Midterm Solutions

1-(a). We have
$$\mathbb{E}(\hat{\Delta}) = \mathbb{E}(\bar{X}) - \mathbb{E}(\bar{Y}) = \mu - \theta = \Delta.$$
As $\bar{X}$ and $\bar{Y}$ are independent, we have
$$\mathsf{Var}(\hat{\Delta}) = \mathsf{Var}(\bar{X}) + \mathsf{Var}(\bar{Y}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m} = \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right).$$
Hence,
$$\mathbb{E}(\hat{\Delta} - \Delta)^2 = \mathsf{Var}(\hat{\Delta}) + (\mathbb{E}(\hat{\Delta}) - \Delta)^2 = \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right).$$

1-(b). We have learned that
$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{and} \quad \frac{\sum_{j=1}^m (Y_j - \bar{Y})^2}{\sigma^2} \sim \chi^2_{m-1}.$$
Also,
$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \quad \text{and} \quad \frac{\sum_{j=1}^m (Y_j - \bar{Y})^2}{\sigma^2} \quad \text{are independent.}$$
Hence,
$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

1-(c). As $\mathbb{E}\chi^2_{n+m-2} = n + m - 2$, we have
$$\mathbb{E}\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{n + m - 2}\right) = \sigma^2.$$

1-(d). We show that $(X_1 - \bar{X}, \ldots, X_n - \bar{X}, Y_1 - \bar{Y}, \ldots, Y_n - \bar{Y}, \bar{X} - \bar{Y}) \in \mathbb{R}^{n+m-1}$ follows a normal distribution. To see this, observe that
$$\begin{bmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \\ Y_1 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \\ \bar{X} - \bar{Y} \end{bmatrix} = \underbrace{\begin{bmatrix} I_n - \frac{1}{n} 1_n 1_n^T & 0_{n\times m} \\ 0_{m\times n} & I_m - \frac{1}{m} 1_m 1_m^T \\ \frac{1}{n} 1_n^T & -\frac{1}{m} 1_m^T \end{bmatrix}}_{=:A} \begin{bmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \end{bmatrix},$$
where $A \in \mathbb{R}^{(n+m+1)\times(n+m)}$ is a block matrix, where $I_n \in \mathbb{R}^{n\times n}$ and $I_m \in \mathbb{R}^{m\times m}$ are identity matrices, $0_{n\times m} \in \mathbb{R}^{n\times m}$ and $0_{m\times n} \in \mathbb{R}^{m\times n}$ are zero matrices, and $1_n \in \mathbb{R}^n$ and $1_m \in \mathbb{R}^m$ are

vectors of all ones. Since

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \end{bmatrix} \sim N\left(\begin{bmatrix} \mu 1_n \\ \theta 1_m \end{bmatrix}, \begin{bmatrix} \sigma^2 I_n & 0_{n\times m} \\ 0_{m\times n} & \sigma^2 I_m \end{bmatrix} \right),$$

we have

$$\begin{bmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \\ Y_1 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \\ \bar{X} - \bar{Y} \end{bmatrix} \sim N\left(A \begin{bmatrix} \mu 1_n \\ \theta 1_m \end{bmatrix}, \sigma^2 \begin{bmatrix} I_n - \frac{1}{n} 1_n 1_n^T & 0_{n\times m} & 0 \\ 0_{m\times n} & I_m - \frac{1}{m} 1_m 1_m^T & 0 \\ 0 & 0 & \frac{1}{n} + \frac{1}{m} \end{bmatrix} \right).$$

From the covariance matrix of $(X_1 - \bar{X}, \ldots, X_n - \bar{X}, Y_1 - \bar{Y}, \ldots, Y_n - \bar{Y}, \bar{X} - \bar{Y})$, we can see that

$$(X_1 - \bar{X}, \ldots, X_n - \bar{X}, Y_1 - \bar{Y}, \ldots, Y_n - \bar{Y}) \quad \text{and} \quad \bar{X} - \bar{Y} \quad \text{are independent.}$$

Therefore, $\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2$ and $\bar{X} - \bar{Y}$ are independent.

1-(e). Since

$$\hat{\Delta} = \bar{X} - \bar{Y} \sim N\left(\mu - \theta, \sigma^2 \left(\frac{1}{n} + \frac{1}{m} \right) \right),$$

we have

$$\frac{\bar{X} - \bar{Y} - (\mu - \theta)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1).$$

Meanwhile, from part (b), we have

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

By the independence shown in part (d), we have

$$\frac{\frac{\bar{X} - \bar{Y} - (\mu - \theta)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{\sigma^2 (n+m-2)}}} \sim t_{n+m-2}.$$

Therefore,

$$\sqrt{\frac{n+m-2}{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y} - (\mu - \theta)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}} \sim t_{n+m-2}.$$

Under the null $H_0 : \mu = \theta$, we have

$$\sqrt{\frac{n+m-2}{\frac{1}{n}+\frac{1}{m}}}\frac{\bar{X}-\bar{Y}}{\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2+\sum_{j=1}^m (Y_j-\bar{Y})^2}} \sim t_{n+m-2}.$$

Now, let $t_{n+m-2,1-\frac{\alpha}{2}}$ be the $1-\frac{\alpha}{2}$ quantile of t_{n+m-2} distribution. Then, we have a level-α test by rejecting the null H_0 if

$$\left|\sqrt{\frac{n+m-2}{\frac{1}{n}+\frac{1}{m}}}\frac{\bar{X}-\bar{Y}}{\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2+\sum_{j=1}^m (Y_j-\bar{Y})^2}}\right| > t_{n+m-2,1-\frac{\alpha}{2}}.$$

Remark From part (b), one may notice that

$$\mathbb{E}\left(\sum_{i=1}^n (X_i-\bar{X})^2\right) = (n-1)\sigma^2 \quad \text{and} \quad \mathbb{E}\left(\sum_{j=1}^m (Y_j-\bar{Y})^2\right) = (m-1)\sigma^2.$$

From this, one can deduce that

$$\mathbb{E}\left(\frac{1}{2}\times\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1}+\frac{1}{2}\times\frac{\sum_{j=1}^m (Y_j-\bar{Y})^2}{m-1}\right) = \sigma^2.$$

Therefore,

$$\frac{1}{2}\times\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1}+\frac{1}{2}\times\frac{\sum_{j=1}^m (Y_j-\bar{Y})^2}{m-1}$$

is an unbiased estimator of σ^2 using both X_i's and Y_j's. In fact, for any constant $\lambda \in (0,1)$,

$$\lambda\times\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n-1}+(1-\lambda)\times\frac{\sum_{j=1}^m (Y_j-\bar{Y})^2}{m-1}.$$

Letting $\lambda = \frac{n-1}{n+m-2}$ gives the above answer

$$\frac{\sum_{i=1}^n (X_i-\bar{X})^2+\sum_{j=1}^m (Y_j-\bar{Y})^2}{n+m-2}.$$

While any $\lambda \in (0,1)$ is valid for part (c), the choice $\lambda = \frac{n-1}{n+m-2}$ desirable as the resulting estimator follows the chi-squared distribution as shown in part (b), which is supposed to used in part (e) to construct a test based on t distribution. If you try to solve part (e) with other λ, e.g., $\lambda = \frac{1}{2}$, then it is not a valid approach as we no longer have a chi-squared distribution.