# Problem 1

## (a)

First we obtained the $\chi^2$ proportion $P(s^2 < 4.8)$ by using `pchisq(9*4.8/4,df=9)`, which gave us a value of 0.7103325. Then we used normal distribution to simulate $P(s^2 < 4.8)$ . First, we sampled $n$ variates of the normal distribution with $\sigma = 2$ and found the following variance $s^2$:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Then, we repeated the first process and found the proportion of variances that were less than 4.8. This proportion turned out to be around .76, which was close to the value we obtained using the $\chi^2$ proportion.

## (b)

Similar to our variate sampling in part a, except using an exponential distribution. We found the proportion of variances calculated to be $\approx 1$.

# Problem 2

Let us begin by finding the variances of the two variables. First, note that the probability of choosing someone with the trait in the entire population is $P(T) = P(Sub_1, T \cup Sub_2, T) = qp_1 + (1-q)p_2$, where $p_i$ is the probability of finding the trait in the $i^{th}$ population, $T$ is an indicator variable for the trait, and $Sub_i$ is an indicator for whether we are in subpopulation $i$ or not. Because $X$ consists of indicator random variables, we can easily find variance.

$$\begin{aligned}
Var(X) &= \frac{1}{n}\sigma^2 \\
&= \frac{1}{n}p(1-p) \\
&= \frac{1}{n}\left\{[qp_1 + (1-q)p_2][1 - qp_1 - (1-q)p_2]\right\}
\end{aligned}$$

Now to find $Var(Y)$, let $S_1$ be the sampled number of people with the trait in population 1 and let $S_2$ be the sampled number of people with the trait in population 2.

$$\begin{aligned}
Var(Y) &= Var(S_1 + S_2) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) \\
&= \frac{1}{n^2}\left[qnVar(S_1) + (1-q)nVar(S_2)\right] \\
&= \frac{1}{n}[qVar(S_1) + (1-q)Var(S_2)] \\
&= \frac{1}{n}[qp_1(1-p_1) + (1-q)p_2(1-p_2)]
\end{aligned}$$

We can also make the observation that, in the case of the problem at hand, $q = (1-q)$ and $p_1 = (1-p_2)$. Finding

$Var(Y)/Var(X)$ is a matter of algebra at this point.

$$\begin{aligned}
Var(Y)/Var(X) &= \frac{qp_1(1-p_1)+(1-q)p_2(1-p_2)}{[qp_1+(1-q)p_2][1-qp_1-(1-q)p_2]} \\
&= \frac{qp_1p_2+qp_2p_1}{[qp_1+q(1-p_1)][1-qp_1+q(1-p_1)]} \\
&= \frac{2qp_1p_2}{q(1-q)} \\
&= \frac{2p_1p_2}{q}
\end{aligned}$$

Plugging in our values for $p_1$, $p_2$, and $q$, we get

$$Var(Y)/Var(X) = 0.75$$

# Problem 3

First we sampled each of the $X, Y, Z$ proportions 15 times with probabilities $p1 = 0.11, p2 = 0.16, p3 = 0.05$ respectively. We then found whether $Z > 2XY$ for a number of iterations, and found that the proportion of times that the array has `TRUE` values is $\approx 0.46$.