

BÀI TẬP THỐNG KÊ MÔ TẢ (THƯ VIỆN STATISTICS)

Câu 1

Điều tra năng suất lúa vụ đông - xuân năm 1990 của 31 tỉnh, thành phố từ Nghệ An trở vào, người điều tra lập được bảng sau:

STT	Tỉnh, thành phố	Năng suất (tạ/ha)
1	Nghệ An	30
2	Hà Tĩnh	30
3	Quảng Bình	20
4	Quảng Trị	25
5	Thừa Thiên - Huế	25
6	Đà Nẵng	45
7	Quảng Nam	40
8	Quảng Ngãi	40
9	Bình Định	35
10	Phú Yên	50
11	Khánh Hòa	45
12	TP. Hồ Chí Minh	35
13	Lâm Đồng	25
14	Ninh Thuận	45
15	Tây Ninh	30
16	Bình Dương	30
17	Đồng Nai	30
18	Bình Thuận	40
19	Bà Rịa - Vũng Tàu	30
20	Long An	25
21	Đồng Tháp	35
22	An Giang	35
23	Tiền Giang	45
24	Vĩnh Long	35

STT	Tỉnh, thành phố	Năng suất (tạ/ha)
25	Bến Tre	35
26	Kiên Giang	35
27	Cần Thơ	30
28	Trà Vinh	40
29	Sóc Trăng	40
30	Bạc Liêu	40
31	Cà Mau	35

- Lập bảng tần số?
- Vẽ biểu đồ cột biểu diễn sự phân bố tần suất?
- Tính số trung bình, độ lệch chuẩn, số trung vị và mốt?

Câu 2

Một nhà nghiên cứu ghi lại tuổi của 30 bệnh nhân mắc bệnh đau mắt hột. Kết quả thu được mẫu số liệu như sau:

Mẫu số liệu bệnh nhân đau mắt hột									
21	17	22	18	20	17	15	13	15	20
15	12	18	17	25	17	21	15	12	18
16	23	14	18	19	13	16	19	18	17

- Lập bảng tần số?
- Vẽ biểu đồ cột biểu diễn sự phân bố tần suất?
- Tính số trung bình, độ lệch chuẩn, số trung vị, tứ phân vị và mốt?

Câu 3

Trong kì thi Tiếng Anh, điểm thi của 32 học sinh (thang điểm 100) như sau:

Mẫu số liệu điểm thi Tiếng Anh									
68	79	65	85	52	81	55	65	49	42
68	66	56	57	65	72	69	60	50	63
74	88	78	95	41	87	61	72	59	47
90	74								

- Lập bảng tần số?
- Vẽ biểu đồ cột biểu diễn sự phân bố tần suất?
- Tính số trung bình, độ lệch chuẩn, số trung vị, tứ phân vị và mốt?
- Hãy trình bày mẫu số liệu trên dưới dạng bảng phân số tần số ghép lớp với các nửa khoảng: $[40, 50)$, $[50, 60)$, ..., $[90, 100)$?

Câu 4

Bác Dũng và bác Thu ghi lại số cuộc điện thoại mà mỗi người gọi mỗi ngày trong 10 ngày được lựa chọn ngẫu nhiên từ tháng 01/2021 ở bảng sau:

Bác Dũng	2	7	3	6	1	4	1	4	5	1
Bác Thu	1	3	1	2	3	4	1	2	20	2

- Hãy tìm số trung bình, độ lệch chuẩn, số trung vị, tứ phân vị và mốt của số cuộc điện thoại mà mỗi bác gọi theo số liệu trên?
- Nếu so sánh theo số trung bình thì ai gọi nhiều cuộc điện thoại hơn?
- Nếu so sánh theo số trung vị thì ai gọi nhiều cuộc điện thoại hơn?
- Theo bạn, nên dùng số trung bình hay số trung vị để so sánh xem ai gọi nhiều cuộc điện thoại hơn mỗi ngày?

Câu 5

Tổng số điểm mà các thành viên đội tuyển Olympic Toán quốc tế (IMO) của Việt Nam đạt được trong 20 kì thi được cho ở bảng sau:

Năm	Tổng điểm	Năm	Tổng điểm	Năm	Tổng điểm
2020	150	2013	180	2006	131
2019	177	2012	148	2005	143
2018	148	2011	113	2004	196
2017	155	2010	133	2003	172
2016	151	2009	161	2002	166
2015	151	2008	159	2001	139

Năm	Tổng điểm	Năm	Tổng điểm	Năm	Tổng điểm
2014	157	2007	168		

Có ý kiến cho rằng điểm thi của đội tuyển giai đoạn 2001 – 2010 cao hơn giai đoạn 2011 – 2020. Hãy sử dụng số trung bình và số trung vị để kiểm nghiệm xem ý kiến trên có đúng không?

Câu 6

Bạn Châu cân lần lượt 50 quả vải thiều Thanh Hà được lựa chọn ngẫu nhiên từ vườn nhà mình và được kết quả như sau:

Cân nặng (gram)	Số quả	Cân nặng (gram)	Số quả
8	1	21	17
19	10	22	3
20	19		

Tính số trung bình, độ lệch chuẩn, số trung vị, tứ phân vị và một?

Câu 7

Một cửa hàng bán xe ô tô thay đổi chiến lược kinh doanh vào cuối năm 2019. Số xe của hàng bán được mỗi tháng trong năm 2019 và 2020 được ghi lại ở bảng sau:

Tháng	Năm 2019	Năm 2020
1	54	45
2	22	28
3	24	31
4	30	34
5	35	32
6	40	35
7	31	37
8	29	33
9	29	33
10	37	35
11	40	34
12	31	37

- Tính số trung bình, độ lệch chuẩn, số trung vị, tứ phân vị và một của số lượng xe bán được trong năm 2019 và năm 2020?
- Vẽ biểu đồ cột đôi biểu diễn số xe của hàng bán được mỗi tháng trong năm 2019 và 2020?
- Nêu nhận xét về tác động của chiến lược kinh doanh mới lên số lượng xe bán ra hàng tháng?

Câu 8

Bảng dưới đây thống kê tổng số giờ nắng trong năm 2019 theo từng tháng được đo bởi hai trạm quan sát khí tượng đặt ở hai tỉnh Tuyên Quang và Cà Mau:

Tháng	1	2	3	4	5	6	7	8	9	10	11	12
Tuyên Quang	25	89	72	117	106	177	156	203	227	146	117	145
Cà Mau	180	223	257	245	191	111	141	134	130	122	157	173

- Tính số trung bình, độ lệch chuẩn, số trung vị, tứ phân vị và một?
- Vẽ biểu đồ cột đôi biểu diễn số giờ nắng theo từng tháng của hai tỉnh?
- Nêu nhận xét về sự thay đổi tổng số giờ nắng theo từng tháng ở mỗi tỉnh?

Câu 9

Bài báo trong tháng 1 năm 1990 của Arizona Trend có bổ sung mô tả 12 cú đánh gôn "tốt nhất" trong tiểu bang Arizona. Độ dài (đơn vị tính: yard) của các cú đánh gôn này như sau:

12 cú đánh gôn "tốt nhất" trong tiểu bang Arizona					
6981	7099	6930	6992	7518	7100
6935	7518	7013	6800	7041	6890

Tính giá trị trung bình mẫu, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?

Câu 10

Tám phép đo được thực hiện trên đường kính bên trong của vòng

piston rèn được sử dụng trong một động cơ ô tô. Dữ liệu (tính bằng milimét) được ghi trong bảng sau:

Kết quả đo đường kính bên trong của vòng piston rèn							
74.001	74.003	74.015	74.000	74.005	74.002	74.005	74.004

- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?
- Vẽ biểu đồ hộp biểu diễn sự phân bố của số liệu và đưa ra nhận xét về dữ liệu?

Câu 11

"Thời gian bắt lửa lạnh" của một động cơ ô tô đang được nhà sản xuất xăng điều tra. Các quan trắc sau (tính bằng giây) thu được cho một phương tiện thử nghiệm:

Thống kê "thời gian bắt lửa lạnh" của động cơ ô tô							
1.75	1.92	2.62	2.35	3.09	3.15	2.53	1.91

- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?
- Vẽ biểu đồ hộp thể hiện sự phân bố dữ liệu?

Câu 12

Một nhà sản xuất các thành phần điện tử quan tâm đến việc xác định tuổi thọ của một loại pin. Một mẫu, tính theo giờ, như sau:

Tuổi thọ của mẫu pin									
123	116	122	110	175	126	125	111	118	117

- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?
- Đặc điểm nào trong tập dữ liệu này là nguyên nhân của sự khác biệt giữa hai đặc trưng ở câu a?

Câu 13

Các số đo sau được ghi lại cho thời gian khô (theo giờ) của một loại sơn nước nào đó:

Thời gian khô của một loại sơn							
3.4	2.5	4.8	2.9	3.6	2.8	3.3	5.6
3.7	2.8	4.4	4.0	5.2	3.0	4.8	

- Cỡ mẫu của mẫu trên là bao nhiêu?
- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?
- Vẽ biểu đồ hộp biểu diễn mẫu trên?

Câu 14

Trong một nghiên cứu được thực hiện bởi Khoa Kỹ thuật Cơ khí tại Học viện Kỹ thuật Virginia, các thanh thép được cung cấp bởi hai công ty khác nhau được so sánh. 10 mẫu lò xo được làm từ các thanh thép được cung cấp từ mỗi công ty, và một phép đo về độ đàn hồi được ghi lại cho mỗi lò xo. Dữ liệu như sau:

Công ty A:	9.3	8.8	6.8	8.7	8.5
	6.7	8.0	6.5	9.2	7.0
Công ty B:	11.0	9.8	9.9	10.2	10.1
	9.7	11.0	11.1	10.2	9.6

- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn cho dữ liệu của hai công ty?
- Vẽ dữ liệu của hai công ty trên cùng một đường thẳng và nêu nhận xét về bất kỳ sự khác biệt dữ liệu nào giữa hai công ty?

Câu 15

Đo lượng cholesterol (đơn vị mg%) cho một số người, ta thu được bảng tần số dưới đây:

X (mg%)	150–160	160–170	170–180	180–190	190–200	200–210
Số người	2	4	5	6	4	3

- Có bao nhiêu người được khảo sát?
- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?

- c. Một mẫu thứ nhì Y có 30 người cho trung bình 180 mg% và độ lệch chuẩn 16 mg%. Nhập hai mẫu lại, tính trung bình và độ lệch chuẩn của mẫu nhập?

Câu 16

Độ bền của cao su silicone được xem là một hàm theo nhiệt độ sấy. Một nghiên cứu được thực hiện, trong đó 12 mẫu cao su được chuẩn bị, sử dụng các nhiệt độ sấy ở mức 20°C và 45°C. Dữ liệu bên dưới thể hiện các giá trị độ bền (tính theo megapascals):

20°C:	2.07	2.14	2.22	2.03	2.21	2.03
	2.05	2.18	2.09	2.14	2.11	2.02
45°C:	2.52	2.15	2.49	2.03	2.37	2.05
	1.99	2.42	2.08	2.42	2.29	2.01

- a. Vẽ biểu đồ hộp biểu diễn dữ liệu với các giá trị độ bền ở nhiệt độ cao và thấp?
- b. Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của độ bền cho cả hai mẫu?
- c. Có thấy được nhiệt độ sấy có tác động lên độ bền dựa vào đồ thị hay không? Nhận xét thêm?

Câu 17

Một loại polymer nào đó được sử dụng cho các hệ thống chân không cho máy bay. Điều quan trọng là polymer phải chịu được quá trình lão hóa. 29 mẫu vật polymer được dùng trong một thí nghiệm. 10 mẫu được gán một cách ngẫu nhiên để tiếp xúc với một quá trình lão hóa từng đợt tăng cường liên quan đến việc tiếp xúc với nhiệt độ cao trong 10 ngày. Các đo đạc độ bền của các mẫu vật được ghi lại, và dữ liệu sau đây được ghi lại về độ bền theo psi:

Không lão hóa:	227	222	218	217	225
	218	216	229	228	221
Lão hóa:	219	214	215	211	209
	218	203	204	201	205

- Vẽ biểu đồ hộp biểu diễn dữ liệu?
- Từ đồ thị trên, có thấy được quá trình lão hóa đã tác động lên độ bền của polymer này hay không? Vì sao?
- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mẫu?
- Thảo luận về tính tương đồng hoặc không tương đồng giữa trung bình và trung vị cho mỗi nhóm?

Câu 18

Một nghiên cứu về tác động của việc hút thuốc lên giấc ngủ được thực hiện. Phép đo được quan trắc là thời gian (theo phút) cần để rơi vào giấc ngủ. Các dữ liệu này là:

Nhóm hút thuốc:	69.3	56.0	22.1	47.6	53.2	48.1
	52.7	34.4	60.2	43.8	23.2	13.8
Nhóm không hút thuốc:	28.6	25.1	26.4	34.9	29.8	28.4
	38.5	30.2	30.6	31.8	41.6	21.1
	36.0	37.9	13.9			

- Tính trung bình, trung vị, tứ phân vị, phương sai và độ lệch chuẩn của mỗi nhóm?
- Vẽ biểu đồ hộp biểu diễn hai tập dữ liệu trên cùng một trục tọa độ?
- Nhận xét về sự tác động của việc hút thuốc đến thời gian cần để rơi vào giấc ngủ?

Câu 19

- Với một mẫu quan trắc x_1, x_2, \dots, x_n chứng minh rằng:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

- Phát sinh mẫu dữ liệu ngẫu nhiên gồm 1000 quan sát $x_1, x_2, \dots, x_{1000}$ và kiểm nghiệm công thức trên theo các trường hợp sau:
 - Mẫu theo phân phối chuẩn?
 - Mẫu theo phân phối chuẩn tắc?
 - Mẫu theo phân phối đều?
 - Mẫu theo phân phối nhị phân?

Câu 20

Một nghiên cứu về khối lượng đường bị biến đổi (y) (theo mg) trong một quá trình nào đó ở các nhiệt độ (x) (theo $^{\circ}\text{C}$) khác nhau. Cho các dữ liệu được mã hóa và ghi lại như sau:

Nhiệt độ ($^{\circ}\text{C}$)	Khối lượng đường bị biến đổi (mg)
1.0	8.1
1.1	7.8
1.2	8.5
1.3	9.8
1.4	9.5
1.5	8.9
1.6	8.6
1.7	10.2
1.8	9.3
1.9	9.2
2.0	10.5

- Tính hệ số tương quan giữa nhiệt độ và khối lượng đường bị biến đổi? Nêu nhận xét về mối tương quan giữa hai đại lượng này?
- Ước lượng đường hồi quy tuyến tính? Vẽ các điểm dữ liệu và đường hồi quy trên cùng một đồ thị? Ước lượng khối lượng trung bình của đường bị biến đổi được tạo ra khi nhiệt độ được mã hóa là 1.75°C ?

Câu 21

Điểm thi giữa kỳ (x) và điểm thi cuối kỳ (y) của một lớp có 9 sinh viên được ghi lại trong bảng sau:

x	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	68

- Tính hệ số tương quan giữa điểm thi giữa kỳ và điểm thi cuối kỳ? Nêu nhận xét về mối tương quan giữa hai đại lượng này?

- b.** Ước lượng đường hồi quy tuyến tính? Vẽ các điểm dữ liệu và đường hồi quy trên cùng một đồ thị? Ước lượng điểm bài thi cuối kỳ của một sinh viên có điểm thi giữa kỳ là 85?

Câu 22

Ở một loại mẫu kim loại nào đó, ứng suất pháp tuyến (x) trên một mẫu vật phụ thuộc hàm vào độ kháng cắt (y). Sau đây là một tập dữ liệu thí nghiệm đã được mã hóa về hai biến:

Ứng suất pháp tuyến	Độ kháng cắt
26.8	26.5
25.4	27.3
28.9	24.2
23.6	27.1
27.7	23.6
23.9	25.9
24.7	26.3
28.1	22.5
26.9	21.7
27.4	21.4
22.6	25.8
25.6	24.9

- a.** Tính hệ số tương quan giữa ứng suất pháp tuyến và độ kháng cắt? Nêu nhận xét về mối tương quan giữa hai đại lượng này?
- b.** Ước lượng đường hồi quy tuyến tính? Vẽ các điểm dữ liệu và đường hồi quy trên cùng một đồ thị? Ước lượng độ kháng cắt ứng với ứng suất pháp tuyến 24.5?

Câu 23

Khối lượng của một hợp chất hóa học (y) hòa tan trong 100 gram nước ở các nhiệt độ khác nhau (x) được ghi lại như sau:

x ($^{\circ}\text{C}$)	y (gram)		
0	8	6	8
15	12	10	14

x (°C)	y (gram)		
30	25	21	24
45	31	33	28
60	44	39	42
75	48	51	44

- a. Tính hệ số tương quan giữa nhiệt độ và khối lượng hòa tan của chất đó trong 100 gram nước? Nêu nhận xét về mối tương quan giữa hai đại lượng này?
- b. Ước lượng đường hồi quy tuyến tính? Vẽ các điểm dữ liệu và đường hồi quy trên cùng một đồ thị? Ước lượng khối lượng của hợp chất hóa học sẽ hòa tan trong 100 gram nước ở 50°C?

Câu 24

Xem bảng sau để biết dữ liệu về xếp hạng cầu thủ ném bóng bầu dục trong mùa giải bóng quốc gia Mỹ năm 2008 (The Sports Network). Người ta nghi ngờ rằng điểm số (y) có liên quan đến số yard đạt được trong mỗi lượt ném (x).

Cầu thủ		Đội	Số yard trong mỗi lượt ném	Điểm số
Philip	Rivers	SD	8.39	105.5
Chad	Pennington	MIA	7.67	97.4
Kurt	Warner	ARI	7.66	96.9
Drew	Brees	NO	7.98	96.2
Peyton	Manning	IND	7.21	95
Aaron	Rodgers	GB	7.53	93.8
Matt	Schaub	HOU	8.01	92.7
Tony	Romo	DAL	7.66	91.4
Jeff	Garcia	TB	7.21	90.2
Matt	Cassel	NE	7.16	89.4
Matt	Ryan	ATL	7.93	87.7
Shaun	Hill	SP	7.1	87.5
Seneca	Wallace	SEA	6.33	87
Eli	Manning	NYG	6.76	86.4
Donovan	McNabb	PHI	6.86	86.4

Cầu thủ		Đội	Số yard trong mỗi lượt ném	Điểm số
Jay	Cutler	DEN	7.35	86
Trent	Edwards	BUF	7.22	85.4
Jake	Delhomme	CAR	7.94	84.7
Jason	Campbell	WAS	6.41	84.3
David	Garrard	JAC	6.77	81.7
Brett	Favre	NYJ	6.65	81
Joe	Flacco	BAL	6.94	80.3
Kerry	Collins	TEN	6.45	80.2
Ben	Roethlisberger	PTT	7.04	80.1
Kyle	Orton	CHI	6.39	79.6
Jamacus	Russell	OAK	6.58	77.1
Tyler	Thigpen	KC	6.21	76
Gus	Freotte	MIN	7.17	73.7
Dan	Orlovsky	DET	6.34	72.6
Marc	Bulger	STL	6.18	71.4
Ryan	Fitzpatrick	CIN	5.12	70
Derek	Anderson	CLE	5.71	66.5

- Tính hệ số tương quan giữa điểm số và số yard trung bình đạt được trong mỗi lượt ném? Nêu nhận xét về mối tương quan giữa hai đại lượng này?
- Ước lượng đường hồi quy tuyến tính? Vẽ các điểm dữ liệu và đường hồi quy trên cùng một đồ thị? Ước lượng số điểm đạt được khi cầu thủ ném được 7.21 yard?

Câu 25

Một động cơ tên lửa được sản xuất bằng cách liên kết hai thành phần với nhau: một bộ đánh lửa và một máy phát điện. Độ bền cắt của liên kết (theo psi) (y) được cho là một hàm tuyến tính của tuổi động cơ đẩy (theo tuần) (x) khi động cơ được đúc. Bảng số liệu 20 quan trắc như sau:

STT	Độ bền cắt (psi)	Tuổi động cơ (tuần)
-----	---------------------	------------------------

1	2158.7	15.5
2	1678.15	23.75
3	2316	8
4	2061.3	17
5	2207.5	5
6	1708.3	19
7	1784.7	24
8	2575	2.5
9	2357.9	7.5
10	2277.7	11
11	2165.2	13
12	2399.55	3.75
13	1779.8	25
14	2336.75	9.75
15	1765.3	22
16	2052.5	18
17	2414.4	6
18	2200.5	12.5
19	2654.2	2
20	1753.7	21.5

- a. Vẽ đồ thị phân tán của dữ liệu trên?
- b. Từ đồ thị, hãy cho biết mô hình hồi quy tuyến tính có phù hợp để mô hình hóa dữ liệu trên không? Nếu có, hãy ước lượng đường hồi quy tuyến tính, sau đó vẽ các điểm dữ liệu và đường hồi quy trên cùng một đồ thị?