1. On a multiple-choice exam with four possible answers for each of the five questions, what is the probability that a student would get 16 or more (out of 20, i.e., get at least four questions correct) just by guessing randomly?

2. Suppose two teams, E and W, are playing the NBA finals (a series of 7 games), where the series is done when E or W wins four matches firstly. If each match is independently won by team E with probability p and by team W with probability $1-p$. Find the expected number of matches that are played, and evaluate this expected number when $p = 1/2$.

3. A fountain show starts every 80 minutes, you arrive the place at random and decide to wait for 20 minutes, what's the probability you will witness the show?

Question 1:

The Probability of guessing a multiple Question right is 0.25. If guessing randomly, the P of rewarding 16 points is $0.25^4 \cdot 0.75 \cdot C_4^5 = \frac{15}{1024}$, and P of rewarding 20 points is $0.25^5 \cdot C_5^5 = \frac{1}{1024}$. Adding up together, the answer is $\frac{1}{64}$ #

Question 2:

In this system, if they only play four games, it means that one team has won four games in a row. Since both teams have chances to win in a row, the chance of this series ending in four games is $p^4 + (1-p)^4$. So, how high is the chance that the system will end with five games? Regardless of which team wins this game, five games means that in the first four games, the winning team has 3 wins and 1 loss. This probability of E winning the game is $C_3^4 \times p^3 \times (1-p) \times p$. The meaning of the last p is that the team E MUST win the last game to win the whole series. Similarly, the probability of W winning the game is $C_3^4 \times (1-p)^3 \times p \times (1-p)$ The answer is

$$4 \times (p^4 + (1-p)^4) + 5 \times C_3^4 \times (p^4 \times (1-p) + p \times (1-p)^4)$$

$$+6 \times C_3^5 \times (p^4 \times (1-p)^2 + p^2 \times (1-p)^4) + 7 \times C_3^6 \times (p^4 \times (1-p)^3 + p^3 \times (1-p)^4)$$ #

The expected number when p = 1/2 is

$$4 \times 1/8 + 5 \times 1/4 + 6 \times 5/16 + 7 \times 5/16$$

, which is **93/16.** #

Question 3:

Intuitively, the chances of waiting the show in 20 minutes equals to the chances of arriving the fountain 60 minutes after last show ended. Therefore, the answer is $20/80 = \mathbf{1/4}$. #

In the post office of A city, there are two clerks working with different efficiencies: clerk 1 has service time following an exponential distribution with mean $1/\mu_1$ while the service time of clerk 2 follows **a different** exponential distribution with mean $1/\mu_2$. One day, John enters the post office and he is served by clerk 1 at 8h00.

a. Mary enters at 8h10, what is the probability she sees John is still being served by clerk 1?
b. Since John is still in service, Mary goes to clerk 2 to be served. What is the probability that Mary finishes her service **before** John does?

* Hint: Memoryless Property

Question 4:

① Using exponential distribution, $f(x) = \lambda e^{-\lambda x}$ or $\frac{1}{\beta} e^{-\frac{1}{\beta}x}$.

Since β means the average of time period a event takes, the average service time $\frac{1}{\mu_1}$ and $\frac{1}{\mu_2}$ should be β, not λ.

Therefore, $P(\{x \le t\}) = 1 - e^{-\frac{t}{\beta}} = 1 - e^{-\mu_1 t}$, $P(\{x \ge t\}) = e^{-\mu_1 t}$.

$P(\{x_1 \ge 10\}) = e^{-10\mu_1}$ #

2. $\int_0^\infty \mu_2 e^{-\mu_2 t} \cdot (e^{-\mu_1 t})\, dt$

$= \mu_2 \cdot \int_0^\infty e^{-(\mu_1+\mu_2)t}\, dt$

$= \mu_2 \cdot \frac{1}{-(\mu_1+\mu_2)} e^{-(\mu_1+\mu_2)t} \Big|_0^\infty$

$= \mu_2 \cdot \frac{1}{\mu_1+\mu_2}$ #

John lives in A city and goes to work every morning by taking one train and then connecting to a local bus in B city. To avoid being late for work, he must arrive no later than 8h30. John always takes the train at 8h00. The trajectory between A and B takes **exactly 10 minutes**. According to the long-term observation, the train is of a delay probability distribution as the table below:

delay (min)	4	6	8	10	12
probability	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{16}$

Another uncertainty to his office is that the bus is not stable either. Averagely, a bus arrives at the train station of B city to pick up passengers at 8h20 and then the trip is **exactly 10 minutes** to arrive John's office. According to another survey, the bus schedule is following the Normal distribution with the average departure time at 8h20 from the train station and a **standard deviation of 2 minutes**.
Assume the train delay and the bus uncertainty are independent. What is the probability that John will be late for work?

Question 5:

Since both the bus and the train takes exactly 10 mins, so John has only $30 - 20 = 10$ mins to waste. Assuming that the 68-95-99.7 rule is implemented.

If the train delays for 4 mins, the probability that the bus delays for less than 6 mins is $50 - \frac{100-99.7}{2} = 49.85\%$. (3 s.d.)

6 mins — 4 mins is $50 - \frac{100-95}{2} = 47.5\%$ (2 s.d.)

8 mins. — 2 mins is $50 - \frac{100-68}{2} = 34\%$ (1 s.d.)

The successful rate is $\frac{1}{16} \times 49.85\% + \frac{1}{8} \times 47.5\% + \frac{1}{2} \times 34\% \doteq 26.053\%$ #

Therefore, the answer $= 1 - 26.053\% = 73.947\%$.

6. Assume in average 1 out of 100 people has cancer (1%) and the cancer detection rate with current X-ray scan is accurate up to 99%. One day, John is diagnosed as positive in the X-ray cancer scan.
   a. What is the probability that John really has the cancer?
   b. Since John knows Bayes' theorem very well, he decides to make an MRI check to know more details of his health. It is known that MRI check has 99.9% accuracy detecting cancer. Unfortunately, John is diagnosed again positive after the MRI check. What is the probability that John has the cancer now?

   * Hint: Bayes' Theorem

Question 6:

1. 在X-ray中,會出現[有病+驗對]、[沒病+驗錯] 2種情型

False Positive Rate : $99\% \times 1\% = 0.0099$

True Positive rate $= 1\% \times 99\% = 0.0099$

Ans. $\frac{0.0099}{0.0099+0.0099} = 50\%$

2. 在X-ray scan rate中被測試為陽性者: MRI複驗結果如下:

False positive Rate $= 50\% \times 0.1\% = 0.0005$

True positive Rate $= 50\% \times 99.9\% = 0.4995$, $\frac{0.4995}{0.0005+0.4995} = 99.9\%$

The amount of time that a customer spends waiting at an airport check-in counter is a random variable with mean 8.5 minutes and standard deviation 3.5 minutes. Suppose that a random sample of $n = 49$ customers is observed. Find the probability that the "average time waiting in line for these 49 customers" is

a. less than 10 minutes;
b. between 7 and 10 minutes;
c. less than 7.5 minutes;

* Hint: Central Limit Theorem

Question 7.

CLT: If we randomly choose a lot of independent random variables from any population, it's mean value will approach to normal distribution. $\frac{\sqrt{n}(\bar{x}_i - \mu_0)}{\sigma} \sim N(0,1)$, $\frac{\sqrt{49}(\bar{x}_i - 8.5)}{3.5} \sim N(0,1)$.

(a) less than 10 minute: $\bar{x}_i = 10$, $\frac{7(10-8.5)}{3.5} = 3 = z$ score, 查表 ⇒ 99.87% #

(b). " 7 " : $\bar{x}_i = 7$, $\frac{7(7-8.5)}{3.5} = -3 = z$ score 查表 $= 0.0013$, $99.87\% - 0.13\% = 99.74\%$ #

(c). 7.5 $\bar{x}_i = 7.5$, $\frac{7(7.5-8.5)}{3.5} = -2$, z score, 查表 $= 0.0228 \Rightarrow 2.28\%$ #

Question 8:

8. The proportion of people living in A city who are iPhone users is estimated to be $p = 0.4$. To test this hypothesis, a random survey of 600 people is conducted. After the statistical analysis, we decide that if the number of iPhone users is between 216 to 264, the hypothesis will be accepted; otherwise, we will conclude that $p \neq 0.4$. Please find the **type I error probability** for this analysis procedure, assuming that $p = 0.4$ for real.

* Hint: z-test

Type I error: H_0 hypothesis is True but the statistics falls in reject zone, thus we accidently reject H_0.

$H_0: p_0 = 0.4$ 又已知 According to CLT, T_n的極限分配是 $N(0,1)$.

If the number of iPhone user is 264, then $S_n^2 = \frac{n}{n-1}(\hat{p}_n)(1-\hat{p}_n) = \frac{600}{599} \times \left(\frac{264}{600}\right) \times \left(\frac{336}{600}\right) \doteq 0.2468$

$T_n = \frac{\sqrt{600}\left(\frac{264}{600} - \frac{240}{600}\right)}{\sqrt{0.2468}} = \sqrt{\frac{600}{0.2468}} \cdot \frac{24}{600} \doteq 1.9726 = z$ score, 查表 ⇒ 97.573%

If the number of iPhone use is 216, then $S_n^2 = \frac{600}{599} \times \frac{216}{600} \times \frac{384}{600} \doteq 0.2308$.

$T_n = \frac{\sqrt{600}\left(\frac{216}{600} - \frac{240}{600}\right)}{\sqrt{0.2308}} \doteq -2.0395 = z$ score, 查表 ⇒ 2.073%

Ans: $(100 - 97.573) + 2.073 = 4.5\%$ #

Question 9.

Define X as the number of under-filled beer bottles from a filling operation in a carton of 24 bottles. 75 cartons are inspected and the following observation on X are recorded:

X	0	1	2	3
Frequency	39	23	12	1

Based on these 75 observations, is a "Binomial distribution" an appropriate model? Perform a Goodness-of-Fit procedure with $\alpha = 0.05$.

1. Goodness-of-fit: whether a sample obey one theorical distribution 若需 obey 某 distribution, 則此分配的期望值應和實際分配相似 (Condition: $\alpha = 0.05$). 由於二項分佈所需之 n, p 皆很可求得各情況下的期望值 ⇒ Yes!

Step 1: the number of under-filled beer bottles: $0\times39+1\times23+2\times12+3\times1=48$.

all bottle $=75\times24=1800$, so the probability of occuring a under-filled beer bottle: $\frac{48}{1800}=\frac{2}{75}$.

We assume that $X\sim B(24,\frac{2}{75})$。

H_0: the distribution obeys $B(24,\frac{2}{75})$.

H_1: " does not obeys $B(24,\frac{2}{75})$.

x_i	O_i	$P(X=x_i)$	E_i	(O_i-E_i)	$(O_i-E_i)^2$	$\frac{(O_i-E_i)^2}{E_i}$
0	39	$(\frac{73}{75})^{24}\doteq.523$	$75\times52.3\%\doteq39.2$	0.2	0.04	0.001
1	23	$C^{24}_1\cdot\frac{2}{75}\cdot(\frac{73}{75})^{23}\doteq34.2\%$	25.6	-2.6	6.76	0.264
2	12	$C^{24}_2\cdot(\frac{2}{75})^2\cdot(\frac{73}{75})^{22}\doteq8.1\%$	8.1	3.9	15.21	1.878
3	1	$C^{24}_3(\frac{2}{75})^3(\frac{73}{75})^{21}\doteq1.5\%$	1.5	-0.5	0.25	0.167
4	0	$C^{24}_4(\frac{2}{75})^4(\frac{73}{75})^{21}\doteq0.3\%$.	0.2	-0.2	0.04	0.2
						⋮
						$\sum x^2\doteq2.51$.

degree of freedom $=5-1=4$.

when significance level $=5\%$, critical values $=9.49$.

$\because 2.51<9.49$ $\therefore$ we see that there is no evidence to reject the null hypothesis. Thus, we retain H_0. and conclude that there is insufficient evident to suggest that the distribution isn't binomial.

Question 10.

Two courses: Probability & Statistics (Prob) and Operations Research (OR), are given in the same semester to the same group of students. Suppose we have 100 students with the following grades summary.

Prob \ OR	A	B	C	
A	15.75 24	13.5 11	15.75 10	45%
B	8.75 7	7.5 13	8.75 5	25%
C	10.5 4	9 6	10.5 20	30%
	35%	30%	35%	

Please setup a hypothesis testing with $\alpha = 0.01$ to conclude if the grades in Prob and OR are related. (Grades are classified into Three levels, where A: excellent; B: average; C: failed.)

* Hint: χ^2 test

H_0: two variables are independent, sampling distribution $=\chi^2$ distribution.

$df=(3-1)(3-1)=4$, $\alpha=0.01$, $\chi^2_{(critical)}=13.277$

若OR和Prob無关, 則不管學生在OR獲得A.B.C, 他們在Prob的A.B.C比例就會和OR成績的邊緣總數比例相同。⇒ expect: 如表上藍字。

$\chi^2_{(obtained)}=\sum(f_o-f_e)^2/f_e=$

$(15.75-24)^2/15.75+(8.75-7)^2/8.75+(10.5-4)^2/10.5+(13.5-11)^2/13.5+(7.5-13)^2/7.5+(9-6)^2/9$

$+(15.75-10)^2/15.75+(8.75-5)^2/8.75+(10.5-20)^2/10.5=26.493126$

$\because 26.493\ldots>13.277$

$\therefore$ We accept H_0 and Prob and OR are not related.