Exercise

- Create a population, which is normally distributed with the mean and standard deviation of your choice
- Draw a new random sample 1000 times (the sample size should be 30, i.e., n=30) and calculate sample variances
- Prove that $E(s^2) = \sigma^2$ and $Var(s^2) = Var(s^2) = \frac{2\sigma^4}{n-1}$
  where s^2 refers to the sample variances, and σ^2 refers to the population variance
- Plot the dist. of $s^2(n-1)/\sigma^2$. Does it look like a chi-square distribution?
- Calculate the degrees of freedom in this example
- For the degrees of freedom that you have calculated, what is the chi-square value at the lower 0.01 interval and the upper 0.05 interval.
- What is the probability that a sample (out of 1000) will have a variance below 35?
- What is the probability that a sample (out of 1000) will have a variance above 15?
- Repeat question no. 2 but this draw a new random sample 1000 times (the sample size should be 5, i.e., n=5).
- Plot the dist. of $s^2(n-1)/\sigma^2$. Does it look like a chi-square distribution? See how the distribution has changed when we decreased the degrees of freedom.

Solution

- Create a population, which is normally distributed with the mean and standard deviation of your choice

```
set.seed(213)
pop <- rnorm(1e+05, mean = 5, sd = 5)
```

- Draw a new random sample 1000 times (the sample size should be 30, i.e., n=30) and calculate sample variances

```
sam_var <- replicate(1000, var(sample(pop, 30)))
```

- Prove that $E(s^2) = \sigma^2$

```
mean(sam_var);    var(pop)

## [1] 24.84
## [1] 24.81
```

- Prove that $Var(s^2) = \frac{2\sigma^4}{n-1}$

```
var(sam_var)

## [1] 42.75
```

Now we check if the variance of sample variances $Var(s^2)$ is equal to $\frac{2\sigma^4}{n-1}$

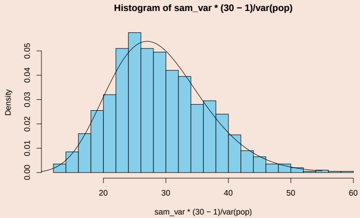
```
2 * (var(pop)^2)/29

## [1] 42.44
```

Solution

- Plot the dist. of $s^2(n-1)/\sigma^2$. Does it look like a chi-square distribution?

```
hist(sam_var * (30 - 1)/var(pop), breaks = 20, prob = TRUE, col = "skyblue")
curve(dchisq(x, df = 29), from = 0, to = 55, add = TRUE)
```



- Calculate the degrees of freedom in this example (df = n-1 = 29)
- For the degrees of freedom that you have calculated, what is the chi-square value at the lower 0.01 interval and the upper 0.05 interval

```
qchisq(0.01, df = 29);          qchisq(0.95, df = 29)

## [1] 14.26
## [1] 42.56
```

Solution

- What is the probability that a sample (out of 1000) will have a variance below 35?

$$P(s^2 < 30) = P\left[s^2(n-1)/\sigma^2 < 35(29)/\sigma^2\right]$$

- We earlier calculated $\sigma^2 = 24.81$

$$P(s^2 < 30) = P\left[\chi^2_{29} < 35(29)/24.81\right]$$

$$P(s^2 < 30) = P\left[\chi^2_{29} < 40.91\right]$$

```
pchisq(40.91, df = 29)

## [1] 0.9299
```

The above probability is associated with the area to the left of 40.91 on the χ^2 dist.

Solution

- What is the probability that a sample (out of 1000) will have a variance above 15?

$$P(s^2 > 15) = P\left[s^2(n-1)/\sigma^2 > 15(29)/\sigma^2\right]$$

- We earlier calculated $\sigma^2 = 24.81$

$$P(s^2 > 15) = P\left[\chi^2_{29} > 15(29)/24.81\right]$$

$$P(s^2 > 15) = P\left[\chi^2_{29} > 17.533\right] = 1 - P\left[\chi^2_{29} < 17.533\right]$$

```
1 - pchisq(17.533, df = 29)

## [1] 0.9532
```

The above probability is associated with the area to the right of 17.533 on the χ^2 dist.

Note: We can also empirically verify by seeing how many percent of our samples had variance above 15 (this should give something close to 95 percent):

```
100 * mean(sam_var > 15)

## [1] 95.8
```

Solution

- Repeat question no. 2 but this time draw a new random sample 1000 times (the sample size should be 5, i.e., n=5).

```
sam_var1 <- replicate(1000, var(sample(pop, 5)))
```

- Plot the dist. of $s^2(n-1)/\sigma^2$. Does it look like a chi-square distribution?

```
curve(dchisq(x, df = 4), from = 0, to = 55, add = TRUE)
```