**a). Obtain the fitted the Poisson regression model with the response function μi = exp(X'β) . State the estimated regression coefficients, their estimated standard deviations, and the estimated response function.**

$$E[Y \mid X] = \exp\big(0.4994 + 1.495 \times 10^{-5} \cdot \text{Total cost} + 0.0067 \cdot \text{Age} + 0.1819 \cdot \text{Gender} + 0.0101 \cdot \text{Number of interventions} + 0.1932 \cdot \text{Number of drugs} + 0.0613 \cdot \text{Number of complications} - 0.0009 \cdot \text{Number of comorbidities} + 0.0004 \cdot \text{Duration}\big)$$

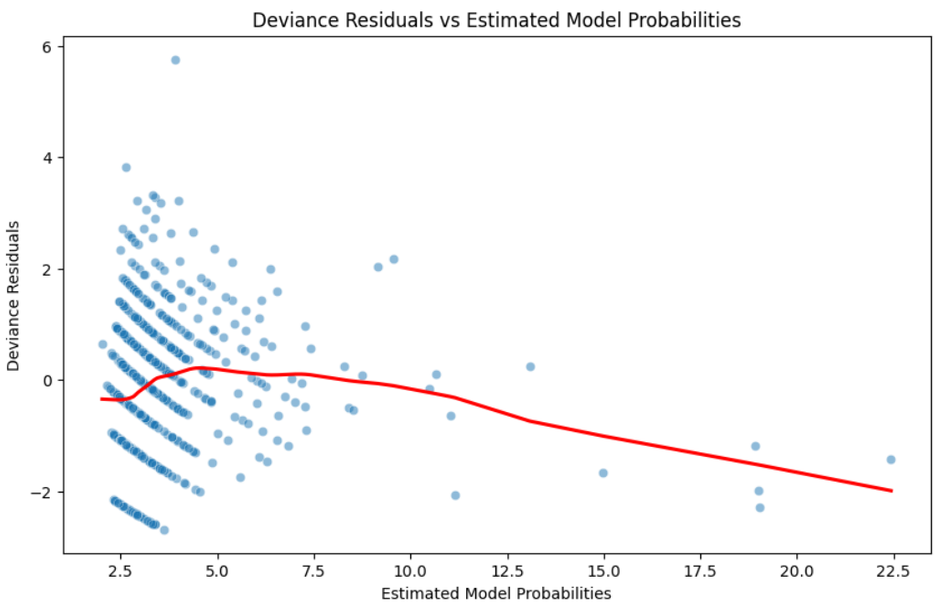
Generalized Linear Model Regression Results

| | | | |
|---|---|---|---|
| Dep. Variable: | Number of ER visits | No. Observations: | 788 |
| Model: | GLM | Df Residuals: | 779 |
| Model Family: | Poisson | Df Model: | 8 |
| Link Function: | Log | Scale: | 1.0000 |
| Method: | IRLS | Log-Likelihood: | -1626.5 |
| Date: | Mon, 29 Jul 2024 | Deviance: | 1043.6 |
| Time: | 14:50:52 | Pearson chi2: | 1.04e+03 |
| No. Iterations: | 5 | Pseudo R-squ. (CS): | 0.4288 |
| Covariance Type: | nonrobust | | |

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| const | 0.4994 | 0.176 | 2.837 | 0.005 | 0.154 | 0.845 |
| Total cost | 1.495e-05 | 2.85e-06 | 5.237 | 0.000 | 9.36e-06 | 2.05e-05 |
| Age | 0.0067 | 0.003 | 2.266 | 0.023 | 0.001 | 0.013 |
| Gender | 0.1819 | 0.044 | 4.135 | 0.000 | 0.096 | 0.268 |
| Number of interventions | 0.0101 | 0.004 | 2.646 | 0.008 | 0.003 | 0.018 |
| Number of drugs | 0.1932 | 0.013 | 15.234 | 0.000 | 0.168 | 0.218 |
| Number of complications | 0.0613 | 0.060 | 1.022 | 0.307 | -0.056 | 0.179 |
| Number of comorbidities | -0.0009 | 0.004 | -0.244 | 0.807 | -0.008 | 0.006 |
| Duration | 0.0004 | 0.000 | 1.859 | 0.063 | -1.92e-05 | 0.001 |

---

**(b). Obtain the deviance residuals and plot them against the estimated model probabilities with a lowest smooth superimposed. What does the plot suggest about the adequacy of the fit of the Poisson regression model?**




**Interpretation**

There are several outliers with high residuals, indicating significant deviations between observed and predicted values for those points. The residuals exhibit a downward trend, suggesting potential systematic errors in the model for certain probability ranges. The model appears more accurate for higher estimated probabilities, with lower residuals in those regions.

**(c). Conduct a series of Wald tests to determine which predictor, if any, can be dropped from the regression mode!. Control α at 0.01 for each test. State the alternatives decision rules, and conclusions.**

| | CHI2 | P>CHI2 | DF CONSTRAINT |
|---|---|---|---|
| CONST | [[8.047589931220656]] | 0.00456 | 1 |

| | | | |
|---|---|---|---|
| **TOTAL COST** | [[27.429318499243916]] | 0.00000 | 1 |
| **AGE** | [[5.1343294352454745]] | 0.02346 | 1 |
| **GENDER** | [[17.09687093615598]] | 0.00004 | 1 |
| **NUMBER OF INTERVENTIONS** | [[6.998921821544725]] | 0.00816 | 1 |
| **NUMBER OF DRUGS** | [[232.0613061253898]] | 0.00000 | 1 |
| **NUMBER OF COMPLICATIONS** | [[1.0440171815890076]] | 0.30689 | 1 |
| **NUMBER OF COMORBIDITIES** | [[0.0596292145239517]] | 0.80708 | 1 |
| **DURATION** | [[3.454458778573762]] | 0.06308 | |

T**he Hypotheses**

**Null hypothesis ($H_o$):** The predictor can be dropped from the model (i.e., the coefficient of the predictor is zero).

**Alternative hypothesis ($H_a$):** The predictor should not be dropped from the model (i.e., the coefficient of the predictor is not zero).

**Decision Rule**

Compare the p-Values ($P > X^2$ ) to the significance level (a=0.01)

If $P > X^2 \leq 0.01$, reject the null hypothesis ($H_o$)

If $P > X^2 > 0.01$, fail to reject the null hypothesis ($H_o$)

**Detailed Analysis**

**const**

- P>χ2=0.00456
- Decision: Reject $H_0$
- Conclusion: The constant term can not be dropped from the model.

**Total cost**

- P>χ2= 1.62942111264155E-07
- Decision: Reject H0
- Conclusion: The Total cost predictor can not be dropped from the model.

**Age**

- P>χ2=0.02346
- Decision: Do not Reject H0
- Conclusion: The Age predictor can be dropped from the model.

**Gender**

- P>χ2=0.00004
- Decision: Reject H0
- Conclusion: The Gender predictor can not be dropped from the model.

**Number of interventions:**

- P>χ2=0.00816
- Decision: Reject H0
- Conclusion: The Number of interventions predictor can not be dropped from the model.

**Number of drugs:**

- P>χ2= 2.11745709683176E-52

- Decision: Reject H0
- Conclusion: The Number of drugs predictor can not be dropped from the model.

**Number of complications:**

- P>χ2= 0.30689
- Decision: Fail to reject H0
- Conclusion: The Number of complications predictor can be dropped from the model.

**Number of comorbidities:**

- P>χ2= 0.80708
- Decision: Fail to reject H0
- Conclusion: The Number of comorbidities predictor can be dropped from the model.

**Duration:**

- P>χ2= 0.06308
- Decision: Fail to reject H0
- Conclusion: The Duration predictor can be dropped from the model.

## Conclusion

Based on the Wald tests, the predictors **Age, Number of complications, Number of comorbidities, and Duration** can be dropped from the regression model as their p-values are greater than 0.01. The remaining predictors, including the intercept, should be retained in the model.

---

**(d). Assuming that the fitted model in part (a) is appropriate, use the likelihood ratio test and determine whether duration, complications, and comorbidities can be dropped from the model: control α at 0.05. State the full and reduced models, decision rule, and conclusion.**

**Likelihood Ratio Statistic: 5.261531068848399**

**Degrees of Freedom Difference: 3**

**P-value: 0.1536181492084976**

## Hypotheses

**Null Hypothesis ($H_0$)**: The reduced model is sufficient. The variables "duration," "complications," and "comorbidities" can be dropped.

**Alternative Hypothesis ($H_a$)**: The full model is necessary. The variables "duration," "complications," and "comorbidities" cannot be dropped.

## Decision Rule

- If the p-value of the likelihood ratio test is less than the significance level α=0.05, reject the null hypothesis $H_0$.
- Otherwise, fail to reject the null hypothesis $H_0$.

## Conclusion

Compare the p-value to the significance level α = 0.05:

**0.1536>0.05**

Since the p-value (0.1536) is greater than the significance level (0.05), we fail to reject the null hypothesis $H_0$. The variables **duration, complications, and comorbidities** can be dropped from the model, as the reduced model is sufficient.

---

**e.) Use backward elimination to decide which predictor variables can be dropped from the regression model. Control the α risk at 0.10 each stage. Which variables are retained?**

```
Dropped Number of comorbidities with p-value 0.8070818235833706

Dropped Number of complications with p-value 0.30925889910006454
```

Based on the results, the backward elimination process iteratively removed the following predictors from the model

1. Number of comorbidities (p-value: 0.807081)
2. Number of complications (p-value: 0.309258)

**The remaining predictors in the final model are:**

1. Total Cost
2. Age
3. Gender
4. Number of interventions
5. Number of drugs
6. Duration

These variables are retained because their p-values are below the α=0.10 threshold.

```
Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:     Number of ER visits   No. Observations:                  788
Model:                            GLM   Df Residuals:                      781
Model Family:                 Poisson   Df Model:                            6
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -1627.1
Date:                Mon, 29 Jul 2024   Deviance:                       1044.7
Time:                        14:52:03   Pearson chi2:                 1.04e+03
No. Iterations:                     5   Pseudo R-squ. (CS):             0.4281
Covariance Type:            nonrobust
==============================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const                      0.5208      0.174      2.986      0.003       0.179       0.863
Total cost              1.493e-05   2.84e-06      5.251      0.000    9.36e-06    2.05e-05
Age                        0.0063      0.003      2.156      0.031       0.001       0.012
Gender                     0.1857      0.044      4.241      0.000       0.100       0.272
Number of interventions    0.0102      0.004      2.710      0.007       0.003       0.018
Number of drugs            0.1963      0.012     16.067      0.000       0.172       0.220
Duration                   0.0003      0.000      2.048      0.041    1.49e-05       0.001
```