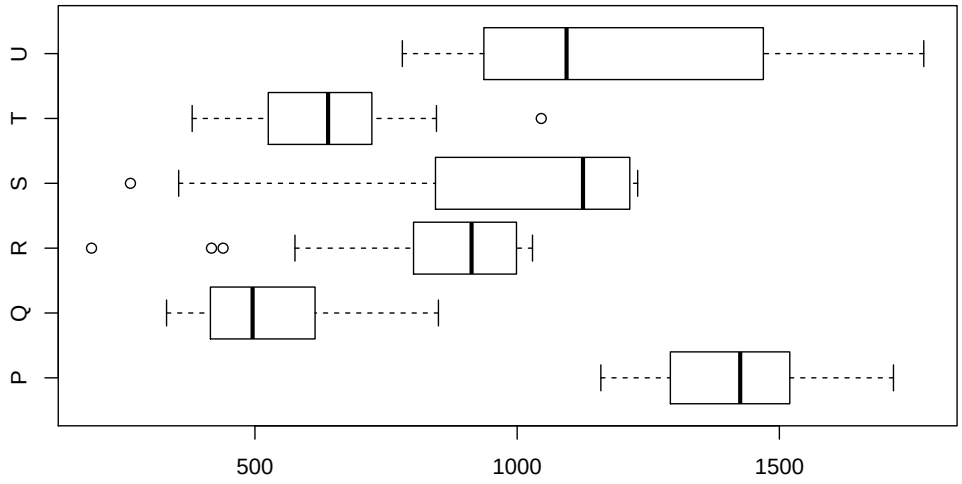
1. **Problem**
   Six random variables were each measured 25 times. The resulting boxplots are shown.



(a) Which variable produced the largest measurment?
(b) Which variable produced the smallest measurment?
(c) Which distribution has the largest median?
(d) Which distribution has the smallest median?
(e) Which distribution has the largest 25th percentile?
(f) Which distribution has the smallest 25th percentile?
(g) Which distribution has the largest 75th percentile?
(h) Which distribution has the smallest 75th percentile?
(i) Which distribution has the largest IQR?
(j) Which distribution has the smallest IQR?

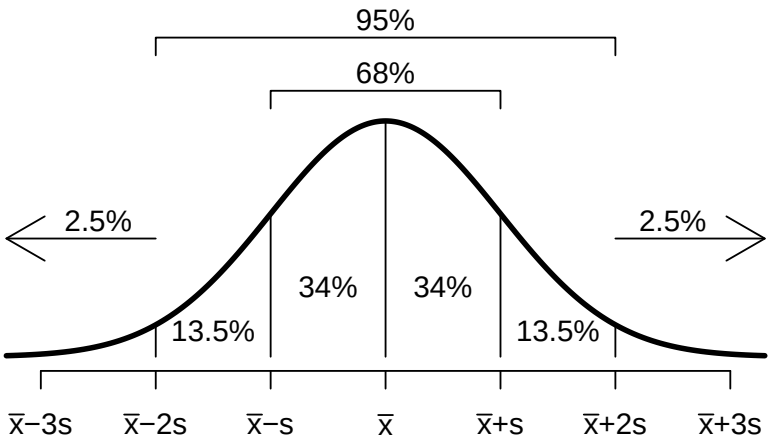
2. **Problem**
   From a very large population, a small sample of measurements was taken.

53	67	59	57

   Please calculate the Average Absolute Deviation (also called the Average Distance from Mean: ADM) using the following formula:

$$\text{AAD} = \frac{\sum |x - \bar{x}|}{n}$$

3. **Problem**
   The figure below summarizes the *standard deviation rule* for normal distributions. In the figure, $\bar{x}$ is the mean and s is the standard deviation. The percentages show the fraction of measurements that fall within various intervals.



A specific distribution is approximately normal with mean $\bar{x} = 17$ and standard deviation $s = 5$.

(a) What percent of the measurements are greater than 22?

(b) What percent of the measurements are less than 12?

(c) What measurement is greater than 2.5% of the measurements?

(d) What measurement is less than 50% of the measurements?

(e) What percent of the measurements are between 7 and 27?

4. **Problem**
   From a very large population, a small sample of measurements was taken.

64	66	58	64	63

   Please calculate the (Bessel corrected) sample standard deviation using the following formula:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

5. **Problem**
   A continuous random variable was measured 500 times. The resulting frequency distribution is shown below.

class	frequency
10–15	150
15–20	37
20–25	34
25–30	15
30–35	31
35–40	32
40–45	41
45–50	116
50–55	44

   (a) Describe the overall shape of the distribution. (symmetric mound, skew left, skew right, uniform, or bimodal)

   (b) Estimate the range of the distribution (range = max-min).

   (c) What percent of the measurements are less than 40?

   (d) What percent of the measurements are greater than 25?

   (e) What percent of the measurements are between 25 and 40?

   (f) What percent of the measurements are within 7.5 of 27.5? In other words, what percent of measurements satisfy $|x - 27.5| \leq 7.5$?

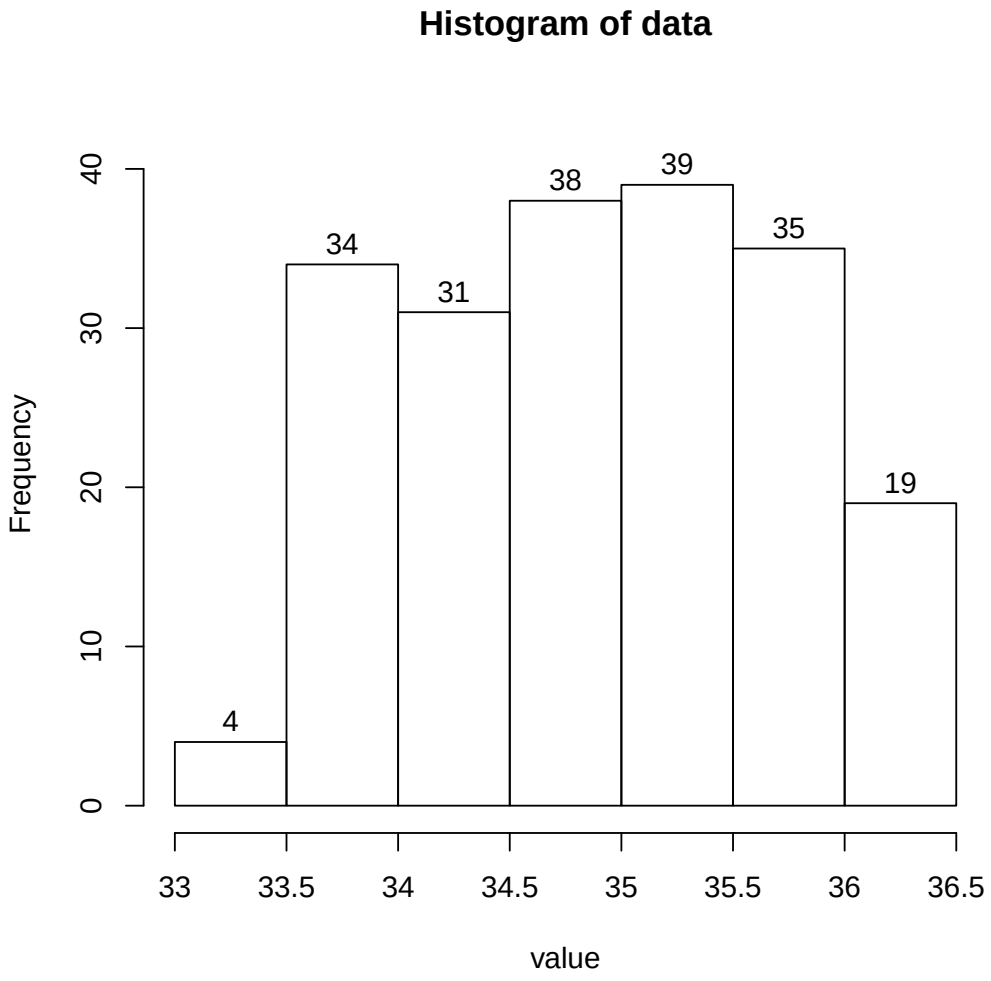
   (g) Of the measurements less than 40, what percent are greater than 25?

   (h) Estimate the value of the 53.4th percentile. In other words, determine a value such that 53.4% of the measurements are less than or equal to it.

6. **Problem**
   Please make a frequency table and a dot plot from the following (unsorted) data.

57	50	50	54	51	50
50	53	57	57	57	51
50	52	53	51	56	54
53	57	51	57	51	56
50	52	53	53	53	56

7. **Problem**
A continuous random variable was measured 200 times. The resulting histogram is shown below.



(a) Describe the overall shape of the distribution. (symmetric mound, skew left, skew right, uniform, or bimodal)

(b) Estimate the range of the distribution (range = max-min).

(c) What percent of the measurements are less than 34?

(d) What percent of the measurements are greater than 33?

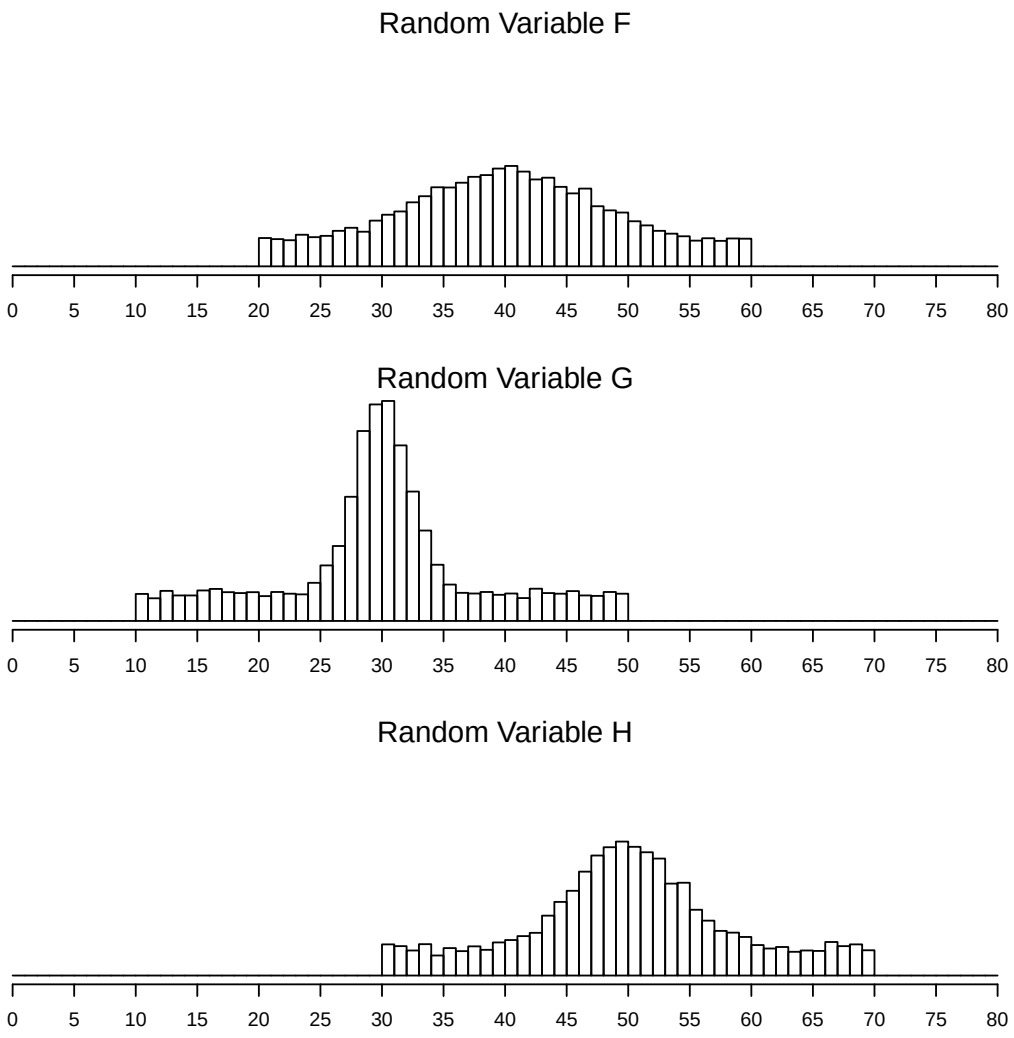
(e) What percent of the measurements are between 33 and 34?

(f) What percent of the measurements are within 0.5 from 34.5? In other words, what percent of measurements satisfy $|x - 34.5| \leq 0.5$?

(g) Of the measurements less than 34, what percent are greater than 33?

(h) Estimate the value of the 73th percentile. In other words, determine a value such that 73% of the measurements are less than or equal to it.

8. **Problem**
Three random variables (F, G, and H) were measured 1000 times each. The resulting histograms show the three distributions.



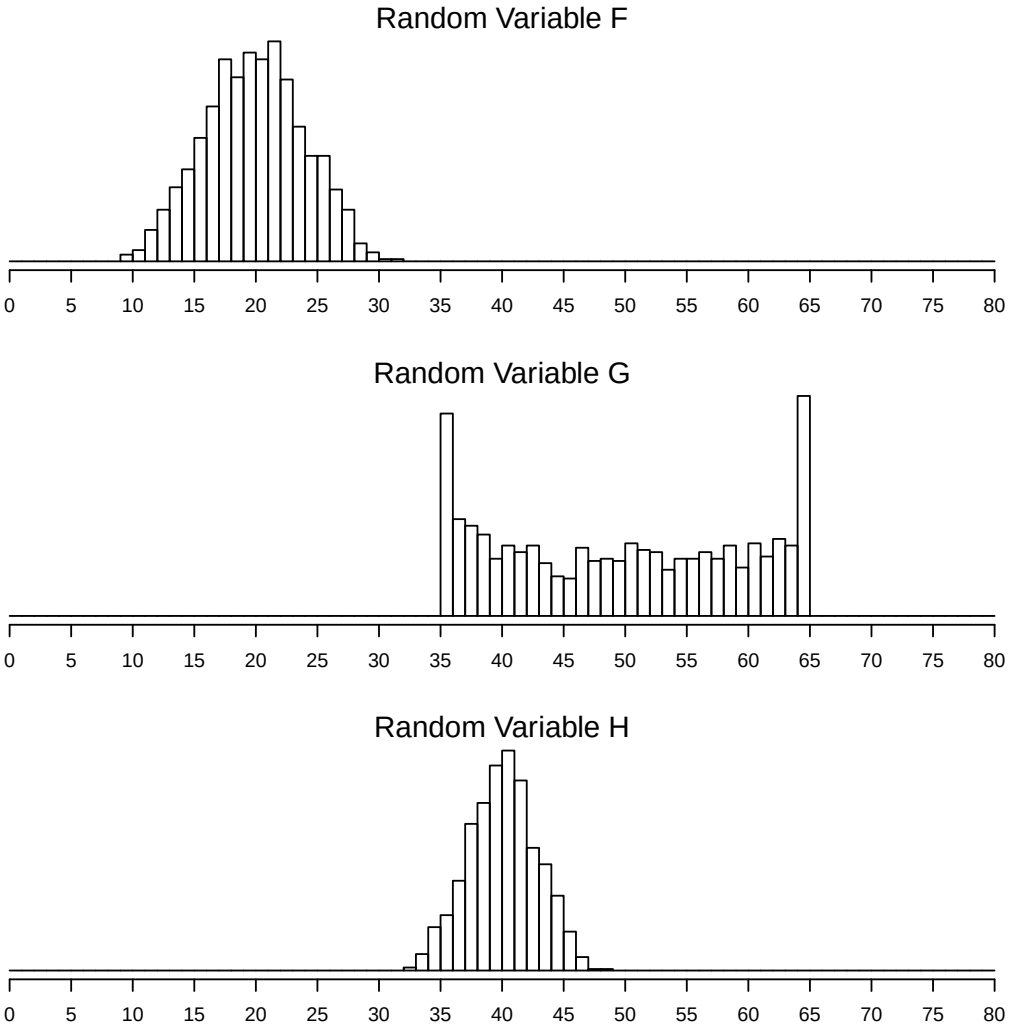
(a) Which distribution has the highest mean? (F, G, or H)

(b) Which distribution has the lowest mean? (F, G, or H)

(c) Which distribution has the largest standard deviation? (F, G, or H)

(d) Which distribution has the smallest standard deviation? (F, G, or H)

9. **Problem**
Three random variables (F, G, and H) were measured 1000 times each. The resulting histograms show the three distributions.



(a) Which distribution has the highest mean? (F, G, or H)

(b) Which distribution has the lowest mean? (F, G, or H)

(c) Which distribution has the largest standard deviation? (F, G, or H)

(d) Which distribution has the smallest standard deviation? (F, G, or H)

10. **Problem**
A continuous random variable X was measured 67 times. The sorted measurements are shown below.

70.03	70.06	70.08	70.14	70.22	70.22	70.29	70.32	70.41	70.42
70.47	70.61	70.64	70.7	70.77	70.78	70.79	70.81	70.91	70.92
70.96	71.14	71.23	71.25	71.3	71.57	71.58	71.72	71.72	71.85
71.87	71.9	72.08	72.3	72.32	72.38	72.47	72.47	72.53	72.53
72.8	72.92	72.95	72.97	73.02	73.17	73.25	73.32	73.34	73.38
73.68	73.81	74.04	74.19	74.34	74.37	74.38	74.42	74.48	74.5
74.52	74.56	74.72	74.72	74.74	74.75	74.85			

The total of the measurements is 4842.95.

(a) Determine the percentile rank of the measurement 74.19. In other words, determine what percent of data are less than or equal to 74.19.

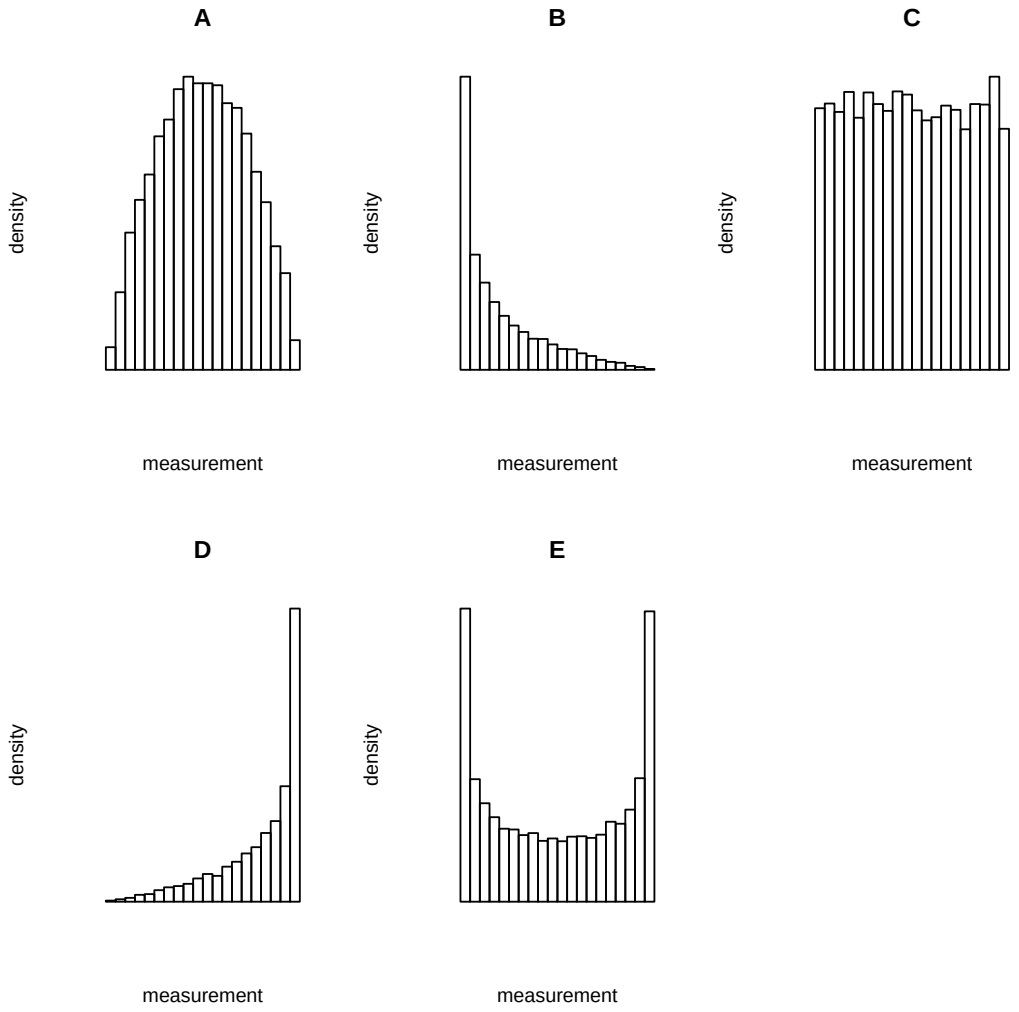
(b) Determine the measurement corresponding to a percentile rank of 0.433. In other words, determine x such that 43.3% of the data are less than or equal to x.

(c) Determine the mean of the measurements.

(d) Determine the median of the measurements.

11. **Problem**
For **each** of the histograms:

- Determine if the mean is higher than, lower than, or equal to the median.
- Would you caution against using the mean?



plot of chunk hists

(a) Answer both questions about distribution A.
(b) Answer both questions about distribution B.
(c) Answer both questions about distribution C.
(d) Answer both questions about distribution D.
(e) Answer both questions about distribution E.

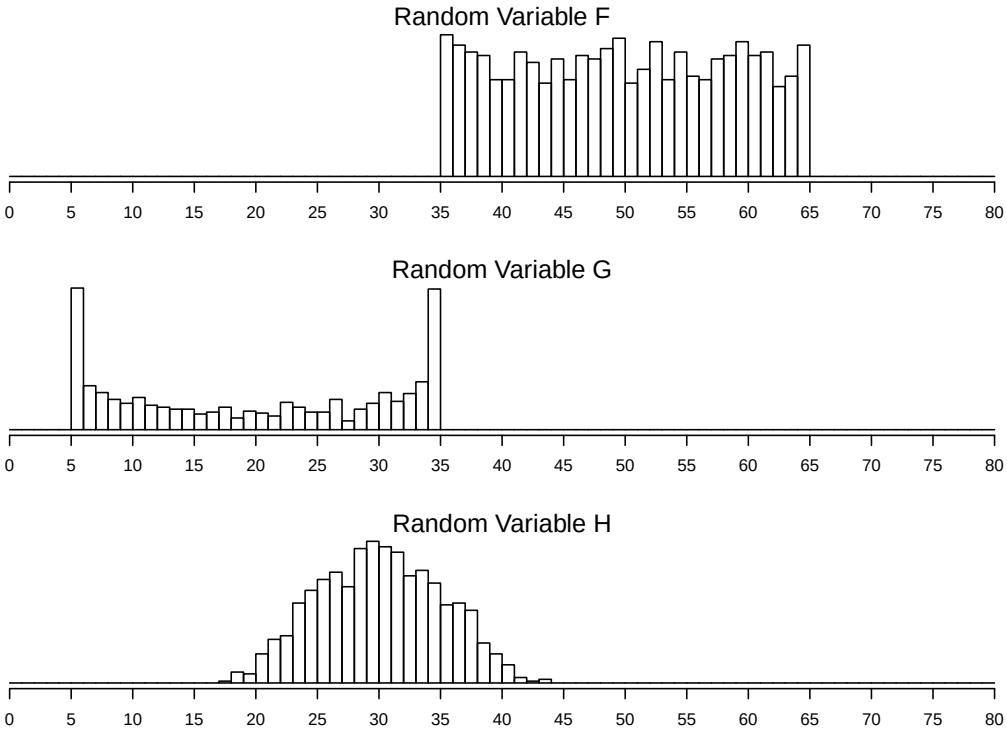
12. **Problem**
We can estimate the mean of **symmetric** distributions.

$$\bar{x} \approx \frac{\max(x) + \min(x)}{2}$$

We can **roughly** estimate the standard deviation of certain distributions.

Shape	SD estimate
bell	range/6
uniform	range/4
bimodal	range/2

Three random variables (F, G, and H) were measured 1000 times each. The resulting histograms show the three distributions.



(a) Estimate the mean of F.

(b) Estimate the mean of G.

(c) Estimate the mean of H.

(d) Estimate the standard deviation of F.

(e) Estimate the standard deviation of G.

(f) Estimate the standard deviation of H.

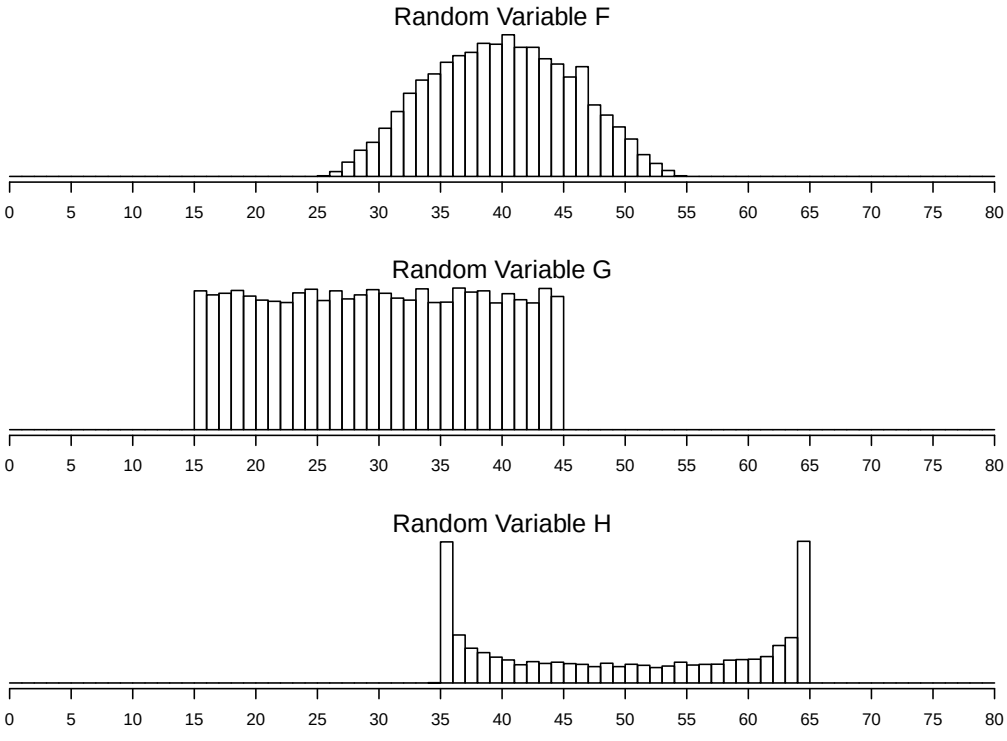
13. **Problem**
We can estimate the mean of a **symmetric** distribution.

$$\bar{x} \approx \frac{\max(x) + \min(x)}{2}$$

We can **roughly** estimate the standard deviation of certain distributions.

Shape	SD estimate
bell	range/6
uniform	range/4
bimodal	range/2

Three random variables (F, G, and H) were measured 10000 times each. The resulting histograms show the three distributions.



.image

(a) Estimate the mean of F.

(b) Estimate the mean of G.

(c) Estimate the mean of H.

(d) Estimate the standard deviation of F.

(e) Estimate the standard deviation of G.

(f) Estimate the standard deviation of H.

14. **Problem**
Two random variables (A and B) are both approximately normal (bell-shaped). Their means and standard deviations are shown in the table.

variable	mean	standard deviation
A	59.4	15
B	69.9	23

Let the *interval of typical measurements* be defined as within 1 SD from the mean.

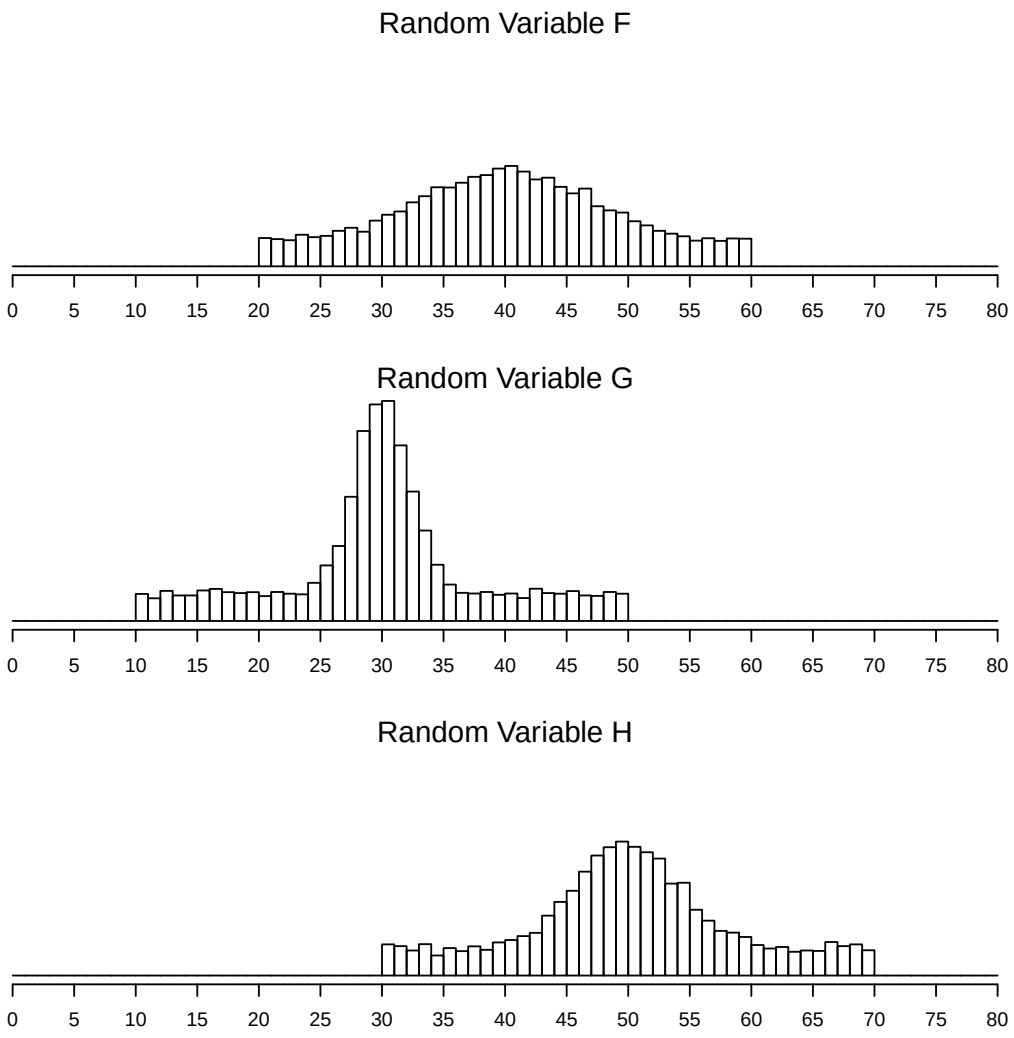
$$\text{interval of typical measurements} = (\text{mean} - \text{SD}, \text{mean} + \text{SD})$$

For each variable, provide an interval of typical measurements. Notice that an interval requires two numbers: the bottom and the top.

(a) Determine the interval of typical measurements for A.

(b) Determine the interval of typical measurements for B.

15. **Problem**
Three random variables (F, G, and H) were measured 1000 times each. The resulting histograms show the three distributions.



(a) Which distribution has the highest mean? (F, G, or H)

(b) Which distribution has the lowest mean? (F, G, or H)

(c) Which distribution has the largest standard deviation? (F, G, or H)

(d) Which distribution has the smallest standard deviation? (F, G, or H)

1. (a) U
   (b) R
   (c) P
   (d) Q
   (e) P
   (f) Q
   (g) P
   (h) Q
   (i) U
   (j) R

2. We fill out the table column by column.

x	$x - \bar{x}$	$\lvert x - \bar{x}\rvert$
53	-6	6
67	8	8
59	0	0
57	-2	2
========	========	========
$\sum x = 236$		$\sum \lvert x - \bar{x}\rvert = 16$
$\bar{x} = 59$		

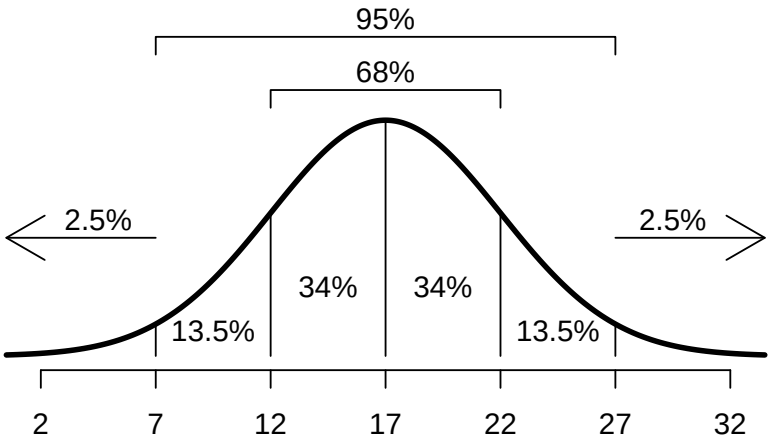
   We are ready for the formula.

$$s = \frac{\sum |x - \bar{x}|}{n}$$

$$= \frac{16}{4}$$

$$= \boxed{4}$$

3. It is probably best to start by redrawing (relabeling) the normal distribution with the specific values.



(a) Because we are asked for the percent of measurements *greater* than 22, we add the areas to the right of 22.

$\boxed{16\%}$

(b) Because we are asked for the percent of measurements *less* than 12, we add the areas to the left of 12.

$\boxed{16\%}$

(c) We determine which leftward area has a total of 2.5%. This occurs at $\boxed{7}$.

(d) We determine which rightward area has a total of 50%. This occurs at $\boxed{17}$.

(e) We add the areas from 7 to 27.

$\boxed{95\%}$

4. We fill out the table column by column.

x	$x - \bar{x}$	$(x - \bar{x})^2$
64	1	1
66	3	9
58	-5	25
64	1	1
63	0	0
========	========	========
$\sum x = 315$		$\sum (x - \bar{x})^2 = 36$
$\bar{x} = 63$		

We are ready for the formula.

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{36}{5-1}}$$

$$= \sqrt{9}$$

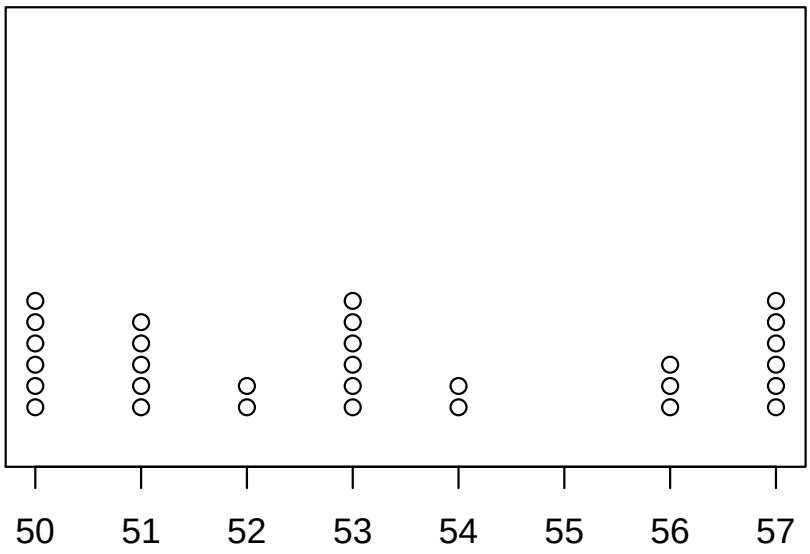
$$= \boxed{3}$$

5. (a) bimodal
   (b) 45
   (c) 59.8%
   (d) 55.8%
   (e) 15.6%
   (f) 16%
   (g) 26.09%
   (h) 35

6. Make a frequency table.

value	frequency
50	6
51	5
52	2
53	6
54	2
55	0
56	3
57	6

Make the dot plot.



.image

7. (a) uniform
   (b) 3.5
   (c) 19%
   (d) 100%
   (e) 19%
   (f) 34.5%
   (g) 100%
   (h) 35.5

8. (a) H
   (b) G
   (c) F
   (d) G

9. (a) G
   (b) F
   (c) G
   (d) H

10. Let x represent a measurement of interest. Let i represent that measurement's index. Let ℓ represent that measurement's percentile. Let n represent the sample size (number of measurements). In general,

$$\ell = \frac{i}{n}$$

(a) We are given $x = 74.19$. This means $i = 54$. We know $n = 67$. Determine the percentile ℓ.

$$\ell = \frac{54}{67}$$

$$\ell = 0.806$$

So, the percentile rank is $\boxed{0.806}$, or 80.6th percentile.

(b) We are given $\ell = 0.433$. We can use algebra to solve for i.

$$\ell = \frac{i}{n}$$

Multiply both sides by n.

$$n \cdot (\ell) = n \cdot \left(\frac{i}{n}\right)$$

Simplify both sides.

$$n\ell = i$$

To make me happy, switch the sides.

$$i = n\ell$$

Now, we can evaluate i.

$$i = (67)(0.433)$$

$$i = 29$$

Determine the x associated with $i = 29$.

$$x = \boxed{71.72}$$

(c) The mean: $\bar{x} = \frac{4842.95}{67} = \boxed{72.283}$

(d) If n is odd, then median is $x_{i=\frac{n+1}{2}}$, the value of x when $i = \frac{n+1}{2}$. Otherwise, if n is even, the median is mean of $x_{i=\frac{n}{2}}$ and $x_{i=\frac{n}{2}+1}$. In this case, $n = 67$ and so n is odd.

$$\text{median} = x_{(67+1)/2,} = x_{34}$$

So, median = $\boxed{72.3}$.

11. (a) The mean is equal to the median. Also, I would not caution against using the mean.

(b) The mean is higher than the median. Also, I would caution against using the mean.

(c) The mean is equal to the median. Also, I would not caution against using the mean.

(d) The mean is lower than the median. Also, I would caution against using the mean.

(e) The mean is equal to the median. Also, I would not caution against using the mean.

12. (a) 50

(b) 20

(c) 30

(d) 7.5

(e) 15

(f) 5

13. (a) 40

(b) 30

(c) 50

(d) 5

(e) 7.5

(f) 15

14. (a)

$$\begin{aligned}\text{interval of typical measurements for } A &= (\text{mean} - \text{SD},\ \text{mean} + \text{SD}) \\ &= (59.4 - 15,\ 59.4 + 15) \\ &= \boxed{(44.4,\ 74.4)}\end{aligned}$$

(b)

$$\begin{aligned}\text{interval of typical measurements for } B &= (\text{mean} - \text{SD},\ \text{mean} + \text{SD}) \\ &= (69.9 - 23,\ 69.9 + 23) \\ &= \boxed{(46.9,\ 92.9)}\end{aligned}$$

15. (a) G

(b) F

(c) G

(d) H