

Sample First Midterm Examination
(without solutions)

The exam was closed book, but one 8.5" x 11" (note book sized) sheet of notes and a calculator was permitted. There was a table equivalent to Table A in *IPS* at the end of the exam

1. Baseball pitcher Nolan Ryan played in 20 games or more in the 24 seasons from 1968 through 1992. Here are the numbers of games he played in during these years.

21	39	39	35	30	33
25	41	37	21	35	32
27	42	31	35	30	30
30	28	34	29	34	27

(a) Make a stemplot of these data, ordering the leaves on each stem

Note: Use 4 or more stems

(b) Find the quartiles of these data (first or lower quartile, third or upper quartile, and median). Here is some MacAnova output:

```
Cmd> ryan <- vector(21,39,39,35,30,33,25,41,37,21,35,32,\
27,42,31,35,30,30,30,28,34,29,34,27) # enter data
```

```
Cmd> sort(ryan)
```

(1)	21	21	25	27	27
(6)	28	29	30	30	30
(11)	30	31	32	33	34
(16)	34	35	35	35	37
(21)	39	39	41	42	

(c) More MacAnova output

```
Cmd> n <- 24
```

```
Cmd> sum(ryan)
```

```
(1) 765
```

```
Cmd> write(sum((ryan - sum(ryan)/n)^2))
```

```
NUMBER:
```

```
(1) 722.625
```

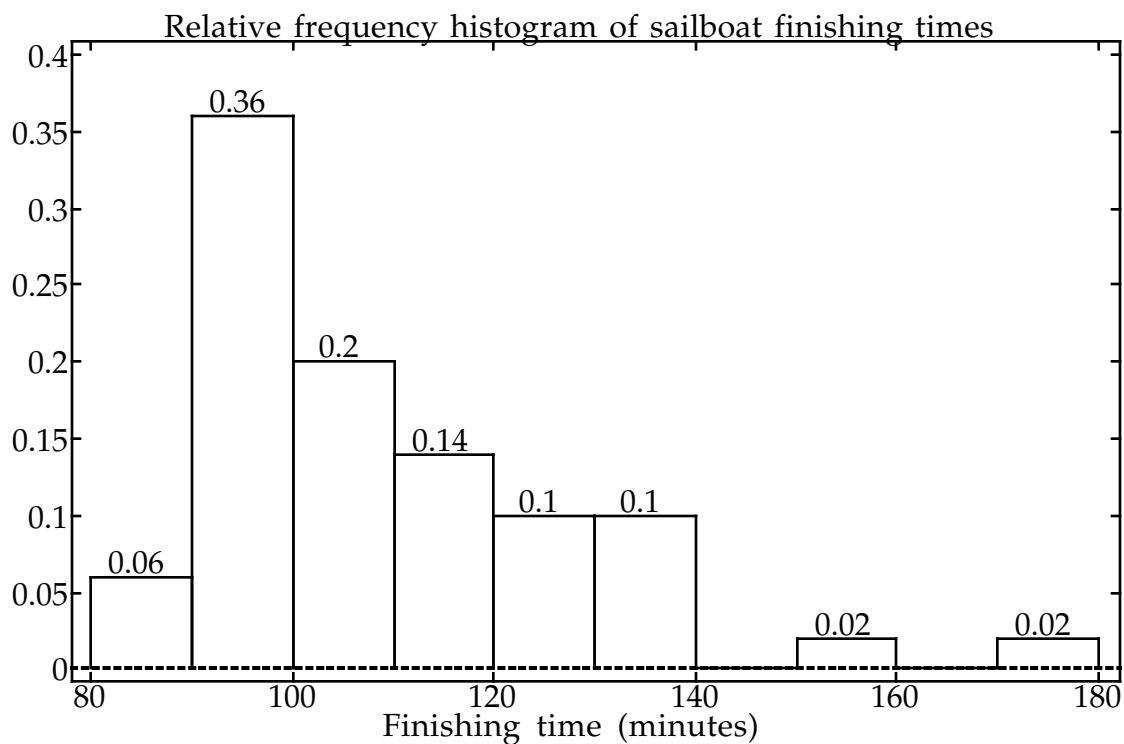
Use these values to calculate the sample mean \bar{x} and standard deviation s . Do not enter the data in your calculator. Show your work.

$$\bar{x} = \underline{\hspace{2cm}} \qquad s = \underline{\hspace{2cm}}$$

2. Fifty boats, all of the same design, competed in a six mile sailboat race on Lake Superior.

(a) Suppose you are told (not necessarily correctly) that the distribution of finishing times (times from start to finish) of these boats was approximately normal, with mean 110 minutes and standard deviation 20 minutes. Approximately what proportion of the finishes would you guess were between 90 and 130 minutes?

Here is a relative frequency histogram of the actual finishing times of the 50 boats. The number above each bar is the relative frequency of sailboats in that time class.



(b) Describe in no more than two sentences the shape of the distribution of finishing times. Would you expect the median to be larger than the mean, smaller than the mean, or about the same?

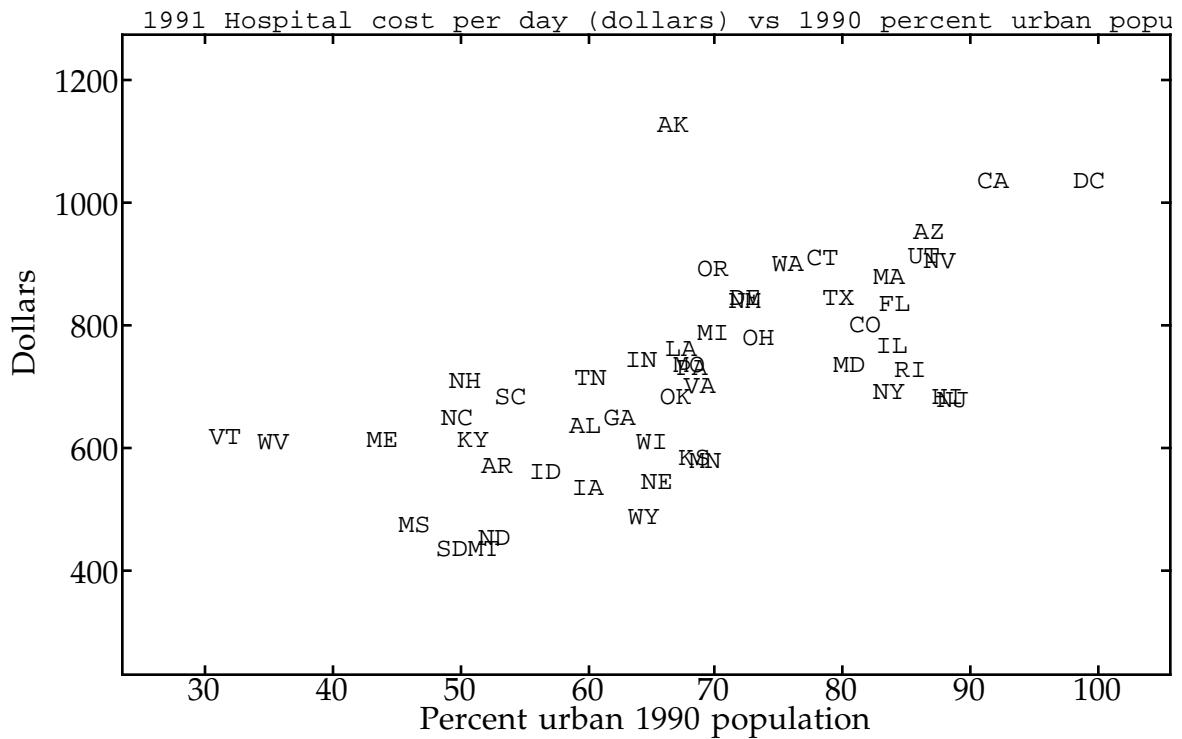
(c) What percent of the sail boats actually finished between 90 and 130 minutes after the start.

Statistics 5021 Sample First Midterm Examination

3. The following table lists data for the 50 states and the District of Columbia from the 1993-1994 Statistical Abstract of the United States. There are two variables, the percent of the population living in urban areas in 1990 (% urban) and the average 1991 hospital cost per day (Hosp \$/day). Also included are the two letter abbreviations for the states.

	% urban	Hosp \$/day		% urban	Hosp \$/day		% urban	Hosp \$/day		% urban	Hosp \$/day
AK	67.5	1130	ID	57.4	565	MT	52.5	437	RI	86.0	730
AL	60.4	637	IL	84.6	770	NC	50.4	651	SC	54.6	684
AR	53.5	571	IN	64.9	745	ND	53.3	454	SD	50.0	436
AZ	87.5	955	KS	69.1	585	NE	66.1	546	TN	60.9	716
CA	92.6	1037	KY	51.8	616	NH	51.0	713	TX	80.3	846
CO	82.4	801	LA	68.1	764	NJ	89.4	680	UT	87.0	915
CT	79.1	910	MA	84.3	880	NM	73.0	840	VA	69.4	701
DC	100.0	1038	MD	81.3	740	NV	88.3	907	VT	32.2	621
DE	73.0	845	ME	44.6	617	NY	84.3	694	WA	76.4	904
FL	84.8	836	MI	70.5	792	OH	74.1	782	WI	65.7	611
GA	63.2	652	MN	69.9	582	OK	67.7	684	WV	36.1	612
HI	89.0	686	MO	68.7	737	OR	70.5	896	WY	65.0	489
IA	60.6	538	MS	47.1	479	PA	68.9	732			

Here is a scatterplot of these data.



Statistics 5021 Sample First Midterm Examination

3. (continued)

(a) Describe in about two sentences any regular pattern to the data and any apparent deviations from the pattern.

(b) Alaska (AK) has the highest hospital cost of any state. In a regression of hospital cost on percent rural population, is Alaska likely to be an influential case? Why or why not.

(c) Here are various summary numbers ($x = \% \text{ urban}$, $y = \text{Hosp } \$/\text{day}$).

$n = 51$	$\sum x = 3509.0$	$\sum y = 36,789$
$\sum(x - \bar{x})^2 = 11,539.94$	$\sum(y - \bar{y})^2 = 1,272,643.7$	$\sum(x - \bar{x})(y - \bar{y}) = 80,509.43$
$s_x = 15,192$	$s_y = 159.54$	$r_{xy} = 0.6643$

Find the intercept and slope of the least squares regression line. Use the summary numbers given. *Do not key in the data in your calculator.*

4. Here are counts from a study of all gun related deaths in Milwaukee, between 1990 and 1994.

	Handgun	Shotgun	Rifle	Unknown	Totals
Homicides	468	28	15	13	524
Suicides	124	22	24	5	175
Totals	592	50	39	18	699

(a) Find the conditional distribution of gun type given cause of death.

(b) On the basis of these distributions, in not more than 2 or 3 sentences, describe the differences in types of weapons used in gun related homicides and suicides.