STAT430 Homework #2: Due Friday, February 11, 2022.

Name: KEY

For this homework, you *must* submit a pdf file to Canvas. You can still submit handwritten solutions if you wish to solve some problems by hand. However, I instead encourage you to learn LaTeX and write up your solutions in the R Markdown file. I have not had any luck yet getting you paid acounts for RStudio cloud, but for now you can use the free version (https://rstudio.cloud/). Let me know if you have problems with this.

Question 1

Let X_i, $i = 1, 2, 3$, be independent with $\mathrm{N}(i, i^2)$ distributions. For each of the following situations, use the X_is to construct a statistic with the indicated distribution.

1. chi-squared with 3 degrees of freedom

For $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable, and for $Z_i, \ldots, Z_n \overset{\text{iid}}{\sim} N(0,1)$, $Z_i^2 + \cdots + Z_n^2$ is a χ^2 random variable with n degrees of freedom. We can therefore construct a χ^2_3 random variable with

$$\left(\frac{X_1 - 1}{1}\right)^2 + \left(\frac{X_2 - 2}{2}\right)^2 + \left(\frac{X_3 - 3}{3}\right)^2$$

2. t distribution with 2 degrees of freedom

A t distributed random variable T with n degrees of freedom can be constructed with a standard normal random variable Z and an independent χ^2_n random variable W as

$$T = \frac{Z}{\sqrt{W/n}} \sim t_n$$

As before, we can construct a standard normal random variable using X_1 and an independent χ^2_2 random variable with X_2 and X_3. Then we can get a t_2 distributed random variable with

$$\frac{\frac{X_1-1}{1}}{\sqrt{\left[\left(\frac{X_2-2}{2}\right)^2 + \left(\frac{X_3-3}{3}\right)^2\right]\Big/2}}$$

3. F distribution with 1 and 2 degrees of freedom

An $F_{n,m}$ distributed random variable V can be constructed as the ratio of a χ^2_n divided by its degrees of freedom random variable W and an independent χ^2_m random variable U divided by its degrees of freedom:

$$V = \frac{W/n}{U/m} \sim F_{n,m}.$$

So,

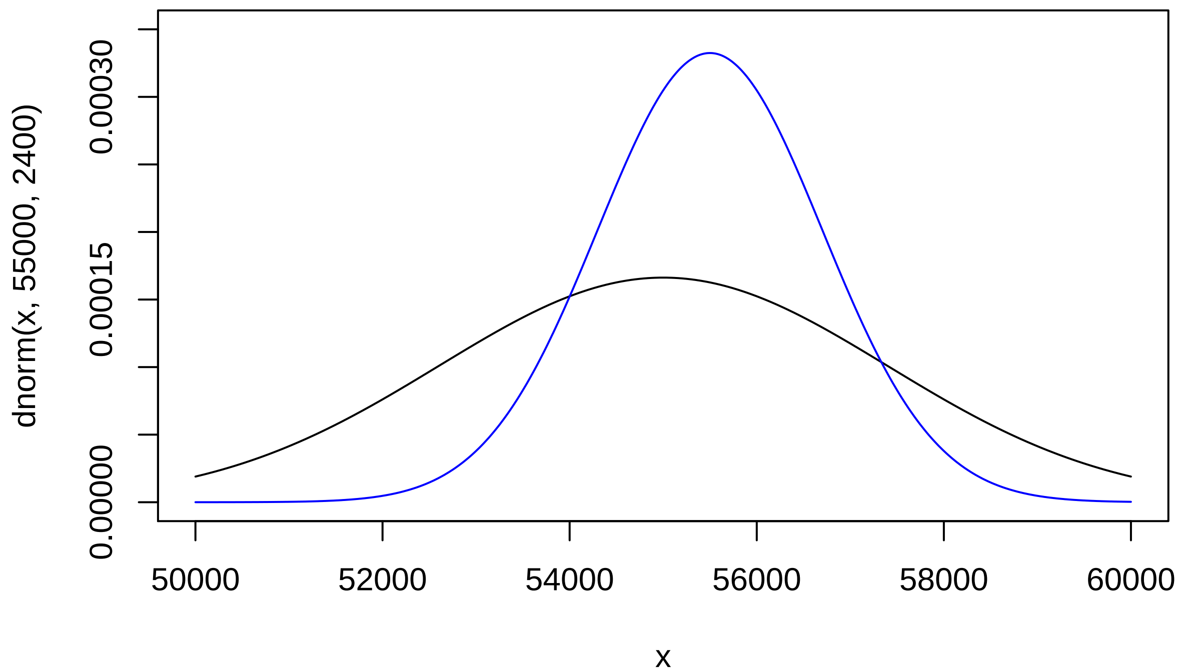
$$\frac{\left(\frac{X_1-1}{1}\right)^2\Big/1}{\left(\left(\frac{X_2-2}{2}\right)^2 + \left(\frac{X_3-3}{3}\right)^2\right)\Big/2} \sim F_{1,2}$$

Question 2

Actuary students from CSU graduate and find a job paying an amount of money that is distributed $N(55000, 2400^2)$, while CU students earning are distributed $N(55500, 1200^2)$.

We can plot these distributions by creating a long vector of `x` values, here ranging from 50K to 60K, and evaluating each of the normal probability density functions using the `dnorm` function at these values of `x`. We then ask `R` to "connect the dots" to make a curve for each density. Note that the functions for the normal distribution in `R` use take a mean and a standard deviation as arguments.

```
x <- seq(50000,60000,length=1000) # vector of values from 50K - 60K
plot(x,dnorm(x,55000,2400),type="l",ylim=c(0,.00035)) # CSU students
lines(x,dnorm(x,55500,1200),type="l",col="blue") # CU students
```



We can also use `R` to calculate the probability that a normal random variable is within a given range using the `pnorm` function. Let's look at the help file for this function.

```
help(pnorm)
```

The default argument is `lower.tail=TRUE` indicating you will get the probability the normal random variable is less than the specified value `q`. Thus, by default is it the cumulative distribution function. Use the `pnorm` command to answer the questions below.

1. You are offered an entry level job paying \$53,800. At which institution do you rank higher? In other words, what percentile do you represent at each institution and which is higher?

```
pnorm(53800, 55000, 2400) # Percentile for CSU Students
```

```
## [1] 0.3085375
```

```
pnorm(53800, 55000, 1200) # Percentile for CU Students
```

```
## [1] 0.1586553
```

\$53,800 is higher than 31% of the salaries of CSU actuary graduates and 16% of CU actuary graduates. Therefore, you rank higher among CSU actuary graduates.

2. Your friend is offered a job paying \$58,000. At which institution does your friend rank higher?

```r
pnorm(58000, 55000, 2400) # Percentile for CSU Students
```

```
## [1] 0.8943502
```

```r
pnorm(58000, 55000, 1200) # Percentile for CU Students
```

```
## [1] 0.9937903
```

\$58,000 is higher than 89% of the salaries of CSU actuary graduates and 99% of CU actuary graduates. Therefore, your friend ranks higher among CU actuary graduates.

Question 3

Suppose $X \sim \chi^2_n$ distribution. Prove that

$$\frac{X-n}{\sqrt{2n}} \to_d N(0,1) \quad \text{as } n \to \infty.$$

The central limit says that for an iid sample $X_1, \dots, X_n$ of random variables

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1).$$

X can be expressed as the sum of $X_1, \dots, X_n \overset{\text{iid}}{\sim} \chi^2_1$ random variables: $X = \sum_{i=1}^n X_i$. Define $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Using the knowledge that $E[X_i] = 1$ and $\text{Var}(X_i) = 2$, we know that

$$\frac{\bar{X}-1}{2/\sqrt{n}} \xrightarrow{d} N(0,1)$$

multiplying the numerator and denominator by n, we get

$$\frac{\sum_{i=1}^n X_i - n}{2\sqrt{n}} = \frac{X-n}{2\sqrt{n}} \xrightarrow{d} N(0,1).$$

Question 4

Let $\bar{X}$ be the average of 16 iid standard normal random variables. Find c such that

$$P(|\bar{X}| < c) = 0.5.$$

We can find c as follows:

$$P(|\bar{X}| < c) = P(-c < \bar{X} < c)$$
$$0.5 = P\left(\frac{-c-0}{1/4} < \frac{\bar{X}-0}{1/4} < \frac{c-0}{1/4}\right) = P\left(-4c < \frac{\bar{X}-0}{1/4} < 4c\right).$$

$\frac{\bar{X}-0}{1/4}$ is t_{15} distributed, which is symmetric so we can use R to find the lower 25th percentile and use this to find c:

```r
qt(.25, 15)
```

```
## [1] -0.6911969
```

So $-4c \approx -0.6912$ and $c \approx 0.1728$.

Question 5

If $X \sim F_{n,m}$, find the distribution of X^{-1}. Explain your reasoning.

If $X \sim F_{n,m}$ we can express X as

$$X = \frac{W/n}{V/m}$$

where $W \sim \chi_n^2$ and $V \sim \chi_m^2$. Taking the reciprocal, we have

$$X^{-1} = \frac{V/m}{W/n}$$

which is again an F distributed random variable, but with m and n degrees of freedom.

Question 6

If $X \sim T_n$, find the distribution of X^2. Explain your reasoning.

If $X \sim t_n$, then X can be expressed as

$$X = \frac{Z}{\sqrt{W/n}}$$

where $Z \sim N(0,1)$ and $W \sim \chi_n^2$. Squaring X, we get

$$X^2 = \frac{Z^2}{W/n} = \frac{Z^2/1}{W/n}.$$

which is the ratio of a χ_1^2 random variable (divided by its degrees of freedom) and a χ_n^2 random variable (divided by its degrees of freedom). Therefore $X^2 \sim F_{1,n}$.