

# Normal distribution



**ASSIGN  
BUSTER**

Final Exam Review Questions Solutions Guide You will probably want to PRINT THIS so you can carefully check your answers. Be sure to ask your instructor if you have questions about any of the solutions given below. 1.

Explain the difference between a population and a sample. In which of these is it important to distinguish between the two in order to use the correct formula? mean; median; mode; range; quartiles; variance; standard deviation. Solution: A sample is a subset of a population. A population consists of every member of a particular group of interest.

The variance and the standard deviation require that we know whether we have a sample or a population. 2. The following numbers represent the weights in pounds of six 7year old children in Mrs. Jones' 2nd grade class.

{25, 60, 51, 47, 49, 45} Find the mean; median; mode; range; quartiles; variance; standard deviation. Solution: mean = 46. 166.... median = 48 mode does not exist range = 35 Q1 = 45 Q2 = median = 48 Q3 = 51

variance = 112. 1396 standard deviation = 10. 59 3. If the variance is 846, what is the standard deviation? Solution: standard deviation = square root of variance =  $\sqrt{846} = 29.086$  4.

If we have the following data 34, 38, 22, 21, 29, 37, 40, 41, 22, 20, 49, 47, 20, 31, 34, 66 Draw a stem and leaf. Discuss the shape of the distribution.

Solution: 2 3 4 5 6 | | | | 219200 48714 0197 6 This distribution is right skewed (positively skewed) because the " tail" extends to the right. 5. What type of relationship is shown by this scatter plot? 45 40 35 30 25 20 15 10 5 0 0 5 10 15 20 Solution: Weak positive linear correlation 6. What values can r take in linear regression? Select 4 values in this interval and describe how they would be interpreted. Solution: the values are between -1 and +1

inclusive. 1 means strong negative correlation +1 means strong positive correlation 0 means no correlation . 5 means moderate positive correlation etc. 7. Does correlation imply causation? Solution: No. 8. What do we call the  $r$  value. Solution: The correlation coefficient. 9. To predict the annual rice yield in pounds we use the equation  $y = 859 + 5.76x_1 + 3.82x_2$ , where  $x_1$  represents the number of acres planted (in thousands) and where  $x_2$  represents the number of acres harvested (in thousands) and where  $r^2 = .94$ . a) Predict the annual yield when 3200 acres are planted and 3000 are harvested. ) Interpret the results of this  $r^2$  value. c) What do we call the  $r^2$  value? Solution: ? (a)  $y = 859 + 5.76 \cdot 3200 + 3.82 \cdot 3000 = 859 + 18432 + 11460 = 30751$  which is 30,751,000 pounds of rice (b) 94% of the variation in the annual rice yield can be explained by the number of acres planted and harvested. The remaining 6% is unexplained and is due to other factors or to chance. (c) It is the coefficient of determination. 10. The Student Services office did a survey of 500 students in which they asked if the student is part-time or full-time. Another question asked whether the student was a transfer student.

The results follow.

	Transfer	Non-Transfer	Row Totals
Part-Time	100	170	270
Full-Time	120	230	350
Column Totals	220	280	500

a) If a student is selected at random (from this group of 500 students), find the probability that the student is a transfer student.  $P(\text{Transfer})$  b) If a student is selected at random (from this group of 500 students), find the probability that the student is a part time student.  $P(\text{Part Time})$  c) If a student is selected at random (from this group of 500 students), find the probability that the student is a transfer student and a part time student.  $P(\text{transfer} \cap \text{part})$

time). ) If a student is selected at random (from this group of 500 students), find the probability that the student is a transfer student if we know he is a part time student.  $P(\text{transfer} \mid \text{part time})$ . e) If a student is selected at random (from this group of 500 students), find the probability that the student is a part time given he is a transfer student.  $P(\text{part time} \mid \text{transfer})$  f)

Are the events part time and transfer independent? Explain mathematically.

g) Are the events part time and transfer mutually exclusive. Explain

mathematically. Solution: (a) The total number of transfer students is 270.

The total number of students in the survey is 500.  $P(\text{Transfer}) = 270/500 = .54$

(b) The total number of part time students is 210. The total number of

students in the survey is 500.  $P(\text{Part Time}) = 210/500 = .42$

(c) From the table we see that there are 100 students which are both transfer and part

time. This is out of 500 students in the sample.  $P(\text{transfer} \cap \text{part time}) =$

$100/500 = .20$  (d) This is conditional probability and so we must change the

denominator to the total of what has already happened. There are 100

students which are both transfer and part time. There are 210 part time

students.

$P(\text{transfer} \mid \text{part time}) = 100/210 = .4762$  (e)  $P(\text{part time} \mid \text{transfer}) =$

$100/270 = .3704$  (f) The definition of independent is  $P(A \mid B) = P(A)$ . To test

we ask if  $P(\text{part time} \mid \text{transfer}) = P(\text{part time})$ ? Is  $.3704 = .42$ ? No, there for

the events are not independent. We could also test  $P(\text{transfer} \mid \text{part time}) =$

$P(\text{transfer})$ . Is  $.4762 = .54$ ? Again, the answer is no. (g) For events to be

mutually exclusive their intersection must be 0. In part c we found that

$P(\text{transfer} \cap \text{part time}) = 100/500 = .20$ . Therefore the events are not

mutually exclusive. 11. A shipment of 40 television sets contains 3 defective units.

How many ways can a vending company can buy five of these units and receive no defective units? Solution: There are 37 sets which are not defective. There are  ${}^{37}C_5$  ways to get 5 sets with none defective.  ${}^{37}C_5 = 435,897$ . Thus, there are 435,897 ways to get 5 sets with non defective. 12.

How do you recognize a discrete distribution? Solution: The outcomes are integers and they are “countable”. 13. The random variable  $X$  represents the annual salaries in dollars of a group of teachers. Find the expected value  $E(X)$ .  $X = \{\$35,000; \$45,000; \$55,000\}$ .  $P(35,000) = .4$ ;  $P(45,000) = .$  ;

$P(55,000) = .3$  Solution:  $E(X) = 35,000 \cdot .4 + 45,000 \cdot .3 + 55,000 \cdot .3 = \$44,000$  14. How do you recognize a binomial experiment? Solution: There

are exactly 2 outcomes: success and failure The trials are independent. The probability of success is the same in each trial. There are a fixed number of trials. 15. An advertising agency is hired to introduce a new product. The agency claims that after its campaign 61% of all consumers are familiar with the product. We ask 7 randomly selected customers whether or not they are familiar with the product. a) Is this a binomial experiment?

Explain how you know. b) Use the correct formula to find the probability that, out of 7 customers, exactly 4 are familiar with the product. Show your calculations. Solution: (a) answers vary, but must discuss the assumptions: Fixed number of independent trials, only two possible outcomes in each trial  $\{S, F\}$ , probability of success is the same for each trial, and random variable  $x$  counts the number of successful trials. (b)  $n = 7$ ;  $p = .61$ ; success =

consumers familiar with product We want  $P(4)$ . From the Binomial Formula or Excel,  $P(x=4) = 0.287516$ . How do you recognize a Poisson experiment?

Solution: You are measuring things in an interval and you know the average from past experience. 17. The mean number of cars per minute going through the Eisenhower turnpike automatic toll is about 7. Find the probability that exactly 3 will go through in a given minute using the correct table, formula, or Excel function. Solution: Poisson with average of 7. want  $P(3)$ .  $P(3) = .052$  18. How do you recognize a normal distribution? Solution: It is symmetric about the mean and has a bell-shape to it. 19. Label a) b) c) d) the following as continuous or discrete distributions.

The lengths of fish in a certain lake. The number of fish in a certain lake. The diameter of 15 trees in a forest. How many trees are on a farmer's acre.

Solution: (a) continuous; (b) discrete; (c) continuous; (d) discrete 20. Jack weighs 160 pounds and his sister weighs 110 pounds. If the mean weight for men his age is 175 with a standard deviation of 14 pounds and the mean weight for women is 145 with a standard deviation of 10 pounds, determine whose weight is closer to "average." Write your answer in terms of z-scores and areas under the normal curve. Solution: for Jack  $z = 1.07$ , area = .423 for his sister = .0062  $x? \mu = x? \mu ? = 160 ? 175 ? 15 = ? ? 1.07$ , when  $z = 1.07$ , area = .423  $14 ? 110 ? 145 ? 25 = = ? 2.50$ , when  $z = -2.50$ , area = .0062 Jack is closer to average, but he is still in the lower 14%, while his sister is in the bottom less than 1% of the population. 21. On a dry surface, the braking distance (in meters) of a certain car is a normal distribution with  $\mu = 45.1$  m and  $\sigma = 0.5$  (a) Find the braking distance that corresponds to  $z = 1.8$  (b) Find the braking distance that represents the 91st percentile.

(c) Find the z-score for a braking distance of 46. m (d) Find the probability that the braking distance is less than or equal to 45 m (e) Find the probability that the braking distance is greater than 46. 8 m (f) Find the probability that the braking distance is between 45 m and 46. 8 m. Solution:

(a)  $\rightarrow x = 46$   $z = \frac{x - \mu}{\sigma} = \frac{46 - 45.1}{1.1} = 0.818$   $\rightarrow .9 = \frac{x - 45.1}{1.1} \rightarrow x = 45.775$  (b) We need to look in the table for the z-score which corresponds to an area to the left of z of .9100. The closest is  $z = 1.34$ . Using the same technique as part a we have  $x - 45.1 = 1.34 \times 1.1 \rightarrow x = 45.775$  (c) Using

the same formula a little differently  $\frac{x - 46.1}{1.1} = 2.55$

The z score corresponding to  $x = 45$  is  $z = \frac{45 - 45.1}{1.1} = -.2$

$P(x < 45) = P(z < -.2) =$  area to the left of  $x = -.2$  which is .4207 (d) The z score corresponding to  $x = 46.8$  is  $\frac{46.8 - 45.1}{1.1} = 1.7$   $z = 1.7$  The area to the left of  $z = 1.7$  is .9554  $P(x > 46.8) = P(z > 1.7) = 1 - P(z < 1.7) = 1 - .9554 = .0446$  (f)  $P(45 < x < 46.8) = P(-.2 < z < 1.7) = P(z < 1.7) - P(z < -.2) = .9554 - .4207 = .5347$  (e) 22. A drug manufacturer wants to estimate the mean heart rate for patients with a certain heart condition.

Because the condition is rare, the manufacturer can only find 14 people with the condition currently untreated. From this small sample, the mean heart rate is 101 beats per minute with a standard deviation of 8. (a) Find a 99% confidence interval for the true mean heart rate of all people with this untreated condition. Show your calculations. (b) Interpret this confidence interval and write a sentence that explains it. Solution: (a). Since sample size  $= n = 14 < 30$  and the population standard deviation is unknown, we must use a t-value. For a 99% confidence level and 13 degrees of freedom (from  $14-1=13$ ), t-value = 3.012.

Also, sample mean =  $\bar{x} = 101$ ; sample standard deviation =  $s = 8$ .  $E = t^* s / \sqrt{n} = 3.012 * 8 / \sqrt{14} = 6.44$  (rounded from 6.4399)  $\bar{x} + E = 101 + 6.44 = 107.44$   $\bar{x} - E = 101 - 6.44 = 94.56$  Thus, 99% confidence interval = (94.56, 107.44) (b) We are 99% confident that the true mean heart rate of all people with this heart condition is between 94.56 and 107.44.

23. Determine the minimum required sample size if you want to be 80% confident that the sample mean is within 2 units of the population mean given  $\sigma = 9.4$ . Assume the population is normally distributed. Solution:  $n = (Z_c * \sigma / E)^2 = [(1.28 * 9.4) / 2]^2 = (6.16)^2 = 36.192256 = 37$  rounded up. (We always round up to the next whole number.)

24. A social service worker wants to estimate the true proportion of pregnant teenagers who miss at least one day of school per week on average. The social worker wants to be within 5% of the true proportion when using a 90% confidence interval. A previous study estimated the population proportion at 0.21. (a) Using this previous study as an estimate for  $p$ , what sample size should be used? (b) If the previous study was not available, what estimate for  $p$  should be used? Solution: (a). The critical  $z$ -value for a 90% confidence interval is 1.645.

Since a previous study is known, we can use it to estimate  $p = 0.21$ . The maximum error is 0.05. Sample size =  $n = p(1-p) * (z / \text{error})^2 = 0.21 * (0.79) * (1.645 / 0.05)^2 = 179.5718 = 180$  rounded up Thus, at least 180 pregnant teenagers must be sampled. (b). If no estimate of  $p$  is known, we must use  $p = 0.5$  to have a large enough sample size to meet the desired maximum error.

25. Suppose you are performing a hypothesis test on a claim about a population proportion. Using an  $\alpha = .04$  and  $n = 90$ , what



two critical values determine the rejection region if the null hypothesis is  $H_0: p = 0.54$ ? (a)  $\pm 1.96$  (b)  $\pm 2.05$  (c)  $\pm 2.3$  (d) none of these Solution:

Answer is (b). If  $H_0: p = 0.54$ , then  $H_a: p \neq 0.54$  and the test is a two-tailed test. Since  $\alpha = .04$ , each tail has an area =  $.02$  (since  $.04/2 = .02$ ). The closest z-value corresponding to an area of  $.02$  to the left is  $z = -2.05$ . ( $z = -2.05$  corresponds to area =  $.0202$ ; this is the closest to  $.02$  available in our tables.) Since there are two tails, the upper z-value is  $2.05$ . Thus, the two z-values are  $\pm 2.05$ .

26. A restaurant claims that its speed of service time is less than 15 minutes. A random selection of 49 service times was collected, and their mean was calculated to be 14.5 minutes.

Their standard deviation is 2.7 minutes. Is there enough evidence to support the claim at  $\alpha = .07$ . Perform an appropriate hypothesis test, showing each important step. (Note: 1st Step: Write  $H_0$  and  $H_a$ ; 2nd Step: Determine Rejection Region; etc.) Solution:  $H_0: \mu \geq 15$  min.  $H_a: \mu < 15$  min. (claim). Therefore, it is a left-tailed test.  $n = 49$ ;  $\bar{x} = 14.5$ ;  $s = 2.7$ ;  $\alpha = 0.07$

Since  $n > 30$ , we can use a z-value test and replace  $\sigma$  with  $s$ . Since  $\alpha = 0.07$  and this is a left-tailed test, the critical z-value that corresponds to a left tail area of  $0.07$  is  $z_c = -1.48$ . (In the Standard Normal Distribution Table note that  $0.694$  (corresponding to  $z = 1.48$ ) is closer to  $0.07$  than  $0.0708$  is.) Standardized test statistic  $z = (\bar{x} - \mu) / (s / \sqrt{n})$   $z = (14.5 - 15) / (2.7 / \sqrt{49})$   $z = -0.5 / 0.3857142857$   $z = -1.296296$   $z = -1.30$  (rounded to two decimals) Since  $-1.30 > -1.48$ , we DO NOT REJECT  $H_0$ . That is, at  $\alpha = 0.07$ , there is NOT sufficient statistical evidence to conclude the true mean service time is less than 15 minutes. Thus, we can not support the restaurant's claim. (p-Value Method) For  $z = -1.30$ , the area

to the left = p-value = .0968. Since p-value = .0968 ; .07 = alpha, we do not reject  $H_0$ .