1. [20 points] For each of the statements below, circle **T** or **F** for "True" or "False," respectively. (4 pts. each)

T **F** : The correlation can be any number, but it is usually between -1 and $+1$.

FALSE: Correlation must be between -1 and $+1$.
See item 1. at the bottom of p. 166 in the text.

T **F** : If the distribution is bell shaped with no outliers, we expect IQR will be smaller than s.

FALSE: It was stated in class that IQR is approximately equal to $1.35s$ for the normal distribution. You can also figure this out from the tables.

T **F** : Nonresponse bias refers to systematic error in sampling from a population due to subjects being unavailable or refusing to reply.

TRUE: See bottom of p. 251 in the text.

T **F** : One purpose of randomization in experimental design is to eliminate confounding effects from lurking variables that might be present in an observational study.

TRUE: This was stated in lecture.

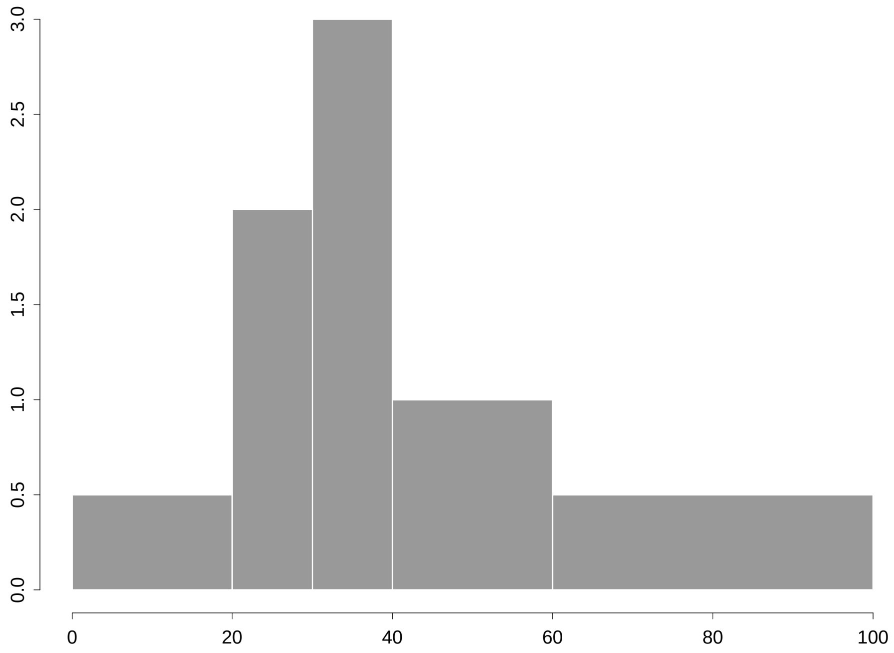
T **F** : Both the mean and the median are resistant measures of the center of a distribution of data.

FALSE: See the bottom of p. 37 to the top of p. 38.

2. [30 points] Use the table below to sketch a density histogram for the data.

Class	Percentage		Class Width	Bar Height
0 – 20	10%	$\longrightarrow$	**20**	**0.5**
20 – 30	20%	$\longrightarrow$	**10**	**2.0**
30 – 40	30%	$\longrightarrow$	**10**	**3.0**
40 – 60	20%	$\longrightarrow$	**20**	**1.0**
60 – 100	20%	$\longrightarrow$	**40**	**0.5**

We have added two extra columns in the table : one for class width and one for the height of the histogram bars. The plot appears below.



3. [30 points] Suppose a data set has approximately a normal distribution with mean $\bar{x} = 200$ and standard deviation $s = 20$.

(a) Estimate the percentage of the data which are between 170 and 210.

Computing the corresponding z values:

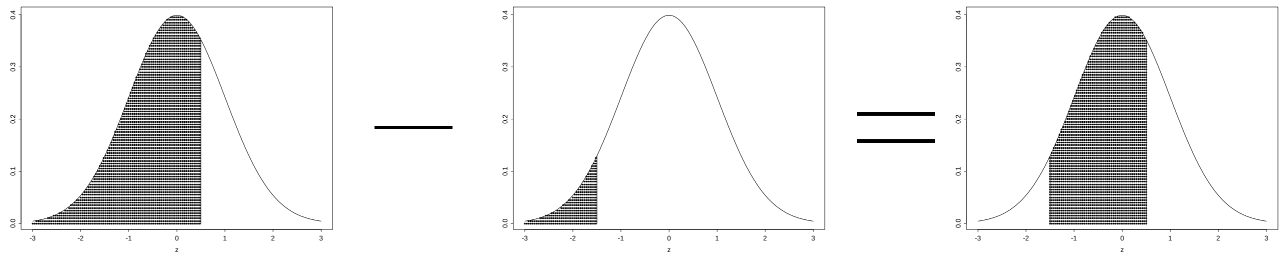
$$z_1 = \frac{170 - 200}{20} = -1.5 \qquad (1)$$

$$z_2 = \frac{210 - 200}{20} = 0.5 \qquad (2)$$

From the tables provided, the area under the curve to the left of $z_1 = -1.5$ is 0.0668, and the area under the curve to the left of $z_2 = 0.5$ is 0.6915. Thus, the area between them is

$$0.6915 \; - \; 0.0668 \; = \; 0.6247.$$

The calculation is depicted pictorially below.



3(b) Find approximately the 80'th percentile of the data.

Using the tables, the 80'th **percentile of the** $N(0,1)$ **distribution is** 0.84. **The area corresponding the** $z = 0.84$ **is** 0.7995, **which is the closest we can get to** 0.8. **The corresponding data value is obtained by the "inverse"** z**-value transformation:**

$$x = \bar{x} + z * s = 200 + 0.84 * 20 = 216.8.$$

Thus, our estimate of the 80'th **percentile of the data is** 216.8.

4. [20 points] Below are 5 values of r, the correlation of a sample, and 4 scatterplots. Match the value of r with the scatterplot by writing the plot label (A, B, C, or D) next to the value of r. Obviously, one value of r will be unmatched.

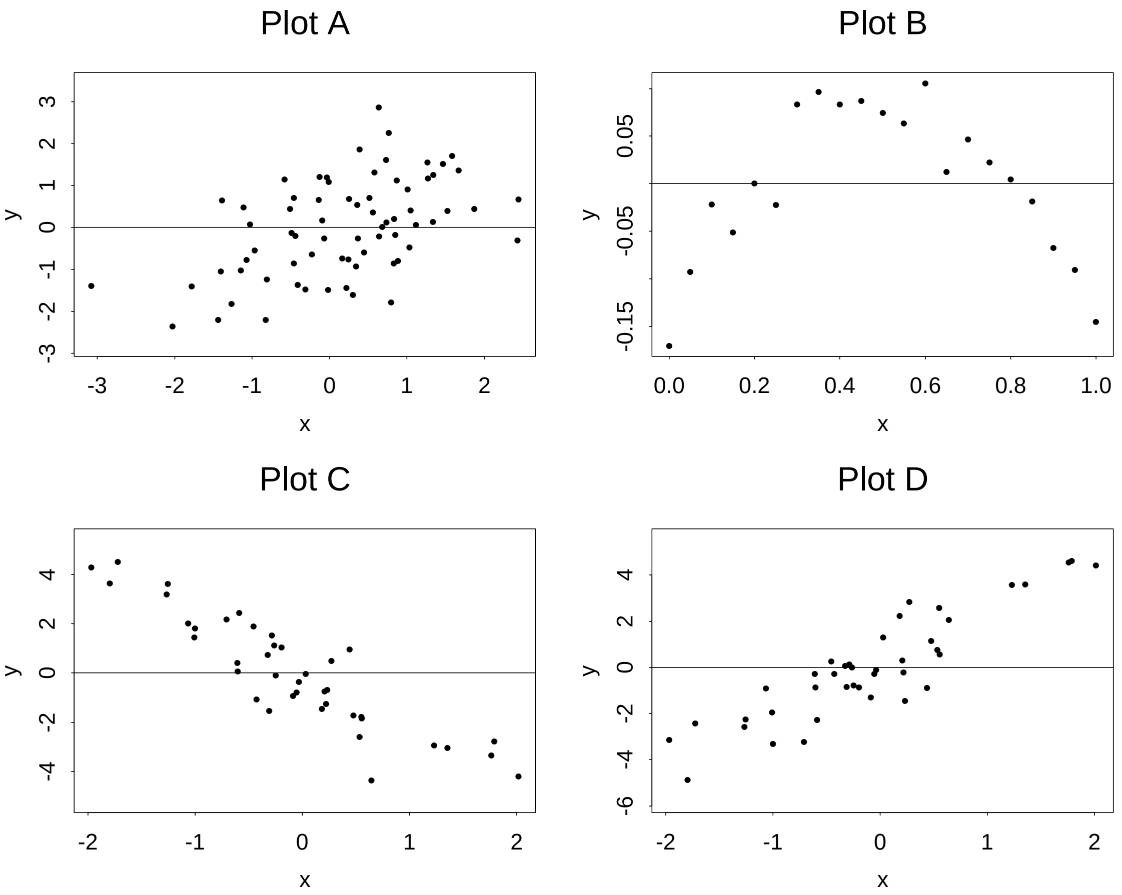
(i) $r = -0.9$ Plot C. Clearly this plot has a negative association and the points fall close to a straight line, so the corresponding correlation is near -1

(ii) $r = -0.5$ No Plot.

(iii) $r = 0.0$ Plot B. There is no upward or downward "tilt" in Plot B, so there is no linear association, although there is a clear nonlinear association.

(iv) $r = +0.5$ Plot A. Plot A displays a fairly clear upward tilt, hence has a positive correlation, but it is not so strongly positive as Plot D, so by process of elimination Plot A must go with $r = +0.5$.

(v) $r = +0.9$ Plot D. Plot D has a very strong positive correlation, near 1.


Plot A
Plot B
Plot C
Plot D
x
y
-3
-2
-1
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
-0.15
-0.05
0.05
-6
-4
4

Summary Statistics for Scores

For $n = 75$ persons taking the exam before Fri., 13 Feb.:

$$\bar{x} \;=\; 87.01, \quad s \;=\; 13.07.$$

Five Number summary:

min	Q_1	M	Q_3	max
56	77	92	99	100

Histogram is shown below. Approximate letter grades are

A-B	$85 - -100$	(68%)
C	$65 - 84$	(25%)
D	$56 - 64$	(7%)