1. (i) (2 pts) Given a normal probability distribution, which of the following is true?

(a) The median is to the right of the mean.

(b) The mean is to the right of the median.

(c) The mean is equal to the median.

(d) Any of the above can occur depending on whether the normal distribution is symmetric, negatively or positively skewed.

**Answer.** C, normal distributions are always symmetric about their means.

(ii) (2 pts) According to Chebyshev’s theorem, at least what proportion of data in any distribution lies within 5 standard deviations of the mean?

**Answer.** $1 - \frac{1}{5^2} = .96$.

(iii) (1 pt) (True or False) Two events in a sample space are independent if they never occur at the same time.

**Answer.** False, the above statement describes mutually exclusive events.

(iv) (1 pt) (True or False) In an hypothesis test on the mean, a Type I error occurs if the null hypothesis is rejected when it is true.

**Answer.** True.

(v) (2 pts) For events $A$ and $B$ in a sample space $S$, we are told $P(A) = .6$ and $P(B) = .3$ and $P(A \text{ and } B) = .2$. Which of the following is true?

(a) $P(A|B) = 2/3$

(b) $P(A \text{ or } B) = .9$

(c) $A$ and $B$ are mutually exclusive.

(d) None of the above are true.

**Answer.** A, $P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{.2}{.3}$.

(vi) (2 pts) Which of the following is true about a binomial random variable for 10 trials with probability of success on each trial given as .1?

(a) The probability of no successes is $(.1)^{10}$

(b) The 10 trials must be identical and independent from one another.

(c) The probability of 10 successes is $(.9)^{10}$

(d) All of the above.

**Answer.** B
2. (i) (2 pts) Suppose that the maximum error $E = 4$ for a 95% confidence interval for the mean using a sample of size $n = 100$. What sample size would you expect to use from the same population for a 95% confidence interval of the mean to have a maximum error $E = 1$?

Answer. To cut the error by a factor of 4, the sample size must be multiplied by $4^2$. Therefore, we need a sample size $n = 1600$.

(ii) (2 pts) According to the empirical rule for bell shaped distributions, approximately what percentage of data lies within one standard deviation of the mean?

Answer. 68%

(iii) (2 pts) Suppose a weight of 50 pounds is at the 67th percentile for Springer Spaniel dogs. In a collection of 500 randomly selected Springer Spaniel dogs, how many would you expect to weigh more than 50 pounds?

Answer. Approximately 33% of 500 which is 165.

(iv) (1 pt) What z-score would you assign to a number that lies 3.5 standard deviations below the mean?

Answer. $-3.5$.

(v) (1 pt) (True or False) In hypothesis testing, the null hypothesis should be rejected if the level of significance is smaller than the $P$-value.

Answer. False, it should be rejected when the $P$-value $\leq$ level of significance.

(vi) (2 pts) Sketch a box and whisker plot for a collection of data given that $L = 10$, $Q_1 = 17$, $Q_2 = 35$, $Q_3 = 40$ and $H = 50$.

Answer. See Section 3.4 in your text.
3. (a) (2 pts) In how many ways can the state of Nevada make license plates of the form yxx-aaa where
the first character is a number 1 - 9, the second and third characters are numbers 0 - 9, and the last three
characters are letters A - Z (all uppercase)?

(b) (2 pts) Given that the probability that a newborn child is female is 48% and 52% that the child is male.
Use binomial probabilities to find the probability that exactly 11 of the first 20 babies born in 2007 were
female?

(c) (2 pts) In how many ways can three medals (gold, silver and bronze) be awarded in a 100m race that
contains 9 runners?

(d) Consider the random variable \( x \) whose probability distribution is given by the following table.

<table>
<thead>
<tr>
<th>( x )</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(x) )</td>
<td>.3</td>
<td>?</td>
<td>.3</td>
<td>.1</td>
</tr>
</tbody>
</table>

(i) (1 pt) Find \( P(x = 4) \). Answer =

(ii) (3 pts) Find the expected value of \( x \) and the standard deviation of \( x \).

Answer. (a) \( 9 \cdot 10 \cdot 10 \cdot 26 \cdot 26 \cdot 26 = 15,818,400 \) ways.

(b) \( P(11) = C_{20,11} \cdot (0.48)^{11} \cdot 0.52^{(20-11)} = 0.1455 \).

(c) \( P_{9,3} = 9 \cdot 8 \cdot 7 = 504 \) ways.

(d) (i) \( P(x = 4) = 1 - .7 = 0.3 \)

(ii) The expected value is

\[ \mu = E(x) = \sum x \cdot p(x) = 2(.3) + 4(.3) + 6(.3) + 8(.1) = 4.4. \]

The variance is

\[ \sigma^2 = \left( \sum x^2 \cdot p(x) \right) - \mu^2 \]

\[ = 4(.3) + 16(.3) + 36(.3) + 64(.1) - 4.4^2 \]

\[ = 3.84 \]

Therefore, the standard deviation is \( \sigma = \sqrt{3.84} = 1.959592. \)

4. Consider the sample of 30 numbers

31 37 44 46 47 51 52 65 68 69 70 71 72 73
75 75 76 76 77 78 79 80 81 82 83 83 85 88 89

(a) (3 pts) Construct a stem and leaf plot for this data with stems 3,4,5,6,7,8

(b) (7 pts) Construct a relative frequency histogram for the data where the first class has limits 30 – 39.
State the class width, and list the limits, boundaries and relative frequencies for all classes.

Answer. See Text Sections 2.2 and 2.3 for information on how to do questions like this.
5. (10 pts) Given the data 9, 12, 15, 17, 18, 19, 23, 45, 52, 61, 63, 63. One has \( \sum x = 397 \) and \( \sum (x - \mu)^2 \approx 5206.917 \). Find:

(a) the mean is \( \frac{397}{12} = 33.08 \).

(b) the sample variance is \( s^2 = \frac{5205.917}{11} = 473.36 \).

(c) the population variance is \( \sigma^2 = \frac{5205.917}{12} = 433.90915 \).

(d) the sample standard deviation is \( s = \sqrt{473.36} = 21.76 \).

(e) the median is \( \frac{19 + 23}{2} = 21 \).

(f) the 50th percentile of the data is 21 (the same as the median).

(g) the mode is 63 the most common data.

6. The following data is for investigating the relation between salary in thousands \( (x) \) and average number of absences per year \( (y) \).

<table>
<thead>
<tr>
<th>Salary ((x))</th>
<th>20</th>
<th>23</th>
<th>28</th>
<th>30</th>
<th>33</th>
<th>35</th>
<th>37</th>
<th>40</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences ((y))</td>
<td>2.4</td>
<td>2.2</td>
<td>1.9</td>
<td>2.1</td>
<td>1.5</td>
<td>1.4</td>
<td>1.3</td>
<td>0.5</td>
<td>0.4</td>
</tr>
</tbody>
</table>

For this data: \( \sum x = 288 \), \( \sum x^2 = 9660 \), \( \sum y = 13.7 \), \( \sum y^2 = 24.93 \), \( \sum xy = 398.2 \).

(a) (4 pts) Find the equation of the least squares regression line.

(b) (2 pts) Does the data appear to be positively or negatively correlated? Explain.

(c) (2 pts) Use the regression line to find: (i) How many absences per year would be expected from an employee that makes $38,000 per year? (ii) The expected salary for someone with 2 absences per year.

(d) (2 pts) Given that the correlation coefficient is \(-.945\), determine how well the data fits the line of best fit. Explain your answer.

Answer. (a) Therefore, the slope is \( b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} = -0.09054 \), and the \( y \)-intercept is \( a = \frac{13.7}{9} - (-0.09054) \left( \frac{288}{9} \right) = 4.4195 \).

Thus the equation of the line is \( y = -0.09054x + 4.4195 \).

(b) Negatively correlated because as the \( x \)'s increase, the \( y \)'s generally decrease.

(c) (i) Solve \( y = -0.09054(38) + 4.4195 = .98 \) absences per year. (ii) Solve \( 2 = -0.09054x + 4.4195 \) and so \( x = 26.723 \), or approximately $26,723 per year.

(d) A correlation coefficient of \( r = -1 \) would mean the data is on a line of negative slope, values close to \(-1\) indicate good fits to a line of negative slope. Since \(-.945\) is quite close to \(-1\), the fit should be quite good.
7. A business employs 300 men and 700 women. Of these employees, 60 men and 210 women have been working there for more than 10 years. Let $A$ be the event the employee is a woman, and $B$ be the event the employee has been employed for more than 10 years.

(a) (3 pts) Find $P(A)$ and $P(B)$.

(b) (3 pts) Find $P(A \text{ and } B)$ and $P(A \text{ or } B)$.

(c) (2 pts) Are the events $A$ and $B$ independent? Explain.

(d) (2 pts) Find the probability an employee is female, given that the length of employment is more than 10 years.

**Answer.**
(a) $P(A) = 0.7$ and $P(B) = 0.27$.

(b) $P(A \text{ and } B) = P(A)P(B|A) = (0.7) \left( \frac{210}{700} \right) = 0.21$.

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.76$.

(c) No, $P(A \text{ and } B) \neq P(A) \cdot P(B)$ since $0.21 \neq (0.7)(0.27)$.

(d) $\frac{210}{270} = 0.777778$.

8. In a recent survey of 180 randomly selected engineering graduates starting work in the United States, it was found that the average annual starting salary for the sample was $43,200 with a sample standard deviation of $6,000.

(a) (7 pts) Conduct an hypothesis test to determine whether the average starting salary for engineers is more than $42,500 per year using a significance level of $\alpha = 0.05$. Make sure to state $H_0$, $H_1$, the $P$-value, and your conclusion.

(b) (3 pts) If the true mean starting salary for engineers in the United States is actually $43,000 with population standard deviation $6,000, what is the probability that a random sample of 180 starting engineers will have a sample mean starting salary less than $44,000?

**Answer.**
(a) $H_0 : \mu = 42,500$, $H_1 : \mu > 42,500$. This is a right-tailed test with since the population standard deviation is not given, we use the $t$-distribution with $d.f. = 179$. The sample test statistic is

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{43,200 - 42,500}{\frac{6000}{\sqrt{180}}} = 1.565.$$ 

We go to $d.f = 100$ on the table, and find $1.451 < 1.564 < 1.660$ and find that $0.050 < P\text{-value} < 0.075$.

Because the $P$-value is larger than .05, we fail to reject $H_0$; there is not enough evidence to show that the true mean starting salary is more than $42,500 at the 5% level of significance.

(b) $P(\bar{x} < 44,000) = P\left( z < \frac{44,000 - 43,000}{\frac{6000}{\sqrt{180}}} \right) = P(z < 2.24) = 0.9875$. 

9. A “Gallop” poll of 1100 randomly selected Americans found that 47.3 percent think the Elian Gonzales story was the top news event of 2000.

(a) (5 pts) Conduct an hypothesis test to determine if the true population proportion of Americans who think the Elian Gonzales story was the top news event of 2000 is less than .50. Report the P-value of your test and conclusion at a level of significance of $\alpha = .05$.

(b) (2 pts) Based on this information, would you be comfortable saying that less than 50 percent of Americans think the Elian Gonzalez story was the top news event in 2000. Explain.

(c) (3 pts) What sample size would be needed so that you could estimate the true proportion in a population within ±.02 with 99 percent confidence?

**Answer.** (a) $H_0 : p = 0.5$ and $H_1 : p < 0.5$. This is a left-tailed test at a level of significance of 0.05. We compute the sample statistic

$$ z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{0.473 - 0.5}{\sqrt{(0.5)(0.5)/n}} = -1.79. $$

The P-value is $P(z < -1.79) = 0.0367$. Because the P-value is less than $\alpha$ we reject the null hypothesis.

(b) Yes, based on the data we are about 96% sure that the population percentage is less than 50 percent.

(c) If we do not assume an initial estimate for $p$, we use the formula $n = \frac{1}{4} \left(\frac{z_c}{E}\right)^2 = \frac{1}{4} \left(\frac{2.58}{.02}\right)^2 = 4160.25$, therefore, we go to the next larger whole number and select a sample size of $n = 4161$. (If you assume an initial estimate of .473 for $p$, you will find a slightly smaller sample size).

10. In a 1993 survey of 150 Education graduates and 200 Social Science graduates, the following data were obtained for their average starting salaries.

<table>
<thead>
<tr>
<th>Major</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>24,500</td>
<td>2600</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>21,500</td>
<td>2400</td>
</tr>
</tbody>
</table>

(a) (6pts) Find a 99% confidence interval for the difference the means $\mu_1 - \mu_2$. Assume the standard deviations given are the population standard deviations.

(b) (2 pts) Explain in ordinary language what the interval in (a) means.

(c) (2 pts) Would you be comfortable claiming that the mean starting salary for education graduates in 1993 was at least $2000 more the mean staring salary for social science graduates?

**Answer.** (a) First, $E = z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 2.58 \sqrt{\frac{2600^2}{150} + \frac{2400^2}{200}} \approx 701.2$ The confidence interval has endpoints $(\bar{x}_1 - \bar{x}_2) \pm E$, and so

$$ 2298.80 < \mu_1 - \mu_2 < 3701.20 \quad \text{with 99\% confidence.} $$

(b) In 1993, the average starting salary for Education graduates was from $2298.80 to $3701.20 more than the average starting salary for Social Science graduates, and we are 99% confident of this.

(c) Yes, I would, am at least 99% sure that the mean starting salaries for Education graduates was at least $2298.80 more than the mean starting salaries of Social Science graduates in 1993.
11. (a) (2 pts) Find the value $z_c$ needed in the formula for constructing a 94% confidence interval from a large sample.

(b) (2 pts) What assumptions must be made on the population and/or sample when constructing a confidence interval for the mean using a large sample?

(c) (2 pts) If a $t$-value of $-2.3$ is reported for a two-tailed hypothesis test at level of significance $\alpha = .05$ on a mean using a sample of size $n = 18$, should the null hypothesis be rejected? Explain.

(d) (2 pts) What assumptions must be made on the population and/or sample to do a hypothesis test on a proportion?

(e) (2 pts) A right-tailed test on a proportion computed a test statistic of $z = 1.37$. What is the $P$-value for this test?

Answer. (a) $z_c = 1.88$ (look on normal table corresponding to an area to the left of $.5 + .94/2 = .97$).

(b) The sample size is large, usually $n \geq 30$. If $\sigma$ is known, we use the normal distribution, i.e. $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, if $\sigma$ is not known, we use the $t$-distribution with $n - 1$ degrees of freedom, and the test statistic is $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$.

(c) Use $d.f. = 17$, and find $.020 < P\text{-value} < .050$ (use two tailed area on chart). Because the $P$-value is less than $\alpha$, we should reject the null hypothesis.

(d) The number of successes and the number of failures in the sample are both larger than 5, which provides the estimate $np > 5$ and $nq > 5$.

(e) $P\text{-value} = 1 - .9147 = .0853$.

12. In a sample of 519 judges it was found that 285 were introverts.

(a) Let $p$ represent the proportion of all judges who are introverts. Find a point estimate for $p$.

(b) Find a 99% confidence interval for $p$, and explain what it means.

(c) Do you think $np > 5$ and $nq > 5$ for this problem? Why, and why is it important?

Answer. (a) $\hat{p} = \frac{285}{519} \approx .5491$.

(b) $z_c = 2.58$, $n = 519$, $\hat{p} = .5491$, $\hat{q} = .4509$ and so $E = 2.58\sqrt{\frac{(.549)(.451)}{519}} \approx .05635$, and so $.4927 < p < .6054$ with 99% confidence.

This interval means that we are 99% sure that between 49.27% and 60.54% of all judges are introverts.

(c) Yes, because there are far more than 5 successes (285) and far more than 5 failures (234) in the sample. This is important, because the confidence interval in (b) relies on the fact that the distribution is approximately normal.
13. (10 pts) (a) Suppose that in an attempt to determine whether a die is fair, you tossed it 120 times
with the following results:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>24</td>
<td>18</td>
<td>15</td>
<td>26</td>
<td>21</td>
<td>16</td>
</tr>
</tbody>
</table>

Does this provide sufficient evidence at a level of significance of $\alpha = .05$ to conclude that the die is not fair? Be sure to find the test statistic, P-value, and state the conclusion.

(b) Explain why the random sample was sufficiently large to do a goodness of fit test.

**Answer.** (a) We test $H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$ versus $H_1 : \text{not all of the proportions are } \frac{1}{6}$.

Since there are 120 tosses, $E = \frac{1}{6}(120) = 20$ in each category, and we use the formula

$$\chi^2 = \sum \frac{(0 - E)^2}{E},$$

where there are $n - 1 = 6 - 1 = 5$ degrees of freedom. The sample test statistic is

$$\chi^2 = \frac{(24 - 20)^2 + (18 - 20)^2 + (15 - 20)^2 + (26 - 20)^2 + (21 - 20)^2 + (16 - 20)^2}{20}$$

$$= \frac{98}{20} = 4.9.$$

Now with $d.f. = 5$ on the $\chi^2$ distribution, we see $1.61 < 4.9 < 9.24$, which means $.10 < \text{P-value} < .90$. Because the P-value $>.05$ we do not reject the null hypothesis, that is, there is not sufficient evidence to conclude that the die is not fair.

(b) This is because the expected number in each category is 20 which is larger than 5.