Q1. (10 points)
A sample of 3 items is selected at random from a box containing 20 items of which 4 are defective. Find the probability to have only one defective in the sample.

Denote $E_i (i = 1, 2, 3)$: the event that the *i*-th item in the sample is defective. We have:

$$P = P\left(E_1\bar{E}_2\bar{E}_3\right) + P\left(\bar{E}_1E_2\bar{E}_3\right) + P\left(\bar{E}_1\bar{E}_2E_3\right)$$
$$= P\left(\bar{E}_2\bar{E}_3|E_1\right)P\left(E_1\right) + P\left(E_2\bar{E}_3|\bar{E}_1\right)P\left(\bar{E}_1\right) + P\left(\bar{E}_2E_3|\bar{E}_1\right)P\left(\bar{E}_1\right)$$
$$= \frac{16}{19}*\frac{15}{18}*\frac{4}{20} + \frac{4}{19}*\frac{15}{18}*\frac{16}{20} + \frac{15}{19}*\frac{4}{18}*\frac{16}{20} = 3*\frac{16}{19}*\frac{15}{18}*\frac{4}{20} = \frac{8}{19}$$
$$= 0.421$$

The result can also be found by: $\dfrac{C_2^{16}C_1^4}{C_3^{20}} = \dfrac{8}{19} = 0.421$

Q2. (10 points)
Diskettes produced by a company will be defective with probability 0.01, independently of each other. The company sells the diskettes in the package of size 10 and offers a money-back guarantee if there exists any defective in the package. If you buy three packages, what will be the probability that you will return exactly one of them?

Prob. to have no defective in one package: 0.99^{10}

Prob. to have defective in one package: $p = 1 - 0.99^{10} = 0.0956$

The probability of interest: $C_1^3 p\left(1-p\right)^2 = 0.235$

Q3. (10 points)
The expected number of typographical errors on a page of a certain book is 0.01. What is the probability to find 2 or more typographical errors when you read one page of that book.

$$P\left(X \geq 2\right) = 1 - P\left(X = 0\right) - P\left(X = 1\right)$$
$$= 1 - \left(1+\lambda\right)e^{-\lambda} = 0.00005$$

Q4. (10 points)
Find the probability mass function of X, given the cumulative distribution function:

$$F(b)=\begin{cases} 0 & b<0 \\ \frac{1}{2} & 0\le b<1 \\ \frac{3}{5} & 1\le b<2 \\ \frac{4}{5} & 2\le b<3 \\ \frac{9}{10} & 3\le b<5 \\ 1 & b\ge 5 \end{cases}$$

$$p(0)=\frac{1}{2};p(1)=\frac{1}{10};p(2)=\frac{1}{5};p(3)=\frac{1}{10};p(5)=\frac{1}{10}$$

Q5. (10 points)
The random variable X has probability density function

$$f(x)=\begin{cases} cx^4 & 0<x<2 \\ 0 & otherwise \end{cases}$$

Find $c,\ E[X]$.

$$\int_0^2 cx^4dx=1 \qquad \Rightarrow c=\frac{5}{32};\ E[X]=\int_0^2 xcx^4dx=\int_0^2 cx^5dx=c\left.\frac{x^6}{6}\right|_0^2=\frac{32c}{3}=\frac{5}{3}$$

Q6. (10 points)
Life tests performed on a sample of 13 batteries of a new model indicated: (1) an average life of 75 months, and (2) a standard deviation of 5 months. Other battery models, produced by similar processes, have normally distributed life spans. Construct the 90% confidence interval for the population mean life of the new model

72.53 to 77.47

Q7. (10 points)
A study is going to be conducted in which a population mean will be estimated using a 92% confidence interval. The estimate needs to be within 12 of the actual population mean. The population variance is estimated to be around 2500. Determine the necessary sample size.

53

Q8. (10 points)
In performing a hypothesis test where the null hypothesis is that the population mean is 6.9 against the alternative hypothesis that the population mean is not equal to 6.9, a random sample of 16 items is selected. The sample mean is 7.1 and the sample standard deviation is 2.4. It can be assumed that the population is normally distributed. The level of significance is selected as 0.05

a. Find the table "t" value for this problem
b. Find the computed "t" value for this problem
c. How about the decision for this problem

a. 2.131   b. 0.33   c. Do not reject the null hypothesis

Q9.(10 points)
Consider the two populations with means μ_1, μ_2 and known variances σ_1^2, σ_2^2. Two samples are taken from these populations with sizes n_1, n_2 $(n_1, n_2 \geq 30)$. Let develop the formula to construct the confidence interval of $\mu_1 - 2\mu_2$ (*Hint*: The variance of $2\bar{x}_2$ is $4\sigma_{\bar{x}_2}^2$).

$$(\bar{x}_1 - 2\bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{4\sigma_2^2}{n_2}} \leq 2\mu_1 - \mu_2 \leq (\bar{x}_1 - 2\bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{4\sigma_2^2}{n_2}}$$

Q10.(10 points)
Sales Director of a specialty e-tailer that operates 87 catalog Web sites on the Internet. feels that the style (color scheme, graphics, fonts, etc.) of a Web site may affect its sales. He chooses three levels of design style (neon, old world, and sophisticated) and randomly assigns six catalog Web sites to each design style. Analysis of data yielded the following ANOVA table

Source of Variation	SS
Treatment	384.3333
Error	1359.667
Total	1744

a. Determine the total degrees of freedom, treatment degrees of freedom and error degrees of freedom.
b. Determine $MS_{Treatments}, MS_E$.
c. Using $\alpha = 0.05$, how about the appropriate decision?

a. 17, 2, 15
b. 192.1667   90.64444
c. $F = 2.12; F_{.05,2,15} = 3.68$ $\Rightarrow$ Do not reject the null hypothesis

Q1. (10 points)
A box contains 24 items, of which four are defective.

a. If we select four items from the box, without replacement, what is the probability that all four items will be defective?
b. If we select four items from the box, with replacement, what is the probability that all four items will be defective?

$$P_1 = \frac{4}{24} * \frac{3}{23} * \frac{2}{22} * \frac{1}{21} = 0.000094$$

b.
$$P_2 = \left(\frac{4}{24}\right)^4 = 0.00077$$

Q2. (10 points)
An electronic system contains 10 independent components. The probability that each component will fail is 0.1. Given that at least one of the components has failed, find the probability that at least two of the components have failed.

Denote: X: number of failed components among 10 components

E_1: the event that at least one component has failed

E_2: the event that at least two components have failed

We have:

$$P(E_2|E_1) = \frac{P(E_1E_2)}{P(E_1)} = \frac{P(E_2)}{P(E_1)}$$

$$P(E_1) = P(X \geq 1) = 1 - C_0^{10} p^0 (1-p)^{10} = 1 - 0.9^{10} = 0.651$$

$$P(E_2) = P(X \geq 2) = 1 - C_0^{10} p^0 (1-p)^{10} - C_1^{10} p^1 (1-p)^9 = 1 - 0.9^{10} - 10 * 0.1 * 0.9^9 = 0.264$$

$$\Rightarrow \quad P(E_2|E_1) = 0.405$$

Q3. (10 points)
An airline usually sells 200 tickets for a flight on an airplane that has only 198 seats due to the fact that, on average, 1% of passengers who booked the tickets will not appear for the departure of their flight. For a certain flight, find the probability that all passengers who appear for the departure will have a seat.

Denote X: number of passengers who appear for a certain flight

X follows Binomial distribution with $n = 200, p = 0.99$

The probability of interest is:

$$P(X \leq 198) = 1 - P(X = 199) - P(X = 200)$$

$$= 1 - \frac{200!}{199! * 1!} 0.99^{199} * 0.01^1 - \frac{200!}{200! * 0!} 0.99^{200} * 0.01^0$$

$$= 0.595$$

Q4. (10 points)
A medical test is developed for detecting a new type of cancer. If the test is applied for a person who has this type of cancer, the probability to have a positive reaction is 0.95. However, the test also gives a positive reaction with the probability of 0.1 when applying for a person who does not have this type of cancer. Given that the test is positive for a randomly selected person, find the probability that he/she does not have this type of cancer.

Denote: P: The test gives a positive reaction
N: The test gives a negative reaction
E: The selected person has the new type of cancer

We have:

$$P(\bar{E}|P)=\frac{P(\bar{E}P)}{P(P)}=\frac{P(P|\bar{E})P(\bar{E})}{P(P|\bar{E})P(\bar{E})+P(P|E)P(E)}$$

$$=\frac{0.1*P(\bar{E})}{0.1*P(\bar{E})+0.95*P(E)}=\frac{0.1(1-p)}{0.1(1-p)+0.95p}=\frac{1-p}{1+8.5p}$$

In which p is the proportion of people who has the new type of cancer in the population.

Q5. (10 points)
The random variable X has probability density function

$$f(x)=\begin{cases} ce^{-2x} & x>0 \\ 0 & \text{otherwise} \end{cases}$$

Find c and determine $P\{1\le X\le 3\}$.

a. $$\int_0^\infty ce^{-2x}dx=1 \quad \Rightarrow c=2$$

b $$P\{1\le X\le 3\}=\int_1^3 2e^{-2x}dx=0.133$$

Q6. (10 points)
A rod shearing process produces rods with a mean length of 120 inches when "in control". Periodically, quality control inspector selects a random sample of 36 rods. If the mean length of sampled rods is too long or too short, the shearing process is shut down. The last sample showed a mean and standard deviation of 120.5 and 1.2 inches, respectively. Using $\alpha = 0.05$

a. Construct the confidence interval for the mean length of rods.
b. Conduct the test on the null hypothesis that the mean length of rods does not change recently. What will be the appropriate decision?
c. Determine the observed significant level for the test conducted above.

a. $n = 36$ $\quad \bar{x} = 120.5$ $\quad s = 1.2$

Confidence interval:

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$120.5 - 1.96\frac{1.2}{\sqrt{36}} \le \mu \le 120.5 - 1.96\frac{1.2}{\sqrt{36}}$$

$$120.108 \le \mu \le 120.892$$

b. $z = \dfrac{\bar{x} - 120}{s/\sqrt{n}} = 2.5 > z_{\alpha/2} = 1.96$: reject the null hypothesis that $\mu = 120$

c. Observed significant level: $p = 2*0.00621 = 0.0124$

Q7. (10 points)

Consider a random sample taken from a normal distribution for which the mean μ is unknown and the variance σ^2 is known. How large the sample should be taken in order that the confidence interval for μ with confidence coefficient 0.95 has the width less than 0.01σ.

$$n \ge \left(\frac{2z_{\alpha/2}\sigma}{W}\right)^2 = \left(\frac{2*1.96}{0.01}\right)^2 = 153{,}664$$

Q8. (10 points)

A certain drug "A" is administered to eight patients selected at random, and after a fixed time period, the concentration of drug in body cells of each patient was measured in appropriate units. The results were as follows:

1.23 1.42 1.41 1.62 1.55 1.51 1.60 1.76

The second drug "B" was also tested for another group of six randomly selected patients and the results were:

1.76 1.41 1.89 1.49 1.67 1.81

Assuming that all the observations have a normal distribution with a common variance, let test the hypothesis that the mean concentration of drug "A" among all patients is *at least* as large as the mean concentration of drug "B" (using $\alpha = 0.01$).

$H_0:\ \mu_A = \mu_B$

$H_0:\ \mu_A < \mu_B$

We have:

$\bar{x}_A = 1.5125$ $\quad s_A = 0.1607$

$\bar{x}_B = 1.6717$ $\quad s_A = 0.1877$

Pooled standard deviation:

$$S = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} = \sqrt{\frac{7*0.1607^2 + 5*0.1877^2}{12}} = 0.1725$$

Test statistic:

$$t = \frac{\overline{x}_A - \overline{x}_B}{S\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{1.5125 - 1.6717}{0.1725\sqrt{\frac{1}{8} + \frac{1}{6}}} = -1.709$$

Rejection region: $t < -t_{0.01,12} = -2.681$: Cannot reject the null hypothesis

Q9.(10 points)
Answer and give explanations for the following questions

a. A null hypothesis was rejected at the 0.10 level of significance. If the level of significance was changed to 0.05 and the same sample results were obtained, what decision should be made?
b. In a two-tailed test, the null hypothesis was rejected at the 0.10 level of significance. If the test is changed to one-tailed test at the same significance level, what will be the new decision?

a. The decision might be: "Reject" or "Not Reject" the null hypothesis
b. The decision is still "Reject"

Q10.(10 points)
Data from a completely randomized design are shown in the following table.

Treatment Level		
1	2	3
27	26	27
26	22	29
23	21	27
24	23	26

($SS_{Treatment} = 36.167,\ SS_T = 64.917$)

For a one-way ANOVA using $\alpha = 0.05$, what will be the appropriate decision?

Degrees of freedom: $df_{Treatment} = a - 1 = 2$, $df_E = N - a = 9$
Hence:

$$MS_{Treatment} = \frac{36.167}{2} = 18.084, \qquad MS_E = \frac{64.917 - 36.167}{9} = 3.194$$

$$\Rightarrow \quad F_0 = \frac{MS_{Treatment}}{MS_E} = 5.66$$

Critical value: $F_{0.05,2,9} = 4.26$: $\Rightarrow$ Treatment means are not equal.

Q1. (10 points) Toss a die two times. Find the probabilities that

a. The sum of the two numbers that appear will be odd
b. The difference between the two numbers that appear will be less than 3

a. $P_1 = \frac{18}{36} = 0.5$

b. $P_2 = \frac{24}{36} = \frac{2}{3} = 0.667$

Q2. (10 points)

Three letters are placed at random in three envelopes prepared for them.

a. Find the probability that exactly one letter will be placed in the correct envelop.
b. Find the probability that no letter will be placed in the correct envelop.

Denote: $E_i\ (i = 1,2,3)$: the event that the i^{th} letter is placed in the correct envelop

a. We have: $P\left(E_1\overline{E}_2\overline{E}_3\right) = P\left(\overline{E}_2\overline{E}_3 | E_1\right)P\left(E_1\right) = \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$

Similarly, $P\left(\overline{E}_1E_2\overline{E}_3\right) = P\left(\overline{E}_1\overline{E}_2E_3\right) = \frac{1}{6}$

Therefore, the probability that exactly one letter is placed in the correct envelop is:

$$P\left(E_1\overline{E}_2\overline{E}_3\right) + P\left(\overline{E}_1E_2\overline{E}_3\right) + P\left(\overline{E}_1\overline{E}_2E_3\right) = \frac{1}{2}$$

b. We have :

$$P\left(E_1E_2E_3\right) = P\left(E_2E_3 | E_1\right)P\left(E_1\right) = \frac{1}{2} * \frac{1}{3} = \frac{1}{6}$$

So, the probability that no letter will be placed in the correct envelop is:

$$1 - \left\{P\left(E_1\overline{E}_2\overline{E}_3\right) + P\left(\overline{E}_1E_2\overline{E}_3\right) + P\left(\overline{E}_1\overline{E}_2E_3\right)\right\} - P\left(E_1E_2E_3\right) = \frac{1}{3}$$

Q3. (10 points)

A manufacturing plant has two production lines A and B. In order to control the quality of the production process, two samples should be taken every day for inspection. On a given day, the quality manager of the plant is equally likely to choose either line A or line B for the first sample. However, after the first sample, the probability that the quality manager will take the second sample on the same production line will be $1/3$ and the probability is $2/3$ that he will switch to the other line.

a. Find the probability that both samples of the day are from line A
b. Find the probability that among the two samples of the day, one is from line A and the other is from line B.

Denote A: the sample is taken from line A
B: the sample is taken from line B

a. $P(AA) = P(AA|A)P(A) + P(AA|B)P(B) = \frac{1}{3}*\frac{1}{2} + 0*\frac{1}{2} = \frac{1}{6} = 0.167$

b. $P(A \text{ and } B) = P(A \text{ and } B|A)P(A) + P(A \text{ and } B|B)P(B) = \frac{2}{3}*\frac{1}{2} + \frac{2}{3}*\frac{1}{2} = \frac{2}{3} = 0.667$

Q4. (10 points)

Three different machines M_1, M_2, and M_3 were used for producing a large batch of similar items. Suppose that 20% of the items were produced by machine M_1, 30% by machine M_2, and 50% by machine M_3. It is also known that 1%, 2%, and 3% of the items produced by machines M_1, M_2, and M_3 are defective.

a. Randomly select one item from the batch, find the probability that the item is defective.
b. One item is randomly select from the batch and it is known that the posterior probability that the item was produced by machine M_i $(i = 1, 2, \text{ or } 3)$ is larger than the prior probability that the item was produced by machine M_i $(i = 1, 2, \text{ or } 3)$. Find the machine M_i.

$$P(D) = P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)$$
$$= 0.2*0.01 + 0.3*0.02 + 0.5*0.03 = 0.023$$

.

b.

$$P(M_1|D) = \frac{P(D|M_1)P(M_1)}{P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)} = \frac{2}{23} < 0.2 = P(M_1)$$

$$P(M_2|D) = \frac{P(D|M_2)P(M_2)}{P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)} = \frac{6}{23} < 0.3 = P(M_2)$$

$$P(M_3|D) = \frac{P(D|M_3)P(M_3)}{P(D|M_1)P(M_1) + P(D|M_2)P(M_2) + P(D|M_3)P(M_3)} = \frac{15}{23} > 0.5 = P(M_3)$$

$\Rightarrow$ Machine M_3.

Q5. (10 points)
The random variable *X* has probability density function

$$f(x) = ce^{-|x|} \qquad -\infty < x < \infty$$

Find c and determine x_0 such that $F(x_0) = 0.9$.

a. $$\int_{-\infty}^{\infty} ce^{-|x|}dx = 1 \quad \Rightarrow c = \frac{1}{2}$$

b Note that

$$F(x) = \begin{cases} \frac{1}{2}e^{x} & x < 0 \\ 1 - \frac{1}{2}e^{-x} & x \geq 0 \end{cases}$$

Therefore $x_0 = -\ln(0.2) = 1.6094$

Q6. (10 points)
An airport limousine can accommodate up to four passengers on one trip. The company will accept a maximum of six reservations for a trip, and a passenger must have a reservation. From previous records, 20% of the passengers who make reservation do not appear for the trip.

a. If six reservations are made, what is the probability that at least one passenger cannot be accommodated on the trip?
b. If six reservations are made, what is the expected number of available places when the limousine departs?

Let *X* denote the number of passengers among the six reservation who do not appear.
X follows the binomial distribution $B(6, 0.2)$

a. The probability that at least one passenger cannot be accommodated is:

$$P\{X \leq 1\} = C_0^6 (0.2)^0 (0.8)^6 + C_1^6 (0.2)^1 (0.8)^5 = 0.6554$$

b. It is noted that the limousine departs only when $X \leq 5$: $P(X \leq 5) = 0.9999$

The available places: 0	when $X \leq 2$:	$P(X \leq 2) = 0.9011$
1	when $X = 3$:	$P(X = 3) = 0.0819$
2	when $X = 4$:	$P(X = 4) = 0.0154$
3	when $X = 5$:	$P(X = 5) = 0.0015$

So, the expected number of places available when the limousine departs is:

$$\frac{0 * P\{X \leq 2\} + 1 * P\{X = 3\} + 2 * P\{X = 4\} + 3 * P\{X = 5\}}{P\{X \leq 5\}} = 0.1173$$

Q7. (20 points)
Consider the following test related to the average lifetime of a certain type of tire.

$$H_0: \quad \mu = \mu_0 = 20000$$
$$H_a: \quad \mu > \mu_0 = 20000$$

The sample size is $n = 16$ and the standard deviation of the population is $\sigma = 1500$. It is assumed that the lifetime of the tire follows normal distribution and hence, the test can be conducted based on *normal distribution* (i.e., z-test can be employed).

a. If $\bar{x} = 20960$ and the significance level $\alpha = 0.01$ is used, what is the decision? How about the observed significant level of the test in this situation?
b. It is noted that the β error (type II error) of the above right-tailed test is defined as the probability that we cannot reject H_0 given that H_0 is not correct. Assume that the actual average lifetime of the tire is $\mu_1 > \mu_0$. Let prove that

$$\beta = \Phi\left(z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$$

In which $\Phi(.)$ is the cumulative distribution function of the standardized normal distribution.

c. Find β when $\mu_1 = 21000$.

a. $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{20960 - 20000}{1500/\sqrt{16}} = 2.56 > z_{0.01} = 2.33$$: reject the null hypothesis

Observed significant level: $p = 0.0052$

b. By definition, we have:

$$\beta = P\left\{\bar{x} < \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right\} = P\left\{\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right\}$$

$$= P\left\{z < z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right\} = \Phi\left\{z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right\}$$

c. $$\beta = \Phi\left\{2.33 + \frac{20000 - 21000}{1500/\sqrt{16}}\right\} = \Phi\{-0.3367\} = 0.368$$

Q8. (20 points)

The maximum concentration of a substance in the flow of water coming out of the treatment plant of a manufacturer located in a city is 1.2 mg/l. Periodically, a quality inspector from the environmental protection department of the city will come to check whether the above manufacturer conforms to the quality standard or not. Considering the following two establishments of the testing hypotheses:

(A): $H_0: \quad \mu = 1.2\,\text{mg/l}$; $H_a: \quad \mu > 1.2\,\text{mg/l}$

(B): $H_0: \quad \mu = 1.2\,\text{mg/l}$; $H_a: \quad \mu < 1.2\,\text{mg/l}$

a. If you are the quality inspector, which of the above establishments will be selected? Explain why?
b. If you are the production manager of the above manufacturer, which of the above establishments will you prefer? Explain why?

a. Select (A): because the quality inspector usually suspect that the quality standard is violated.

b. It depends:

- If you believe that your treatment plant still work well: select (B)
- Otherwise, select (A): because reducing the significance level might help to "not reject" the null hypothesis

Q1. (10 points)
Urn 1 contains 3 red and 3 back balls, whereas urn 2 contains 4 red and 6 black balls. If a ball is randomly selected from each urn, what is the probability that the balls will be the same color?

$$
\begin{aligned}
P(R_1R_2 \cup B_1B_2) &= P(R_1R_2)+P(B_1B_2) \\
&= P(R_1)P(R_2)+P(B_1)P(B_2) \\
&= \frac{1}{2}*\frac{2}{5}+\frac{1}{2}*\frac{3}{5}=\frac{1}{2}
\end{aligned}
$$

Q2. (10 points)
Each product produced at factory A is defective with probability 0.05, whereas each one produced at factory B is defective with probability 0.01. A sample of 2 products is randomly selected. This sample is taken from the same factory, which is equally likely to be either factory A or factory B. If the first product is defective, what will be the conditional probability that the second product is also defective?

$$
\begin{aligned}
P(D_2|D_1) &= \frac{P(D_2D_1)}{P(D_1)}=\frac{P(D_2D_1|A)P(A)+P(D_2D_1|B)P(B)}{P(D_1|A)P(A)+P(D_1|B)P(B)} \\
&= \frac{(0.05)^2*0.5+(0.01)^2*0.5}{0.05*0.5+0.01*0.5}=\frac{13}{300}=0.043
\end{aligned}
$$

Q3. (10 points)
Two students A and B are both registered for a certain course. Assume that student A attends class 80% of the time and student B attends class 60% of the time.

a. Find the probability that at least one of the two students will be in class on a given day.
b. If at least one of the two students is in class on a given day, what is the probability that A is in class that day?

Denote X: At least one of the students is in class

a. $$
\begin{aligned}
P(X) &= P(A\cup B)=P(A)+P(B)-P(AB) \\
&= 0.8+0.6-0.8*0.6=0.92
\end{aligned}
$$

b. $$P(A|X)=\frac{P(AX)}{P(X)}=\frac{P(A)}{P(X)}=\frac{0.8}{0.92}=0.870$$

Q4. (10 points)

Suppose that 30% of the bottles produced in a certain plant are defective. If a bottle is defective, the probability is 0.9 that an inspector will notice it and remove it from the filling line. If a bottle is not defective, the probability is 0.2 that the inspector will think that it is defective and remove it from the filling line.

a. If a bottle is removed from the filling line, what is the probability that it is defective?

b. If the customer buys a bottle that has not been remove from the filling line, what is the probability that it is defective?

Denote D: The bottle is defective

$\bar{D}$: The bottle is not defective

R: The bottle is removed

$\bar{R}$: The bottle is not removed

a. $$P(D|R) = \frac{P(R|D)P(D)}{P(R)}$$

$$= \frac{P(R|D)P(D)}{P(R|D)P(D) + P(R|\bar{D})P(\bar{D})}$$

$$= \frac{0.9 * 0.3}{0.9 * 0.3 + 0.2 * 0.7} = 0.659$$

b. $$P(D|\bar{R}) = \frac{P(\bar{R}|D)P(D)}{P(\bar{R})}$$

$$= \frac{P(\bar{R}|D)P(D)}{P(\bar{R}|D)P(D) + P(\bar{R}|\bar{D})P(\bar{D})}$$

$$= \frac{0.1 * 0.3}{0.1 * 0.3 + 0.8 * 0.7} = 0.051$$

Q5. (10 points)

Does there exist any number c such that the following function would be a probability mass function

$$p(x) = \frac{c}{x} \qquad x = 1, 2, \ldots$$

Explain why?

The value of c does not exist due to the fact that $\sum_{x=1}^{\infty} p(x) = \sum_{x=1}^{\infty} \frac{c}{x} > \frac{c}{[c]} \geq 1$

Q6. (10 points)

The lifetimes of interactive computer chips produced by a certain semiconductor manufacturer are normally distributed with $\mu = 1.4*10^6$ hours and $\sigma = 3*10^5$ hours.

a. What is the probability that the lifetime of a randomly selected chip will be less than $1.8*10^6$ hours?

b. If a batch of 100 chips is selected, find the probability that there exist at least 20 chips whose lifetimes are less than $1.8*10^6$ hours.

Denote X: Lifetime of the chips

N: Number of chips with lifetime less than $1.8*10^6 h$ in the batch

a. The probability that the lifetime of the chip is less than $1.8*10^6 h$:

$$p = P\left\{X \le 1.8*10^6\right\} = P\left\{Z \le \frac{1.8*10^6 - 1.4*10^6}{3*10^5} = \frac{4}{3}\right\} = 0.909$$

b. It is noted that N follow binomial distribution with parameters $n = 100$ and $p = 0.909$

Hence, the probability of interest is: $P = P(N \ge 20)$

Approximate the binomial distribution by the normal distribution with parameters:

$$\mu = np = 90.9 \qquad \sigma = \sqrt{npq} = 2.876$$

We have:

$$P \approx P\left(N\left(\mu, \sigma^2\right) \ge 19.5\right) = P\left(Z \ge \frac{19.5 - 90.9}{2.876}\right) = P\left(Z \ge -24.83\right) = 1$$

Q7. (20 points)

The tensile strength of a fiber used in manufacturing cloth is of interest to the purchaser. Previous experience indicates that the standard deviation of tensile strength is 2 *psi*. A random sample of eight fiber specimens is selected, and the average tensile strength is found to be 127 *psi* . Assume that the tensile strength follows normal distribution, and hence, the z-test can be employed in this case.

a. Test the hypothesis that the mean tensile strength equals 125 *psi* versus the alternative that the mean exceeds 125 *psi*. Use $\alpha = 0.05$?

b. What is the *p*-value for this test?

c. Discuss why a one-sided test was used in part (a).

a. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{127 - 125}{2/\sqrt{8}} = 2.828 > z_{0.05} = 1.645$: reject the null hypothesis. It means that the tensile strength is higher than the nominal value 125 *psi*

b. Observed significant level: $p = 0.0023$

c. Because the higher tensile strength is preferred

Q8. (20 points)

Consider the following two-sided test on the mean value:

$$H_0: \quad \mu = \mu_0$$
$$H_a: \quad \mu \neq \mu_0$$

Suppose that the z-test can be employed, and the sample average $\bar{x}$ is known to be greater than μ_0. If an analyst really wants to reject the null hypothesis, he/she will have two options:

1. Increase the normal value of α from 0.05 to 0.1
2. Keep α at 0.05, but convert to one-sided test

a. In your opinion, if the second option is chosen, the left-sided test or the right-sided test should be used? Explain why?

The right-sided test should be used because there is no way to reject the null hypothesis with the use of left-sided test when $\bar{x} > \mu_0$.

b. Which option should be selected?

The probabilities of Type I error are the same for both options (0.05). The probabilities of Type II error are also the same (noted that $\bar{x} > \mu_0$). Hence, the two options are equivalent.

Q1. (10 points)
Shortly after being put into service, some buses manufactured by a company have developed cracks on the underside of the main frame. Suppose a particular city has 20 of these buses, and cracks have actually appeared in 8 of them.

a. How many ways are there to select a sample of 5 buses which contain exactly 4 with visible cracks?

a. $C_1^{12} * C_4^8 = 840$

b. If a sample of 5 buses is chosen at random. What is the probability that exactly 4 of the 5 will have visible cracks?

b. $\dfrac{C_1^{12} * C_4^8}{C_5^{20}} = \dfrac{840}{15504} = 0.054$

Q2. (10 points)
One box contains 6 red balls and 4 green balls. Another box contains 7 red balls and 3 green balls. A ball is randomly chosen from the first box and placed in the second box. Then a ball is randomly selected form the second box and placed in the first box.

a. What is the probability that "a red ball is selected from the first box and a red ball is selected from the second box"?
b. What is the probability that the numbers of red and green balls in the first box are identical to the numbers at the beginning?

a. $P\{R_1R_2\} = P\{R_2|R_1\}P\{R_1\} = \dfrac{8}{11} * \dfrac{6}{10} = \dfrac{24}{55} = 0.436$

a. $P\{G_1G_2\} = P\{G_2|G_1\}P\{G_1\} = \dfrac{4}{11} * \dfrac{4}{10} = \dfrac{8}{55} = 0.145$

So, the solution is 0.436+0.145 = 0.581

Q3. (10 points)
Four engineers A, B, C, and D have been scheduled for job interviews at a company. The personnel manager has scheduled the four for interview rooms 1, 2, 3, and 4, respectively. However, the manager's secretary does not know this, so assigns them to the four rooms in a completely random fashion. What is the probability that.

a. All four end up in the correct rooms?
b. None of the four ends up in the correct room?

a. The probability that all four end up in the correct rooms: $\dfrac{1}{4!} = \dfrac{1}{24} = 0.0417$

b. The probability that only one ends up in the correct rooms: $\dfrac{C_1^4 * 2}{4!} = \dfrac{1}{3}$

The probability that two end up in the correct rooms: $\frac{C_2^4}{4!} = \frac{1}{4}$

So, the probability that none of the four ends up in the correct room is:

$$1 - \left(\frac{1}{24} + \frac{1}{3} + \frac{1}{4} \right) = \frac{9}{24} = 0.375$$

Q4. (10 points)
Suppose that only 20% of all drivers come to a complete stop at an intersection having flashing red lights in all directions when no other cars are visible. What is the probability that, of 20 randomly chosen drivers coming to an intersection under these conditions

a. At most 6 will come to a complete stop?

$$P(\text{At most } 6) = \sum_{i=0}^{6} C_i^{20} (0.2)^i (0.8)^{20-i} = 0.913$$

b. At least 6 will come to a complete stop?

$$P(\text{At least } 6) = \sum_{i=6}^{\infty} C_i^{20} (0.2)^i (0.8)^{20-i} = 1 - \sum_{i=0}^{5} C_i^{20} (0.2)^i (0.8)^{20-i} = 0.196$$

Q5. (10 points)
A college professor never finishes his lecture before the bell rings to end the period and always finishes his lectures within 2 minutes after the bell rings. Let X denote the time that elapses between the bell and the end of the lecture and suppose that the probability density function of X is

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the value of k
b. What is the probability that the lecture continues beyond the bell for between 60 seconds and 90 seconds?

a. $k = \frac{3}{8} = 0.375$

b. The probability of interest is: $\int_1^{1.5} f(x)dx = 0.297$

Q6. (10 points)
Suppose that only 40% of all drivers regularly wear a seatbelt. A random sample of 500 drivers is selected. What is the probability that.

a. Between 180 and 230 (inclusive) of the drivers regularly wear a seatbelt?
b. Fewer than 175 of those in the sample regularly wear the seatbelt?

Let X be the number of drivers in the sample that regularly wear a seatbelt. Then X follows binomial distribution with parameters $n = 500$, $p = 0.4$. If normal approximation is used then: $\mu = np = 200$, $\sigma = \sqrt{npq} = \sqrt{120} = 10.95$

a. $$P\{180 \le X \le 230\} = P\{179.5 \le N(\mu,\sigma^2) \le 230.5\}$$
$$= P\{-1.87 \le Z \le 2.78\}$$
$$= \Phi(2.78) - \Phi(-1.87) = 0.997 - 0.030 = 0.967$$

b. $$P\{X < 175\} = P\{N(\mu,\sigma^2) \le 174.5\}$$
$$= P\{Z < -2.33\} = \Phi(-2.33) = 0.010$$

Q7. (20 points)

a. A sample of 50 lenses used in the eyeglasses yields a sample mean thickness of 3.05 mm and a sample standard deviation 0f 0.34 mm. The desired true average thickness of such lenses is 3.20 mm. Does the data strongly suggest that the true average thickness of such lenses is something other than what is desired? Test using $\alpha = 0.05$
b. In the above problem, suppose that the experimenter had believed before collecting the data that the value of σ was approximately 0.30 mm. If the experimenter wished the probability of a type II error, i.e., β, to be 0.05 when $\mu = 3.00$, was a sample size 50 unnecessarily large?

a. $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \approx \dfrac{3.05 - 3.20}{0.34/\sqrt{50}} = -3.12$: reject the null hypothesis because $z_{0.025} = 1.96$.

b. We have

$$\beta = P\{\text{Not reject } \mathrm{H}_0 | \mu = 3.00\}$$
$$= P\left\{\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} | \mu = 3.00\right\}$$
$$= P\left\{\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \le z \le \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\}$$
$$= P\{2.754 \le z \le 6.674\} = \Phi(6.674) - \Phi(2.754)$$
$$\approx 1 - 0.997 = 0.003$$

So, no need to use sample size of 50 (the sample size of 30 is OK - $\beta = 0.0454$).

Q8. (20 points)

Suppose that the population distribution is normal with known σ. Let γ be such that $0 < \gamma < \alpha$. For testing $H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$, consider the test that reject H_0 if $z < -z_{\alpha-\gamma}$ or $z > z_\gamma$, where the test statistic is $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

a. What is the probability of type I error for this test?

b. If the actual mean value of the population is $\mu' \neq \mu_0$, let derive the following expression for the probability of type II error

$$\beta = \Phi\left(z_\gamma + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \right) - \Phi\left(-z_{\alpha-\gamma} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \right)$$

($\Phi(.)$: cumulative distribution function of the standardized normal distribution)

a. Probability of Type I error is α

b. $\beta = P\left\{ \text{Not reject } \mathrm{H}_0 \,|\, \mu = \mu' \right\}$

$$= P\left\{ \mu_0 - z_{\alpha-\gamma} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_\gamma \frac{\sigma}{\sqrt{n}} \,\Big|\, \mu = \mu' \right\}$$

$$= P\left\{ \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} - z_{\alpha-\gamma} \leq z \leq \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} + z_\gamma \right\}$$

$$= \Phi\left(z_\gamma + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \right) - \Phi\left(-z_{\alpha-\gamma} + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}} \right)$$

Q1. (10 points)

a. How many different 7-place license plates are possible if the first 2 places are for letters (A to Z) and the other 5 for numbers (0 to 9)?

b. Repeat part (a) under the assumption that no letter or number can be repeated in a single license plate.

a. $26^2 * 10^5 = 67,600,000$

b. $26*25*10*9*8*7*6 = 19,656,000$

Q2. (10 points)

One box contains 3 red and 3 black balls. Another box contains 4 red and 6 black balls. A ball is randomly selected from each box. Find the probability that the balls will be the same color.

$$P\{R_1R_2\} = P\{R_1\}P\{R_2\} = \frac{3}{6} * \frac{4}{10} = 0.2$$

$$P\{B_1B_2\} = P\{B_1\}P\{B_2\} = \frac{3}{6} * \frac{6}{10} = 0.3$$

So, the solution is 0.2+0.3 = 0.5

Q3. (10 points)

A box contains 24 light bulbs of which four are defective. If one person selects 10 bulbs from the box in a random manner, and a second person then takes the remaining 14 bulbs. What are the probabilities that.

a. All four defective bulbs will be with the first person?

b. At least one defective bulb is with each person?

a. The probability that all four defective bulbs are with the first person

$$\frac{C_4^4 * C_6^{20}}{C_{10}^{24}} = 0.0198$$

b. The probability that all four defective bulbs are with the second person

$$\frac{C_4^4 * C_{10}^{20}}{C_{14}^{24}} = 0.0942$$

So, the probability that at least one defective bulb is with each person is

$$1-(0.0198+0.0942) = 0.8860$$

Q4. (20 points)
When a machine is adjusted properly, 50% of the items produced by it are of high quality and the other 50% are of medium quality. However, when the machine is improperly adjusted, only 25% of the items produced by it are of high quality and the remaining 75% are of medium quality. Suppose that the machine is improperly adjusted during 10% of its working time.

a. Five items produced by the machine at a certain time are selected at random and inspected. What is the probability that four of these items are of high quality and one item is of medium quality?
b. If the situation in (a) occurs, what is the posterior probability that the machine was adjusted properly?

Denote H: A randomly selected item is of high quality
G: The machine is adjusted properly
Y: There are 4 high quality items in the sample

a. We have

$$P(Y) = P(Y|G)P(G) + P(Y|\bar{G})P(\bar{G})$$
$$= \left(C_4^5 * 0.5^4 * 0.5\right) * 0.9 + \left(C_4^5 * 0.25^4 * 0.75\right) * 0.1 = 0.142$$

b. The posterior probability that the machine was adjusted properly

$$P(G|Y) = \frac{P(Y|G)P(G)}{P(Y)} = 0.990$$

Q5. (10 points)
A system can function for a random amount of time X. If the density function of X (in months) is given by:

$$f(x) = \begin{cases} Cxe^{-x/2} & x > 0 \\ 0 & x \le 0 \end{cases}$$

a. Find the value of C
b. Find the probability that the system functions for at least 5 months

a. $\int_0^\infty f(x)dx = 1 \quad \Rightarrow \quad C = 1/4$

b. Therefore,

$$P\{X \ge 5\} = \int_5^\infty \frac{1}{4} x e^{-\frac{x}{2}} dx = \frac{14}{4} e^{-\frac{5}{2}} = 0.287$$

Q6. (10 points)
To determine whether the pipe welds in a nuclear power plant meet specifications, a random sample of welds is selected, and tests are conducted on each weld in the sample. Weld strength is measured as the force required to break the weld. Suppose the specifications state that mean strength of welds should exceed 100 lb/in^2. Consider the following two establishments of the test

(**A**) $H_0 : \mu = 100$ versus $H_a : \mu < 100$

or (**B**) $H_0 : \mu = 100$ versus $H_a : \mu > 100$

a. Describe the meanings of Type I & Type II errors in each case
b. Which one should be used? Explain why?

a. (**A**) $H_0 : \mu = 100$ versus $H_a : \mu < 100$

Type I error: Concluding that the strength of welds is not satisfactory while it is
Type II error: Concluding that the strength of welds is satisfactory while it is not

(**B**) $H_0 : \mu = 100$ versus $H_a : \mu > 100$

Type I error: Concluding that the strength of welds is satisfactory while it is not
Type II error: Concluding that the strength of welds is not satisfactory while it is

b. Select (**B**) because if (**A**) is selected, the consequence of Type II error (if it occurs) is catastrophic (noted also that Type I error is small and controllable while Type II error is usually large and uncontrollable)

Q7. (20 points)

A sample of 16 wires is selected and tested to determine the tensile strength (in $\mathrm{N/mm^2}$). The resulting sample mean and standard deviation are 2160 and 30, respectively.

a. In principle, what hypotheses should be tested to determine whether the mean tensile strength exceed 2150 $\mathrm{N/mm^2}$?

$$H_0 : \mu = 2150 \text{ versus } H_a : \mu > 2150$$

b. Determine the *p*-value of the test. Can we reject the null hypothesis? What do you think about the final conclusion?

We have

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2160 - 2150}{30/\sqrt{16}} = 1.33 \quad \Rightarrow \quad p\text{-value} = 0.102$$

So, we cannot reject the null hypothesis. This means that we have to accept that $\mu \le 2150$ although the sample mean is larger than 2150. It looks strange! But it is good for manufacturer of the wire because this conclusion will request for quality improvement.

Q8. (20 points)

Consider the two-tailed test: $H_0 : \mu = \mu_0$ versus $H_a : \mu \ne \mu_0$ in which $\mu_0 = 100$. Suppose that the standard deviation of the population is $\sigma = 30$ and actual value of population mean is $\mu_1 = 110$.

Let determine the sample size such that Type II error β of the z-test is at most 0.2

We have:

$$\beta = P\left\{\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

$$= P\left\{\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \le \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} \le \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\}$$

$$= P\left\{\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \le \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} \le \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\}$$

$$= P\left\{-\frac{\sqrt{n}}{3} - 1.96 \le z \le -\frac{\sqrt{n}}{3} + 1.96\right\}$$

$$\approx P\left\{z \le -\frac{\sqrt{n}}{3} + 1.96\right\}$$

With $\beta = 0.2$: $-\frac{\sqrt{n}}{3} + 1.96 = -0.84 \quad \Rightarrow n = 71$

2010

Q1. (20 points) A box contains 10 balls of which five are red. Suppose that the balls are drawn from the box one at a time without replacement.

a. Determine the probability that the fist ball drawn will be red.
b. Determine the probability that the last ball drawn will be red.
c. Determine the probability that the fifth ball drawn will be red.

a. 0.5

b. $\frac{C_4^5}{C_9^{10}} = 0.5$

c. $\sum_{i=0}^{4} \frac{C_i^5 C_{4-i}^5}{C_4^{10}} * \frac{5-i}{6} = 0.5$ (in fact, this result can be found immediately by the fact that the events of "red ball" and "non-red ball" are equally likely!)

Q2. (20 points)
A production facility employs 20 workers on the day shift, 15 workers on the swing shift, and 10 workers on the graveyard shift. A quality control consultant is to randomly select 6 of these workers for in-depth interviews

a. Determine the probability that all 6 selected workers will be from the day shift.
b. Determine the probability that all 6 selected workers will be from the same shift.
c. Determine the probability that at least two different shifts will be represented among the selected workers.
d. What is the probability that at least one of the shifts will be unrepresented in the sample of workers?

a. $\dfrac{C_6^{20}}{C_6^{45}}$

b. $\dfrac{C_6^{20}+C_6^{15}+C_6^{10}}{C_6^{45}}$

c. $1-\dfrac{C_6^{20}+C_6^{15}+C_6^{10}}{C_6^{45}}$

d. The probability that two shifts will be represented in the sample is:

$$\frac{\left(C_6^{35}-C_6^{20}-C_6^{15}\right)+\left(C_6^{30}-C_6^{20}-C_6^{10}\right)+\left(C_6^{25}-C_6^{15}-C_6^{10}\right)}{C_6^{45}}$$

So, the probability that at least one shift will not be represented in the sample, or equivalently, only one or two shifts is presented in the sample, can be determined as

$$\frac{C_6^{20}+C_6^{15}+C_6^{10}}{C_6^{45}}+\frac{\left(C_6^{35}-C_6^{20}-C_6^{15}\right)+\left(C_6^{30}-C_6^{20}-C_6^{10}\right)+\left(C_6^{25}-C_6^{15}-C_6^{10}\right)}{C_6^{45}}$$

$$=\frac{\left(C_6^{35}+C_6^{30}+C_6^{25}\right)-\left(C_6^{20}+C_6^{15}+C_6^{10}\right)}{C_6^{45}}$$

Q3. (20 points)

A work station has two independent machines A and B. The station will work properly if none of the machine experiences failure. From historical records on maintenance, it is estimated that the probability for machine A to face a failure during a work shift is 0.1; and the corresponding figure for machine B is 0.2. In addition, if both machines work well during a work shift, the whole batch of product produced during the shift will pass the quality acceptance test with probability of 0.95; however, if one of the machines experienced failure during the shift the probability for the batch of product to pass the test will be 0.5, and if both machines experienced failures during the shift the probability for the batch of product to pass the test will be only 0.1.

a. Find the probability that a batch of product produced in a specific work shift will pass the quality test.
b. If a batch of product produced in a work shift has been rejected by quality test, which machine the maintenance crew should investigate for the occurrence of failure during the work shift first?

Hint: Find the posterior probabilities that machines A, B experienced failure during the shift given that the batch of product produced has not passed the quality test.

a. Denote:

X_1: both machines worked properly during the work shift

X_2: machine A worked properly but machine B experienced failure during the work shift

X_3: machine B worked properly but machine A experienced failure during the work shift

X_4: both machines experienced failures during the work shift

Y: The batch of product produced passes the quality test

We have $P(X_1) = 0.9 * 0.8 = 0.72 \qquad P(X_2) = 0.9 * 0.2 = 0.18$

$P(X_3) = 0.1 * 0.8 = 0.08 \qquad P(X_4) = 0.1 * 0.2 = 0.02$

So, the probability that the batch of product produced passes the quality test is:

$$P(Y) = \sum_{i=1}^{4} P(Y|X_1)P(X_1)$$
$$= 0.95 * 0.72 + 0.5 * 0.18 + 0.5 * 0.08 + 0.1 * 0.02 = 0.816$$

b. The probability that machine A experienced failure during the work shift given that the batch of product produced has not passed the quality test:

$$P\left(X_3 \text{ or } X_4 \middle| \bar{Y}\right) = P\left(X_3 \middle| \bar{Y}\right) + P\left(X_4 \middle| \bar{Y}\right)$$
$$= \frac{P(\bar{Y}|X_3) * P(X_3)}{P(\bar{Y})} + \frac{P(\bar{Y}|X_4) * P(X_4)}{P(\bar{Y})}$$
$$= \frac{0.5 * 0.08}{0.184} + \frac{0.9 * 0.02}{0.184} = 0.315$$

The probability that machine B experienced failure during the work shift given that the batch of product produced has not passed the quality test:

$$P\left(X_2 \text{ or } X_4 \middle| \bar{Y}\right) = P\left(X_2 \middle| \bar{Y}\right) + P\left(X_4 \middle| \bar{Y}\right)$$
$$= \frac{P(\bar{Y}|X_2) * P(X_2)}{P(\bar{Y})} + \frac{P(\bar{Y}|X_4) * P(X_4)}{P(\bar{Y})}$$
$$= \frac{0.5 * 0.18}{0.184} + \frac{0.9 * 0.02}{0.184} = 0.587$$

So, machine B should be investigated first.

Q4. (10 points)
Twenty percent of all telephones of a certain type are submitted for service while under warranty. Of these, 60% can be repaired whereas the other 40% must be replaced with new units. If a company purchases ten of these telephones,

a. What is the probability that exactly two will end up to be replaced under warranty?
b. What is the probability that exactly five will be submitted for service under warranty in which only two will end up to be replaced?

Probability for a purchased phone to be replaced: $p_1 = 0.2*0.4 = 0.08$

Probability for a purchased phone to be repaired: $p_2 = 0.2*0.6 = 0.12$

Probability for a purchased phone to work properly: $p_3 = 0.8$

a. Probability that exactly two phones will be replaced: $C_2^{10} p_1^2 (1 - p_1)^8 = 0.1478$

b. Probability that exactly five will be submitted for service under warranty in which only two will end up to be replaced

$$\frac{10!}{2!3!5!} p_1^2 p_2^3 p_3^5 = 0.009$$

Q5. (10 points)
The incidence of a certain type of chromosome defect in adult male population is believed to be 1 in 75. A random sample of 800 individuals (only male) living in a big city reveals 16 who have such defects.

a. Can it be concluded that the incident rate of this defect in the examined city differs from the presumed rate for the entire adult male population (using $\alpha = 0.05$)
b. Determine the *p*-value associated with the test. Based on this *p*-value, let discuss whether the conclusion in (a.) is a strong or weak conclusion?

a. We have: $\hat{p} = 16/800 = 0.02 \quad \Rightarrow \quad z = \dfrac{0.02 - 1/75}{\sqrt{\dfrac{1/75 * 74/75}{800}}} = 1.644$

With $\alpha = 0.05$, rejection region: $z < -1.96$ or $z > 1.96$.
So, the null hypothesis cannot be rejected, i.e., there is no strong evidence that the incident rate of the investigated chromosome defect in the examined city differs from the presumed rate for the entire adult male population
b. The *p*-value of the test:

$$p \approx 2*0.05 = 0.1$$

The conclusion in (a) is a strong conclusion.

Q6. (10 points)
Consider the two-tailed test for mean value of a characteristic of a product produced from a manufacturing process: $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$. Suppose that the examined product characteristic follows normal distribution with a fixed standard deviation σ. Assume also that the actual mean value of the investigated characteristic is $\mu_1 = \mu_0 + \sigma$.

Let determine the sample size of the test such that Type II error β of the z-test is at most 0.25 (using $\alpha = 0.01$)

$$\beta = P\left\{\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

$$= P\left\{\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \le \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} \le \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\}$$

$$= P\left\{\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \le \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} \le \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\}$$

$$= P\left\{-\sqrt{n} - 2.576 \le z \le -\sqrt{n} + 2.576\right\}$$

$$\approx P\left\{z \le -\sqrt{n} + 2.576\right\}$$

With $\beta = 0.25$: $-\sqrt{n} + 2.576 = -0.674 \;\Rightarrow\; n = 11$

Q7. (10 points)
The nominal mean weight of a type of toothpaste produced by a process is μ_0. A sample has been drawn and it was found that the sample average $\bar{x}$ is greater than μ_0. Based on the sample statistics, the process engineer believes that the process mean has been shifted. However, in order to persuade the plant manager about this belief, a test should be conducted.

a. Which test the process engineer should select? Two-tailed, left-tailed, or right-tailed? Why?
b. Small or large value of α should be used? Why?

a. Right-tailed test because the one-tailed test is more powerful than the two-tailed test (i.e., the probability to reject H_0 is higher at the same value of α)
b. Large value of α should be used so that the probability of rejecting H_0 is higher

2012

Q1. (20 points) A lot of 100 semiconductor chips contains 25 that are defective. Two are selected randomly, without replacement, from the lot.

a. Determine the probability that the first one selected is defective.
b. Determine the probability that the second one selected is defective given that the first one was defective.
c. Determine the probability that both are defective.
d. How does the answer in (b) change if chips selected were replaced prior to the next selection?

a. 0.25

b. $P\{D_2/D_1\} = \frac{24}{99} = 0.24$

c. $P\{D_1 D_2\} = P\{D_1\} P\{D_2/D_1\} = 0.25 * \frac{24}{99} = 0.061$

d. $P\{D_2/D_1\} = 0.25$ (due to independence!)

Q2. (20 points)

In a chemical plant, 24 holding tanks are used for final product storage. Suppose that 6 tanks contain materials in which the viscosity exceeds the customer requirements. Four tanks are selected randomly without replacement.

a. Determine the probability that exactly one tank in the sample contains high-viscosity material.
b. What is the probability that at least one tank in the sample contains high-viscosity material?
c. In addition to the six tanks with high viscosity levels, four different tanks contain material with high impurities. What is the probability that exactly one tank in the sample contains high-viscosity material and exactly one tank in the sample contains material with high impurities?

a. $\frac{C_1^6 * C_3^{18}}{C_4^{24}} = 0.461$

b. $1 - \frac{C_4^{18}}{C_4^{24}} = 0.712$

c. $\frac{C_1^6 * C_1^4 * C_2^{14}}{C_4^{24}} = 0.206$

Q3. (20 points)
An inspector working for a manufacturing company has a 99% chance of correctly identifying defective items and a 0.5% chance of incorrectly classifying a good item as defective. The company has evidence that its line produces 0.9% of defective items.

a. Find the probability that a selected item for inspection is classified as defective.
b. If an item selected at random is classified as nondefective, what is the probability that it is indeed good?

Denote:
X_1: the selected item is defective
X_2: the selected item is good
Y: the selected item is classified as defective

a. We have

$$P(Y) = P(Y|X_1)P(X_1) + P(Y|X_2)P(X_2)$$
$$= 0.99 * 0.009 + 0.005 * 0.991 = 0.01387$$

b. The probability that the item is good given that it was classified as nondefective

$$P(X_2|\bar{Y}) = \frac{P(\bar{Y}|X_2) * P(X_2)}{P(\bar{Y})} = \frac{0.995 * 0.991}{1 - 0.01387} = 0.9999$$

Q4. (10 points)
Due to the fact that not all airline passengers show up for their reserved seat, an airline sells 130 tickets for a flight that can hold only 120 passengers. The probability that a passenger does not show up is 0.1

a. What is the probability that every passenger who shows up can take the flight?
b. What is the probability that the flight departs with empty seats?

Denote *X*: The number of passengers who do not show up
It is noted that *X* follows Binomial distribution with $n = 130, p = 0.1$

a. Probability that every passenger who shows up can take the flight: $P\{X \geq 10\}$
Approximate by normal distribution: $\mu = np = 13$, $\sigma = \sqrt{npq} = 3.42$

$$P\{X \geq 9.5 | \mu = 13, \sigma = 3.42\} = P\left\{Z \geq \frac{9.5 - 13}{3.42}\right\} = P\left\{Z \geq \frac{9.5 - 13}{3.42}\right\}$$
$$= P\{Z \geq -1.023\} = 0.8469$$

b. Probability that the flight departs with empty seats: $P\{X > 10\}$
Approximate by normal distribution:

$$P\{X \geq 10.5 | \mu = 13, \sigma = 3.42\} = P\left\{Z \geq \frac{10.5 - 13}{3.42}\right\} = P\left\{Z \geq \frac{10.5 - 13}{3.42}\right\}$$

Q5. (10 points)
A company is formulating a new shampoo and is interested in foam height (in millimeters). Foam height is approximately normally distributed and has a standard deviation of 25 millimeters. The company wishes to test $H_0 : \mu = 175$ mm versus $H_a : \mu > 175$ mm, using the results of n = 15 samples.

a. Find the Type I error probability α if the rejection region is $\bar{x} > 185$

b. What is the probability of Type II error β if the true mean foam height is 185mm?

Due to the fact that the population follows normal distribution and the standard deviation is known, the z-test can be used in this case even though the sample size is small.

a. The rejection region is defined by: $\bar{x} > \mu_0 + z_\alpha \dfrac{\sigma}{\sqrt{n}}$. So, we must have

$$\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} = 185 \quad \Rightarrow \quad z_\alpha = \frac{(185-\mu_0)*\sqrt{n}}{\sigma} = \frac{(185-175)*\sqrt{15}}{25} = 1.549$$

So, $\alpha = 0.061$

b. The probability of Type II error can be determined as follows:

$$\beta = P\left\{\bar{x} \le 185 \Big|_{\mu=\mu_1=185}\right\}$$

$$= P\left\{ z = \frac{\bar{x}-\mu_1}{\sigma/\sqrt{n}} \le \frac{185-\mu_1}{\sigma/\sqrt{n}} \Bigg|_{\mu=\mu_1=185} \right\} = P\{z \le 0\} = 0.5$$

Q6. (20 points)

In a random sample of 85 automobile crankshaft bearings, 10 have a surface finish roughness that exceeds the specifications.

a. Does the data present the strong evidence that the proportion of crankshaft bearings exhibiting excess surface roughness exceeds 0.1 (use α =0.05)?

b. If it is really the situation that p = 0.15, how about the probability that the test procedure in (a) will not reject the null hypothesis?

c. If p = 0.15, how large would the sample size has to be for us to have a probability of correctly rejecting the null hypothesis of 0.9?

Hypotheses: $H_0 : p = 0.1$

$H_a : p > 0.1$

a. We have

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{10/85 - 0.1}{\sqrt{\frac{0.1*0.9}{85}}} = 0.5423 < z_{0.05} = 1.645$$

There is no strong evidence to say that the proportion of crankshaft bearings exhibiting excess surface roughness exceeds 0.1

b. The critical value for rejection decision is: $p_0 + z_{0.05}\sqrt{\frac{p_0 q_0}{n}}$. So, if $p = p_1 = 0.15$ the test will not reject the null hypothesis with the probability defined as follows:

$$P\left\{\hat{p} \le p_0 + z_{0.05}\sqrt{\frac{p_0 q_0}{n}}\right\} = P\left\{\frac{\hat{p} - p_1}{\sqrt{\frac{p_1 q_1}{n}}} \le \frac{(p_0 - p_1) + z_{0.05}\sqrt{\frac{p_0 q_0}{n}}}{\sqrt{\frac{p_1 q_1}{n}}}\right\}$$

$$= P\left\{z \le \frac{(0.1 - 0.15) + z_{0.05}\sqrt{\frac{0.1*0.9}{85}}}{\sqrt{\frac{0.15*0.85}{85}}}\right\} = P\{z \le 0.091\} = 0.536$$

c. We must have :

$$1 - P\left\{z \le \frac{(0.1 - 0.15) + z_{0.05}\sqrt{\frac{0.1*0.9}{n}}}{\sqrt{\frac{0.15*0.85}{n}}}\right\} \ge 0.9$$

Q1. (15 points)

A batch of 500 containers for frozen orange juice contains ten that are defective. Two containers are selected, at random, without replacement from the batch

a. What is the probability that the second one selected is defective given that the first one was defective?
b. What is the probability that both are defective? Both are acceptable?
c. If another container is selected, what is the probability that it is defective given that the first one selected was defective and the second one selected was okay?

Denote F_i : the i^{th} selected container is defective

a. $P\left\{F_2 | F_1\right\} = \dfrac{9}{499} = 0.018$

b. $P\left\{F_1 F_2\right\} = \dfrac{C_2^{10}}{C_2^{500}} = 0.0004$

$P\left\{\overline{F}_1 \overline{F}_2\right\} = \dfrac{C_2^{490}}{C_2^{500}} == 0.9604$

c. $P\left\{F_3 | F_1 \overline{F}_2\right\} = \dfrac{9}{498} = 0.018$

Q2. (10 points)

An encryption-decryption system consists of three elements: encode, transmit, and decode. A faulty encode occurs in 0.5% of the messages processed, transmission errors occur in 1% of the messages, and a decode error occurs in 0.1% of the messages. Assume the errors are independent.

a. What is the probability of a completely defect-free message?
b. What is the probability of a message that has either an encode or a decode error?

Denote: *E*: the message has an encode error
T: the message has a transmission error
D: the message has a decode error

a. $P\left\{\overline{E}\overline{T}\overline{D}\right\} = 0.995 * 0.99 * 0.999 = 0.9841$

b. $P\left\{E \cup D\right\} = P\left\{E\right\} + P\left\{D\right\} - P\left\{ED\right\} = 0.005 + 0.001 - 0.005 * 0.001 = 0.006$

Q3. (20 points)

A lot of 50 spacing washers contains 30 washers that are thicker than the target dimension. Suppose that thee washers are selected at random, without replacement from the lot

a. What is the probability that the third washer selected is thicker than the target given that the first two washers selected are thinner than the target?
b. What is the probability that the third washer selected is thicker than the target?

c. What is the minimum number of washers that need to be selected so that the probability that one or more washers are thicker than the target is at least 0.9?
d. Give the answer for question c. if the lot of washers is large enough that it can be assumed that the sampling is done with replacement. Assume also that 60% of the washers exceed the target thickness.

Denote F_i : the i^{th} selected washer is thicker than the target

a. $P\left\{F_3\middle|\bar{F}_1\bar{F}_2\right\}=\frac{30}{48}=0.625$

b. The probability that the third washer is thicker than the target

$$P\left\{F_3\middle|F_1F_2\right\}P\left\{F_1F_2\right\}+P\left\{F_3\middle|\bar{F}_1F_2\right\}P\left\{\bar{F}_1F_2\right\}+P\left\{F_3\middle|F_1\bar{F}_2\right\}P\left\{F_1\bar{F}_2\right\}+P\left\{F_3\middle|\bar{F}_1\bar{F}_2\right\}P\left\{\bar{F}_1\bar{F}_2\right\}$$

$$=\frac{28}{48}*\frac{30}{50}*\frac{29}{49}+\frac{29}{48}*\frac{20}{50}*\frac{30}{49}+\frac{29}{48}*\frac{30}{50}*\frac{20}{49}+\frac{30}{48}*\frac{20}{50}*\frac{19}{49}=0.6$$

c. The statement "The probability that one or more washers are thicker than the target is at least 0.9" is equivalent to the statement "The probability that no washer is thicker than the target is at most 0.1". If we select n washers, the above condition can be expressed as:

$$P\left\{\bar{F}_1\bar{F}_2...\bar{F}_2\right\}\le 0.1$$

or

$$P\left\{\bar{F}_1\right\}P\left\{\bar{F}_2\middle|\bar{F}_1\right\}...P\left\{\bar{F}_n\middle|\bar{F}_1\bar{F}_2...\bar{F}_{n-1}\right\}\le 0.1$$

or

$$\frac{20}{50}*\frac{19}{49}*...\frac{21-n}{51-n}\le 0.1$$

By trial and error, we can find $n=3$

d. The probability that no washer is thicker than the target: 0.4^n . We must have

$0.4^n \le 0.1$ $\Rightarrow n \ge \frac{\ln(0.1)}{\ln(0.4)}=2.51$. So, $n=3$

Q4. (25 points)

A steel plate contains 20 bolts. Assume that 5 bolts are not torqued to the proper limit. Four bolts are selected at random, without replacement, to be checked for torque

a. What is the probability that all four of the selected bolts are torqued to the proper limit?
b. What is the probability that at least one of the selected bolts is not torqued to the proper limit?
c. If an operator checks a bolt, the probability that an incorrectly torqued bolt is identified is 0.95. If a checked bolt is correctly torqued, the operator's conclusion is always correct. What is the probability that at least one bolt in the sample of four is identified as being incorrectly torqued? If the operator identifies that all four bolts in the sample are okay, what is the probability that at least one bolt in the sample was incorrectly torqued?

a. The probability that all four selected bolts are torqued to the proper limit

$$\frac{C_4^{15}}{C_4^{20}} = 0.2817$$

b. The probability that at least one of the selected bolts is not torqued to the proper limit

$$1-\frac{C_4^{15}}{C_4^{20}} = 0.7183$$

c. Denote E_i ($i = 0,1,\ldots,4$) : there are i bolts that were not torqued to the proper limit in the sample, and F : At least one bolt in the sample is identified as being incorrectly torqued

We have:

$$P\{E_0\}=\frac{C_4^{15}}{C_4^{20}} = 0.2817 \qquad P\{E_1\}=\frac{C_3^{15}C_1^5}{C_4^{20}} = 0.4696$$

$$P\{E_2\}=\frac{C_2^{15}C_2^5}{C_4^{20}} = 0.2167 \qquad P\{E_3\}=\frac{C_1^{15}C_3^5}{C_4^{20}} = 0.0310$$

$$P\{E_4\}=\frac{C_4^5}{C_4^{20}} = 0.0010$$

Also,

$$P\{F|E_0\}=0 \qquad P\{F|E_1\}=0.95$$

$$P\{F|E_2\}=1-0.05^2 = 0.9975 \qquad P\{F|E_3\}=1-0.05^3 = 0.9999$$

$$P\{F|E_4\}=1-0.05^4 \approx 1$$

So, the probability that at least one bolt is identified as being incorrectly torqued is:

$$P\{F\}=\sum_{i=0}^{4} P\{F|E_i\}P\{E_i\}=0.6942$$

The probability that at least one bolt was incorrectly torqued given that the operator identified that all four bolts are okay:

$$P\{E_1\cup E_2\cup E_3\cup E_4|\overline{F}\}=1-P\{E_0|\overline{F}\}$$

$$=1-\frac{P\{\overline{F}E_0\}}{P\{\overline{F}\}}=1-\frac{P\{\overline{F}|E_0\}*P\{E_0\}}{P\{\overline{F}\}}$$

$$=1-\frac{1*0.2817}{1-0.6942}=0.0786$$

Q5. (10 points)
A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kg with a standard deviation of 0.5 kg. The company wishes to test the hypothesis $H_0 : \mu = 12$ against $H_a : \mu < 12$, using a random sample of four specimens.

a. Find the Type I error probability α if the rejection region is $\bar{x} < 11.5$ kg

b. What is the probability of Type II error β if the true mean elongation is 11.8kg?

a. The rejection region is defined by: $\bar{x} < \mu_0 - z_\alpha \dfrac{\sigma}{\sqrt{n}} = 12 - z_\alpha \dfrac{0.5}{\sqrt{4}} = 12 - 0.25 z_\alpha$.

So, we must have

$$12 - 0.25 z_\alpha = 11.5 \quad \Rightarrow \quad z_\alpha = 2$$

So, $\alpha = 0.02275$

b. The probability of Type II error can be determined as follows:

$$\beta = P\left\{\bar{x} > 11.5 \big|_{\mu = \mu_1 = 11.5}\right\}$$

$$= P\left\{ z = \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} > \frac{11.5 - 11.8}{0.5/\sqrt{4}} \right\} = P\{z > -1.2\} = 0.885$$

Q6. (20 points)

The life in hours of a battery is known to be approximately normally distributed, with a standard deviation $\sigma = 1.25$ hours. A random sample of 20 batteries has a mean life of $\bar{x} = 40.5$ hours.

a. Is there any evidence to support the claim that battery life exceeds 40 hours? Use α =0.05?

b. What is the *p*-value for the test in part (a)?

c. What is the β error for the test in part (a) if the true mean life is 42 hours?

d. What sample size should be required to ensure that β does not exceed 0.1 if the true mean life is 42 hours?

Hypotheses: $H_0 : \mu = 40$

$H_a : \mu > 40$

a. We have

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{40.5 - 40}{1.25 / \sqrt{20}} = 1.789 > z_{0.05} = 1.645$$

So, it can be claimed that the battery life exceed 40 hours

b. The observed significance level $p = 0.037$

c. The critical value of the test: $\bar{x}^u = \mu_0 + z_\alpha \sigma / \sqrt{n}$

$$\beta = P\left\{\bar{x} \le \mu_0 + z_\alpha \sigma / \sqrt{n} \Big|_{\mu=\mu_1=42}\right\}$$

$$= P\left\{z = \frac{\bar{x} - \mu_1}{\sigma / \sqrt{n}} \le \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}} + z_\alpha\right\}$$

$$= P\left\{z \le \frac{(40-42)}{1.25 / \sqrt{20}} + 1.645\right\} = P\{z \le -5.51\} \approx 0$$

c. We must have :

$$\beta = P\left\{z \le \frac{(40-42)}{1.25 / \sqrt{n}} + 1.645\right\} \le 0.1$$

So $\frac{(40-42)}{1.25 / \sqrt{n}} + 1.645 \le -1.282 \quad \Rightarrow n = 4$

2014

Q1. (10 points)

A batch of 140 semiconductor chips contains 10 chips that do not conform to the customer requirements. A sample of 5 chips is selected from the batch

a. What is the probability that the sample contains exactly one nonconforming chip?

b. What is the probability that the sample contains at least one nonconforming chip?

a. The probability that the sample contains one nonconforming chip

$$= \frac{C_1^{10} * C_4^{130}}{C_5^{140}} = 0.272$$

b. The probability that the sample contains at least one nonconforming chip

$$= 1 - \frac{C_0^{10} * C_5^{130}}{C_5^{140}} = 0.314$$

Q2. (10 points)

Products of a company are produced from 2 production lines. The two production lines produce with nonconforming percentages of 1% and 1.5%, respectively. The company receives an order of size $n = 50$ from a customer. In order to fulfill that order, production line 1 or production line 2 will equally likely be operated. At receiving inspection, the customer will accept the whole lot if and only if the number of nonconforming units does not exceed 2.

a. What is the probability that the lot is accepted by the customer?

b. If the lot is rejected, what is the probability that production line 2 has been operated to fulfill the order?

Denote: L_1: the lot is produced by line 1

L_2: the lot is produced by line 2

A: the lot is accepted

a. We have

$$P\{A\} = P\{A|L_1\}P\{L_1\} + P\{A|L_1\}P\{L_1\}$$

In which

$$P\{A|L_1\} = \sum_{i=0}^{2} C_i^{50}(0.01)^i(0.99)^{50-i} = 0.9862$$

$$P\{A|L_2\} = \sum_{i=0}^{2} C_i^{50}(0.015)^i(0.985)^{50-i} = 0.9607$$

So,

$$P\{A\} = 0.9862 * \frac{1}{2} + 0.9607 * \frac{1}{2} = 0.9735$$

b.

$$P\{L_2|\overline{A}\} = \frac{P\{\overline{A}|L_2\}P\{L_2\}}{P\{\overline{A}\}} = \frac{(1-.9607)*0.5}{1-.9735} = 0.740$$

Q3. (20 points)

An employee of the records office at a university currently has ten forms on his desk awaiting processing. Six of these are withdrawal petitions and the other four are course substitution requests.

a. If he randomly selects five of these forms to give to a subordinate, what is the probability that only one of the two types of forms remains on his desk?

If the five remaining forms are of only one type, they should be withdrawal petitions. So, the probability that all five remaining forms are of only one type is: $\frac{C_5^6}{C_5^{10}} = 0.0238$

b. Suppose he has time to process only four of these forms before leaving for the day. If these four are randomly selected one by one, what is the probability that each succeeding form is of a different type from its predecessor?

b. The probability of interest is $P = P\{WSWS\} + P\{SWSW\}$

in which S = course substitution request, W = withdrawal petition

Hence, $P = \frac{6}{10} * \frac{4}{9} * \frac{5}{8} * \frac{3}{7} + \frac{4}{10} * \frac{6}{9} * \frac{3}{8} * \frac{5}{7} = 0.143$

Q4. (20 points)

A test for the presence of a certain disease has probability .20 of giving a false positive reading (i.e., indicating that an individual has the disease when this is not the case) and probability .10 of giving a false negative result. Suppose that ten individuals are tested, five of whom have the disease and five of whom do not. Let X = the number of positive readings that results.

a. Does X have a binomial distribution? Explain your reasoning.

b. What is the probability that exactly three of the ten test results are positive?

c.

a. No. Because the probability of a positive reading is not a constant in this case, it depends on whether the tested individual has the disease or not.

b. Let i be the number of infected individuals in the three positive test results, then among the ten tests we must have:

i	true positive test results (with probability = 0.9)
$5-i$	false negative test results (with probability = 0.1)
$2+i$	true negative test results (with probability = 0.8)
$3-i$	false positive test results (with probability = 0.2)

So, the probability that there are three positive test results is:

$$\sum_{i=0}^{3} \left(C_i^5 * 0.9^i * 0.1^{5-i}\right) * \left(C_{3-i}^5 * 0.2^{3-i} * 0.8^{2+i}\right) = 0.0273$$

Q5. (20 points)

A manufacturer is interested in the output voltage of a power supply used in a PC. Output voltage is assumed to be normally distributed, with standard deviation 0.2 volt, and the manufacturer wishes to test $H_0 : \mu = 5\,\text{volts}$ against $H_a : \mu \neq 5\,\text{volts}$, using $n = 8$ units.

a. Although the sample size is small, the manufacturer still uses the z-test (instead of t-test) in this case? Is it correct? Explain.

b. The "not rejection" region in this test is $4.85 \le \bar{x} \le 5.15$. Find the value of α .

c. Find the power of the test for detecting a true mean output voltage of 5.1 volts.

Hint: The power of the test is $(1-\beta)$ *in which* $\beta = P\{\text{Not reject } H_0 | \mu = \mu_1 \neq \mu_0\}$ (*with* $\mu_0 = 5\,\text{volts}$ *and* $\mu_1 = 5.1\,\text{volts}$).

a. Although the sample size is small, it is known that the voltage output follows normal distribution, and hence, the sample average will follow normal distribution. In addition, the standard deviation is known, therefore, the test can be conducted based on normal distribution, i.e., the z-test

b. The "not rejection" region is defined by: $\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

So, we must have

$$\mu_0 = \frac{4.85 + 5.15}{2} = 5.0$$

and $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 0.15 \quad \Rightarrow \quad z_{\alpha/2} = \frac{0.15\sqrt{n}}{\sigma} = \frac{0.15\sqrt{8}}{0.2} = 2.12$

Therefore, $\alpha = 0.0339$

c. The probability of Type II error can be determined as follows:

$$\begin{aligned}
\beta \quad &= P\left\{4.85 \le \bar{x} \le 5.15 \big|_{\mu=\mu_1=5.1}\right\} \\
&= P\left\{\frac{4.85-5.1}{0.2/\sqrt{8}} \le z \le \frac{5.15-5.1}{0.2/\sqrt{8}}\right\} \\
&= P\{-3.536 \le z \le 0.707\} \\
&= \Phi(0.707) - \Phi(-3.536) = 0.76
\end{aligned}$$

The power of the test: $1 - \beta = 0.24$

Q6. (20 points)

An engineer who is studying the tensile strength of a steel alloy knows that tensile strength is approximately normally distributed with $\sigma = 415\,\text{kN/m}^2$. A random sample of 12 specimens has a mean tensile strength of $\overline{x} = 23800\,\text{kN/m}^2$.

a. Test the hypothesis that mean strength is $24150\,\text{kN/m}^2$, use α =0.01?

b. What is the smallest level of significant at which you would be willing to reject the null hypothesis?

c. What is the β error for the test in part (a) if the true mean is 23925 hours?

d. Suppose that we want to reject the null hypothesis with probability at least 0.8 if mean strength $\mu = 23925$. What sample size should be used?

d. Confirm the answer in part (a) by constructing a confidence interval on mean tensile strength

The test: $H_0: \mu = 24150\ \text{kN/m}^2$; $H_a: \mu \neq 24150\ \text{kN/m}^2$

a. We have

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{23800 - 24150}{415/\sqrt{12}} = -2.922$$

With $\alpha = 0.01$: $z_{\alpha/2} = z_{0.005} = 2.576$

So, the null hypothesis is rejected (the tensile strength is less than 24150 kN/m^2)

b. The observed significance level $p = 2*0.00174 \approx 0.0035$. So, the smallest level of significant at which the null hypothesis can be rejected is 0.0035. However, this value is impracticable.

c. $$\beta = P\left\{\mu_0 - z_{\alpha/2}\sigma/\sqrt{n} \le \overline{x} \le \mu_0 + z_{\alpha/2}\sigma/\sqrt{n}\Big|_{\mu=\mu_1=23925}\right\}$$

$$= P\left\{\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \le z \le \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right\}$$

$$= P\left\{\frac{24150 - 23925}{415/\sqrt{12}} - 2.576 \le z \le \frac{24150 - 23925}{415/\sqrt{12}} + 2.576\right\}$$

$$= P\{-0.698 \le z \le 4.454\}$$

$$= \Phi(4.454) - \Phi(-0.698) = 0.757$$

d. The probability to reject the null hypothesis is :

$$P\left\{\bar{x} < \mu_0 - z_{\alpha/2}\,\sigma/\sqrt{n}\right\} + P\left\{\bar{x} > \mu_0 + z_{\alpha/2}\,\sigma/\sqrt{n}\right\}$$

However, due to the fact that $\mu_1 = 23925 < \mu_0 = 24150$, the above probability can be approximated as:

$$P\left\{\bar{x} < \mu_0 - z_{\alpha/2}\,\sigma/\sqrt{n}\right\} = P\left\{z < \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right\}$$

We must have:

$$\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \approx z_{0.2}$$

$$\Rightarrow \quad n \approx \left(\frac{\sigma\left(z_{\alpha/2} + z_{0.2}\right)}{\mu_0 - \mu_1}\right)^2 = \left(\frac{415(2.576 + 0.842)}{24150 - 23925}\right)^2 = 40$$

e. Confidence Interval:

$$\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n}$$

Or

$$23491.42 \le \mu \le 24108.58$$

The hypothesized value of 24150 is located outside the confidence interval. So, the null hypothesis is rejected.

Q1. (10 points)

The probability that a customer's order is not shipped on time is 0.04. A particular customer places three orders, and the orders are placed independently.

a. What is the probability that exactly one order is not shipped on time?

b. What is the probability that two or more orders are not shipped on time?

a. The probability that exactly one order is not shipped on time

$$= C_1^3 (0.04)(0.96)^2 = 0.1106$$

b. The probability that two or more orders are not shipped on time

$$= C_2^3 (0.04)^2 (0.96) + C_3^3 (0.04)^3 (0.96)^0 = 0.0047$$

Q2. (10 points)

A component is delivered from a supplier to a manufacturer in batch of size $N = 10000$ units. On receiving a specific batch with percentage of defective component $p = 0.01$, the manufacturer randomly selects a sample of size $n_1 = 50$ for inspection. The whole batch will be accepted if there is no defective component detected and rejected if there exist at least two defective components. However, when there is only one defective component detected, another sample of size $n_2 = 50$ will be selected for inspection and the batch will be accepted only if there is no defective component found in the second sample

a. What is the probability that the batch is accepted?

b. If the batch is accepted, what is the probability that it is accepted by using only the inspection result of the first sample?

Denote: A_1: the batch is accepted on the first sample

N_1: no conclusion on the first sample

A_2: the batch is accepted

a. We have

$$P\{A_2\} = P\{A_1\} + P\{A_2 | N_1\} * P\{N_1\}$$

In which

$$P\{A_1\} = P\{A_2 | N_1\} = 0.99^{50} = 0.6050$$

$$P\{N_1\} = C_1^{50} * 0.01 * 0.99^{49} = 0.3056$$

Hence, the probability that the batch is accepted can be determined as:

$$P\{A_2\} = 0.6050 + 0.3056 * 0.6050 = 0.7899$$

b. $$P\{A_1 | A_2\} = \frac{P\{A_1 A_2\}}{P\{A_2\}} = \frac{P\{A_1\}}{P\{A_2\}} = \frac{0.6050}{0.7899} = 0.7660$$

Q3. (20 points)

Password of a computer is set with seven *different* characters in which the first four characters are selected among the 26 letters (a to z) and the last three characters are selected among 10 integers (0 to 9).

a. If a hacker tries to access the computer by selecting a random password, what is the probability that he has the correct password?
b. If the hacker knew that the letters used in the password is selected in the set of vowels only (i.e., either *a*, *e*, *i*, *o*, *u*). What is the probability that he has the correct password?
c. What will be the answers for a. & b. if the selected characters can be repeated.

a. Number of possible passwords: $\left(4!*C_4^{26}\right)*\left(3!*C_3^{10}\right)=258,336,000$

So, the probability that the hacker has the correct password is: $\dfrac{1}{258336000}=3.871*10^{-9}$

b. Number of possible passwords: $\left(4!*C_4^{5}\right)*\left(3!*C_3^{10}\right)=86,400$

So, the probability that the hacker has the correct password is: $\dfrac{1}{86400}=1.157*10^{-5}$

c. The answers will be changed to:

For a.: $\dfrac{1}{26^4*10^3}=2.188*10^{-9}$

For b.: $\dfrac{1}{5^4*10^3}=1.6*10^{-6}$

Q4. (20 points)

In a multiple choice test, a student may understand or not understand the questions. The probability that the student understands a specific question is 0.6

- When the student understands the question, the probability that he/she gives the correct answer is 0.75
- When the student does not understand the question, he/she is equally likely to select either "give no answer" or "give an answer by randomly select one among the four available options".

a. For a specific question, what will be the probability that the student gives the correct answer.
b. If the test consists of 10 questions, what is the probability that the student give correct answers for 7 questions.
c. Given that the student understands five among ten questions, what is the probability that he/she gives correct answers for 9 questions?

a. Denote: U : the student understands the question

$\bar{U}$: the student does not understand the question

A : the student selects to answer the question

C : the student has the correct answer

We have:

$$p = P\{C\} = P\{C|U\} * P\{U\} + P\{C|\bar{U}A\} * P\{\bar{U}A\}$$

Noted that $P\{\bar{U}A\} = P\{A|\bar{U}\} * P\{\bar{U}\} = 0.5 * 0.4 = 0.2$

Hence,

$$p = 0.75 * 0.6 + 0.25 * 0.2 = 0.5$$

b. The probability that the student give correct answers for 7 questions

$$= C_7^{10} p^7 (1 - p)^3 = 0.1172$$

c. To give correct answers for 9 questions, the student must answer at least 4 questions among the five questions that he don't understand. Let i be the number of "not understand" questions that the student decides to answer

- When $i = 4$: The probability to have 9 correct answers is:
  $0.75^5 * 0.25^4 = 0.000927$
- When $i = 5$: The probability to have 9 correct answers is:
  $\left(C_4^5 * 0.75^4 * 0.25\right) * 0.25^5 + 0.75^5 * \left(C_4^5 * 0.25^4 * 0.75\right) = 0.003862$

So, the probability that the student give correct answer for 9 questions is:

$$0.000927 * P\{i = 4\} + 0.003862 * P\{i = 5\}$$
$$= 0.000927 * \left\{C_4^5 * 0.5^4 * 0.5\right\} + 0.003862 * 0.5^5 = 0.000266$$

Q5. (20 points)

The heat evolved (in calories per gram) of a cement mixture is approximately normally distributed. The mean is thought to be 100 and the standard deviation is 2. We wish to test $H_0: \mu = 100$ versus $H_a: \mu \neq 100$ with a sample size of $n = 9$ specimens.

a. Can we use the z-test (instead of t-test) in this case? Explain.

b. The "not rejection" region in this test is $98.5 \le \bar{x} \le 101.5$. Find the value of α.

c. Find the value of β if the true mean heat evolved is $\mu_1 = 103$ (noted that $\beta = P\{\text{Not reject } H_0 | \mu = \mu_1 \neq \mu_0\}$).

d. Find the value of β if the true mean heat evolved is $\mu_1 = 105$. Why the value of β here is less than the value of β found in c.?

a. Although the sample size is small, it is known that heat evolved follows normal distribution, and hence, the sample average will follow normal distribution. In addition, the standard deviation is known, therefore, the z-test can be employed.

b. The "not rejection" region is defined by: $\mu_0 - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu_0 + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.

So, we must have

$$\mu_0 = \frac{98.5 + 101.5}{2} = 100.0$$

and $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1.5 \quad \Rightarrow \quad z_{\alpha/2} = \dfrac{1.5\sqrt{n}}{\sigma} = \dfrac{1.5\sqrt{9}}{2} = 2.25$

Therefore, $\alpha = 0.0244$

c. The probability of Type II error can be determined as follows:

$$\begin{aligned}
\beta \quad &= P\{98.5 \le \bar{x} \le 101.5|_{\mu=\mu_1=103}\} \\
&= P\left\{\frac{98.5-103}{2/\sqrt{9}} \le z \le \frac{101.5-103}{2/\sqrt{9}}\right\} \\
&= P\{-6.75 \le z \le -2.25\} \\
&= \Phi(-2.25) - \Phi(-6.75) = 0.0122
\end{aligned}$$

d. The probability of Type II error can be determined as follows:

$$\beta \quad = P\{98.5 \le \bar{x} \le 101.5|_{\mu=\mu_1=105}\}$$

$$= P\left\{\frac{98.5-105}{2/\sqrt{9}} \le z \le \frac{101.5-105}{2/\sqrt{9}}\right\}$$
$$= P\{-9.75 \le z \le -5.25\}$$
$$= \Phi(-5.25) - \Phi(-9.75) \approx 0$$

The value of β is reduced in comparison to the answer in c. because the difference between the actual mean value and the hypothesized mean value increases, and hence, it becomes easier to detect the shift.

Q6. (20 points) Let give answers for the following theoretical questions

a. Why the *p*-value method is preferable over other hypothesis testing methods?

b. When conducting a hypothesis test, which conclusion, i.e., "Reject H_0" or "Not reject H_0", you really want to achieve in principle?

c. If you really want to reject H_0, high or low value of α should be selected? Why?

d. In which situation, the one-tailed test should be used even though the nature of the problem requests a two-tailed test? Explain why?

e. Why Fisher test is always conducted on the right tail of Fisher distribution? If we still want to conduct the Fisher test directly on the left tail, can we?

All questions have been discussed in class. -____-

Q1. (10 points)
It is known that two defective copies of a commercial software program were sent to a shipping lot that has a total of 80 copies. A sample of 3 copies is selected from the lot and inspected.

a. Determine the probability that exactly one of the defective copies will be found.
b. Determine the probability that both defective copies will be found.

a. The probability that exactly one of the defective copies will be found

$$= \frac{C_1^2 * C_2^{78}}{C_3^{80}} = 0.0731$$

b. The the probability that both defective copies will be found

$$= \frac{C_2^2 * C_1^{78}}{C_3^{80}} = 0.00095$$

Q2. (15 points)
A master student has asked his advisor for a letter of recommendation for a doctoral scholarship. He estimates that there is an 80% chance that he will get the scholarship if he receives a strong recommendation, a 40% chance if he receives a moderately good recommendation, and a 10% chance if he receives a weak recommendation. He also estimates that probabilities that the recommendation will be strong, moderate, or weak are 0.7, 0.2, and 0.1, respectively.

a. How certain is the student that he will receive the scholarship?
b. Given that the student does receive the scholarship, find the probabilities that he receives a strong recommendation, a moderate recommendation and a weak recommendation.
c. Given that the student does not receive the scholarship, find the probabilities that he receives a strong recommendation, a moderate recommendation and a weak recommendation.

Notation:
- O: The scholarship is offered
- NO: The scholarship is not offered
- SR: Strong recommendation
- MR: Moderate recommendation
- WR: Weak recommendation

a. The probability to receive the scholarship

$$\begin{aligned} P(O) &= P(O|SR)P(SR) + P(O|MR)P(MR) + P(O|WR)P(WR) \\ &= 0.8*0.7 + 0.4*0.2 + 0.1*0.1 = 0.65 \end{aligned}$$

b. $$P(SR|O) = \frac{P(O|SR)P(SR)}{P(O)} = \frac{0.8*0.7}{0.65} = \frac{56}{65} = 0.862$$

$$P(MR|O) = \frac{P(O|MR)P(MR)}{P(O)} = \frac{0.4*0.2}{0.65} = \frac{8}{65} = 0.123$$

$$P(WR|O) = \frac{P(O|WR)P(WR)}{P(O)} = \frac{0.1*0.1}{0.65} = \frac{1}{65} = 0.015$$

c.

$$P(SR|NO) = \frac{P(NO|SR)P(SR)}{P(NO)} = \frac{0.2*0.7}{0.35} = \frac{14}{35} = 0.4$$

$$P(MR|NO) = \frac{P(NO|MR)P(MR)}{P(NO)} = \frac{0.6*0.2}{0.35} = \frac{12}{35} = 0.343$$

$$P(WR|NO) = \frac{P(NO|WR)P(WR)}{P(NO)} = \frac{0.9*0.1}{0.35} = \frac{9}{35} = 0.257$$

Q3. (15 points)

A congested computer network has a 0.002 probability of losing a data package and packet losses are independent events. A lost package must be resent.

a. What is the probability that an e-mail message with 100 packets will need any to be resent?
b. What is the probability that an e-mail message with 3 packages will need exactly one to be resent?
c. If 10 e-mail messages are sent, each with 100 packets, what is the probability that at least one message will need some packets to be resent?

a. The probability that an e-mail message with 100 packets will need any to be resent

$$p_1 = 1 - 0.998^{100} = 0.1814$$

b. The probability that an e-mail message with 3 packages will need exactly one to be resent

$$p_2 = C_1^3 (0.002) * (0.998)^2 = 0.00598$$

c. Let X denote the number of e-mail messages with some packets need to be resent, then $X \boxed{?} B(10, p_1)$. Therefore, the probability that at least one message will need some packets to be resent is:

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - (1 - p_1)^{10} = 0.8649$$

Q4. (20 points)

Urn 1 contains 5 back balls and 5 white balls, urn 2 contains 7 black balls and 3 white balls. One ball is drawn from each urn and put into the other urn simultaneously. One urn will be selected using the following rules

- If the two swapped balls are both black or both white, urn 1 and urn 2 are equally likely to be selected
- If the transferred ball from urn 1 is back and the transferred ball from urn 2 is white, urn 1 will be selected
- If the transferred ball from urn 1 is white and the transferred ball from urn 2 is back, urn 2 will be selected

A ball will then be drawn from the selected urn

a. Find the probability that the final selected ball is a white ball.

b. If the final selected ball is back, what is the probability that the two swapped balls at the beginning are of different colors?

Denote:

BB: the transferred balls from urn1 and urn 2 are both black

BW: the transferred balls from urn1 and urn 2 are black and white, respectively

WB: the transferred balls from urn1 and urn 2 are white and black, respectively

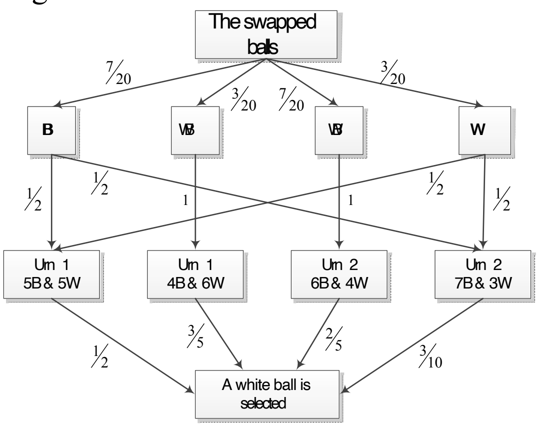
WW: the transferred balls from urn1 and urn 2 are both white

We have:

$$P\{BB\} = \frac{1}{2} * \frac{7}{10} = \frac{7}{20} \qquad P\{BW\} = \frac{1}{2} * \frac{3}{10} = \frac{3}{20}$$

$$P\{WB\} = \frac{1}{2} * \frac{7}{10} = \frac{7}{20} \qquad P\{WW\} = \frac{1}{2} * \frac{3}{10} = \frac{3}{20}$$

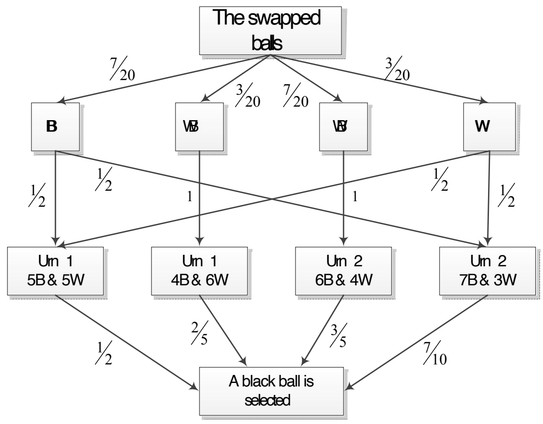
a. Using the following diagram



The probability that the final selected ball is a white ball can be determined as:

$$P\{White\} = \frac{7}{10} * \frac{1}{2} * \frac{1}{2} + \frac{7}{10} * \frac{1}{2} * \frac{3}{10} + \frac{3}{20} * 1 * \frac{3}{5} + \frac{7}{20} * 1 * \frac{2}{5} + \frac{3}{20} * \frac{1}{2} * \frac{1}{2} + \frac{3}{20} * \frac{1}{2} * \frac{3}{10} = 0.43$$

b. Using the following diagram



The probability that the final selected ball is a black ball can be determined as:

$$P\{Black\} = \frac{7}{10}*\frac{1}{2}*\frac{1}{2}+\frac{7}{10}*\frac{1}{2}*\frac{7}{10}+\frac{3}{20}*1*\frac{2}{5}+\frac{7}{20}*1*\frac{3}{5}+\frac{3}{20}*\frac{1}{2}*\frac{1}{2}+\frac{3}{20}*\frac{1}{2}*\frac{7}{10} = 0.57$$

(or $P\{Black\} = 1 - P\{White\}$)

So, the probability that the two swapped balls at the beginning are of different colors is

$$P\{BW \cup WB|_{Black}\} = \frac{\frac{3}{20}*1*\frac{2}{5}+\frac{7}{20}*1*\frac{3}{5}}{0.57} = 0.4737$$

Q5. (20 points)

In order to monitor the length of steel rods produced from a shearing machine with the nominal cutting length of 2.15 in., a sample of 5 steel rods will be selected every hour to test the following hypotheses:

$$H_0 : \mu = 2.15 \text{ in.} \qquad\qquad H_a : \mu \neq 2.15 \text{ in.}$$

For quality control purpose, the null hypothesis is rejected if $\bar{x} < 2.1$ in. or $\bar{x} > 2.2$ in. If the null hypothesis cannot be rejected, the shearing machine is believed to be in-control; otherwise, it is believed to be out-of-control. The length of the produced steel rods is assumed to follow normal distribution with a standard deviation σ of 0.025 in.

The specification range for the steel rods is $[2.120, 2.195]$. A steel rod with the length located in the specification range is considered a *conforming* product; otherwise, it is a *nonconforming* product.

a. If the shearing machine is in-control, let determine the fraction nonconforming, i.e., the probability that a produced steel rod is a nonconforming product.
b. If the process mean shifts to 2.175 in. and the standard deviation σ does not change, what will be the fraction nonconforming?
c. In part (b), what is the probability of catching the shift on the first sample right after the shift has occurred?
d. In part (b), what is the probability of missing the shift on the next two samples and catching it on the third sample after the shift has occurred?

a. The fraction of nonconforming product:

$$P(x < 2.120) + P(x > 2.195)$$

$$= \Phi\left(\frac{2.120 - 2.15}{0.025}\right) + 1 - \Phi\left(\frac{2.195 - 2.15}{0.025}\right)$$

$$= \Phi(-1.2) + 1 - \Phi(1.8)$$

$$= 0.1151 + 1 - 0.9641 = 0.151$$

b. The fraction nonconforming when $\mu = 2.175$

$$P(x < 2.120) + P(x > 2.195)$$

$$= \Phi\left(\frac{2.120 - 2.175}{0.025}\right) + 1 - \Phi\left(\frac{2.195 - 2.175}{0.025}\right)$$

$$= \Phi(-2.2) + 1 - \Phi(0.8)$$

$$= 0.0139 + 1 - 0.7881 = 0.226$$

c. The probability of catching the shift on the first sample

$$p = P(\bar{x} < 2.1) + P(\bar{x} > 2.2)$$

$$= \Phi\left(\frac{2.1 - 2.175}{0.025/\sqrt{5}}\right) + 1 - \Phi\left(\frac{2.2 - 2.175}{0.025/\sqrt{5}}\right)$$

$$= \Phi(-6.71) + 1 - \Phi(2.24)$$

$$= 0.0 + 1 - 0.9873 = 0.0127$$

d. The probability of missing the shift on the next two samples and catching it in the third sample is: $p(1-p)^2 = 0.0124$

Q6. (20 points) Let give answers for the following theoretical questions

a. In which ranges of the *p*-value, the conclusion of a hypothesis test is a strong conclusion, i.e., it does not depend on the selected value of significance level α ?
b. Why we should avoid saying that we accept the null hypothesis if we cannot reject it?
c. Can we control type II error β of a hypothesis test? Let derive the formula to help determine β in a two-tailed hypothesis test of mean value with large sample if we know the actual value μ_1 of μ which is different from the hypothesized value μ_0 ?
d. If you really want to reject a null hypothesis, what you will do?

All questions have been discussed in class.

Q1. (15 points)

In a batch of product of size $N = 100$ there are $a = 10$ defective units. A sample of size n is randomly selected from the lot. Denote X to be a random number that represents the number of defective units in the sample.

a. Given $n = 5$, what is the domain of X? Given $n = 15$, what is the domain of X?

b. Determine the domain of X in terms of $N, a,$ and n

c. Given $n = 5$, find $P\{X = i\}$ for all values i in the domain of X.

a. When $n = 5$: the domain of X is $D_X = [0,5]$

When $n = 15$: the domain of X is $D_X = [0,10]$

b. The domain of X is $D_X = [\text{Max}\{0, n - N + a\}, \text{Min}\{n, a\}]$

c. With $n = 5$:

$$P\{X = 0\} = \frac{C_0^{10} C_5^{90}}{C_5^{100}} = 0.5838$$

$$P\{X = 1\} = \frac{C_1^{10} C_4^{90}}{C_5^{100}} = 0.3394$$

$$P\{X = 2\} = \frac{C_2^{10} C_3^{90}}{C_5^{100}} = 0.0702$$

$$P\{X = 3\} = \frac{C_3^{10} C_2^{90}}{C_5^{100}} = 0.0064$$

$$P\{X = 4\} = \frac{C_4^{10} C_1^{90}}{C_5^{100}} = 0.0003$$

$$P\{X = 5\} = \frac{C_5^{10} C_0^{90}}{C_5^{100}} \approx 0$$

Q2. (10 points)

A manufacturing plant operates three shifts per day. The probabilities that a product produced in shift 1, shift 2, and shift 3 is a defective are 0.01, 0.015 and 0.012, respectively. For quality control purpose, a sample of size 10 will be selected equally likely from the three shifts each day for inspection purpose. If there are more than one defective in the sample, the production line is considered to be out-of-control on that day.

a. Determine the probability that the production line is believed to be in-control on a specific day.

b. If the production line is considered out-of-control on a specific day, what are the probabilities that the sample is selected from shift 1, shift 2, and shift 3 on that day?

Notation: S_1, S_2, S_3 : The sample is selected from shifts 1,2,3, respectively
Y : The production line is in-control
N : The production line is out-of-control

a. The probability that the production line is believed to be in-control

$$P(Y) = P(Y|S_1)P(S_1) + P(Y|S_2)P(S_2) + P(Y|S_3)P(S_3)$$

In which

$$P(Y|S_1) = C_0^{10} * 0.01^0 * 0.99^{10} + C_1^{10} * 0.01^1 * 0.99^9 = 0.9957$$

$$P(Y|S_2) = C_0^{10} * 0.015^0 * 0.985^{10} + C_1^{10} * 0.015^1 * 0.985^9 = 0.9907$$

$$P(Y|S_3) = C_0^{10} * 0.012^0 * 0.988^{10} + C_1^{10} * 0.012^1 * 0.988^9 = 0.9939$$

So,

$$P(Y) = 0.9957 * \frac{1}{3} + 0.9907 * \frac{1}{3} + 0.9939 * \frac{1}{3} = 0.9934$$

b. The probabilities of interest are

$$P(S_1|N) = \frac{P(N|S_1)P(S_1)}{P(N)} = \frac{(1-0.9957) * \frac{1}{3}}{1-0.9934} = 0.2183$$

$$P(S_2|N) = \frac{P(N|S_2)P(S_2)}{P(N)} = \frac{(1-0.9907) * \frac{1}{3}}{1-0.9934} = 0.4721$$

$$P(S_3|N) = \frac{P(N|S_3)P(S_3)}{P(N)} = \frac{(1-0.9939) * \frac{1}{3}}{1-0.9934} = 0.3096$$

Q3. (15 points)
A supplier sends lots of component of the same lot size to a manufacturer based on a long-term supply contract. The probability for any component sent by the supplier to be a defective is 0.001. Under normal inspection scheme, on receiving a specific lot the manufacturer will select a random sample of size $n = 100$ for testing, and the whole lot will be accepted only if there is no more than 2 defective components found; otherwise, the whole lot will be rejected and returned to the supplier. When a lot is rejected, the manufacturer will switch from normal inspection scheme to tightened inspection scheme. Suppose that the normal inspection scheme is applied at present

a. What is the probability that the normal inspection scheme will still be used in the next three lots sent from the supplier?
b. What is the probability that the normal inspection scheme will be switched to the tightened inspection scheme in the third lots sent from the supplier from now?
c. What is the probability that the normal inspection scheme will be switched to the tightened inspection scheme within the next three lots sent from the supplier from now?

a. The probability for the lot to be accepted under normal inspection scheme

$$p_1 = C_0^{100} * 0.001^0 * 0.999^{100} + C_1^{100} * 0.001^1 * 0.999^{99} + C_2^{100} * 0.001^2 * 0.999^{98} = 0.9998$$

So, the probability that the normal inspection scheme is used in the three slots is:

$$P_1 = p_1^3 = 0.9995$$

b. The probability that the inspection scheme will be switched to tightened scheme in the third slot is

$$P_2 = p_1^2(1 - p_1) = 0.0002$$

c. The probability that the normal inspection scheme will be switched to the tightened inspection scheme within the next three slots is:

$$P_3 = 1 - P_1 = 0.0005$$

Q4. (20 points)

The first child of a family is equally likely to be a boy or a girl. However, the probability for the next child to be a boy or a girl depends on the immediate preceding child. If the preceding child is a boy then the next child will be a boy with probability of 0.7, but if the preceding child is a girl, the next child will be a boy with probability of 0.4 only. Consider a family with 3 children.

a. Find the probability that there are 2 boys and 1 girl in that family.
b. If the third child of the family is a girl, what is the probability that the others are boys?
c. If the third child of the family is a girl, what is the probability that the first one is also a girl?

Draw the tree diagram for the problem. Then from the tree diagram, we can find all probabilities of interest as follows

a. The probability to have 2 boys and 1 girl:

$$P\{2B \& 1G\} = \frac{1}{2} * 0.7 * 0.3 + \frac{1}{2} * 0.3 * 0.4 + \frac{1}{2} * 0.4 * 0.7 = 0.305$$

b. The probability for the third child is a girl:

$$P\{G_3\} = \frac{1}{2} * 0.7 * 0.3 + \frac{1}{2} * 0.3 * 0.6 + \frac{1}{2} * 0.4 * 0.3 + \frac{1}{2} * 0.6 * 0.6 = 0.435$$

So, the probability that the others are boys given the third child is girl can be determined as:

$$P\{B_1 B_2 | G_3\} = \frac{\frac{1}{2} * 0.7 * 0.3}{0.435} = 0.241$$

c. The probability that the first child is a girl given that the third child is a girl:

$$P\{G_1 | G_3\} = \frac{\frac{1}{2} * 0.4 * 0.3 + \frac{1}{2} * 0.6 * 0.6}{0.435} = 0.552$$

Q5. (20 points)
Consider the following hypothesis testing problem:

$$H_0: 2\mu_1 - 3\mu_2 = a \qquad\qquad H_a: 2\mu_1 - 3\mu_2 \neq a$$

a. What will be the testing procedures for large sample & small sample cases (assume that $\sigma_1^2 = \sigma_2^2$ for small sample case)?

b. Let construct the confidence intervals for $2\mu_1 - 3\mu_2$ in large sample & small sample cases. Can we use these confidence intervals for testing purpose in question a.?.

a. <u>For large sample</u>:

Test statistic:
$$z = \frac{(2\bar{x}_1 - 3\bar{x}_2) - a}{\sqrt{\frac{4\sigma_1^2}{n_1} + \frac{9\sigma_2^2}{n_2}}} \approx \frac{(2\bar{x}_1 - 3\bar{x}_2) - a}{\sqrt{\frac{4s_1^2}{n_1} + \frac{9s_2^2}{n_2}}}$$

Rejection region:
$$z < -z_{\alpha/2} \text{ or } z > z_{\alpha/2}$$

<u>For small sample</u>:

Test statistic:
$$t = \frac{(2\bar{x}_1 - 3\bar{x}_2) - a}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}}\sqrt{\frac{4}{n_1} + \frac{9}{n_2}}}$$

Rejection region:
$$t < -t_{\alpha/2, n_1+n_2-2} \text{ or } z > t_{\alpha/2, n_1+n_2-2}$$

b. Confidence intervals:

<u>For large sample</u>:
$$(2\bar{x}_1 - 3\bar{x}_2) - z_{\alpha/2}\sqrt{\frac{4s_1^2}{n_1} + \frac{9s_2^2}{n_2}} \le 2\mu_1 - 3\mu_2 \le (2\bar{x}_1 - 3\bar{x}_2) + z_{\alpha/2}\sqrt{\frac{4s_1^2}{n_1} + \frac{9s_2^2}{n_2}}$$

<u>For small sample</u>:
$$(2\bar{x}_1 - 3\bar{x}_2) - t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}}\sqrt{\frac{4}{n_1} + \frac{9}{n_2}} \le 2\mu_1 - 3\mu_2 \le (2\bar{x}_1 - 3\bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}}\sqrt{\frac{4}{n_1} + \frac{9}{n_2}}$$

We can used those confidence intervals to conduct the test in question a. The null hypothesis will be rejected if the hypothesized value *a* is not located within the confidence intervals.

Q6. (20 points) Let give answers for the following theoretical questions

a. The *p*-value is considered to be very important in hypothesis testing procedure, why?
b. What you will do if you don't want to reject a null hypothesis?
c. In principle, selection of one-tailed test or two-tailed test should be based on the nature of the problem. However, it is sometimes observed that one-tailed test has been used instead of two-tailed test. Why?
d. When we reject the null hypothesis in the two-tailed test with $H_0 : \mu = \mu_0$, can we conclude $\mu < \mu_0$ or $\mu > \mu_0$?

a. The *p*-value can guide the selection of significant level α
b. If we don't want to reject a null hypothesis, we should use small value of α.
c. This is done so as to enlarge the rejection area when we really want to reject the null hypothesis
d. Yes, we can. It depends on if the null hypothesis is rejected on the right or the left tail of the distribution of sample mean. If it is rejected on the right tail, we can conclude that $\mu > \mu_0$. If it is rejected on the left, we can conclude that $\mu < \mu_0$

2018

Q1. (15 points)

In a master class there are 5 male students and 10 female students. The lecturer would like to form a working group with 2 male students and 3 female students. Mr. A and Ms. B really want to be selected to work in the group. However, the probability that a specific male/female student to be selected is equally likely in the group of male/female students.

a. Find the probability that Mr. A will be selected.
b. Find the probability that Ms. B will be selected.
c. Find the probability that both Mr. A and Ms. B will be selected.

a. Noted that among the two male students selected, Mr. A can be selected as the first or the second. Therefore, the probability that Mr. A will be selected can be determined as:

$$P_1 = \frac{1}{5} + \frac{4}{5} * \frac{1}{4} = \frac{2}{5}$$

Another way: $$P_1 = \frac{C_1^4}{C_2^5} = \frac{4}{10} = \frac{2}{5}$$

(C_1^4 : number of ways to select another male student in the remaining 4 male students together with Mr. A to form a group of 2 male students)

b. Similarly, the probability that Ms. B will be selected can be determined as:

$$P_2 = \frac{1}{10} + \frac{9}{10} * \frac{1}{9} + \frac{9}{10} * \frac{8}{9} * \frac{1}{8} = \frac{3}{10}$$

Another way: $$P_2 = \frac{C_2^9}{C_3^{10}} = \frac{3}{10}$$

c. The probability that both Mr. A and Ms. B will be selected: $$P = P_1 * P_2 = \frac{3}{25}$$

Q2. (15 points) A master student has asked his advisor for a letter of recommendation for a doctoral scholarship. He estimates that there is an 80% chance that he will get the scholarship if he receives a strong recommendation, a 40% chance if he receives a moderately good recommendation, and a 10% chance if he receives a weak recommendation. He also estimates that probabilities that the recommendation will be strong, moderate, or weak are 0.7, 0.2, and 0.1, respectively.

a. How certain is the student that he will receive the scholarship?
b. Given that the student does receive the scholarship, find the probabilities that he receives a strong recommendation, a moderate recommendation and a weak recommendation.
c. Given that the student does not receive the scholarship, find the probabilities that he receives a strong recommendation, a moderate recommendation and a weak recommendation.

Notation:
- O: The scholarship is offered
- NO: The scholarship is not offered
- SR: Strong recommendation
- MR: Moderate recommendation
- WR: Weak recommendation

a. The probability to receive the scholarship

$$\begin{aligned} P(O) &= P(O|SR)P(SR) + P(O|MR)P(MR) + P(O|WR)P(WR) \\ &= 0.8*0.7 + 0.4*0.2 + 0.1*0.1 = 0.65 \end{aligned}$$

b.

$$P(SR|O) = \frac{P(O|SR)P(SR)}{P(O)} = \frac{0.8*0.7}{0.65} = \frac{56}{65} = 0.862$$

$$P(MR|O) = \frac{P(O|MR)P(MR)}{P(O)} = \frac{0.4*0.2}{0.65} = \frac{8}{65} = 0.123$$

$$P(WR|O) = \frac{P(O|WR)P(WR)}{P(O)} = \frac{0.1*0.1}{0.65} = \frac{1}{65} = 0.015$$

c.

$$P(SR|NO) = \frac{P(NO|SR)P(SR)}{P(NO)} = \frac{0.2*0.7}{0.35} = \frac{14}{35} = 0.4$$

$$P(MR|NO) = \frac{P(NO|MR)P(MR)}{P(NO)} = \frac{0.6*0.2}{0.35} = \frac{12}{35} = 0.343$$

$$P(WR|NO) = \frac{P(NO|WR)P(WR)}{P(NO)} = \frac{0.9*0.1}{0.35} = \frac{9}{35} = 0.257$$

Q3. (15 points)
Items produced from a process are classified into two categories, i.e, conforming or nonconforming (i.e., defective). If an item produced is a conforming item, the next item produced will also be a conforming item with probability 0.8. However, if an item produced is a nonconforming item then the next item produced will be a conforming item with probability 0.6 only. A sample of three items produced successively are selected from the process

a. If the first item is a conforming item, what is the probability that the remaining two items are also conforming items?
b. If the first item is a nonconforming item, what is the probability that the remaining two items are conforming items?
c. Given that the first and the last item are conforming items, what is the probability that the second item is also a conforming item?

a. The probability that the remaining two items are conforming items given that the first item is a conforming items: 0.8*0.8 = 0.64

b. The probability that the remaining two items are conforming items given that the first item is a nonconforming items: 0.6*0.8 = 0.48

c. The probability that the last item is a conforming item given that the first item is a conforming item: 0.8*0.8 + 0.2*0.6 = 0.76.
So, the probability that the second item is a conforming item given that the first and the last item are conforming items: $\frac{0.64}{0.76} = 0.86$

Q4. (20 points) A true-false question is to be posed to a husband and wife team on a quiz show. Both the husband and the wife will, independently, give the correct answer with probability $p = 0.6$.

a. Which of the following is a best strategy for this couple?
    a. Randomly choose one of them and let that person answer the question
    b. Have them both consider the question and then either give the common answer if they agree, or, if they disagree, flip a coin to determine which answer to give.
b. If the second strategy is used, what is the conditional probability that the couple gives the correct answer given that they 1./ agree; 2./ disagree?
c. If the second strategy is used, what is the conditional probability that the couple disagreed given that they give the correct answer?

Notation:
- AR: The couple agrees on the right answer
- AW: The couple agrees on the wrong answer
- DA: The couple disagrees
- C: The answer is correct

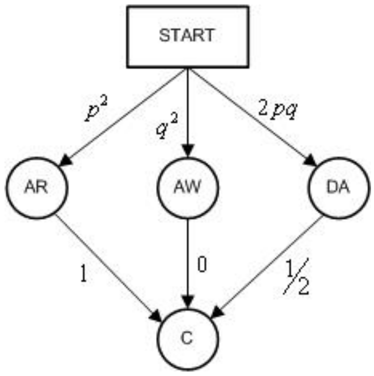
a. We have

1. If the first strategy is used, the probability to answer correctly is

$$\frac{1}{2}p+\frac{1}{2}p = p = 0.6$$

2. If the second strategy is used, the probability to answer correctly is:

$$P\{C\} = 1*p^2 + 0*q^2 + \frac{1}{2}*2pq = p^2 + pq = p = 0.6$$



So, both strategies are equivalent

b. If the second strategy is used

1. The conditional probability that the couple gives the correct answer given that they agree is:

$$P\{C|AR\,or\,AW\} = \frac{P\{C\cap(AR\,or\,AW)\}}{P\{AR\,or\,AW\}} = \frac{P\{C\cap AR\}+P\{C\cap AW\}}{P\{AR\,or\,AW\}}$$

$$= \frac{P\{C\cap AR\}}{P\{AR\,or\,AW\}} = \frac{P\{AR\}}{P\{AR\,or\,AW\}} = \frac{p^2}{p^2+q^2} = \frac{9}{13}$$

2. The conditional probability that the couple gives the correct answer given that they disagree is: $\frac{1}{2}$

c. The conditional probability that the couple disagreed given that they give the correct answer is

$$P\{DA|C\} = \frac{P\{C\cap DA\}}{P\{C\}} = \frac{P\{C|DA\}P\{DA\}}{P\{C\}}$$

$$= \frac{\frac{1}{2}*2pq}{p} = q = 0.4$$

Q5. (15 points)

Consider the following hypothesis testing problem:

$$H_0: \mu_1 = 3\mu_2 \qquad H_a: \mu_1 \neq 3\mu_2$$

a. What will be the testing procedures for large sample & small sample cases (assume that $\sigma_1^2 = \sigma_2^2$ for small sample case)?

b. Let construct the confidence intervals for $\mu_1 - 3\mu_2$ in large sample & small sample cases.

a. For large sample:

Test statistic:
$$z = \frac{\bar{x}_1 - 3\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{9\sigma_2^2}{n_2}}} \approx \frac{\bar{x}_1 - 3\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{9s_2^2}{n_2}}}$$

Rejection region:
$$z < -z_{\alpha/2} \text{ or } z > z_{\alpha/2}$$

For small sample:

Test statistic:
$$t = \frac{\bar{x}_1 - 3\bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}}\sqrt{\frac{1}{n_1} + \frac{9}{n_2}}}$$

Rejection region:
$$t < -t_{\alpha/2, n_1+n_2-2} \text{ or } z > t_{\alpha/2, n_1+n_2-2}$$

b. Confidence intervals:

For large sample:
$$(\bar{x}_1 - 3\bar{x}_2) - z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{9s_2^2}{n_2}} \le \mu_1 - 3\mu_2 \le (\bar{x}_1 - 3\bar{x}_2) + z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{9s_2^2}{n_2}}$$

For small sample:
$$(\bar{x}_1 - 3\bar{x}_2) - t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}}\sqrt{\frac{1}{n_1} + \frac{9}{n_2}} \le \mu_1 - 3\mu_2 \le (\bar{x}_1 - 3\bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}}\sqrt{\frac{1}{n_1} + \frac{9}{n_2}}$$

Q6. (20 points) Let give answers for the following theoretical questions

a. In conducting hypothesis test, it is better to gather the sample data before defining the null and the alternative hypotheses as well as the value of α . Explain why?
b. Why we should not say that we "accept" the null hypothesis?
c. If the nature of the test requires a one-tailed test, can we conduct a two-tailed test instead?
d. When we cannot reject the null hypothesis in a hypothesis test, what does it really mean?

a. Knowing the sample data, selection of the right-tailed test or the left-tailed test can be done appropriately. Also, the *p*-value can be determined so that an appropriate value of α can be selected

b. Because the β error can be very high.

c. We should not do this because the test can give useful result only when the null hypothesis is rejected. Changing the one-tailed test to two-tailed test will reduce the probability of rejection.

d. When we cannot reject the null hypothesis, it means that the assumption stated in the null hypothesis is still valid, but it does not mean that the null hypothesis is true. In other words, the situation before the test remains the same.

2019

Q1. (15 points)

You receive products delivered from a supplier in boxes of 10 units. In each box, there are 2 defective units and 8 conforming units. For quality checking purpose, a sample of 3 units will be selected from each box and the box is considered of acceptable quality if there is no more than one defective unit in the sample.

a. What is the probability that a box is classified as of acceptable quality level?
b. What is the probability that a box is classified as of unacceptable quality level?
c. Suppose that you receive a batch of 10 boxes from the supplier. The whole batch will be accepted if and only if at most 2 boxes are classified as of unacceptable quality level. What is the probability for the whole batch to be accepted?

a. The probability that a box is classified as of acceptable quality level:

$$p = \frac{C_3^8 + C_2^8 C_1^2}{C_3^{10}} = \frac{56+56}{120} = \frac{14}{15} = 0.933$$

b. The probability that a box is classified as of unacceptable quality level:

$$q = 1 - p = \frac{1}{15} = 0.067$$

c. The probability that the whole batch is accepted

$$P = C_0^{10} q^0 (1-q)^{10} + C_1^{10} q^1 (1-q)^9 + C_2^{10} q^2 (1-q)^8 = 0.975$$

Q2. (15 points)
In an oral exam, a student can be equally likely assigned to one of the three exam rooms. The probabilities that the student will pass the exam if he is assigned to room 1, room 2, room 3 are estimated at 50%, 60% and 70%, respectively.

a. What is the probability that the student will pass the exam?
b. Given that the student has passed the exam. What are the probabilities that he has been assigned to room 1, room 2, and room 3?
c. Given that the student has not passed the exam. What are the probabilities that he has been assigned to room 1, room 2, and room 3?

Notation:	$R1, R2, R3$:	The student is assigned to room 1, room 2, room 3
	P:	The scholarship passes the exam
	NP:	The student does not pass the exam

a. The probability the the student will pass the eaxm

$$P(P) = P(P|R1)P(R1) + P(P2)P(R2) + P(P|R3)P(R3)$$

$$= 0.5 * \frac{1}{3} + 0.6 * \frac{1}{3} + 0.7 * \frac{1}{3} = 0.6$$

b.

$$P(R1|P) = \frac{P(P|R1)P(R1)}{P(P)} = \frac{0.5 * 1/3}{0.6} = \frac{5}{18} = 0.278$$

$$P(R2|P) = \frac{P(P|R2)P(R2)}{P(P)} = \frac{0.6 * 1/3}{0.6} = \frac{1}{3} = 0.333$$

$$P(R3|P) = \frac{P(P|R3)P(R3)}{P(P)} = \frac{0.7 * 1/3}{0.6} = \frac{7}{18} = 0.389$$

c.

$$P(R1|NP) = \frac{P(NP|R1)P(R1)}{P(NP)} = \frac{0.5 * 1/3}{0.4} = \frac{5}{12} = 0.417$$

$$P(R2|NP) = \frac{P(NP|R2)P(R2)}{P(NP)} = \frac{0.4 * 1/3}{0.4} = \frac{1}{3} = 0.333$$

$$P(R3|NP) = \frac{P(NP|R3)P(R3)}{P(NP)} = \frac{0.3 * 1/3}{0.4} = \frac{1}{4} = 0.25$$

Q3. (15 points)

Items of a product produced in a manufacturing plant may come from two independent production lines. The percentages of defective unit produced from the two lines are 1% and 2%, respectively. A sample of two units are selected as follows: the first unit is randomly selected from one of the two production lines; if the first unit selected is a conforming unit then the second unit will be selected from the other production line; but if the first unit is a defective unit then the second unit will be selected from the same production line.

a. What is the probability that the two units in the sample are conforming units?
b. What is the probability that the second unit in the sample is a defective unit?
c. Given that the second unit is a defective one, what is the probability that the first unit is a conforming unit?

a. The probability that the two units in the sample are conforming unit:

$$0.5*0.99*0.98 + 0.5*0.98*0.99 = 0.9702$$

b. The probability that the second unit in the sample is a defective unit

$$0.5*0.99*0.02 + 0.5*0.01*0.01 + 0.5*0.98*0.01 + 0.5*0.02*0.02 = 0.0151$$

c. The probability that the first unit is a defective one given that the second unit is a defective one

$$p = \frac{0.5*0.99*0.02 \ + \ 0.5*0.98*0.01}{0.5*0.99*0.02 \ + \ 0.5*0.01*0.01 \ + \ 0.5*0.98*0.01 \ + \ 0.5*0.02*0.02} = 0.9834$$

Q4. (20 points)

A patient who was infected by a disease is undergoing a blood test. The test can give a false negative result with probability 0.05. However, if the test gives a negative result, another test will be conducted on the patient.

a. What is the probability that the final result is "positive" (i.e., the final conclusion is that the patient is actually infected by the disease)?
b. If the final result is "positive", what is the probability that this conclusion comes from the first test?
c. If the final result is "positive", what is the probability that this conclusion is based on the second test?

Notation:	$PO1$:	The first test gives positive result
	$NPO1$:	The first test gives a false negative result
	PO:	The final test result is positive

a. The probability that the final result is positive

$$P\{PO\} = P\{PO|PO1\}P\{PO1\} + P\{PO|NPO1\}P\{NPO1\} = 1*0.95 + 0.95*0.05 = 0.9975$$

b. The probability that the final positive result comes from the first test

$$P\{PO1|PO\} = \frac{P\{PO|PO1\}P\{PO1\}}{P\{PO\}} = \frac{0.95}{0.9975} = 0.9524$$

c. The probability that the final positive result comes from the first test:

$$1-0.9524 = 0.0476$$

Q5. (15 points)
Consider the following hypothesis testing problem:

$$H_0 : 2\mu_1 = 3\mu_2 \qquad\qquad H_a : 2\mu_1 \neq 3\mu_2$$

a. What will be the testing procedures for large sample & small sample cases (assume that $\sigma_1^2 = \sigma_2^2$ for small sample case)?

b. Let construct the confidence intervals for $2\mu_1 - 3\mu_2$ in large sample & small sample cases.

a. For large sample:

Test statistic:

$$z = \frac{2\overline{x}_1 - 3\overline{x}_2}{\sqrt{\frac{4\sigma_1^2}{n_1} + \frac{9\sigma_2^2}{n_2}}} \approx \frac{2\overline{x}_1 - 3\overline{x}_2}{\sqrt{\frac{4s_1^2}{n_1} + \frac{9s_2^2}{n_2}}}$$

Rejection region:

$$z < -z_{\alpha/2} \text{ or } z > z_{\alpha/2}$$

For small sample:

Test statistic:

$$t = \frac{2\overline{x}_1 - 3\overline{x}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{4}{n_1} + \frac{9}{n_2}}}$$

Rejection region:

$$t < -t_{\alpha/2, n_1+n_2-2} \text{ or } z > t_{\alpha/2, n_1+n_2-2}$$

b. Confidence intervals:

For large sample:

$$\left(2\overline{x}_1 - 3\overline{x}_2\right) - z_{\alpha/2}\sqrt{\frac{4s_1^2}{n_1} + \frac{9s_2^2}{n_2}} \le 2\mu_1 - 3\mu_2 \le \left(2\overline{x}_1 - 3\overline{x}_2\right) + z_{\alpha/2}\sqrt{\frac{4s_1^2}{n_1} + \frac{9s_2^2}{n_2}}$$

For small sample:

$$\left(2\overline{x}_1 - 3\overline{x}_2\right) - t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{4}{n_1} + \frac{9}{n_2}} \le 2\mu_1 - 3\mu_2 \le \left(2\overline{x}_1 - 3\overline{x}_2\right) + t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{4}{n_1} + \frac{9}{n_2}}$$

Q6. (20 points) Let give answers for the following theoretical questions

a. In conducting hypothesis test, when the conclusion (i.e., "reject" or "not reject" the null hypothesis) will not depend on the value of α ? Explain why?

b. Can we use the *t*-test for the case of large sample when conducting the test on mean value of the population? Why?

c. In principle, when conducting hypothesis test which conclusion (i.e., reject H_0 or not reject H_0) we would like to achieve? Why?

d. Can we conduct a two-tailed test for a problem that requires an one-tailed test by its nature?

a. When the observed significant level p is less than 0.01 or greater than 0.1, the conclusion of the test does not depend on the value of α. This is because of

- When p is less than 0.01: $p<\alpha\ \forall\alpha\in[0.01,0.1]$, and hence, H_0 will be rejected for sure
- When p is greater than 0.1: $p>\alpha\ \forall\alpha\in[0.01,0.1]$, and hence, H_0 cannot be rejected for sure

b. Yes, we can because the *t*-test can be used for any value of sample size, and for large sample the *t*-test and the *z*-test are exactly the same.

c. When conducting hypothesis test, we would like to reject H_0 because if we cannot reject H_0, the situation after the test is the same as the situation before the test (i.e., H_0 is still an assumption).

d. No, we cannot because converting an one-tailed test to a two-tailed test will lead to the reduction of the rejection area, and hence, it will be more difficult to reject the null hypothesis.

Q1. (15 points)

In each work shift, a production line can produce a batch of 50 units in which the probability for each unit in the batch to be a defective unit is 0.01, independently from each other. A batch is considered acceptable if there are no more than 2 defective units.

a. What is the probability that a batch is classified as of acceptable quality level?
b. What is the probability that a batch is classified as of unacceptable quality level?
c. In an order of 10 batches delivered to a customer, there is a requirement that at least 9 batches should be of acceptable quality level; otherwise, the supply contract will be terminated in the future. What is the probability for the above requirement is fulfilled?

a. The probability that a batch is classified as of acceptable quality level:

$$p = \sum_{i=0}^{2} C_i^{50} \left(0.01\right)^i \left(0.99\right)^{50-i} = 0.9862$$

b. The probability that a batch is classified as of unacceptable quality level:

$$q = 1 - p = 0.0138$$

c. The probability that the requirement is fulfilled

$$P = \sum_{i=9}^{10} C_i^{10} p^i q^{10-i} = 0.992$$

Q2. (15 points)

You ask your friend to water a sickly plant while you are on vacation. Without water it will die with probability 0.8; with water it will die with probability 0.15. You are 90 percent certain that your neighbor will remember to water the plant.

a. What is the probability that the plant will be alive when you return?
b. If the plant is dead when you return, what is the probability that your friend forgot to water it?

Notation:
- F: The friend remembered to water the plant
- $\bar{F}$: The friend forgot to water the plant
- E: The plant is still alive
- $\bar{E}$: The plant is dead

a. The probability that the plant is still alive

$$\begin{aligned} P(E) &= P\left(E \middle| F\right)P(F) + P\left(E \middle| \bar{F}\right)P\left(\bar{F}\right) \\ &= 0.85*0.9 + 0.2*0.1 = 0.785 \end{aligned}$$

b. The probability that the plant is dead: $P\left(\bar{E}\right) = 0.215$

So, the probability of interest is

$$P\left(\bar{F} \middle| \bar{E}\right) = \frac{P\left(\bar{E} \middle| \bar{F}\right)P\left(\bar{F}\right)}{P(O)} = \frac{0.8*0.1}{0.215} = 0.372$$

Q3. (30 points)
Among 2000 recently admitted patients in an intensive care unit (ICU) of a hospital, there are:

- 120 patients infected by both covid-19 and seasonal influenza
- 250 patients infected by only covid-19
- 380 patients infected by only seasonal influenza
- 1250 patients infected by other diseases

Among the patients infected by both covid-19 and influenza, 80% will have a fever
Among the patients infected by only covid-19, 70% will have a fever
Among the patients infected by only seasonal influenza, 90% will have a fever
Among the patients who are not infected by both covid-19 and influenza, only 50% will have a fever.

a. A new patient has just been admitted in the ICU unit of the hospital and was confirmed by a test that he/she was infected by only seasonal influenza, what will be the probability that the patient will have a fever?
b. A new patient has just been admitted in the ICU unit of the hospital, what will be the probability that the patient will have a fever?
c. A new patient was admitted to the ICU unit of the hospital yesterday, and he/she starts to have a fever today. Before a test is conducted to confirm the disease(s) the patient was suffered from, what will be the most likely anticipated classification of the patient? (i.e., is the patient suffered from both covid-19 & influenza, or just only covid-19, or just only influenza, or another disease?).

(*Hint*: find the *posterior probabilities* of the events: "the patient suffered from both covid-19 & influenza", "the patient suffered from only covid-19", "the patient suffered from only influenza", "the patient suffered from another disease" under the condition that he/she has a fever. Then select the one with highest posterior probability)

Denote

C: the new patient was infected by covid-19
I: the new patient was infected by seasonal influenza
F: the new patient will have a fever

Then

$C \cap I$: the new patient was infected by both covid-19 and influenza
$C \cap \bar{I}$: the new patient was infected by only covid-19
$\bar{C} \cap I$: the new patient was infected by only influenza
$\bar{C} \cap \bar{I}$: the new patient was infected by another disease

a. The probability that the new patient will have a fever given that he/she was confirmed to be infected by only seasonal influenza is 0.9

b. The probability that the new patient will have a fever:

$$P(F)=P(F|C\cap I)P(C\cap I)+P(F|C\cap \bar{I})P(C\cap \bar{I})$$
$$+P(F|\bar{C}\cap I)P(\bar{C}\cap I)+P(F|\bar{C}\cap \bar{I})P(\bar{C}\cap \bar{I})$$
$$=0.8*\frac{120}{2000}+0.7*\frac{250}{2000}+0.9*\frac{380}{2000}+0.5*\frac{1250}{2000}=0.619$$

c. We have:

$$P(C\cap I|F)=\frac{P(F|C\cap I)*P(C\cap I)}{P(F)}=\frac{0.8*120/2000}{0.619}=0.0775$$

$$P(C\cap \bar{I}|F)=\frac{P(F|C\cap \bar{I})*P(C\cap \bar{I})}{P(F)}=\frac{0.7*250/2000}{0.619}=0.1414$$

$$P(\bar{C}\cap I|F)=\frac{P(F|\bar{C}\cap I)*P(\bar{C}\cap I)}{P(F)}=\frac{0.9*380/2000}{0.619}=0.2763$$

$$P(\bar{C}\cap \bar{I}|F)=\frac{P(F|\bar{C}\cap \bar{I})*P(\bar{C}\cap \bar{I})}{P(F)}=\frac{0.5*1250/2000}{0.619}=0.5048$$

So, the most likely classification of the new patient is "$\bar{C}\cap\bar{I}$", i.e., he/she suffered from another disease

Q4. (20 points)
Consider the following hypothesis testing problem:

$H_0: 3\mu_1 = 4\mu_2$ $\qquad\qquad H_a: 3\mu_1 \neq 4\mu_2$

a. What will be the testing procedures for large sample & small sample cases (assume that $\sigma_1^2=\sigma_2^2$ for small sample case)?
b. Let construct the confidence intervals for $3\mu_1-4\mu_2$ in large sample & small sample cases.

a. <u>For large sample</u>:

Test statistic: $$z=\frac{3\bar{x}_1-4\bar{x}_2}{\sqrt{\frac{9\sigma_1^2}{n_1}+\frac{16\sigma_2^2}{n_2}}}\approx\frac{3\bar{x}_1-4\bar{x}_2}{\sqrt{\frac{9s_1^2}{n_1}+\frac{16s_2^2}{n_2}}}$$

Rejection region: $z<-z_{\alpha/2}$ or $z>z_{\alpha/2}$

<u>For small sample</u>:

Test statistic: $$t=\frac{3\bar{x}_1-4\bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{9}{n_1}+\frac{16}{n_2}}}$$

Rejection region: $t < -t_{\alpha/2, n_1+n_2-2}$ or $z > t_{\alpha/2, n_1+n_2-2}$

b. Confidence intervals:

<u>For large sample</u>:

$$(3\bar{x}_1 - 4\bar{x}_2) - z_{\alpha/2}\sqrt{\frac{9s_1^2}{n_1} + \frac{16s_2^2}{n_2}} \le 3\mu_1 - 4\mu_2 \le (3\bar{x}_1 - 4\bar{x}_2) + z_{\alpha/2}\sqrt{\frac{9s_1^2}{n_1} + \frac{16s_2^2}{n_2}}$$

<u>For small sample</u>:

$$(3\bar{x}_1 - 4\bar{x}_2) - t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{9}{n_1} + \frac{16}{n_2}} \le 3\mu_1 - 4\mu_2 \le (3\bar{x}_1 - 4\bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{9}{n_1} + \frac{16}{n_2}}$$

Q5. (20 points) Let give answers for the following theoretical questions

a. Why determining the $p-\text{value}$ is considered the best approach when conducting hypothesis tests?
b. When determine the sample size needed to construct a confidence interval for population proportion with a predefined width, the analysts usually use $p = 0.5$. Why? What is the main drawback of this practice?
c. What should be the appropriate conclusion when we cannot reject a null hypothesis? Explain why?
d. When conducting a one-tailed test (for instance, $H_0 : \mu \le \mu_0$ versus $H_0 : \mu > \mu_0$), the null hypothesis will be converted to: $H_0 : \mu = \mu_0$. What are the reasons for this transformation?

a. Knowing the $p-\text{value}$ will help to select the appropriate value of α so that the analysts can manipulate the statistical conclusion.

b. The analysts usually use $p = 0.5$ because they cannot use a sample to estimate p (if this approach is used, the sample must be large, and it is a waste of time and money!). The drawback of this approach is that the sample size will be over-estimated.

c. When we cannot reject the null hypothesis, the appropriate conclusion is "the null hypothesis cannot be rejected". We should not "accept" the null hypothesis because it is just a valid assumption if we cannot reject it.

d. The null hypothesis must be a clear statement (it should be noted that the statement $H_0 : \mu \le \mu_0$ is not clear because it covers many hypothesized values of μ). Also, the principal of the test is to reject the null hypothesis in order to confirm the alternative hypothesis. So, "reject $H_0 : \mu = \mu_0$ to confirm $H_0 : \mu > \mu_0$" is equivalent to "reject $H_0 : \mu \le \mu_0$ to confirm $H_0 : \mu > \mu_0$".