

---

**Assignment due 7-23-10**

**Probability histogram, population and sample sd, etc.**

**1. For the following data:**

2.6 6.3 6.5 2.9 6.7 6.5 5.8 2.6 8.8 4.6 5.5 4.3  
incl                   incl                   incl                   incl

**a. The height of the probability histogram over the class interval [2.5, 4.5].**

$$\text{height} = \frac{\text{fraction of data in the interval}}{\text{interval width}} = \frac{4/12}{4.5-2.5} = \frac{1}{6}$$

**b. The median (sort the list).**

sorted list {2.6, 2.6, 2.9, 4.3, 4.6, 5.5, 5.8, 6.3, 6.5, 6.5, 6.7, 8.8}

$$\text{median} = \frac{1}{2}(5.5 + 5.8) = 5.65$$

**c. All modes.**

2.6 and 6.5 are modes, each with frequency two

**d. Median of the list  $-2x + 4$  where  $x$  is the above list.**

$$-2(\text{median of } x) + 4 = -2(5.65) + 4$$

**e.  $\sigma$  for the list  $x$ .**

1.83005

**f.  $\sigma$  for the list  $-2x + 4$ .**

$$\sigma \text{ for list } (-2x) = |-2| \sigma \text{ for list } x = 2(1.83005)$$

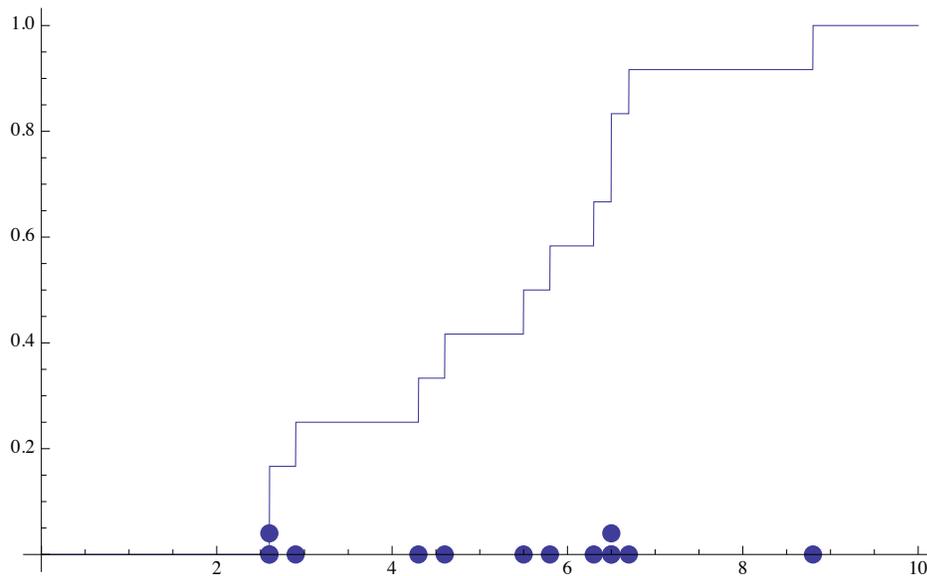
**g. s for the list x.**

$$s \text{ for list } x = \sqrt{\frac{n}{n-1}} \sigma \text{ for list } x = \sqrt{\frac{12}{12-1}} 1.83005 = 1.91142$$

**h. s for the list  $-2x + 4$ .**

$$s \text{ for list } (-2x + 4) = |-2| s \text{ for list } x = 2 \cdot 1.91142$$

**i. On a number line place dots above the numbers in list x. Identify  $\mu$  as the balance point (I've included the answer to (j) in this plot).**



**j. Make a cumulative frequency plot for list x (see key 13).**

Two points to keep in mind:

a. The plot should only jump upward by  $1/n$  (as it passes over a single value such as 4.3 which only appears once in the list) or some multiple of  $1/n$  (such as  $2/n$  when passing over 2.6 which appears twice on the list). Elsewhere, the plot does not rise but remains flat (it only appears to rise continuously when the data is tightly packed as in the illustrations shown in the textbook).

b. The textbook speaks of a cumulative frequency plot but is actually showing cumulative relative frequency plots since they rise from zero at the extreme left to one at the extreme right.

2. There is a 0.003 probability the left engine fails and a 0.006 probability the right engine fails. Suppose these events are **independent**.

a. **P(right engine fails |<sub>IF</sub> left engine fails).**

0.006 since knowledge that the left has failed does not affect the probability that the right fails if the engine failures are independent.

b. **P(both engines fail).**

$$P(\text{left fails}) P(\text{right fails} |_{IF} \text{left fails}) = 0.003 \cdot 0.006$$

c. **P(at least one engine fails).**

$$P(\text{left fails}) + P(\text{right fails}) - P(\text{both fail})$$

$$0.003 \quad 0.004 \quad 0.003 \cdot 0.004$$

d. **P(neither engine fails).**

$$P(\text{left does not fail}) P(\text{right does not fail} |_{IF} \text{left does not})$$

$$(1 - 0.003) \quad (1 - 0.006) \quad \text{by independence}$$

**Notice that "independence of failures" enables us to work from the reliability of individual components through to the reliability of systems involving those components. Otherwise, we cannot easily evaluate the reliability of the system.**

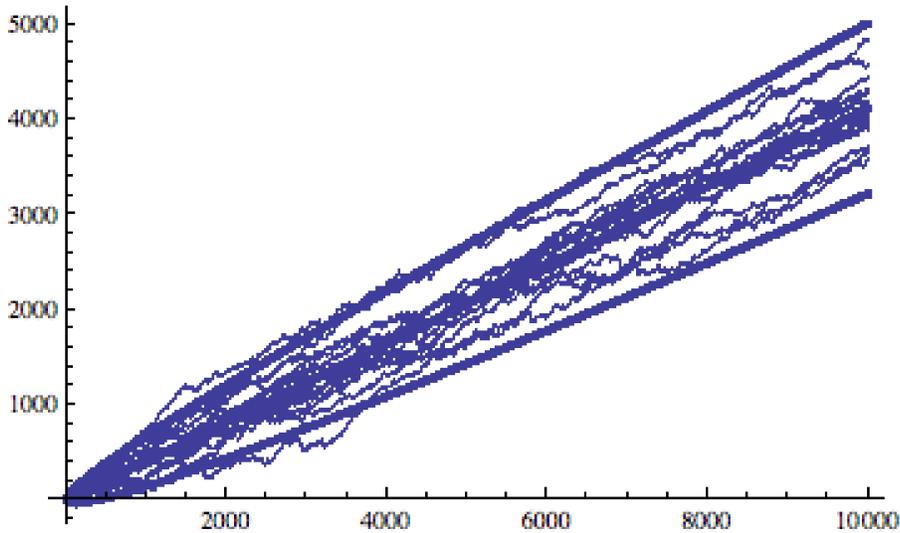
e. **For commercial airliner engines the conditions at airports introduce dependencies among engine failures. What is the reason for this?**

One reason is that many failures are caused by encounters with bird flocks, possibly affecting both engines.

f. **Refer to (e). Is the actual probability P(right engine fails |<sub>IF</sub> left engine fails) likely to be larger or smaller than suggested by your answer to (a)?**

Failure of left tends to raise the chance of birds which in turn raises the chance of right failure.

3. The following pertains to 15 runs, each of which charts the accumulating revenues coming to a casino over 10000 plays of one of its games.



a. What is the per-play mean return to the casino?

From the trend line it appears to have averaged around 4000 over 10000 plays, around 0.4 per play.

b. Name five important features shown in this example that we've claimed are usual for such plots.

Early straying of some plots outside the envelope.

Infrequent and negligible later straying.

Many plots tend to shade the enveloped region.

Plots tend to spend most of the time to one side of trend.

Around half of plots dwell above trend, around half below.

**c. If you double the standard deviation of the per-play return but keep the mean the same what happens to the plot?**

Since per-play  $\sigma$  is a multiplier of the gap between trend line and envelopes that gap will double.

**4. Random variables X, Y are independent.**

**a. Express  $E(aX + bY + c)$  in terms of the constants a, b, c and  $E X$ ,  $E Y$ . Is independence required?**

$E(aX + bY + c) = a E X + b E Y + c$   
independence is not required.

**b. Suppose the average money held on the person of MSU students is \$4.37. Suppose the average coin money held on the person of MSU students is \$0.52. What is the average paper money held on the person of MSU students?**

It is an example of (a),  $4.37 + 0.52 = 4.89$ .

**c. Refer to (b). If we multiply the money held by each MSU student by the last digit of their student number what do you think would be the average result?**

Around 4.89 times 4.5 since last digit of student number is statistically independent of money held, in which case  
 $E(\text{money held times last digit})$   
 $= E(\text{money held}) \text{ times } E(\text{last digit})$

**d. Express  $\text{Variance}(aX + bY + c)$  in terms of the constants a, b, c and  $\text{Variance } X$ ,  $\text{Variance } Y$ . Is independence (or something close to it) required?**

From independence we get  $E(XY) = E X \text{ times } E Y$ .  
 This has the consequence that  $\text{Var}(X + Y) = \text{Var } X + \text{VAR } Y$ .  
 So  $\text{Var}(aX + bY + c) = \text{Var}(aX) + \text{Var}(bY)$   
 $= a^2 \text{Var } X + b^2 \text{Var } Y$