# (Key) Biostat Exam 2, S01, Apr 12, 2001

## Part A: (Closed Book)

Coverage: *StatPrimer* 4 - 6 and Labs 4 - 6.

General Instructions: Please write your name in usual location. For Part A, use the scantron to record your answers, selecting the single best answer in each instance. When you complete part A, turn in your scantron and pick up your lab notebook, and then begin part B. Time limit: 70 minutes

1. *Over many trials*, we observe that the relative frequency of success of a particular intervention in a given population is about 15%. We therefore conclude that the probability of success is about .15. What type of probability is this?
   a. Probability based on logic
   b. Probability based on experience*
   c. Probability based on subjective opinion
   d. none of the above

2. What is the probability of having a May 2nd birthday?
   a. 1 in 365
   b. 1 in 366
   c. 1 in 1461
   d. 4 in 1461*

3. A *single trial* that can have only two possible outcomes: either a “success” or “failure” is referred to as a:
   a. Bernoulli trial*
   b. binomial random variable
   c. continuous random variable
   d. normal random variable

4. The statement that declares that an observed difference is due to chance is called the:
   a. null hypothesis*
   b. alternative hypothesis
   c. alpha level
   d. test statistic

5. The “choose function” tells you the number of different ways to choose *i* objects of *n* without repetition.
   a. True*
   b. False

6. Probability function curves for continuous random variables are constructed so
   a. the total area under the curve = 1
   b. the probability of any exact value = 0
   c. the area under the curve between any two points represents the probability of that range.
   d. “a” and “c”
   e. “a” “b” and “c”*

7. Normal distributions
   a. are also called Gaussian distributions
   b. have a mean of 0 and standard deviation of 1
   c. have various means and standard deviations
   d. “a” and “c”*
   e. “a” “b” and “c”

8. A distance of ±1 standard deviations from the mean of a normal distribution encloses about
   a. 68% of the area under the curve.*
   b. 95% of the area under the curve.
   c. virtually all of the area under the curve.
   d. none of the above

9. To standardize a normal random variable we subtract the distribution’s mean and divide by its:
   a. mean
   b. variance
   c. standard deviation*
   d. probability

10. A statistic that is used to assess the evidence against the null hypothesis is called a:
    a. confidence interval
    b. test statistic*
    c. type I error
    d. type II error

11. The theory that postulates sampling distributions of means tend to be normally distributed is:
    a. the law of large numbers
    b. the law of unbiasedness
    c. the central limit theorem*
    d. the law of the farm

12. The standard error of a mean quantifies:
   a. the spread of a measurement
   b. the center of a measurement
   c. the shape of a measurement
   d. how greatly a particular sample mean is likely to differ from the population mean*

13. The probability of avoiding a type II error is:
   a. alpha
   b. beta
   c. confidence
   d. power*

14. The two forms of statistical *inference* are:
   a. estimation and hypothesis testing*
   b. *p* values and test statistics
   c. parameters and statistics
   d. Ren and Stimpy

15. *Population-based* statistical characteristics are called:
   a. variables
   b. statistics
   c. parameters*
   d. tests

16. Which of the following is the *point estimator* of the population mean?
   a. *SEM*
   b. σ
   c. $\bar{x}$ *
   d. :

17. To construct a (1 - α)100% confidence interval for μ when the population standard deviation is known (or assumed), we use the formula:
   a. $\bar{x}$ ± (1.96)(*SEM*)
   b. $\bar{x}$ ± ($z_{1 ! \alpha/2}$)(*SEM*)*
   c. $\bar{x}$ ± ($t_{n-1,.975}$)(*sem*)
   d. $\bar{x}$ ± ($t_{n-1,1 ! \alpha/2}$)(*sem*)

18. To construct a (1 - α)100% confidence interval for μ when the population standard deviation is calculated from the sample we use the formula:
   a. $\bar{x}$ ± (1.96)(*SEM*)
   b. $\bar{x}$ ± ($z_{1 ! \alpha/2}$)(*SEM*)
   c. $\bar{x}$ ± ($t_{n-1,.975}$)(*sem*)
   d. $\bar{x}$ ± ($t_{n-1,1 ! \alpha/2}$)(*sem*)*

19. We reject $H_0$ when:
   a. *p* # α*
   b. *p* $ α
   c. *p* # .05
   d. *p* $ .05

**Use this table to answer the next 4 questions:**

| | **$H_0$ True** | **$H_0$ False** |
|---|---|---|
| **Retain $H_0$** | (1) | (2) |
| **Reject $H_0$** | (3) | (4) |

20. The table cell indicated by (1) can be labeled:
   a. Correct retention of $H_0$*
   b. Correct rejection of $H_0$
   c. Type I error
   d. Type II error

21. The table cell indicated by (2) can be labeled:
   a. Correct retention of $H_0$
   b. Correct rejection of $H_0$
   c. Type I error
   d. Type II error*

22. The table cell indicated by (3) can be labeled:
   a. Correct retention of $H_0$
   b. Correct rejection of $H_0$
   c. Type I error*
   d. Type II error

23. The table cell indicated by (4) should be labeled:
   a. Correct retention of $H_0$
   b. Correct rejection of $H_0$*
   c. Type I error
   d. Type II error

24. The term statistical power refers to:
   a. the probability of making a type I error
   b. the probability of making a type II error
   c. the probability of avoiding a type I error
   d. the probability of avoiding a type II error*

25. A jury that declares an innocent person "guilty" is making an error that is analogous to a ____ error.
   a. type I error*
   b. type II error
   c. random
   d. non-random

## Part B (Lab Notebook)

1. Calculate the following probabilities for a *binomial random variable* based on 2 trials with the probability of success for each trial being .90. Show all work. [8 pts]

   a. $Pr(X = 0) = ({}_2C_0)(.9^0)(.1^2) = {}_2C_0(.90^0)(.1^2) = (1)(1)(.01) = .01$
   b. $Pr(X = 1) = ({}_2C_1)(.9^1)(.1^1) = {}_2C_1(.90^1)(.1^1) = (2)(.9)(.1) = .18$
   c. $Pr(X = 2) = ({}_2C_2)(.9^2)(.1^0) = {}_2C_2(.90^2)(.1^0) = (1)(.81)(1) = .81$
   d. $Pr(X <= 1) = .01 + .18 = .19$

2. Find the following *z* and *t* percentiles: [8 pts]

   a. $z_{.85}$ = ______+1.04
   b. $z_{.15}$ = ______-1.04
   c. $z_{.995}$ = ______+2.57 or +2.58
   d. $z_{.02}$ = ______ -2.05
   e. $t_{14,.975}$ = ______+2.14
   f. $t_{14,.025}$ = ______ -2.41
   g. $t_{16,.5}$ = ______0
   h. 1.35_____ < $t_{14,.925}$ < 1.76______

3. A normal random variable with a mean of 60 and standard deviation of 18 shows a value of 76. Find the Pr(X > 76). Show all work. [4 pts]

$z = (76 - 60) / 18 = +0.89$
$Pr(X > 76) = Pr(Z > 0.89) = .1867$
(Curve should be drawn but is not shown on Web.)

4. A school psychologist wishes to estimate the mean IQ of students in her student population. She is willing to assume that the IQ scores are approximately normally distributed with a standard deviation ($\sigma$) of 15. A sample of 8 subjects participate in a testing program which yields a means IQ score of 97.5. Calculate a 95% confidence interval for $\mu$. Show all work. [4 pts]

$97.5 \pm (1.96)(15/\text{sqrt}(8)) = 97.5 \pm 10.4 = (87.1, 107.9)$

5. Redo problem 4 but now calculate a 99% confidence interval for $\mu$. [4 pts]

$97.5 \pm (z_{1-.01/2})(15/\text{sqrt}(8)) = 97.5 \pm (2.58)(5.30) = (83.8, 111.2)$

6. Why is the confidence interval in question 5 constructed in such a way so that it is broader than the confidence interval in question 4? [1 pt]

So that it has a greater likelihood of capturing $\mu$. (Answer should reference confidence of capturing the population mean $\mu$.)

7. Based on a sample of 30 invoices, a health care administrator finds that the average fee collected on in-patient services is \$64.57 with a standard deviation of \$25.06. (The standard deviation is calculated from the sample.) The administrator wants to estimate the average fee for administered in the clinical. Calculate a 95% confidence interval for μ. Show all work. [5 pts]

$64.57 \pm (t_{29,.975})(25.06/\text{sqrt}(30)) = 64.57 \pm (2.05)(4.575) = 64.57 \pm 9.40 = (55.17, 73.97)$

8. A study is performed to determine whether average length of stay in a hospital has decreased since last year. The researcher is certain that the length of hospital stays has not increased, and is willing to assume that the mean length of hospital stays used to be 5 days with a standard deviation (σ) of 2 days. A sample of 20 recent hospital stays shows a mean of 4 days. Conduct a test to see if there has been a significant decrease in hospital stay length. Let α = .05. Show all work and hypothesis testing steps including (a) a statement of the null hypothesis and alternative hypothesis (b) restatement of the alpha level (c) calculation of the test statistic and p value and (d) statement of the conclusion [6 pts]

$H_0$: μ >= 5 days
$H_1$: μ < 5 days
α = .05
$z_{stat} = (4 - 5)/(2/\text{sqrt}(20)) = -2.24$

[Curve not drawn on the Web]

$p = .0125$
Reject $H_0$