5.1 a) why prefer an unbiased estimator to a biased estimator?
the mean squared error is minimized where
$\mu_E = \mu$

b) does an unbiased statistic ensure its a good estimator?
it is one of two criteria, the other is for it to be "stable" where the variation of Σ is small $\sigma_\Sigma << \sigma$. If another random sample is taken we dont want $\bar{x}$ to be random

c) when biased might be a better choice?
if the bias can reduce the variance such that the mean square error decreases

5.2 normal distribution, std dev = 0.5mm
res lengths: 75.3, 76.0, 75.0, 77.0, 75.4, 76.3, 77.0, 74.9, 76.5, 75.8
a) what is the parameter of interest?
The mean length of changed parts
b) 759.2/10
75.92
c) .99 confidence
lower upper $m\varepsilon = .407$
(75.51, 76.33)

5.3 why is it necessary to test a classifier with independent data
The second alg is better because the 5% error in training data allows for learning & therefore has less error for test data. A classifier needs to be tested with independent data to increase the generalizing capability of the model for unseen data.
$e = .632\, e_{test} + .368\, e_{resub}$
$e_1 = 0 + .368((.2) = .0736$ $\min(e_1, e_2) = e_2$
$e_2 = .632(.05) + .368(.1) = .0684$ therefore 2nd alg is better

Predicted class

		a	b	c	d
5.4	a	15	2	3	5
actual class	b	1	18	6	2
	c	3	5	9	0
	d	2	2	7	12

$R_s = 1 - R_e$

$$\frac{15+18+9+12}{1+3+2+2+5+2+7+3+6+5+2+0+54} \quad \frac{54}{38+54} = .587 = R_s$$

$R_e = .413$

5.5 # cases = 609

cancer = 84

not cancer = 525

	C	NC
C	TP	FP
NC	FN	TN

total # of positive = TP+FN

Success = 75%

Error = 25%

Sensitivity $= \frac{TP}{TP+FN}$ Specificity $= \frac{TN}{FP+TN}$

$FP = 525 \cdot .25 = 131$

$TN = 525 \cdot .75 = 354$

$Sens = \frac{75}{75+9} = .892$

$Spec = \frac{354}{525} = .674$

7.5 $\frac{423}{21} = 20.14 = \mu$

median $= 16$

outliers $= 69$ & 55

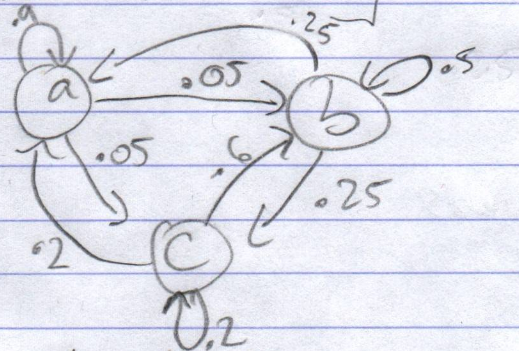
$\frac{299}{19} = 15.74$ median $= 16$

A will have a higher avg than B because the two outliers are in the upper bound of samples which increases the mean the median is the same as outliers do not outweigh the # of middle #s

7.6 Class code 2 is better because more bits offers a more accurate representation of hamming distance

B would be correct class as they have least # of differences a)

$$\begin{array}{c|ccc} & a & b & c \\ \hline a & .9 & .05 & .05 \\ b & .25 & .5 & .25 \\ c & .2 & .6 & .2 \end{array} = P$$



b)

$$\begin{array}{cccc} & L & L1 & L2 \\ S_0 = & [.8 & .1 & .1] \end{array}$$

$$S_1 = S_0 P = \underset{1\times 3}{[.8 \quad .1 \quad .1]} \underset{3\times 3}{\begin{bmatrix} .9 & .05 & .05 \\ .25 & .5 & .25 \\ .2 & .6 & .2 \end{bmatrix}} =$$

$$\begin{array}{cccc} & L & L1 & L2 \\ S_1 = & [.765 & .15 & .085] \end{array}$$

$$\begin{array}{cccc} & L & L1 & L2 \\ S_1 P = S_2 = & [.743 & .16425 & .09275] \end{array}$$

$$S_6 = S_0 P^6 = [.7078 \quad .1900 \quad .1025]$$

$$S_0 P^{99} = [.6993 \quad .1958 \quad .1049]$$

$K = 15 \quad SP = S$

$$S_{15} = \underset{S}{[.6995 \quad .1956 \quad .1048]} \overset{P}{\begin{bmatrix} .9 & .05 & .05 \\ .25 & .5 & .25 \\ .2 & .6 & .2 \end{bmatrix}} =$$

$$SP = S = [.6994 \quad .1957 \quad .1049]$$