1. A hardware manufacturer produces bolts with a diameter of approximately 10 mm. A random sample of 20 bolts were measured and a stem-and-leaf plot of the data was constructed. Note that the stem unit is 0.1 mm and the leaf unit is 0.01 mm.

98	9
99	45678899
100	0123335588
101	0

(a) (3 points) Find the five-number summary.

(b) (2 points) Are there any outliers in this data? Show all your work.

2. You draw two cards randomly (one after the other, *with* replacement) from a standard deck of 52 cards. Define the following events:

A = first card is red. C = both cards are hearts.
B = first card is a queen. D = second card is a spade.

(a) (3 points) Are A and B disjoint events? Are A and B independent? Justify.

(b) (3 points) Are C and D disjoint events? Are C and D independent? Justify.

(c) (2 points) Find $P(B \cup C)$. (Recall B = first card is a queen, C = both cards are hearts.)

(d) (2 points) Find $P(C|A)$ and $P(A|C)$. (Recall A = first card is red, C = both cards are hearts.)

(e) (2 points) Find $P(A' \cap D')$. (Recall A = first card is red, D = second card is a spade.)

3. (4 points) A biometric security device uses fingerprints to allow or deny access to a secret research lab. For authorized personnel, the system sometimes denies access (1 in $1,000$ chance). Also, for unauthorized personnel, the device will sometimes erroneously allow access to the research lab (1 in $1,000,000$ chance). Assume that 95 percent of those who seek access are authorized personnel. If the alarm goes off and a person is denied access to the lab, what is the probability that the person was really authorized?

4. Santa Claus visits a shopping mall with 50 gifts in his bag, all containing different toys. Suppose that 30 gifts are wrapped in **red** paper, and 20 gifts are wrapped in **green** paper. When he arrives, a line-up of 10 children are waiting for him.

(a) (2 points) If Santa picks a random gift from his bag for each child, calculate the probability that all of the 10 children will get a gift of the same color.

(b) (2 points) If Santa picks a random gift from his bag for each child, calculate the probability that exactly 6 of the 10 children will receive a red gift.

(c) (2 points) Suppose instead that Santa selects 10 gifts from his bag, and gives them to children randomly, allowing some children to get multiple gifts and some children to get none. Calculate the probability that at least one child will not receive a gift.

5. Suppose X is a discrete random variable with the following probability distribution.

x	-1	0	0.5	1	2
$p(x)$	0.3	0.1		0.3	0.2

(a) (1 point) Fill-in the missing probability in the table.

(b) (2 points) Calculate the expected value and the variance of X.

(c) (2 points) Sketch the graph of the cumulative distribution.

(d) (1 point) Find $P(X \geq 0 \,|\, X \leq 0)$.

(e) (2 points) Suppose Y is another random variable with the same probability distribution as X. Also, X and Y are independent. Find the probability that the product of X and Y is equal to 1.

6. In a hospital, the number of babies born during a specific hour is a random variable X that follows a Poisson distribution with an average of 1.25 babies per hour. Suppose a nurse is starting an 8-hour shift.

(a) (2 points) What is the probability that this nurse will see at least 5 but less 15 babies born during her shift?

(b) (2 points) What is the probability that this nurse will see at least 1 baby born during each hour of her shift?

7. Two different gambling games are offered at a casino:

Game A	*Game B*
- costs 3$ to play,	- also costs 3$ to play,
- you roll 4 dice,	- first, you select 5 different numbers from 1 to 20,
- you win 5$ for each 6 that you roll.	- then the host randomly draws 5 different numbers from 1 to 20,
	- you win 3$ for each matching number (order doesn't matter).

(a) (5 points) Calculate the expected value and the variance of the net earnings for Game A and for Game B. Compare your answers; which game would you play?

(b) (3 points) Suppose you play Game A repeatedly 60 times in a row. What is the probability that the sum of all your earnings will be greater than 30$?

8. (4 points) For a certain brand of batteries, it has been discovered that 90% of batteries have an acceptable voltage. Suppose you buy a package containing 150 batteries. Use a normal approximation to calculate the probability that more than 125 of these batteries will have an acceptable voltage. Verify that it is valid to use such an approximation in this case.

9. Let X be a continuous random variable with the following probability density function:

$$f(x) = \begin{cases} 4(x - x^3) & \text{for } 0 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) (3 points) Find the mean of X.

(b) (3 points) Find the median of X.

10. (3 points) Suppose the grades given by a teacher follow a normal distribution with an average of 70 and a standard deviation of 8. A new semester is starting and this teacher has a group of 25 students. What is the probability that the class average will be between 70 and 75?

11. (3 points) Question 1 on this exam presented a sample of 20 bolt diameters. The sample average diameter is 10.0065 mm and the sample standard deviation of this data is 0.05224 mm. Construct a 95% confidence interval for the true standard deviation of the diameter of bolts manufactured.

12. (3 points) In the Great Sand Dunes National Park, a researcher is interested in estimating the true average wind speed by constructing a 94% confidence interval. Find the minimum sample size required in order for the margin of error to be at most 10 cm/s. Suppose it is known that the standard deviation of wind speeds is approximately $\sigma \approx 100$ cm/s, and you may assume that the sample size will be above 40.

13. For several years, evidence has been mounting that folic acid reduces major birth defects. One particular study that was published in the *New England Journal of Medicine* gave the follow results. A group of 2701 women took daily multivitamins containing folic acid during their pregnancy, and major birth defects occurred in 35 cases. A second group of 2052 women did not take any supplements containing folic acid, and birth defects occurred in 47 cases.

(a) (5 points) Use an appropriate hypothesis test to see if folic acid significantly reduces the occurrence of birth defects. Use a 1% significance level and follow the P-value approach.

(b) (1 point) Was the test applied in part (a) valid in this case? Check the requirements of the test or state any necessary assumptions.

14. Costs to hospitals (measured in US dollars, per patient, per day) are reported by the American Hospital Association. A sample of hospitals in Georgia is compared to a sample of hospitals in Illinois.

	Georgia	Illinois
n	12	15
$\overline{x}$	797.2	987.9
s	250	300.2

(a) (4 points) Construct a 98% confidence interval for the difference in the true average costs to hospitals in Georgia and Illinois.

(b) (1 point) Interpret your answer in the context of this application.

15. Gardemann et al. (1998) surveyed specific genotypes and their possible effect on coronary artery disease. The following two-way table summarizes the results.

	No Disease	Artery Disease	Total
Genotype A	268	807	1075
Genotype B	199	759	958
Genotype C	42	184	226
Total	509	1750	2259

(a) (6 points) Does this data indicate a significant relationship between the genotype and the occurrence of coronary artery disease? Use a 5% significance level and follow the P-value approach.

(b) (1 point) Interpret your answer in the context of this application.

16. (2 points) Find the 10^{th} and the 90^{th} percentiles of an F-distribution with a numerator degree of freedom of 20 and a denominator degree of freedom of 3.

17. A physician recorded the age and peak heart rate (during intensive exercise) for 10 randomly selected people.

age (**X**)	30	38	41	38	29	39	46	41	42	24
peak heart rate (**Y**)	186	183	171	177	191	177	175	176	171	196

$$\sum x_i = 368, \quad \sum x_i^2 = 13968, \quad \sum y_i = 1803, \quad \sum y_i^2 = 325723, \quad \sum x_i y_i = 65865.$$

(a) (4 points) Estimate the three parameters in the linear regression model (y-intercept, slope and variance of the error term).

(b) (3 points) For a given adult of age 34, calculate the probability that he/she has a peak heart rate between 180 and 190 during intense exercise.

18. Suppose you are planning a test on the correlation coefficient with the alternative $H_a : \rho < 0$, using a 5% significance level.

(a) (3 points) If the sample size is $n = 15$, what values of the sample correlation coefficient would be considered significantly negative (rejecting H_0)?

(b) (3 points) Suppose that the correlation coefficient $r = -0.5$ is claimed to be a significantly negative in the framework this test. What is the minimum sample size n for which this can be true?
Hint: changing n will modify the rejection region!

19. (1 point) State the definition of a type II error.

ANSWERS

1. (a) five-number summary: 9.89, 9.975, 10.005, 10.04, 10.10
   (b) no data outside the interval $[9.8775, 10.1375]$, therefore no outliers

2. (a) not disjoint, since we can draw the queen of hearts, satisfying both events
   independent, since $P(A \cap B) = P(A) \cdot P(B)$
   (b) disjoint, since the second card cannot be a heart and a spade
   not independent, since $P(C|D) = 0 \neq P(C)$
   (c) $P(B \cup C) = 0.1346$
   (d) $P(C|A) = 0.125$ and $P(A|C) = 1$
   (e) $P(A' \cap D') = 0.375$

3. 0.01865

4. (a) $\dfrac{C_{10}^{30}}{C_{10}^{50}} + \dfrac{C_{10}^{20}}{C_{10}^{50}} = 0.002943$ (b) $\dfrac{C_6^{30} \cdot C_4^{20}}{C_{10}^{50}} = 0.2801$ (c) $1 - \dfrac{P_{10}^{10}}{10^{10}} = 0.9996$

5. (a) 0.1 (b) $E(X) = 0.45$ and $V(X) = 1.2225$ (d) 0.25 (e) 0.22

6. (a) 0.888 (b) $(0.7135)^8 = 0.06717$

7. (a) Game A:
   X = number of sixes rolled $\sim$ binomial with $n = 4$, $p = \frac{1}{6}$
   Y = net earnings $= 5X - 3$
   $\Rightarrow \quad E(Y) = 5E(X) - 3 = 0.33 \quad \text{and} \quad V(Y) = 25V(X) = 13.8889$
   Game B:
   X = number of matching numbers $\sim$ hypergeometric with $N = 20$, $M = 5$, $n = 5$
   Y = net earnings $= 3X - 3$
   $\Rightarrow \quad E(Y) = 3E(X) - 3 = 0.75 \quad \text{and} \quad V(Y) = 9V(X) = 6.6612$
   Decision: would play Game B, yields a higher average earning.

   (b) $P(\text{sum of earnings} > 30) = P(\overline{Y} > 0.5) = P(Z > 0.35) = 0.3632$

8. 0.9962

9. (a) $E(X) = \frac{8}{15}$ (b) median $= 0.5412$

10. 0.4991

11. (0.03973, 0.07630)

12. $n \geq 354$

13. (a) difference of proportions, $z^* = -2.61$, P-value $= 0.0045$, reject H_0
    (b) yes, test is valild: $n_1\hat{p}_1 = 35 \geq 10, \quad n_1(1-\hat{p}_1) = 2666 \geq 10, \quad n_2\hat{p}_2 = 47 \geq 10, \quad n_2(1-\hat{p}_2) = 2005 \geq 10$

14. (a) $(-454.6208, 73.2208)$ (using t-distribution with $\nu = 24$)

    (b) We are 98% confident that the difference between the true average cost to hospitals in Georgia and Illinois is between -454.6208 and 73.2208 US dollars per patient per day.

15. (a) test of independence (χ^2-distribution with $\nu = 2$)
    $\chi^{2,*} = 7.2504 \quad \Rightarrow \quad 0.025 < P\text{-value} < 0.05 \quad \Rightarrow \quad \text{reject } \mathrm{H}_0$

    (b) There is sufficient evidence to conclude that there is a relationship between genotype and the occurrence of coronary artery disease, at the 5% significance level.

16. 90^{th} percentile $= F_{0.10,20,3} = 5.18$
    10^{th} percentile $= F_{0.90,20,3} = \dfrac{1}{F_{0.10,3,20}} = \dfrac{1}{2.38} = 0.4202$

17. (a) $\hat{\beta}_1 = -1.1405, \quad \hat{\beta}_0 = 222.2704, \quad s^2 = 11.0627$
    (b) $P(180 < Y < 190) = P(-1.05 < Z < 1.96) = 0.8281$

18. (a) $r \leq -0.4409$

    (b) to reject H_0, we need $t^* = \dfrac{-0.5\sqrt{n-2}}{\sqrt{1-0.5^2}} \leq -t_{n-2,0.05}$
    this inequality can be rewritten as $n \geq 3t^2_{n-2,0.05} + 2$
    the smallest value of n that satisfies this inequality is $n = 12$

19. A type II error is when we *fail to reject* the null hypothesis, when in fact the null hypothesis is *false*.