

统计案例

【学习目标】

- 1.能对已知数据进行基础的分析，掌握平均数、方差、中位数、标准差、极差的概念，能画出频率分布直方图，能够撰写简单的统计分析报告；
- 2.会用统计的方法分析与解决问题，体会各种数字特征的作用，感受图表在数据分析中的直观性作用；
- 3.经历数据—图表分析—数字特征分析—结论的过程，体会统计方法对实际问题中隐含信息的挖掘作用，提升数据分析的素养。

【学习指导】

在一个统计问题中，如何依据现有的数据挖掘隐含的信息，从而提供切实可行的建议？

【案例】近年来，我国肥胖人群的规模急速增长，肥胖人群有很大的心血管安全隐患.目前，国际上常用身体质量指数(Body Mass Index，缩写 BMI)来衡量人体胖瘦程度以及是否健康，其计算公式是 $BMI = \frac{\text{体重} (kg)}{\text{身高}^2 (m^2)}$ ，中国成人的 BMI 数值标准为: $BMI < 18.5$ 为偏瘦； $18.5 \leq BMI < 23.9$ 为正常； $24 \leq BMI < 27.9$ 为偏胖； $BMI \geq 28$ 为肥胖.

为了解某公司员工的身体肥胖情况，研究人员从公司员工体检数据中，采用比例分配的分层随机抽样方法抽取了 90 名男员工、50 名女员工的身高和体重数据，计算得到他们的 BMI 值如下：

男员工	23.5	21.6	30.6	22.1	23.7	20.6	24.0	23.9	20.8	21.5
	22.1	21.6	19.0	20.2	19.6	17.3	17.9	23.4	18.7	23.1
	17.3	22.4	20.8	25.1	21.3	27.7	23.5	23.6	19.4	23.1
	18.6	24.1	21.3	19.5	18.7	21.0	22.6	16.0	18.0	17.9
	22.1	19.3	19.3	22.8	29.0	21.4	22.3	18.8	19.7	27.4
	23.5	23.6	30.5	22.3	21.6	17.6	21.5	29.1	25.5	18.7
	22.1	18.9	25.8	27.8	35.3	17.5	27.0	19.9	22.2	24.5
	18.0	19.0	21.1	21.3	18.7	23.9	20.8	34.2	16.6	19.3
	20.9	23.7	23.7	23.0	18.7	27.3	21.2	17.3	23.5	30.1
女员工	21.8	18.2	25.2	28.1	21.5	19.1	25.7	24.4	17.6	20.8
	20.5	20.2	17.4	21.6	18.4	20.3	30.8	23.6	23.3	22.8
	20.8	16.8	19.0	16.4	18.7	26.1	20.2	17.6	15.4	21.5
	19.5	31.6	19.1	20.4	13.9	18.6	16.6	15.9	18.3	18.0
	29.7	18.9	16.9	25.8	19.8	18.5	16.0	17.6	19.1	26.5

二、任务与要求

根据上面的数据，写一份该公司员工肥胖情况的统计分析报告.要求

- 1. 比较男、女员工在肥胖状况上的差异；
- 2. 分析公司员工胖瘦程度的整体情况；
- 3. 提出控制体重的建议

三、统计分析报告的主要组成部分

- 1.标题
- 2.前言

简单交代调查的目的、方法、范围等背景情况，使读者了解调查的基本情况.

3.主体

展示数据分析的全过程:首先要明确所关心的问题是什么,说明数据蕴含的信息;根据数据分析的需要,说明如何选择合适的图表描述和表达数据;从样本数据中提取能刻画其特征的量,如均值、方差等,用于比较男、女员工在肥胖状况上的差异;通过样本估计总体的统计规律,分析公司员工胖瘦程度的整体情况

4.结尾

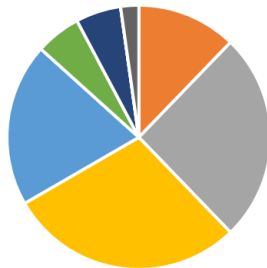
对主体部分的内容进行概括,结合控制体重的一般方法(可以查阅有关文献),提出控制公司员工体重的建议

一、数据的直观性分析

问题 1. 如何让数据所反映的信息更加直观?

任务一、饼图与折线图

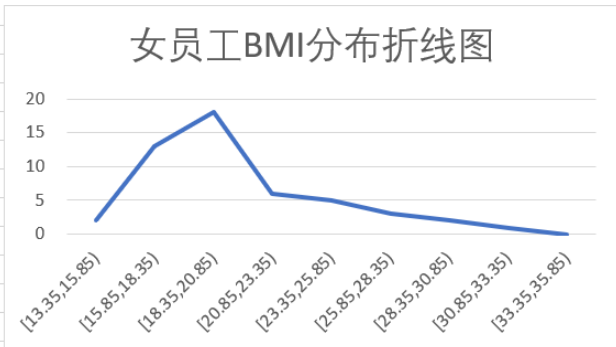
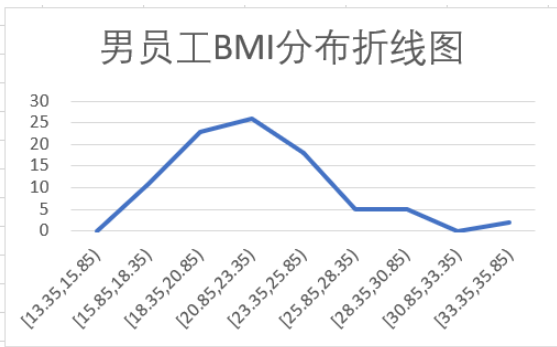
男员工BMI分布饼形图



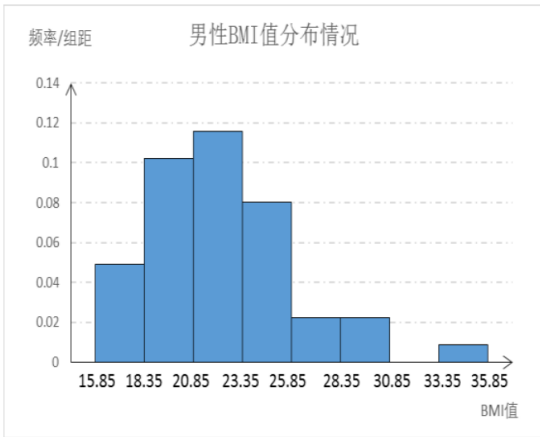
女员工BMI分布饼形图



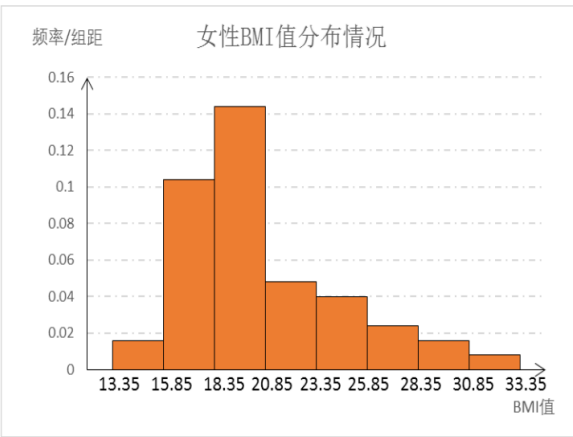
■ [13.35,15.85) ■ [15.85,18.35) ■ [18.35,20.85) ■ [20.85,23.35) ■ [23.35,25.85) ■ [13.35,15.85) ■ [15.85,18.35) ■ [18.35,20.85) ■ [20.85,23.35) ■ [23.35,25.85)
■ [25.85,28.35) ■ [28.35,30.85) ■ [30.85,33.35) ■ [33.35,35.85) ■ [25.85,28.35) ■ [28.35,30.85) ■ [30.85,33.35) ■ [33.35,35.85)



任务二、频率分布直方图



(1)



(2)

问题 2. 几种统计图各适用于什么情境？

- 折线图更能反映 BMI 变化的趋势；
- 饼图与频率分布直方图更能体现 BMI 在各个区间的分布情况；
- 频率分布直方图更方便做两组数据的对比；

问题 3. 从统计图中你能得到哪些信息？

- 男员工 BMI 值在区间(15.85, 25.85]内比较多；
- 数据较集中，大于 25.85 的较少；
- 而女员工的 BMI 值主要集中在(15.85, 20.85]内，后面呈阶梯式下降.
- 总的来说男员工的 BMI 值要比女员工的 BMI 值大一些；
- 男员工 BMI 值没有在区间(13.35, 15.85]内的，即没有偏瘦的；
- 男女员工的频率分布直方图都不对称，峰值都是左偏的，即男、女员工中都有偏胖的.

小结：

- 数据堆积不容易找到其中的规律，可将数据放入统计图中作直观化处理；
- 折线图更能反应 BMI 变化的趋势，饼图与频率分布直方图更能体现 BMI 在各个区间的分布情况；

二、数据的数字特征分析

问题 2. 数据的哪些信息可以反映男、女员工在肥胖状态上的差异？

任务一、计算男女员工的各个数字特征值；

	中位数	平均数	标准差	方差	极差
男员工	21.6	22.18	3.78	14.32	19.3
女员工	19.65	20.70	4.06	16.47	17.7

任务二、每组的百分比分布

	偏瘦	正常	偏胖	肥胖
男员工	12%	68%	12%	8%
女员工	32%	48%	12%	8%

问题 2. 从各个数据信息中你能得到哪些有价值的信息？

- 可以看出男员工的中位数和平均数都比女员工的大，但都在正常值范围之内.
- 男员工的 BMI 值变化范围比女员工的变化范围大，这是由某个极端值引起的，
- 男员工 BMI 的最大值为 35.3，已经达到了重度肥胖的标准；
- 从标准差上看，男员工整体的分散程度比女员工略小.
- 男、女员工偏胖和肥胖的比例差不多，但女员工偏瘦的比例较大，这可能与女性更追求身材好有关.

三、样本估计总体

分析公司员工胖瘦的整体情况

问题 3. 如何由样本数据估计整体数据？

[要点 1]可由几部分数据的均值、方差计算整体的均值与方差

[要点 2]均值与方差可以反映

$$\bar{x} = \frac{1}{n} \sum_{j=1}^2 n_j \bar{x}_j, \quad s^2 = \frac{n_1}{n} (s_1^2 + (\bar{x}_1 - \bar{x})^2) + \frac{n_2}{n} (s_2^2 + (\bar{x}_2 - \bar{x})^2).$$

$$\begin{aligned}\bar{x} &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n} \\ &= \frac{22.18 \times 90 + 20.7 \times 50}{140} = \frac{3031.2}{140} = 21.65.\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{1}{140} \{90 \times [14.32 + (22.18 - 21.65)^2] + 50 \times [16.47 + (20.7 - 21.65)^2]\} \\ &= 15.59.\end{aligned}$$

平均数为 $\bar{x} = 21.65$ ，这个值在正常值范围内，可见这个公司员工 BMI 值的平均水平是正常的，方差为 15.59，得标准差 $s = 3.95$ ，

$$\bar{x} - 2s = 13.75, \quad \bar{x} + 2s = 29.55.$$

可见，绝大部分员工的 BMI 在区间[13.75, 29.55]内，还有个别员工的 BMI 值属于肥胖但比例很小.

问题 4. 提出控制体重的建议。

[提示]

员工应注意保持体重；

个别肥胖的员工可适当加强训练；

偏瘦的女员工可为健康着想，适当增重；

四、统计分析报告的撰写

问题 1. 统计分析报告需要有哪些环节？

[提示]标题、前言、主体、结尾

问题 2. 主体应当包含哪些部分？

[提示]

1、对数据的直观性分析及其结论；

2、对数据的数字特征分析及其结论；

3、针对实际问题给出的建议；