

## Midterm 1

You have until 10:20am to complete this exam. Please remember to put your name, section and ID number on both your scantron sheet and the exam. Fill in test form A on the scantron sheet. Answer all multiple choice questions on your scantron sheet. Choose the single best answer for each multiple choice question. Answer the long answer questions directly on the exam. Keep your answers complete but concise. For the long answer questions, you must show your work where appropriate for full credit.

**Name:**

**ID Number:**

**Section:**

### (POTENTIALLY) USEFUL FORMULAS

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	$Pr[ T_{n-1}  > t_{\frac{\alpha}{2}, n-1}] = \alpha$
$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	$\sum_{i=1}^n a = na$
$CV = \frac{s}{\bar{x}}$	$\sum_{i=1}^n (ax_i) = a \sum_{i=1}^n x_i$
$skew = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^3$	$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$
$kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$	$s^2 = \bar{x}(1 - \bar{x})$ for proportions data
$\mu = E(X)$	$t_{\alpha, n-1} = TINV(2\alpha, n-1)$
$z^* = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$Pr( T_{n-1}  \geq  t^* ) = TDIST( t^* , n-1, 2)$
$t^* = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$	$Pr(T_{n-1} \geq t^*) = TDIST(t^*, n-1, 1)$
$Pr[T_{n-1} > t_{\alpha, n-1}] = \alpha$	$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

### (POTENTIALLY) USEFUL EXCEL OUPUT

TINV(.005,999)=2.81	TINV(.005,99)=2.87
TINV(.01,999)=2.58	TINV(.01,99)=2.63
TINV(.02,999)=2.33	TINV(.02,99)=2.36
TINV(.025,999)=2.24	TINV(.025,99)=2.28
TINV(.05,999)=1.96	TINV(.05,99)=1.98
TINV(.10,999)=1.65	TINV(.10,99)=1.66
TINV(.20,999)=1.28	TINV(.20,99)=1.29

## SECTION I: MULTIPLE CHOICE (60 points)

1. The variance of the sample mean of  $X$  will always be:
  - (a) Greater than or equal to the variance of  $X$ .
  - (b) Less than or equal to the variance of  $X$ .
  - (c) Equal to the variance of  $X$ .
  - (d) None of the above.
2. The nominal interest rate is equal to the real interest rate plus the inflation rate. If the mean of the real interest rate over the past year is positive and the mean of the inflation rate over the past year is positive, we can say for certain that:
  - (a) The mean of the nominal interest rate over the past year is positive.
  - (b) The variance of the nominal interest rate over the past year is greater than or equal to zero.
  - (c) Neither (a) nor (b) is necessarily true.
  - (d) Both (a) and (b) are true.
3. Which of the following is an example of panel data:
  - (a) A dataset of income and educational attainment for 1,000 random individuals.
  - (b) A dataset with ten years of monthly unemployment rates for each county in California.
  - (c) A dataset giving current GDP for 100 different countries.
  - (d) A dataset with daily caloric intake for one person over a six month period.
4. Suppose that  $X_t$  is a variable measuring monthly interest rates and  $\tilde{X}_t$  is a five-month moving average of  $X_t$ . Which of the following would be the best way to look at just the high frequency fluctuations in the interest rates?
  - (a) A line graph of  $X_t$ .
  - (b) A line graph of  $\tilde{X}_t$ .
  - (c) A line graph of  $X_t + \tilde{X}_t$ .
  - (d) A line graph of  $X_t - \tilde{X}_t$ .
5. The true population mean for the height of American males is 1.776 meters. Suppose that a researcher takes a random sample of 100 American males that has a sample mean is 1.655 meters and uses this information to reject the null hypothesis that the mean height in the population is greater than or equal to 1.700 meters. Which of the following can we say for certain:
  - (a) The researcher has done the hypothesis testing incorrectly.
  - (b) The researcher has committed a Type I error.
  - (c) The researcher has committed a Type II error.
  - (d) Both (a) and (b).
6. Suppose that for a random variable  $Y$ , the mean of the distribution of  $Y$  is greater than the median of the distribution of  $Y$ . The distribution of the sample mean of  $Y$  will be:
  - (a) Left skewed.
  - (b) Right skewed.
  - (c) Symmetric.
  - (d) Either (a) or (b) could be true.

7. Which of the following would not affect the width of a confidence interval for the mean of  $X$ ?
  - (a) The standard deviation of the sample.
  - (b) The mean of the sample.
  - (c) The significance level chosen for the confidence interval.
  - (d) All of the above would influence the width of the confidence interval.
8. If you can reject the null hypothesis that  $\mu = 10$  at the 5% significance level, which of the following is definitely true?
  - (a) You can reject the null hypothesis that  $\mu = 10$  at the 1% significance level.
  - (b) You can reject the null hypothesis that  $\mu = 10$  at the 10% significance level.
  - (c) You can reject the null hypothesis that  $\mu \leq 10$  at the 10% significance level.
  - (d) Both (b) and (c) are true.
9. Suppose that the population growth rate is a constant 15% per year. Which of the following statements is true?
  - (a) A graph with year on the horizontal axis and population on the vertical axis will have a slope of 15.
  - (b) A graph with year on the horizontal axis and population on the vertical axis will have a slope of 0.15.
  - (c) A graph with year on the horizontal axis and  $\ln(\text{population})$  on the vertical axis will have a slope of 15.
  - (d) A graph with year on the horizontal axis and  $\ln(\text{population})$  on the vertical axis will have a slope of 0.15.
10. Suppose that we take a sample of incomes for 1,000 people. Which of the following is not realization of a random variable?
  - (a) The mean income in the sample.
  - (b) The sample size.
  - (c) The variance of income in the sample.
  - (d) All of the above are realizations of random variables.
11. Suppose that we create a new variable  $Z$  that is equal to the variable  $X$  plus 10 (so  $z_i = x_i + 10$  for observation  $i$ ). Which of the following statements is definitely true?
  - (a) The mean of  $Z$  is equal to the mean of  $X$ .
  - (b) The standard deviation of  $Z$  is equal to the standard deviation of  $X$ .
  - (c) The median of  $Z$  is equal to the median of  $X$ .
  - (d) None of the above are definitely true.
12. Suppose that you are doing a univariate hypothesis test and calculate a p-value of 0.08. You would reject the null hypothesis if:
  - (a) You are using a 5% significance level.
  - (b) You are using a 10% significance level.
  - (c) Both (a) and (b) are true if you are doing a two-tailed test.
  - (d) Both (a) and (b) are true if your are doing a one-tailed test.

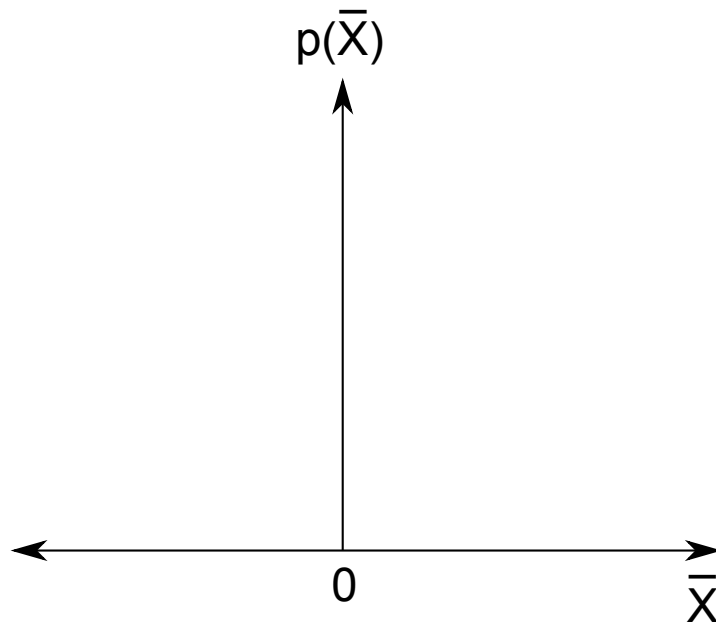
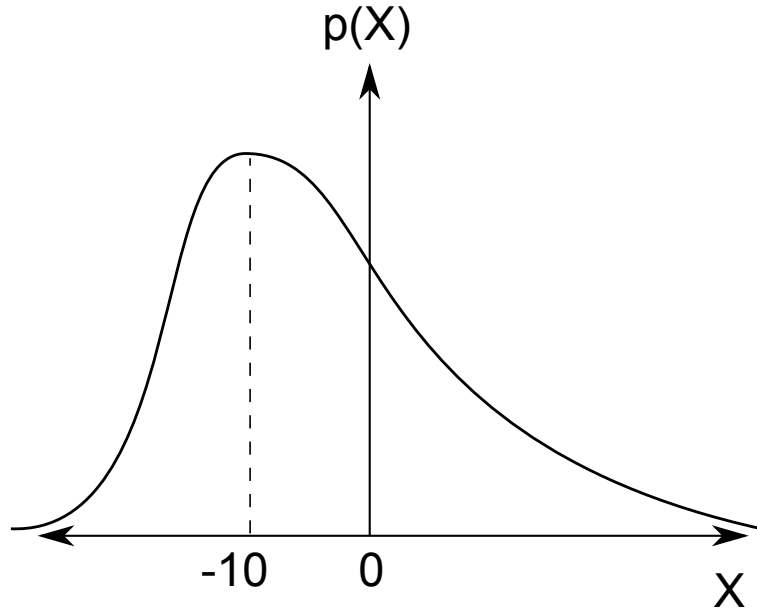
13. Suppose that 99% of the population in a country have annual incomes between \$0 and \$80,000. The other 1% of the population all have annual incomes over \$10,000,000. Which of the following is not true?
- (a) The mean income for the population is less than \$100,000.
  - (b) The median income for the population is less than \$100,000.
  - (c) The median income for the population is less than the mean income.
  - (d) In a random sample of 100 people, the expected value of the sample mean is greater than \$100,000.
14. Increasing the significance level used for our hypothesis testing from 5% to 10% will:
- (a) Decrease the probability of a Type I error but have no effect on the probability of a Type II error.
  - (b) Decrease the probability of a Type I error and increase the probability of a Type II error.
  - (c) Decrease the probability of both a Type I and Type II error.
  - (d) None of the above.
15. Suppose that weight increases with age. We take a sample of 20-year-olds and use that sample to estimate the mean weight in the population of all people 20 years old and older. Which of the following statements is not true?
- (a) The sample mean will be an unbiased estimator of the population mean.
  - (b) The sample variance will depend on the size of the sample we use.
  - (c) The expected value of the sample mean will be smaller than the population mean.
  - (d) The sample variance will be greater than or equal to zero.
16. Which of the following would not decrease the probability of a Type II error?
- (a) A larger sample size.
  - (b) A smaller variance of the variable of interest.
  - (c) Using a smaller significance level.
  - (d) All of the above would decrease the probability of a Type II error.
17. The probability of getting a sample mean  $\bar{x}$  that is greater than the median of  $x$  is:
- (a) Equal to 50%.
  - (b) Greater than 50%.
  - (c) Less than 50%.
  - (d) Not enough information.
18. A dataset of 1,000 voters that contains information on which presidential candidate people voted for in the last election is an example of:
- (a) Categorical data.
  - (b) Cross-sectional data.
  - (c) Both (a) and (b).
  - (d) Neither (a) nor (b).

19. Suppose that we have a discrete random variable  $Z$  that can take on the values 1, 2, 3, 4 or 5. For a sample of 800 observations of  $Z$ , we plot a histogram with a separate bar for each of these five values. The value on the horizontal axis that corresponds to the highest point on this histogram is:
- (a) Equal to the mean of  $Z$  in the sample.
  - (b) Equal to the median of  $Z$  in the sample.
  - (c) Equal to the mode of  $Z$  in the sample.
  - (d) None of the above are necessarily true.
20. Which of the following datasets would be the best achoice for examining the differences in the standard of living over time in the United States?
- (a) Cross-sectional data on GDP.
  - (b) Time series data on GDP.
  - (c) Cross-sectional data on GDP per capita.
  - (d) Time series data on GDP per capita.

## SECTION II: SHORT ANSWER (40 points)

1. (15 points) Suppose that there is a ballot initiative to overturn Proposition 13. In a sample of 1000 voters, 470 voters say they would vote to overturn Prop 13 and 530 voters say they would vote to keep Prop 13 intact.
  - (a) Suppose we create a variable  $X$  that equals one if a person votes to overturn Proposition 13 and equals zero if the person votes to keep Proposition 13 intact. What is the mean and standard deviation of  $X$  in the 1000 person sample?
  - (b) Based on the information above, calculate a 90% confidence interval for the percentage of votes that will be cast on election day in favor of overturning Proposition 13.
  - (c) Suppose that if the vote is tied, there will need to be another election. The California government needs to decide whether or not to print up a second set of ballots. Write down the set of hypotheses the government will use to determine whether they will need to print a second set of ballots. Your hypotheses should be written in term of  $X$  (or parameters relating to the distribution of  $X$ ).
  - (d) Use the information in the problem to test your hypotheses from part (c) using a 10% significance level. Be certain to show your work and clearly state your conclusions.

2. (10 points) The top graph below shows the distribution of the random variable  $X$ . The median of  $X$  is  $-5$ , the mean of  $X$  is  $5$  and the variance of  $X$  is  $25$ . On the lower graph, draw the distribution for the sample mean of  $X$ . Label the mean and median of this distribution on the graph, including their numerical values if possible.



3. (15 points) A researcher is testing for dangerous levels of lead in toys coming into the United States. A dangerous level of lead is considered any level greater than 600 parts per million (ppm). The researcher has a sample of 100 toys that she tests for lead levels and finds a mean level of 595 ppm and a standard deviation of 20 ppm.
- (a) Assume that the researcher wants to be cautious and error on the side of declaring toys dangerous. In other words, the researcher would rather assume the toys are dangerous and place the burden of proof on showing that the toys are safe. Write down the null and alternative hypotheses the researcher would use to test whether the toys are safe.
  - (b) Given your hypotheses in part (a), which would the researcher consider worse, a Type I error or a Type II error? Explain your answer in no more than two sentences.
  - (c) Find the critical value the researcher will use to test the hypotheses from part (a) assuming she wants to use a 5% significance level.
  - (d) Use the critical value you found in part (c) and the information given in the problem to test your hypothesis from part (a), being certain to formally state your conclusions.