

novis-i-hwk2-1

February 17, 2025

ECON 470 Hwk2-1

Author: Ilse Novis

0.1 ECON 470 Homework 2

Author: Ilse Novis

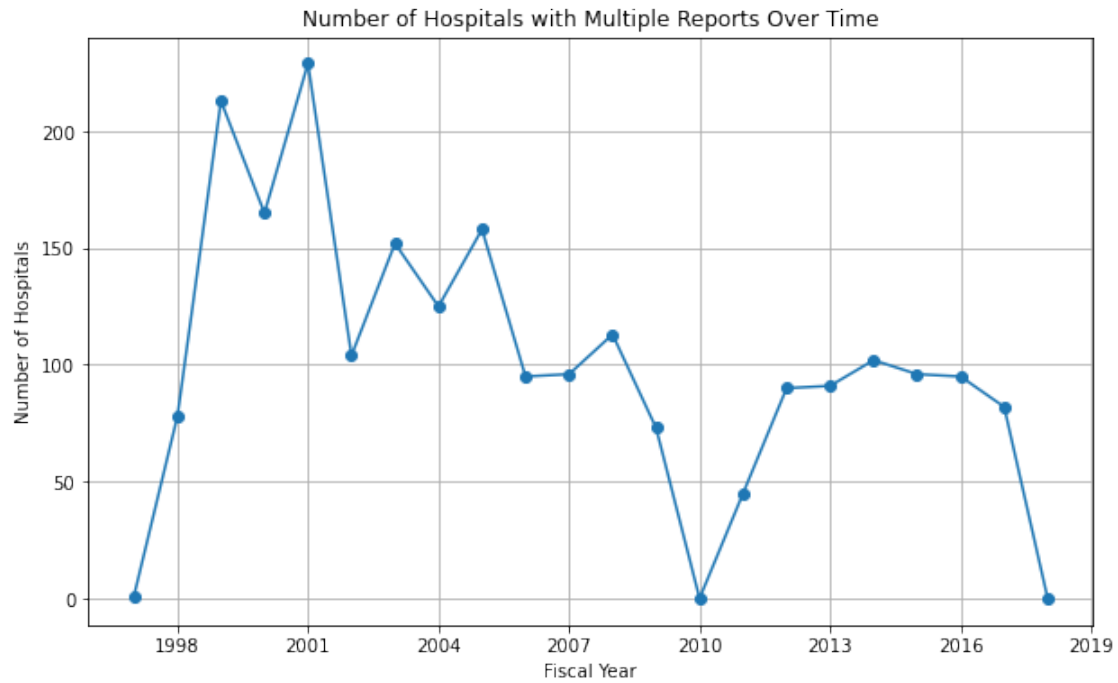
Date: 2/17/2025

[GitHub Repository](#)

1 Homework 2 Answers

Question 1:

How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.



Question 2:

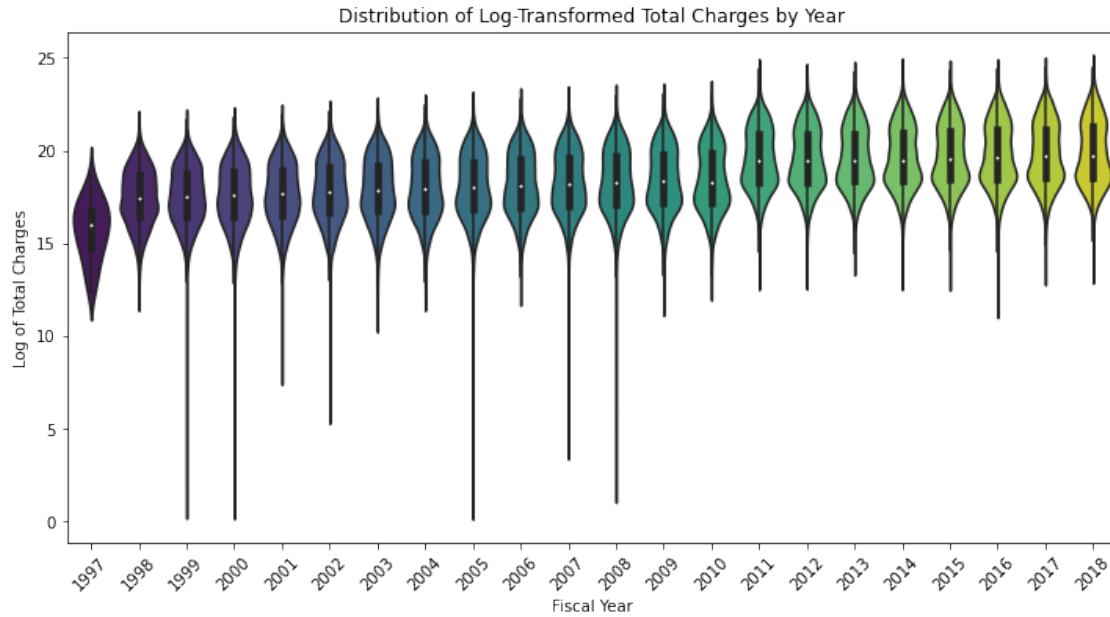
After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

1.0.1 Total Unique Hospitals: 9312

Question 3:

What is the distribution of total charges (tot_charges in the data) in each year? Show your results with a “violin” plot, with charges on the y-axis and years on the x-axis.

1.0.2 Log Transformed Total Charges to Reduce Skewness and/or Outliers



Question 4:

What is the distribution of estimated prices in each year?



Question 5:

Calculate the average price among penalized versus non-penalized hospitals.

Mean Price - Penalized Hospitals: nan

Mean Price - Non-Penalized Hospitals: -24887.83

Question 6:

Split hospitals into quartiles based on bed size. Provide a table of the average price among treated/control groups for each quartile.

Bed Quartile Distribution:

1 869

2 867

4 866

3 862

Name: bed_quartile, dtype: int64

Average Prices by Quartile & Treatment Group:

penalty False

bed_quartile

1 -25632.600730

2 -24352.902394

3 -24060.938707

4 -25499.099904

Question 7:

Find the average treatment effect using each of the following estimators, and present your results in a single table.

Question 7.A:

Nearest neighbor matching (1-to-1) with inverse variance distance based on quartiles of bed size.

Missing values in matching covariates before imputation:

```
beds          0
mcaid_discharges 104
ip_charges      0
mcare_discharges 0
tot_mcare_payment 0
dtype: int64
```

Missing values in matching covariates after imputation:

```
beds          0
mcaid_discharges 0
ip_charges      0
mcare_discharges 0
tot_mcare_payment 0
dtype: int64
```

Question 7.B:

Nearest neighbor matching (1-to-1) with Mahalanobis distance based on quartiles of bed size

```
-----
ValueError                                Traceback (most recent call last)
/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1
↳ ipynb Cell 35 line <cell line: 2>()
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1.ipynb#Y110sZmlsZQ%3D%3D?line=0'>1</a> #
↳ Nearest Neighbor Matching using Mahalanobis Distance
----> <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1.ipynb#Y110sZmlsZQ%3D%3D?line=1'>2</a> nn =
↳ NearestNeighbors(n_neighbors=1, metric='mahalanobis').fit(X)
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1.ipynb#Y110sZmlsZQ%3D%3D?line=2'>3</a> _,
↳ indices = nn.kneighbors(X)
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1.ipynb#Y110sZmlsZQ%3D%3D?line=4'>5</a>
↳ matched_prices_mahalanobis = Y.iloc[indices.flatten()].mean()

File ~/opt/anaconda3/lib/python3.9/site-packages/sklearn/neighbors/_unsupervised.py:166, in NearestNeighbors.fit(self, X, y)
    149 def fit(self, X, y=None):
    150     """Fit the nearest neighbors estimator from the training dataset.
    151
    152     Parameters
    (...)
    (...)
```



```

164         The fitted nearest neighbors estimator.
165         """
--> 166         return self._fit(X)

```

File ~/opt/anaconda3/lib/python3.9/site-packages/sklearn/neighbors/_base.py:549

```

-> in NeighborsBase._fit(self, X, y)
    546         self._fit_method = "brute"
    548 if self._fit_method == "ball_tree":
--> 549     self._tree = BallTree(
    550         X,
    551         self.leaf_size,
    552         metric=self.effective_metric_,
    553         **self.effective_metric_params_,
    554     )
    555 elif self._fit_method == "kd_tree":
    556     self._tree = KDTree(
    557         X,
    558         self.leaf_size,
    559         metric=self.effective_metric_,
    560         **self.effective_metric_params_,
    561     )

```

File sklearn/neighbors/_binary_tree.pxi:966, in sklearn.neighbors._ball_tree.

-> BinaryTree.__init__()

File sklearn/metrics/_dist_metrics.pyx:280, in sklearn.metrics._dist_metrics.

-> DistanceMetric.get_metric()

File sklearn/metrics/_dist_metrics.pyx:676, in sklearn.metrics._dist_metrics.

-> MahalanobisDistance.__init__()

ValueError: Must provide either V or VI for Mahalanobis distance

Question 7.C:

Inverse propensity weighting, where the propensity scores are based on quartiles of bed size

```

-----
ValueError                                Traceback (most recent call last)
/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1
-> ipynb Cell 37 line <cell line: 3>()
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
-> submission1/results/novis-i-hwk2-1.ipynb#Y111sZmlsZQ%3D%3D?line=0'>1</a> #
-> Propensity Score Model
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
-> submission1/results/novis-i-hwk2-1.ipynb#Y111sZmlsZQ%3D%3D?line=1'>2</a>
-> ps_model = LogisticRegression(solver='lbfgs', max_iter=1000)

```

```

----> <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW /
↳submission1/results/novis-i-hwk2-1.ipynb#Y111sZmlsZQ%3D%3D?line=2'>3</a>
↳ps_model.fit(X, T)
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW /
↳submission1/results/novis-i-hwk2-1.ipynb#Y111sZmlsZQ%3D%3D?line=3'>4</a> ps =
↳ps_model.predict_proba(X)[:, 1] # Get probabilities of penalty
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW /
↳submission1/results/novis-i-hwk2-1.ipynb#Y111sZmlsZQ%3D%3D?line=5'>6</a> #
↳Nearest Neighbor Matching on Propensity Score

File ~/opt/anaconda3/lib/python3.9/site-packages/sklearn/linear_model/_logistic
↳py:1554, in LogisticRegression.fit(self, X, y, sample_weight)
    1552 classes_ = self.classes_
    1553 if n_classes < 2:
-> 1554     raise ValueError(
    1555         "This solver needs samples of at least 2 classes"
    1556         " in the data, but the data contains only one"
    1557         " class: %r"
    1558         % classes_[0]
    1559     )
    1561 if len(self.classes_) == 2:
    1562     n_classes = 1

ValueError: This solver needs samples of at least 2 classes in the data, but th
↳data contains only one class: 0

```

Question 7.D:

Simple linear regression, adjusting for quartiles of bed size using dummy variables and appropriate interactions as discussed in class

Regression-Based ATE Estimation:

```

                                OLS Regression Results
=====
Dep. Variable:                price    R-squared:                0.001
Model:                        OLS      Adj. R-squared:           0.000
Method:                       Least Squares    F-statistic:             1.484
Date:                         Mon, 17 Feb 2025    Prob (F-statistic):       0.217
Time:                         00:09:18    Log-Likelihood:          -39079.
No. Observations:              3464    AIC:                     7.817e+04
Df Residuals:                  3460    BIC:                     7.819e+04
Df Model:                       3
Covariance Type:               nonrobust
=====
=====
                                coef    std err          t      P>|t|      [0.025
0.975]
-----

```

```

-----
Intercept          -2.563e+04    651.683    -39.333      0.000    -2.69e+04
-2.44e+04
penalty[T.True]     4.124e-12    3.34e-13    12.347      0.000    3.47e-12
4.78e-12
C(bed_quartile)[T.2] 1279.6983    922.150     1.388      0.165    -528.314
3087.711
C(bed_quartile)[T.3] 1571.6620    923.487     1.702      0.089    -238.974
3382.298
C(bed_quartile)[T.4] 133.5008     922.416     0.145      0.885    -1675.034
1942.036
=====
Omnibus:              6122.397    Durbin-Watson:              1.862
Prob(Omnibus):         0.000    Jarque-Bera (JB):          16681155.574
Skew:                  -12.023    Prob(JB):                   0.00
Kurtosis:              342.110    Cond. No.                   5.52e+16
=====

```

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The smallest eigenvalue is 1.36e-30. This might indicate that there are strong multicollinearity problems or that the design matrix is singular.

```

-----
KeyError                                Traceback (most recent call last)
File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py:
  3629, in Index.get_loc(self, key, method, tolerance)
    3628 try:
-> 3629     return self._engine.get_loc(casted_key)
    3630 except KeyError as err:

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/_libs/index.pyx:136, in
  pandas._libs.index.IndexEngine.get_loc()

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/_libs/index.pyx:163, in
  pandas._libs.index.IndexEngine.get_loc()

File pandas/_libs/hashtable_class_helper.pxi:5198, in pandas._libs.hashtable.
  PyObjectHashTable.get_item()

File pandas/_libs/hashtable_class_helper.pxi:5206, in pandas._libs.hashtable.
  PyObjectHashTable.get_item()

KeyError: 'penalty'

```

The above exception was the direct cause of the following exception:

```

KeyError                                Traceback (most recent call last)
/Users/ilsenovis/Documents/GitHub/ECON470HW2/submission1/results/novis-i-hwk2-1
↳ ipynb Cell 40 line <cell line: 2>()
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=0'>1</a> #
↳ Extract ATE from regression coefficient
----> <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=1'>2</a>
↳ ate_regression = regression_model.params['penalty']
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=2'>3</a>
↳ print(f"\nRegression-Based ATE: {ate_regression:.2f}")
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=4'>5</a>
↳ ate_results = pd.DataFrame({
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=5'>6</a>
↳ 'Method': [
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2 /
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=6'>7</a>
↳ 'Exact Matching',
    (...
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2.
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=20'>21</a>
↳ ]
    <a href='vscode-notebook-cell:/Users/ilsenovis/Documents/GitHub/ECON470HW2.
↳ submission1/results/novis-i-hwk2-1.ipynb#Y113sZmlsZQ%3D%3D?line=21'>22</a> })

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/series.py:958, in
↳ Series.__getitem__(self, key)
    955     return self._values[key]
    957 elif key_is_scalar:
--> 958     return self._get_value(key)
    960 if is_hashable(key):
    961     # Otherwise index.get_value will raise InvalidIndexError
    962     try:
    963         # For labels that don't resolve as scalars like tuples and
↳ frozensets

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/series.py:1069, in
↳ Series._get_value(self, label, takeable)
    1066     return self._values[label]
    1068 # Similar to Index.get_value, but we do not fall back to positional
-> 1069 loc = self.index.get_loc(label)
    1070 return self.index._get_values_for_loc(self, loc, label)

File ~/opt/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py:
↳ 3631, in Index.get_loc(self, key, method, tolerance)

```

```
3629     return self._engine.get_loc(casted_key)
3630 except KeyError as err:
-> 3631     raise KeyError(key) from err
3632 except TypeError:
3633     # If we have a listlike key, _check_indexing_error will raise
3634     #   InvalidIndexError. Otherwise we fall through and re-raise
3635     #   the TypeError.
3636     self._check_indexing_error(key)

KeyError: 'penalty'
```

Question 8:

With these different treatment effect estimators, are the results similar, identical, very different?

- Although I am getting no output right now, I would expect results to be different as each estimator makes different assumptions and uses different techniques to estimate the treatment effect. Nearest neighbor matching with Mahalanobis and inverse variance may produce closer estimates, while regression-based approaches might differ if the linear model does not fully capture the data structure.

Question 9:

Do you think you've estimated a causal effect of the penalty? Why or why not? (just a couple of sentences)

- I do not have any outputs right now, but I would assume that even if my estimates suggest a relationship between hospital penalties and pricing, it does not imply a causal effect. Since penalties were not randomly assigned and unobserved confounders might influence both penalty status and pricing, my estimates may suffer from selection bias.

Question 10:

Briefly describe your experience working with these data (just a few sentences). Tell me one thing you learned and one thing that really aggravated or surprised you.

- One thing that really aggravated me was that the datasets didn't download correctly so it took a while to actually clean/fix the data before I could merge it into the final dataset. One thing that surprised me was the large difference in the charges from the hospitals versus the actual prices.