

作业 3

朱志儒 SA20225085

3.6

(a) 最小-最大规范化:

$$200: \frac{200-200}{1000-200} \times (1-0) + 0 = 0$$

$$300: \frac{300-200}{1000-200} \times (1-0) + 0 = 0.125$$

$$400: \frac{400-200}{1000-200} \times (1-0) + 0 = 0.25$$

$$600: \frac{600-200}{1000-200} \times (1-0) + 0 = 0.5$$

$$1000: \frac{1000-200}{1000-200} \times (1-0) + 0 = 1$$

(b) z 分数规范化:

数据组的平均值:

$$\frac{1}{5} \times (200 + 300 + 400 + 600 + 1000) = 500$$

数据组的方差:

$$\frac{1}{5} \times (200^2 + 300^2 + 400^2 + 600^2 + 1000^2) - 500^2 = 80000$$

数据组的标准差:

$$\sqrt{80000} \approx 282.8427$$

$$200: \frac{200-500}{282.8427} = -1.06066017$$

$$300: \frac{300-500}{282.8427} = -0.70710678$$

$$400: \frac{400-500}{282.8427} = -0.35355339$$

$$600: \frac{600-500}{282.8427} = 0.35355339$$

$$1000: \frac{1000-500}{282.8427} = 1.76776695$$

(c) z 分数规范化, 使用均值绝对偏差

均值绝对偏差:

$$\frac{1}{5} \times (|200 - 500| + |300 - 500| + |400 - 500| + |600 - 500| + |1000 - 500|)$$
$$= 240$$

$$200: \frac{200-500}{240} = -1.25$$

$$300: \frac{300-500}{240} = -0.8333$$

$$400: \frac{400-500}{240} = -0.4167$$

$$600: \frac{600-500}{240} = 0.4167$$

$$1000: \frac{1000-500}{240} = 2.0833$$

(d) 小数定标规范化:

$$200: \frac{200}{10000} = 0.02$$

$$300: \frac{300}{10000} = 0.03$$

$$400: \frac{400}{10000} = 0.04$$

$$600: \frac{600}{10000} = 0.06$$

$$1000: \frac{1000}{10000} = 0.1$$

3.7

(a) 最小-最大规模化:

$$\frac{35 - 13}{70 - 13} \times (1 - 0) + 0 = 0.3860$$

(b) age 的均值:

$$\frac{1}{27} \times (13 + 15 + 16 + 16 + 19 + 20 + 20 + 21 + 22 + 22 + 25 + 25 + 25 + 25$$
$$+ 30 + 33 + 33 + 35 + 35 + 35 + 35 + 36 + 40 + 45 + 46 + 52$$
$$+ 70) = 29.96$$

z 分数规范化:

$$\frac{35 - 29.96}{12.94} = 0.3895$$

(c) 小数定标规范化:

$$\frac{35}{100} = 0.35$$

(d) 对于给定的数据,我愿意采用最小-最大规模化,因为该规范化的方法保持原始数据值之间的联系,并且不会出现负值,年龄不能为负值。