# Midterm Exam

4|25

95

22

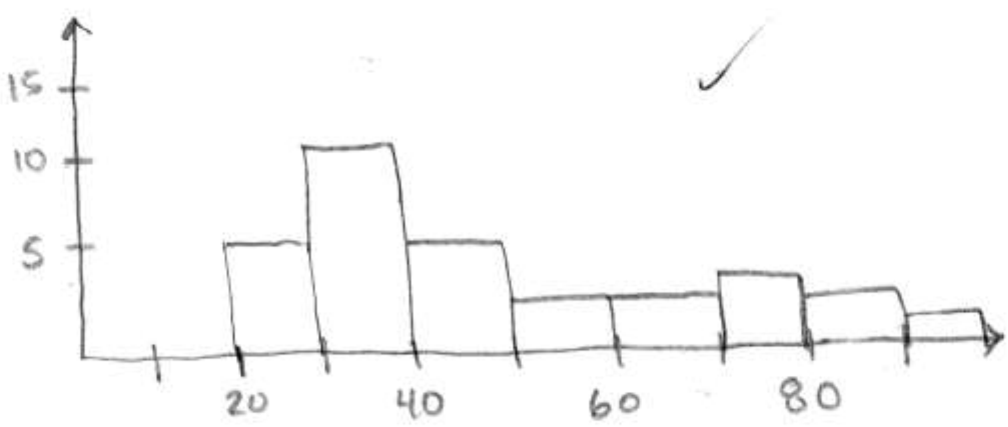
Answer all questions.

Partial credit generously given; show what you know!

1. Daily air quality impact index (AQI) data for Alameda county, for January 2008, sorted from lowest (good) to highest (bad), are:

| | | | | | | |
|---|---|---|---|---|---|---|
| 22 | 32.5 27 | 27 | 28 | 29 5 | 30 | 31 |
| 32 | 33 Median | 33 | 33 | 34 | 38 | 39 |
| 39 | 39 ·11 | 40 | 40 | 44 | 45 | 49 5 |
| 56 | 58 2 | 61 61.5 | 62 2 | 71 | 72 | 74 3 |
| 80 | 81 2 | 90 1 | | | | |

   a) Construct a histogram for these data.
   b) Sketch part a cumulative frequency diagram. Plot it for AQI values up to 27, and then starting at 80.
   c) What as the median AQI for Alameda county in January 2008?
   d) The sum of these index values (22+27+...+90) is 1437, and the sum of the squares of the values ($22^2+27^2+...+90^2$) is 77271. What are the variance and standard deviation for these data?
   e) What is the approximate inter-quartile range for these data?
   f) Comment on the skewness of the data.




2. Major earthquakes have occurred on the Hayward fault in the years 1315, 1470, 1630, 1725, and 1868. Historical records do not allow earthquakes earlier than 1300 to be identified. Suppose an earthquake forecasting system that can predict the occurrence of an earthquake sometime in the next year is devised. It is believed that if a major earthquake occurs, this system will predict it 95% of the time. However, the system also has a false alarm rate of 5%. In other words, if there is a year without a major earthquake, 5% of the time the system will predict that an earthquake will happen in that year. If the system predicts a major earthquake in the next year, what is the probability that a major earthquake will occur?

E = an earthquake occurs in that year

A = the system predicts an earthquake in that year

# of years btwn earthquakes:

155 160 95 143 $\bar{x} = 138$

$$\text{Know} \begin{cases} P(E) = \frac{1}{138} = 0.007246 \\ P(A|E) = 0.95 \\ P(A|E^c) = 0.05 \end{cases}$$

Bayes

$$\text{Want}: \quad P(E|A) = \frac{P(A|E)P(E)}{P(A|E)P(E) + P(A|E^c)P(E^c)}$$

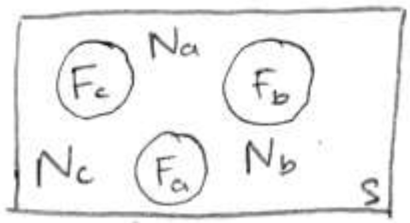
$$= \frac{(0.95)(0.007246)}{(0.95)(0.007246) + (0.05)(1 - 0.007246)}$$

$$= 0.1218$$

$$\boxed{12.18\% \text{ chance of E}}$$

3

3. A structural load bearing test is performed on three concrete beams, labeled a, b, and c. Each beam either fails or does not fail in the test. Tests results for the three beams are independent. Because of different fabrication methods the probability of failure of beams a, b, and c are ¼, ½, and ½ respectively.
   a) What is the sample space for the experiment described above?
   b) What is the event that exactly one beam fails?
   c) Let A be the event that no beam fails and B be the event that exactly one beam fails.
      a. What is P(AB)?
      b. What is P(AUB)?
   d) Let X be the number of beams that fail the test
      a. What is the PMF for X?
      b. What is E(X)?
      c. What is VAR(X)?




$F_a, N_a, F_b, N_b, F_c, N_c$

F = fails

N = not fail = $F^c$

$S(a,b,c) = \{(N_a, N_b, N_c), (F_a, N_b, N_c), (N, F, N), (N, N, F), (F, F, N), (F, N, F), (N, F, F), (F, F, F)\}$

) $\{(F,N,N), (N,F,N), (N,N,F)\}$

$(F_a \cap N_b \cap N_c) \cup (N_a \cap F_b \cap N_c) \cup (N_a \cap N_b \cap F_c)$

) a. $\emptyset$

b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.4375 + 0.1875$
$= \boxed{0.625}$

$P(A) = P(\ )$
$= (\frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{2}) +$
$(\frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2}) +$
$(\frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2}) - 0$
$= 0.4375$

$P(B) = P(F_a^c F_b^c F_c^c)$
$= \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{2}$
$= 0.1875$

d)

a.

| $x$ | $p(x)$ |
|---|---|
| 1 | 0.4375 |
| 2 | 0.3125 |
| 3 | 0.0625 |
| 0 | 0.1875 |

b. $E[x] = \sum_{i=0}^{3} x\,p(x) = \boxed{1.25 \text{ fail}}$

c. $\mathrm{var}(X) = E[(x - E[X])^2] = \sum_{i=0}^{3} (x - E[X])^2 \cdot p(x)$

Answer the following multiple choice questions

4. A joint pmf
   a. Gives the probability that a set of discrete random variables take a particular set of values.
   b. Gives the probability that a set of continuous random variables takes a particular set of values.
   c. Is equivalent to the empty set for a continuous sample space.
5. If f(y) is a PDF then which of the following must be true?
   a. $\int_a^b f(y)dy \leq 1$
   b. $f(y) \leq 1$
   c. $f(\infty) = 1$
6. The graph of a CDF for a discrete random variable normally looks like
   a. An elevator
   b. A parabola
   c. A straight line
   d. A staircase
7. Let X and Y be jointly distributed, independent, continuous random variables with joint PDF $f_{XY}(x, y)$. Then:
   a. $f_{XY}(x,y) = f_X(x) + f_Y(y)$
   b. $f_{XY}(x,y) = f_X(x) / f_Y(y)$
   c. $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$
   d. $f_{XY}(x,y) = f_{XY}(xy)$
8. Let A and B be two events. Suppose $P(A \cup B) = P(A) + P(B)$. Then:
   a. A and B are independent.
   b. P(AB)=0
   c. P(AB)=1
   d. P(A)=P(B)

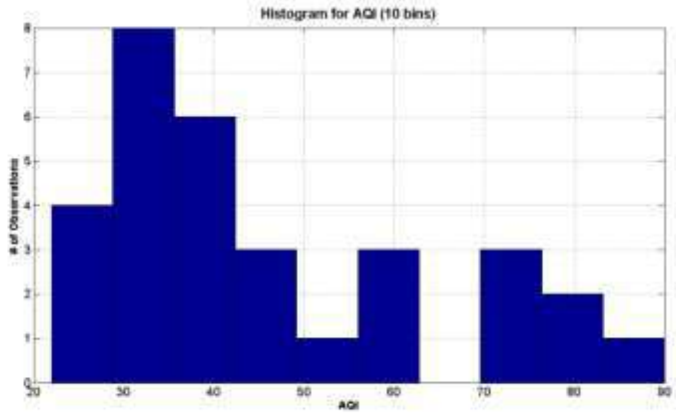
# Midterm Solution

**1. AQI values, sorted from lowest to highest**

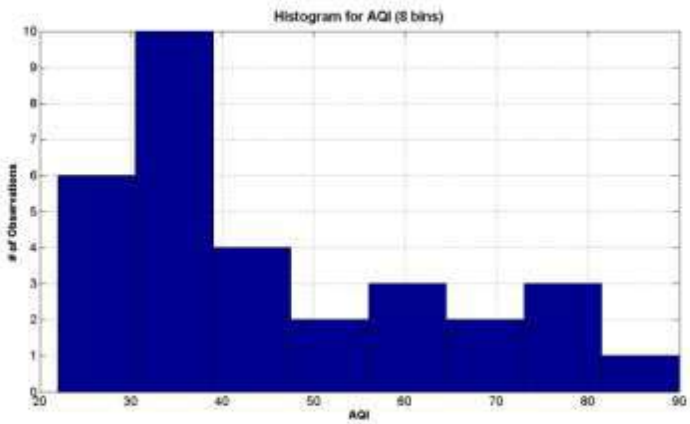
**a. Construct a histogram of these data**

Anything similar to one of these will work. The number of bins shouldn't be too small or too large.

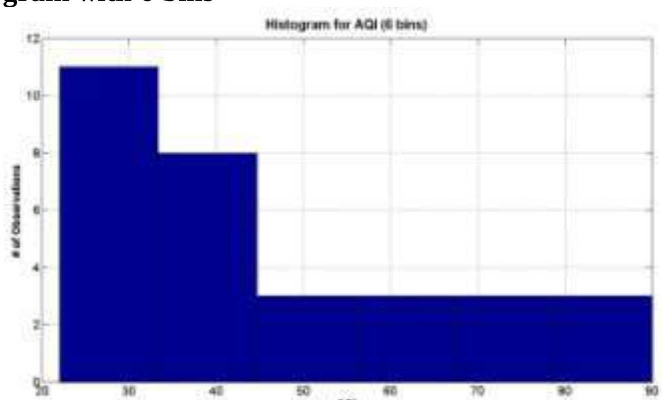
**Histogram with 10 bins**




**Histogram with 8 bins**



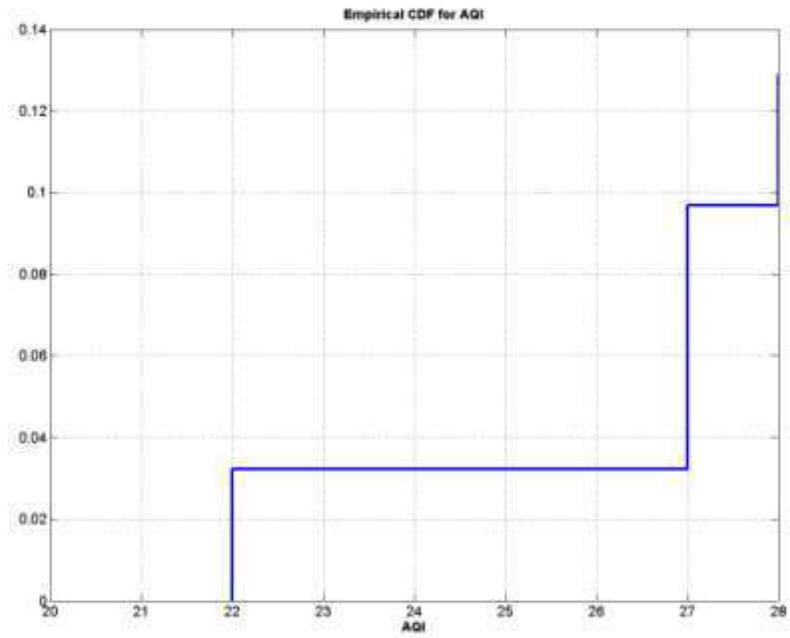

**Histogram with 6 bins**



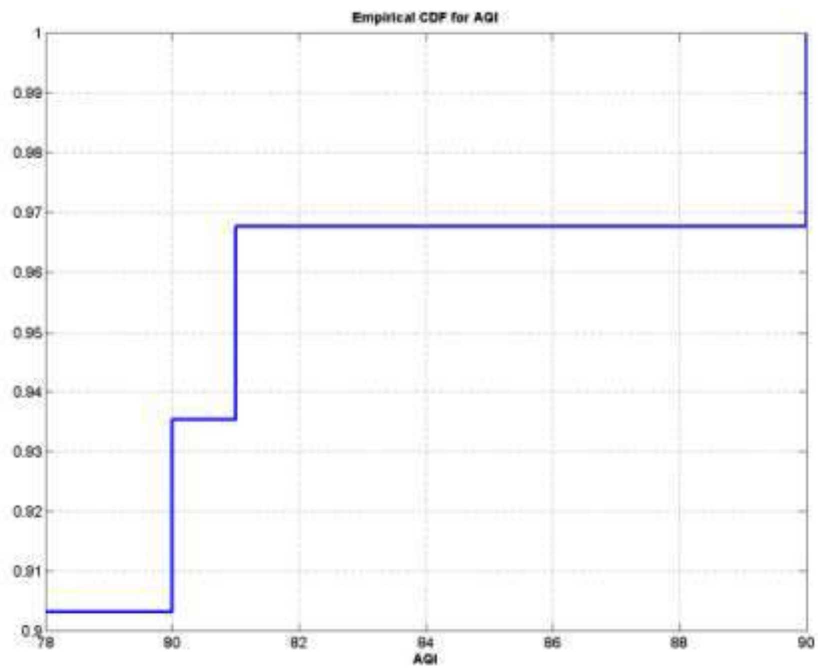



**b. Sketch part of a cumulative frequency diagram. Plot it for AQI values up to 27, then starting at 80.**

**CDF up to 27**




**CDF starting at 80**




**c. What is the median AQI for Alameda county in January 2008?**



There are 31 observations. The 16th observation, when ranked from low to high, is 39.

**d. The sum of these index values (22 + 27 + … + 90) is 1437, and the sum of the squares of the values ($22^2 + 27^2 + … + 90^2$) is 77,271. What are the variance and the standard deviation for these data?**

$$\bar{x} = \frac{\sum_{i=1}^{31} x_i}{n} = \frac{1437}{31} = 46.354$$

$$\text{Variance: } s^2 = \frac{\sum_{i=1}^{31}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{31} {x_i}^2}{n} - \left(\frac{\sum_{i=1}^{31} x_i}{n}\right)^2 = \frac{77271}{31} - \left(\frac{1437}{31}\right)^2 = 2492.6 - 2148.8 = 343.8$$

$$\text{Standard Deviation: } s = \sqrt{s^2} = \sqrt{348.8} = 18.54$$

**e. What is the approximate inter-quartile range for these data?**

Inter-quartile range = IQR = $X_{0.75} - X_{0.25}$, or the difference between the 75th and 25th percentile values.

With 31 observations, the observation 32 is roughly at the 25th percentile (8 observations at or below = about ¼ of all observations at or below). Also accepted would be 31 or 33. For the 75th percentile, we get 61 using the same method (also accepted would be 58 or 62). Using these ranges of percentiles, you could get anywhere from 25 (58-33) to 31 (62-31) for the inter-quartile range.

**f. Comment on the skewness of the data.**

Based on the histogram, the data appears to be positively skewed. That is, there appears to be a much larger tail to the right than there is to the left. Also, the mean of the data (46.3) is larger than the median (39), which is an indication of outliers to the right of the median, and thus positive skewness.

2. E : A major earthquake happens

P : The system predict a major earthquake

$P(P|E) = 0.95$ (so $P(P^c|E) = 0.05$)

$P(P|E^c) = 0.05$ (so $P(P^c|E^c) = 0.95$)

From 1300 to 2010, we have 710 observations

# of major earthquakes = 5

So $P(E) = \frac{5}{710}$, and $P(E^c) = 1 - \frac{5}{710} = \frac{705}{710}$

$$P(E|P) = \frac{P(PE)}{P(P)} = \frac{P(P|E)\cdot P(E)}{P(P|E)\cdot P(E) + P(P|E^c)\cdot P(E^c)}$$

$$= \frac{0.95 \times \frac{5}{710}}{0.95 \times \frac{5}{710} + 0.05 \times \frac{705}{710}}$$

$$= \frac{0.95}{0.95 + 0.05 \times 141} = \frac{0.95}{8} \approx 0.11875$$

3. $F_a$ : concrete beam a fails.

Similarly, we can define $F_b$ & $F_c$.

$P(F_a) = \frac{1}{4}$, $P(F_b) = \frac{1}{2}$, $P(F_c) = \frac{1}{2}$

So $P(F_a^c) = 1 - \frac{1}{4} = \frac{3}{4}$

$P(F_b^c) = P(F_c^c) = \frac{1}{2}$

a) $S = \{F_a F_b F_c,\ F_a F_b^c F_c,\ F_a F_b^c F_c^c,\ F_a F_b F_c^c,$

$F_a^c F_b F_c,\ F_a^c F_b^c F_c,\ F_a^c F_b^c F_c^c,\ F_a^c F_b F_c^c\}$

b) Event = $\{F_a F_b^c F_c^c,\ F_a^c F_b F_c^c,\ F_a^c F_b^c F_c\}$

c) $A = \{F_a^c F_b^c F_c^c\}$, $B = \{F_a F_b^c F_c^c,\ F_a^c F_b F_c^c,\ F_a^c F_b^c F_c\}$

$AB = \phi$

$P(AB) = 0$

$P(A \cup B) = P(A) + P(B) - P(AB) = P(A) + P(B)$

$P(A) = P(F_a^c)\cdot P(F_b^c)\cdot P(F_c^c)$ (By independence)

$= \frac{3}{4} \times \frac{1}{2} \times \frac{1}{2} = \frac{3}{16}$

$P(B) = P(F_a F_b^c F_c^c) + P(F_a^c F_b F_c^c) + P(F_a^c F_b^c F_c)$ (Mutually exclusive)

$= \frac{1}{4} \times \frac{1}{2} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2} \times \frac{1}{2} + \frac{3}{4} \times \frac{1}{2} \times \frac{1}{2} = \frac{7}{16}$

$P(A \cup B) = \frac{3}{16} + \frac{7}{16} = \frac{5}{8}$

d) $X = 0 \Leftrightarrow F_a^c F_b^c F_c^c$

$P(X=0) = P(A) = \frac{3}{16}$

$P(X=1) = P(B) = \frac{7}{16}$

$P(X=3) = P(F_a F_b F_c) = \frac{1}{4} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

$P(X=2) = 1 - P(X=0) - P(X=1) - P(X=3) = \frac{5}{16}$

$$\text{PMF of } X = P(x) = \begin{cases} \frac{3}{16}, & x = 0 \\ \frac{7}{16}, & x = 1 \\ \frac{5}{16}, & x = 2 \\ \frac{1}{16}, & x = 3 \\ 0, & \text{otherwise} \end{cases}$$

$E[X] = \sum_{\text{all } x} x \cdot P(x)$

$= 0 \cdot \frac{3}{16} + 1 \cdot \frac{7}{16} + 2 \cdot \frac{5}{16} + 3 \cdot \frac{1}{16} = \frac{20}{16} = \frac{5}{4} = 1.25$

$var(X) = E[X^2] - (E[X])^2$

$E[X^2] = \sum_{\text{all } x} x^2 \cdot P(x)$

$= 0^2 \cdot \frac{3}{16} + 1^2 \cdot \frac{7}{16} + 2^2 \cdot \frac{5}{16} + 3^2 \cdot \frac{1}{16} = \frac{36}{16} = \frac{9}{4}$

So $var(X) = \frac{9}{4} - (\frac{5}{4})^2 = \frac{9}{4} - \frac{25}{16} = \frac{11}{16} = 0.6875$

4. a
5. a
6. d
7. c
8. b