Last Name:

First Name:

SCIPER:

Sample Test 1 – Microengineering 110

Spring 2024

Prof. Vivek Subramanian

- **During the actual test, you would be entering your answers on moodle and on the paper. For this sample test, only the paper version is being provided. During the actual test you MUST enter your answers BOTH on this paper version and on the moodle. The paper version will be used as a backup in case your moodle version is corrupted, etc.**
- **No communications are allowed with others. Therefore, YOU MUST EXIT ALL COMMUNICATION PROGRAMS (Chat windows, etc.). Should a proctor notice that you have an open communication program, it will be considered to be a violation of academic integrity standards.**
- **In the real test, the moodle will allow you to enter information for exactly 45 minutes, at which point it will lock you out. This practice test is much longer than the real test (>2X in length), to give you more practice opportunities.**

1. Cerium Oxide (CeO2) is an important particle used as a catalyst in many electrochemical reactions. The accompanying summary data on CeO2 particle sizes (nm) under certain experimental conditions was read from a graph in the article "Nanoceria— Energetics of Surfaces, Interfaces and Water Adsorption" (J. of the Amer. Ceramic Soc., 2011, 3992–3999):

Diameter	3.0–<3.5	3.5–<4.0	4.0–<4.5	4.5–<5.0	5.0–<5.5	5.5–<6.0	6.0–<6.5	6.5–<7.0	7.0–<7.5	7.5–<8.0
Frequency	5	15	27	34	22	14	7	2	4	1

a. What fraction of the observations are less than 5nm in diameter ? Give your answer as a decimal number.

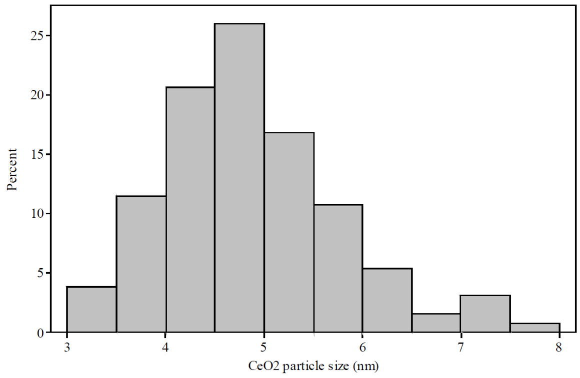
The sample size for this data set is $n = 5 + 15 + 27 + 34 + 22 + 14 + 7 + 2 + 4 + 1 = 131$.

a. The first four intervals correspond to observations less than 5, so the proportion of values less than 5 is $(5 + 15 + 27 + 34)/131 = 81/131 = .618$.

b. What fraction of the observations are at least 6nm in diameter? Give your answer as a decimal number.

The last four intervals correspond to observations at least 6, so the proportion of values at least 6 is $(7 + 2 + 4 + 1)/131 = 14/131 = .107$.

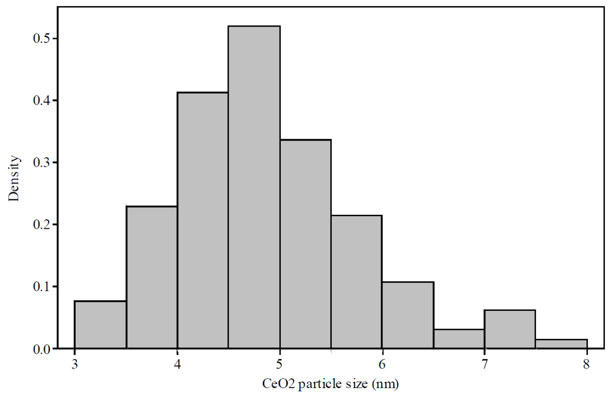
c. Construct a histogram with relative frequency on the vertical axis and comment on interesting features.



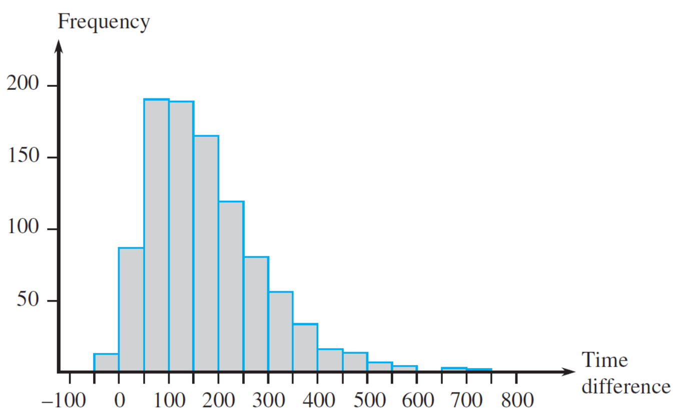
d. Does the distribution of particle sizes appear to be reasonably symmetric or somewhat skewed?

The histogram shows that the data is slightly positively skewed.

e. Construct a histogram with density on the vertical axis and compare to the histogram in (c).



2. How does the speed of a runner vary over the course of a marathon (a distance of 42.195 km)? To examine this, we calculate the difference in time (in seconds) taken to run the first 5km and the 5km starting at the 35th km. A positive value of this difference corresponds to a runner slowing down toward the end of the race. The accompanying histogram is based on times of runners who participated in several different Japanese marathons ("Factors Affecting Runners' Marathon Performance," Chance, Fall, 1993: 24–30).



a. Based on this data, do most runners slow down or speed up over the course?

Most runners slow down at the end of the race, since a huge percentage of the plotted differences are positive.

b. Which of the following is a typical difference value?

- -50 seconds
- 10 seconds
- <u>200 seconds</u>
- 400 seconds

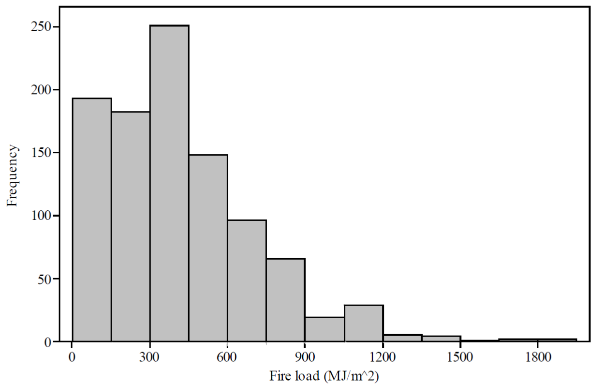
c. Roughly what fraction (give your answer as a decimal) of the runners ran the late distance more quickly than the early distance?

About 0.01. The number of runners that ran more quickly is about 10, and the total data set appears to include about 1000 runners.

3. Prof. Subramanian is from California, where there are major wildfires every year. Californians always worry bout the possibility of their homes catching fire. Fire load (MJ/m2) is the heat energy that could be released per square meter of floor area by combustion of contents and the house structure itself. The article "Fire Loads in Office Buildings" (J. of Structural Engr., 1997: 365–368) gave the following cumulative percentages (read from a graph) for fire loads in a sample of 388 rooms:

Value	0	150	300	450	600	750	900	1050	1200	1350	1500	1650	1800	1950
Cumulative%	0	19.3	37.6	62.7	77.5	87.2	93.8	95.7	98.6	99.1	99.5	99.6	99.8	1000

a. Construct a relative frequency histogram from the above data



b. What fraction (give your answer as a decimal) of fire loads are less than 600?

The proportion of the fire loads less than 600 is .193 + .183 + .251 + .148 = .775.

c. What fraction (give your answer as a decimal) of fire loads are at least 1200?

.005 + .004 + .001 + .002 + .002 = .014.

d. What fraction (give your answer as a decimal) of the loads are between 600 and 1200?

1 – .775 – .014 = .211

4. Exposure to microbial products, especially endotoxin, may have an impact on vulnerability to allergic diseases. The article "Dust Sampling Methods for Endotoxin— An Essential, But Underestimated Issue" (Indoor Air, 2006: 20–27) considered various issues associated with determining endotoxin concentration. The following data on concentration (EU/mg) in settled dust for one sample of urban homes (U) and another of farm homes (F) was obtained.

U: 6.0, 5.0, 11.0, 33.0, 4.0, 5.0, 80.0, 18.0, 35.0, 17.0, 23.0
F: 4.0, 14.0, 11.0, 9.0, 9.0, 8.0, 4.0, 20.0, 5.0, 8.9, 21.0, 9.2, 3.0, 2.0, 0.3

a. Determine the sample mean for each sample. Which one is larger, i.e., E(U) or E(F)?

For urban homes, $\bar{x} = 21.55$ EU/mg; for farm homes, $\bar{x} = 8.56$ EU/mg. The average endotoxin concentration in urban homes is more than double the average endotoxin concentration in farm homes.

b. Determine the sample median for each sample. Which one is larger? Why is the median for the urban sample so different from the mean for that sample?

For urban homes, $\tilde{x} = 17.00$ EU/mg; for farm homes, $\tilde{x} = 8.90$ EU/mg. The median endotoxin concentration in urban homes is nearly double the median endotoxin concentration in farm homes. The mean and median endotoxin concentration for urban homes are so different because the few large values, especially the extreme value of 80.0, raise the mean but not the median.

c. Calculate the trimmed mean for each sample by deleting the smallest and largest observation. What are the corresponding trimming percentages?

For urban homes, deleting the smallest ($x = 4.0$) and largest ($x = 80.0$) values gives a trimmed mean of $\bar{x}_{tr} = 153/9 = 17$ EU/mg. The corresponding trimming percentage is $100(1/11) \approx 9.1\%$. The trimmed mean is less than the mean of the entire sample, since the sample was positively skewed. Coincidentally, the median and trimmed mean are equal.

For farm homes, deleting the smallest ($x = 0.3$) and largest ($x = 21.0$) values gives a trimmed mean of $\bar{x}_{tr} = 107.1/13 = 8.24$ EU/mg. The corresponding trimming percentage is $100(1/15) \approx 6.7\%$. The trimmed mean is below, though not far from, the mean and median of the entire sample.

d. Determine the value of the sample standard deviation for each sample, interpret these values, and then compare the variability in the two samples.

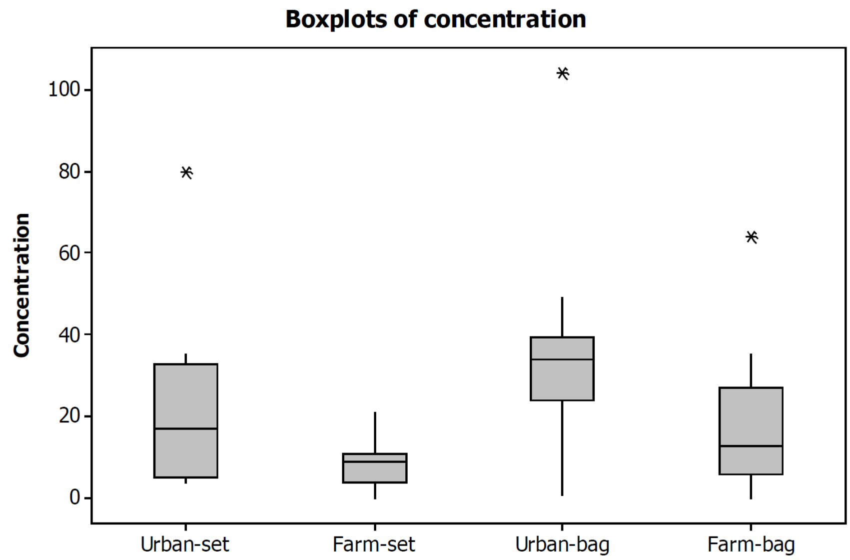
Using the sums provided for urban homes, $S_{xx} = 10{,}079 - (237.0)^2/11 = 4972.73$, so $s = \sqrt{\dfrac{4972.73}{11-1}} = 22.3$ EU/mg. Similarly for farm homes, $S_{xx} = 518.836$ and $s = 6.09$ EU/mg.
The endotoxin concentration in an urban home "typically" deviates from the average of 21.55 by about 22.3 EU/mg. The endotoxin concentration in a farm home "typically" deviates from the average of 8.56 by about 6.09 EU/mg. (These interpretations are very loose, especially since the distributions are not symmetric.) In any case, the variability in endotoxin concentration is far greater in urban homes than in farm homes.

e. The authors of the cited article also provided endotoxin concentrations in dust bag dust, taken from vacuum cleaners used to clean homes in these environments:

UB: 34.0, 49.0, 13.0, 33.0, 24.0, 24.0, 35.0, 104.0, 34.0, 40.0, 38.0, 1.0
FB: 2.0, 64.0, 6.0, 17.0, 35.0, 11.0, 17.0, 13.0, 5.0, 27.0, 23.0, 28.0, 10.0, 13.0, 0.2

construct a comparative boxplot to compare and contrast the four samples.

In the real test ,you would be given options to choose from in moodle. Here, you should generate box plots using python.



5. Poly(3-hydroxybutyrate) (PHB), a semicrystalline polymer that is fully biodegradable and biocompatible, is obtained from renewable resources. From a sustainability perspective, PHB offers many attractive propertie though it is more expensive to produce than standard plastics. The accompanying data on melting point (°C) for each of 12 specimens of the polymer using a differential scanning calorimeter appeared in the article "The Melting Behaviour of Poly(3-Hydroxybutyrate) by DSC. Reproducibility Study" (Polymer Testing, 2013: 215–220).

180.5, 181.7, 180.9, 181.6, 182.6, 181.6, 181.3, 182.1, 182.1, 180.3, 181.7, 180.5

Compute the following:

a. The sample range

The maximum and minimum values are 182.6 and 180.3, respectively, so the range is 182.6 – 180.3 = 2.3°C.

b. The sample variance s^2

	x_i	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	x_i^2
	180.5	–0.90833	0.82507	32580.3
	181.7	0.29167	0.08507	33014.9
	180.9	–0.50833	0.25840	32724.8
	181.6	0.19167	0.03674	32978.6
	182.6	1.19167	1.42007	33342.8
	181.6	0.19167	0.03674	32978.6
	181.3	–0.10833	0.01174	32869.7
	182.1	0.69167	0.47840	33160.4
	182.1	0.69167	0.47840	33160.4
	180.3	–1.10833	1.22840	32508.1
	181.7	0.29167	0.08507	33014.9
	180.5	–0.90833	0.82507	32580.3
sums:	2176.9	0	5.769167	394913.6
	$\overline{x}$ = 181.41			

$s^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2/(n-1) = 5.769167/(12 – 1) = 0.52447.$

c. The sample standard deviation

$$s = \sqrt{0.52447} = 0.724$$

6. Four hocket teams —1, 2, 3, and 4—are participating in a hockey tournament. In the first round, 1 will play 2 and 3 will play 4. Then the two winners will play for the championship, and the two losers will also play. The outcome of the tournament is a final ranking. One possible outcome can be denoted by 1324 (1 beats 2 and 3 beats 4 in first-round games, and then 1 beats 3 and 2 beats 4).

   a. List all outcomes in S

S = {1324, 1342, 1423, 1432, 2314, 2341, 2413, 2431, 3124, 3142, 4123, 4132, 3214, 3241, 4213, 4231}.

   b. What is the size of the sample space in terms of number of outcomes?

16.

   c. Let A denote the event that team 1 wins the tournament. List outcomes in A.

Event A contains the outcomes where 1 is first in the list:
A = {1324, 1342, 1423, 1432}.

   d. Let B denote the event that 2 gets into the championship game. List outcomes in B.

Event B contains the outcomes where 2 is first or second:
B = {2314, 2341, 2413, 2431, 3214, 3241, 4213, 4231}.

e. What are the outcomes in $A \cup B$ and in $A \cap B$?

The event $A \cup B$ contains the outcomes in A or B or both:
$A \cup B$ = {1324, 1342, 1423, 1432, 2314, 2341, 2413, 2431, 3214, 3241, 4213, 4231}.
$A \cap B = \emptyset$, since 1 and 2 can't both get into the championship game.

f. What are the outcomes in A'?

$A' = S - A$ = {2314, 2341, 2413, 2431, 3124, 3142, 4123, 4132, 3214, 3241, 4213, 4231}.

7. Suppose that 55% of all adults regularly consume coffee, 45% regularly consume carbonated soda, and 70% regularly consume at least one of these two products.

a. What is the probability that a randomly selected adult regularly consumes both coffee and soda?

Let A = an adult consumes coffee and B = an adult consumes carbonated soda. We're told that $P(A) = .55$, $P(B) = .45$, and $P(A \cup B) = .70$.

a. The addition rule says $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, so $.70 = .55 + .45 - P(A \cap B)$ or $P(A \cap B) = .55 + .45 - .70 = .30$.

b. What is the probability that a randomly selected adult doesn't regularly consume at least one of these two products?

There are two ways to read this question. We can read "does not (consume at least one)," which means the adult consumes neither beverage. The probability is then P(neither A nor B) $= P(A' \cap B') = 1 - P(A \cup B) = 1 - .70 = .30$.

The other reading, and this is presumably the intent, is "there is at least one beverage the adult does not consume, i.e. $A' \cup B'$. The probability is $P(A' \cup B') = 1 - P(A \cap B) = 1 - .30$ from **a** $= .70$. (It's just a coincidence this equals $P(A \cup B)$.)

Both of these approaches use *deMorgan's laws*, which say that $P(A' \cap B') = 1 - P(A \cup B)$ and $P(A' \cup B') = 1 - P(A \cap B)$.

8. A certain system can experience three different types of defects. Let A_i (i = 1,2,3) denote the event that the system has a defect of type i. Suppose that:

$P(A_1) = .12$ $P(A_2) = .07$ $P(A_3) = .05$
$P(A_1 \cup A_2) = .13$ $P(A_1 \cup A_3) = .14$
$P(A_2 \cup A_3) = .10$ $P(A_1 \cap A_2 \cap A_3) = .01$

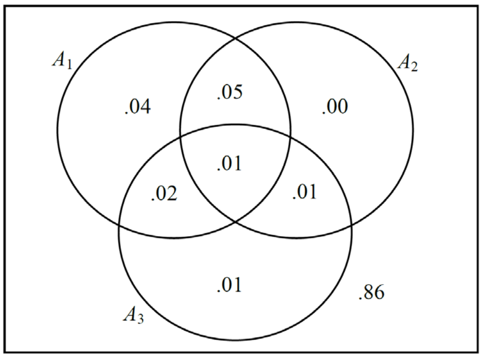
a. What is the probability that the system does not have a type 1 defect?

$P(A_1') = 1 - P(A_1) = 1 - .12 = .88$.

b. What is the probability that the system has both type 1 and type 2 defects?

The addition rule says $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Solving for the intersection ("and") probability, you get $P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) = .12 + .07 - .13 = .06$.

c. What is the probability that the system has both type 1 and type 2 defects but not a type 3 defect?



A Venn diagram shows that $P(A \cap B') = P(A) - P(A \cap B)$. Applying that here with $A = A_1 \cap A_2$ and B $= A_3$, you get $P([A_1 \cap A_2] \cap A_3') = P(A_1 \cap A_2) - P(A_1 \cap A_2 \cap A_3) =$ .06 – .01 = .05.

d. What is the probability that the system has at most two of these defects?

The event "at most two defects" is the complement of "all three defects," so the answer is just 1 – $P(A_1 \cap A_2 \cap A_3)$ = 1 – .01 = .99.

e. Given that the system has a type 1 defect, what is the probability that it has a type 2 defect?

$$P(A_2 \mid A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{.06}{.12} = .50$$

f. Given that the system has a type 1 defect, what is the probability that it has all three types of defects?

$$P(A_1 \cap A_2 \cap A_3 \mid A_1) = \frac{P([A_1 \cap A_2 \cap A_3] \cap A_1)}{P(A_1)} = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1)} = \frac{.01}{.12} = .0833.$$ The numerator simplifies because $A_1 \cap A_2 \cap A_3$ is a subset of A_1, so their intersection is just the smaller event.

g. Given that the system has at least one type of defect, what is the probability that it has exactly one type of defect?

For this example, you definitely need a Venn diagram. The seven pieces of the partition inside the three circles have probabilities .04, .05, .00, .02, .01, .01, and .01. Those add to .14 (so the chance of no defects is .86).

Let E = "exactly one defect." From the Venn diagram, $P(E)$ = .04 + .00 + .01 = .05. From the addition above, P(at least one defect) = $P(A_1 \cup A_2 \cup A_3) = .14$. Finally, the answer to the question is

$$P(E \mid A_1 \cup A_2 \cup A_3) = \frac{P(E \cap [A_1 \cup A_2 \cup A_3])}{P(A_1 \cup A_2 \cup A_3)} = \frac{P(E)}{P(A_1 \cup A_2 \cup A_3)} = \frac{.05}{.14} = .3571.$$ The numerator simplifies because E is a subset of $A_1 \cup A_2 \cup A_3$.

h. Given that the system has both of the first two types of defects, what is the probability that it does not have the third type of defect?

$$P(A_3' \mid A_1 \cap A_2) = \frac{P(A_3' \cap [A_1 \cap A_2])}{P(A_1 \cap A_2)} = \frac{.05}{.06} = .8333$$

9. An experimenter is studying the effects of temperature, pressure, and type of catalyst on yield from a certain chemical reaction. Three different temperatures, four different pressures, and five different catalysts are under consideration.

   a. If any particular experimental run involves the use of a single temperature, pressure, and catalyst, how many experimental runs are possible?

$n_1 = 3$, $n_2 = 4$, and $n_3 = 5$, there are (3)(4)(5) = 60 runs.

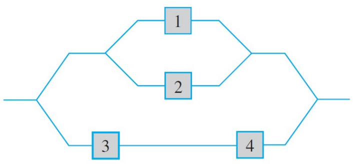
   b. How many experimental runs are there that involve use of the lowest temperature and two lowest pressures?

With $n_1 = 1$ (just one temperature), $n_2 = 2$, and $n_3 = 5$, there are (1)(2)(5) = 10 such runs.

   c. Suppose that five different experimental runs are to be made on the first day of experimentation. If the five are randomly selected from among all the possibilities, so that any group of five has the same probability of selection, what is the probability that a different catalyst is used on each run?

For each of the 5 specific catalysts, there are (3)(4) = 12 pairings of temperature and pressure. Imagine we separate the 60 possible runs into those 5 sets of 12. The number of ways to select exactly one run from each of these 5 sets of 12 is $\binom{12}{1}^5 = 12^5$. Since there are $\binom{60}{5}$ ways to select the 5 runs overall, the desired probability is $\binom{12}{1}^5 / \binom{60}{5} = 12^5 / \binom{60}{5} = .0456$.

10. Consider the system of components connected as in the accompanying picture. Components 1 and 2 are connected in parallel, so that subsystem works iff (as a reminder, this means if *and only* if) either 1 or 2 works; since 3 and 4 are connected in series, that subsystem works iff both 3 and 4 work. If components work independently of one another and P(component i works) = .9 for i = 1,2 and = .8 for i = 3,4, calculate P(system works)



Let A_i denote the event that component #i works (i = 1, 2, 3, 4). Based on the design of the system, the event "the system works" is $(A_1 \cup A_2) \cup (A_3 \cap A_4)$. We'll eventually need $P(A_1 \cup A_2)$, so work that out first: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = (.9) + (.9) - (.9)(.9) = .99$. The third term uses independence of events. Also, $P(A_3 \cap A_4) = (.8)(.8) = .64$, again using independence.

Now use the addition rule and independence for the system:

$$P((A_1 \cup A_2) \cup (A_3 \cap A_4)) = P(A_1 \cup A_2) + P(A_3 \cap A_4) - P((A_1 \cup A_2) \cap (A_3 \cap A_4))$$
$$= P(A_1 \cup A_2) + P(A_3 \cap A_4) - P(A_1 \cup A_2) \times P(A_3 \cap A_4)$$
$$= (.99) + (.64) - (.99)(.64) = .9964$$

(You could also use deMorgan's law in a couple of places.)

11. A company that produces fine crystal knows from experience that 10% of its goblets have cosmetic flaws and must be classified as "seconds."

a. Among six randomly selected goblets, how likely is it that only one is a second?

$$P(X=1) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{6}{1} (.1)^1 (.9)^5 = .3543$$

b. Among six randomly selected goblets, what is the probability that at least two are seconds?

$$P(X \geq 2) = 1 - [P(X=0) + P(X=1)] = 1 - \left[\binom{6}{0} (.1)^0 (.9)^6 + \binom{6}{1} (.1)^1 (.9)^5 \right] = 1 - [.5314 + .3543] =$$

$.1143$.

c. If goblets are examined one by one, what is the probability that at most five must be selected to find four that are not seconds?

Either 4 or 5 goblets must be selected.

Select 4 goblets with zero defects: $P(X=0) = \binom{4}{0} (.1)^0 (.9)^4 = .6561$.

Select 4 goblets, one of which has a defect, and the 5th is good: $\left[\binom{4}{1} (.1)^1 (.9)^3 \right] \times .9 = .26244$

So, the desired probability is .6561 + .26244 = .91854.

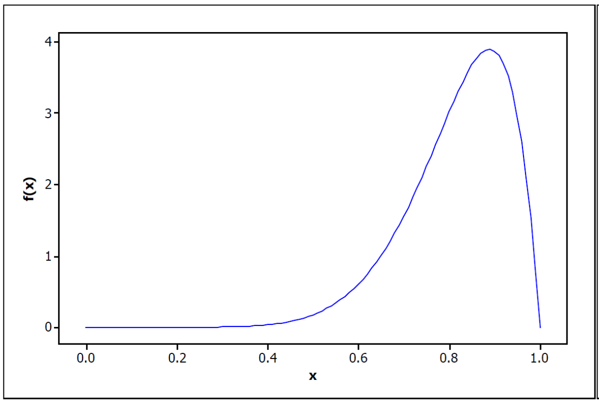
12. Let X denote the amount of space occupied by an article placed in a 1m^3 box. The pdf of X is

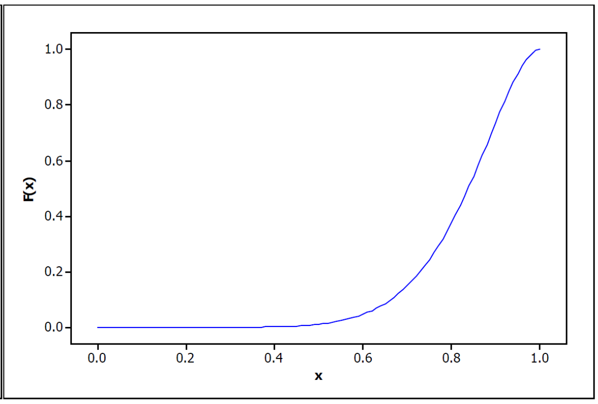
$$f(x) = \begin{cases} 90x^8(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Graph the pdf. Then obtain the cdf of X and graph it.

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_0^x 90y^8(1-y)dy = \int_0^x (90y^8 - 90y^9)dy = 10y^9 - 9y^{10} \Big]_0^x = 10x^9 - 9x^{10} .$$

The graphs of the pdf and cdf of X appear below.





b. What is P(X <= .5) [i.e., F(.5)]?

$F(.5) = 10(.5)^9 - 9(.5)^{10} = .0107.$

c. Using the cdf from (a), what is P(.25 < X <= .5)?

$P(.25 < X \leq .5) = F(.5) - F(.25) = .0107 - [10(.25)^9 - 9(.25)^{10}] = .0107 - .0000 = .0107$

d. What is P(.25 < X < .5)?

Since X is continuous, $P(.25 \leq X \leq .5) = P(.25 < X \leq .5) = .0107.$

e. What is the 75th percentile of the distribution?

The 75th percentile is the value of x for which $F(x) = .75$: $10x^9 - 9x^{10} = .75 \Rightarrow x = .9036$

f. Compute E(X) and σ_X.

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx = \int_0^1 x \cdot 90x^8(1-x)dx = \int_0^1 (90x^9 - 90x^{10})dx = 9x^{10} - \frac{90}{11}x^{11}\Big]_0^1 = 9 - \frac{90}{11} = \frac{9}{11} = .8182.$$

Similarly, $E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot 90x^8(1-x)dx = \ldots = .6818$, from which $V(X) = .6818 - (.8182)^2 = .0124$ and $\sigma_X = .11134$.

g. What is the probability that X is more than 1 standard deviation from its mean value?

$\mu \pm \sigma = (.7068, .9295)$. Thus, $P(\mu - \sigma \leq X \leq \mu + \sigma) = F(.9295) - F(.7068) = .8465 - .1602 = .6863$, and the probability X is more than 1 standard deviation from its mean value equals $1 - .6863 = 3137$.