**Bài tập chương 2: Trình bày và mô tả dữ liệu**

**Dạng 1: Bài tập vẽ biểu đồ**

**Bài 1:** Chính phủ Mỹ chi ra hơn 2.5 nghìn tỷ đô la Mỹ trong năm 2006. Bảng dưới đây liệt kê các khoản chi tiêu chính của chính phủ trong năm qua:

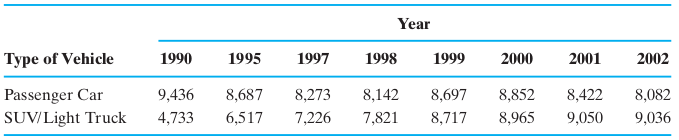
|  |  |
| --- | --- |
| **Các khoản chi tiêu** | **Số tiền đã chi (tỷ đô la)** |
| Quốc phòng | 525 |
| An ninh xã hội | 500 |
| Y tế | 500 |
| Trả lãi nợ quốc gia | 300 |
| Các chương trình xã hội | 200 |
| Các chi phí khác | 475 |

a. Vẽ biểu đồ tròn với các dữ liệu trên

b. Vẽ biểu đồ cột với các dữ liệu trên

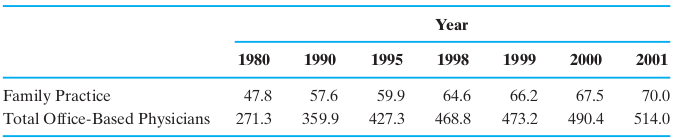
c. Vẽ lại 2 biểu đồ trên sử dụng giá trị phần trăm thay cho số tiền

**Bài 2:** Một thay đổi lớn dường như đang diễn ra đối với loại phương tiện mà người dân đang mua. Cục Phân tích Kinh tế Hoa Kỳ trong ấn phẩm Survey of Current Business (tháng 2 năm 2002) cung cấp dữ liệu được đưa ra trong bảng sau. Các con số được báo cáo là với số lượng hàng nghìn chiếc — tức là 9.436 tương ứng với 9.436.000 xe bán ra trong năm 1990.

a. Có nên sử dụng biểu đồ tròn để hiển thị các số liệu trên. Tại sao?

b. Hãy vẽ biểu đồ cột thể hiện sự thay đổi trong vòng 12 năm về sự thay đổi của công chúng khi lựa chọn xe

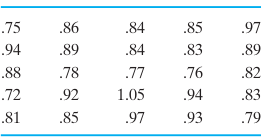
**Bài 3:** Người ta đã báo cáo rằng đã có một sự thay đổi trong loại hình hành nghề của các bác sĩ lựa chọn cho sự nghiệp của họ. Đặc biệt, có lo ngại sẽ thiếu gia đình. hành nghề bác sĩ trong những năm tới. Bảng sau đây chứa dữ liệu về tổng số văn phòng- bác sĩ dựa trên cơ sở và số lượng những bác sĩ tuyên bố mình là hành nghề gia đình các thầy thuốc. Các con số trong bảng được đưa ra bằng hàng nghìn bác sĩ. (Nguồn: Statistical Tóm tắt của Hoa Kỳ: 2003)

a. Sử dụng biểu đồ cột để biểu diễn sự tăng lên của số bác sĩ gia đình từ năm 1990 đến 2002

b. Hãy tính tỷ lệ số bác sĩ gia đình trên số bác sĩ văn phòng và hiển thị số liệu này thông qua biểu đồ cột

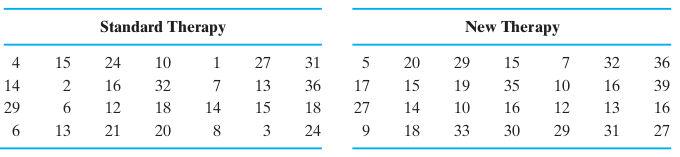
c. Liệu có sự khác biệt nào giữa hai biểu đồ trên hay không?

**Bài 4:** Các quy định của hội đồng y tế ở một tiểu bang cụ thể quy định rằng mức florua không được vượt quá 1,5 phần triệu (ppm). 25 phép đo được đưa ra ở đây đại diện cho mức florua cho một mẫu trong 25 ngày. Mặc dù mức florua được đo nhiều hơn một lần mỗi ngày, những dữ liệu này đại diện cho các bài đọc vào buổi sáng sớm trong 25 ngày được lấy mẫu



Vẽ biểu đồ histogram cho dữ liệu này

**Bài 5:** Thời gian sống sót (tính bằng tháng) đối với hai phương pháp điều trị cho bệnh nhân trái mãn tính nặng- suy tim thất được đưa ra trong bảng sau.



Vẽ hai biểu đồ histogram cho thời gian sống sót đối với hai phương pháp điều trị bệnh. Liệu phương pháp điều trị mới có kéo dài thời gian sống? Giải thích?

**Bài 6:** Datasets AHCAVote2017 (gói Stat2Data)

   1. Hãy xây dựng bảng tần số về số thành viên của quốc hội theo từng bang

   2. Xây dựng bảng tần số giữa kết quả phiếu bầu đạo luật này (AHCAvote) với kết quả bầu tổng thống trước đó (Trump)

   3. Hãy vẽ biểu đồ cột về số thành viên quốc hội theo đảng Dân chủ (Democrat) và đảng Cộng hòa (Republican)

   4. Hãy vẽ biểu đồ tán xạ giữa cột uniChange và AHCAvote

**Bài 7:** Datasets BirdNest (gói Stat2Data)

   1. Xây dựng bảng tần số về số loài chim theo từng loại tổ (Nesttype)

   2. Loài chim nào có chiều dài trung bình lớn nhất, nhỏ nhất?

   3. Vẽ biểu đồ cột thể hiện vị trí làm tổ của các loài chim

   4. Hãy tính thời gian trung bình các loài chim này chăm sóc chim non trong tổ

**Bài 8:** Datasets BlueJays (gói Stat2Data)

   1. Có bao nhiêu con chim đực, chim cái trong mẫu dữ liệu trên?

   2. Tính trung bình và độ lệch chuẩn các thông số của loài chim trên (BillDepth, BillWidth, BillLength, Head, Mass, Skull)

   3. Vẽ biểu đồ histogram về trọng lượng của loài chim trên

   4. Vẽ biểu đồ tán xạ giữa BillWidth và BillLength

**Bài 9:** Datasets BritishUnions (gói Stat2Data)

    1. Vẽ biểu đồ thể hiện tỷ lệ thất nghiệp theo thời gian.

    2. Vẽ biểu đồ tán xạ giữa các thuộc tính AgreePct, DisagreePct, NetSupport với Unemployment.

**Bài 10:** Datasets CAFE (gói Stat2Data)

    1. Hãy thống kê số thành viên ở Thượng viện theo từng bang

    2. Điều luật trên có được thông qua hay không?

    3. Hãy vẽ biểu đồ cột thể hiện kết quả bỏ phiếu theo từng đảng (Đảng Dân Chủ / Đảng Cộng Hòa)

    4. Hãy vẽ biểu đồ hộp thể hiện mức đóng góp cho ngành công nghiệp ô tô với những thành viên không thông qua dự luật trên.

**Bài 11:** Datasets Clothing (gói Stat2Data)

    1. Lập bảng tần số thể hiện số khách hàng có thẻ thành viên (Card = 1) và không có thẻ thành viên.

    2. Tính số tiền trung bình mà 1 khách hàng bỏ ra trong lần mua gần đây nhất với nhà bán lẻ này

    3. Hãy tính số tiền trung bình trong 1 lần thanh toán của 60 khách hàng này trong 12 và 24 tháng qua

    4. Vẽ biểu đồ cột thể hiện số tiền trung bình trong 1 lần thanh toán giữa các khách hàng có thẻ thành viên (Card = 1) và không có thẻ.

**Bài 12:** Datasets CloudSeeding2 (gói Stat2Data)

    1. Tính lượng mưa trung bình tại các khu vực trên khi Seeded = U

    2. Tính lượng mưa trung bình tại các khu vực trên theo mùa khi Seed = U

    3. Tính lượng mưa tại các khu vực TE và TW khi Seeded = S

    4. Tính lượng mưa tại các khu vực TE và TW theo mùa khi Seeded = S

    5. So sánh các kết quả trên và đưa ra kết luận.

**Bài 13:** Datasets Diamonds (gói Stat2Data)

    1. Tính giá trị trung bình của 1 viên kim cương.

    2. Tính giá trị trung bình của 1 carat kim cương.

    3. Vẽ biểu đồ histogram về trọng lượng của các viên kim cương,

    4. Vẽ biểu đồ thể hiện tổng khối lượng kim cương theo từng màu

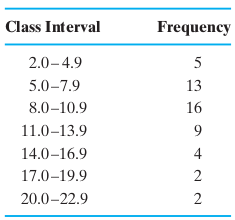
    5. Vẽ biểu đồ tán xạ giữa Depth và PricePerCt, Depth và TotalPrice.

**Dạng 2: Tính toán các số đặc trưng**

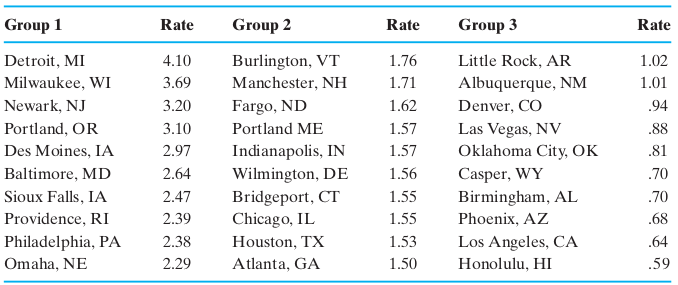
**Bài 1:** Tính mean, median và mode với dữ liệu sau:



**Bài 2:** Tính mean, median, mode cho các số liệu trong bảng sau:



**Bài 3:** Thuế suất hiệu dụng (trên 100 đô la) đối với bất động sản nhà ở cho ba nhóm thành phố lớn, được xếp hạng theo thuế suất bất động sản nhà ở, được thể hiện trong bảng sau.

Hãy tính mean, median và mode của từng nhóm, cũng như của toàn bộ 30 số liệu trên.

**Dạng 3: Phát hiện giá trị bất thường bằng biểu đồ hộp**

**Bài 1:** Hãy tìm median, các tứ phân vị thứ nhất và tứ phân vị thứ ba của dãy số liệu sau: 13, 21, 9, 15, 13, 17, 21, 9, 19, 23, 11, 9, 21

**Bài 2:** Vẽ boxplot với các dãy số liệu sau:

a) 1, 0, 3, 2, 1, 1, 7, 8, 6, 6, 7, 7

b) 16, 55, 78, 80, 81, 82, 84, 86, 88, 90, 105

c) 90, 77, 79, 60, 87, 87, 80, 80, 83