Exercise 1: Below are three sample of a random variable. Which sample displays the greatest variance? Why?

Sample 1: 21.45, 22.93, 31.86, 19.37, 20.87
Sample 2: 182, 186, 179, 184, 187
Sample 3: 14.3, 14.7, 10.0, 14.8, 14.6

Solution:
I would say, just from looking at the samples number 3 has the lowest variance since 4 of the entries seemed to be grouped around 14.5 and the only outlier is 10, which isn't very far either. Out of samples 1 and 2 I'm not really sure which one has greater variance but I would guess that its sample 1 since the 31.86 term is a very large (relatively) outlier, while the entries in sample 2 are all more or less equally spaced.

Exercise 2: A random sample of dogs studied by veterinarians was found to have the following lifespans in years: $14, 18, 6, 9, 13$. The veterinarians inquired about each dog's mother and found the mother' lifespans to be: $10, 19, 13, 8, 24$. What is the sample covariance between the offsprings' lifespans and mothers' lifespans?

Solution:
First we need to compute the sample mean for each of our samples.

$$\bar{x}_{dogs} = \frac{14 + 18 + 6 + 9 + 13}{5} = 12,$$

$$\bar{x}_{mothers} = \frac{10 + 19 + 13 + 8 + 24}{5} = 14.8,$$

Now we can compute the sample covariance,

$$Cov(X_{dogs}, X_{mothers}) = \frac{\sum(x_{dogs} - \bar{x}_{dogs})(x_{mothers} - \bar{x}_{mothers})}{n - 1} = 14$$

I omitted the large fraction and algebra, but I double checked it with r and it does check out.

Exercise 3: Consider X, a random variable with a standard normal distribution. What is the probability that X is greater than $-.4$.

Solution:
So given that $X \approx N(1, 0)$ we want to find the value of $P(X > -.4)$. Using r or a table we can compute the probability, and we get that,

$$P(X > -.4) \approx 0.655$$

Code:

```
> pnorm(-.4, lower.tail = FALSE)
[1] 0.6554217
```

Exercise 4: Consider V, a random variable with a $N(13, 400)$. What is the probability that V is less than or equal to -7?

Solution:
First we want to get our probability in terms of the standard normal,

$$P(V \leq -7) = P\left(\frac{V - 13}{\sqrt{400}} \leq \frac{-7 - 13}{\sqrt{400}}\right) = P(Z \leq -1)$$

Using r or a table we can compute the probability, and we get that,

$$P(V \leq -7) \approx 0.159$$

Code:

```
> pnorm(-1)
[1] 0.1586553

> pnorm(-7, 13,sqrt(400))
[1] 0.1586553
```

Exercise 5: Let the expected value of random variable X be a, the expected value of Y be b, and the expected value of Z be c. Find $E(4 - 2X + 3Y - 10Z)$.

Solution:
Applying the rule of the expected values we get,

$$E(4 - 2X + 3Y - 10Z) = 4 - 2E(X) + 3E(Y) - 10E(Z).$$

By substitution we get,

$$E(4 - 2X + 3Y - 10Z) = 4 - 2a + 3b - 10c.$$

Exercise 6: Let the variance of random variable X be 3, the variance of Y be 12, and the variance of Z be 9, and let X, Y, and Z be uncorrelated . Find $V(4 - 2X + 3Y - 10Z)$.

Solution:
Applying the rule of variances we get,

$$V(4 - 2X + 3Y - 10Z) = V(4) + (-2)^2 V(X) + 3^2 V(Y) + (-10)^2 V(Z)$$

By substitution we get,

$$V(4 - 2X + 3Y - 10Z) = V(4) + 2^2(3) + 3^2(12) + 10^2(9) = V(4) + 1020.$$

Note that, $V(4) = E(4^2) - E(4)^2 = 16 - 16 = 0$ so we get that,

$$V(4 - 2X + 3Y - 10Z) = 1020.$$

Exercise 7: If a random variable A is distributed $N(3, 8)$ and an independent random variable B is distributed $N(-3, 10)$, what is the distribution of $mA + nB$ for some constants m and n?

Solution:
Recall that if A and B are independent random variables that are normally distributed, their sum is also normally distributed. Therefore we know that $mA + nB$ is a normally distributed random variable. Solving for $E(mA + nB)$,

$$E(mA + nB) = E(mA) + E(nB) = E(A)m + E(B)n = 3m - 3n = 3(m - n)$$

Solving for $V(mA + nB)$,

$$V(mA + nB) = V(mA) + V(nB) = V(A)m^2 + V(B)n^2 = 8m^2 + 10n^2$$

So finally we get that $mA + nB$ is distributed $N(3(m - n), 8m^2 + 10n^2)$.

Exercise 8: Angie records the lengths of 15 poems from ancient Greece and finds the average wordcount of the poems is 456 with a standard deviation of 192. What is the standard error of Angie's estimator of the mean wordcount for all Greek poems?

Solution:
Recall the formula for computing the standard error of the sample mean,

$$SE_{\overline{x}} = \frac{S}{\sqrt{n}}$$

By substitution we get that,

$$SE_{\overline{x}} = \frac{192}{\sqrt{15}} \approx 48.57$$

Exercise 9: Consider the mean of a random sample of size 75, $\overline{X}$. If S^2 is the sample variance and the population is normally distributed with mean μ, what is the distribution of,

$$\frac{\overline{X} - \mu}{S / \sqrt{75}}$$

Solution:
The formula above describes the normalizing of a random sample. If the estimators were replaced with their population parameters we would have a variable with a standard normal distribution, instead we have a variable with a t-distribution with 74 degrees of freedom.

Exercise 10: The mean weight of peanuts in a sample of size 16 from a barrel is 0.09 ounces. The standard deviation of the sample is 0.015 ounces. What is a 90% confidence interval for the mean weight of all peanuts in the barrel? Assume peanut weight s in the barrel are normally distributed.

Solution:
Recall we compute the 90% CI with the following formula,

$$90\% CI_{\overline{x}} = \overline{x} \pm (t_{.1/2,n-1} \cdot \frac{s}{\sqrt{n}})$$

Finding the critical t-value for a $90\% CI$ with 15 degrees of freedom using r or a table we get,

$$t_{.5,15} = 1.75305$$

Now we can find the confidence interval,

$$\begin{aligned} 90\% CI_{\overline{x}} &= .09 \pm (1.75305 \cdot \frac{.015}{\sqrt{16}}) \\ &\approx .09 \pm 0.00657 \end{aligned}$$

Exercise 11: The weights, in pounds, of a team of sixteen male athletes are as follows:

188.5, 183.0, 194.5, 185.0, 214.0, 203.5, 186.0, 178.5, 186.0, 184.5, 204.0, 184.5, 195.5, 202.5, 174.0, 183.0

Assume that this team is representative of all athletes in the league. Test the hypothesis that the mean weight of the league is greater than 190. Assume that athlete weights in the league are normally distributed and use an α level of 0.01.

Solution:
First we must define our experiment,

$$H_0 : \mu = 190$$
$$H_a : \mu > 190$$

Computing the sample mean and standard deviation we get that,

$$\overline{x} = 190.4375,$$
$$s = 10.84416$$

Now we need to solve for the one-sample t test statistic,

$$t = \frac{\overline{x} - 190}{s/\sqrt{(n)}} = 0.1613772$$

To compute the p-value we need to find the area to the right of t in the T_{n-1} distribution. Using r or a table we get,

$$p \approx 0.4369749$$

Since $p > .01$ we fail to reject the null hypothesis. There is not sufficient evidence at the 1% level to conclude that the mean athlete weight is greater than 190 pounds.

Code:

```
> x<-c(188.5, 183.0, 194.5, 185.0, 214.0, 203.5, 186.0 ,
    178.5, 186.0, 184.5, 204.0, 184.5, 195.5, 202.5, 174.0, 183.0);

> mean(x)
[1] 190.4375

> sd(x)
[1] 10.84416

> t.value= (mean(x)-190)/(sd(x)/sqrt(length(x)))
> t.value
[1] 0.1613772

> p.value = pt(t.statistic , df=length(x)-1, lower.tail = FALSE)
> p.value
```

```
[1] 0.4369749

## Double Checking
> t.test(x, alternative = 'greater', mu=190, conf.level = .99)

        One Sample t-test

data:  x
t = 0.16138, df = 15, p-value = 0.437
alternative hypothesis: true mean is greater than 190
99 percent confidence interval:
 183.3821      Inf
sample estimates:
mean of x
 190.4375
```

Exercise 12: A doctor is performing a clinical trian of research medication for treating symptoms of emphysema. She recruits 50 patients with the disease and randomly assigns 25 to the treatment group and 25 to the placebo group. After 6 weeks, she measures each patient's blood oxygen level and compares the sample means of the two groups. She calculates a test statistic of 4.38 and a p-value of 0.0001. She concludes that, since $0.0001 < 0.025$, the medication significantly raises blood oxygen levels compared to the placebo. She calculates that the power of her test is .7.

What would constitute a type I error in this scenario and what is it;s probability? What would constitute a type II error in this scenario and what is it's probability?

Solution:
Let's consider the doctors experiment, suppose that μ_1 is the average blood oxygen level of the medication users and μ_2 is the placebo group.

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 > \mu_2$$

A type I error constitutes the doctor concluding that the medication does alleviate symptoms of emphysema, when in reality it does not. The probability of this happening is $\alpha = .025$

A type II error constitutes the doctor concluding that the medication fails to alleviate symptoms of emphysema, when in reality it does. The probability of this happening is $\beta = 1 - power = .3$