

SEER数据分析报告

回归分析小组

2023 年 12 月 31 日

目录

1	问题背景	1
1.1	肺癌	1
1.2	对肺癌进行数据分析的重要意义	1
2	数据来源	1
2.1	基本信息	1
2.2	个人生活信息	2
2.3	2
3	描述性统计	2
4	数据建模	2
5	结论及建议	2

1 问题背景

1.1 肺癌

肺癌通常被认为是一种严重的癌症类型，因为它通常在早期没有明显的症状，导致在诊断时已经进展到晚期。由于这个原因，肺癌的治疗可能面临更大的挑战，治疗成功率相对较低。

1.2 对肺癌进行数据分析的重要意义

通过对肺癌患者的统计数据进行分析，可以识别患病的风险因素、疾病发展的模式以及不同人群之间的差异。这有助于预测患者的疾病风险，为预防和早期筛查提供信息。统计分析可以揭示不同治疗方法的效果，有助于优化治疗方案。了解患者对特定治疗的反应，可以个体化治疗计划，提高治疗的效果和患者的生存率。

2 数据来源

我们通过 SEER*Stat 软件下载了 1975 – 2020 年期间的肺癌数据 [1]。

2.1 基本信息

变量名
Patient ID
Age recode with single ages
Sex
Race recode(W,B,AI,API)

表 1: 基本信息变量

变量名
Rural Urban Continuum Code
Marital status at diagnosis
Median household income inflation adj to 2021

表 2: 个人生活变量

2.2 个人生活信息

2.3

3 描述性统计

4 数据建模

5 结论及建议

参考文献

- [1] SEER Program. Surveillance, epidemiology, and end results (seer) program. <http://www.seer.cancer.gov>, 2023. SEER*Stat Database: Incidence - SEER Research Data, 8 Registries, Nov 2021 Sub (1975-2020) - Linked To County Attributes - Time Dependent (1990-2020) Income/Rurality, 1969-2020 Counties, National Cancer Institute, DC-CPS, Surveillance Research Program, released April 2023, based on the November 2022 submission.