

DIRECTIONS:

- This exam contains 50 multiple choice questions, each worth 2 points.
- Circle the correct response for each question. Make sure that your answer is clearly marked. You will not receive partial credit for any work done.
- The standard normal table appears on Page 13 of this exam.
- This is a closed-book, closed notes examination. You may use a calculator if you wish. However, cell phones are not permitted for use in any way.
- Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material or cell-phone usage, will be dealt with severely. Violations may result in an “F” for this exam, “F” for the class, suspension, or expulsion.
- This exam is worth a total of 100 points. **Print your name at the top of this page in the upper right hand corner.** *Good Luck!!*

HONOR PLEDGE FOR THIS EXAM:

After you have finished the exam, please read the following statement and sign your name below it.

I promise that I did not discuss any aspect of this exam with anyone other than the instructor, that I neither gave nor received any unauthorized assistance on this exam, and that the work presented herein is entirely my own.

HELPFUL FORMULAS

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$z = \frac{x - \mu}{\sigma}$$

$$x = \mu + z\sigma$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$y = a + bx$$

$$\hat{p} \sim \mathcal{N} \left(p, \sqrt{\frac{p(1-p)}{n}} \right)$$

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$n = \left(\frac{z^*}{m} \right)^2 p^*(1-p^*)$$

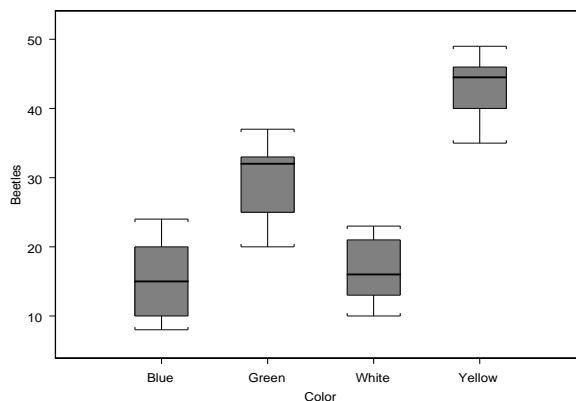
C	80	90	95	99
z^*	1.28	1.64	1.96	2.58

$$\bar{x} \sim \mathcal{N} \left(\mu, \frac{\sigma}{\sqrt{n}} \right)$$

$$\bar{x} \pm z^* \left(\frac{s}{\sqrt{n}} \right)$$

$$n = \left(\frac{z^* \sigma}{m} \right)^2$$

- Loosely speaking, what does the **Central Limit Theorem** say?
 - The area under a normal density curve is one.
 - Measures of central tendency should always be computed with and without outliers.
 - The sampling distribution of \bar{x} is approximately normal.
 - Confidence intervals have zero margin of error for large sample sizes.
- In class, we examined the following graph. The graph depicts the number of beetles captured on boards of different colors.



The median for the distribution of beetles captured on **white** boards is closest to

- 10
 - 15
 - 25
- Refer to the graph in Question 2. The interquartile range for the distribution of beetles captured on **blue** boards is closest to
 - 10
 - 15
 - 25
 - Refer to the graph in Question 2. The shape of the distribution of beetles captured on **green** boards is
 - skewed left
 - symmetric
 - skewed right
 - Which value is closest to the **90th percentile** for the standard normal distribution?
 - 1.3
 - 0.7
 - 0.7
 - 1.3

6. We have computed the probability of an event to be 0.01. Which of the following statements is **correct**?
- (a) The event is unlikely to occur.
 - (b) We would expect the event to occur about 10 percent of the time.
 - (c) The event can not occur.
 - (d) All of the above.
7. In a simple random sample (SRS) of $n = 100$ Columbia residents, 38 of them said that they had attended a USC football game this year. Which value is the closest to the **margin of error** for a 95 percent confidence interval for the proportion of Columbia residents who have attended a game this year?
- (a) 0.01
 - (b) 0.10
 - (c) 0.38
 - (d) 0.62
8. Using data from the $n = 100$ Columbia residents in Question 7, a 99 percent confidence interval for the mean age of Columbia residents was computed to be (29.8, 38.5). What is the **interpretation** attached to this interval?
- (a) We are 99 percent confident that the mean age of Columbia residents is between 29.8 and 38.5.
 - (b) Ninety-nine percent of the residents in our sample had ages between 29.8 and 38.5.
 - (c) We are 99 percent confident that the mean age of the residents in our sample is between 29.8 and 38.5.
 - (d) All of the above are valid interpretations.
9. In Question 8, what is one way to **decrease** the length of the confidence interval?
- (a) Increase the sample size
 - (b) Use a smaller confidence level
 - (c) Both (a) and (b) are correct
 - (d) Neither (a) nor (b) are correct
10. Which principle of experimental design is described by using **impersonal chance** to assign individuals to treatments?
- (a) blocking
 - (b) equipoise
 - (c) randomization
 - (d) control

11. If the **sample mean** of a data set is 15 and the **sample standard deviation** is 9, what percent of the data would you expect to fall between 6 and 24, assuming that the data distribution is fairly symmetric?

- (a) 68 percent
- (b) 81.5 percent
- (c) 95 percent
- (d) 99.7 percent

12. What is the **key difference** between an observational study and an experiment?

- (a) Institutional review boards are not needed with observational studies because ethical issues do not arise with them.
- (b) Experiments always provide unbiased results.
- (c) In observational studies, subjects do not receive any intervention designed to influence their responses.
- (d) Placebos can not be used in observational studies.

13. In a clinical trial, the researchers guess that the proportion of patients responding to a certain drug is around 0.6 (but this is just a **guess**). To engage in a larger trial, the researchers would like to know how many patients they should recruit into the study. Their resulting 90 percent confidence interval for p , the true population proportion of patients responding to the drug, should have a margin of error no greater than $m = 0.05$. What is the smallest **sample size** they would need for the Phase III trial?

- (a) 111
- (b) 259
- (c) 387
- (d) 404

14. True or False. The **correlation** measures the strength and direction of the linear relationship between two quantitative variables.

- (a) True
- (b) False

15. What is a **matched-pairs** experiment?

- (a) It is an experiment that matches control subjects with experimental subjects.
- (b) It is an experiment where subjects are randomly assigned to one of two treatments.
- (c) It is an experiment where each subject provides two responses; one to one treatment, and one to another.
- (d) It is an experiment where a statistically significant difference is observed.

16. True or False. We use sample statistics to estimate population parameters.

- (a) True
- (b) False

17. What is a **sampling distribution**?

- (a) It describes all the non-sampling errors that occur when sampling from a population.
- (b) It describes how a statistic's value will change from sample to sample.
- (c) It describes how randomization is used to guarantee unbiased estimates.
- (d) It summarizes all sources of variability not explained by a least-squares regression model.

18. In the bird-oxygen rate example we did in class, we examined how much oxygen birds consumed at different temperatures. We computed the least-squares regression equation to be

$$\text{O2rate} = 3.5 - 0.1 (\text{Temperature})$$

What oxygen rate (measured in ml/g/hr) would you **predict** for a bird subjected to a temperature of 8 degrees?

- (a) -0.1
- (b) 2.7
- (c) 3.5
- (d) 4.3

19. If my weekly Jillian's bar tab has **increased 100 percent**, what has happened?

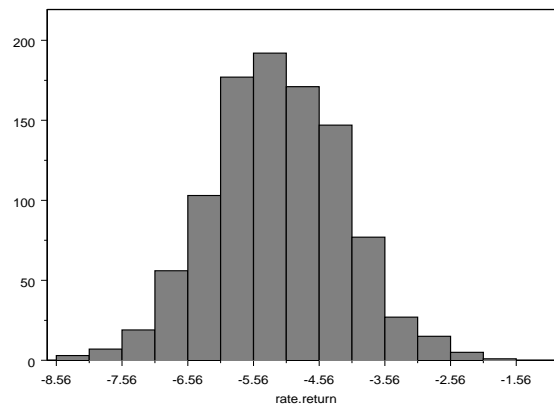
- (a) My bar tab has increased 100 dollars.
- (b) My bar tab has doubled.
- (c) My bar tab has remained the same.
- (d) It makes no sense for an amount to increase by 100 percent.

20. On November 18, 2006, Ohio State University beat the University of Michigan 42-39 in a college football match-up. This game was of particular interest because OSU was ranked #1 and UM was ranked #2 at the time. The very next day, the Ohio State lottery number chosen was 4239—identical to the score of the game! A CNN announcer, when reporting the anomalous lottery selection said, "I'm not a mathematician, but what is the chance of that?" Of course, it is not difficult to conclude that the probability of this occurrence is $1/10000$. What is the lesson?

- (a) Probability does not describe what happens in real life.
- (b) Even events with small probability can occur.
- (c) There is clear evidence that someone rigged the results.
- (d) There is a strong correlation between lottery number selections and college football scores.

21. The text (in Chapter 2) discusses how one might use a **Table of Random Digits**. At a local consulting firm, when would it be appropriate to use this table?
- (a) to determine whether or not a set of finance charge data is approximately symmetric is shape
 - (b) to decide which accounting invoices are sampled
 - (c) to determine whether or not a population frame of customers is reliable
 - (d) to investigate whether or not a response variable is quantitative
22. A simple random sample of size $n = 25$ is drawn from a population with mean 50 and standard deviation 5. What is the **standard deviation** of the sample mean \bar{x} ?
- (a) 1
 - (b) 2
 - (c) 5
 - (d) 10
23. One example we looked at in class involved HIV positive rates among rural Chinese farmers. One of the variables studied was **education level**, which was recorded as “illiterate,” “primary,” and “secondary.” What type of variable is education level in this example?
- (a) a quantitative variable
 - (b) a categorical variable
 - (c) a response variable
 - (d) a lurking variable
24. We recently spent some time discussing the Women’s Interagency HIV Study (WIHS) which examined the effects of abusive violence on HIV status. The study was a **longitudinal** study. What does this mean?
- (a) Subjects were randomized to different treatments.
 - (b) Subjects’ confidentiality was preserved by using a randomized response technique.
 - (c) The effects of violence were statistically significant.
 - (d) The subjects were monitored over time.
25. Using the data from the study cited in Question 24, we computed the sample proportion of childhood abuse victims to be $\hat{p} = 399/1288 \approx 0.31$. What is the **upper limit** of a 99 percent confidence interval for the true proportion of childhood abuse victims?
- (a) 0.22
 - (b) 0.32
 - (c) 0.34
 - (d) 0.47

26. A random sample of $n = 1000$ portfolio rate returns (measured as a percentage) is taken; here are the data:



The distribution of the rate of returns is best described as

- (a) skewed left
- (b) symmetric
- (c) skewed right

27. In Question 26, the **sample mean** rate of return is closest to

- (a) 1
- (b) -1
- (c) 5
- (d) -5

27. In Question 26, the **sample standard deviation** is closest to

- (a) -1
- (b) 1
- (c) 4
- (d) 11

29. A phenomenon or process that produces results that can not be predicted **with certainty** is

- (a) unbiased
- (b) systematic
- (c) random
- (d) reliable

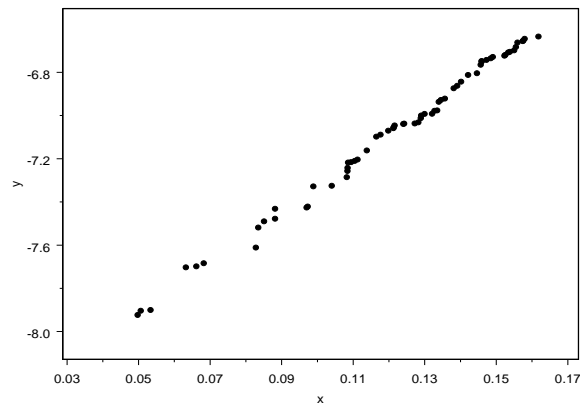
30. A measurement process is said to be **biased** if

- (a) it does not preserve subject anonymity.
- (b) it produces nearly the same result every time.
- (c) it consistently understates or overstates the true value.
- (d) None of the above.

31. Which statistic provides a measure of **variability** (or spread)?

- (a) mean
- (b) median
- (c) maximum
- (d) interquartile range

32. In the scatterplot below, the **correlation** is closest to



- (a) $r = -0.87$
- (b) $r = 0.25$
- (c) $r = 0.99$
- (d) $r = 1.03$

33. True or False. **Outliers** can greatly affect the value of the sample mean and sample standard deviation.

- (a) True
- (b) False

34. In any normal distribution, the proportion of observations that are **within 2 standard deviations of the mean** is closest to

- (a) 0.03
- (b) 0.68
- (c) 0.95
- (d) 0.98

35. What does it mean for a result to be **statistically significant**?

- (a) The result is too large to be consistent with chance variability.
- (b) The result proves that our study design was unbiased.
- (c) The data from the study were within the margin of error.
- (d) The study produced a result that is likely biased.

36. In the language of experiments, what is a **treatment**?

- (a) a data frame that consists of all experimental units
- (b) a group of individuals that are similar in some way
- (c) a randomization procedure
- (d) a condition or intervention which is applied to individuals

37. We would like to estimate the mean GPA of the population of all USC undergraduates. We would like to compute a 90 percent confidence interval for this mean, and we would like our interval to have margin of error no greater than 0.1. The Registrar's Office says that a reliable guess of the population standard deviation is 0.5. What is the smallest **number of students** we should recruit to achieve our goals?

- (a) 10
- (b) 68
- (c) 97
- (d) 144

38. We have learned that the sample proportion \hat{p} is an unbiased estimator for the population proportion p . Which one of the following statements is a result of this fact?

- (a) The sample proportion and population proportion will be equal.
- (b) The sample proportion will always be within the margin of error.
- (c) The variability in the values of \hat{p} increases as the sample size increases.
- (d) None of the above.

39. Cardiac pacemaker batteries will stop working if they are not properly calibrated and charged. Suppose that the length of time a battery operates follows a normal distribution with mean 25.0 days and standard deviation 2.0 days. The **first quartile** of this distribution is closest to

- (a) 1.4
- (b) 23.6
- (c) 28.4
- (d) 29.0

40. The density curve for starting salaries in Major League Baseball is highly skewed right. If the median starting income is \$500,000 per year, what is the **mean** starting income?

- (a) It is greater than \$500,000 per year.
- (b) It is equal to \$500,000 per year.
- (c) It is less than \$500,000 per year.

41. Suppose that we take a **random sample** of size n from a large population (e.g., USC students) and compute \hat{p} , the sample proportion of individuals who are left handed. Which of the following statements is **true**?

- (a) The larger the sample size, the smaller the margin of error.
- (b) The larger the sample size, the smaller the bias.
- (c) The larger the sample size, the larger the length of a 95 percent confidence interval for the proportion of left-handed individuals.
- (d) All of the above are true.

42. Last Friday, I collected some data (an observational study) to study the effects of pool accuracy (as measured by the percentage of made shots) and the number of (alcoholic) beers consumed for 25 pool players at Jillian's. For each player, I recorded x = the number of beers consumed, and y = the percentage of made shots. For these 25 data pairs, I then computed the correlation to be $r = 0.07$. What does this correlation tell me about the **relationship** between alcohol intake and pool accuracy?

- (a) There is almost no relationship between alcohol and pool accuracy.
- (b) An increase in alcohol intake causes players to play worse.
- (c) Players will play their best before they have consumed any alcohol.
- (d) None of the above.

43. True or False. Judith's Singer's *Chance* article "Afraid to Discuss Evolution" argued that it is not politically correct to teach evolution in Louisiana high schools.

- (a) True
- (b) False

44. You are working at General Electric and are participating in a study that is examining the **starting time** of jet engines (y , measured in seconds) and the **amount of thrust** exerted by them (x , measured in a scale similar to "horsepower"). In a recent study, we tested 10 engines and computed the least-squares regression line to be

$$\text{TIME} = 3.80 - 0.002 \text{ THRUST}.$$

What is the **slope** of the regression line for these data?

- (a) 3.80
- (b) -0.002
- (c) 0.002
- (d) It is impossible to tell without more information.

45. True or False. The use of poorly-worded questions is an example of a **nonsampling error**.

- (a) True
- (b) False

46. What is one of the **most important** things we have learned this semester?

- (a) Clinical trials are always ethical.
- (b) Be very skeptical of what you read and hear in the news.
- (c) President Bush has done a terrible job handling the Iraq war.
- (d) Statistics is an exact science.

47. The length of human pregnancies (with no complications) from conception to birth varies according to a normal distribution with mean $\mu = 260$ days and standard deviation $\sigma = 10$. **Five percent** of pregnancy lengths are **larger** than what value?

- (a) 244
- (b) 266
- (c) 276
- (d) 280

48. In recent years, “nondrowsy” cold and allergy medicines have been developed. One such product, Hismanal, is claimed “to cause drowsiness in no more than 3 percent of the public.” In a controlled preliminary study, out of $n = 200$ workers that took Hismanal, 20 experienced drowsiness. Treat this sample as an SRS from a large population of industrial workers. What is the **sample proportion** of workers that experienced drowsiness?

- (a) 0.01
- (b) 0.04
- (c) 0.10
- (d) 0.14

49. In Question 48, what is the **lower limit** of the 95 percent confidence interval for the true proportion of workers that will experience drowsiness with Hismanal?

- (a) 0.06
- (b) 0.08
- (c) 0.10
- (d) 0.17

50. Coach Spurrier likes statistics. In fact, after each game he examines many variables to prepare for his next opponent. Which one of the following variables is **categorical**?

- (a) the number of passing yards for the quarterback
- (b) the color of the opposing team’s jerseys
- (c) the attendance
- (d) the number of plays ran by the offense.

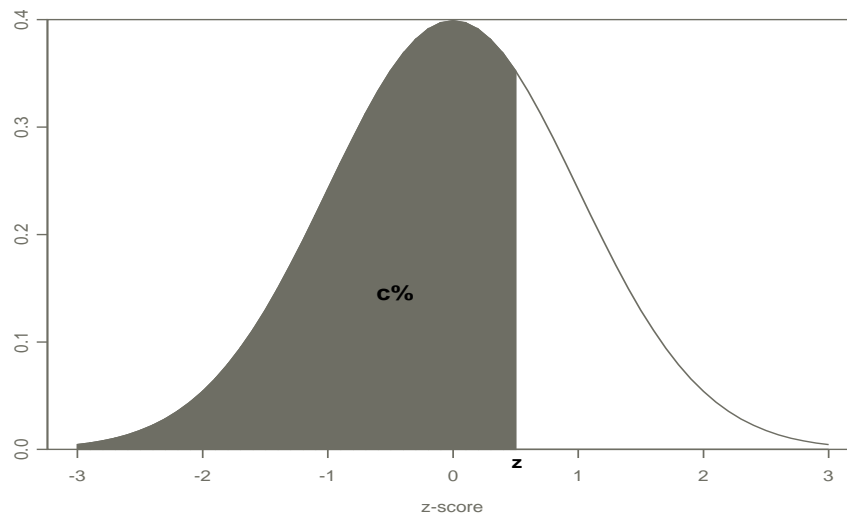


Table 1: *Percentiles of the standard normal distribution.*

Standard score (z)	Percentile (c)	Standard score (z)	Percentile (c)	Standard score (z)	Percentile (c)
-3.4	0.03	-1.1	13.57	1.2	88.49
-3.3	0.05	-1.0	15.87	1.3	90.32
-3.2	0.07	-0.9	18.41	1.4	91.92
-3.1	0.10	-0.8	21.19	1.5	93.32
-3.0	0.13	-0.7	24.20	1.6	94.52
-2.9	0.19	-0.6	27.42	1.7	95.54
-2.8	0.26	-0.5	30.85	1.8	96.41
-2.7	0.35	-0.4	34.46	1.9	97.13
-2.6	0.47	-0.3	38.21	2.0	97.73
-2.5	0.62	-0.2	42.07	2.1	98.21
-2.4	0.82	-0.1	46.02	2.2	98.61
-2.3	1.07	0.0	50.00	2.3	98.93
-2.2	1.39	0.1	53.98	2.4	99.18
-2.1	1.79	0.2	57.93	2.5	99.38
-2.0	2.27	0.3	61.79	2.6	99.53
-1.9	2.87	0.4	65.54	2.7	99.65
-1.8	3.59	0.5	69.15	2.8	99.74
-1.7	4.46	0.6	72.58	2.9	99.81
-1.6	5.48	0.7	75.80	3.0	99.87
-1.5	6.68	0.8	78.81	3.1	99.90
-1.4	8.08	0.9	81.59	3.2	99.93
-1.3	9.68	1.0	84.13	3.3	99.95
-1.2	11.51	1.1	86.43	3.4	99.97