Question 1

A lady asserts that by tasting a cup of tea made with milk she can determine whether the milk or the tea infusion was first added to the cup. Suppose that an experiment is designed to allow the lady to provide evidence in support her claim. 10 cups of tea are made and presented to the lady in a random order. 5 of the cups are made with tea first, and the other 5 are made with milk first. The lady is tasked with identifying which 5 of the 10 cups are made with tea first.

(a) What is the null hypothesis in this experiment?

(b) What is the probability of the lady correctly selecting all 5 tea first cups by pure chance?

(c) **(Optional)** Obtain an upper bound on the probability in (b) to three decimal places.

(d) What is the probability that, out the 5 cups the lady selects, exactly 4 are made with tea first, and 1 is made with milk first?

(e) What is the probability that out of the 5 cups the lady select, at least 4 of the 5 cups are made with tea first?

(f) **(Optional)** Obtain a lower bound on the probability in (e) to one decimal place.

(g) Suppose the experiment takes place and, out of the 5 cups the lady selects, exactly 4 are made with tea first. Would you accept the lady's claim that she can taste the difference between the two processes of making tea?

Solution to Question 1

Part (a) :

The null hypothesis is :

$$H_0 : \text{The lady has no ability to discriminate between the two processes for making tea.}$$

Of course different words can be used, e.g. 'discriminate' can be swapped with 'distinguish', etc. But the important point is that the null hypothesis is that the lady **no ability** to taste the difference between the cups of tea.

When the experiment is run, if the cups she identifies as being made with tea first actually have a high proportion of tea-first cups, this might provide evidence to reject the null hypothesis and therefore support her assertion that she is able to taste the difference between the two ways of making tea with milk.

Part (b) :

Since this is a fundamental problem in probablity theory, although you have seen the answer before in the first term, we shall go through the calculation again carefully.

First, let's count how many ways there are choosing a group of 5 cups out of a total of 10 cups.

- Choosing the five cups sequentially, There are 10 choices for the first cup, then 9 choices for the second, 8 choices for the third, 7 choices for the fourth and 6 for the fifth; in other words, $10 \times 9 \times 8 \times 7 \times 6 = 30240$ ways of sequentially choosing the five cups.
- However, this number counts the ways of choosing the five cups in a specific order, i.e. choosing the five cups in the order $(8, 3, 7, 6, 1)$ is counted separately to the choosing the five cups in the order $(1, 3, 6, 7, 8)$.
- The number of different ways for ordering five cups is (by the same argument above) $5 \times 4 \times 3 \times 2 \times 1 = 5! = 120$.
- Therefore, there are

$$\frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1} = \frac{30240}{120} = 252$$

ways of choosing 5 out of 10 cups.

Of course, the above calculation can be written

$$\frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2 \times 1} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times (5 \times 4 \times 3 \times 2 \times 1)}{5 \times 4 \times 3 \times 2 \times 1 \times (5 \times 4 \times 3 \times 2 \times 1)} = \frac{10!}{5!5!} = \binom{10}{5} = 252$$

Where $\binom{10}{5}$ is the familiar 'choose' notation; this gives clear meaning to why we read that symbol as '10 choose 5'.

However, we are not yet done. We still need to compute how many ways there are for choosing the cups so that the 5 cups that are chosen are exactly the 5 cups that are made with tea first. In this case, there is exactly 1 way to make this choice.

This is equivalent to saying the number of ways of the 5 chosen cups being the 5 tea-first cups is: the number of ways of choosing 5 out of the 5 tea-first cups is $\binom{5}{5}$, and the number of ways of choosing 0 out of the 5 milk-first cups is $\binom{5}{0}$.

Therefore, the number of ways of choosing the the 5 cups so that exactly 5 are tea-first is

$$\binom{5}{5}\binom{5}{0} = 1 \cdot 1 = 1.$$

Therefore, the probability of the 5 chosen cups all being tea-first is $p = \frac{1}{252}$.

Model answer Although this is a long explanation, a model answer to such a question would be much shorter, such as the following:

The number of ways of choosing 5 cups out of 10 is $\binom{10}{5}$.

The number of ways of choosing the 5 cups so that all 5 are tea-first, is the same as the number of ways of choosing 5 out of 5 tea-first cups, multiplied by the number of ways of choosing 0 out of 5 milk-first cups, i.e. $\binom{5}{5}\binom{5}{0}$.

Therefore, the probability of choosing 5 cups out of 10 so that all 5 cups are the 5 tea-first cups is

$$p = \frac{\binom{5}{5}\binom{5}{0}}{\binom{10}{5}} = \frac{1}{252}.$$

Part (c): (Optional)

$$\frac{1}{252} = \frac{4}{1008} < \frac{4}{1000} = 0.004$$

(Can actually show $\frac{1}{252} \in (0.00396, 0.004)$, if desired.)

Part (d):

Using the explanation in Part (b),

$$p = \frac{\binom{5}{4}\binom{5}{1}}{\binom{10}{5}} = \frac{5 \cdot 5}{252} = \frac{25}{252}.$$

Part (e):

Adding the probabilities from Parts (d) and (e):

$$p = \frac{\binom{5}{4}\binom{5}{1}}{\binom{10}{5}} + \frac{\binom{5}{5}\binom{5}{0}}{\binom{10}{5}} = \frac{25+1}{252} = \frac{26}{252}.$$

Part (f): (Optional)

$$\frac{26}{252} > \frac{26}{260} = 0.1.$$

Part (g):

Computing the p-value exactly,

$$p = \frac{26}{252} = 0.103.$$

There is only about a 10% chance of the lady having chosen at least 4 out of the 5 tea-first cups by pure chance.

Whether or not this is statistically significant enough for **YOU** is, literally, up to you.

However, for historical reasons, p-values above 0.05 are generally not regarded as statistically significant.

Question 2

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed according to a normal distribution with unknown mean μ and known variance $\sigma^2 = 9$. Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is observed as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, where

$$\sum_{i=1}^{n} x_i = 740, \qquad n = 100.$$

Using the tables in your notes, test the hypothesis that $\mu = 8$ at the significance level $\alpha = 0.05$.

Solution to Question 2

If one considers the estimator for the sample mean $\overline{X}$, recall that by linearity of expectation,

$$\mathrm{E}\left(\overline{X}\right) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

since all $\mathrm{E}\left(X_i\right) = \mu$ by assumption, although the value of μ is unknown. We also know, using properties of the variance, and since the random variables are independent, that

$$\mathrm{Var}\left(\overline{X}\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

Therefore, the random variable

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathrm{N}\left(0, 1\right)$$

Assuming the null hypothesis

$$H_0 : \mu = 8$$

is true, since

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{100}(740) = 7.4, \qquad \frac{\sigma}{\sqrt{n}} = \frac{3}{10} = 0.3,$$

the realisation of Z is

$$z = \frac{7.4 - 8}{0.3} = -\frac{0.6}{0.3} = -2.$$

Looking at our table which shows

$$\begin{aligned}
\mathrm{P}\left(Z \leq 1.96\right) &= 0.975 \\
\Rightarrow \mathrm{P}\left(Z \leq -1.96\right) &= 0.025 \\
\Rightarrow \mathrm{P}\left(Z < -1.96\right) &= 0.025 \\
\Rightarrow \mathrm{P}\left(-1.96 \leq Z \leq 1.96\right) &= 0.95 \\
\Rightarrow \mathrm{P}\left(|Z| \leq 1.96\right) &= 0.95 \\
\Rightarrow \mathrm{P}\left(|Z| > 1.96\right) &= 0.05
\end{aligned}$$

Therefore, since $|z| = 2 > 1.96$, we would reject the null hypothesis at the level $\alpha = 0.05$.

Alternatively, one could compute $\mathrm{P}\left(|Z| \geq 2\right) = 0.0455 < \alpha = 0.05$, and so we would reject the null hypothesis at significance level $\alpha = 0.05$.

Question 3

Suppose that the random variables $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed according to a normal distribution with unknown mean μ and unknown variance σ^2. Suppose that $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ is observed as $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, where

$$\sum_{i=1}^{n} y_i = 32, \qquad \sum_{i=1}^{n} y_i^2 = 124, \qquad n = 16.$$

(a) Test the hypothesis that $\mu = 0.9$ at the significance level $\alpha = 0.01$.

(b) If this hypothesis is not rejected at $\alpha = 0.01$, find the smallest significance level at which it is rejected.

Solution to Question 3

Part (a):

Since the variance σ^2 is unknown, if S^2 is the sample variance of the random variables $Y_1, Y_2, \ldots, Y_n$, then

$$T = \frac{\overline{Y} - \mu}{\frac{S}{\sqrt{n}}}$$

follows a t-distribution with $n - 1 = 15$ degrees of freedom.

One can compute the sample mean as

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{16}(32) = 2,$$

and the sampel variance as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2 \right) = \frac{1}{15} \left(124 - 16(2)^2\right) = \frac{1}{15} (124 - 64) = \frac{1}{15} (60) = 4.$$

Therefore, assuming the null hypothesis $\mu = 0.9$ is true, T is realised as

$$t = \frac{\overline{y} - \mu}{\frac{s}{\sqrt{n}}} = \frac{2 - 0.9}{\frac{2}{\sqrt{16}}} = \frac{1.1}{0.5} = 2.2.$$

The table gives us that for 15 degrees of freedom, and $1 - \alpha = 1 - 0.01 = 0.99$,

$$\mathrm{P}\left(|T| \leq 2.947\right) = 0.99$$

Since $t = 2.2 < 2.947$, we would not reject the null hypothesis at level $\alpha = 0.01$.

Part (b):

The table gives us that for 15 degrees of freedom, and $\alpha = 0.05$,

$$\mathrm{P}\left(|T| \leq 2.131\right) = 0.95$$

Since $t = 2.2 > 2.131$, we would reject the null hypothesis at level $\alpha = 0.05$.