**포아송 모형 적합 예제**

한 품질 엔지니어가 성형 수지 부품의 결점 유형(변색, 찌꺼기)에 관심이 있습니다. 호스 오염과 수지 펠릿의 마찰로 인해 최종 제품에 변색된 줄이 나타날 수 있습니다. 공정이 더 높은 온도에서 더 빠른 이동률로 실행될 경우 찌꺼기가 발생할 수 있습니다. 엔지니어는 반응(결점)에 대한 세 가지 가능한 예측 변수를 식별합니다. 엔지니어는 예측 변수 수준에 변화를 주면서 각 결점 유형의 1시간 세션당 결점 수를 기록합니다.

엔지니어는 여러 예측 변수가 변속 결정에 미치는 영향을 조사하려고 합니다. 반응 변수가 유한 관측 공간에서 사건이 발생하는 횟수를 설명하기 때문에 엔지니어는 포아송 모형을 적합합니다.

**포아송 회귀 분석: 변색 결점 대 세척 후 시간, 온도, 나사 크기**

**방법**

|  |  |
| --- | --- |
| 연결 함수 | 자연 로그 |
| 범주형 예측 변수 코드화 | (1, 0) |
| 사용된 행 | 36 |

**회귀 방정식**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 변색 결점 | = | | exp(Y') | | |
| **나사 크기** | |  | |  |  | |
| 대 | | Y' | | = | 4.398 + 0.01798 세척 후 시간 - 0.001974 온도 | |
|  | |  | |  |  | |
| 소 | | Y' | | = | 4.244 + 0.01798 세척 후 시간 - 0.001974 온도 | |

**계수**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **항** | **계수** | **SE 계수** | **Z-값** | **P-값** | **VIF** |
| 상수 | 4.3982 | 0.0628 | 70.02 | 0.000 |  |
| 세척 후 시간 | 0.01798 | 0.00826 | 2.18 | 0.029 | 1.00 |
| 온도 | -0.001974 | 0.000318 | -6.20 | 0.000 | 1.00 |
| 나사 크기 |  |  |  |  |  |
| 소 | -0.1546 | 0.0427 | -3.62 | 0.000 | 1.00 |

**모형 요약**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **이탈도 R-Sq** | **이탈도 R-Sq(수정)** | **AIC** | **AICc** | **BIC** |
| 64.20% | 60.80% | 253.29 | 254.58 | 259.62 |

**적합도 검정**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **검정** | **DF** | **추정치** | **평균** | **카이-제곱** | **P-값** |
| 이탈도 | 32 | 31.60722 | 0.98773 | 31.61 | 0.486 |
| Pearson | 32 | 31.26713 | 0.97710 | 31.27 | 0.503 |

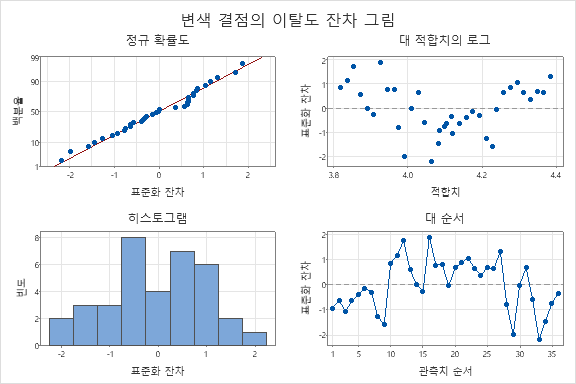
**분산 분석**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Wald 검정** | |
| **출처** | **DF** | **카이-제곱** | **P-값** |
| 회귀 | 3 | 56.29 | 0.000 |
| 세척 후 시간 | 1 | 4.74 | 0.029 |
| 온도 | 1 | 38.46 | 0.000 |
| 나사 크기 | 1 | 13.09 | 0.000 |

**비정상적 관측치에 대한 적합치 및 진단**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **관측** | **변색 결점** | **적합치** | **잔차** | **표준화 잔차** |  |
| 33 | 43.00 | 58.18 | -2.09 | -2.18 | R |

*R  큰 잔차*



**분석 방법**

* **연결 함수**: 자연 로그(log-link) 함수가 사용되었습니다.
* **범주형 예측 변수 코드화**: 범주형 변수인 나사 크기는 (1, 0)으로 코드화되었습니다.
* **사용된 행**: 총 36개의 데이터 행이 사용되었습니다.

**회귀 방정식**

변색 결점의 수는 지수 함수로 변환된 Y′Y'Y′ 값으로 예측됩니다.

* **대형 나사**:

Y′=4.398+0.01798×세척 후 시간−0.001974×온도Y' = 4.398 + 0.01798 \times \text{세척 후 시간} - 0.001974 \times \text{온도}Y′=4.398+0.01798×세척 후 시간−0.001974×온도

* **소형 나사**:

Y′=4.244+0.01798×세척 후 시간−0.001974×온도Y' = 4.244 + 0.01798 \times \text{세척 후 시간} - 0.001974 \times \text{온도}Y′=4.244+0.01798×세척 후 시간−0.001974×온도

변색 결점 수 = exp⁡(Y′)\exp(Y')exp(Y′)

**계수 해석**

* **상수항 (Intercept)**:
  + 대형 나사: 4.3982
  + P-값이 0.000으로 매우 유의미합니다.
* **세척 후 시간**:
  + 계수: 0.01798
  + P-값: 0.029로 유의미함. 세척 후 시간이 증가할수록 변색 결점이 증가하는 경향을 보입니다.
* **온도**:
  + 계수: -0.001974
  + P-값: 0.000으로 매우 유의미함. 온도가 증가할수록 변색 결점이 감소하는 경향을 나타냅니다.
* **나사 크기 (소형 나사)**:
  + 계수: -0.1546
  + P-값: 0.000으로 매우 유의미함. 대형 나사에 비해 소형 나사는 변색 결점이 적게 나타납니다.

**모형 요약**

* **이탈도 R-Squared**: 64.20% (수정된 R-Squared: 60.80%)
  + 모형이 데이터의 64.20%의 변동성을 설명할 수 있음을 의미합니다.
* **AIC (Akaike Information Criterion)**: 253.29
  + AIC는 모형의 적합도를 평가하는 지표로, 낮을수록 더 나은 모형으로 간주됩니다.

**적합도 검정**

* **이탈도**와 **Pearson 검정**의 P-값이 각각 0.486과 0.503으로, 모형이 데이터에 잘 맞는다는 것을 나타냅니다.

**분산 분석 (Wald 검정)**

* 회귀 분석 전체는 P-값 0.000으로 매우 유의미합니다.
* **세척 후 시간**: P-값 0.029로 유의미한 영향을 미침.
* **온도**: P-값 0.000으로 매우 유의미한 영향을 미침.
* **나사 크기**: P-값 0.000으로 매우 유의미한 영향을 미침.

**비정상적 관측치**

* 관측치 33번이 큰 잔차를 가지고 있으며, 표준화 잔차가 -2.18로 나타나, 비정상적 관측치로 간주될 수 있습니다.

**결론**

이 포아송 회귀 분석 결과는 세척 후 시간, 온도, 나사 크기가 변색 결점 수에 중요한 영향을 미친다는 것을 보여줍니다. 특히, 온도가 상승할수록 변색 결점이 감소하고, 세척 후 시간이 길어질수록 변색 결점이 증가하는 경향이 있습니다. 또한, 소형 나사는 대형 나사에 비해 변색 결점이 적게 발생하는 경향이 확인되었습니다. 이러한 결과를 바탕으로 공정 최적화를 통해 변색 결점을 줄일 수 있는 개선 방안을 도출할 수 있습니다.

**결과 해석**

표준화 이탈도 잔차 대 적합치 그림은 고유 곡선을 표시합니다. 잔차 대 순서 그림에서 데이터 집합의 중간에 있는 잔차는 처음과 끝에 있는 잔차보다 더 높은 경향이 있습니다. 이런 데이터의 경우, 두 패턴의 원인은 모두 나사 크기와 온도 간에 교호작용 항이 없기 때문입니다. 잔차 대 순서 그림에서 패턴이 보이는 이유는 엔지니어가 데이터를 임의적인 순서로 수집하지 않았기 때문입니다. 엔지니어는 온도와 나사 크기 사이의 교호작용을 통해 모형을 다시 적합시켜서 결점을 더 정확하게 모형화합니다.

**수정후 포아송 회귀 분석: 온도, 나사 크기의 교초작용 항 추가**

**방법**

|  |  |
| --- | --- |
| 연결 함수 | 자연 로그 |
| 범주형 예측 변수 코드화 | (1, 0) |
| 사용된 행 | 36 |

**회귀 방정식**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 변색 결점 | = | | exp(Y') | | |
| **나사 크기** | |  | |  |  | |
| 대 | | Y' | | = | 4.576 + 0.01798 세척 후 시간 - 0.003285 온도 | |
|  | |  | |  |  | |
| 소 | | Y' | | = | 4.032 + 0.01798 세척 후 시간 - 0.000481 온도 | |

**계수**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **항** | **계수** | **SE 계수** | **Z-값** | **P-값** | **VIF** |
| 상수 | 4.5760 | 0.0736 | 62.15 | 0.000 |  |
| 세척 후 시간 | 0.01798 | 0.00826 | 2.18 | 0.029 | 1.00 |
| 온도 | -0.003285 | 0.000441 | -7.46 | 0.000 | 1.92 |
| 나사 크기 |  |  |  |  |  |
| 소 | -0.5444 | 0.0990 | -5.50 | 0.000 | 5.37 |
| 온도\*나사 크기 |  |  |  |  |  |
| 소 | 0.002804 | 0.000640 | 4.38 | 0.000 | 6.64 |

**모형 요약**

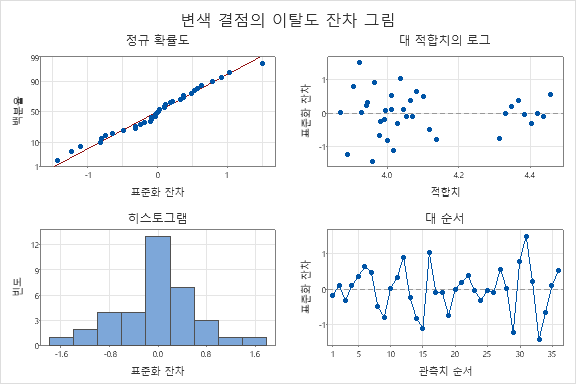
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **이탈도 R-Sq** | **이탈도 R-Sq(수정)** | **AIC** | **AICc** | **BIC** |
| 85.99% | 81.46% | 236.05 | 238.05 | 243.97 |

**적합도 검정**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **검정** | **DF** | **추정치** | **평균** | **카이-제곱** | **P-값** |
| 이탈도 | 31 | 12.36598 | 0.39890 | 12.37 | 0.999 |
| Pearson | 31 | 12.31611 | 0.39729 | 12.32 | 0.999 |

**분산 분석**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Wald 검정** | |
| **출처** | **DF** | **카이-제곱** | **P-값** |
| 회귀 | 4 | 78.77 | 0.000 |
| 세척 후 시간 | 1 | 4.74 | 0.029 |
| 온도 | 1 | 55.60 | 0.000 |
| 나사 크기 | 1 | 30.21 | 0.000 |
| 온도\*나사 크기 | 1 | 19.17 | 0.000 |



**1. 회귀 방정식**

* **변색 결점**은 나사 크기, 세척 후 시간, 온도에 따라 다음과 같이 예측됩니다:
  + **대형 나사**: Y′=4.576+0.01798×세척 후 시간−0.003285×온도Y' = 4.576 + 0.01798 \times \text{세척 후 시간} - 0.003285 \times \text{온도}Y′=4.576+0.01798×세척 후 시간−0.003285×온도
  + **소형 나사**: Y′=4.032+0.01798×세척 후 시간−0.000481×온도Y' = 4.032 + 0.01798 \times \text{세척 후 시간} - 0.000481 \times \text{온도}Y′=4.032+0.01798×세척 후 시간−0.000481×온도

변색 결점의 예측값은 위 식에서 계산된 Y′Y'Y′ 값을 기반으로 exp⁡(Y′)\exp(Y')exp(Y′)로 변환되어 얻어집니다.

**2. 계수 해석**

* **상수항**:
  + 대형 나사: 4.5760 (P-값: 0.000)
  + 이 값은 나사 크기가 "대형"일 때의 기준값을 의미합니다.
* **세척 후 시간**:
  + 계수: 0.01798 (P-값: 0.029)
  + 세척 후 시간이 증가할수록 변색 결점 수가 증가하는 경향이 있습니다. 이 값은 두 나사 크기 모두 동일하게 적용됩니다.
* **온도**:
  + 대형 나사 계수: -0.003285 (P-값: 0.000)
  + 소형 나사 계수: -0.000481
  + 온도가 증가할수록 변색 결점 수가 감소하는 경향이 있습니다. 이 효과는 대형 나사에서 더 크게 나타납니다.
* **나사 크기 (소형 나사)**:
  + 계수: -0.5444 (P-값: 0.000)
  + 소형 나사는 대형 나사에 비해 변색 결점 수가 적게 나타납니다.
* **온도\*나사 크기 상호작용**:
  + 계수: 0.002804 (P-값: 0.000)
  + 소형 나사에서 온도와의 상호작용 효과가 있으며, 온도가 변할 때 변색 결점 수에 미치는 영향이 상호작용을 통해 변할 수 있음을 시사합니다.

**3. 모형 요약**

* **이탈도 R-Squared**: 85.99% (수정된 R-Squared: 81.46%)
  + 모형이 데이터 변동의 85.99%를 설명할 수 있음을 나타냅니다. 매우 높은 설명력을 가지고 있습니다.
* **AIC (Akaike Information Criterion)**: 236.05
  + AIC 값은 모형의 적합도를 평가하는 데 사용되며, 값이 낮을수록 더 나은 모형임을 의미합니다.

**4. 적합도 검정**

* **이탈도**와 **Pearson 검정**에서 P-값이 0.999로 나타났으며, 이는 모델이 데이터에 잘 맞고 있음을 나타냅니다.

**5. 분산 분석 (Wald 검정)**

* **회귀 전체**: P-값 0.000으로 매우 유의미합니다.
* **세척 후 시간**: P-값 0.029로 유의미한 영향을 미칩니다.
* **온도**: P-값 0.000으로 매우 유의미한 영향을 미칩니다.
* **나사 크기**: P-값 0.000으로 매우 유의미한 영향을 미칩니다.
* **온도\*나사 크기 상호작용**: P-값 0.000으로 유의미한 상호작용이 존재함을 나타냅니다.

**결론**

이 포아송 회귀 분석 결과, 변색 결점에 대한 주요 영향 요인으로 세척 후 시간, 온도, 나사 크기가 모두 유의미하게 작용하고 있음을 알 수 있습니다. 특히, 온도와 나사 크기 간의 상호작용이 존재하여, 소형 나사에서는 온도 변화가 변색 결점에 더 큰 영향을 미칠 수 있음을 시사합니다.

이 분석을 통해 공정 관리에 있어 온도와 세척 후 시간을 적절히 조절하는 것이 중요하며, 나사 크기에 따라 이러한 변수들의 최적화를 고려해야 함을 알 수 있습니다. 이를 통해 변색 결점을 줄이고 제품 품질을 향상시킬 수 있습니다.

**결과 해석**

교호작용이 있는 모형의 경우, AIC는 약 236으로, 교호작용이 없는 모형보다 낮습니다. AIC 기준은 교호작용이 있는 모형이 교호작용이 없는 모형보다 나음을 나타냅니다. 전차 대 적합치 그림의 곡면성이 없어집니다. 엔지니어는 일부 계수가 > 5인 VIF 값을 가지고 있음을 알 수 있습니다. 이 경우, 공선화의 효과를 줄이기 위해 표준화된 연속 예측변수를 가진 분석은 모델내 용어의 통계적 유의성에 대해 동일한 결론을 내린다. (자세한 내용은 [회귀 분석의 다중 공선성](https://support.minitab.com/ko-kr/minitab/help-and-how-to/statistical-modeling/regression/supporting-topics/model-assumptions/multicollinearity-in-regression/)에서 확인하십시오.) 엔지니어는 교호작용이 없는 모형 대신 이 모형을 해석하기로 결정합니다.