短视频是指在各种新媒体上播放的、适合在移动状态和短时休闲状态下观看的、高频推送的视频内容，时长一般在5分钟以内。艾瑞咨询数据显示，2018年，中国短视频用户规模达5.01亿人，预计未来中国短视频行业用户规模仍将保持稳定增长态势。短视频的内容价值、传播热度、传播广度可通过点赞数等关键指标反映。本案例研究短视频点赞数的影响因素，从而加深人们对短视频偏好影响因素的理解，总结出一套能够有效提高短视频点赞数的可行手段，帮助运营者更好地做好账号与视频管理。本分析案例包含的短视频数据data.csv，共6400条观测，9个变量，其中每一列分别对应：序号、作者编号、点赞数、评论数、分享数、背景音乐、时长、发布日期、发布时间、类别、标题字数。数据说明表如下所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 变量类型 | | 变量名称 | 详细说明 | 取值范围 | 备注 |
| 因变量 | | 点赞数 | 定量变量 | 1.5-455.7 | 单位：万 |
| 自变量 | 视频 要素 | 视频时长 | 定量变量 | 4.4-60.0 | 单位：秒 |
| 标题字数 | 定量变量 | 0-72 |  |
| 发布日期 | 时间变量 | 2019/1/1-2019/5/9 | 单位：年/月/日 |
| 发布时间 | 时间变量 | 00:00-23:59 | 单位：时/分 |
| 视频类型 | 定性变量 | 美妆、游戏等7水平 |  |
| 背景音乐 | 文本变量 | 例：2018小恋曲\_沈虫虫 |  |
| 互动 要素 | 分享数 | 定量变量 | 0.0003-34.8 | 单位：万 |
| 评论数 | 定量变量 | 0.0028-12.0 | 单位：万 |

请完成以下分析任务：

1. 读入样本数据data.csv（提前设置好工作路径），命名为video。使用summary()函数查看数据记录的变量等情况。

2. 进行描述性分析，包括但不限于以下方面：

（1）绘制因变量点赞分布直方图，并进行适当解读；

（2）利用数据集中的发布时间变量，提取小时变量，命名为hour。使用折线图展示不同小时发布的短视频点赞数的差异，并进行适当解读；

（3）请使用分组箱线图对比不同类别的视频的点赞数差异，并进行适当解读。

3.由于汽车类短视频的点赞数远低于其他类别视频，对**非汽车类数据**进行线性回归建模分析。首先提取非汽车类数据命名为video1。再将视频发布时间分为6个时段：“20:00-次日5:59”、“6:00-8:59”、“9:00-11:59”、“12:00-13:59”、“14:00-16:59”、“17:00-19:59”，分别命名为“夜间”“早间”“上午”“中午”“下午”“傍晚”，将该变量命名为“视频发布时间”。建立点赞数的双对数线性回归模型，因变量为对数点赞数，自变量包括：视频发布时段（基准组：夜间）、对数分享数、对数评论数、时长、类别（基准组：宠物）、标题字数，并对结果进行适当解读。

4. 对第3问中建立的模型进行模型诊断，输出结果，并进行适当解读。