

糖尿病数据集变量解释

人口统计学特征

1. Age (年龄)

- 描述：患者的实际年龄（以年为单位）
- 重要性：年龄是糖尿病的重要风险因素，2型糖尿病风险随年龄增长而上升，尤其是45岁以上人群
- 数据类型：连续变量

2. Ethnicity (种族/民族)

- 描述：患者的种族或民族背景，通常以0-3的编码表示不同种族群体
- 重要性：某些种族群体（如非裔美国人、拉丁裔、亚裔、太平洋岛民等）糖尿病发病率更高
- 数据类型：分类变量

3. SocioeconomicStatus (社会经济状况)

- 描述：反映患者的经济和社会地位，通常以0-2的编码表示不同水平
- 重要性：低社会经济地位与糖尿病风险增加相关，可能受医疗资源获取、健康食品可及性等因素影响
- 数据类型：分类变量

4. EducationLevel (教育水平)

- 描述：患者的最高教育程度，以0-3的编码表示不同水平
- 重要性：教育水平可能影响健康意识、自我管理能力和医疗素养
- 数据类型：分类变量

身体状况和生活方式

5. BMI (体重指数)

- 描述：体重(kg)除以身高(m)的平方，用于评估体重状况
- 重要性：BMI \geq 25为超重， \geq 30为肥胖，肥胖是2型糖尿病的主要风险因素
- 数据类型：连续变量
- 正常范围：18.5-24.9

6. Smoking (吸烟状况)

- 描述：患者是否吸烟，0表示不吸烟，1表示吸烟
- 重要性：吸烟增加胰岛素抵抗和糖尿病风险，也加剧糖尿病并发症
- 数据类型：二分类变量

7. AlcoholConsumption (酒精消费量)

- 描述：患者的饮酒量，可能以标准饮酒单位或频率表示
- 重要性：过量饮酒可能增加糖尿病风险，而适度饮酒可能有保护作用
- 数据类型：连续变量

8. PhysicalActivity (身体活动水平)

- 描述：患者的运动频率和强度
- 重要性：规律运动可降低胰岛素抵抗，减少糖尿病风险，建议每周至少150分钟中等强度活动
- 数据类型：连续变量

9. DietQuality (饮食质量)

- 描述：评估患者饮食的健康程度
- 重要性：高纤维、低糖、低饱和脂肪的饮食有助于预防糖尿病
- 数据类型：连续变量

10. SleepQuality (睡眠质量)

- 描述：患者的睡眠质量评分
- 重要性：睡眠不足或质量差与胰岛素抵抗和糖尿病风险增加相关
- 数据类型：连续变量

医疗状况和生理指标

11. Hypertension (高血压)

- 描述：患者是否患有高血压，0表示无，1表示有
- 重要性：高血压与糖尿病高度相关，两者常共存于代谢综合征
- 数据类型：二分类变量

12. SystolicBP (收缩压)

- 描述：心脏收缩时的血压值，单位为mmHg
- 重要性：高血压是糖尿病的风险因素和常见并发症
- 数据类型：连续变量
- 正常范围：<120 mmHg

13. DiastolicBP (舒张压)

- 描述：心脏舒张时的血压值，单位为mmHg
- 重要性：与收缩压一起评估心血管健康状况
- 数据类型：连续变量
- 正常范围：<80 mmHg

14. FastingBloodSugar (空腹血糖)

- 描述：至少8小时未进食后的血糖水平，单位通常为mg/dL或mmol/L
- 重要性：糖尿病诊断的核心指标之一
- 数据类型：连续变量
- 正常范围：70-100 mg/dL (3.9-5.6 mmol/L)
- 前驱糖尿病：100-125 mg/dL (5.6-6.9 mmol/L)
- 糖尿病：≥126 mg/dL (≥7.0 mmol/L)

15. HbA1c (糖化血红蛋白)

- 描述：反映过去2-3个月平均血糖水平的指标，以百分比表示
- 重要性：糖尿病诊断和长期血糖控制的重要指标
- 数据类型：连续变量
- 正常范围：<5.7%
- 前驱糖尿病：5.7-6.4%
- 糖尿病：≥6.5%

16. SerumCreatinine (血清肌酐)

- 描述：肾功能的标志物，单位通常为mg/dL或μmol/L
- 重要性：评估肾功能，糖尿病肾病是常见并发症
- 数据类型：连续变量

- 正常范围：男性0.74-1.35 mg/dL，女性0.59-1.04 mg/dL

17. BUNLevels (血尿素氮水平)

- 描述：另一种肾功能指标，单位通常为mg/dL
- 重要性：与血清肌酐一起评估肾功能状态
- 数据类型：连续变量
- 正常范围：7-20 mg/dL

血脂指标

18. CholesterolTotal (总胆固醇)

- 描述：血液中所有类型胆固醇的总量，单位通常为mg/dL
- 重要性：血脂异常常与糖尿病共存
- 数据类型：连续变量
- 理想范围：<200 mg/dL

19. CholesterolLDL (低密度脂蛋白胆固醇)

- 描述：俗称"坏胆固醇"，单位通常为mg/dL
- 重要性：高LDL增加心血管疾病风险，糖尿病患者风险更高
- 数据类型：连续变量
- 理想范围：<100 mg/dL

20. CholesterolHDL (高密度脂蛋白胆固醇)

- 描述：俗称"好胆固醇"，单位通常为mg/dL
- 重要性：低HDL是代谢综合征和糖尿病的风险因素
- 数据类型：连续变量
- 理想范围：男性>40 mg/dL，女性>50 mg/dL

药物使用

21. Statins (他汀类药物使用)

- 描述：患者是否使用降脂药物，0表示不使用，1表示使用
- 重要性：评估血脂管理状况，也可能影响血糖水平
- 数据类型：二分类变量

症状

22. FrequentUrination (频繁排尿)

- 描述：患者是否有尿频症状，0表示无，1表示有
- 重要性：糖尿病的典型症状之一，由于高血糖导致肾脏过滤更多水分
- 数据类型：二分类变量

23. ExcessiveThirst (过度口渴)

- 描述：患者是否有明显口渴感，0表示无，1表示有
- 重要性：糖尿病的典型症状之一，常与尿频同时出现
- 数据类型：二分类变量

24. UnexplainedWeightLoss (原因不明的体重减轻)

- 描述：患者是否有非故意的体重减轻，0表示无，1表示有

- 重要性：1型糖尿病的典型症状，也可见于未控制的2型糖尿病
- 数据类型：二分类变量

25. FatigueLevels (疲劳程度)

- 描述：患者感到疲劳的程度
- 重要性：高血糖会导致能量利用效率下降，引起疲劳
- 数据类型：连续变量

26. BlurredVision (视力模糊)

- 描述：患者是否有视力模糊症状，0表示无，1表示有
- 重要性：高血糖可导致眼部晶状体暂时变形，引起视力变化
- 数据类型：二分类变量

生活质量和健康管理

27. QualityOfLifeScore (生活质量评分)

- 描述：评估患者整体生活质量的分数
- 重要性：反映疾病对患者日常生活的影响程度
- 数据类型：连续变量

28. MedicalCheckupsFrequency (体检频率)

- 描述：患者接受常规医疗检查的频率
- 重要性：反映健康意识和预防性医疗行为
- 数据类型：连续变量

29. MedicationAdherence (药物依从性)

- 描述：患者按照医嘱服用药物的程度
- 重要性：良好的药物依从性对疾病管理至关重要
- 数据类型：连续变量

30. HealthLiteracy (健康素养)

- 描述：患者理解和应用健康信息的能力
- 重要性：高健康素养有助于自我管理和预防疾病
- 数据类型：连续变量

数据应用建议

这组变量提供了全面的糖尿病风险评估框架，包括：

- 预测模型构建**：可用于开发糖尿病风险预测算法
- 患者分层**：识别高风险人群，进行针对性干预
- 早期筛查**：重点关注 FastingBloodSugar、HbA1c、BMI 等核心指标
- 干预重点**：可修改的风险因素如 PhysicalActivity、DietQuality、BMI 是干预的关键目标
- 综合评估**：结合生理指标、症状和生活方式因素进行全面评估

这些变量共同构成了一个多维度的糖尿病风险和管理评估体系，适用于临床决策支持和公共卫生干预规划。