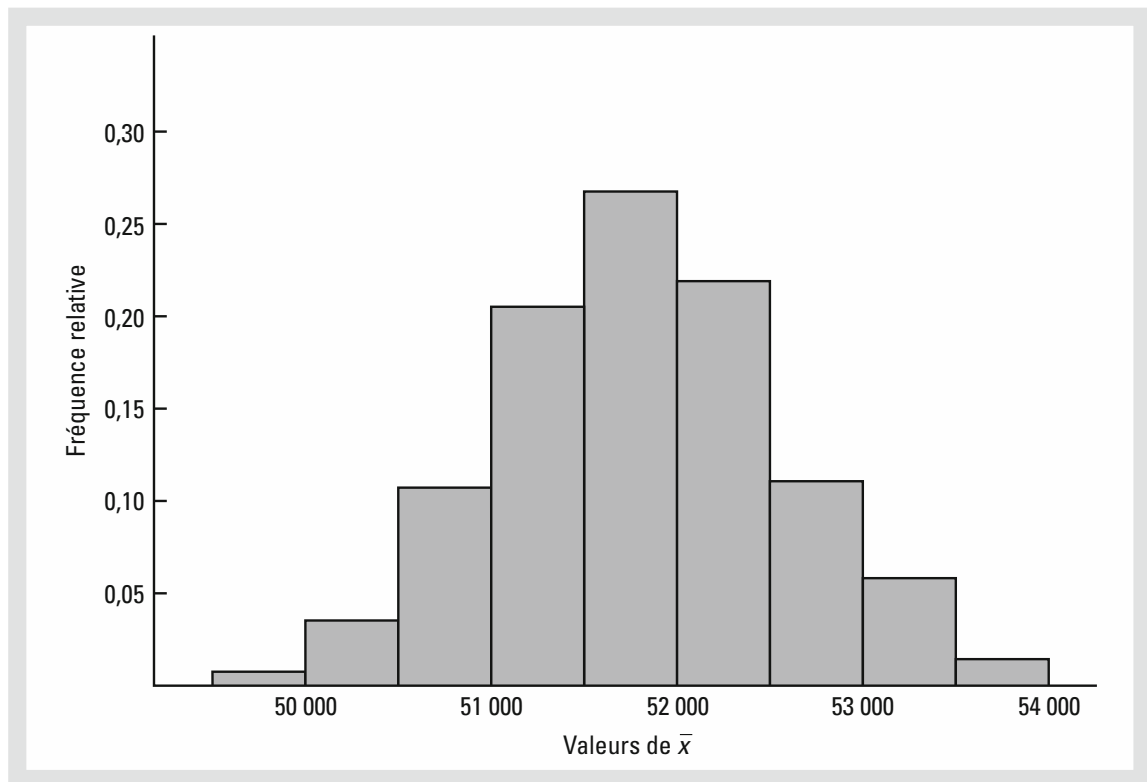


Figure 7.1 Histogramme de la fréquence relative des valeurs de \bar{x} obtenues à partir de 500 échantillons aléatoires simples de taille égale à 30

et calculer chaque moyenne d'échantillon. Cependant, l'histogramme des 500 valeurs de \bar{x} fournit une approximation de cette distribution d'échantillonnage. Grâce à cet histogramme, nous observons que la distribution est en forme de cloche. Notons que la plus forte concentration des valeurs de \bar{x} et la moyenne des 500 valeurs de \bar{x} sont proches de la moyenne de la population, $\mu = 51\,800$ dollars. Nous décrirons les propriétés de la distribution d'échantillonnage de \bar{x} plus longuement dans la section suivante.

Les 500 valeurs de la proportion d'échantillon \bar{p} sont résumées par l'histogramme de la fréquence relative, représenté à la figure 7.2. Comme dans le cas de \bar{x} , \bar{p} est une variable aléatoire. Si tous les échantillons de taille 30 possibles étaient sélectionnés à partir de la population et si une valeur de \bar{p} était calculée pour chaque échantillon, la distribution de probabilité associée correspondrait à la distribution d'échantillonnage de \bar{p} . L'histogramme de la fréquence relative des 500 valeurs d'échantillon (figure 7.2) reflète la forme générale de la distribution d'échantillonnage de \bar{p} .

En pratique, on ne constitue qu'un seul échantillon aléatoire simple à partir de la population. Nous avons répété le processus d'échantillonnage 500 fois dans cette section simplement pour illustrer le fait que de nombreux échantillons différents sont possibles et qu'ils génèrent diverses valeurs pour les statistiques d'échantillon \bar{x} et \bar{p} . La distribution de probabilité d'une statistique d'échantillon particulière est appelée distribution d'échantillonnage de cette statistique. Dans les sections 7.5 et 7.6, nous verrons respectivement les caractéristiques de la distribution d'échantillonnage de \bar{x} et de \bar{p} .

7.5 DISTRIBUTION D'ÉCHANTILLONNAGE DE \bar{x}

Dans la section précédente, nous avons vu que la moyenne d'échantillon \bar{x} est une variable aléatoire et sa distribution de probabilité est appelée distribution d'échantillonnage de \bar{x} .

► Distribution d'échantillonnage de \bar{x}

La distribution d'échantillonnage de \bar{x} correspond à la distribution de probabilité de toutes les valeurs possibles de la moyenne d'échantillon \bar{x} .

Cette section décrit les propriétés de la distribution d'échantillonnage de \bar{x} . Comme pour d'autres distributions de probabilité, la distribution d'échantillonnage de \bar{x} a une espérance mathématique, un écart type et une forme caractéristique. Commençons en considérant la moyenne de toutes les valeurs possibles de \bar{x} , qui correspond à l'espérance mathématique de \bar{x} .

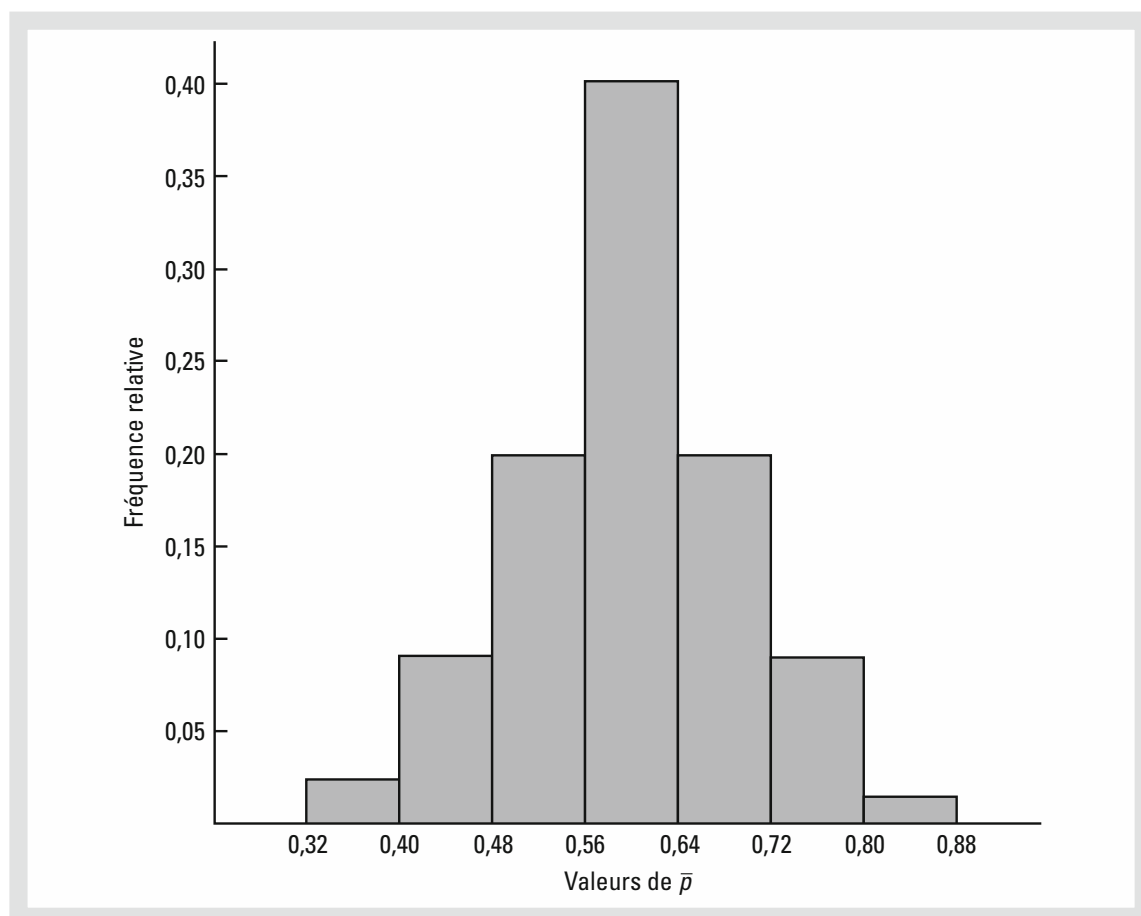


Figure 7.2 Histogramme de la fréquence relative des valeurs de \bar{p} obtenues à partir de 500 échantillons aléatoires simples de taille égale à 30

7.5.1 Espérance mathématique de \bar{x}

Dans le problème d'échantillonnage de la société EAI, nous avons constaté que différents échantillons aléatoires simples conduisent à diverses valeurs de la moyenne d'échantillon \bar{x} . Puisque de nombreuses valeurs différentes de la variable aléatoire \bar{x} sont possibles, on s'intéresse souvent à la moyenne de toutes les valeurs possibles de \bar{x} générées par les divers échantillons aléatoires simples. La moyenne de la variable aléatoire \bar{x} correspond à l'espérance mathématique de \bar{x} . Soient $E(\bar{x})$ l'espérance mathématique de \bar{x} et μ la moyenne de la population d'où est issu un échantillon aléatoire simple. On peut montrer qu'avec un échantillonnage aléatoire simple, $E(\bar{x})$ et μ sont égaux.

► **Espérance mathématique de \bar{x}**

$$E(\bar{x}) = \mu \quad (7.1)$$

où

$E(\bar{x})$ correspond à l'espérance mathématique de \bar{x}
 μ correspond à la moyenne de la population

L'espérance mathématique de \bar{x} est égale à la moyenne de la population d'où est issu l'échantillon.

Ce résultat indique qu'avec un échantillonnage aléatoire simple, l'espérance mathématique ou la moyenne de la distribution d'échantillonnage de \bar{x} est égale à la moyenne de la population. Dans la section 7.1, nous avons calculé le salaire annuel moyen pour la population des employés de la société EAI : il est égal à 51 800 dollars. Ainsi, selon l'équation (7.1), la moyenne de toutes les moyennes d'échantillons possibles dans le cadre du problème de la société EAI est également égale à 51 800 dollars.

Lorsque l'espérance mathématique d'un estimateur ponctuel est égale au paramètre de la population, on dit que l'estimateur ponctuel est **sans biais**. Ainsi, l'équation (7.1) indique que \bar{x} est un estimateur sans biais de la moyenne de la population μ .

7.5.2 Écart type de \bar{x}

Définissons l'écart type de la distribution d'échantillonnage de \bar{x} . Nous utilisons la notation suivante :

$\sigma_{\bar{x}}$ pour l'écart type de \bar{x}
 σ pour l'écart type de la population
 n pour la taille de l'échantillon
 N pour la taille de la population

On peut montrer que la formule de l'écart type de \bar{x} dépend du type de population considérée, finie ou infinie. Les deux formules de l'écart type de \bar{x} correspondent à :

► **Écart type de \bar{x}**

<i>Population finie</i>	<i>Population infinie</i>	
$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \left(\frac{\sigma}{\sqrt{n}} \right)$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	(7.2)

En comparant les deux formules, on voit que le facteur $\sqrt{(N-n)/(N-1)}$ est nécessaire pour calculer l'écart type de \bar{x} dans le cas d'une population finie mais pas dans le cas d'une population infinie. Ce facteur est communément appelé **facteur de correction pour une population finie**. Dans de nombreux cas d'échantillonnage, la population, bien que finie, est « grande », alors que la taille de l'échantillon est relativement « petite ». Dans de tels cas, le facteur de correction $\sqrt{(N-n)/(N-1)}$ est proche de 1. En conséquence, la différence entre les deux valeurs de l'écart type de \bar{x} pour les cas de population finie et infinie devient négligeable. Alors, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ devient une bonne approximation de l'écart type de \bar{x} même si la population est finie. Cette observation conduit à la règle générale suivante pour calculer l'écart type de \bar{x} .

► **Utiliser l'expression suivante pour calculer l'écart type de \bar{x}**

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (7.3)$$

Lorsque

1. La population est infinie ; ou
2. La population est finie et la taille de l'échantillon est inférieure ou égale à 5 % de la taille de la population ; c'est-à-dire si $n/N \leq 0,05$.

Dans les cas où $n/N > 0,05$, la version de la formule (7.2) pour population finie devrait être utilisée pour calculer $\sigma_{\bar{x}}$. Sauf mention contraire, à travers l'ouvrage, nous supposons que la population est suffisamment grande pour que $n/N \leq 0,05$ et l'expression (7.3) peut être utilisée pour calculer $\sigma_{\bar{x}}$.

Le problème 21 montre que lorsque $n/N \leq 0,05$, le facteur de correction pour une population finie a peu d'impact sur la valeur de $\sigma_{\bar{x}}$.

Pour calculer $\sigma_{\bar{x}}$, il nous faut connaître σ , l'écart type de la population. Pour bien souligner la différence entre $\sigma_{\bar{x}}$ et σ , nous nommerons l'écart type de \bar{x} , $\sigma_{\bar{x}}$, l'**erreur type** de la moyenne. En général, le terme d'*erreur type* est employé pour désigner l'écart type d'un estimateur ponctuel. Plus tard, nous verrons que la valeur de l'erreur type de la moyenne est utile pour déterminer l'écart entre la moyenne d'échantillon et la moyenne de la population. Revenons maintenant au problème de la société EAI et déterminons l'erreur type de la moyenne associée aux échantillons aléatoires simples de 30 employés de la société EAI.

Le terme erreur type est employé en inférence statistique pour désigner l'écart type d'un estimateur ponctuel.

Dans la section 7.1, nous avons montré que l'écart type du salaire annuel de la population des 2 500 employés de EAI est égal à 4 000 dollars. Dans ce cas, la population est finie, avec $N = 2\,500$. Cependant, avec un échantillon de taille 30, nous avons $n/N = 30/2500 = 0,012$. Puisque la taille de l'échantillon est inférieure à 5 % de la taille de la population, nous pouvons ignorer le facteur de correction pour une population finie et utiliser l'expression (7.3) pour calculer l'erreur type de \bar{x} .

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730,3$$

7.5.3 Forme de la distribution d'échantillonnage de \bar{x}

Les résultats précédents concernant l'espérance mathématique et l'écart type de la distribution d'échantillonnage de \bar{x} sont applicables à toutes populations. La dernière étape dans l'identification des caractéristiques de la distribution d'échantillonnage de \bar{x} correspond à la détermination de la forme de la distribution d'échantillonnage. Nous considérons deux cas : (1) la population a une distribution normale ; (2) la population n'a pas une distribution normale.

La population a une distribution normale. Dans de nombreuses situations, il est raisonnable de supposer que la population à partir de laquelle est sélectionné un échantillon aléatoire simple, a une distribution normale ou presque normale. Lorsque la population a une distribution normale, la distribution d'échantillonnage de \bar{x} est normalement distribuée quelle que soit la taille de l'échantillon.

La population n'a pas une distribution normale. Lorsque la population à partir de laquelle est sélectionné un échantillon aléatoire simple, n'a pas une distribution normale, le théorème central limite permet d'identifier la forme de la distribution d'échantillonnage de \bar{x} . Une définition du théorème central limite applicable à la distribution d'échantillonnage de \bar{x} est donnée ci-dessous.

► **Théorème central limite**

En sélectionnant des échantillons aléatoires simples de taille n à partir d'une population, la distribution d'échantillonnage de la moyenne d'échantillon \bar{x} peut être approchée par une distribution de probabilité normale lorsque la taille de l'échantillon devient importante.

La figure 7.3 montre comment s'applique le théorème central limite pour trois populations différentes ; chaque colonne correspond à l'une des populations. En haut de la figure, aucune des populations n'est normalement distribuée. La population I suit une loi uniforme. La population II est souvent qualifiée de distribution en forme d'oreilles de lapin. Elle est symétrique, mais les valeurs les plus vraisemblables se situent dans les queues de la distribution. La population III a une forme similaire à une loi exponentielle ; elle est asymétrique à droite.

La partie inférieure de la figure 7.3 représente la forme de la distribution d'échantillonnage pour des échantillons de taille $n = 2$, $n = 5$ et $n = 30$. Lorsque la taille de

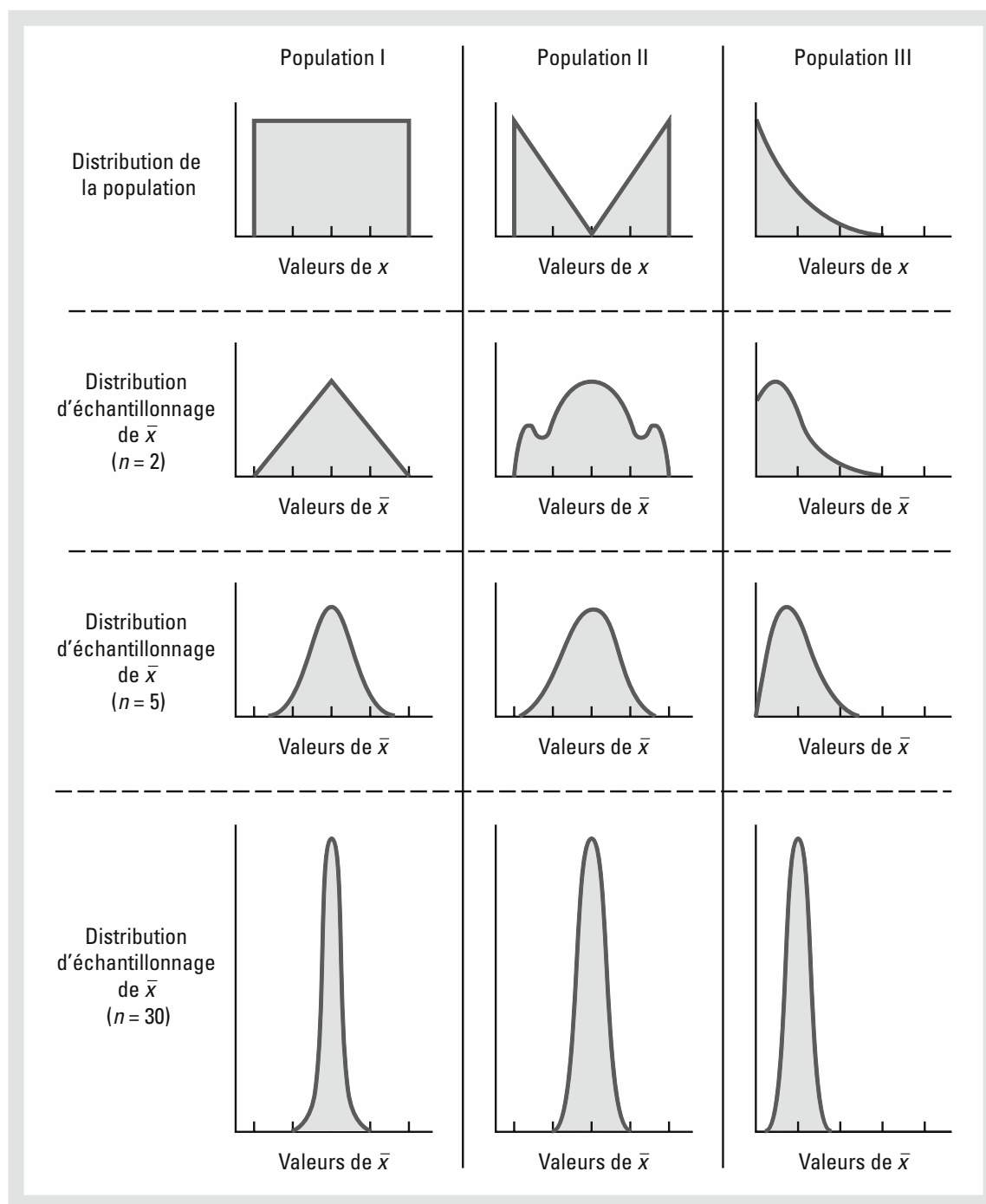


Figure 7.3 Illustration du théorème central limite pour trois populations

l'échantillon est égale à 2, la forme de chaque distribution d'échantillonnage est différente de la forme de la distribution de la population correspondante. Pour des échantillons de taille égale à 5, la forme des distributions d'échantillonnage des populations I et II commence à ressembler à la forme d'une distribution normale. Même si la forme de la distribution d'échantillonnage de la population III commence à ressembler à la forme d'une

distribution normale, une asymétrie à droite est encore présente. Finalement, pour des échantillons de taille égale à 30, les formes de chacune des trois distributions d'échantillonnage sont approximativement normales.

D'un point de vue pratique, nous souhaitons connaître la taille minimale de l'échantillon nécessaire pour appliquer le théorème central limite et supposer que la forme de la distribution d'échantillonnage est approximativement normale. Les statisticiens ont étudié cette question en observant la distribution d'échantillonnage de \bar{x} pour une variété de populations et de tailles d'échantillon. La pratique veut que, pour la plupart des applications, la distribution d'échantillonnage de \bar{x} puisse être approchée par une loi normale lorsque la taille de l'échantillon est supérieure ou égale à 30. Dans les cas où la population est fortement asymétrique ou lorsque des valeurs aberrantes sont présentes, une taille d'échantillon de 50 est nécessaire. Finalement, si la population est discrète, la taille de l'échantillon nécessaire pour une approximation normale dépend souvent de la proportion de la population. Nous en dirons plus à ce sujet dans la section 7.6 consacrée à la distribution d'échantillonnage de \bar{p} .

7.5.4 Distribution d'échantillonnage de \bar{x} pour le problème de la société EAI

Dans l'étude la société EAI, nous avons montré que $E(\bar{x}) = 51\,800$ et $\sigma_{\bar{x}} = 730,3$. Nous n'avons pas d'information concernant la distribution de la population ; elle peut être normale ou non. Si la population a une distribution normale, la distribution d'échantillonnage de \bar{x} est normale. Si la population n'a pas une distribution normale, l'échantillon aléatoire simple de 30 employés et le théorème central limite nous permettent de conclure que la distribution d'échantillonnage de \bar{x} est approximativement normale. Dans chacun des cas, nous pouvons conclure que la distribution d'échantillonnage de \bar{x} peut être décrite par une loi normale, représentée par la figure 7.4.

7.5.5 Intérêt pratique de la distribution d'échantillonnage de \bar{x}

Lorsqu'un échantillon aléatoire simple est sélectionné et que la valeur de la moyenne d'échantillon \bar{x} est utilisée pour estimer la valeur de la moyenne de la population μ , on ne peut s'attendre à ce que la moyenne d'échantillon soit exactement égale à la moyenne de la population. La raison pour laquelle on s'intéresse à la distribution d'échantillonnage de \bar{x} , est qu'elle peut fournir des informations probabilistes sur l'écart entre la moyenne d'échantillon et la moyenne de la population. Pour le démontrer, revenons au problème de la société EAI.

Supposez que le directeur du personnel considère la moyenne d'échantillon comme une estimation acceptable de la moyenne de la population, si la différence en valeur absolue entre la moyenne d'échantillon et la moyenne de la population est inférieure ou égale à 500 dollars. Cependant, il n'est pas possible de garantir que cette condition est satisfaite. Au contraire, le tableau 7.5 et la figure 7.1 montrent que certaines moyennes d'échantillon, parmi les 500 échantillons, s'écartent de plus de 2 000 dollars de la moyenne de la population. Aussi, devons nous interpréter la requête du directeur du personnel en termes de probabilité. Autrement dit, le directeur du personnel s'intéresse à la question suivante : Quelle est

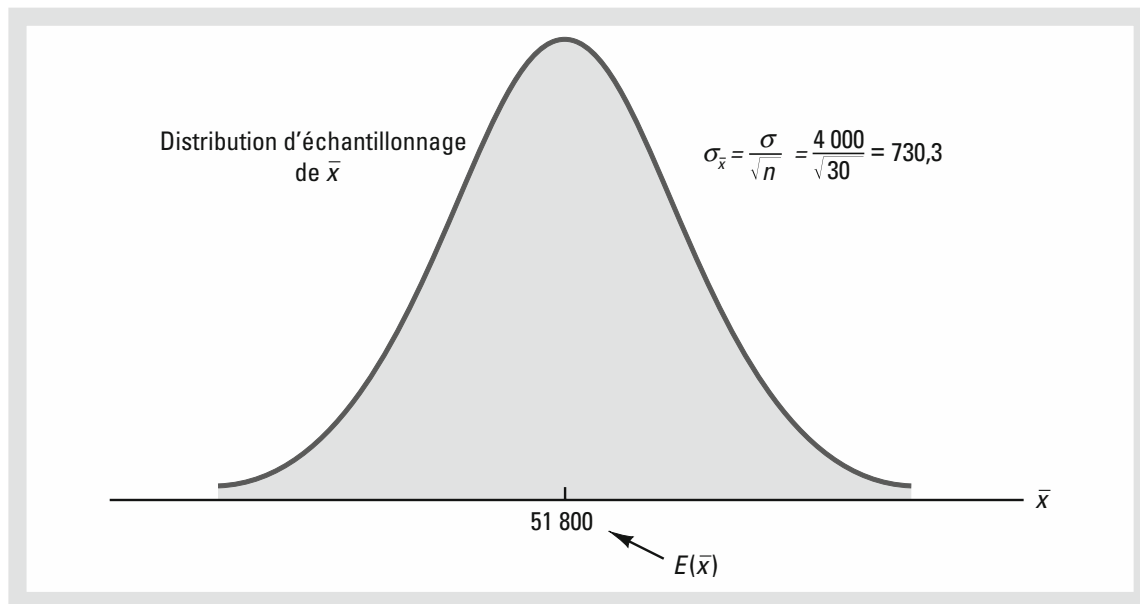


Figure 7.4 Distribution d'échantillonnage de \bar{x} pour le salaire annuel moyen d'un échantillon aléatoire simple de 30 employés de la société EAI

la probabilité que la moyenne d'un échantillon de 30 employés de la société EAI s'écarte, au plus, de 500 dollars en valeur absolue de la moyenne de la population ?

Puisque nous avons identifié les propriétés de la distribution d'échantillonnage de \bar{x} (voir figure 7.4), nous utiliserons cette distribution pour déterminer la probabilité recherchée. Référez-vous à la distribution d'échantillonnage de \bar{x} représentée de nouveau à la figure 7.5. La moyenne de la population étant égale à 51 800 dollars, le directeur du

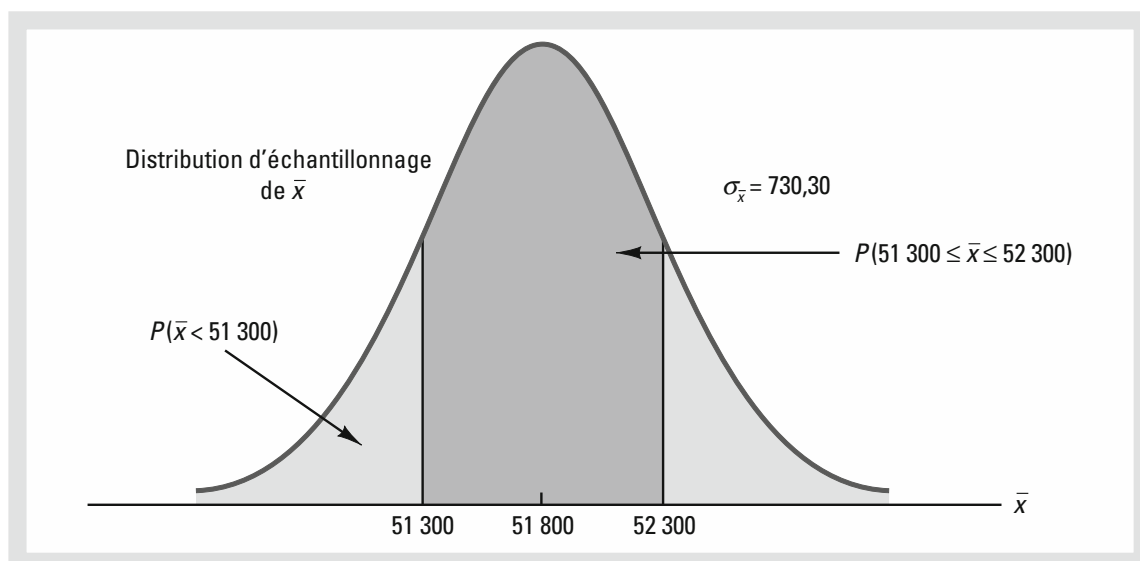


Figure 7.5 Probabilité qu'une moyenne d'échantillon s'écarte d'au plus 500 dollars de la moyenne de la population, en valeur absolue, pour un échantillon aléatoire simple de 30 employés de la société EAI

personnel cherche à déterminer la probabilité que la moyenne d'échantillon soit comprise entre 51 300 et 52 300 dollars. Cette probabilité correspond à l'aire de la partie grisée de la distribution d'échantillonnage représentée à la figure 7.5. Puisque la distribution d'échantillonnage est normale, de moyenne égale à 51 800 et d'écart type égal à 730,3, nous pouvons utiliser la table de la loi normale centrée réduite pour trouver la probabilité recherchée.

Nous calculons tout d'abord la valeur z associée à la limite supérieure de l'intervalle (52 300) et utilisons la table pour déterminer l'aire sous la courbe à gauche de ce point (l'aire dans la queue gauche). Ensuite, nous calculons la valeur z associée à la limite inférieure de l'intervalle (51 300) et utilisons la table pour déterminer l'aire sous la courbe à gauche de ce point (l'aire dans une autre queue gauche). En soustrayant la seconde aire à la première nous obtenons la probabilité souhaitée.

$$\begin{aligned} \text{Au point } \bar{x} = 52\,300, \text{ nous avons} \\ z = \frac{52\,300 - 51\,800}{730,3} = 0,68 \end{aligned}$$

En se référant à la table des probabilités normales centrées réduites, nous trouvons une probabilité cumulée (l'aire à gauche de $z = 0,68$) égale à 0,7517.

$$\begin{aligned} \text{Au point } \bar{x} = 51\,300, \text{ nous avons} \\ z = \frac{51\,300 - 51\,800}{730,3} = -0,68 \end{aligned}$$

L'aire sous la courbe à gauche de $z = -0,68$ est égale à 0,2483. Ainsi, $P(51\,300 \leq \bar{x} \leq 52\,300) = P(z \leq 0,68) - P(z \leq -0,68) = 0,7517 - 0,2483 = 0,5034$.

Les calculs précédents indiquent qu'un échantillon aléatoire simple de 30 employés de la société EAI a une probabilité de 0,5034 de fournir une moyenne d'échantillon \bar{x} qui ne s'écarte pas de plus de 500 dollars, en valeur absolue, de la moyenne de la population. Ainsi, il y a une probabilité de 0,4966 ($1 - 0,5034 = 0,4966$) que la moyenne d'échantillon sous- ou surestime la moyenne de la population de plus de 500 dollars. En d'autres termes, un échantillon aléatoire simple de 30 employés de la société EAI a presque une chance sur deux d'être dans l'intervalle acceptable de 500 dollars autour de la moyenne de la population. Peut-être faudrait-il envisager une taille plus importante de l'échantillon. Explorons cette hypothèse en considérant la relation entre la taille de l'échantillon et la distribution d'échantillonnage de \bar{x} .

La distribution d'échantillonnage de \bar{x} peut fournir des informations probabilistes sur l'écart entre la moyenne d'échantillon \bar{x} et la moyenne de la population μ .

7.5.6 Relation entre la taille de l'échantillon et la distribution d'échantillonnage de \bar{x}

Supposez que dans le problème de la société EAI, nous sélectionnons un échantillon aléatoire simple de 100 employés de la société au lieu des 30 considérés à l'origine. Intuitivement, il est vraisemblable qu'avec un échantillon plus grand de taille égale à 100,

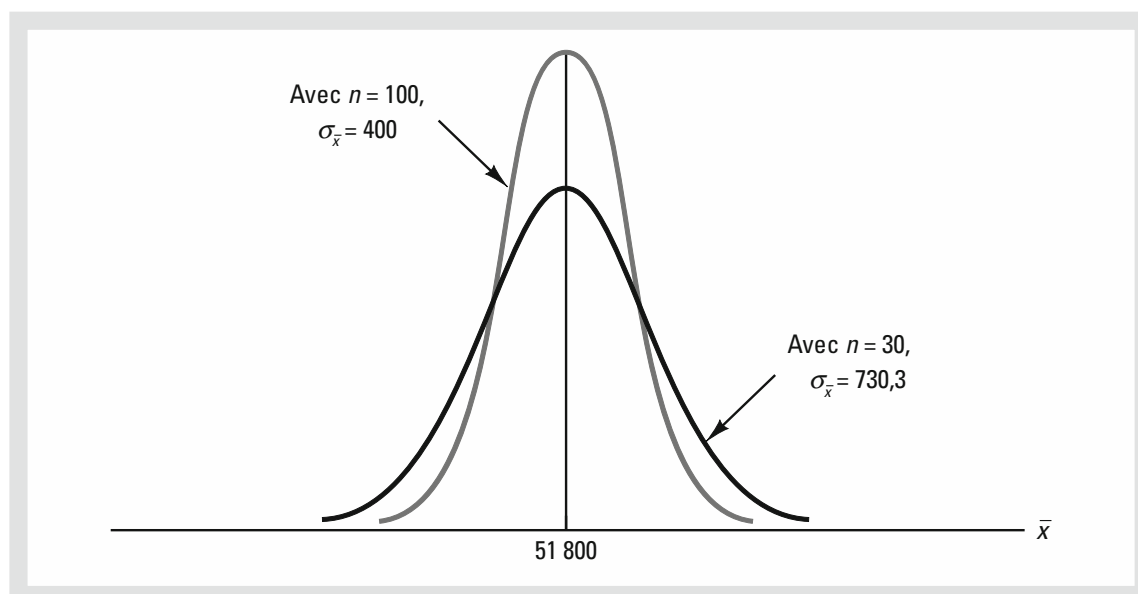


Figure 7.6 Comparaison des distributions d'échantillonnage de \bar{x} pour des échantillons aléatoires simples de taille $n = 30$ et $n = 100$ employés de la société EAI

donc avec plus de données, la moyenne d'échantillon fournit une meilleure estimation de la moyenne de la population qu'une moyenne d'échantillon basée sur un échantillon de 30 employés. Pour mesurer l'importance de l'amélioration, considérons la relation entre la taille de l'échantillon et la distribution d'échantillonnage de \bar{x} .

Tout d'abord, notez que $E(\bar{x}) = \mu$ quelle que soit la taille de l'échantillon. Ainsi, la moyenne de toutes les valeurs possibles de \bar{x} est égale à la moyenne de la population μ , quelle que soit la taille n de l'échantillon. Cependant, notez que l'erreur type de la moyenne, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, est liée à la racine carrée de la taille de l'échantillon. Lorsque la taille de l'échantillon augmente, l'erreur type de la moyenne $\sigma_{\bar{x}}$ diminue. Avec $n = 30$, l'erreur type de la moyenne pour le problème de la société EAI est égale à 730,3. Cependant, avec l'augmentation de la taille de l'échantillon à 100, l'erreur type de la moyenne diminue à

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{100}} = 400$$

Les distributions d'échantillonnage de \bar{x} pour $n = 30$ et $n = 100$ sont représentées à la figure 7.6. Puisque la distribution d'échantillonnage pour $n = 100$ a une plus petite erreur type, les valeurs de \bar{x} varient moins et ont tendance à être plus proches de la moyenne de la population que les valeurs de \bar{x} obtenues avec un échantillon de taille $n = 30$.

Nous pouvons utiliser la distribution d'échantillonnage de \bar{x} dans le cas où $n = 100$ pour calculer la probabilité qu'un échantillon aléatoire simple de 100 employés de la société EAI fournisse une moyenne d'échantillon qui ne s'écarte pas de plus de 500 dollars, en valeur absolue, de la moyenne de la population. Puisque la distribution d'échantillonnage est normale, de moyenne égale à 51 800 et d'erreur type égale à 400,

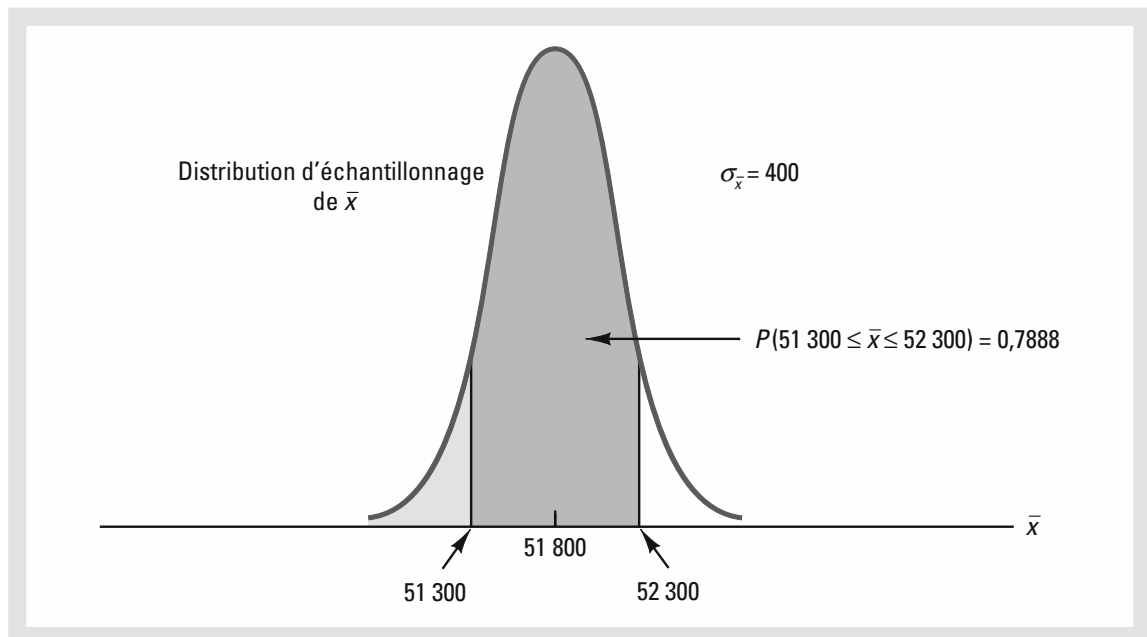


Figure 7.7 Probabilité qu'une moyenne d'échantillon s'écarte d'au plus 500 dollars de la moyenne de la population, en valeur absolue, pour un échantillon aléatoire simple de 100 employés de la société EAI

nous pouvons utiliser la table de la loi normale centrée réduite pour trouver la probabilité cherchée.

Au point $\bar{x} = 52\,300$ (figure 7.7), nous avons

$$z = \frac{52300 - 51800}{400} = 1,25$$

En nous référant à la table de la loi normale centrée réduite, nous trouvons que la probabilité cumulée correspondant à $z = 1,25$ est égale à 0,8944.

Au point $\bar{x} = 51\,300$, nous avons

$$z = \frac{51300 - 51800}{400} = -1,25$$

La probabilité cumulée correspondant à $z = -1,25$ est égale à 0,1056. Ainsi, $P(51300 \leq \bar{x} \leq 52300) = P(z \leq 1,25) - P(z \leq -1,25) = 0,8944 - 0,1056 = 0,7888$. En augmentant la taille de l'échantillon de 30 à 100 employés de la société EAI, la probabilité d'obtenir une moyenne d'échantillon dans un intervalle de 500 dollars de part et d'autre de la moyenne de la population, est passée de 0,5034 à 0,7888.

Le point important de cette discussion est que l'erreur type de la moyenne diminue lorsque la taille de l'échantillon augmente. Par conséquent, plus l'échantillon est grand, plus la probabilité que la moyenne d'échantillon soit comprise dans un intervalle précis autour de la moyenne de la population est élevée.

REMARQUES

1. En présentant la distribution d'échantillonnage de \bar{x} dans le cadre du problème de la société EAI, nous avons tiré parti du fait que la moyenne de la population, $\mu = 51\,800$, et l'écart type de la population, $\sigma = 4\,000$, étaient connus. Cependant, en général, les valeurs de la moyenne de la population μ et de l'écart type de la population σ , nécessaires pour déterminer la distribution d'échantillonnage de \bar{x} , ne sont pas connues. Dans le chapitre 8, nous verrons comment sont utilisés la moyenne d'échantillon \bar{x} et l'écart type d'échantillon s lorsque μ et σ sont inconnus.
2. L'application théorique du théorème central limite nécessite que les observations de l'échantillon soient indépendantes. Cette condition est satisfaite pour des populations infinies ou des populations finies dans lesquelles l'échantillonnage est fait avec remise. Bien que le théorème central limite ne s'adresse pas directement à l'échantillonnage sans remise effectué à partir de populations finies, dans la pratique, on applique les résultats du théorème central limite à ce cas, lorsque la taille de la population est grande.

EXERCICES

Méthode

18. Une population est caractérisée par une moyenne égale à 200 et un écart type égal à 50. Un échantillon aléatoire simple de taille égale à 100 est sélectionné et la moyenne d'échantillon \bar{x} est utilisée pour estimer la moyenne de la population.
 - a) Quelle est l'espérance mathématique de \bar{x} ?
 - b) Quel est l'écart type de \bar{x} ?
 - c) Représenter la distribution d'échantillonnage de \bar{x} .
 - d) Que montre la distribution d'échantillonnage de \bar{x} ?
19. Une population est caractérisée par une moyenne égale à 200 et un écart type égal à 50. Un échantillon aléatoire simple de taille égale à 100 est sélectionné et \bar{x} est utilisé pour estimer μ .
 - a) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 5 de la moyenne de la population ?
 - b) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 10 de la moyenne de la population ?
20. Supposez que l'écart type de la population soit $\sigma = 25$. Calculer l'erreur type de la moyenne, $\sigma_{\bar{x}}$, pour des échantillons de taille égale à 50, 100, 150 et 200. Que pouvez-vous dire quant à l'ampleur de l'erreur type de la moyenne lorsque la taille de l'échantillon augmente ?
21. Supposez qu'un échantillon aléatoire simple de taille 50 soit constitué à partir d'une population caractérisée par $\sigma = 10$. Trouver la valeur de l'erreur type de la moyenne dans chacun des cas suivants (utiliser le facteur de correction pour population finie, si nécessaire).



- a) La taille de la population est infinie.
- b) La taille de la population est $N = 50\,000$.
- c) La taille de la population est $N = 5\,000$.
- d) La taille de la population est $N = 500$.

Applications

22. Référez-vous au problème d'échantillonnage de la société EAI. Supposez qu'un échantillon aléatoire simple de 60 employés soit sélectionné.
- a) Dessiner la distribution d'échantillonnage de \bar{x} lorsque des échantillons aléatoires simples de taille 60 sont utilisés.
 - b) Que devient la distribution d'échantillonnage de \bar{x} si des échantillons aléatoires simples de taille 120 sont utilisés ?
 - c) Quelle conclusion générale pouvez-vous tirer concernant la distribution d'échantillonnage de \bar{x} lorsque la taille de l'échantillon augmente ? Est-ce que cela semble logique ? Expliquer.
23. Dans le problème d'échantillonnage de la société EAI (cf. figure 7.5), nous avons montré que pour $n = 30$ il y avait une probabilité de 0,5034 d'obtenir une moyenne d'échantillon qui s'écarte au plus de ± 500 dollars de la moyenne de la population.
- a) Quelle est la probabilité que \bar{x} s'écarte au plus de ± 500 dollars de la moyenne de la population si un échantillon de taille 60 est utilisé ?
 - b) Répondre à la question (a) pour un échantillon de taille 120.
24. Le magazine Barron's a rapporté que le nombre moyen de semaines passées au chômage par un individu est égal à 17,5 (Barron's, 18 février 2008). Supposez que pour la population de tous les chômeurs, la durée moyenne de chômage de la population soit de 17,5 semaines et que l'écart type de la population soit de 4 semaines. Supposez que vous vouliez sélectionner un échantillon aléatoire de 50 chômeurs pour effectuer une étude.
- a) Représenter la distribution d'échantillonnage de \bar{x} , la moyenne d'échantillon pour un échantillon de 50 chômeurs.
 - b) Quelle est la probabilité qu'un échantillon aléatoire simple de 50 chômeurs fournisse une moyenne d'échantillon qui s'écarte au plus de ± 1 semaine de la moyenne de la population ?
 - c) Quelle est la probabilité qu'un échantillon aléatoire simple de 50 chômeurs fournisse une moyenne d'échantillon qui s'écarte au plus de $\pm 1/2$ semaine de la moyenne de la population ?
25. Le conseil d'éducation des lycées américains a rapporté la moyenne des notes obtenues aux trois épreuves du test d'aptitude scolaire SAT (*The World Almanac*, 2009) :

Lecture critique :	502
Mathématiques :	515
Rédaction :	494

Supposez que l'écart type de la population pour chaque épreuve soit égal à $\sigma = 100$.



- a) Quelle est la probabilité qu'un échantillon aléatoire de 90 lycéens fournisse une note moyenne qui s'écarte au plus de ± 10 de la moyenne de la population égale à 502 pour l'épreuve de lecture critique ?
- b) Quelle est la probabilité qu'un échantillon aléatoire de 90 lycéens fournisse une note moyenne qui s'écarte au plus de ± 10 de la moyenne de la population égale à 515 pour l'épreuve de mathématiques ? Comparer cette probabilité à celle calculée à la question (a).
- c) Quelle est la probabilité qu'un échantillon aléatoire de 100 lycéens fournisse une note moyenne qui s'écarte au plus de ± 10 de la moyenne de la population égale à 494 pour l'épreuve de rédaction ? Commenter les différences entre cette probabilité et les valeurs calculées aux questions (a) et (b).
26. Pour l'année 2010, 33 % des contribuables dont le revenu brut imposable est compris entre 30 000 et 60 000 dollars, ont fourni une liste d'éléments déductibles de leurs impôts (*The Wall Street Journal*, 25 octobre 2012). Le montant moyen des déductions pour cette population de contribuables s'élevait à 16 642 dollars. Supposez que l'écart type soit égal à 2 400 dollars.
- a) Quelle est la probabilité qu'un échantillon de contribuables qui appartiennent à ce groupe de revenus et qui ont fourni une liste d'éléments déductibles, fournisse une moyenne d'échantillon qui s'écarte de plus ou moins 200 dollars de la moyenne de la population pour chacune des tailles d'échantillon suivantes : 30, 50, 100 et 400 ?
- b) Quel est l'avantage d'avoir une taille d'échantillon assez importante, lorsque l'on souhaite estimer la moyenne d'une population ?
27. L'institut de politique économique publie périodiquement des rapports sur les salaires des travailleurs lors de leur entrée dans la vie active. L'institut a rapporté que les salaires de départ des hommes diplômés de l'université étaient de 21,68 dollars de l'heure et celui des femmes diplômées de l'université de 18,80 dollars de l'heure en 2011 (site Internet de l'institut de politique économique, 30 mars 2012). Supposez que l'écart type pour les hommes diplômés soit égal à 2,30 dollars et pour les femmes diplômés à 2,05 dollars.
- a) Quelle est la probabilité qu'un échantillon de 50 hommes diplômés fournisse une moyenne d'échantillon qui s'écarte au plus de $\pm 0,50$ dollar de la moyenne de la population égale à 21,68 dollars ?
- b) Quelle est la probabilité qu'un échantillon de 50 femmes diplômées fournisse une moyenne d'échantillon qui s'écarte au plus de $\pm 0,50$ dollar de la moyenne de la population égale à 18,80 dollars ?
- c) Dans lequel des deux cas précédents (a) ou (b), avons-nous la probabilité la plus élevée d'obtenir une estimation de la moyenne qui s'écarte au plus de $\pm 0,50$ dollar de la moyenne de la population ? Pourquoi ?
- d) Quelle est la probabilité qu'un échantillon aléatoire simple de 120 femmes diplômées fournisse une moyenne d'échantillon inférieure de plus de 0,30 dollar par rapport à la moyenne de la population ?
28. Les précipitations annuelles moyennes sont de 22 pouces en Californie et de 42 pouces dans l'État de New York (site Internet de Current Results, 27 octobre 2012). Supposez que l'écart type pour les deux États soit de 4 pouces. Un échantillon de 30 années de précipitations pour la Californie et un échantillon de 45 années de précipitations pour New York ont été sélectionnés.

- a) Déterminer la distribution de probabilité de la moyenne d'échantillon des précipitations annuelles pour la Californie.
 - b) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 1 pouce de la moyenne de la population pour la Californie ?
 - c) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 1 pouce de la moyenne de la population pour New York ?
 - d) Dans quel cas, (b) ou (c), la probabilité d'obtenir une moyenne d'échantillon s'écartant au plus de ± 1 pouce de la moyenne de la population est-elle la plus élevée ? Pourquoi ?
29. Les frais de préparation moyens que H&R Block a fait payer à ses clients l'année dernière s'élevaient à 183 dollars (*The Wall Street Journal*, 7 mars 2012). Utilisez ce prix comme la moyenne de la population et supposez que l'écart type de la population des frais de préparation soit de 50 dollars.
- a) Quelle est la probabilité que le prix moyen pour un échantillon de 30 clients de H&R Block s'écarte au plus de ± 8 dollars de la moyenne de la population ?
 - b) Quelle est la probabilité que le prix moyen pour un échantillon de 50 clients de H&R Block s'écarte au plus de ± 8 dollars de la moyenne de la population ?
 - c) Quelle est la probabilité que le prix moyen pour un échantillon de 100 clients de H&R Block s'écarte au plus de ± 8 dollars de la moyenne de la population ?
 - d) Recommanderiez-vous d'utiliser un échantillon de taille égale à 30, 50 ou 100 pour avoir une probabilité de 0,95 que la moyenne d'échantillon s'écarte au plus de ± 8 dollars de la moyenne de la population ?
30. Pour estimer l'âge moyen d'une population de 4 000 employés, un échantillon aléatoire simple de 40 employés est sélectionné.
- a) Utiliseriez-vous le facteur de correction pour population finie pour calculer l'erreur type de la moyenne ? Expliquer.
 - b) Si l'écart type de la population est $\sigma = 8,2$ ans, calculer l'erreur type avec et sans le facteur de correction pour population finie. Quel est le raisonnement pour expliquer l'abandon du facteur de correction pour population finie lorsque $n/N \leq 0,05$?
 - c) Quelle est la probabilité que l'âge moyen des employés de l'échantillon s'écarte au plus de ± 2 ans de l'âge moyen de la population ?

7.6 DISTRIBUTION D'ÉCHANTILLONNAGE DE \bar{p}

La proportion d'échantillon \bar{p} est l'estimateur ponctuel de la proportion de la population p . La formule de calcul de la proportion d'échantillon est

$$\bar{p} = \frac{x}{n}$$

où x est le nombre d'éléments dans l'échantillon qui possèdent la caractéristique à laquelle on s'intéresse et n est la taille de l'échantillon.

Comme noté dans la section 7.4, la proportion d'échantillon \bar{p} est une variable aléatoire et sa distribution de probabilité est appelée distribution d'échantillonnage de \bar{p} .

► **Distribution d'échantillonnage de \bar{p}**

La distribution d'échantillonnage de \bar{p} correspond à la distribution de probabilité de toutes les valeurs possibles de la proportion d'échantillon \bar{p} .

Pour déterminer l'écart entre la proportion d'échantillon \bar{p} et la proportion de la population p , il est nécessaire de connaître les propriétés de la distribution d'échantillonnage de \bar{p} : l'espérance mathématique de \bar{p} , l'écart type de \bar{p} et la forme de la distribution d'échantillonnage de \bar{p} .

7.6.1 Espérance mathématique de \bar{p}

L'espérance mathématique de \bar{p} , la moyenne de toutes les valeurs possibles de \bar{p} , est égale à la proportion de la population p .

► **Espérance mathématique de \bar{p}**

$$E(\bar{p}) = p \quad (7.4)$$

où

$E(\bar{p})$ correspond à l'espérance mathématique de \bar{p}
 p correspond à la proportion de la population

Puisque $E(\bar{p}) = p$, \bar{p} est un estimateur sans biais de p . Rappelons que dans la section 7.1, nous avons noté que $p = 0,60$ pour la population de la société EAI, où p correspond à la proportion de la population des employés qui ont suivi le programme de formation au management, dispensé par la société. Ainsi, l'espérance mathématique de \bar{p} dans le cadre du problème de la société EAI est égale à 0,60.

7.6.2 Écart type de \bar{p}

Comme nous l'avons montré pour l'écart type de \bar{x} , l'écart type de \bar{p} dépend du caractère fini ou infini de la population. Les deux formules de calcul de l'écart type de \bar{p} suivent.

► **Écart type de \bar{p}**

$$\begin{array}{cc} \text{Population finie} & \text{Population infinie} \\ \sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p(1-p)}{n}} & \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \end{array} \quad (7.5)$$

En comparant les deux formules de l'équation (7.5), on voit que la seule différence est l'utilisation d'un facteur de correction pour population finie $\sqrt{(N-n)/(N-1)}$.

Comme dans le cas de la moyenne d'échantillon \bar{x} , la différence entre les expressions pour population finie et infinie devient négligeable lorsque la taille de la population finie est importante comparativement à la taille de l'échantillon. Nous suivons la même règle pratique que celle recommandée dans le cas de la moyenne d'échantillon. C'est-à-dire, si la population est finie avec $n/N \leq 0,05$, nous utiliserons $\sigma_{\bar{p}} = \sqrt{p(1-p)/n}$.

Cependant, si la population est finie avec $n/N > 0,05$, le facteur de correction pour population finie devra être utilisé. De nouveau, sauf mention contraire, à travers l'ouvrage nous supposons que la taille de la population est importante comparativement à la taille de l'échantillon et donc que le facteur de correction pour population finie est inutile.

Dans la section 7.5, nous avons utilisé le terme d'*erreur type de la moyenne* pour faire référence à l'écart type de \bar{x} . En général, le terme d'erreur type est employé pour désigner l'écart type d'un estimateur ponctuel. Ainsi, pour la proportion, nous utilisons le terme d'*erreur type de la proportion* pour désigner l'écart type de \bar{p} . Revenons à présent à l'exemple de la société EAI et calculons l'erreur type de la proportion associée aux échantillons aléatoires simples de 30 employés de la société EAI.

Pour l'étude du problème de la société EAI, nous savons que la proportion de la population des employés qui ont suivi le programme de formation au management est $p = 0,60$. Avec $n/N = 30/2500 = 0,012$, nous pouvons ignorer le facteur de correction pour population finie pour calculer l'erreur type de la proportion. Pour l'échantillon aléatoire simple de 30 employés, $\sigma_{\bar{p}}$ est égal à

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,60(1-0,60)}{30}} = 0,0894$$

7.6.3 La forme de la distribution d'échantillonnage de \bar{p}

Maintenant que nous connaissons la moyenne et l'écart type de la distribution d'échantillonnage de \bar{p} , déterminons la forme de la distribution d'échantillonnage de \bar{p} . La

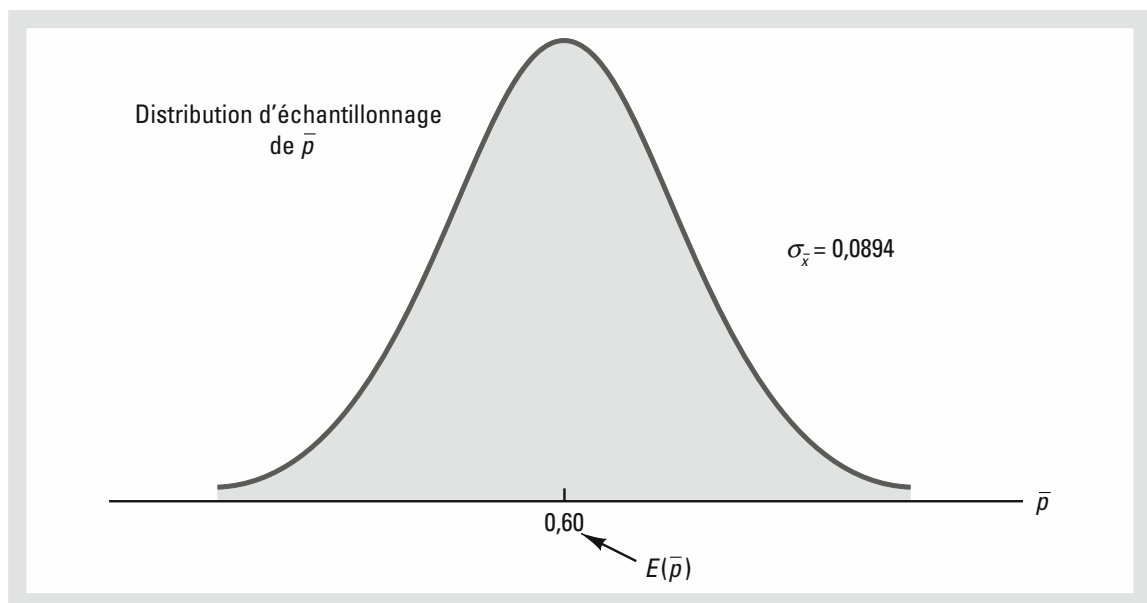


Figure 7.8 Distribution d'échantillonnage de \bar{p} pour la proportion des employés de la société EAI qui ont suivi le programme de formation au management

proportion d'échantillon est $\bar{p} = x/n$. Pour un échantillon aléatoire simple issu d'une population de grande taille, la valeur de x est une variable aléatoire binomiale, indiquant le nombre d'éléments dans l'échantillon possédant la caractéristique à laquelle on s'intéresse. Puisque n est constant, la probabilité de x/n est la même que la probabilité binomiale de x , ce qui signifie que la distribution d'échantillonnage de \bar{p} est également une distribution de probabilité discrète et que la probabilité de chaque valeur x/n est la même que la probabilité binomiale de x .

Dans le chapitre 6, nous avons également montré qu'une distribution binomiale peut être approchée par une distribution normale si la taille de l'échantillon est suffisamment grande pour satisfaire les deux conditions suivantes :

$$np \geq 5 \quad \text{et} \quad n(1-p) \geq 5$$

Supposant que ces deux conditions sont satisfaites, la distribution de probabilité du nombre d'éléments dans l'échantillon possédant la caractéristique à laquelle on s'intéresse, peut être approchée par une distribution normale. Et puisque n est constant, la distribution d'échantillonnage de $\bar{p} = x/n$ peut aussi être approchée par une distribution normale. Cette approximation est établie ci-dessous :

- ▶ La distribution d'échantillonnage de \bar{p} peut être approchée par une distribution normale lorsque $np \geq 5$ et $n(1-p) \geq 5$.

Dans des applications pratiques, lorsqu'on désire estimer une proportion d'échantillon, on cherche les tailles d'échantillon qui sont presque toujours assez grandes pour permettre l'utilisation d'une approximation normale de la distribution d'échantillonnage de \bar{p} .

Rappelons que dans le cadre du problème de la société EAI, nous savons que la proportion de la population des employés qui ont suivi le programme de formation est $p = 0,60$. Avec un échantillon aléatoire simple de taille 30, nous avons $np = 30(0,60) = 18$ et $n(1-p) = 30(0,40) = 12$. Ainsi, la distribution d'échantillonnage de \bar{p} peut être approchée par une distribution de probabilité normale, comme indiqué à la figure 7.8.

7.6.4 Intérêt pratique de la distribution d'échantillonnage de \bar{p}

L'intérêt pratique de la distribution d'échantillonnage de \bar{p} est qu'elle peut fournir des informations probabilistes concernant l'écart entre la proportion d'échantillon et la proportion de la population. Supposez, dans le cadre du problème de la société EAI, que le directeur du personnel veuille connaître la probabilité d'obtenir une valeur de \bar{p} qui s'écarte d'au plus 0,05, en valeur absolue, de la proportion de la population des employés de la société EAI qui ont suivi le programme de formation. En d'autres termes, quelle est la probabilité d'obtenir un échantillon dont la proportion \bar{p} sera comprise entre 0,55 et 0,65 ? L'aire grisée de la figure 7.9 correspond à cette probabilité. En utilisant le fait que la distribution d'échantillonnage de \bar{p} puisse être approchée par une distribution de probabilité normale de moyenne égale à 0,60 et d'erreur

type égale à $\sigma_{\bar{p}} = 0,0894$, la variable aléatoire normale centrée réduite correspondant à $\bar{p} = 0,65$ a une valeur égale à $z = (0,65 - 0,60)/0,0894 = 0,56$. En se référant à la table des probabilités normales centrées réduites, nous voyons que la probabilité cumulée correspondant à $z = 0,56$ est égale à 0,7123. De même, au point $\bar{p} = 0,55$, nous trouvons $z = (0,55 - 0,60)/0,0894 = -0,56$. D'après la table des probabilités normales centrées réduites, la probabilité cumulée correspondant à $z = -0,56$ est égale à 0,2877. Ainsi, la probabilité de sélectionner un échantillon qui fournisse une proportion d'échantillon \bar{p} qui s'écarte d'au plus 0,05, en valeur absolue, de la proportion de la population p est égale à $0,7123 - 0,2877 = 0,4246$.

Si l'on considère un échantillon de taille $n = 100$, l'erreur type de la proportion devient

$$\sigma_{\bar{p}} = \sqrt{\frac{0,60(1 - 0,60)}{100}} = 0,049$$

Avec un échantillon de 100 employés de la société EAI, la probabilité d'obtenir une proportion d'échantillon qui s'écarte d'au plus 0,05, en valeur absolue, de la proportion de la population peut maintenant être calculée. Puisque la distribution d'échantillonnage est approximativement normale, de moyenne égale à 0,60 et d'écart type égal à 0,049, nous pouvons utiliser la table des probabilités normales centrées réduites pour trouver la probabilité cherchée. Au point $\bar{p} = 0,65$, nous avons $z = (0,65 - 0,60)/0,049 = 1,02$. En se référant à la table des probabilités normales centrées réduites, la probabilité cumulée correspondant à $z = 1,02$ est égale à 0,8461. De même, au point $\bar{p} = 0,55$, nous avons $z = (0,55 - 0,60)/0,049 = -1,02$. La probabilité cumulée correspondant à $z = -1,02$ est égale à 0,1539. Ainsi, si la taille de l'échantillon augmente de 30 à 100, la probabilité que la proportion d'échantillon \bar{p} s'écarte d'au plus 0,05, en valeur absolue, de la proportion de la population p passe à 0,6922 ($0,8461 - 0,1539 = 0,6922$).

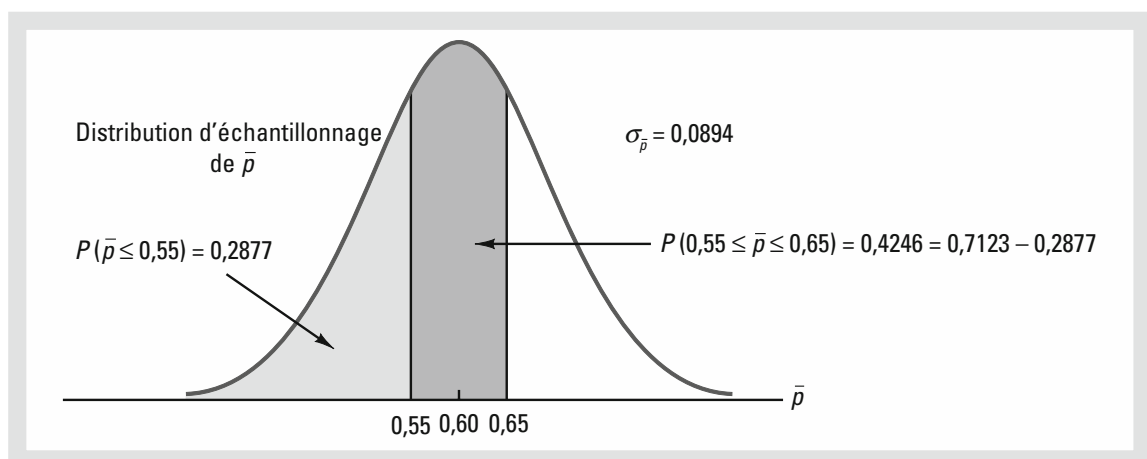


Figure 7.9 Probabilité d'obtenir \bar{p} entre 0,55 et 0,65

EXERCICES

Méthode

31. Un échantillon aléatoire simple de taille 100 est sélectionné à partir d'une population caractérisée par $p = 0,40$.

- a) Quelle est l'espérance mathématique de \bar{p} ?
- b) Quel est l'erreur type de \bar{p} ?
- c) Déterminer la distribution d'échantillonnage de \bar{p} .
- d) Que montre la distribution d'échantillonnage de \bar{p} ?



32. La proportion d'une population est égale à 0,40. Un échantillon aléatoire simple de taille 200 est sélectionné et la proportion d'échantillonnage \bar{p} sera utilisée pour estimer la proportion de la population.

- a) Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,03$ de la proportion de la population ?
- b) Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,05$ de la proportion de la population ?

33. Supposez que la proportion d'une population soit égale à 0,55. Calculer l'erreur type de la proportion, $\sigma_{\bar{p}}$, pour des échantillons de taille 100, 200, 500 et 1 000. Que pouvez-vous dire concernant l'ampleur de l'erreur type de la proportion lorsque la taille de l'échantillon augmente ?

34. La proportion de la population est de 0,30. Quelle est la probabilité que la proportion d'un échantillon s'écarte au plus de $\pm 0,04$ de la proportion de la population pour chacune des tailles d'échantillon suivantes ?

- a) $n = 100$
- b) $n = 200$
- c) $n = 500$
- d) $n = 1\,000$
- e) Quel est l'avantage d'avoir une taille d'échantillon importante ?

Applications



35. Le président de la société Doerman Distributors estime que 30 % des commandes de l'entreprise proviennent de nouveaux clients. Un échantillon aléatoire simple de 100 commandes est utilisé pour estimer la proportion de nouveaux clients.

- a) Supposez que le président ait raison et que $p = 0,30$. Quelle est la distribution d'échantillonnage de \bar{p} dans cette étude ?
- b) Quelle est la probabilité que la proportion d'échantillon \bar{p} soit comprise entre 0,20 et 0,40 ?
- c) Quelle est la probabilité que la proportion d'échantillon soit comprise entre 0,25 et 0,35 ?

- 36.** *The Wall Street Journal* a rapporté que 55 % des entrepreneurs avaient au plus 29 ans lorsqu'ils ont fondé leur première start-up et 45 % avaient au moins 30 ans (*The Wall Street Journal*, 19 mars 2012).
- Supposez qu'un échantillon de 200 entrepreneurs soit sélectionné pour en savoir davantage sur les qualités les plus importantes d'un entrepreneur. Déterminer la distribution d'échantillonnage de la proportion d'échantillon \bar{p} correspondant à la proportion d'entrepreneurs qui ont fondé leur première start-up au plus tard à 29 ans.
 - Quelle est la probabilité que la proportion d'échantillon de la question (a) s'écarte d'au plus $\pm 0,05$ de la proportion de la population ?
 - Supposez qu'un échantillon de 200 entrepreneurs soit sélectionné pour en savoir davantage sur les qualités les plus importantes d'un entrepreneur. Déterminer la distribution d'échantillonnage de la proportion d'échantillon \bar{p} correspondant à la proportion d'entrepreneurs qui ont fondé leur première start-up à 30 ans ou plus.
 - Quelle est la probabilité que la proportion d'échantillon de la question (c) s'écarte d'au plus $\pm 0,05$ de la proportion de la population ?
 - La probabilité obtenue aux questions (b) et (d) est-elle différente ? Pourquoi ?
 - Répondre à la question (b) pour un échantillon de taille égale à 400. La probabilité est-elle inférieure ? Pourquoi ?
- 37.** Les gens finissent par jeter 12 % de ce qu'ils achètent chez l'épicier (*Reader's Digest*, mars 2009). Supposez qu'il s'agit de la vraie proportion de la population et que vous envisagez de constituer un échantillon de 540 consommateurs pour étudier davantage leur comportement.
- Déterminer la distribution d'échantillonnage de \bar{p} , la proportion de biens d'épicerie jetés par les clients échantillonnés.
 - Quelle est la probabilité que votre étude fournisse une proportion d'échantillon qui s'écarte au plus de $\pm 0,03$ de la proportion de la population ?
 - Quelle est la probabilité que votre étude fournisse une proportion d'échantillon qui s'écarte au plus de $\pm 0,015$ de la proportion de la population ?
- 38.** Quarante-deux pourcents des médecins pensent que leur patients reçoivent des soins médicaux inutiles (*Reader's Digest*, décembre 2011/janvier 2012).
- Supposez qu'un échantillon de 300 médecins soit sélectionné. Déterminer la distribution d'échantillonnage de la proportion de médecins qui pensent que leurs clients ont reçu des soins médicaux inutiles.
 - Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,03$ de la proportion de la population ?
 - Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,05$ de la proportion de la population ?
 - Quel est l'impact de prendre un échantillon plus large sur les probabilités des questions (b) et (c) ? Pourquoi ?
- 39.** En 2008, le bureau Better Business a traité 75 % des plaintes reçues (*USA Today*, 2 mars 2009). Supposez que vous êtes embauché par le bureau Better Business pour étudier les plaintes reçues relatives à des concessionnaires automobiles. Vous envisagez de sélectionner

un échantillon des plaintes impliquant des concessionnaires automobiles pour estimer la proportion de plaintes que le bureau Better Business est en mesure de traiter. Supposez que la proportion de plaintes traitées dans la population, impliquant des concessionnaires automobiles, est égale à 0,75, identique à la proportion globale de plaintes traitées en 2008.

- a) Supposez que vous sélectionnez un échantillon de 450 plaintes impliquant des concessionnaires automobiles. Déterminer la distribution d'échantillonnage de \bar{p} .
 - b) En vous basant sur un échantillon de 450 plaintes, quelle est la probabilité que la proportion de l'échantillon s'écarte au plus de $\pm 0,04$ de la proportion de la population ?
 - c) Supposez que vous sélectionnez un échantillon de 200 plaintes impliquant des concessionnaires automobiles. Déterminer la distribution d'échantillonnage de \bar{p} .
 - d) En vous basant sur un échantillon de 200 plaintes, quelle est la probabilité que la proportion de l'échantillon s'écarte au plus de $\pm 0,04$ de la proportion de la population ?
 - e) En termes de probabilité, combien gagnez-vous en précision en utilisant un échantillon plus grand ?
40. Les producteurs de biens d'épicerie américains ont indiqué que 76 % des consommateurs lisent les étiquettes indiquant la composition des produits. Supposez que la proportion de la population soit $p = 0,76$ et qu'un échantillon de 400 consommateurs soit issu de cette population.
- a) Déterminer la distribution d'échantillonnage de la proportion d'échantillon \bar{p} correspondant à la proportion des consommateurs de l'échantillon qui lisent l'étiquette de composition des produits.
 - b) Quelle est la probabilité que la proportion d'échantillon s'écarte d'au plus $\pm 0,03$ de la proportion de la population ?
 - c) Répondre à la question (b) pour un échantillon de 750 clients.
41. L'institut de marketing alimentaire révèle que 17 % des ménages dépensent plus de 100 dollars par semaine en épicerie. Supposez que la proportion de la population soit $p = 0,17$ et qu'un échantillon aléatoire simple de 800 ménages soit sélectionné parmi cette population.
- a) Déterminer la distribution d'échantillonnage de \bar{p} , la proportion des ménages de l'échantillon qui dépensent plus de 100 dollars par semaine en épicerie.
 - b) Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,02$ de la proportion de la population ?
 - c) Répondre à la question (b) pour un échantillon de 1 600 ménages.

7.7 AUTRES MÉTHODES D'ÉCHANTILLONNAGE

Nous avons décrit la procédure d'échantillonnage aléatoire simple comme une procédure d'échantillonnage à partir d'une population finie et discuté des propriétés des distributions d'échantillonnage de \bar{x} et de \bar{p} , lorsqu'on utilise un échantillon aléatoire simple. Des méthodes telles que l'échantillonnage aléatoire stratifié, l'échantillonnage par grappes et l'échantillonnage systématique sont des méthodes d'échantillonnage alternatives qui présentent, dans certaines situations, des avantages par rapport à l'échantillonnage aléatoire

simple. Dans cette section, nous introduirons brièvement ces méthodes alternatives d'échantillonnage.

Cette section fournit une brève introduction aux méthodes d'échantillonnage autres que l'échantillonnage aléatoire simple.

7.7.1 Échantillonnage aléatoire stratifié

Dans l'échantillonnage aléatoire stratifié, la population est tout d'abord divisée en groupes d'éléments appelés *strates*, de façon à ce que chaque élément de la population appartienne à une et une seule strate. L'élément de base qui définit une strate, tel qu'un service, un lieu, un âge, un type d'industrie, etc., est laissé à la discrétion du créateur de l'échantillon. Cependant, de meilleurs résultats sont obtenus lorsque les éléments de chaque strate sont aussi semblables que possible. La figure 7.10 représente un diagramme d'une population divisée en H strates.

Après avoir formé les strates, un échantillon aléatoire simple est sélectionné dans chaque strate. Des formules permettent de combiner les résultats obtenus à partir des échantillons individuels en une estimation du paramètre de la population auquel on s'intéresse. La valeur de l'échantillonnage aléatoire stratifié dépend de l'homogénéité des éléments contenus dans une strate. Si les éléments contenus dans une strate sont semblables (homogénéité), la strate aura une faible variance. Ainsi, des échantillons relativement petits pourront être utilisés pour obtenir de bonnes estimations des caractéristiques de la strate. Si les strates sont homogènes, la procédure d'échantillonnage aléatoire stratifié fournira des résultats aussi précis que ceux obtenus par la procédure d'échantillonnage aléatoire simple en utilisant un échantillon total plus petit.

L'échantillonnage aléatoire stratifié fournit de meilleurs résultats lorsque la variance parmi les éléments de chaque strate est relativement faible.

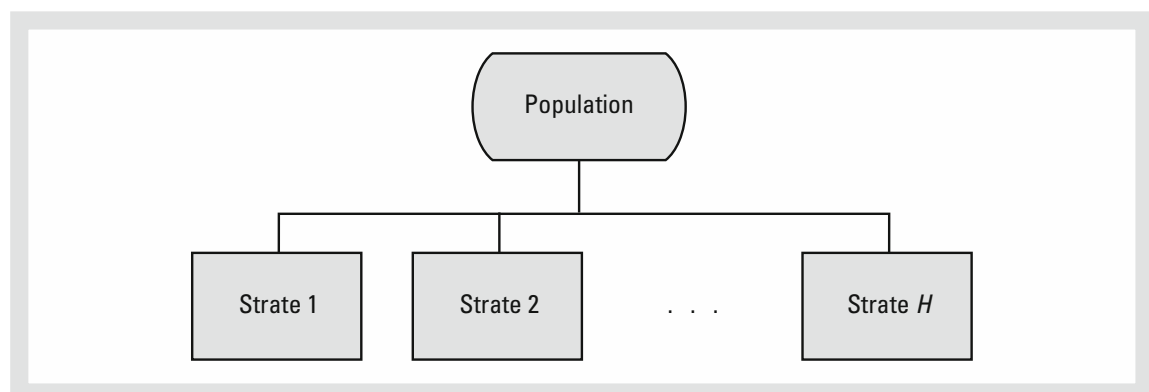


Figure 7.10 Diagramme pour l'échantillonnage aléatoire stratifié

7.7.2 Échantillonnage par grappes

Dans l'échantillonnage par grappes, la population est tout d'abord divisée en groupes d'éléments séparés, appelés *grappes*. Chaque élément de la population appartient à une et une seule grappe (cf. figure 7.11). Un échantillon aléatoire simple des grappes est ensuite sélectionné. Tous les éléments contenus dans une grappe sélectionnée forment l'échantillon. L'échantillonnage par grappes tend à fournir de meilleurs résultats lorsque les éléments contenus dans les grappes sont hétérogènes (dissemblables). Dans le cas idéal, chaque grappe est une représentation à petite échelle de la population entière. La valeur de l'échantillonnage par grappes dépend du degré de représentativité de la population entière dans chaque grappe. Si toutes les grappes représentent la population, échantillonner un petit nombre de grappes fournira de bonnes estimations des paramètres de la population.

L'échantillonnage par grappes fournit de meilleurs résultats lorsque chaque grappe représente, à plus petite échelle, la population.

L'une des applications principales de l'échantillonnage par grappes est l'échantillonnage de régions, où les grappes sont les quartiers d'une ville ou d'autres zones bien définies. L'échantillonnage par grappes nécessite généralement un échantillon total plus grand que l'échantillonnage aléatoire simple ou stratifié. Cependant, il peut générer des économies de coût, du fait que lorsqu'une personne sonde une grappe sélectionnée (par exemple, un quartier), beaucoup d'observations peuvent être obtenues en un temps relativement court. Par conséquent, un échantillon de taille plus importante peut être obtenu avec un coût total significativement plus faible.

7.7.3 Échantillonnage systématique

Dans certaines situations, spécialement lorsque les populations sont importantes, il est coûteux (en temps) de sélectionner un échantillon aléatoire simple en trouvant tout d'abord un nombre aléatoire et ensuite en cherchant dans la liste de la population l'élément

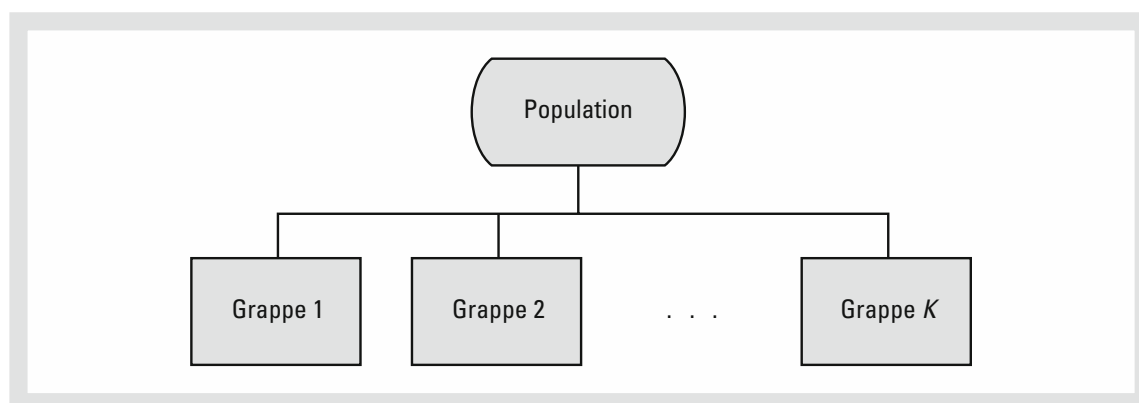


Figure 7.11 Diagramme pour l'échantillonnage par grappes

correspondant. Une alternative à l'échantillonnage aléatoire simple est l'**échantillonnage systématique**. Par exemple, si l'on souhaite sélectionner un échantillon de taille 50 parmi une population contenant 5 000 éléments, cela revient à sélectionner un élément tous les $5000/50 = 100$ éléments de la population. Constituer un échantillon systématique dans ce cas consiste à sélectionner aléatoirement un élément parmi les 100 premiers de la liste de la population. Les autres éléments de l'échantillon sont identifiés de la façon suivante : le deuxième élément sélectionné correspond au 100^e élément qui suit le premier élément sélectionné dans la liste de la population ; le troisième élément sélectionné correspond au 100^e élément qui suit dans la liste de la population le deuxième élément sélectionné, et ainsi de suite. En fait, l'échantillon de taille 50 est identifié en se déplaçant systématiquement dans la population et en identifiant le 100^e, le 200^e, le 300^e, etc. élément qui suivent le premier élément choisi aléatoirement. L'échantillon de taille 50 est généralement plus facile à identifier de cette manière qu'en utilisant l'échantillonnage aléatoire simple. Puisque le premier élément sélectionné l'est aléatoirement, un échantillon systématique est généralement supposé avoir les propriétés d'un échantillon aléatoire simple. Cette hypothèse est particulièrement appropriée lorsque la liste de la population est une énumération aléatoire des éléments de la population.

7.7.4 Échantillonnage de commodité

Les méthodes d'échantillonnage présentées jusqu'à présent sont dites techniques *d'échantillonnage probabiliste*. Les éléments sélectionnés parmi la population ont une probabilité connue de faire partie de l'échantillon. L'avantage de l'échantillonnage probabiliste est que la distribution d'échantillonnage de la statistique d'échantillon appropriée peut généralement être identifiée. Des formules comme celles présentées dans ce chapitre pour l'échantillonnage aléatoire simple, permettent de déterminer les propriétés de la distribution d'échantillonnage. Ensuite, la distribution d'échantillonnage permet de tirer des conclusions en termes de probabilité sur l'erreur d'échantillonnage associée aux résultats.

L'**échantillonnage de commodité** est une technique *d'échantillonnage non-probabiliste*. Comme son nom l'indique, l'échantillon est principalement identifié par commodité. Les éléments sont inclus dans l'échantillon sans probabilité connue ou pré-spécifiée d'être choisis. Par exemple, un professeur qui mène une expérience à l'université peut utiliser des étudiants volontaires pour constituer un échantillon simplement parce qu'ils sont disponibles et participeront en tant que sujets à l'expérience pour un coût très faible ou même nul. De même, un inspecteur peut échantillonner une cargaison d'oranges en sélectionnant les oranges au hasard parmi plusieurs caisses. Étiqueter chaque orange et utiliser une méthode probabiliste d'échantillonnage seraient irréalisables. Des échantillons tels que les animaux sauvages en captivité et les panels de consommateurs volontaires sont des échantillons de commodité.

Les échantillons de commodité ont l'avantage d'être facilement constitués et les données sont facilement collectées ; cependant, il est impossible d'évaluer le degré de représentativité de l'échantillon au regard de la population. Un échantillon de commodité peut fournir de bons résultats aussi bien que des mauvais ; aucune procédure statistique

ne permet de faire une analyse probabiliste ou de l'inférence sur la qualité des résultats de l'échantillon. Parfois, des chercheurs appliquent des méthodes statistiques conçues pour des échantillons probabilistes aux échantillons de commodité, déclarant que l'échantillon de commodité peut être traité comme un échantillon probabiliste. Cependant, cet argument ne peut être soutenu, et il faut rester prudent en interprétant les résultats obtenus grâce à un échantillon de commodité, dans le but de faire de l'inférence sur les populations.

7.7.5 Échantillonnage subjectif

Une autre technique d'échantillonnage non-probabiliste est *l'échantillonnage subjectif*. Dans cette approche, la personne la mieux documentée sur le sujet de l'étude sélectionne des éléments de la population qu'elle pense être les plus représentatifs de la population. Souvent, cette méthode est une manière relativement facile de sélectionner un échantillon. Par exemple, un journaliste peut choisir deux ou trois sénateurs, en jugeant que l'opinion de ces sénateurs reflète l'opinion générale. Cependant, la qualité des résultats de l'échantillon dépend des croyances de la personne qui sélectionne l'échantillon. De nouveau, il faut faire très attention en tirant des conclusions concernant les populations, lorsqu'on se fonde sur des échantillons subjectifs.

REMARQUES

Nous recommandons l'utilisation des méthodes d'échantillonnage probabilistes lorsque l'on cherche à constituer des échantillons à partir de populations finies : l'échantillonnage aléatoire simple, l'échantillonnage aléatoire stratifié, l'échantillonnage par grappes ou l'échantillonnage systématique. Pour ces méthodes, des formules permettent d'évaluer la qualité des estimations des caractéristiques de la population, fournies par les résultats de l'échantillon. Une évaluation de la justesse des résultats ne peut pas être faite avec des échantillons de commodité ou des échantillons subjectifs. Aussi, une attention particulière doit-elle être portée à l'interprétation des résultats lorsque des méthodes d'échantillonnage non-probabilistes sont utilisées.

RÉSUMÉ

Dans ce chapitre, nous avons présenté les concepts d'échantillonnage et de distributions d'échantillonnage. Nous avons montré comment constituer un échantillon aléatoire simple à partir d'une population finie et discuté de la constitution d'un échantillon aléatoire à partir d'une population infinie. Les données collectées dans de tels échantillons peuvent être utilisées pour développer des estimations ponctuelles des paramètres de la population. Puisque différents échantillons aléatoires simples fournissent diverses valeurs des estimateurs ponctuels, les estimateurs ponctuels tels que \bar{x} et \bar{p} sont des variables aléatoires. La distribution de probabilité de telles variables aléatoires est appelée distribution d'échantillonnage. En particulier, nous avons décrit les distributions d'échantillonnage de la moyenne d'échantillon \bar{x} et la proportion d'échantillon \bar{p} .

En considérant les caractéristiques des distributions d'échantillonnage de \bar{x} et \bar{p} , nous avons établi que $E(\bar{x}) = \mu$ et $E(\bar{p}) = p$. Après avoir développé les formules de l'écart type ou erreur type de ces estimateurs, nous avons décrit les conditions nécessaires sous lesquelles les distributions d'échantillonnage de \bar{x} et \bar{p} suivent une loi normale. D'autres méthodes d'échantillonnage dont l'échantillonnage aléatoire stratifié, l'échantillonnage par grappes, l'échantillonnage systématique, l'échantillonnage de commodité et l'échantillonnage subjectif, ont été présentées.

GLOSSAIRE

POPULATION ÉCHANTILLONNÉE. La population à partir de laquelle l'échantillon est constitué.

CADRE. Une liste d'éléments à partir desquels l'échantillon est sélectionné.

PARAMÈTRE. Caractéristique numérique d'une population, telle que la moyenne de la population μ , l'écart type de la population σ , la proportion de la population p , etc.

ÉCHANTILLON ALÉATOIRE SIMPLE. Un échantillon aléatoire simple de taille n issu d'une population finie de taille N est un échantillon sélectionné de façon à ce que chaque échantillon possible de taille n ait la même probabilité d'être choisi.

ÉCHANTILLONNAGE SANS REMISE. Une fois qu'un élément a été inclus dans l'échantillon, il est retiré de la population et ne peut pas être choisi une seconde fois.

ÉCHANTILLONNAGE AVEC REMISE. Une fois qu'un élément a été inclus dans l'échantillon, il est remis dans la population. Un élément déjà sélectionné peut de nouveau être choisi et donc peut apparaître plus d'une fois dans l'échantillon.

ÉCHANTILLON ALÉATOIRE. Un échantillon aléatoire issu d'une population infinie est un échantillon sélectionné de telle façon que les deux conditions suivantes sont satisfaites : (1) chaque élément sélectionné est issu de la même population ; (2) chaque élément est sélectionné indépendamment des autres.

STATISTIQUE D'ÉCHANTILLON. Caractéristique d'échantillon, telle que la moyenne d'échantillon \bar{x} ,

l'écart type d'échantillon s , la proportion d'échantillon \bar{p} , etc. La valeur de la statistique d'échantillon est utilisée pour estimer la valeur du paramètre de la population.

ESTIMATEUR PONCTUEL. Statistique d'échantillon, telle que \bar{x} , s ou \bar{p} , qui fournit l'estimation ponctuelle d'un paramètre de la population.

ESTIMATION PONCTUELLE. Valeur d'un estimateur ponctuel utilisée en tant qu'estimation d'un paramètre de la population.

POPULATION CIBLE. Population pour laquelle est faite de l'inférence statistique telle que des estimations ponctuelles. Il est important que la population cible soit aussi proche que possible de la population échantillonnée.

DISTRIBUTION D'ÉCHANTILLONNAGE. Distribution de probabilité composée de toutes les valeurs possibles d'une statistique d'échantillon.

SANS BIAIS. Propriété d'un estimateur ponctuel caractérisée par l'égalité entre l'espérance mathématique de l'estimateur ponctuel et la valeur du paramètre de la population qu'il estime.

FACTEUR DE CORRECTION POUR POPULATION FINIE. Terme $\sqrt{(N-n)/(N-1)}$ utilisé dans les formules de $\sigma_{\bar{x}}$ et de $\sigma_{\bar{p}}$ lorsqu'une population finie, et non infinie, est échantillonnée. La règle pratique généralement acceptée est d'ignorer le facteur de correction pour population finie lorsque $n/N \leq 0,05$.

ERREUR TYPE. Écart type d'un estimateur ponctuel.

THÉORÈME CENTRAL LIMITE. Théorème qui permet d'utiliser la distribution de probabilité normale pour estimer la distribution d'échantillonnage de \bar{x} lorsque l'échantillon est de grande taille.

ÉCHANTILLONNAGE ALÉATOIRE STRATIFIÉ. Méthode d'échantillonnage probabiliste dans laquelle la population est tout d'abord divisée en strates et un échantillon aléatoire simple est ensuite sélectionné parmi chaque strate.

ÉCHANTILLONNAGE PAR GRAPPES. Méthode d'échantillonnage probabiliste dans laquelle la population est tout d'abord divisée en grappes et un échantillon aléatoire simple de grappes est ensuite sélectionné.

ÉCHANTILLONNAGE SYSTÉMATIQUE. Méthode d'échantillonnage probabiliste dans laquelle on choisit aléatoirement un des k premiers éléments, puis tous les k° éléments qui suivent.

ÉCHANTILLONNAGE DE COMMODITÉ. Méthode d'échantillonnage non-probabiliste dans laquelle les éléments de l'échantillon sont sélectionnés en fonction de leur commodité.

ÉCHANTILLONNAGE SUBJECTIF. Méthode d'échantillonnage non-probabiliste dans laquelle les éléments de l'échantillon sont sélectionnés en fonction des croyances de la personne qui fait l'étude.

FORMULES CLÉ

Espérance mathématique de \bar{x}

$$E(\bar{x}) = \mu \quad (7.1)$$

Écart type de \bar{x} (erreur type)

Population finie	Population infinie
$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \left(\frac{\sigma}{\sqrt{n}} \right)$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (7.2)$

Espérance mathématique de \bar{p}

$$E(\bar{p}) = p \quad (7.4)$$

Écart type de \bar{p} (erreur type)

Population finie	Population infinie
$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p(1-p)}{n}}$	$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \quad (7.5)$

EXERCICES SUPPLÉMENTAIRES

42. *U.S. News & World Report* publie des informations sur les meilleures écoles américaines (*America's Best Colleges*, 2009). Entre autre, le rapport fournit une liste des 133 meilleures universités du pays. Vous souhaitez sélectionner un échantillon de ces universités pour une

- étude sur les étudiants. Commencez par le bas de la troisième colonne des nombres aléatoires du tableau 7.1. En ignorant les deux premiers chiffres des groupes de nombres à cinq chiffres et en utilisant les nombres aléatoires à trois chiffres commençant par 959, remontez dans la colonne pour identifier le numéro (compris entre 1 et 133) des sept premières universités qui seront incluses dans un échantillon aléatoire simple. Continuez en commençant en bas de la quatrième puis de la cinquième colonne, en remontant si nécessaire.
- 43.** Les dernières données disponibles indiquent que les dépenses de santé s'élevaient à 8 086 dollars par personne aux États-Unis, soit 17,6 % du produit intérieur brut (PIB) (site Internet des Centres de services Medicare & Medicaid, 1^{er} avril 2012). Utilisez 8 086 dollars comme la moyenne de la population et supposez qu'une entreprise de conseil sélectionne un échantillon de 100 personnes pour déterminer la nature de leurs dépenses de santé. Supposez que l'écart type de la population est égal à 2 500 dollars.
- Déterminer la distribution d'échantillonnage du montant moyen des dépenses de santé pour un échantillon de 100 personnes.
 - Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 200 dollars de la moyenne de la population ?
 - Quelle est la probabilité que la moyenne d'échantillon soit supérieure à 9 000 dollars ? Si le consultant vous dit que la moyenne d'échantillon est supérieure à 9 000 dollars, vous demanderiez-vous s'il a correctement suivi la procédure d'échantillonnage ? Pourquoi ?
- 44.** Foot Locker utilise les ventes par mètre carré pour mesurer la productivité de ses magasins. Les ventes annuelles sont actuellement de l'ordre de 406 dollars par mètre carré (*The Wall Street Journal*, 7 mars 2012). La direction vous a demandé de mener une étude sur un échantillon de 64 magasins Foot Locker. Supposez que l'écart type des ventes annuelles par mètre carré pour la population des 3 400 magasins Foot Locker soit égal à 80 dollars.
- Déterminer la distribution d'échantillonnage de \bar{x} correspondant à la moyenne d'échantillon des ventes annuelles par mètre carré pour un échantillon de 64 magasins Foot Locker.
 - Quelle est la probabilité que la moyenne de l'échantillon s'écarte au plus de ± 15 dollars de la moyenne de la population ?
 - Supposez que vous trouviez une moyenne d'échantillon égale à 380 dollars. Quelle est la probabilité de trouver une moyenne d'échantillon inférieure ou égale à 380 dollars ? Considérez-vous cet échantillon comme un groupe inhabituellement sous-performant de magasins ?
- 45.** Allegiant Airlines pratique un tarif de base moyen de 89 dollars. En plus, la compagnie tarifie la réservation sur son site Internet, l'enregistrement des bagages et les boissons consommées en vol. Ces frais supplémentaires coûtent en moyenne 39 dollars par passager (*Bloomberg Businessweek*, 8-14 octobre 2012). Supposez qu'un échantillon aléatoire de 60 passagers soit sélectionné pour déterminer le coût total de leur vol avec la compagnie Allegiant Airlines. L'écart type de la population du coût total des vols est égal à 40 dollars.
- Quel est le coût moyen d'un vol au niveau de la population ?

- b) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 10 dollars du coût moyen d'un vol au niveau de la population ?
- c) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de ± 5 dollars du coût moyen d'un vol au niveau de la population ?
- 46.** Déduction faite des bourses accordées sous condition de ressources, le coût moyen d'inscription à l'Université de Californie du Sud (USC) est de 27 175 dollars (*U.S. News & World Report, America's Best Colleges, 2009*). Supposez que l'écart type de la population s'élève à 7 400 dollars. Supposez qu'un échantillon aléatoire de 60 étudiants soit issu de cette population.
- a) Quelle est la valeur de l'erreur type de la moyenne ?
- b) Quelle est la probabilité que la moyenne d'échantillon soit supérieure à 27 175 dollars ?
- c) Quelle est la probabilité que la moyenne d'échantillon s'écarte au plus de $\pm 1 000$ dollars de la moyenne de la population ?
- d) Quelle serait la probabilité de la question (c) si la taille d'échantillon était égale à 100 ?
- 47.** Trois entreprises ont des inventaires différents par leur taille. L'entreprise A a une population de 2 000 pièces, l'entreprise B a une population de 5 000 pièces et l'entreprise C a une population de 10 000 pièces. L'écart type de la population pour le coût des pièces est $\sigma = 144$. Un consultant recommande que chaque entreprise prenne un échantillon de 50 pièces parmi sa population pour fournir des estimations statistiques valables sur le coût moyen par pièce. Les responsables de la petite entreprise pensent pouvoir obtenir les données à partir d'un échantillon plus petit que celui nécessaire aux grandes entreprises, du fait de sa plus petite population. Cependant, selon le consultant, pour obtenir la même erreur type et donc la même précision dans les résultats de l'échantillon, toutes les entreprises doivent utiliser un échantillon de même taille, quelle que soit la taille de la population.
- a) En utilisant le facteur de correction pour population finie, calculer l'erreur type pour chacune des trois entreprises, étant donné un échantillon de taille 50.
- b) Quelle est la probabilité que pour chaque entreprise, la moyenne d'échantillon \bar{x} s'écarte au plus de ± 25 de la moyenne de la population μ ?
- 48.** Un chercheur rapporte les résultats d'une étude en révélant que l'erreur type de la moyenne est de 20. L'écart type de la population est égal à 500.
- a) Quelle est la taille de l'échantillon utilisé dans cette étude ?
- b) Quelle est la probabilité que l'estimation s'écarte au plus de ± 25 de la moyenne de la population ?
- 49.** Un processus de production est vérifié périodiquement par un inspecteur du contrôle de la qualité. L'inspecteur sélectionne des échantillons aléatoires simples de 30 produits finis et calcule la moyenne d'échantillon des poids des produits \bar{x} . Si les résultats de test sur une longue période révèlent que 5 % des valeurs de \bar{x} sont supérieures à 2,1 livres et que 5 % sont inférieures à 1,9 livre, quels sont la moyenne et l'écart type pour la population des produits fabriqués avec ce procédé ?

50. Quinze pourcent des Australiens fument. En introduisant des lois rigoureuses interdisant de faire apparaître la marque sur les paquets de cigarette, l'Australie espère réduire le pourcentage de la population qui fume de 10 % d'ici 2018 (site Internet de Reuters, 23 octobre 2012). Répondre aux questions suivantes basées sur un échantillon de 240 Australiens.
- Déterminer la distribution d'échantillonnage de \bar{p} , la proportion d'échantillon des Australiens qui fument.
 - Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,04$ de la proportion de la population ?
 - Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,02$ de la proportion de la population ?
51. Une société d'études de marché effectue des sondages par téléphone, avec historiquement un taux de réponse de 40 %. Quelle est la probabilité que dans un nouvel échantillon de 400 numéros de téléphone, au moins 150 individus coopèrent et répondent aux questions ? En d'autres termes, quelle est la probabilité que la proportion d'échantillon soit au moins égale à $150/400 = 0,375$?
52. Les annonceurs publicitaires concluent des contrats avec les fournisseurs d'accès à Internet et les moteurs de recherche pour placer leur publicité sur les sites web. Ils paient une taxe forfaitaire basée sur le nombre de clients potentiels qui s'intéresseront à leur publicité. Malheureusement, la fraude – le fait de cliquer sur une publicité uniquement pour accroître les revenus publicitaires – est devenue un réel problème. Quarante pourcents des annonceurs prétendent avoir été victimes de fraude (*Business Week*, 13 mars 2006). Supposez qu'un échantillon aléatoire simple de 380 annonceurs soit constitué pour déterminer plus précisément l'impact de cette pratique sur les annonceurs.
- Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,04$ de la proportion de la population des annonceurs victimes de fraude ?
 - Quelle est la probabilité que la proportion d'échantillon soit supérieure à 0,45 ?
53. La proportion d'individus assurés par la compagnie d'assurance automobile All-Driver, qui ont reçu au moins une contravention au cours des cinq dernières années, est de 0,15.
- Déterminer la distribution d'échantillonnage de \bar{p} , si un échantillon aléatoire de 150 assurés est utilisé pour estimer la proportion d'individus ayant reçu au moins une contravention.
 - Quelle est la probabilité que la proportion d'échantillon s'écarte au plus de $\pm 0,03$ de la proportion de la population ?
54. Lori Jeffrey est l'une des meilleures représentantes commerciales d'un important éditeur de manuels scolaires. Historiquement, Lori décroche une vente sur 25 % de ses appels. En considérant ses ventes par téléphone pendant un mois comme un échantillon de toutes les ventes par téléphone possibles, supposez qu'une étude statistique des données fournisse une erreur type de la proportion de 0,0625.
- Quelle est la taille de l'échantillon utilisé dans cette étude ? C'est-à-dire, combien d'appels Lori a-t-elle passé au cours du mois considéré ?
 - Soit \bar{p} la proportion des ventes effectuées au cours du mois. Déterminer la distribution d'échantillonnage de \bar{p} .

Tableau 7.6 Évaluation des 10 premières métropoles

Métropole	Évaluation	Métropole	Évaluation
Albany	64,18	Baltimore	69,75
Albuquerque	66,16	Birmingham	69,59
Appleton	60,56	Boise City	68,36
Atlanta	69,97	Boston	68,99
Austin	71,48	Buffalo	66,10



- c) En utilisant la distribution d'échantillonnage de \bar{p} , calculer la probabilité que Lori décroche des ventes sur 30 % ou plus de ses appels au cours d'un mois.

ANNEXE 7.1 ÉCHANTILLONNAGE ALÉATOIRE AVEC MINITAB

Si une liste des éléments d'une population est disponible dans un fichier Minitab, Minitab peut être utilisé pour sélectionner un échantillon aléatoire simple. Par exemple, une liste des 100 plus importantes métropoles américaines et canadiennes est fournie dans la colonne 1 du fichier *Métropoles* (*Places Rated Almanac* – Édition du millénaire 2000). La colonne 2 contient l'évaluation de chaque métropole. Les 10 premières métropoles et leurs évaluations sont présentées dans le tableau 7.6.

Supposez que vous vouliez sélectionner un échantillon aléatoire simple de 30 métropoles pour réaliser une étude approfondie du coût de la vie aux États-Unis et au Canada. Les étapes suivantes permettent de sélectionner l'échantillon.

- Étape 1.** Sélectionner le menu **Calc**
- Étape 2.** Choisir **Random Data**
- Étape 3.** Choisir **Sample From Columns**
- Étape 4.** Lorsque la boîte de dialogue apparaît :
 - Entrer 30 dans la boîte **Number of rows to sample**
 - Entrer C1 C2 dans la boîte **From columns**
 - Entrer C3 C4 dans la boîte **Store samples in**
- Étape 5.** Cliquer sur **OK**

L'échantillon aléatoire de 30 métropoles apparaît dans les colonnes C3 et C4.

ANNEXE 7.2 ÉCHANTILLONNAGE ALÉATOIRE AVEC EXCEL

Si une liste des éléments d'une population est disponible dans un fichier Excel, Excel peut être utilisé pour sélectionner un échantillon aléatoire simple. Par exemple, une liste des 100 plus importantes métropoles américaines et canadiennes est fournie dans la colonne A du fichier *Métropoles (Places Rated Almanac – Édition du millénaire 2000)*. La colonne B contient l'évaluation de chaque métropole. Les 10 premières métropoles et leurs évaluations sont présentées dans le tableau 7.6. Supposez que vous vouliez sélectionner un échantillon aléatoire simple de 30 métropoles pour réaliser une étude approfondie du coût de la vie aux États-Unis et au Canada.

Les lignes d'un fichier Excel peuvent être placées dans un ordre aléatoire en ajoutant une colonne supplémentaire au fichier et en remplissant cette colonne par des nombres aléatoires en utilisant la fonction =RAND(). Ensuite en réarrangeant la colonne des nombres aléatoires par ordre croissant, le fichier est réordonné de façon aléatoire. L'échantillon aléatoire de taille n correspond alors aux n premières lignes de ce fichier réordonné.

Pour le fichier *Métropoles*, la première ligne contient l'intitulé des colonnes et les 100 métropoles sont inscrites dans les lignes 2 à 101. Les étapes suivantes permettent de sélectionner un échantillon aléatoire simple de 30 métropoles.

- Étape 1.** Entrer =RAND() dans la cellule C2
- Étape 2.** Copier la cellule C2 dans les cellules C3:C101
- Étape 3.** Sélectionner une cellule de la colonne C
- Étape 4.** Cliquer sur le bouton **Home** dans la barre des tâches
- Étape 5.** Dans le groupe **Editing**, cliquer sur **Sort & Filter**
- Étape 6.** Cliquer sur **Sort Smallest to Largest**

L'échantillon aléatoire de 30 métropoles apparaît dans les lignes 2 à 31 du fichier réordonné. Les nombres aléatoires figurant dans la colonne C ne sont plus nécessaires et peuvent être effacés.

ANNEXE 7.3 ÉCHANTILLONNAGE ALÉATOIRE AVEC STATTOOLS

Si une liste des éléments d'une population est disponible dans un fichier Excel, StatTools Random Sample Utility peut être utilisé pour sélectionner un échantillon aléatoire simple. Par exemple, une liste des 100 plus importantes métropoles américaines et canadiennes est fournie dans la colonne A du fichier *Métropoles (Places Rated Almanac – Édition du millénaire 2000)*. La colonne B contient l'évaluation de chaque métropole. Supposez que vous vouliez sélectionner un échantillon aléatoire simple de 30 métropoles pour réaliser une étude approfondie du coût de la vie aux États-Unis et au Canada.

Commencez par utiliser Data Set Manager pour créer un ensemble de données StatTools pour ces données en utilisant la procédure décrite dans l'annexe du chapitre 1. Les étapes suivantes permettent de créer un échantillon aléatoire simple de 30 métropoles.

- Étape 1.** Cliquer sur **StatTools** dans la barre des tâches
- Étape 2.** Dans **Data Group** cliquer sur **Data Utilities**
- Étape 3.** Choisir l'option **Random Sample**
- Étape 4.** Lorsque la boîte de dialogue apparaît :
 - Dans la section **Variables**
 - Sélectionner **Métropoles**
 - Sélectionner **Rating**
 - Dans la section **Options**
 - Entrer 1 dans la boîte **Number of Samples**
 - Entrer 30 dans la boîte **Sample Size**
 - Cliquer sur **OK**

L'échantillon aléatoire de 30 métropoles apparaîtra dans les colonnes A et B d'une feuille de calcul intitulée Échantillon aléatoire.