Question 1

Suppose that $X_1, X_2, \ldots, X_n$ follow a $\mathrm{N}\left(\theta_1, \sigma^2\right)$ distribution where θ_1 is unknown and σ^2 is unknown, and the independent random variables $Y_1, Y_2, \ldots, Y_m$ follow a $\mathrm{N}\left(\theta_2, \sigma^2\right)$ distribution where θ_2 is unknown (note that both sets of random variables are assumed to have the same unknown variance σ^2). We further assume that each X_i is independent of each Y_j.

Use Theorem 3.4.4 in the notes to show that if we assume $\mu_1 = \mu_2$ then the statistic

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

where

$$S_p^2 = \frac{1}{n+m-2}\left((n-1)S_X^2 + (m-1)S_Y^2\right),$$
$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \qquad S_Y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2.$$

follows Student's t-distribution with $n+m-2$ degrees of freedom, i.e. $T \sim t_{n+m-2}$.

Solution to Question 1

Theorem 3.4.4 states that if $U \sim \mathrm{N}(0,1)$ and $V \sim \chi^2_\nu$, and U and V are independent random variables, then the random variable T defined by

$$T = \frac{U}{\sqrt{V/\nu}},$$

follows t_ν.

First, we use Theorem 3.2.2 to recall that

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2_{n-1},$$
$$\frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{m-1}.$$

Furthermore, since the X_i and the Y_j random variables are independent, $\frac{(n-1)S_X^2}{\sigma^2}$ and $\frac{(m-1)S_Y^2}{\sigma^2}$ are independent and so their sum

$$V = \frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} = \frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

since $(n-1)+(m-1) = n+m-2$. This can be shown using the property of sums of independent chi-squared random variables, or by using moment generating functions. Furthermore, V is independent of the X_i and Y_j random variables, and so is independent of $\overline{X} - \overline{Y}$. Now, let us rewrite T:

$$T = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$= \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n+m-2)S_p^2}{\sigma^2} \cdot \frac{\sigma^2}{(n+m-2)}\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

$$= \frac{\overline{X} - \overline{Y}}{\sqrt{V \cdot \frac{1}{(n+m-2)}}\sqrt{\sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

Now, since

$$\mathrm{E}[\overline{X}-\overline{Y}] = \mathrm{E}[\overline{X}] - \mathrm{E}[\overline{Y}] = \mu_1 - \mu_2 = 0$$

since we assume $\mu_1 = \mu_2$, and

$$\begin{aligned}
\mathrm{Var}\left(\overline{X}-\overline{Y}\right) &= \mathrm{Var}\left(\overline{X}\right) + \mathrm{Var}\left(\overline{Y}\right) \\
&= \frac{\sigma^2}{n} + \frac{\sigma^2}{m} \\
&= \sigma^2\left(\frac{1}{n}+\frac{1}{m}\right),
\end{aligned}$$

we therefore have that

$$\begin{aligned}
U &= \frac{\overline{X}-\overline{Y}}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}} \\
\Rightarrow U &\sim \mathrm{N}\left(0,1\right).
\end{aligned}$$

So,

$$\begin{aligned}
T &= \frac{\overline{X}-\overline{Y}}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}} \cdot \frac{1}{\sqrt{V \cdot \frac{1}{(n+m-2)}}} \\
&= \frac{U}{\sqrt{V/(n+m-2)}}
\end{aligned}$$

and so by Theorem 3.4.4 $T \sim t_{n+m-2}$.

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 $\overline{X}$, again we have

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Therefore, the random variable

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

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

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

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

Alternatively, one could compute $\mathrm{P}(|Z| \geq 2) = 0.0455 < \alpha = 0.05$, 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 sample 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 $1 - \alpha/2 = 0.95$, $t_{15,0.995} = 2.947$

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

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 $\alpha = 0.05$, (so $1 - \alpha/2 = 0.975$), $t_{15,0.975} = 2.131$,

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

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