MSDS 6370 Project

Team members:

Abbott, Andrew and Taylor, Celia

Date: August 1, 2017

What Proportion Does Medicare Pay for the Top 100 Diagnosis-Related Groups of Inpatient Charges?

**Introduction**

# The direct title of the data set is, “Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG) - FY2011”. This data set is located at [www.kaggle.com](http://www.kaggle.com) and at <https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3>. We chose our research question to be, “What is the average proportion Medicare pays for the Top 100 Diagnosis Related Groups (DRG) in FY2011?” This is a calculated value obtained from taking the parameters,

# AverageProportionPaid = AverageMedicarePayments/AverageTotPayments.

# The advantage of this research question is that it utilizes the entire population of the data set, which is 163065 observations, instead of just one or a few procedures or geographical areas.

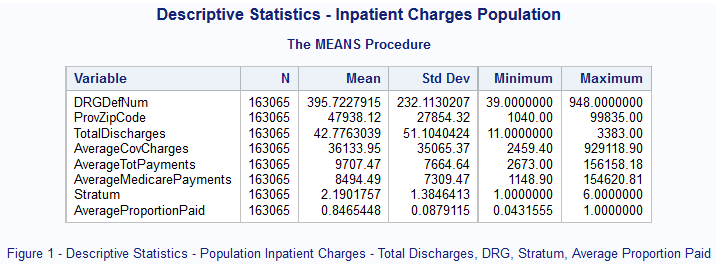
**Descriptive Statistics – Data Inspection and Handling**

Table 1 shows the descriptions of the parameters in this data set. Of primary interest are the DRG Definition, Total Discharges, Average Total Payments and Average Medicare Payments.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Type** |
| DRG Definition | Code and description identifying the DRG. DRGs are a classification system that groups similar clinical conditions (diagnoses) and the procedures furnished by the hospital during the stay. | Plain Text |
| Provider Id | Provider Identifier billing for inpatient hospital services. | Number |
| Provider Name | Name of the provider. | Plain Text |
| Provider Street Address | Street address in which the provider is physically located. | Plain Text |
| Provider City | City in which the provider is physically located. | Plain Text |
| Provider State | State in which the provider is physically located. | Plain Text |
| Provider Zip Code | Zip code in which the provider is physically located. | Number |
| Hospital Referral Region Description | HRR in which the provider is physically located. | Plain Text |
| Total Discharges | The number of discharges billed by the provider for inpatient hospital services. | Number |
| Average Covered Charges | The provider's average charge for services covered by Medicare for all discharges in the DRG. These will vary from hospital to hospital because of differences in hospital charge structures. | Money |
| Average Total Payments | The average of Medicare payments to the provider for the DRG including the DRG amount, teaching,  disproportionate share, capital, and outlier payments for all cases. Also included are co-payment and deductible amounts that the patient is responsible for | Money |
| Average Medicare Payments |  | Money |
|  | Table information directly from [**https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3**](https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3) |  |

Table 1: Data Set Parameter Descriptions

Figure 1 shows the numeric parameters that are primarily used in this analysis to get AverageProportionPaid.



Additional formatting and data cleaning was done to include another column, DRGDefNum from the DRG definition column. This column only contains the numeric part of the DRG definition column. The columns that were originally formatted as currency were converted to numeric with 2 digits to the right of the decimal point. The column headers were shortened so that they would not contain spaces. The AverageProportionPaid column was formatted to contain 9 places to the right of the decimal point.

**Strata Development and Allocation**

The strata are based on the Total Discharges column. We determined the total sample size by using the empirical method to estimate the population standard deviation. The strata allocation was found using both the proportional and Neyman allocation methods. To determine the total sample size the maximum and minimum values of the AverageProportionPaid, 1.00 for the maximum and 0.04 for the minimum, were used to find the range, 0.96, which was then used as an estimate for 6 standard deviations based on a normal distribution. For a normal distribution, 99.7% of the data is contained within 3 standard deviations of the mean or 6 standard deviations total. This method produced an estimate for the standard deviation of 0.1595. With a standard deviation, the sample size could then be calculated using the formula:

For , the margin of error, 0.01 was chosen.

The finite population correction was then applied:

Table 2 shows the allocation of the total sample size to strata based on the column Total Discharges according to the proportional method and the Neyman allocation method. The calculation steps for the proportional allocation are shown in Table 2. The calculation steps for the Neyman allocation are shown in Table 3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| stratum = h | Nh | Nh/N | nh\* | rounded | nh = Proportional Allocation Strata Size |
| 1 | 71475 | 0.44 | 425.6107994 | 426 | 426 |
| 2 | 37914 | 0.23 | 225.7657621 | 226 | 226 |
| 3 | 23779 | 0.15 | 141.5963511 | 142 | 142 |
| 4 | 15545 | 0.10 | 92.56551069 | 93 | 92 |
| 5 | 9791 | 0.06 | 58.30227823 | 58 | 58 |
| 6 | 4561 | 0.03 | 27.15929844 | 27 | 27 |
| Total | 163065 | 1 | 971 | 972 | 971 |

Table 2: Proportional Allocation

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| stratum = h | Nh | Sh | Nh\*Sh | (ShNh)/  ( Σfrom h=1toH (Nh\*Sh ) ) | nh\* = n (ShNh)/  ( Σfrom h=1toH (Nh\*Sh ) ) | rounded | nh= Neyman Allocation Strata Size |
| 1 | 71475 | 0.097 | 6957.07 | 0.49 | 474.43 | 474 | 475 |
| 2 | 37914 | 0.086 | 3254.16 | 0.23 | 221.92 | 222 | 222 |
| 3 | 23779 | 0.079 | 1879.26 | 0.13 | 128.15 | 128 | 128 |
| 4 | 15545 | 0.075 | 1160.22 | 0.08 | 79.12 | 79 | 79 |
| 5 | 9791 | 0.070 | 688.74 | 0.05 | 46.97 | 47 | 47 |
| 6 | 4561 | 0.066 | 299.26 | 0.02 | 20.41 | 20 | 20 |
| Total | 163065 |  | 14238.70 | 1 | 971 | 970 | 971 |

Table 3: Neyman Allocation

**Task One- Compare Stratified Method To Another Sampling Technique**

We executed Proc Surveyselect with seed = 91118 to obtain the proportional allocation sample (Figure 2). We also executed Proc Surveyselect with seed = 91118 to get the sample size and sample allocation for Neyman allocation (Figure 3).

|  |  |
| --- | --- |
|  |  |

For this project, after preparing the data, the first step is to run 4 exploratory executions of SAS code. These executions include a stratified proportional allocation (Figure 4) verses a stratified Neyman allocation (Figure 5) and a clustered stratified proportional allocation (Figure 6) verses a clustered Neyman allocation (Figure 7). Table 4 shows the composite results of these exploratory executions. For both the stratified and the cluster designs the proportional allocation performed marginally better.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistics Including Confidence Limits** | | | | | | | | | | |
| **Exploratory Executions** | | **Variable** | **Mean** | **Std Error of Mean** | **95% CL for Mean** | | **Sum** | **Std Error of Sum** | **95% CL for Sum** | |
| **Figure 4** | **Proportional** | **Average Proportion Paid** | **0.848400** | **0.002683** | **0.84313** | **0.85366** | **138344** | **437.47678** | **137485.775** | **139202.806** |
| **Figure 5** | **Neyman** | **Average ProportionPaid** | **0.850274** | **0.002715** | **0.84495** | **0.85560** | **138650** | **442.77684** | **137781.043** | **139518.876** |
| **Figure 6** | **Proportional Cluster** | **Average ProportionPaid** | **0.848400** | **0.004014** | **0.84050** | **0.85630** | **138344** | **4746.51675** | **129006.940** | **147681.642** |
| **Figure 7** | **Neyman Cluster** | **Average ProportionPaid** | **0.850274** | **0.004205** | **0.84200** | **0.85855** | **138650** | **4385.64248** | **130021.315** | **147278.604** |

Table 4: Composite Statistics for Exploratory Executions to Compare Proportional, Neyman, Proportional Cluster, and Neyman Cluster

**Task Two- Compare 5 Different Samples From Each Technique to Another**

The five different samples of Proportional Stratified are depicted in Figures 8-12. For each of the samples taken a different seed was used. For the five different examples of Proportional Clustered, depicted in Figures 13-17, the same five seeds were used as in the Proportional sampling. The means and 95% confidence intervals for each of the ten samples is shown in Table 5. The actual mean of 0.8465448 is within 100% of the 95% confidence intervals tested.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistics Including Confidence Limits** | | | | | | | | | | | |
| **Proportional Stratified Executions** | | **Variable** | **Mean** | **Std Error of Mean** | **95% CL for Mean** | | **Sum** | **Std Error of Sum** | **95% CL for Sum** | | **Includes True Mean?** |
| **Figure 8** | **Proportional Stratified Seed 91118** | **Average Proportion Paid** | **0.848400** | **0.00268** | **0.84313** | **0.85366** | **138344** | **437.47678** | **137485.775** | **139202.806** | **Yes** |
| **Figure 9** | **Proportional Stratified Seed 91119** | **Average Proportion Paid** | **0.847121** | **0.00278** | **0.84166** | **0.852583** | **138136** | **453.84793** | **137245.111** | **139026.397** | **Yes** |
| **Figure 10** | **Proportional Stratified Seed 91120** | **Average Proportion Paid** | **0.846693** | **0.00288** | **0.84105** | **0.85234** | **138066** | **468.94888** | **137145.784** | **138986.338** | **Yes** |
| **Figure 11** | **Proportional Stratified Seed 91121** | **Average Proportion Paid** | **0.845439** | **0.00292** | **0.83970** | **0.85118** | **137862** | **476.64707** | **136926.148** | **138796.917** | **Yes** |
| **Figure 12** | **Proportional Stratified Seed 91122** | **Average Proportion Paid** | **0.848771** | **0.00277** | **0.84333** | **0.85421** | **138405** | **451.97032** | **137517.829** | **139291.745** | **Yes** |
| **Figure 13** | **Proportional Clustered Seed 91118** | **Average Proportion Paid** | **0.848400** | **0.00268** | **0.84313** | **0.85366** | **138344** | **437.47678** | **137485.775** | **139202.806** | **Yes** |
| **Figure 14** | **Proportional Clustered Seed 91119** | **Average Proportion Paid** | **0.847121** | **0.00278** | **0.84166** | **0.852583** | **138136** | **453.84793** | **137245.111** | **139026.397** | **Yes** |
| **Figure 15** | **Proportional Clustered Seed 91120** | **Average Proportion Paid** | **0.846693** | **0.00288** | **0.84105** | **0.85234** | **138066** | **468.94888** | **137145.784** | **138986.338** | **Yes** |
| **Figure 16** | **Proportional Clustered Seed 91121** | **Average Proportion Paid** | **0.845439** | **0.00292** | **0.83970** | **0.85118** | **137862** | **476.64707** | **136926.148** | **138796.917** | **Yes** |
| **Figure 17** | **Proportional Clustered Seed 91122** | **Average Proportion Paid** | **0.848771** | **0.00277** | **0.84333** | **0.85421** | **138405** | **451.97032** | **137517.829** | **139291.745** | **Yes** |

Table 5: Composite Statistics for 5 Stratified and 5 Cluster Test Executions

**Conclusion**

What Proportion Does Medicare Pay for the Top 100 Diagnosis-Related Groups of Inpatient Charges? Using four different sampling designs we have been able to consistently estimate the mean average proportion paid such that the actual value has been within the 95% confidence interval for all tested designs. The designs that were used were a proportional stratified design, a Neyman allocation stratified design, a cluster design using the same proportional strata allocation in addition to the clusters and a cluster design using the Neyman allocation. For both stratified and cluster methods the proportional performed slightly better than the Neyman. A simple proportionally allocated sampling design is well suited to estimating the mean proportion paid by Medicare for these 100 Diagnosis-Related Groups of Inpatient charges.

**References**

[**https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3**](https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3)

[**https://www.kaggle.com/**](https://www.kaggle.com/)

# Appendix

|  |  |  |  |
| --- | --- | --- | --- |
| **Surveyselect Sample Generation Using Different Seeds (In Addition To Figure 2)** | | | |
|  |  |  |  |
| **Table 5A: Composite Surveyselect For Proportional Sample Generation** | | | |

|  |
| --- |
|  |
|  |
|  |

|  |
| --- |
|  |

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |

|  |
| --- |
|  |
|  |
|  |
|  |

**Appendix**

|  |
| --- |
| */\**  *Celia Taylor and Andrew Abbott*  *MSDS6370 Project Summer 2017*  *data set https://data.cms.gov/Medicare-Inpatient/Inpatient-Prospective-Payment-System-IPPS-Provider/97k6-zzx3*  *\*/*  *data incharges;*  */\*Andrew's directory path*  *infile '/folders/myfolders/sasuser.v94/inpatientChargesAddedDRG\_5.csv' firstobs=2 dlm = ',' DSD;*  *\*/*  */\**  *\* Celia's directory path*  *\*/*  *infile '/folders/myfolders/SASDATA/inpatientChargesAddedDRG\_5FromAA.csv' firstobs=2 dlm = ',' DSD;*  *input DRGDef $ DRGDefNum ProvStreetAdd $ ProvCity $ ProvState $ ProvZipCode HospRefRegionDesc $ TotalDischarges AverageCovCharges AverageTotPayments AverageMedicarePayments Stratum ;*  *AverageProportionPaid = AverageMedicarePayments/AverageTotPayments;*  *output incharges;*  *run;*  *title "Descriptive Statistics - Inpatient Charges Population";*  *proc means data=incharges;*  *footnote "Figure 1 - Descriptive Statistics - Population Inpatient Charges - Total Discharges, DRG, Stratum, Average Proportion Paid";*  *run;*  *title "Select the Sample Proportionally";*  *proc surveyselect data=incharges method = srs out = propsample*  *sampsize = (426,226,142,92,58,27) seed=91118;*  *strata Stratum;*  *footnote "Figure 2 - Sample Selected";*  *run;*  *title "Select the Sample with Neyman Allocation";*  *proc surveyselect data=incharges method = srs out = neymansample*  *sampsize = (475,222,128,79,47,20) seed=91118;*  *strata Stratum;*  *footnote "Figure 3 - Sample Selected";*  *run;*  *data strsizes;*  *input Stratum \_total\_;*  *datalines;*  *1 71475*  *2 37914*  *3 23779*  *4 15545*  *5 9791*  *6 4561*  *;*  *run;*  */\**  *\* CLSUM 100(1-α)% two-sided confidence limits for the SUM,*  *\* where α is determined by the ALPHA= option; the default is α=0.05*  *\**  *\* CLM 100(1-α)% two-sided confidence limits for the MEAN,*  *\* where α is determined by the ALPHA= option; the default is α=0.05*  *\*/*  *title "Proportional allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 4 - Population Inpatient Charges - Proportional Allocation Confidence Limits";*  *run;*  *title "Neyman allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = neymansample sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 5 - Population Inpatient Charges - Neyman Allocation Confidence Limits";*  *run;*  *title "Proportional Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *cluster DRGDefNum;*  *footnote "Figure 6 - Population Inpatient Charges - Proportional Cluster Design Confidence Limits";*  *run;*  *title "Neyman Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = neymansample sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *cluster DRGDefNum;*  *footnote "Figure 7 - Population Inpatient Charges - Neyman Cluster Design Confidence Limits";*  *run;*  *proc surveyselect data=incharges method = srs out = propsample2*  *sampsize = (426,226,142,92,58,27) seed=91119;*  *strata Stratum;*  *footnote "Surveyselect Seed 91119";*  *run;*  *proc surveyselect data=incharges method = srs out = propsample3*  *sampsize = (426,226,142,92,58,27) seed=91120;*  *strata Stratum;*  *footnote "Surveyselect Seed 91120";*  *run;*  *proc surveyselect data=incharges method = srs out = propsample4*  *sampsize = (426,226,142,92,58,27) seed=91121;*  *strata Stratum;*  *footnote "Surveyselect Seed 91121";*  *run;*  *proc surveyselect data=incharges method = srs out = propsample5*  *sampsize = (426,226,142,92,58,27) seed=91122;*  *strata Stratum;*  *footnote "Surveyselect Seed 91122";*  *run;*  *title "Proportional allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 8 - Proportional Stratified - Seed=91118";*  *run;*  *title "Proportional allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample2 sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 9 - Proportional Stratified - Seed=91119";*  *run;*  *title "Proportional allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample3 sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 10 - Proportional Stratified - Seed=91120";*  *run;*  *title "Proportional allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample4 sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 11 - Proportional Stratified - Seed=91121";*  *run;*  *title "Proportional allocation with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample5 sum clsum total = strsizes mean CLM;*  *var AverageProportionPaid;*  *weight SamplingWeight;*  *strata Stratum / list;*  *footnote "Figure 12 - Proportional Stratified - Seed=91122";*  *run;*  *title "Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample sum clsum total = strsizes mean CLM;*  *strata Stratum;*  *weight SamplingWeight;*  *cluster DRGDefNum;*  *var AverageProportionPaid;*  *footnote "Figure 13 - Proportional Clustered - Seed=91118";*  *run;*  *title "Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample2 sum clsum total = strsizes mean CLM;*  *strata Stratum;*  *weight SamplingWeight;*  *cluster DRGDefNum;*  *var AverageProportionPaid;*  *footnote "Figure 14 - Proportional Clustered - Seed=91119";*  *run;*  *title "Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample3 sum clsum total = strsizes mean CLM;*  *strata Stratum;*  *weight SamplingWeight;*  *cluster DRGDefNum;*  *var AverageProportionPaid;*  *footnote "Figure 15 - Proportional Clustered - Seed=91120";*  *run;*  *title "Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample4 sum clsum total = strsizes mean CLM;*  *strata Stratum;*  *weight SamplingWeight;*  *cluster DRGDefNum;*  *var AverageProportionPaid;*  *footnote "Figure 16 - Proportional Clustered - Seed=91121";*  *run;*  *title "Cluster Sample with CLM two-sided confidence limits for the MEAN and CLSUM for two-sided confidence limits for the SUM";*  *proc surveymeans data = propsample5 sum clsum total = strsizes mean CLM;*  *strata Stratum;*  *weight SamplingWeight;*  *cluster DRGDefNum;*  *var AverageProportionPaid;*  *footnote "Figure 17 - Proportional Clustered - Seed=91122";*  *run;* |