

Formation et développement IPUMS: la demande de données (Exercice 2 pour SAS)



Objectifs

- Créer et télécharger un extrait de données IPUMS PMA
- Décompresser le fichier de données et lire les données dans SAS
- Analyser les données en utilisant un exemple de code

Résumé

Comprendre la structure des ensembles de données de points de livraison de services IPUMS PMA et la manière dont ils peuvent être exploités pour explorer vos intérêts de recherche. Cet exercice utilisera ensemble de données IPUMS PMA pour explorer la relation entre les caractéristiques des points de livraison de services et en rupture de stock de contraceptives. Vous allez créer un extrait de données avec les variables EAID, FACILITYTYPEGEN, FACILITYADV, PILLOSBS, PILLOUTDAY, et URBAN; puis vous allez utiliser le code d'exemple à analyser les données.

Code SAS à réviser

Code	Procédure
<u>PROC FREQ;</u>	Générer une procédure de fréquence
<u>PROC MEANS;</u>	Générer une procédure de moyen, rendre le moyen d'une variable, peut être utilisé pour regrouper les données
<u>tables</u>	Syntaxe requise pour afficher les résultats en format tableau
<u>where</u>	Sélectionner certains cas à inclure dans la procédure
<u>weight;</u>	Fournir la pondération de chaque observation dans l'ensemble de données saisies
<u>data newdata;</u> <u>set olddata;</u>	Utiliser le « data step » pour renommer ou classer les données selon opérateurs conditionnels, commandes prémisses-conclusion, ou transformation.
PROC UNIVARIATE	A utiliser afin de créer histogrammes
PROC GCHART	A utiliser avec « hbar » afin de créer histogrammes un graphique à barres horizontale
PROC SQL	A utiliser afin de faire sortir la valeur de la variable d'intérêt
merge	A utiliser dans le « data step » pour fusionner plusieurs ensembles de données

Erreurs courantes à éviter

- Pas complètement décompresser les données.
- Ne pas changer le répertoire de travail où vos données sont stockées
- Oublier de lancer la procédure avec « run »
- Oublier de terminer la commande avec point-virgule « ; »



Enregistrement avec IPUMS

- Allez sur <http://pma.ipums.org>, cliquez sur « Register » pour utiliser IPUMS PMA sur le côté gauche de l'écran. Cliquez sur le bouton « Register for IPUMS PMA » et remplissez le formulaire pour demander l'accès. Vous devrez attendre que votre compte soit approuvé pour accéder aux données. Une fois que vous avez reçu l'email d'approbation, cliquez sur « Connect » en haut de la page et utilisez votre email et votre mot de passe.
- Aller à « Browse and Select Data »
- Choisissez l'unité d'analyse « Person »

CHOOSE THE UNIT OF ANALYSIS FOR DATA BROWSING	
PERSON	EACH RECORD WILL BE A PERSON DESCRIPTION
SERVICE DELIVERY POINT	EACH RECORD WILL BE A SERVICE DELIVERY POINT DESCRIPTION

- Cliquez sur la case « Select Samples », cochez la case pour Kenya 2016 R5.

☐ Kenya ☒ 2016 R5 ☐ 2015b R4 ☐ 2014b R2
☐ 2015a R3 ☐ 2014a R1

- Faites défiler vers le bas de la page et cliquez sur l'option du bouton radio pour « All Cases (Respondents and Non-respondents to Service Delivery Point Questionnaires) ». La valeur par défaut est Facility Respondents.
- Cliquez sur la case « Submit Sample Selections. »

Sample Members

- ☐ Facility Respondents
☒ All Cases (Respondents and Non-respondents to Service Delivery Point Questionnaires)



À l'aide du menu déroulant ou de la fonction de recherche, sélectionnez les variables suivantes:

FACILITYTYPEGEN: Type d'établissement

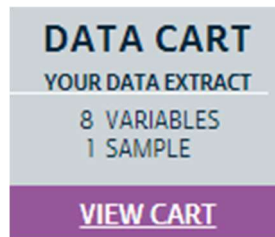
FACILITYADV: Etablissement avancé

PILLOBS: pilules contraceptives en stock ou en rupture de stock

PILLOUTDAY: Nombre de jours que les pilules contraceptives ont été en rupture de stock

URBAN: statut urbain ou rural

- Cliquez sur le bouton violet VIEW CART sous votre panier de données.



- Vérifier la sélection des variables. Notez que certaines variables apparaissent dans votre panier même si vous ne les avez sélectionnés pas. Ils ne sont pas inclus dans le compte de variables dans votre panier de données. Ces variables présélectionnées sont nécessaires pour la pondération, l'estimation de la variance, ou pour identifier l'année, le pays, et le round d'un échantillon.
- Cliquez sur le bouton « Create Data Extract ».
- Examinez l'écran « Extract Request Summary », décrivez votre extrait et cliquez sur « Submit Extract »
- Vous recevrez un email lorsque les données seront disponibles pour le téléchargement.
- Pour accéder à la page pour télécharger les données, suivez le lien dans l'e-mail ou le lien « My Data Extracts » sur la page d'accueil.



Lancer les données dans votre logiciel de statistiques

- Allez sur <http://pma.ipums.org> et cliquez sur « My Data Extracts »

Extract Number	Date	Formatted Data	Fixed-width Text Files					Codebook	
			Data	Command Files					
51	2018-10-26	--	Download .DAT	SPS	SAS	STATA	R	Basic	DDI

- Faites un clic droit sur le lien de données à côté de l'extrait que vous avez créé.
- Choisissez « Save Target As ... » (ou « Save Link As ... »)
- Enregistrer dans « Documents » (qui devrait apparaître comme localisation par défaut)
- Faites la même opération pour le lien SAS à côté de l'extrait
- Trouver le dossier "Documents" dans le menu de démarrage.
- Faites un clic droit sur le fichier « .dta ».
- Dans la fenêtre qui s'ouvre, appuyez sur le bouton « Extract » pour l'extraire ici.
- Vérifier que le fichier « pma_000... » se trouve dans le dossier "Documents"
- Un logiciel de décompression gratuit est disponible à l'adresse <http://www.iris.net/soft/wingzip/>
- Ouvrir le fichier « pma_000##.sas »
- Dans la fenêtre du fichier « pma_000##.sas » changer la première ligne de « libname IPUMS '.' » à « libname IPUMS "C:\Users\Documents...;" » en remplaçant le répertoire correspondant au fichier où vos données sont sauvegardées.
- Après « filename ASCII DAT », saisir l'emplacement du fichier complet, en terminant en « pma_000##.dat; »
- Choisir « Submit » sous l'option de menu Run.



Analyser l'échantillon

Partie 1: Explorer les types d'établissements de santé

1. Créez un tableau de fréquences pour FACILITYTYPEGEN indiquant la proportion de chaque type d'établissement enquêtée au Kenya, série 5.

```
PROC FREQ;  
  
tables facilitytypegen;  
  
run;
```

2. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers étudié pour FACILITYTYPEGEN? _____
3. Les utilisateurs doivent noter que de nombreuses variables de l'enquête sur les points de livraison de services (PLS) ont un univers défini par FACILITYADV, une désignation spécifique au pays de type « d'établissements avancés ». Créez un tableau croisé pour voir quels types d'établissements de la question précédente ont été désignés « établissements avancés » au Kenya pour 2016.

```
PROC FREQ;  
  
tables facilitytypegen*facilityadv / norow nocol nopercent;  
  
run;
```

4. Consultez l'onglet « Comparability » pour FACILITYADV, en prenant soin de noter que les désignations d'établissements avancés varient d'un pays à l'autre et varient parfois d'une enquête à l'autre dans un pays. Recherchez l'entrée Kenya et déterminez si sa désignation d'établissement avancé correspond à ce que vous avez trouvé à la question 3. La désignation est-elle concordante pour toutes les enquêtes Kényanes comprenant cette variable ? _____



Partie 2: Statistiques descriptives

5. Considérons la variable PILLOBS, qui indique si le PLS disposait d'un stock de pilules contraceptives observable le jour de l'entretien. Selon l'onglet « Codes », quelles sont les réponses possibles pour les PLS enquêtés au Kenya 2016? _____

6. Selon l'onglet « Comparability », les réponses possibles à PILLOBS peuvent varier d'un échantillon à l'autre. Comment? _____
7. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers de l'enquête pour PILLOBS? _____

8. Parmi les établissements qui fournissent généralement les pilules contraceptives indiquées dans PILLOBS, quel type d'établissement était le moins susceptible d'avoir des réserves de pilules contraceptives en stock le jour de l'entretien ? Quelle proportion d'établissements de ce type était en rupture de stock ? (Limitez l'analyse aux seules interviews terminées et aux cas dans l'univers). _____

```
PROC FREQ;  
tables facilitytypegen*pillobs / nocol nopercnt;  
where pillobs <90;  
run;
```

Partie 3: Visualisation des données

9. Pour les établissements qui étaient en rupture de stock de pilules contraceptives, PILLOUTDAY indique le nombre de jours pendant lesquels les réserves étaient indisponibles. Pour certains PLS étant en rupture de stock depuis plus de 90 jours, les codes NIU et les valeurs manquantes pour PILLOUTDAY sont codés sous les valeurs 9994, 9997 et 9999 afin de dépasser la gamme de réponses valides.



Calculez la pénurie moyenne de pilules pour tous les établissements dans l'univers de PILLOUTDAY, en prenant soin d'exclure toute valeur supérieure à 9 000. Recherchez ensuite la moyenne de chaque type d'établissement dans FACILITYTYPEGEN et affichez le résultat sous forme d'histogramme (limitez l'analyse aux réponses valides des PLS dans l'univers pour PILLOUTDAY).

```
PROC MEANS data=IPUMS.pma_000##;  
var pilloutday;  
where pilloutday <9000;  
run;  
  
PROC MEANS data=IPUMS.pma_000##;  
var pilloutday;  
class facilitytypegen;  
where pilloutday <9000;  
run;  
  
PROC GCHART data=IPUMS.pma_000##;  
title1 "Mean days out of stock by facility type";  
hbar facilitytypegen / discrete type=mean sumvar=pilloutday  
nostats;  
where pilloutday <9000;  
run;
```

10. Supposons que vous soupçonnez que la différence apparente entre les établissements dans 9 est en réalité une disparité entre les types d'établissements les plus susceptibles de se trouver dans les zones urbaines par rapport aux zones rurales. Créez une paire de graphiques à barres groupé par URBAN pour vérifier si cela est vrai. Existe-t-il des différences entre les établissements urbaines et rurales de chaque type? _____
- _____




```
PROC GCHART data=IPUMS.pma_000##;

title1 "Mean days out of stock by facility type and urban/rural
status";

hbar facilitytypegen / discrete type=mean sumvar=pilloutday
nostats group=urban;

where pilloutday <9000;

run;
```

Partie 4: Combinaison de données PLS et HHF

Les utilisateurs doivent noter que PMA2020 a enquêté des établissements situés dans les mêmes zones d'échantillonnage que les ménages et les femmes lors de la même enquête. Ces données PLS ne sont pas censées être représentatives au niveau national. Au lieu de cela, ils sont destinés à décrire l'environnement de prestation de soins de santé des ménages et des femmes interrogées. Ainsi, il n'y a pas de poids d'échantillonnage pour les variables PLS.

Les fichiers contiennent un poids pour les unités d'échantillonnage EAWEIGHT, qui est un poids de probabilité représentant la possibilité que le secteur de dénombrement (EA) soit sélectionné pour l'échantillonnage. Les collecteurs des données originales ne recommandent pas d'utiliser EAWEIGHT pour pondérer les variables PLS. La meilleure utilisation des variables PLS consiste plutôt de calculer des statistiques récapitulatives au niveau des SD et à les associer au jeu de données Ménage and Femme (HHF) à l'aide de la variable EAID en tant que source d'informations contextuelles pour l'environnement de prestation de services de chaque femme.

Par exemple, vous pouvez utiliser les variables PILLOBS et PILLOUTDAY pour déterminer si un établissement dans chaque EAID est en rupture de stock de pilules contraceptives et le nombre moyen de jours pendant lesquels les établissements de chaque EAID ont été en rupture de stock. Ces statistiques récapitulatives peuvent être



combinées avec l'ensemble de données HHF afin de montrer si chaque femme répondante avait un accès local fiable aux pilules contraceptives.

11. Générez un tableau avec le nombre des femmes âgées de 15 à 49 ans (ÉLIGIBLE == 1) de l'échantillon Kenya 2016 de la série 5 HHF qui résidaient dans chacun des secteurs de dénombrement où les pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS? Combien de secteurs de dénombrement de Kenya 2016 satisfont à ces critères? _____

```
data IPUMS.pma_000##;  
  set IPUMS.pma_000##;  
    if pilloutday >9000 then pilloutday= .;  
    if pillobs =3 then pillobs = 1;  
    else pillobs =0;  
run;  
PROC MEANS data=IPUMS.pma_000## nway;  
class eaid;  
var pillobs pilloutday;  
output out=IPUMS.pma_000##_1 max(pillobs) = pillobs  
mean(pilloutday) = pilloutday;  
run;  
PROC SQL;  
select count(*) as N_obs  
from IPUMS.pma_000##_1  
where pillobs=1;  
quit;
```

12. Combinez PILLOBS et PILLOUTDAY avec l'ensemble de données Ménage et Femme (HHF) en faisant correspondre les EAID de tous les répondants. Combien des femmes âgées de 15 à 49 ans (ÉLIGIBLE == 1) de l'échantillon Kenya 2016 de la série 5 HHF résidaient dans chacun des secteurs de dénombrement où les



pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS? (Notez que vous devrez remplacer ## par le numéro de votre extrait IPUMS PMA utilisé dans l'exercice 1).

```
PROC SORT data=IPUMS.pma_000##_1;  
by eaid;  
run;  
  
PROC SORT data=IPUMS.pma_000##;  
by eaid;  
run;  
  
data IPUMS.HHF_PLS;  
merge IPUMS.pma_000##_1 IPUMS.pma_000##;  
by eaid;  
run;  
  
PROC FREQ;  
tables eaid;  
where pillob=1 & eligible=1;  
run;
```

13. Effectuez une régression logistique pour estimer le lien entre les femmes qui prennent actuellement la pilule (FPNOWUSPILL) et la durée moyenne de la pénurie (PILLOUTDAY) pour chaque secteur de dénombrement qui était en absence de pilules le jour de l'interview PLS. Ajustez votre modèle pour qu'il soit représentatif des femmes kényanes utilisant FQWEIGHT. Considérez l'effet du nombre moyen de jours en rupture de stock comme une variable continue et un prédicteur de l'effet recherché. Recodez les valeurs pour FPNOWUSPILL qui ne sont pas dans l'univers ou qui manquent à zéro.



Existe-t-il un lien entre le nombre de jours pendant lesquels les établissements du secteur de dénombrement de la femme sont en rupture de stock de pilules et l'utilisation actuelle de la pilule par la femme pour la planification familiale?

```
data IPUMS.HHF_PLS;
set IPUMS.HHF_PLS;
if fpnwuspill >90 then fpnwuspill= 0;
if pilloutday =. then pilloutday = 0;
run;

PROC LOGISTIC data=IPUMS.HHF_PLS;
class fpnwuspill / param=ref;
model fpnwuspill = pilloutday;
weight fqweight;
run;
```



RÉPONSES - Analyse de l'échantillon

Partie 1: Explorer les types d'établissements de santé

1. Créez un tableau de fréquences pour FACILITYTYPEGEN indiquant la proportion de chaque type d'établissement enquêtée au Kenya, série 5.

Hospital: $79/428 = 18.46\%$

Dispensary: $190/428 = 44.39\%$

Health Center: $90/428 = 21.03\%$

Pharmacy / Chemist / Drug Shop: $48/428 = 11.21\%$

Health Clinic: $16/428 = 3.74\%$

Other: $4/428 = 0.93\%$

Other Health Facility: $1/428 = 0.23\%$

2. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers étudié pour FACILITYTYPEGEN? Tous les points de livraison de services
3. Les utilisateurs doivent noter que de nombreuses variables de l'enquête sur les points de livraison de services (PLS) ont un univers défini par FACILITYADV, une désignation spécifique au pays de types « établissements avancés ». Créez un tableau croisé pour voir quels types d'établissements de la question précédente ont été désignées comme « établissements avancés » au Kenya pour 2016. Toutes sont avancées, à l'exception des pharmacies.
4. Consultez l'onglet « Comparability » pour FACILITYADV, en prenant soin de noter que les désignations d'établissements avancés varient d'un pays à l'autre et varient parfois d'une enquête à l'autre dans un pays. Recherchez l'entrée Kenya et déterminez si sa désignation d'établissement avancé correspond à ce que vous avez trouvé à la question C. La désignation est-elle concordante pour toutes les enquêtes Kényanes comprenant cette variable? Ils correspondent, et toutes les enquêtes kényanes interrogés ont la même désignation.



Partie 2: Statistiques descriptives

5. Considérons la variable PILLOBS, qui indique si le PLS disposait d'un stock de pilules contraceptives observable le jour de l'entretien. Selon l'onglet « Codes », quelles sont les réponses possibles pour les PLS enquêtés au Kenya 2016?

1 - In-stock and observed

94 - Not interviewed (PLS questionnaire)

2 - In-stock but not observed

98 - No response or missing

3 - Out of stock

99 - NIU (not in universe)

6. Selon l'onglet « Comparability », les réponses possibles à PILLOBS peuvent varier d'un échantillon à l'autre. Comment? Certains échantillons initiaux incluent moins de détails et fournissent des réponses dichotomiques selon que l'enquêteur a observé la pilule contraceptive en stock. Dans ces échantillons initiaux, si les pilules contraceptives n'étaient pas observées, elles étaient supposées être "en rupture de stock". Lors d'enquêtes ultérieures, les enquêteurs avaient l'option de signaler que les pilules contraceptives étaient "en stock mais pas observées".
7. Selon l'onglet « Universe », quels établissements sont inclus dans l'univers de l'enquête pour PILLOBS? Services de livraison fournissant des pilules contraceptives.
8. Parmi les établissements qui fournissent généralement les pilules contraceptives indiquées dans PILLOBS, quel type d'établissement était le moins susceptible d'avoir des réserves de pilules contraceptives en stock le jour de l'entretien? Quelle proportion d'établissements de ce type était en rupture de stock? (Limitez l'analyse aux seules interviews terminées et aux cas dans l'univers). Les cliniques de santé étaient les plus susceptibles de manquer de médicaments avec 25% de rupture de stock.



Partie 3: Visualisation des données

9. Pour les établissements qui... Calculez la pénurie moyenne de pilules pour tous les établissements dans l'univers de PILLOUTDAY, en prenant soin d'exclure toute valeur supérieure à 9 000. Recherchez ensuite la moyenne de chaque type d'établissement dans FACILITYTYPEGEN et affichez le résultat sous forme d'histogramme (limitez l'analyse aux réponses valides des PLS dans l'univers pour PILLOUTDAY).

Toutes les installations: 87

Clinique de santé: 33

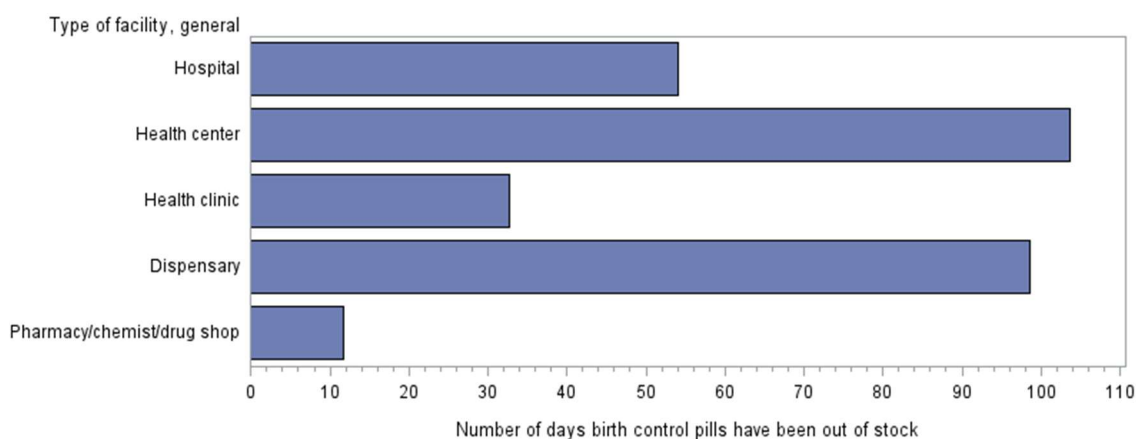
Hôpital: 54

Dispensaire: 99

Centre de santé: 104

Pharmacie: 12

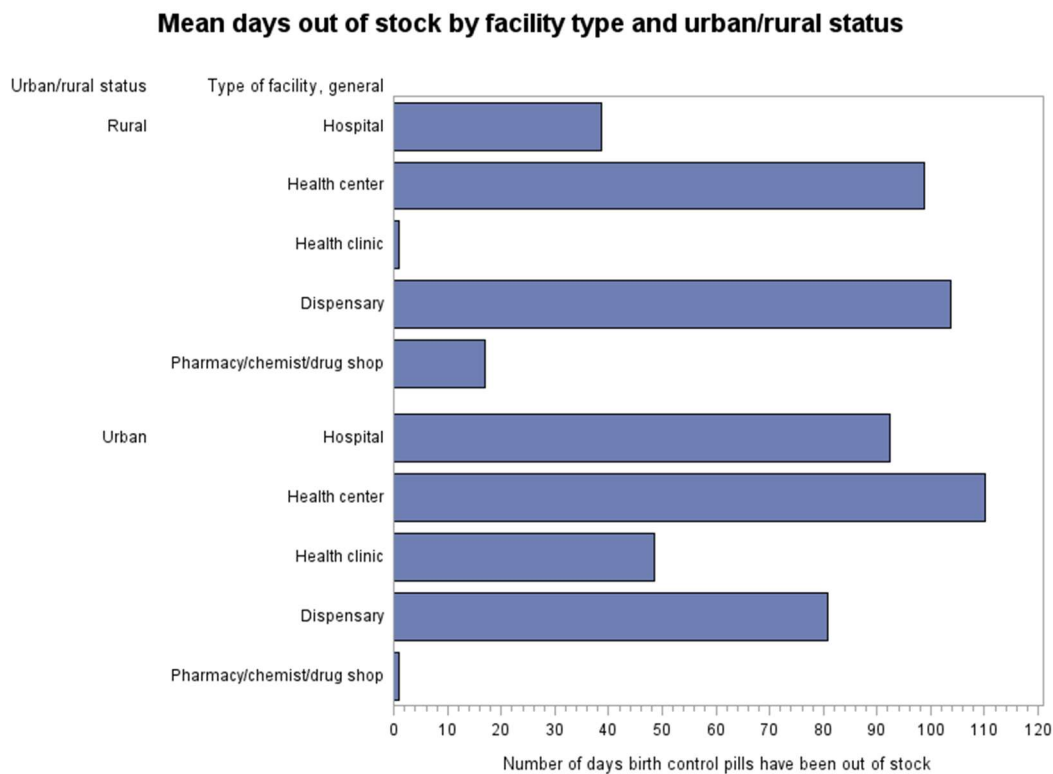
Mean days out of stock by facility type



10. Supposons que vous soupçonnez que la différence apparente entre les établissements dans 9) est en réalité une disparité entre les types d'établissements les plus susceptibles de se trouver dans les zones urbaines par rapport aux zones



rurales. Créez une paire de graphiques à barres groupé par URBAN pour vérifier si cela est vrai. Existe-t-il des différences entre les établissements urbaines et rurales de chaque type?



Partie 4: Combinaison de données PLS et HHF

11. Générez un tableau avec le nombre des femmes âgées de 15 à 49 ans (ÉLIGIBLE == 1) de l'échantillon Kenya 2016 de la série 5 HHF qui résidaient dans chacun des secteurs de dénombrement où les pilules contraceptives n'étaient pas disponibles dans tous les établissements locaux de l'enquête PLS? Combien de secteurs de dénombrement de Kenya 2016 satisfont à ces critères?

43 secteurs de dénombrement

12. Combinez PILLOBS et PILLOUTDAY avec l'ensemble... Effectuez une régression logistique pour estimer le lien entre les femmes qui prennent actuellement la pilule



(FPNOWUSPILL) et la durée moyenne de la pénurie (PILLOUTDAY) pour chaque secteur de dénombrement qui était en absence de pilules le jour de l'interview PLS. Ajustez votre modèle pour qu'il soit représentatif des femmes kényanes utilisant FQWEIGHT. Considérez l'effet du nombre moyen de jours en rupture de stock comme une variable continue et un prédicteur de l'effet recherché. Recodez les valeurs pour FPNOWUSPILL qui ne sont pas dans l'univers ou qui manquent à zéro.

Ce sont les trois premiers:

EAID 4013 = 33 femmes

EAID 4163 = 47 femmes

EAID 4047 = 44 femmes

13. Existe-t-il un lien entre le nombre de jours pendant lesquels les établissements du secteur de dénombrement de la femme sont en rupture de stock de pilules et l'utilisation actuelle de la pilule par la femme pour la planification familiale?

La probabilité qu'une femme de l'échantillon utilise des pilules contraceptives reste la même quel que soit le nombre moyen de jours pendant lesquels son PLS local n'a pas été disponible (rapport de cotes = 1,000). Cependant, ce résultat n'est pas statistiquement significatif (p = 0,728, IC à 95% entre 0,998 et 1,002).

FPNOWUSPILL	Rapport de cotes	Valeur p	IC à 95%
PILLOUTDAY	1,000	0,728	(0,998 et 1,002)

