PRACTICE PROBLEMS
FOR
BIOSTATISTICS

BIOSTATISTICS
DESCRIBING DATA, THE NORMAL DISTRIBUTION

1. The duration of time from first exposure to HIV infection to AIDS diagnosis is called the *incubation period.* The incubation periods of a random sample of 7 HIV infected individuals is given below (in years):

12.0	10.5
9.5	6.3
13.5	12.5
7.2	

   a. Calculate the sample mean.
   b. Calculate the sample median.
   c. Calculate the sample standard deviation.
   d. If the number 6.3 above were changed to 1.5, what would happen to the sample mean, median, and standard deviation? State whether each would increase, decrease, or remain the same.
   e. Suppose instead of 7 individuals, we had 14 individuals. (we added 7 more randomly selected observations to the original 7)

12.0	10.5	5.2
9.5	6.3	13.1
13.5	12.5	10.7
7.2	14.9	6.5
8.1	7.9	

   Make an educated guess of whether the sample mean and sample standard deviation for the 14 observations would increase, decrease, or remain roughly the same compared to your answer in part (c) based on only 7 observations. Now actually calculate the sample mean standard deviation to see if you were right. How does your calculation compare to your educated guess? Why do you think this is?

2. In a random survey of 3,015 boys age 11, the average height was 146 cm, and the standard deviation (SD) was 8 cm. A histogram suggested the heights were approximately normally distributed. Fill in the blanks.
   a. One boy was 170 cm tall. He was above average by __________ SDs.
   b. Another boy was 148 cm tall. He was above average by __________ SDs.
   c. A third boy was 1.5 SDs below the average height. He was __________ cm tall.
   d. If a boy was within 2.25 SDs of average height, the shortest he could have been is __________ cm and the tallest is __________ cm.
   e. Here are the heights of four boys: 150 cm, 130 cm, 165 cm, 140 cm. Which description from the list below best fits each of the boys (a description can be used more than once)? Justify you answer
      – Unusually short.
      – About average.
      – Unusually tall.

3. Assume blood-glucose levels in a population of adult women are normally distributed with mean 90 mg/dL and standard deviation 38 mg/dL.

   a. Suppose the "abnormal range" were defined to be glucose levels outside of 1 standard deviation of the mean (i.e., either at least 1 standard deviation above the mean, or at least 1 standard deviation below mean). Individuals with abnormal levels will be retested. What percentage of individuals would be called "abnormal" and need to be retested? What is the normal range of glucose levels in units of mg/dL?
   b. Suppose the abnormal range were defined to be glucose levels outside of 2 standard deviations of the mean. What percentage of individuals would now be called "abnormal"? What is the normal range of glucose levels (mg/dL)?

4. A sample of 5 body weights (in pounds) is as follows: 116, 168, 124, 132, 110. The *sample median* is:
   a. 124.
   b. 116.
   c. 132.
   d. 130.
   e. None of the above.

5. Suppose a random sample of 100 12-year-old boys were chosen and the heights of these 100 boys recorded. The sample mean height is 64 inches, and the sample standard deviation is 5 inches. You may assume heights of 12-year-old boys are normally distributed. Which interval below includes approximately 95% of the heights of 12-year-old boys?

a. 63 to 65 inches.
b. 39 to 89 inches.
c. 54 to 74 inches.
d. 59 to 69 inches.
e. Cannot be determined from the information given.
f. Can be determined from the information given, but none of the above choices is correct.

6. Cholesterol levels are measured on a random sample of 1,000 persons, and the sample standard deviation is calculated. Suppose a second survey were repeated in the same population, but the sample size tripled to 3,000. Then which of the following is true?

a. The new sample standard deviation would tend to be smaller than the first and approximately about one-third the size.
b. The new sample standard deviation would tend to be larger than the first and approximately about three times the size.
c. The new sample standard deviation would tend to be larger than the first, but we cannot approximate by how much.
d. None of the above is true because there is no reason to believe one standard deviation would tend to be larger than the other.

BIOSTATISTICS
SAMPLING DISTRIBUTIONS, CONFIDENCE INTERVALS

Investigator A takes a random sample of 100 men age 18-24 in a community. Investigator B takes a random sample of 1,000 such men.

a. Which investigator will tend to get a bigger standard deviation (SD) for the heights of the men in his sample? Or, can it not be determined?
b. Which investigator will tend to get a bigger standard error of the mean height? Or, can it not be determined?
c. Which investigator is likely to get the tallest man? Or are the chances about the same for both investigators?
d. Which investigator is likely to get the shortest man? Or are the chances about the same for both investigators?

2. A study is conducted concerning the blood pressure of 60 year old women with glaucoma. In the study 200 60-year old women with glaucoma are randomly selected and the sample mean systolic blood pressure is 140 mm Hg and the sample standard deviation is 25 mm Hg.

a. Calculate a 95% confidence interval for the true mean systolic blood pressure among the population of 60 year old women with glaucoma.

b. Suppose the study above was based on 100 women instead of 200 but the sample mean (140) and standard deviation (25) are the same. Recalculate the 95% confidence interval. Does the interval get wider or narrower? Why?

3. The post-surgery times to relapse of a sample of 500 patients with a particular disease is a skewed distribution. The sampling distribution of the sample mean relapse time:

(a) will be approximately normally distributed.
(b) will be skewed
(c) No general statement can be made

4. A survey is performed to estimate the proportion of 18-year old females who have had a recent sexually transmitted disease (STD) defined as an STD in the past year. In a random sample of 300 women, 200 have agreed to participate. Based on these 200 women, a 95% confidence interval for the proportion who had a recently sexually transmitted disease was .10 to .21.

Which of the following is true about the proportion who had a recent STD among the 100 who did not agree to participate in the survey:

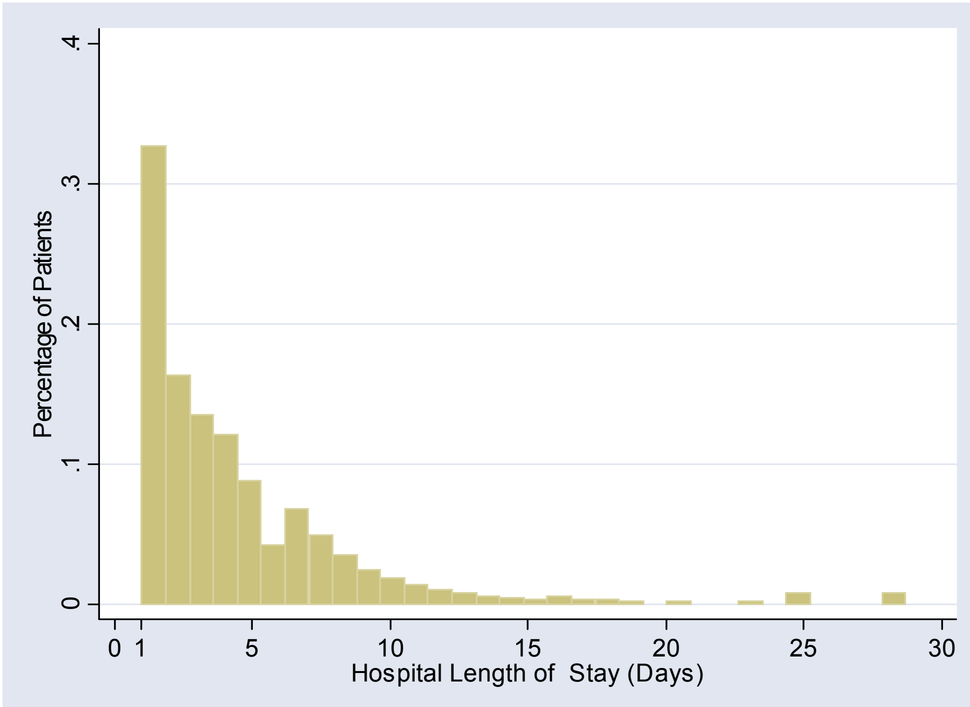
(a) The proportion will definitely be in the interval .10 to .21.
(b) The proportion will definitely not be in the interval .10 to .21.
(c) Proportion will be in the interval .10 to .21 with 95% confidence.
(d) No general statement can be made without additional information.

5. A random sample of 300 diastolic blood pressure measurements are taken. Suppose a 99% confidence interval for the population mean diastolic blood pressure is 68 to 73 mm Hg. If a 95% confidence interval is also calculated, then

(a) The 95% confidence interval will be wider than the 99%.
(b) The 95% confidence interval will be narrower than the 99%.
(c) 95% and 99% confidence interval will be the same.
(d) One cannot make a general statement about whether the 95% confidence interval would be narrower, wider or the same as the 99%.

You will need the following information to answer questions 6 through 8:

There were over 3.5 million hospital discharges in the year 2000 in the U.S. state of California. Patient length of stay summary statistics available on all reported year 2000 hospital discharges in California include a median length of stay of 3.0 days, a mean length of stay of 4.6 days, and a standard deviation of 4.5 days. Below is a histogram that shows the distribution of the length of stay, measured in days, for all hospital discharges in the year 2000 in California. (the California all discharge data set). You may consider this the population distribution of hospital discharges for the year 2000 in California.



6. If a random sample of 1,000 discharges were taken from the California all-discharge database, and a histogram were made of patient length of stay for the sample, which of the following is most likely true:

   a) The histogram will look approximately like a normal distribution because the sample size is large , and the Central Limit Theorem applies.
   b) The histogram will look approximately like a normal distribution because the number of samples is large , and the Central Limit Theorem applies.
   c) The histogram will appear to be right skewed.
   d) The histogram will appear to be left skewed.
   e) The histogram will look like a uniform distribution

7. Suppose we compared 2 random samples taken from the California all-discharge database described above. **Sample A** is a random sample with 100 discharges. **Sample B** is a random sample with 2,000 discharges. What can be said about the relationship between the sample standard error in Sample A ($\boldsymbol{SE_A}$) relative to the sample standard error of length-of-stay value in Sample B (SE_B)?

   a) $\boldsymbol{SE_A < SE_B}$
   b) $\boldsymbol{SE_A > SE_B}$
   c) $\boldsymbol{SE_A}$ is exactly equal to $\boldsymbol{SE_B}$
   d) Not enough information given to determine relationship between the two standard errors.

8. Suppose we took 5,000 random samples from the California all-discharge data set, each sample containing 100 discharges. For each of the 5,000 samples, the sample mean was computed. A histogram was then created with the 5,000 sample mean values. Which of the following statements most likely describes this histogram?

a) The histogram will look approximately like a normal distribution because the size of each sample is large , and the Central Limit Theorem applies.
b) The histogram will look approximately like a normal distribution because the number of samples is large , and the Central Limit Theorem applies.
c) The histogram will appear to be right skewed.
d) The histogram will appear to be left skewed.
e) The histogram will look like a uniform distribution.

9. In a health care utilization journal, results are reported from a study performed on a random sample of 100 deliveries at a large teaching hospital. The sample mean birth weight is reported as 120 ounces, and the sample standard deviation is 25 ounces. The researchers neglected to report a 95% confidence interval for the population birth weight (i.e.: mean birthweight for all deliveries in the hospital). You decide to do so, and find the 95% confidence interval for the population mean birth weight to be:

a) 119.5 ounces to 120.5 ounces
b) 115 ounces to 125 ounces
c) 70 ounces to 170 ounces
d) 117.5 ounces to 122.5 ounces

10. A survey was conducted on a random sample of 1,000 Baltimore residents. Residents were asked whether they have health insurance. 650 individuals surveyed said they do have health insurance, and 350 said they do not have health insurance. A 95% CI for the proportion of Baltimore residents with health insurance is:

a) 60% to 75%
b) 32% to 38%
c) 62% to 68%
d) 36% to 46%

BIOSTATISTICS
HYPOTHESIS TESTING

1. A study was undertaken to evaluate the effect of percutaneous transluminal coronary angioplasty (PTCA) in patients with one-vessel coronary artery disease. A random sample of one hundred and seven patients with coronary artery disease were given PTCA. Patients were given exercise tests at baseline and after 6 months of follow-up. Exercise tests were performed up to maximal effort until symptoms (such as angina) were present. The "change" in the duration of exercise was calculated. "Change" is defined as the 6 month test minus the baseline test. The mean change was 2.1 minutes and the standard deviation of the changes was 3.1.

   (a) What statistical test can be performed to see if there has been a statistically significant change in duration of exercise for this group of patients given PTCA?
   (b) Compute a 95% confidence interval for the mean change in exercise duration.
   (c) Can we conclude from this study that PTCA is effective in increasing exercise duration? Are there any limitations or weaknesses in this study for answering that question?

2. A researcher wishes to determine if vitamin E supplements could increase cognitive ability among elderly women. In 1999 the researcher recruits a sample of elderly women age 75-80. At the time of the enrollment into the study, the women were randomized to either take Vitamin E, or a placebo for six months. At the end of the six month period, the women were given a cognition test. Higher scores on this test indicate better cognition. The mean and standard deviation of the test scores of 81 women who took vitamin E supplements was 27 and 6.9 respectively. The mean and standard deviation of the test scores of the 90 women who took placebo supplements was 24 and 6.2, respectively.

   1. Compute a 95% confidence interval for the mean difference in cognition test scores between Vitamin E and placebo groups.
   2. What statistical test would you perform to compare the mean scores?
   3. Are there limitations to this study for drawing conclusions about whether vitamin E can enhance cognitive ability in elderly women?
   4. What would you conclude from these study results?

3. Measurements on babies of mothers who used marijuana during pregnancy were compared to measurements on babies of mothers who did not. The sample mean head circumference was larger in the group who were not exposed to marijuana and the 95% confidence interval for the difference in mean circumference between the 2 groups was .61 to 1.19 cm. What can be said about the (2-sided) p value for testing the hypothesis of equal means?

   (a) The p value is greater than .05
   (b) The p value is equal to .05
   (c) The p value is less than .05
   (d) The p value is 0
   (e) Cannot be determined from the information given.

4. A study of 100 patients is performed to determine if cholesterol levels are lowered after 3 months of taking a new drug. Cholesterol levels are measured on each individual at the beginning of the study and 3 months later. The cholesterol change is calculated which is the value at 3 months minus the value at the beginning of the study. On average the cholesterol levels among these 100 patients decreased by 15.0 and the standard deviation of the changes in cholesterol was 40. What can be said about the 2 sided p-value for testing the null hypothesis of no change in cholesterol levels?

   (a) The p value is less than. 05
   (b) The p value is greater than .05
   (c) The p value is equal to .05
   (d) Cannot be determined from the information given

5. The standard error of a statistic is

   (a) the mean of the sampling distribution
   (b) the standard deviation of the sampling distribution
   (c) the statistic divided by the square root of the sample size (that is, statistic/$\sqrt{N}$)

6. Two hundred hypertensive patients are randomized to either a diet program or an exercise program. (102 patients to diet program, 98 patients to exercise program) Blood pressure is measured on each patient both before the programs start and 3 months after the start of each program. The correct statistical procedure to determine whether or not the mean blood pressure change (after-before) was statistically significantly different for the two programs is:

   a. Paired t-test
   b. 2 sample unpaired t-test
   c. Fisher's exact test
   d. None of the above

7. Measurements on a (random) sample of babies born to mothers who took "Prescription Drug A" during pregnancy were compared to measurements on a (random) sample of babies born to mothers who did not take "Prescription Drug A". A statistically significant difference in the mean head circumference was found between children born to the two groups of mothers ($p = .03$). Based only on this information you can conclude:

   a. There is a 3% chance the null hypothesis is true.
   b. Taking "Prescription Drug A" during pregnancy causes a reduction in child's head circumference.
   c. Taking "Prescription Drug A" is associated with an increased child head circumference.
   d. The sample mean difference in children's head circumference between children born to the two groups of mothers is clinically important/significant.
   e. None of the above.

8. You are reading an article detailing a study designed to compare the average body mass index among hypertensive individuals (persons with high blood pressure) to normotensive individuals. The study was conducted using participants in a health plan offered by a large insurance carrier. 200 randomly selected hypertensive participants were compared to 200 randomly selected normotensive participants. The study reports a 95% confidence interval (CI) for the mean difference in BMI between hypertensives and normotensives. The 95% CI is (0.15 kg/m^2, 2.43 kg/m^2). However, no p-value is reported for testing the null hypothesis of no mean difference in BMI between the two groups (a mean difference of 0 kg/mg^2) versus the alternative of a non-zero mean difference. Luckily, you have taken this course, so you are able to determine:

a. the p-value is < .05
b. the p-value is > .05
c. the p-value is exactly .05.
d. the p-value is exactly .009
e. None of the above.

Question 9 refers to the following information:

A study was performed on 200 elementary school students to investigate whether regular Vitamin A supplementation was effective in preventing colds during the month of March. 100 were randomized to receive daily Vitamin A supplements during the month of March, and 100 students were randomized to a placebo group (and did not receive Vitamin A) during the same month. The number of students getting at least one cold in March was computed in the two groups, and the results are given in the following 2 X 2 table.

	Cold	No Cold	
Vitamin A	15	85	100
Placebo	25	75	100
	40	160	200

9. If you were interested in testing for a statistical relationship between taking Vitamin A and cold status in the population of elementary school students (ie: testing the null hypothesis of no difference in the proportion of colds in Vitamin A and placebo groups, versus the alternative of a difference in proportions), the correct statistical test is:

a. Two-sample t-test
b. Ad-hoc test for testing the equality of things
c. Chi-squared test
d. Paired t-test

BIOSTATISTICS
PROPORTIONS

1. An article in the Journal of the American Medical Association[1] documents the results of a randomized clinical trial designed to evaluate whether the influenza vaccine is effective in reducing the occurrence of acute otitis media (AOM) in young children. Acute otitis media is an infection that causes inflammation of the middle ear canal. In the study, children were randomized to receive either the influenza vaccine or a placebo. (randomization was done in a 2 to 1 ratio, meaning that two times as many children were randomized to the vaccine treatment as were randomized to the placebo group). The children were followed for one year after randomization, and monitored for AOM during this period. 262 children were randomized to the vaccine group, and 150 of these children experienced at least one incident of AOM during the follow-up period. 134 children were randomized to the placebo group, and 83 of these children experience at least one incident of AOM during the follow-up period. Note that the standard error of the differences in the proportions in the two groups is 0.0519.

   (a) Estimate a 95% confidence interval for the proportion of children experiencing at least one incident of AOM during the follow-up period in each of the randomization groups. How do these 95% CI's compare? (similar range of values? Overlap?)
   (b) Estimate the 95% confidence interval for the difference in the two proportions.
   (c) State the null and alternative hypotheses associated with testing for an association between the influenza vaccine and AOM. Is the p-value for testing these hypotheses significant at the 0.05 level?
   (d) Is this a randomized study? What does this suggest when translating the observed difference in proportions?

[1] based on data from:
Hoberman, A. et al. Effectiveness of Inactivated Influenza Vaccine in Preventing Acute Otitis Media in Young Children: A Randomized Controlled Trial (2003). *Journal of the American Medical Association*, 290, 1608-1616.

2. A study was done to investigate whether there is a relationship between survival of patients with coronary heart disease and pet ownership. A representative sample of 92 patients with CHD was taken. Each of these patients was classified as having a pet or not and by whether they survived one year following their first heart attack. Of 53 pet owners, 50 survived. Of 39 non-pet owners, 28 survived. Note that the standard error of the difference in the two proportions is 0.34.

   Suppose you were interested in doing a statistical analysis of these study results. Answering the follow questions to help you with this goal!

   (a) Estimate a 95% confidence interval for the proportion of patients surviving at least one year beyond the first heart attack in each of the groups (pet/no pet).
   (b) Is this a randomized study? What does this suggest when translating the statistical result from part (a) into a substantive/scientific conclusion?

3. To evaluate the effectiveness of 2 different smoking cessation programs, smokers are randomized to receive either program A or program B. Of 6 smokers on program A, 1 stopped smoking and 5 did not. Of 6 smokers on program B, 4 stopped smoking and 2 did not. Which statistical procedure would you use to test the null hypotheses that the programs are equally effective and to obtain a p-value?

   (a) Paired t- test
   (b) Chi Square Test
   (c) One way Analysis of Variance
   (d) 2 sample t- test
   (e) Fisher's exact test
   (f) Mann-Whitney (Wilcoxon rank sum) nonparametric test

4. Suppose you are asked to help design a study to compare the mean birth weights of 2 groups of infants. The study researcher plans to have the same sample size in both groups. The researcher tells you that she would like to have 80% power to detect a difference in mean birth weights of 5 ounces using a significance level of .05. Has the researcher provided you with enough information for you or your computer to perform the sample size calculation?

a. Yes, there is enough information to calculate the sample size
b. No, but there would be enough information if I was also told the mean birth weight in both groups
c. None of the above

BIOSTATISTICS
LINEAR REGRESSION

The objective of a study is to understand the factors that are associated with systolic blood pressure in infants. Systolic blood pressure, weight (ounces) and age (days) are measured in 100 infants. A multiple linear regression is performed to predict blood pressure (mm Hg) from age and weight. The following results are presented in a journal article. (Questions 1-3 refer to these results)

Multiple Linear Regression Analysis of the Predictors of Systolic Blood Pressure in Infants

	$\hat{b}'s$ **(coefficients)**	**SE(**$\hat{b}'s$**)**
Intercept	**50.**	**4.0**
Birth Weight	**0.10**	**0.3**
Age (days)	**4.0**	**0.60**

1. How much higher would you expect the blood pressure to be of an infant who weighed 120 ounces compared to an infant who weighed 90 ounces if both infants were of exactly the same age?

a. 0.1 mm Hg
b 1.0 mm Hg
c. 2.0 mm Hg
d. 3.0 mm Hg
e. 4.0 mm Hg

2. Which of the following is a 95% confidence for the difference in SBP between two infants of the same weight who differ by 2 days in age (older compared to younger)?

a. 2.8 mmHg to 5.2 mmHg
b. 5.6 mmHg to 10.4 mmHg
c. 6.8 mmHg to 9.6 mmHg
d. –0.5 mmHg to 0.7 mmH

3. Suppose the R^2 from the above regression model is .57, which means that roughly 57% of the variability in the infant's blood pressure measurements is explained by infant's age and weight. What would happen to this R^2 value, if weight had been recorded as kilograms instead of ounces?

a. R^2 would go increase.
b. R^2 would decrease
c. R^2 would equal .57.
d. Not enough information to determine.

4. A recent study of the relationship between Scholastic Aptitude Test (SAT) scores and U.S. state level characteristics found a statistically significant ($p < .05$) relationship between average SAT scores and the percent of high school seniors who actually took the SAT within a state. Linear regression was used to estimate this relationship , and the resulting regression equation was:

$$y = 1024 – 2.3x_1$$

where y represents average SAT score, and x_1 represents the percentage of high school seniors taking the SAT.
The coefficient of determination, R^2 is 0.76. What can be said about the correlation coefficient, r?

a. The correlation coefficient is – 0.76
b. The correlation coefficient is 0.76
c. The correlation coefficient is .87
d. The correlation coefficient is – .87

A data set contains information about the hourly wage (in U.S. dollars) and the gender of each of the 534 workers surveyed in 1985, as well as information about each worker's age, union membership, and education level (measured in years of education). A linear regression analysis was performed to model hourly wage as a function of worker sex, number of years of education, and union membership. Below find the results from this regression:

	Regression Coefficient	**Standard Error of Regression Coefficient**
sex (1 = Female, 0 = Male)	-1.9	0.4
union member (1 = yes, 0 = no)	1.9	0.5
years of education	0.76	0.08
intercept	-0.3	1

5. Using the above regression results, estimate the mean hourly wage in 1985 for male workers with a high school education (12 years of education), who were not union members.

a. $10.72 per hour
b. $6.93 per hour
c. $9.12 per hour
d. $8.82 per hour

6.. Give a 95% confidence interval for the mean difference in hourly wages for male workers with a high school education who were union members when compared to male workers with a high school education who were not union members.

a. $0.90 per hour to $2.90 per hour
b. - $2.70 per hour to -$1.20 per hour
c. Cannot be estimated with the reported regression results.

7. The relationship between forced expiratory volume (FEV), which is measured in liters, and age, which is measured in years, is evaluated in a random sample of 200 men between the ages of 20 and 60. A simple linear regression analysis is performed to predict FEV from age. The following results are published in a paper.

Results of Simple Linear Regression Analysis

	Regression Coefficient	**Standard Error**
Intercept	**-4.0**	**.30**
Age	**0.02**	**.005**

a) Interpret the coefficient of age in words.
b) Give a 95% confidence interval for the coefficient of age. Write a sentence explaining the interpretation of this confidence interval.
c) Given the above results, can you ascertain whether the linear relationship between FEV and age is strong? Why or why not?
d) Suppose above results are used to compare 60 year old men to 50 year old men – what would be the estimated average difference in FEV between the two groups of men?
e) Would it make sense to use the above results to estimated the average FEV levels for men 80 years of age? Why or why not?

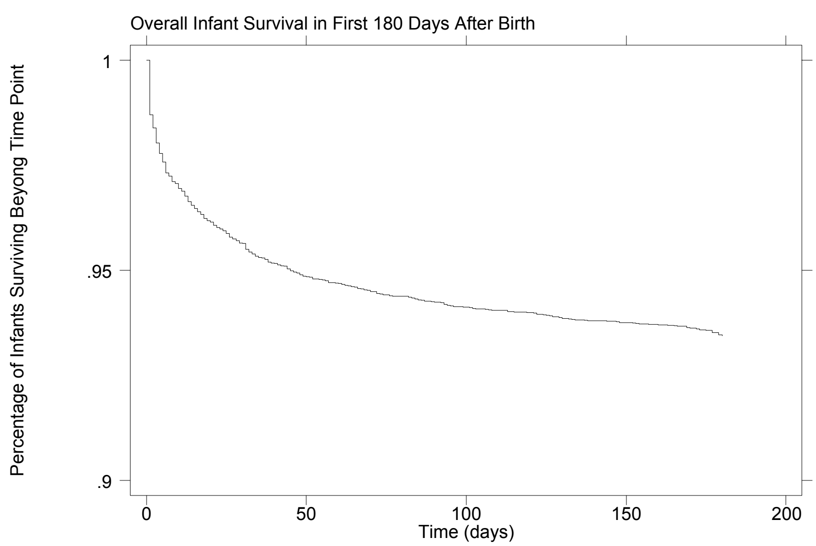
BIOSTATISTICS
SURVIVAL ANALYSIS

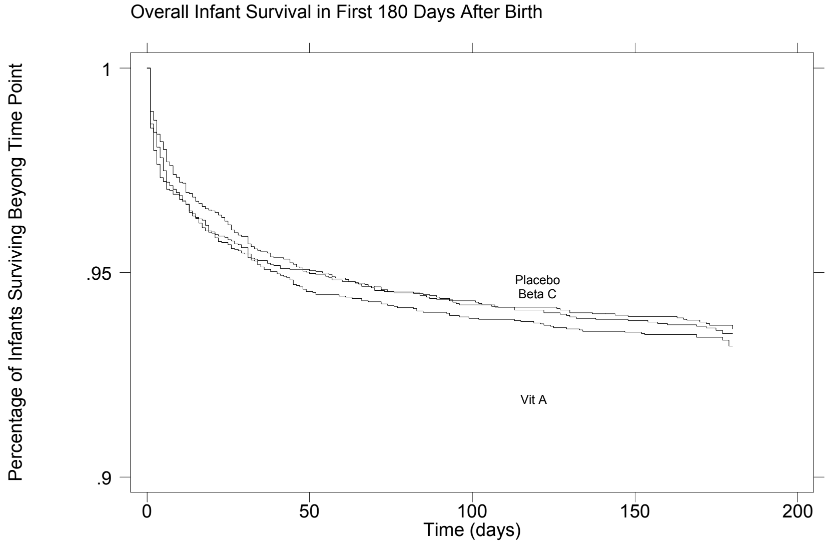
1. Ten health care workers who were accidentally stuck with needles contaminated with HIV were followed for onset of AIDS. Six of the workers developed AIDS at 36, 65, 90, 110, 140 and 121 months. Four of the workers had still not developed AIDS at the time they were last contacted which was 40, 75, 130 and 160 months after the needle stick occurred. What statistical method would you use to find the median incubation period (time from HIV exposure to AIDS) ?

The following information is referenced in questions 2 –3.

NNIPS-II was a randomized study of 15,987 infants born alive to 43,559 women who received Vitamin A, Beta-carotene, or a placebo prior to and during pregnancy. (Investigators: Drs. Keith West, Joanne Katz, Parul Christian et al..) Infants were followed for up to 180 days after birth, until death (the event (outcome) being studied), dropout or study's end (180 days). We are using a representative subset of 10,295 live births from the original data set to investigate factors associated with increased risk of infant mortality in this population.

Kaplan Meier curves for overall infant survival in the first 180 days, and survival stratified by each of the three drug groups are as follows:





2. Based on the first Kaplan Meier curve ("Overall Infant Survival in First 180 days"), an estimate of the median survival time after birth for this sample of infants is:

   a. 47 days
   b. 100 days
   c. 150 days
   d. The median survival time cannot be estimated using the given Kaplan-Meier curve.

3. Based on the first Kaplan Meier curve, an estimate of the 100 day survival rate is

   a. 50 days
   b. 0.94
   c. 1.00
   d. 0.90
   e. Cannot be estimated using the information given

4. Suppose a random subsample of 1,000 infants was followed for up to 10 years after their birth. 122 of these infants died within 10 years of their birth date, and the remaining infants either were lost to followup, or survived the full 10 year study period. Suppose the mean of the recorded death and censoring times was computed for this group of infants to estimate mean survival times in the first ten years following birth, and this sample mean is 6.7 years. Is this sample mean a good estimate of the true mean survival time in the 10 years following birth?

   a. Yes, it is a good estimate.
   b. No, the sample mean will tend to overestimate the true mean survival time.
   c. No, the sample mean will tend to underestimate the true mean survival time.