STAT 420 – Homework 1

1. Meerkats

a. Let M = height of a random adult male meerkat; F = height of a random adult female meerkat. We want $P(M > F) = P(M - F > 0)$ $P(M - F > 0)$, so find the distribution of $M - F$. Since M, F are both Normal, then $M - F$ is Normal with

E[$M - F$] = 11.4 – 11.2 = 0.2,
Var[$M - F$] = $(2.1)^2 + (1.9)^2$ = 8.02, and
SD[$M - F$] = $(2.1)^2 + (1.9)^2$ = 2.83.

Thus,

$$P(M - F > 0) = P\left(Z > \frac{0 - 0.2}{2.83}\right) = P(Z > -0.07) = 1 - \Phi(-0.07) = 1 - 0.4721 = \mathbf{0.5279}.$$

Or, in R,

```
> 1 - pnorm(0, 0.2, 2.83)
[1] 0.5282
```

b. Let M = height of a random adult male meerkat; F = height of a random adult female meerkat. We want $P(M > F) = P(M - F > 0)$ $P(M - F > 0)$, so find the distribution of $M - F$. Since M, F are both Normal, then $M - F$ is Normal with

E[$M - F$] = 11.4 – 11.2 = 0.2,
Var[$M - F$] = $(2.1)^2 + (1.9)^2 + 2(1)(-1)(0.38)(2.1)(1.9)$ = 4.99, and
SD[$M - F$] = $(2.1)^2 + (1.9)^2$ = 2.23.

Thus,

$$P(M - F > 0) = P\left(Z > \frac{0 - 0.2}{2.23}\right) = P(Z > -0.09) = 1 - \Phi(-0.09) = 1 - 0.4641 = \mathbf{0.5359}.$$

Or, in R,

```
> 1 - pnorm(0, 0.2, 2.23)
[1] 0.5357
```

2. Stocks

a. We want $P(Y > 50)$. Since Y is Normal with E[Y] = 46 and SD[Y] = 3, then

$$P(Y > 50) = P\left(Z > \frac{50-46}{3}\right) = P(Z > 1.33) = 1 - \Phi(1.33) = 1 - 0.9082 = \mathbf{0.0918}.$$

Or, in R,

```
> 1 - pnorm(50, 46, 3)
[1] 0.09121
```

b. We want $P(Y > 50 \mid X = 100)$. Y is Normal with

$$\text{E}[Y \mid X] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) = 46 + 0.64 \cdot \frac{3}{4.5}(100 - 95) = 48.13,$$
$$\text{Var}[Y \mid X] = (1 - \rho^2)\sigma_Y^2 = (1 - 0.64^2)3^2 = 5.31, \text{ and}$$
$$\text{SD}[Y \mid X] = 2.31.$$

Thus,

$$P(Y > 50 \mid X = 100) = P\left(Z > \frac{50 - 48.13}{2.31}\right) = P(Z > 0.81) = 1 - 0.7910 = \mathbf{0.2090}.$$

Or, in R,

```
> 1 - pnorm(50, 48.13, 2.31)
[1] 0.2091
```

c. We want $P(4X + 5Y < 600)$. Since $4X + 5Y$ is a linear combination of Normal random variables, then $4X + 5Y$ is Normal with

$$\text{E}[4X + 5Y] = a\mu_X + b\mu_Y = 4(95) + 5(46) = 610$$
$$\text{Var}[4X + 5Y] = \text{Var}[4X] + 2 \cdot \text{Cov}[4X, 5Y] + \text{Var}[5Y]$$
$$= 4^2 \cdot (4.5)^2 + 2 \cdot (4)(5)(0.64)(4.5)(3.0) + 5^2 \cdot (3.0)^2 = 894.6$$
$$\text{SD}[4X + 5Y] = 29.91$$

Thus,

$$P(4X + 5Y < 600) = P\left(Z < \frac{600 - 610}{29.91}\right) = P(Z < -0.33) = \Phi(-0.33) = \mathbf{0.3707}.$$

Or, in R,

```
> pnorm(600, 610, 29.91)
[1] 0.3691
```

d. We want $P(X < 2Y) = P(X - 2Y < 0)$. Since $X - 2Y$ is a linear combination of Normal random variables, then $X - 2Y$ is Normal with

$$\mathrm{E}[X - 2Y] = a\mu_X + b\mu_Y = 1(95) - 2(46) = 3$$
$$\mathrm{Var}[X - 2Y] = \mathrm{Var}[X] - 2\cdot\mathrm{Cov}[X, 2Y] + \mathrm{Var}[2Y]$$
$$= (4.5)^2 - 2\cdot(1)(2)(0.64)(4.5)(3.0) + 2^2\cdot(3.0)^2 = 21.69$$
$$\mathrm{SD}[X - 2Y] = 4.66$$

Thus,

$$P(X - 2Y < 0) = P\left(Z < \frac{0-3}{4.66}\right) = P(Z < -0.64) = \Phi(-0.64) = \mathbf{0.2611}.$$

Or, in R,

```
> pnorm(0, 3, 4.66)
[1] 0.2599
```

3. Multivariate Normal

Suppose $\boldsymbol{X}$ follows a 3-dimensional multivariate normal distribution with mean $\boldsymbol{\mu} = \begin{bmatrix} 20 \\ 30 \\ 25 \end{bmatrix}$ and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 12 & 8 & -6 \\ 8 & 9 & -6 \\ -6 & -6 & 25 \end{bmatrix}$.

a. Find $P(X_2 > 32)$.

$X_2 \sim \mathrm{N}(30, 9)$

$$P(X_2 > 32) = P\left(Z > \frac{32-30}{3}\right) = P(Z > 0.67) = 1 - \Phi(0.67) = 1 - 0.7486 = \mathbf{0.2514}$$

Or, in R,

```
> 1 - pnorm(32, 30, 3)
[1] 0.2525
```

b. Find $P(X_3 > 32)$.

$X_3 \sim \mathrm{N}(25, 25)$

$$P(X_3 > 32) = P\left(Z > \frac{32-25}{5}\right) = P(Z > 1.40) = 1 - \Phi(1.40) = 1 - 0.9192 = \mathbf{0.0808}$$

Or, in R,

```
> 1 - pnorm(32, 25, 5)
[1] 0.08076
```

c. Find $P(X_1 + X_3 > 50)$.

$$\mathrm{E}[X_1 + X_3] = 20 + 25 = 45$$

$$\mathrm{Var}[X_1 + X_3] = \mathrm{Var}[X_1] + 2\cdot\mathrm{Cov}[X_1, X_3] + \mathrm{Var}[X_3] = 12 + 2(-6) + 25 = 25 \text{, or…}$$

$$\mathrm{Var}[X_1 + X_3] = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 12 & 8 & -6 \\ 8 & 9 & -6 \\ -6 & -6 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 & 2 & 19 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = 25$$

$$P(X_1 + X_3 > 50) = P\left(Z > \frac{50-45}{5} \right) = P(Z > 1) = 1 - \Phi(1.00) = 1 - 0.8413 = \mathbf{0.1597}$$

Or, in R,

```
> 1 - pnorm(50, 45, 5)
[1] 0.1587
```

d. Find $P(X_1 - X_3 > 0)$.

$$\mathrm{E}[X_1 - X_3] = 20 - 25 = -5$$

$$\mathrm{Var}[X_1 - X_3] = \mathrm{Var}[X_1] - 2\cdot\mathrm{Cov}[X_1, X_3] + \mathrm{Var}[X_3] = 12 - 2(-6) + 25 = 49 \text{, or…}$$

$$\mathrm{Var}[X_1 - X_3] = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 12 & 8 & -6 \\ 8 & 9 & -6 \\ -6 & -6 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 18 & 14 & -31 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = 49$$

$$P(X_1 - X_3 > 0) = P\left(Z > \frac{0-(-5)}{7} \right) = P(Z > 0.71) = 1 - \Phi(0.71) = 1 - 0.7611 = \mathbf{0.2389}$$

Or, in R,

```
> 1 - pnorm(0, -5, 7)
[1] 0.2375
```

e. Find $P(X_1 + 2X_2 + 3X_3 > 200)$.

$$\mathrm{E}[X_1 + 2X_2 + 3X_3] = 20 + 2\cdot 30 + 3\cdot 25 = 155$$

$$\mathrm{Var}[X_1 + 2X_2 + 3X_3] = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 12 & 8 & -6 \\ 8 & 9 & -6 \\ -6 & -6 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 & 8 & 57 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 197$$

$$\mathrm{SD}[X_1 + 2X_2 + 3X_3] = 14.04$$

$$P(X_1 + 2X_2 + 3X_3 > 200) = P\left(Z > \frac{200-155}{14.04}\right) = P(Z > 3.21) = 1 - 0.9993 = \mathbf{0.0007}$$

Or, in R,

```
> 1 - pnorm(200, 155, 14.04)
[1] 0.00068
```

4. Inference

a. First, calculate the pooled variance.

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(6-1)280 + (4-1)350}{6+4-2} = 306.25$$

Then, calculate the test statistic.

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{64-48}{\sqrt{306.25\left(\frac{1}{6} + \frac{1}{4}\right)}} = 1.416$$

Or, in R,

```
> tt <- t.test(x, y, alternative=c("two.sided"), var.equal=T)
> tt$statistic
    t
1.416
```

b. $df = n_1 + n_2 - 2 = 6 + 4 - 2 = 8$

Or, in R,

```
> tt$parameter
df
 8
```

c. p-value = $2 \cdot P(T > 1.416)$ since the alternative is two-sided.

From the t-table where $df = 8$, we see that $0.05 < P(T > 1.416) < 0.10$,

thus $0.10 < p\text{-value} < 0.20$.

Or, in R,

```
> tt$p.value
[1] 0.1944
```

d. Since the p-value > 0.05, our conclusion is to not reject the null hypothesis. There is not enough evidence to suggest that the means are significantly different.

And here is the complete R output.

```
> tt

   Two Sample t-test

data:  x and y
t = 1.416, df = 8, p-value = 0.1944
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
 -10.05  42.05
sample estimates:
mean of x mean of y
       64        48
```