**Results for the pre-module**

**TOOLS USED**

SAS

Question 1

**Step 2: Identify the most common types of the RA**

A screenshot of a computer

AI-generated content may be incorrect.

In the **Chronic RA** group, the most frequently occurring code is **M069,** with **712 cases**, making it the dominant diagnosis. The second most common code, **M0579,** appears in **15 cases**, while **M059** is recorded in **8 cases**. In the **Systemic RA** group, the most frequently recorded code is **M0510**, with **18 cases.** The codes **M0519** and **M05671** are much less frequent, each appearing in only **2 cases.** This distribution indicates that **Chronic RA cases are significantly more common** than **Systemic RA cases**, with a large portion of patients diagnosed with **M069**

**Step 3: Gender differences in RA**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Fisher's Exact Test: Gender Differences in RA Prevalence |  |  |  |  |  |
|  |  |  |  |  |  |
| The FREQ Procedure |  |  |  |  |  |
|  |  |  |  |  |  |
| **Frequency** |  | **Table of sex by RA\_flag** | | | |
| **Percent** |  | **sex** | **RA\_flag** | | |
| **Row Pct** |  | **0** | **1** | **Total** |
| **Col Pct** |  | **1** | 168199 | 203 | 168402 |
|  | | 45.38 | 0.05 | 45.44 |
|  | | 99.88 | 0.12 |  |
|  | | 45.48 | 25.38 |  |
|  | | **2** | 201623 | 597 | 202220 |
|  | | 54.4 | 0.16 | 54.56 |
|  | | 99.7 | 0.3 |  |
|  | | 54.52 | 74.63 |  |
|  | | **Total** | 369822 | 800 | 370622 |
|  | | 99.78 | 0.22 | 100 |
|  | | **Frequency Missing = 11** | | | |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Statistics for Table of sex by RA\_flag** |  |  |  |  |  |
|  |  |  |  |  |  |
| **Statistic** | **DF** | **Value** | **Prob** |  |  |
| **Chi-Square** | 1 | 130.1658 | <.0001 |  |  |
| **Likelihood Ratio Chi-Square** | 1 | 137.6341 | <.0001 |  |  |
| **Continuity Adj. Chi-Square** | 1 | 129.3561 | <.0001 |  |  |
| **Mantel-Haenszel Chi-Square** | 1 | 130.1655 | <.0001 |  |  |
| **Phi Coefficient** |  | 0.0187 |  |  |  |
| **Contingency Coefficient** |  | 0.0187 |  |  |  |
| **Cramer's V** |  | 0.0187 |  |  |  |
|  |  |  |  |  |  |
| **Fisher's Exact Test** | |  |  |  |  |
| **Cell (1,1) Frequency (F)** | 168199 |  |  |  |  |
| **Left-sided Pr <= F** | 1 |  |  |  |  |
| **Right-sided Pr >= F** | <.0001 |  |  |  |  |
|  |  |  |  |  |  |
| **Table Probability (P)** | <.0001 |  |  |  |  |
| **Two-sided Pr <= P** | <.0001 |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Sample Size = 370622** |  |  |  |  |  |
| **Frequency Missing = 11** |  |  |  |  |  |

Among the **370,622 individuals** inthe dataset, **800 cases** were diagnosed with RA. Women (sex = 2) had a higher **prevalence of RA (0.3%)** compared to men (sex = 1) who had a lower prevalence of **0.12%.** While the overall proportion of RA cases is low, the statistical tests confirm that **women are significantly more likely to have RA than men**, aligning with known epidemiological trends.

**Step 4: Calculate the inter-quartile range of the costs**



The **first quartile (Q1) is 682.48**, meaning 25% of the data falls below this value, while the **median is 1521.62,** representing the middle value of the dataset. The **third quartile (Q3) is 3440.18,** indicating that 75% of the data falls below this point. The **interquartile range (IQR) is 2757.7**, showing the spread of the middle 50% of the data. The large gap between Q1 and Q3 suggests high variability, and the median being closer to Q1 than Q3 further supports a **right-skewed distribution.**

**Step 5: Study of service utilization**

A screenshot of a table

AI-generated content may be incorrect.

The most frequently used service is associated with REVCODE 300, which was utilized 2,255 times, accounting for 27.98% of total usage. REVCODE 636 follows with 890 occurrences (11.04%), and REVCODE 250 ranks third with 874 occurrences (10.84%). Additionally, REVCODE 450 and 320 appear 831 (10.31%) and 291 (3.61%) times, respectively.

Question 2

Lions share in terms of admissions

A screenshot of a computer

AI-generated content may be incorrect.

the lion's share of admissions is associated with the category where MDC is 1 and hnum2 is 5, accounting for 2,039 admissions out of a total of 3,276, making up 62.24% of the total admissions in that subset. Similarly, another significant contributor is the category with MDC 1 and hnum2 8, which records 2,416 admissions, comprising 42.59% of the overall admissions. These two categories dominate the distribution, indicating a higher utilization of healthcare services within these specific groups. Other categories contribute smaller shares, ranging between 10.23% and 2.59%.

Lion share in terms of charges

A screenshot of a data

AI-generated content may be incorrect.

the highest proportion of total charges is attributed to the first row, where MDC 1 and hnum2 5 contribute **$92,200,269.6,** making up **79.27%** of the grand total charges of **$116,309,418.86.** This indicates that this category holds the lion's share of the total charges. The second highest contribution comes from MDC 14 and hnum2 8, amounting to **$30,773,062.49,** which represents **46.53%** of the respective grand total of **$66,142,373.48.** Other categories have significantly lower shares, with percentages dropping below 8%.

Thus it is hospital 5 (University of Vermont Medical Center)