Student : Dane Johnson

Homework policy: you must show your work to receive credit for these exercises. It is your responsibility to convince the grader that you understand how to solve each of these exercises and to explain precisely how you arrived at your solution.

1. Suppose $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ are vectors of data points, and that a and b are scalar constants. Define $a\mathbf{x} := (ax_1, \ldots, ax_n)$. Express your answers in terms of a, b, $\bar{x}$, $\bar{y}$, s_x, s_y, and ρ.

    a) What is the sample mean of $a\mathbf{x}$?

    $\overline{a\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} ax_i = \frac{a}{n}\sum_{i=1}^{n} x_i = a\bar{\mathbf{x}}$. (We factored the constant a out of a finite sum).

    b) What is the sample variance of $a\mathbf{x}$?

$$s_{a\mathbf{x}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(ax_i - \overline{a\mathbf{x}})^2 = \frac{1}{n-1}\sum_{i=1}^{n}[a(x_i - \bar{\mathbf{x}})]^2$$
$$= a^2 \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})^2$$
$$= a^2 s_{\mathbf{x}}^2,$$

    where we factored the constant a^2 out of a finite sum.

    c) What is the sample covariance of $a\mathbf{x}$ and $b\mathbf{y}$?

$$cov(a\mathbf{x}, b\mathbf{y}) = \frac{1}{n-1}\sum_{i=1}^{n}(ax_i - a\bar{x})(by_i - b\bar{y})$$
$$= \frac{1}{n-1}\sum_{i=1}^{n} a(x_i - \bar{x})b(y_i - \bar{y})$$
$$= (ab)\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$
$$= (ab)cov(\mathbf{x}, \mathbf{y}) = (ab)\rho(\mathbf{x}, \mathbf{y})s_x s_y.$$

    We replaced covariance with something in terms of correlation in the last line because the covariance symbol was not allowed in the final answer by the directions of this problem.

d) What is the sample correlation of $a\mathbf{x}$ and $b\mathbf{y}$?

$$\rho(a\mathbf{x}, b\mathbf{y}) = \frac{cov(a\mathbf{x}, b\mathbf{y})}{s_{ax}s_{by}} = \frac{(ab)cov(\mathbf{x}, \mathbf{y})}{\sqrt{s_{ax}^2}\sqrt{s_{by}^2}}$$

$$= \frac{(ab)cov(\mathbf{x}, \mathbf{y})}{\sqrt{a^2 s_x^2}\sqrt{b^2 s_y^2}}$$

$$= \frac{(ab)cov(\mathbf{x}, \mathbf{y})}{(ab)s_x s_y}$$

$$= \rho(\mathbf{x}, \mathbf{y}).$$

e) What is the sample variance of the linear combination $a\mathbf{x} + b\mathbf{y}$?

$$s_{a\mathbf{x}+b\mathbf{y}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(ax_i + by_i - \overline{a\mathbf{x}+b\mathbf{y}})^2 \qquad (*)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i + by_i - (a\bar{x} + b\bar{y}))^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(a(x_i - \bar{x}) + b(y_i - \bar{y}))^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}a^2(x_i - \bar{x})^2 + 2ab(x_i - \bar{x})(y_i - \bar{y}) + b^2(y_i - \bar{y})^2$$

$$= a^2 \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 + 2ab\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + b^2\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$= a^2 s_x^2 + 2ab\, cov(\mathbf{x}, \mathbf{y}) + b^2 s_y^2 = a^2 s_x^2 + 2ab\, \rho(\mathbf{x}, \mathbf{y})s_x s_y + b^2 s_y^2.$$

We replaced covariance with something in terms of correlation in the last line because the covariance symbol was not allowed in the final answer by the directions of this problem.

$(*)$ $\overline{a\mathbf{x}+b\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n} ax_i + by_i = \frac{1}{n}\sum_{i=1}^{n} ax_i + \frac{1}{n}\sum_{i=1}^{n} by_i = a\bar{x} + b\bar{y}$. Therefore we have: $\overline{a\mathbf{x}+b\mathbf{y}} = a\bar{x} + b\bar{y}$.

2. In R, create vectors $\mathbf{x}$ and $\mathbf{y}$ by running the following code snippet.

```
thetas <- seq(0, pi, length = 181)
x <- cos(thetas)
y <- sin(thetas)
```

a) Produce a plot of $\mathbf{y}$ vs. $\mathbf{x}$ being careful to follow the plotting best practices.

b) Describe the dependence of `y` on `x` mathematically.

A general mathematical identity about the dependence of `y` and `x` is :

$$sin(\theta) = cos(\theta - \frac{\pi}{2}).$$

That is, $y = sin(\theta)$ is equal to a cosine wave with the same period but with a phase shift exactly equal to $\frac{\pi}{2}$ radians. This means that in our case that y and x should be orthogonal for any θ input.

c) Use R to compute the sample correlation between `x` and `y`. Are `x` and `y` uncorrelated? Explain your reasoning.

Using R we have the following input and output:

```
> cor(x,y)
[1] -6.324118e-17
```

Using R's command `.Machine` we see that this result is apparently below R's epsilon- or roughly speaking our result is basically 0. We should conclude that `x` and `y` are uncorrelated. This is sensible since as we discussed in part(b), $sin(\theta)$ and $cos(\theta)$ are orthogonal functions on our interval (by this we mean $\int_0^\pi sin(\theta)cos(\theta) \, d\theta = 0$) and so we expect that a sample of points from this interval evaluated by these functions to produce our data vectors should be orthogonal vectors. Since covariance (and thereby correlation) between two vectors can be formulated in terms of a dot product we expect that data in orthogonal vectors be uncorrelated.

3. Use closing price data for iRobot Corp (ticker IRBT) from 1 Jan 2010 through 31 Dec 2014 for the following questions.

a) Produce normal qq plots (see the function `qqnorm`) of daily, weekly, monthly, and quarterly continuously compounded returns. Remember the plotting best practices.

Hint 1: you can add a line to a normal qq plot by calling the function `qqline` imme- diately after `qqnorm` (use the same arguments).

Hint 2: call the function `par(mfrow = c(2, 2))` to plot all 4 plots on one device.

See attached R code.

b) Which plot is the closest to the Gaussian reference? Which is the furthest?

Quarterly appears to be the closest to the Gaussian reference while daily seems to be the furthest. We might expect that fluctations in daily closing prices might be smoothed over time so it is reasonable that monthly data would be closer to normally distributed.