

Exam 2 - STAT 303 Session 201
Summer 2009

Name:

UIN:

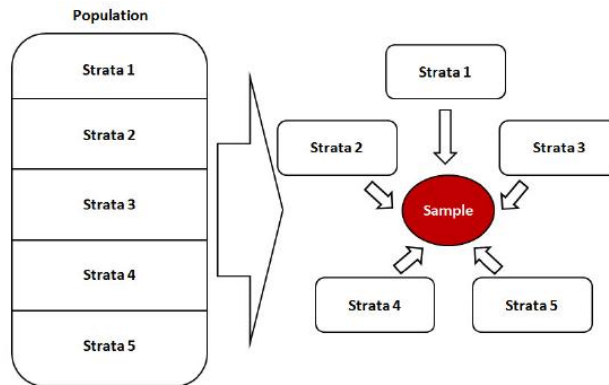
Signature:

1. Do not open this test until told to do so.
2. Turn in your exam with your answers circled when you are done with the exam. You should not take the exam with you.
3. This is a closed book examination. You may use one both-sided sheet of formulas that you have brought with you. You should have no other printed or written material with you on the exam.
4. You have 60 minutes to work on this exam. There are 16 multiple choice questions and 4 work out questions, each worth 5 points.
5. You may use a calculator but not a phone during the exam.
6. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
7. Do not sit directly next to another student.
8. Good Luck!!!

1. Select the FALSE explanation.

- (a) **Voluntary Response:** Voluntary response samples are biased because people with strong opinions, especially negative opinions, are most likely to respond.
- (b) **Undercoverage:** Some groups in the population are left out when the sample is taken.
- (c) **Response Bias:** Results that are influenced by the behavior of the respondent or interviewer
- (d) **Nonresponse:** *An individual chosen for the sample can be contacted or does cooperate.*
- (e) **Convenience Sample:** It is not a probability sample because we do not select the individuals completely at random.

2. Select the answer which can match with a figure below



- (a) Simple Random Sampling
- (b) *Stratified Random Sampling*
- (c) Cluster Sampling
- (d) Multi-stage Sampling

3. Select the right explanation for the sampling technique.

- (a) Simple Random Sampling: Select n individuals from the population chosen in such a way that every set of n individuals has a different chance to be the sample actually selected.
- (b) *Stratified Random Sampling: First divide the population into groups of similar individuals. Then, choose a separate simple random sample in each stratum and combine these simple random samples to form the full sample.*
- (c) Cluster Sampling: Divide the population into several strata. Then take a simple random sample of strata. Next, take simple random samples from EACH of the strata selected on the previous stage.

- (d) Multi-stage Sampling: Divide the population into several strata or clusters. Then take a simple random sample of clusters. Finally, use ALL the individuals in the selected clusters. The technique works best when most of the variation in the population is within the groups, not between them.

4. Select the right explanation about bias and variance for a following figure.



- (a) High Bias, Low Variability
 (b) Low Bias, Low Variability
 (c) High Bias, High Variability
 (d) *Low Bias, High Variability*

5. Select the FALSE probability rules in following.

- (a) $0 \leq P(A) \leq 1$
 (b) $P(S) = 1$ where S is the whole sample space
 (c) For any event A , $P(A^c) = 1 - P(A)$
 (d) If A and B are independent events, then $P(A \text{ and } B) = P(A)P(B)$
 (e) *If A and B are not disjoint, then $P(A \text{ or } B) = P(A) + P(B)$*

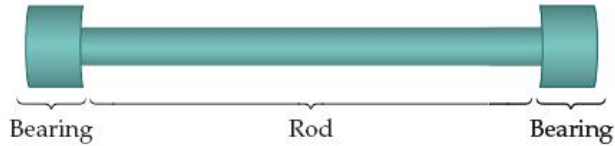
6. Temperature transducers of a certain type are shipped in batches of 50. A sample of 60 batches was selected, and the number transducers in each batch not conforming to design specifications was determined, resulting in the following data.

X	0	1	2	3	4	5	6
p(X)	0.2	0.2	0.25	0.15	0.12	0.05	0.03

What proportion have at least three nonconforming units and what proportion have fewer than two?

- (a) 0.35, 0.65
 (b) 0.20, 0.4

- (c) 0.35, 0.4
(d) 0.20, 0.65
(e) 0.20, 0.20
7. Using the attached Z-table, find the probability $Pr(Z < 1.53)$ and $Pr(Z > -0.52)$
- (a) 0.9370, 0.6985
(b) 0.6030, 0.6985
(c) 0.9370, 0.3015
(d) 0.6030, 0.3015
(e) 0.9357, 0.7019
8. Too much cholesterol in the blood increases the risk of heart disease. Young women are generally less afflicted with high cholesterol than other groups. The cholesterol levels for women aged 20 to 34 follow an approximately normal distribution with mean 200 milligrams per deciliter (mg/dl) and standard deviation 40 mg/dl. Suppose cholesterol levels above 260 mg/dl demand medical attention. What proportion of young women have levels above 260 mg/dl? State the value of the z-score with 2 decimals, the probability with 4 decimals.
- (a) 1.50, 0.1469
(b) 1.05, 0.1469
(c) 1.50, 0.0808
(d) 1.05, 0.0808
(e) 1.50, 0.0668
9. The mean and standard deviation of our exam 1 are 65 and 15. What score should I need to get if my score is above third quartile?
- (a) 72.8
(b) 77.6
(c) 75.2
(d) 70.7
(e) 80.6
10. A mechanical assembly (a following figure) consists of a rod with a bearing on each end. The three parts are manufactured independently, and all vary a bit from part to part. The length of the rod has mean 15 centimeters (cm) and standard deviation 0.15 millimeters (mm). The length of a bearing has mean 1 cm and standard deviation 0.1 mm. What are the mean and standard deviation of the total length of the assembly?



- (a) 16(cm), 0.1700(mm)
 (b) 17(cm), 0.2061(mm)
 (c) 16(cm), 0.2061(mm)
 (d) 17(cm), 0.1700(mm)
 (e) 17(cm), 0.4123(mm)
11. Suppose that 80% of college men and 70% of college women were employed last summer. A sample survey interviews SRSs of 1000 college men and 1000 college women. The two samples are of course independent. What is the approximate distribution of the proportion P_F of women and P_M who worked last summer?
- (a) Women - $N(0.7, 0.00021^2)$, Men - $N(0.8, 0.00016^2)$
 (b) Women - $N(0.8, 0.01265^2)$, Men - $N(0.7, 0.01449^2)$
 (c) Women - $N(0.7, 0.01449^2)$, Men - $N(0.8, 0.01265^2)$
 (d) Women - $N(0.8, 0.00016^2)$, Men - $N(0.7, 0.00021^2)$
12. Using the information in question 11, our survey wants to compare men and women. What is the approximate distribution of the difference in the proportions who worked, $\hat{p}_M - \hat{p}_F$
- (a) $N(-0.1, 0.02714^2)$
 (b) $N(0.1, 0.02714^2)$
 (c) $N(-0.1, 0.01924^2)$
 (d) $N(0.1, 0.01924^2)$
13. The law allows marketers of herbs and other natural substances to make health claims that are not supported by evidence. Brands of ginkgo extract claim to “improve memory and concentration”. A randomized comparative experiment found no evidence for such effects. The subjects were 300 healthy volunteers over 60 years old, with 120 women and 180 men. 60 of the women and 90 of the men were randomly assigned to take a ginkgo extract pill, while the remaining women and men were assigned to take a placebo pill. All subjects took a battery of tests for learning and memory before treatment started and again after six weeks. What element of good experiments is missing from this study?

- (a) Blinding
- (b) Control group
- (c) Randomization
- (d) Using enough subjects
- (e) *All the elements are present*

14. What are the missing probability and the mean for the distribution below?

X	0	1	2	3	4
p(X)	0.4	0.3	0.15	0.1	???

- (a) Without the probability, we cannot determine the mean.
 - (b) $P(X = 4) = 0.5$ and $\mu = 1.4$
 - (c) $P(X = 4) = 0.5$ and $\mu = 2.9$
 - (d) $P(X = 4) = 0.05$ and $\mu = 1.1$
 - (e) $P(X = 4) = 0.05$ and $\mu = 1.5$
15. Using the discrete distribution above, how likely are we to get all 1s if we draw two times, assuming each draw is independent?
- (a) 0.6
 - (b) 0.3
 - (c) 0.09
 - (d) 0.0001
 - (e) We don't know the exact probability using above information.
16. In the fall of 2003, a magazine article reported that about 87% of adults drink milk. A local dairy farmers association is planning a new marketing campaign for the tri-county area they represent. They randomly polled 800 people in the area. In this sample, 654 people said that they drink milk. If 87% is the correct percentage of adults who drink milk, what is the sampling distribution of the sample proportion, p_{800} ?
- (a) We cannot tell if the shape is normal since we do not have a plot of the data.
 - (b) $N(0.87, 0.00014^2)$
 - (c) $N(0.87, 0.01189^2)$
 - (d) $N(0.00109, 0.00014^2)$
 - (e) $N(0.87, 0.00109^2)$

17. The probability to have a A blood type is 0.43 for an American and 0.22 for a Chinese. What is the probability that if we choose an American and Chinese at random, independently of each other; that both will have A blood type? Also at least one of the will have A blood type? (Please write down whole procedure for your calculation for partial or full credit.)

$$P(A \cap B) = P(A)P(B) = .43 \times .22 = 0.0946$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = .43 + .22 - .0946 = .5554$$

18. Spell-checking software catches "nonword errors" that result in a string of letters that is not a word, as when "the" is typed as "teh". When undergraduates are asked to write a 250-word essay (without spell-checking), the number X of nonword errors has the following distribution. (Please write down whole procedure for your calculation for partial or full credit.)

X	0	1	2	3	4
$p(X)$	0.11	0.19	0.31	0.31	0.08

- (a) Find the Mean for random variable X .

$$\mu_X = 0 \times .11 + 1 \times .19 + 2 \times .31 + 3 \times .31 + 4 \times .08 = 2.06$$

- (b) Find the Variance and Standard Deviation for random variable X .

$$\sigma_X^2 = (0-2.06)^2 \times .11 + (1-2.06)^2 \times .19 + (2-2.06)^2 \times .31 + (3-2.06)^2 \times .31 + (4-2.06)^2 \times .08 = 1.2564$$

19. Typographic errors in a text are either nonword errors (as when "the" is typed as "teh") or word errors that result in a real but incorrect word. Spellchecking software will catch nonword errors but not word errors. Suppose human proofreaders catch 65% of nonword errors. You ask a fellow student to read an essay in which you have deliberately made 40 nonwords errors. What is the mean and variance of error caught? (Please write down whole procedure for your calculation for partial or full credit.)

$$\mu_X = 40 \times .65 = 26 \quad \sigma_X^2 = 40 \times .65 \times (1 - .65) = 9.1$$

20. Suppose past studies indicate it takes an average of 6 minutes to memorize a short passage of 20 words with population standard deviation 2. A psychologist claims a new method of memorization will reduce the average time to 4.5 minutes with population standard deviation 1.5. To check the validity of his claim, a random sample of 40 people are to be used. What is the sampling distribution for current method and new method? (Assume that the average is the population mean).

$$\bar{X} \sim N\left(6, \frac{2^2}{40}\right) \quad \bar{Y} \sim N\left(4.5, \frac{1.5^2}{40}\right)$$