Biostatistics 140.621

Some Extra Problems

1. The following graphs show the blood lead level distribution for children aged 1 - 5 in the United States during two time periods of the National Health and Nutrition Examination Survey (NHANES) (JAMA 1994;272:284-291). From these graphs, it can be seen that: (Circle either True or False for each statement.)

T F The median blood lead level is higher in 1988 - 1991.

T F The range of blood lead levels was wider in 1976 -1980.

T F The cumulative relative frequency of blood lead levels less than or equal to 0.48 mol/L (10 ug/dL) differs appreciably by time period.

T F The mean and median blood lead levels are similar in 1988 - 1991.

T F The interquartile range of blood lead levels is wider in 1976 - 1980.

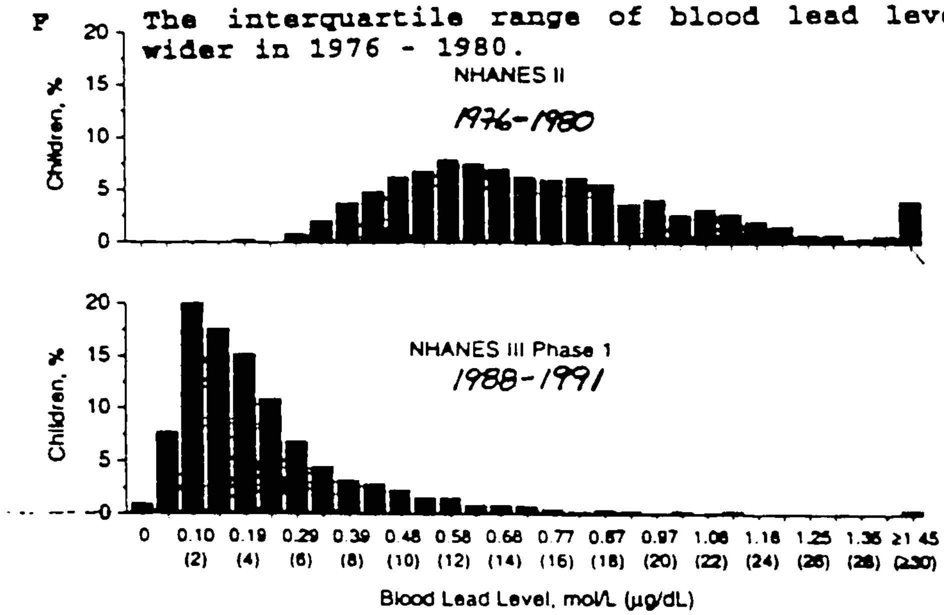
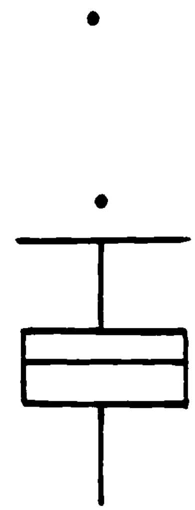


Fig 2.—Blood lead levels for children aged 1 to 5 years: United States, second National Health and Nutrition Examination Survey (1976 to 1980, top) and phase 1 of the third National Health and Nutrition Examination Survey (1988 to 1991, bottom).

2. The following are serum glucose levels (mg/dl) for 26 stroke patients. Given the following stem and leaf display and boxplot (Fill in the blanks):

```
21 | 7
20 |
19 |
18 |
17 |
16 |
15 | 1 9
14 |
13 |
12 | 3 4 7
11 | 4 7 8 9
10 | 3 5 5 7
 9 | 0 3 4 4 5 7 8
 8 | 5 7
 7 | 2 6
 6 | 8
```



a. 75% of the serum glucose levels are above the value of ___________.

b. The mean of this distribution is ___________ than the median.

c. The interquartile range is ___________ mg/dl.

d. Outliers are the values of ___________.

e. The 50th percentile is also known as the ___________.

f. 25% of the serum levels are below the value of ___________.

3. The following figure shows the results of a survey which measured weekly average nicotine exposures in the workplace (JAMA 1995; 274:956-960). From these graphs, it can be seen that: (Fill in the blanks.)

   a. The median nicotine concentration in offices that allowed smoking is approximately ______________ µg/m³.

   b. The value of the 60th percentile in offices in which smoking was restricted is approximately ___ as compared to the 60th percentile value of ____ in offices where smoking is banned.

   c. In offices where smoking is restricted, approximately ___ % have weekly average nicotine concentrations less than 5µg/m³.

   d. Another way of graphing this data is to use a logarithmic scale. Would you recommend this? Why or why not?

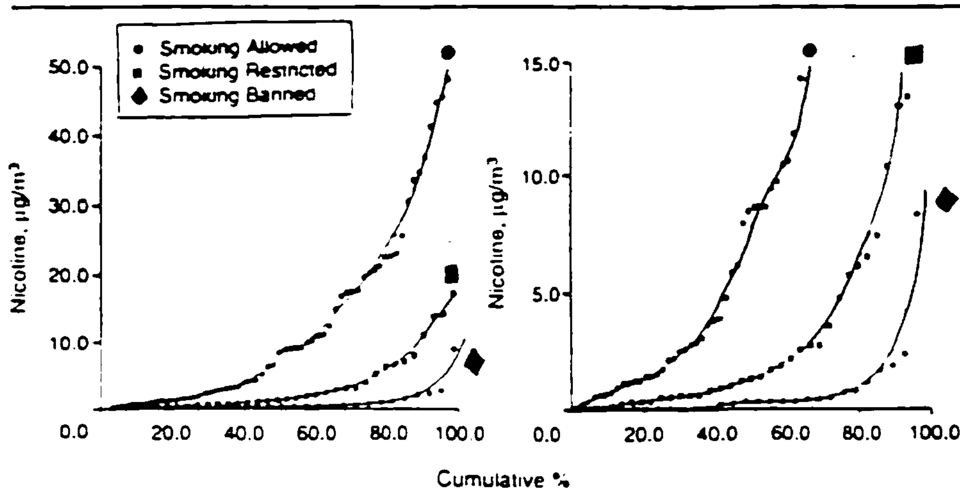


Figure 1 —Left. Cumulative frequency distributions of weekly average nicotine concentrations at the desks of nonsmoking workers in open offices measured at 25 worksites. Median (50th percentile) nicotine concentrations were 8.6, 1.3, and 0.3 µg/m³ at worksites that allowed, restricted, or banned smoking, respectively. Right, Nicotine concentrations at the desks of nonsmoking office workers, with detail provided at concentrations less than 15 µg/m³.

4. The following figure shows more results from this survey which measures exposures to environmental tobacco smoke (ETS) in workplaces where smoking was allowed. Also shown are weekly average nicotine concentrations from personal samples of non-smokers from these offices and home samples from homes of smokers. From these plots, it can be seen that:

(Circle either True or False for each statement.)

T F The distributions of weekly average nicotine concentration are similar in both workplace samples.

T F Fences and outliers were calculated for these boxplots.

T F The whiskers on these boxplots are drawn to the minimum and maximum nicotine concentrations.

T F The boxplots of the personal samples indicate no difference in ETS exposure by smoking status of spouse.

(Fill in the blanks.)

A. The median nicotine concentration in open offices is approximately _____ µg/m³ compared to the median of _____ µg/m³ for shops.

B. The nicotine concentration in personal samples of non-smokers married to smokers is more than ______ µg/m³ in 25% of samples.

C. The interquartile range in the open office samples is approximately ___________ µg/m³.

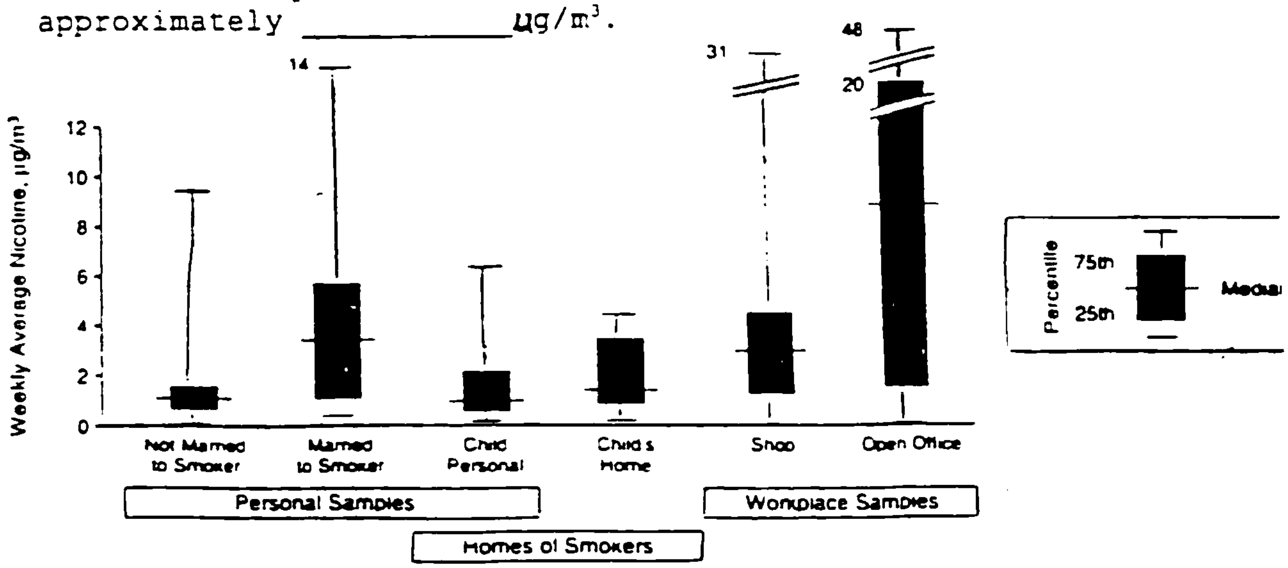


Figure 3 —Weekly average nicotine exposures in the home and at work by personal sampling (first two bars), in smokers' homes, area and personal sampling (middle two bars), and at work (this study) (last two bars)

5. The following figures show the characteristics of households which were included in a large, longitudinal study of risk factors for leprosy in a rural district of Malawi (AJE 1997: 146:91-102). From these figures, it can be seen that: (Fill in the blanks.)

a. The median number of individuals per household is ______________.

b. Approximately _______ % of the households have more than one individual.

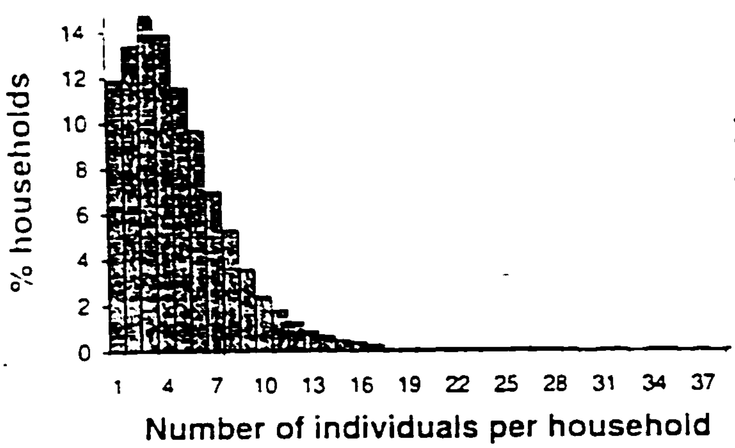
c. The first quartile of the distribution of the number of individuals per household is approximately __________.

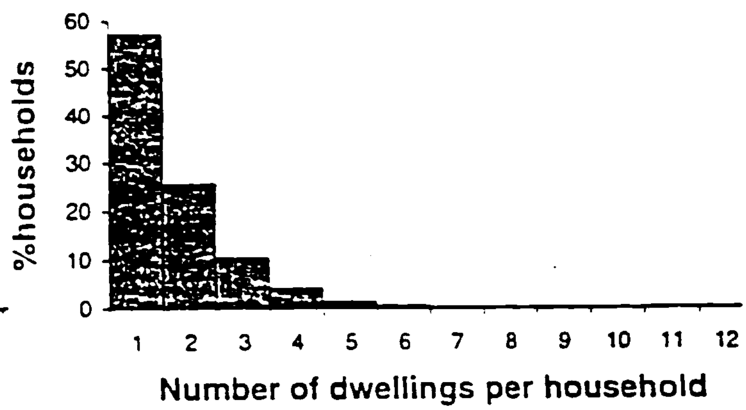
d. The third quartile of the distribution of the number of individuals per household is approximately ___________.

e. The interquartile range of the distribution of the number of individuals per household is approximately __________.

f. The range of the number of individuals per household is ______________.

g. Only __________% of the households have more than one associated dwelling.





6. The following box plots show blood lead levels in the patella (knee) bone by age among participants in the Normative Aging Study (JAMA 1996; 275:1049-1052).

(Circle either True or False for each statement.)

T F Fences are actual observations in a data set.

T F The median patella bone lead level decreases with increasing age.

T F The interquartile range increases with increasing age.

T F The mean blood lead levels are shown in these plots.

(Fill in the blanks.)

a. In the age group 70 to < 75, outliers are values greater than approximately _____ μg/g

b. In the age group > 75 years, 75% of the observations are less than ______μg/g.

c. There are no outliers in age group __________.

d. In age group 65 to < 70, 25% of the observations are less than _________ μg/g.

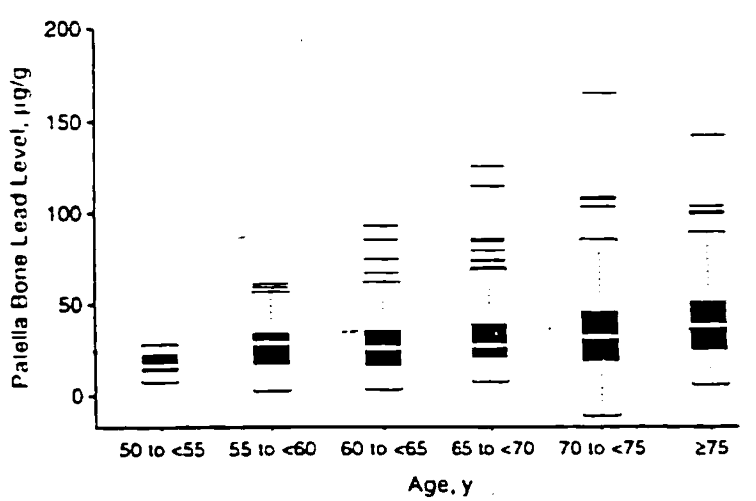


Figure 2.—Box plot of patella lead levels measured by K x-ray fluorescence vs age among participants in the Normative Aging Study. The horizontal line in the interior of the box is located at the median of the data. The box describes the interquartile distance (IQD) between the third quartile of the data and the first quartile. The dotted lines extend a distance of 1.5 × IQD from the center to the "whiskers." The data bars that fall outside the whiskers may be considered outliers.

Problems 7a through 7d concern the following table which shows the results of intragrader variability in assessing cervical smears for the presence or absence of abnormal squamous cells. Each slide was reviewed by a particular grader and then re-graded 6 months later by the same grader.

First Grading	Second Grading Present	Second Grading Absent	Total
Present	1763	489	2252
Absent	403	670	1073
Total	2166	1159	3325

7a. What is the probability that abnormal squamous cells were "absent" in both gradings? (Check only one response.)

( ) a. 0.80
( ) b. 0.53
( ) c. 0.20
( ) d. 0.35
( ) e. 0.68

7b. What is the probability of a grading of "absent" in the second grading given that abnormal cells were "present" in the first grading? (Check only one response.)

( ) a. 0.15
( ) b. 0.22
( ) c. 0.42
( ) d. 0.19
( ) e. 0.27

7c. What is the probability of a grading of "present" in the second grading given that abnormal cells were "absent" in the first grading? (Check only one response.)

( ) a. 0.65
( ) b. 0.19
( ) c. 0.22
( ) d. 0.38
( ) e. 0.78

7d. What is the probability that the gradings disagree? (Check only one response.)

( ) a. 0.27
( ) b. 0.73
( ) c. 0.59
( ) d. 0.41
( ) e. 1.00

Problems 8a through 8 concern the following list showing the primary tumor sites for 2,000 patients:

Site	Frequency
Lung	680
Colon	420
Breast	200
Head and neck	140
Other	560
	2,000

8a. What is the probability that a patient in this study has the primary tumor site in the lung or colon? (Check only one response.)

( ) a. 0.34
( ) b. 0.21
( ) c. 1.00
( ) d. 0.55
( ) e. 0.07

8b. What is the probability that a patient in this study has the primary tumor site in the breast? (Check only one response.\

( ) a. 0.10
( ) b. 0.90
( ) c. 0.20
( ) d. 0.80

8c. In a random group of 5 patients from the study, what is the probability that none of the 5 patients have the primary tumor site in the breast? (Check only one response.)

( ) a. 0.3280
( ) b. 0.5000
( ) c. 0.5905
( ) d. 0.00001
( ) e. 1.0000

8d. In a random group of 5 patients from the study, what is the probability that two patients have the primary tumor site in the breast? (Check only one response.)

( ) a. 0.0729
( ) b. 0.9185
( ) c. 0.9914
( ) d. 0.0073
( ) e. 0.2000

Problems 9a through 9h concern the following: For a particular population, the lifetime probability of developing glaucoma is approximately 0.007 and the lifetime probability of developing diabetes is 0.020. Suppose the probability of an individual developing both diseases in a lifetime is 0.0008.

9a. What is the probability of not developing glaucoma for an individual in this population? (Check only one response.)

( ) a. 0.007
( ) b. 0.020
( ) c. 0.993
( ) d. 0.980
( ) e. 0.9992

9b. What is the probability of not developing diabetes for an individual in this population? (Check only one response.)

( ) a. 0.007
( ) b. 0.020
( ) c. 0.993
( ) d. 0.980
( ) e. 0.9992

9c. What is the probability of developing either diabetes or glaucoma for an individual in this population? (Check only one response.)

( ) a. 0.027
( ) b. 0.0278
( ) c. 0.0262
( ) d. 0.0008
( ) e. 0.35

9d. What is the probability of developing glaucoma given a individual has diabetes? (Check only one response.)

( ) a. 0.04
( ) b. 0.11
( ) c. 0.35
( ) d. 0.0008
( ) e. 0.0192

9e. What is the probability of developing diabetes given an individual has glaucoma? (Check only one response.)

( ) a. 0.04
( ) b. 0.11
( ) c. 0.0008
( ) d. 0.0192
( ) e. 0.35

9f. In this population, is developing glaucoma independent of developing diabetes? (Check only one response.)

( ) a. Yes, because there are few individuals with both conditions.
( ) b. No, because the probability of developing diabetes is not influenced by glaucoma
( ) c. Yes, because the probability of developing diabetes. is influenced by glaucoma.
( ) d. No, because the probability of developing glaucoma is influenced by diabetes

9g. In a random group of 10 people from this population, what is the probability that at least one person develops diabetes? (Check only one response.)

( ) a. 0.817
( ) b. 0.183
( ) c. 0.02
( ) d. 0.167

9h. In a random group of 10 people from this population, what is the probability that only the second and seventh individuals develop diabetes? (Check only one response.)

( ) b. 0.0153
( ) c. 0.00034
( ) d. 0.04
( ) d. 0.183
( ) e. 0.0167

Problems 10a through 10d concern the following: It is given that the national infant mortality rate (IMR) for a certain year in the U.S. is 8 deaths per 1,000 live births.

10a. Suppose it is observed that the IMR in a particular region of the U.S. is 3 deaths per 1,000 live births. The probability of this observation or something smaller is: (Check only one response.)

( ) a. 0.991
( ) b. 0.647
( ) c. 0.042
( ) d. 0.224
( ) e. 0.028

10b. What would you conclude from your response in problem 4a? (Check only one response.)

( ) a. It is unlikely that this region has the same risk of infant mortality as the entire U.S.; the regional risk appears lower.
( ) a. It is unlikely that this region has the same risk of infant mortality as the entire U.S.; the regional risk appears higher.
( ) c. It is likely that this region has the same risk of infant mortality as the entire U.S.
( ) d. No conclusion can be made.

10c. What is the probability of observing 3 or fewer infant deaths per 1,000 live births during a 6-month period in the entire U.S.? (Check only one response.)

( ) a. 0.021
( ) b. 0.433
( ) c. 0.042
( ) d. 0.224
( ) e. 0.195

10d. What probability distribution was used to describe infant mortality in these problems? (Check only one response.)

( ) a. Binomial with $p = 0.008$, $n = 1{,}000$
( ) b. Poisson with $\lambda = 8$
( ) c. Normal with mean = 0.008 and standard deviation = 0.003
( ) d. Poisson with $\lambda = 3$

Problems 11a through 11e concern the following: A study was conducted to evaluate potential tuberculosis screening strategies, including symptoms history (fever, cough, night sweats, or weight loss). In this study, it was determined that the probability of being a tuberculosis case (prevalence) was 0.15. 81% of the tuberculosis cases reported symptoms. Among study participants who were not tuberculosis cases, 51% reported no symptoms.

11a. What is the overall probability of reporting symptoms in this study population? (Check only one response.)

( ) a. 0.15
( ) b. 0.85
( ) c. 0.81
( ) d. 0.51
( ) e. 0.54

11b. What is the probability of having tuberculosis or having symptoms in this study population? (Check only one response.)

( ) a. 0.12
( ) b. 0.69
( ) c. 0.57
( ) d. 0.88
( ) e. 0.43

11c. What is the probability of having tuberculosis for an individual who reports symptoms in this population? (Check only one response.)

( ) a. 0.94
( ) b. 0.22
( ) c. 0.80
( ) d. 0.51
( ) e. 0.12

11d. Which probability distribution best describes the probability of being a tuberculosis case for the group of study participants who report symptoms? (Check only one response.)

( ) a. A Poisson distribution with $\lambda = 0.15$.
( ) b. A binomial distribution with $p = 0.80$.
( ) c. A uniform distribution with probability = 0.15.
( ) d. A binomial distribution with $p = 0.22$.
( ) e. A Poisson distribution with λ equal to 0.12.

11e. In a random sample of 5 individuals from the group of study participants who report symptoms, what is the probability that only one will have tuberculosis? (Check only one response.)

( ) a. 0.9256
( ) b. 0.2297
( ) c. 0.4072
( ) d. 0.5928
( ) e. 0.6959

Problems 12a through 12g concern the following: Human blood can be classified by the ABO blood grouping system depending on whether antigens labelled A and B are present on red blood cells. Suppose for the U.S. population, the probabilities of race and blood type are as follows:

	Blood Group				
Race	A	B	AB	O	Total
Caucasian	0.36	0.06	0.02	0.36	0.80
African American	0.05	0.04	0.01	0.10	0.20
Total	0.41	0.10	0.03	0.46	1.00

12a. Which is the most common blood group? (Check only one response.)

( ) a. A
( ) b. B
( ) c. AB
( ) d. O
( ) e. Cannot tell from this table

12b. What is the probability of being an African American and in blood group B? (Check only one response.)

( ) a. 0.10
( ) b. 0.06
( ) c. 0.04
( ) d. 0.40
( ) e. 0.20

12c. What is the probability of being in blood group A given an individual is Caucasian? (Check only one response.)

( ) a. 0.88
( ) b. 0.36
( ) c. 0.86
( ) d. 0.45
( ) e. 0.64

12d. Which characteristics are mutually exclusive? (Check ALL responses that apply.)

( ) a. Classifications of blood group
( ) b. Classifications of race
( ) c. Classifications of race-blood groups
( ) d. Classifications of gender-blood groups

12e. Is blood group independent of race in this population? Why or why not? (Give a brief answer.)

____

____

12f. In a random group of 6 Caucasians, what is the probability that at least three have blood group O? (Check only one response.)

( ) a. 0.4415
( ) b. 0.5585
( ) c. 0.7447
( ) d. 0.0370
( ) e. 0.2779

12g. In a random group of 6 Caucasians, what is the probability that only the first two or only the last two have blood group O? (Check only one response.)

( ) a. 0.4415
( ) b. 0.5585
( ) c. 0.7447
( ) d. 0.0370
( ) e. 0.2779

Problems 13a through 13d concern the following: A blood screening test for diabetes was performed. Suppose 12% of the people in the population are diabetic. Also suppose that 18% tested positively for the test.

13a. If the probability of testing positively is 83% if the person is a diabetic (sensitivity), what is the probability of being a diabetic and testing positively? (Check only one response.)

( ) a. 0.01
( ) b. 0.10
( ) c. 0.20
( ) d. 0.30
( ) e. 0.40

13b. What is the probability of a person testing negatively if he/she is not diabetic (specificity)? (Check only one response.)

( ) a. .08
( ) b. .80
( ) c. .09
( ) d. .91

13c. What is the probability of being diabetic if he/she tests positively (positive predictive value)? (Check only one response.)

( ) a. 0.44
( ) b. 0.56
( ) c. 0.08
( ) d. 0.67

13d. What is the probability of being diabetic or testing positively? (Check only one response.)

( ) a. 0.08
( ) b. 0.10
( ) c. 0.20
( ) d. 0.91

Problems 14a through 14b concern the following: In a certain Village A, it is hypothesized that the probability of a child being malnourished is 0.04. There are 1,000 children in Village A.

14a. Which probability distribution would best describe the number of malnourished children in the population? (Check only one response.)

( ) a. Binomial with p=0.4 , n=1,000
( ) b. Poisson with λ = 0.04
( ) c. Normal with mean = 4 and standard deviation = 0.006
( ) d. Binomial with p =0.04, n = 1,000
( ) e. Chi-square with 999 degrees of freedom

14b. What assumptions are made with your response to problem 2a? (Check ALL responses that apply.)

( ) a. Independence of occurrences of malnutrition among children.
( ) b. Same probability of malnutrition for each child.
( ) c. Different probabilities of malnutrition for each child.
( ) d. Malnutrition is a continuous measurement.
( ) e. Malnutrition is a dichotomous event.

Problems 15a through 15c concern the following: The distribution of the Rh factor in a Caucasian population is as follows for both males and females. Individuals are assumed to be independent.

Rh Positive (Rh+,Rh+)	Rh Positive (Rh+, Rh-)	Rh Negative (Rh-, Rh-)
0.35	0.48	0.17

15a. What is the probability of any one individual being Rh Positive? (Check only one response.)

( ) a. 0.17
( ) b. 0.35
( ) c. 0.48
( ) d. 0.83

15b. In a random group of 5 individuals, what is the probability that three are Rh Negative? (Check only one response.)

( ) a. 0.17
( ) b. 0.0338
( ) c. 0.1652
( ) d. 0.51

15c. In a random group of 5 individuals, what is the probability that only the first and third individuals are Rh Negative? (Check only one response.)

( ) a. 0.17
( ) b. 0.1652
( ) c. 0.0165
( ) d. 0.34

Problems 16a through 16e concern the following: In a certain population of men aged 45-69, 30 percent are overweight and 20 percent have high blood pressure. Of those with high blood pressure, 50 percent are overweight.

16a. What is the probability that a man selected at random from this population is overweight and has high blood pressure? (Check only one response.)

( ) a. 0.01
( ) b. 0.10
( ) c. 0.20
( ) d. 0.30
( ) e. 0.40

16b. What is the probability of a man being overweight or having high blood pressure? (Check only one response.)

( ) a. 0.01
( ) b. 0.10
( ) c. 0.20
( ) d. 0.30
( ) e. 0.40

16c. What is the probability of a man having low blood pressure if he is not overweight? (Check only one response.)

( ) a. 0.86
( ) b. 0.14
( ) c. 0.20
( ) d. 0.75
( ) e. 0.60

16d. What is the probability of a man not being overweight if he does not have high blood pressure? (Check only one response.)

( ) a. 0.86
( ) b. 0.14
( ) c. 0.20
( ) d. 0.75
( ) e. 0.60

16e. Is blood pressure independent of weight? (Check only one response.)

( ) a. Yes, because there are few overweight men
( ) b. Yes, because the probability of having high blood pressure is not influenced by weight
( ) c. No, because the probability of having high blood pressure is different for the weight groups
( ) d. Yes, because Chi-square equals zero

Problems 17a through 17d concern the following tables showing the probability of death within 5 years by cause for white males ages 20 - 25:

Cause of Death	Probability
Suicide	0.00126
Homicide	0.00063
Auto Accidents	0.00581
Leukemia	0.00023
All Other Causes	0.00788

17a. What is the probability of a white male aged 20-25 dying within 5 years from any cause? (Check only one response.)

( ) a. 0.01581
( ) b. 1.0
( ) c. 0.00788
( ) d. 0.98419

17b. Out of 10,000 white males aged 20-25, how many deaths due to homicide are expected within 5 years? (Check only one response.)

( ) a. 78.88
( ) b. 31.5
( ) c. 6.3
( ) d. 50

17c. Given a person dies, what is the probability of dying from homicide? (Check only one response.)

( ) a. 0.00063
( ) b. 0.03985
( ) c. 0.00001
( ) d. 0.99937

17d. What is the probability of two unrelated individuals dying in automobile accidents? (Check only one response.)

( ) a. 0.01162
( ) b. 0.00003
( ) c. 0.00581
( ) d. 0.98010

18. Problems 18a and 18c concern the following:

In Great Britain, about one in 2000 live births are affected with cystic fibrosis, a genetic disorder. If both parents are heterogeneous for the abnormal gene, there is a 1 in 4 chance of their child having cystic fibrosis. (Ref. D. Altman).

18a. What is the probability that a couple who are both heterogeneous will have two unaffected children? (Check only one response.)

( ) a. 0.0005
( ) b. 0.25
( ) c. 0.0625
( ) d. 0.5625

18b. What is the probability that this couple has 4 unaffected children followed by 2 affected children? (Check only one response.)

( ) a. 0.2966
( ) b. 0.8306
( ) c. 0.1318
( ) d. 0.0198

18c. If this couple already has four unaffected children, what is the probability that their fifth child is unaffected? (Check only one response.)

( ) a. 0.0010
( ) b. 0.25
( ) c. 0.75
( ) d. 0.2373

Problems 19a through 19 concern the following: Research has concluded that individuals experience a common cold approximately two times per year. Assume that the time between colds is normally distributed with a mean of 160 days and a standard deviation of 40 days.

19a. What is the probability of going 200 or more days between colds? (Check only one response.)

( ) a. 0.682
( ) b. 0.138
( ) c. 0.841
( ) d. 0.159
( ) e. 0.318

19b. Approximately 95% of this population will experience colds between: (Check only one response.)

( ) a. 0 and 320 days
( ) b. 120 and 200 days
( ) c. 40 and 280 days
( ) d. 80 and 240 days

Problems 20a through 20c concern the following: Suppose that the distribution of serum Vitamin E is approximately normal with a mean of 860 ug/dl and standard deviation of 340 ug/dl.

20a. If the "normal range" is considered to be within (+/-) 1.5 standard deviations of the mean, this range is between: (Check only one response.)

( ) a. 520 - 1200 ug/dl
( ) b. 180 - 1540 ug/dl
( ) c. 350 - 1370 ug/dl
( ) d. 690 - 1030 ug/dl
( ) e. 340 - 860 ug/dl

20b. What is the probability of an individual with a Vitamin E level falling outside of the normal range? (Check only one response.)

( ) a. 0.13
( ) b. 0.05
( ) c. 0.023
( ) d. 0.15
( ) e. 0.32

20c. What is the probability of an individual with a Vitamin E level more than two standard deviations above the mean value? (Check only one response.)

( ) a. 0.13
( ) b. 0.05
( ) c. 0.023
( ) d. 0.15
( ) e. 0.32

ANSWER KEY

1. FFTFT
2. a. 93 b. higher c. $Q_1 - Q_3 = 119 - 93 = 26$
   d. 159, 217 e. Q_2 = median f. 93
3. a. 8.6 b. 2, <1 c. 75% d. Yes, because of wide range of values (which necessitated the second graph)
4. FFTF A. 9, 3 B. 6 C. 20-1≈19
5. a. 4 b. 88% c. 2-3 d. 6-7 e. 6-2=4 f. 39-1=38 g. 43%
6. FFTF a. 85 b. 50 c. 50 to <55 d. 25
7. c b d a
8. d a c a
9. c d c a b d b c
10. c a b b
11. e c b d c
12. d c d (a,b,c) P(blood group) ≠ P(blood group | Caucasian) ≠ P(blood group | African American)
    b d
13. b d b c
14. d (a,b,e)
15. d b c
16. b e a d c
17. a c b b
18. d d c
19. d d
20. c a c

Binomial Formula

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \quad \text{where } \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Poisson Formula

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$$