Solution_hw1

2. (a) From the theorem mentioned in class, we know Y_1+2Y_2 is still normal with

mean $\mu = E(Y_1+2Y_2) = E(Y_1)+2E(Y_2) = 0+2\cdot 1 = 2$

& variance $\sigma^2 = V(Y_1+2Y_2) = V(Y_1)+2^2V(Y_2)+2\times 2\,Cov(Y_1, Y_2)$

$= 1+2^2\cdot 4+2\times 2\times 1 = 21$

OR: you can get the mean & variance directly from the result in the theorem, $Y_1+2Y_2 = (1 \;\; 2)\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} =: C\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$

So, mean $= C\begin{pmatrix} 0 \\ 1 \end{pmatrix} = (1 \;\; 2)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 2$

variance $= C\begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}C' = (1 \;\; 2)\begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 21$

(b) Since Y_1+2Y_2 & $-3Y_1+Y_2$ are both normal, then independent $\Leftrightarrow$ uncorrelated

Calculate $Cov(Y_1+2Y_2, -3Y_1+Y_2)$

$= -3V(Y_1)+2V(Y_2)+(1-2\times 3)\,Cov(Y_1, Y_2)$

$= -3\times 1+2\times 4-5\times 1 = 0$

So, Y_1+2Y_2 & $-3Y_1+Y_2$ are uncorrelated, thus independent.

3. Since Y_1 & Y_2 are independent normal, then

Y_1+Y_2 are normal with mean $= \mu_1+\mu_2 = 0$

& variance $= \sigma_1^2+\sigma_2^2 = 1+4 = 5$

Y_1-2Y_2 are normal with mean $= \mu_1-2\mu_2 = 0$

& variance $= \sigma_1^2+2^2\sigma_2^2 = 1+2^2\times 4 = 17$

$Cov(Y_1+Y_2, Y_1-2Y_2) = V(Y_1)-2V(Y_2)+(-2+1)\,Cov(Y_1+Y_2)$

$= 1-2\times 4-1\times 0 = -7$

$\Rightarrow \begin{pmatrix} Y_1+Y_2 \\ Y_1-2Y_2 \end{pmatrix}$ are normal distribution with mean $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

& variance-covariance matrix $\Sigma = \begin{pmatrix} 5 & -7 \\ -7 & 17 \end{pmatrix}$

<u>OR</u>: $\begin{pmatrix} Y_1+Y_2 \\ Y_1-2Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} =: C\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$

So, $\begin{pmatrix} Y_1+Y_2 \\ Y_1-2Y_2 \end{pmatrix}$ is normal with

mean $C\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

& variance - covariance matrix $C\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}C'$

$$= \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 1 & \\ & 4 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 5 & -7 \\ -7 & 17 \end{pmatrix}$$

4 (a) Since $\begin{pmatrix} Y_1 \\ Y_3 \end{pmatrix}$ is a subvector of Y, it's still normal with

$N\left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 9 \end{pmatrix}\right)$

(b) $\begin{pmatrix} Y_1+Y_2+2 \\ Y_2-Y_3-3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} + \begin{pmatrix} 2 \\ -3 \end{pmatrix} =: C\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} + b$

It's normal with mean $C\mu+b = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$

& variance $C\Sigma C' = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 9 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}$

8.68 The 95% confidence interval, based on $n-1 = 20$ degrees of freedom, is

$$\bar{y} \pm t_{0.025}\left(\frac{s}{\sqrt{n}}\right) = 26.6 \pm 2.086\left(\frac{7.4}{\sqrt{21}}\right)$$

$$= 26.6 \pm 3.37 = (23.23,\ 29.97)$$

8.76 a. Let μ_1 = mean verbal score for engineering students
& μ_2 = ... language / Literature students
Then the 95% confidence interval for the difference is

$$(\bar{y}_1 - \bar{y}_2) \pm t_{0.025}\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $t_{0.025} = 2.048$ with degrees of freedom $n_1 + n_2 - 2 = 28$

$$\& \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{14(42)^2 + 14(45)^2}{28} = 1894.5$$

$$\Rightarrow 95\% \text{ CI is } 446 - 534 \pm 2.048\sqrt{1894.5\left(\frac{1}{15} + \frac{1}{15}\right)}$$

$$= -88 \pm 32.55 = (-120.55,\ -55.45)$$

b. Similar to a

Let μ_1 = mean math score for engineering students
& μ_2 = ... language / Literature students

$$\text{Here, } S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{14(57)^2 + 14(52)^2}{28} = 2976.5$$

$$\Rightarrow 95\% \text{ CI is } 548 - 517 \pm 2.048\sqrt{2976.5\left(\frac{1}{15} + \frac{1}{15}\right)}$$

$$= 31 \pm 40.80 = (-9.80,\ 71.80)$$

c. The 95% CI in (a) indicates a significant difference between the mean verval scores for students in engineering & language / Literature since both endpoints of the interval are negative.

The 95% CI in (b) indicates that students in engineering have the tendency to have higher mean math scores than students in language / literature, but the difference is not so significant since 0 is in the interval.

d. We assume that the verval / math scores for the two groups are randomly & independently selected from two normal distributions with common variance.

10.53 a. The hypothesis to be tested is

$H_0: \mu = 45 \qquad H_a: \mu < 45$

Calculate $\bar{y} = \Sigma y_i / n = 712.01 / 18 = 39.556$

$$s^2 = \frac{\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}}{n-1} = \frac{29040.4275 - \frac{(712.01)^2}{18}}{17}$$

$$= 50.94593 \Rightarrow s = 7.138$$

The t-statistic is $t = \dfrac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} = \dfrac{39.556 - 45}{\frac{7.138}{\sqrt{18}}} = -3.24$

Under H_0, t follows Student's t distribution with $n-1 = 17$ degrees of freedom, then we have p-value $= P(t_{17} < -3.24) < 0.005$ is very small. Therefore, we reject H_0 & conclude that the average is significantly less than 45¢

b. The 95% CI, based on 17 degrees of freedom, is

$$\bar{y} \pm t_{0.025} \frac{s}{\sqrt{n}} = 39.556 \pm 2.110 \left(\frac{7.138}{\sqrt{18}}\right) = 39.556 \pm 3.550$$

$$= (36.006, 43.106)$$

10.58 The hypothesis to be tested is

$H_0: \mu_1 - \mu_2 = 0$ $\quad H_a: \mu_1 - \mu_2 > 0$

Calculate $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \dfrac{9(.017)^2 + 12(.006)^2}{10+13-2}$

$= 0.0001443$

Then the t-test statistic is

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{.041 - .026}{\sqrt{s^2\left(\frac{1}{10}+\frac{1}{13}\right)}} = 2.97$$

The RR, with $\alpha = .05$ & 21 degrees of freedom is $t > 1.721$, thus the null hypothesis is rejected