**1. A survey of the members of a large professional engineering society is conducted to determine their views on proposed changes to an ASTM measurement standard. Overall 80% of the entire membership favor the proposed changes.**

**(a) If possible, describe the center, dispersion, and shape of the sampling distribution of the proportion of engineers for samples of size 20 who favor the proposed changes. Explain your answer including which of these three aspects of distribution you can & cannot describe and why.**

We know a proportion and $\pi$, so we have to use CLTFP.

Theorem CLTFP:

The sampling distribution of the sample proportion p is approximately Normally distributed with $\mu_p = \pi$ and $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ provided that $n\pi \geq 5$ and $n(1-\pi) \geq 5$.

3 Aspects of Distribution:

1. Center: ($\mu_p = \pi$)
   - $\mu_p = E[p] = \pi = \boxed{0.80}$ (assuming $p \approx \pi$)
2. Shape: (Apply CLTFP -> (normal)?)
   - $n\pi \geq 5$
     - $n\pi = 20 * 0.80 = 16 \geq 5$
     - This is TRUE!!!
   - $n(1-\pi) \geq 5$.
     - $n(1-\pi) = 20 * (1 - 0.80) = 20 * (0.2) = 4 \ngeq 5$
     - 4 is not greater than or equal to 5, so this condition of the CLTFP does not hold true.
   - $\boxed{\text{Therefore, we can conclude that the shape of the sampling distribution of the proportion is not normal.}}$
3. Dispersion: ($\sigma_x = \frac{\sigma}{\sqrt{n}}$)
   - We can find the standard deviation of the sample proportion:
   - $\sigma_p = \sqrt{V[p]} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.80((1-0.80)}{20}} = \sqrt{\frac{0.80((0.20)}{20}} = 0.0894427971 \approx \boxed{0.08944}$

I can describe the center and dispersion, but not the shape because the sampling distribution of proportions fails the CLTFP.

**(b) If possible, describe the center, dispersion, and shape of the sampling distribution of the proportion of engineers for samples of size 50 who favor the proposed changes. Explain your answer including which of these three aspects of distribution you can & cannot describe and why.**

We know a proportion and $\pi$, so we have to use CLTFP.

Theorem CLTFP:

The sampling distribution of the sample proportion p is approximately Normally distributed with $\mu_p = \pi$ and $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ provided that $n\pi \geq 5$ and $n(1-\pi) \geq 5$.

3 Aspects of Distribution:

1. Center: ($\mu_p = \pi$)
   - $\mu_p = E[p] = \pi = \boxed{0.80}$ (assuming $p \approx \pi$)
2. Shape: (Apply CLTFP -> (normal)?)
   - $n\pi \geq 5$
     - $n\pi = 50 * 0.80 = 40 \geq 5$
     - This is TRUE!!!
   - $n(1-\pi) \geq 5$.
     - $n(1-\pi) = 50 * (1-0.80) = 50 * (0.2) = 10 \geq 5$
     - This is TRUE!!!
   - $\boxed{\text{Because both conditions of the CLTFP hold true, we can conclude that the shape of the sampling distribution is approximately normally.}}$
3. Dispersion: ($\sigma_x = \frac{\sigma}{\sqrt{n}}$)
   - We can find the standard deviation of the sample proportion:
   - $\sigma_p = \sqrt{V[p]} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.80(1-0.20)}{50}} = \sqrt{\frac{0.80(0.20)}{50}} = 0.0565685425 \approx \boxed{0.0566}$

I can describe the center and dispersion, and the shape, because the sampling distribution of proportions passes the CLTFP.

**2. The lifetime of a brand of battery is normally distributed with a mean value of 8 hours and a standard deviation of 1 hour. There are four such batteries in a package.**

**(a) Completely describe the average lifetime of the four batteries and justify your answer.**

**You will use this distribution in parts (b) and (c).**

3 Aspects of Distribution:

1. Shape: CLT?
   - Because the population distribution is stated to be normally distributed, we know that the average lifetime of the four batteries (the sampling distribution of the sample mean) is **approximately normally distributed** because of the CLT part 1.
2. Center: ($\mu_{\overline{x}} = \mu$)
   - $\mu_{\overline{x}} = \mu = 8$ hours
3. Dispersion:
   - $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1\text{hour}}{\sqrt{4}} = \frac{1\text{hour}}{2} = 0.5$ hours

**(b) Find the probability that the average lifetime of the four batteries exceeds 9 hours. Express the probability using mathematical notation and illustrate it with a picture.**

$\mu_{\overline{x}} = 8$ hours

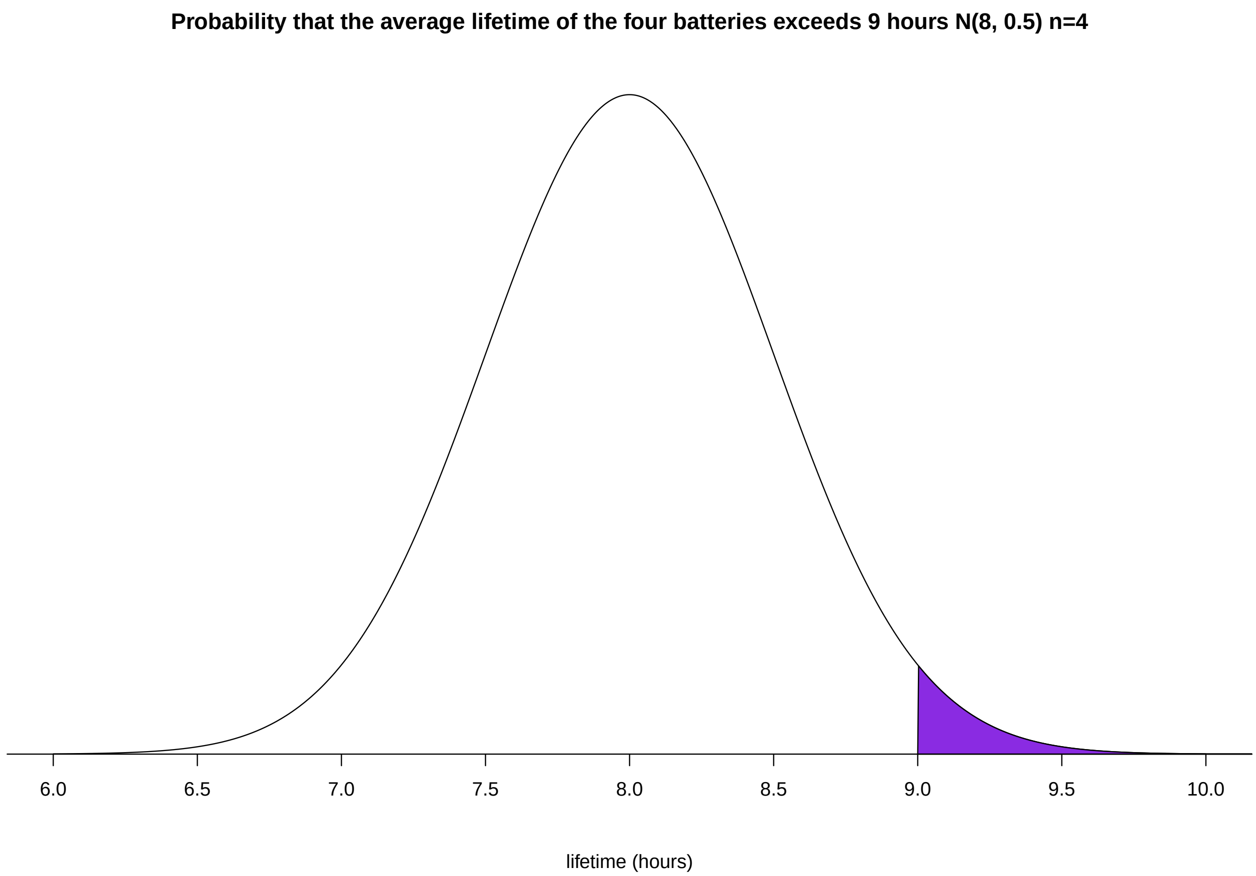
$\sigma_{\overline{x}} = 0.5$ hours

$\overline{x} \sim N(8, 0.5)$

$P(\overline{x} > 9)$

```
1 - pnorm(9,mean=8,sd=0.5)
```

```
## [1] 0.02275013
```



**(c) Let $\overline{x}$ denote the average lifetime of the four batteries in a randomly selected package. Find the numerical value of A for which P $(\overline{x}) \geq$ A) = 0.80. Express the probability in mathematical notation and illustrate it with a picture.**

$\mu_{\overline{x}} = 8$ hours

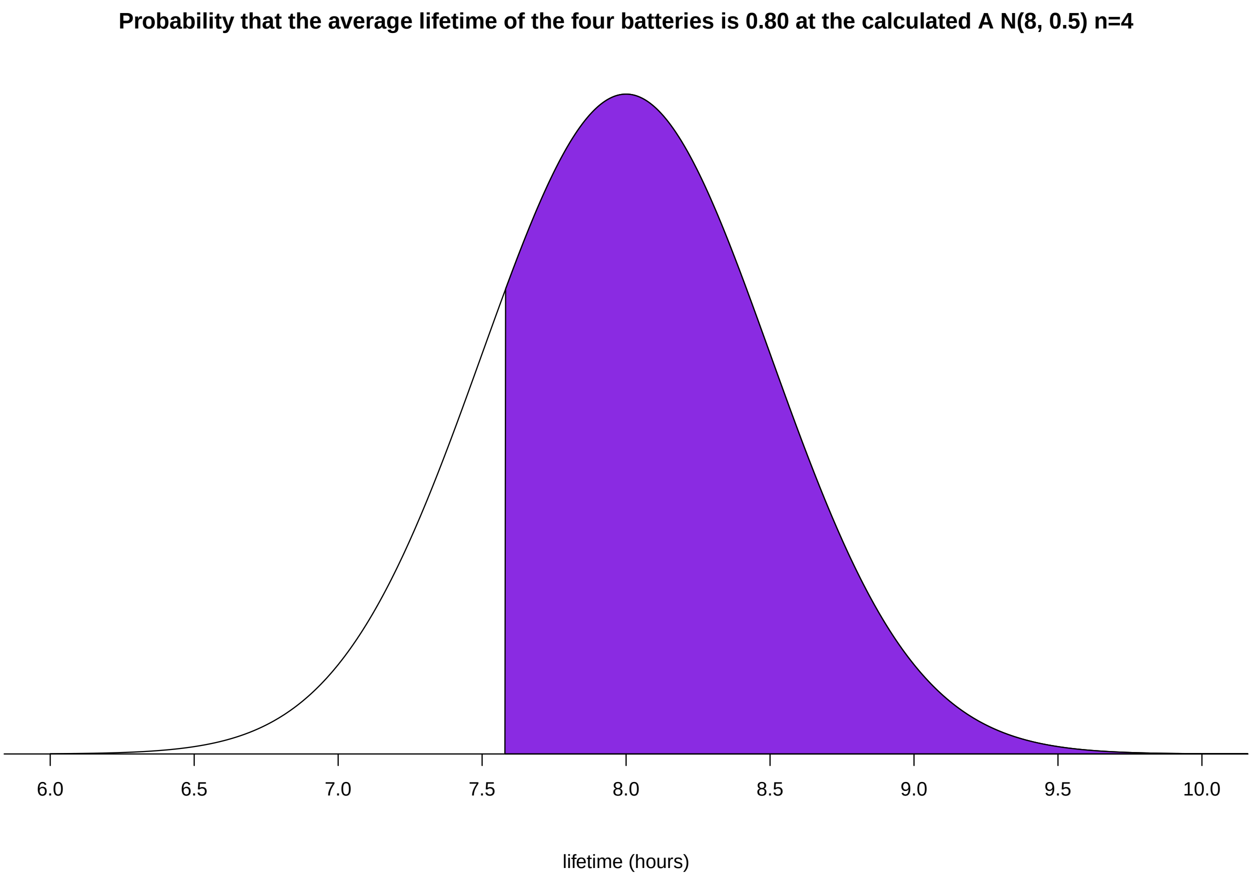
$\sigma_{\overline{x}} = 0.5$ hours

$\overline{x} \sim N(8, 0.5)$

$P(\overline{x} \geq A)$

```
#paste("A = ", qnorm(0.80,mean=8,sd=0.5))
paste("A = ", qnorm(0.20,mean=8,sd=0.5))
```

```
## [1] "A =  7.57918938321354"
```



**3. Consider the the distribution of under-inflated tires on a four-wheel automobile. The probability mass function is:**

**p(0) = 0.4, p(1) = p(2) = p(3) = 0.1, p(4) = 0.3.**

**Let X be the number of under-inflated tires on a randomly selected car.**

See Page 145 of notes.

**(a) Write down the domain of X.**

$D_x = \{0, 1, 2, 3, 4\}$

**(b) Calculate the expected value (mean) of X.**

$E[\overline{X}] = \sum_{x=0}^{4} x \cdot p(x) = 0(0.4) + 1(0.1) + 2(0.1) + 3(0.1) + 4(0.3) = 6(0.1) + 4(0.3) = \boxed{1.8}$ under-inflated tires per four-wheel automobile

(this is which mean of x value in the domain, which x is calculated for the probability function)

**(c) Calculate the standard deviation of X.**

$V[X] = E[\overline{X}^2] - (E[\overline{X}])^2$

We calculated $E[\overline{X}]$ above in part b, and we can plug this value in to solve for $(E[\overline{X}])^2$

$E[\overline{X}^2] = \sum_{x=0}^{4} x^2 p(x) = 0^2(0.4) + 1^2(0.1) + 2^2(0.1) + 3^2(0.1) + 4^2(0.3) = 1(0.1) + 4(0.1) + 9(0.1) + 16(0.3) = 14(0.1) + 16(0.3) = \boxed{6.2}$

$V[X] = E[\overline{X}^2] - (E[\overline{X}])^2 = 6.2 - (1.8)^2 = 2.96$

$sd = \sigma = \sqrt{V[x]} = \sqrt{2.96} = 1.720465053 \approx \boxed{1.7205}$ under-inflated tires per four-wheel automobile

**(d) Calculate the expected value (mean) of the average number of under-inflated tires in a sample of 40 cars.**

$E[\overline{X}] = \mu_{\overline{X}} = E[X] = \boxed{1.8}$ under-inflated tires per four-wheel automobile

**(e) Calculate the standard deviation of the average number of under-inflated tires in a sample of 40 cars.**

$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.720465053}{\sqrt{40}} = 0.2720294101 \approx \boxed{0.2720}$ under-inflated tires per four-wheel automobile

**(f) For a random sample of 40 cars, calculate the approximate probability that the average number of under-inflated tires exceeds 0.8.**

- **Use correct probability notation,**
- **draw a picture that represents the probability you are finding,**
- **and justify your answer.**

$P(\overline{X} > 0.8)$

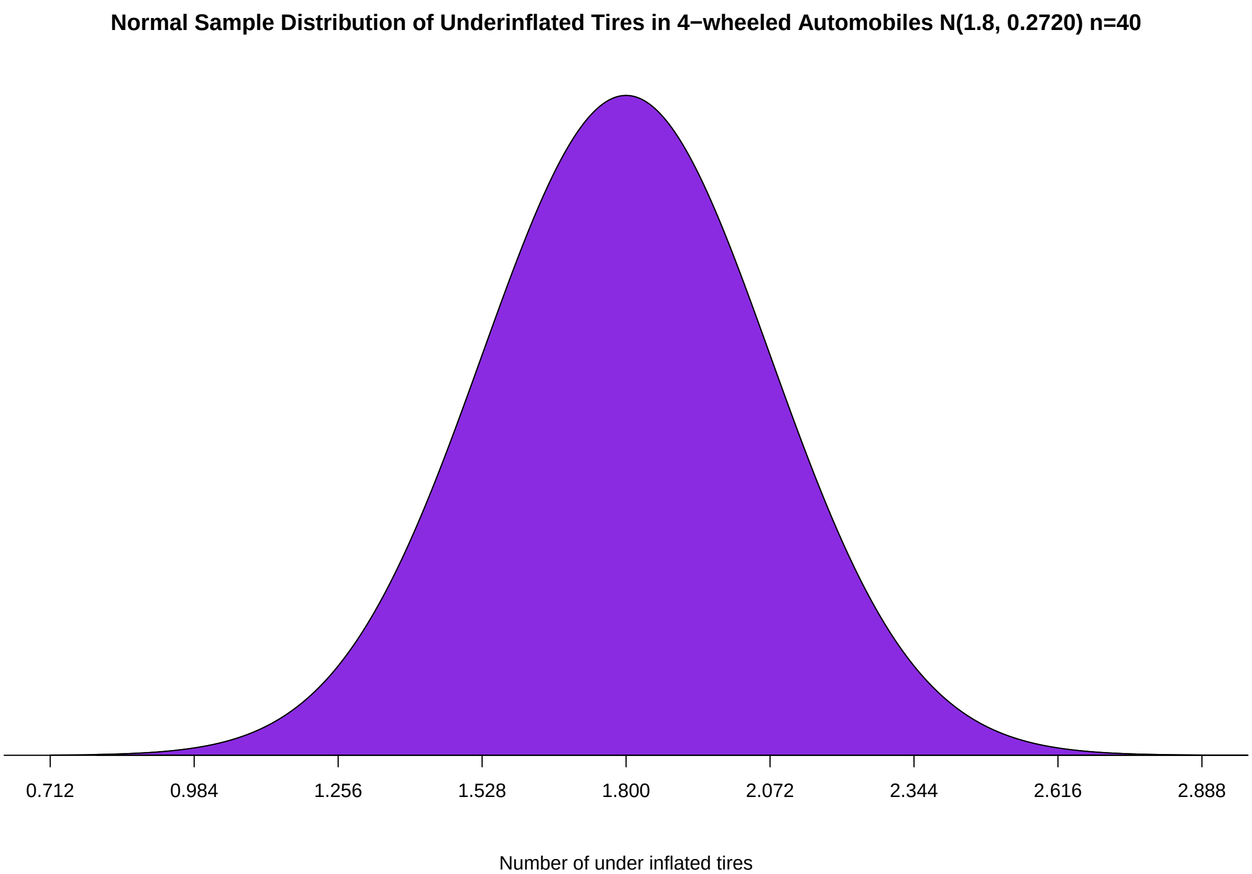
Calcutator Syntax: normalcdf(0.8,1E99,1.8, 0.272) $= 0.9998817265 \approx \boxed{0.9999}$

R Syntax: $P(\overline{X} > 0.8) = 1 - P(\overline{X} <= 0.8)$

```
1 - pnorm(0.8,mean=1.8,sd=0.2720)
```

```
## [1] 0.9998818
```

Because the sample size is greater than 30 (40 cars), we know that this sampling distribution of $\overline{x}$ is approximately normally distributed because of the CLT part 2.



**4. The inside diameter of a randomly selected piston ring is a random variable with a mean of 12 cm and a standard deviation of 0.04 cm.**

**(a) Find the mean and standard deviation of the sampling distribution of the sample mean for samples of size n. Use correct notation and show your work where applicable.**

Sampling Distribution of the Sample Mean -> Use CLT:

1. Sampling Mean: ($\mu_{\overline{x}} = \mu$)
   - $\mu_{\overline{x}} = \mu = 12$ cm
2. Sampling Standard Deviation:
   - $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.04cm}{\sqrt{n}}$

**(b) Describe the shape of the sampling distribution of the sample mean or samples of size n. State any assumptions you make and justify your answer.**

Because our population is described by a mean and standard deviation, we can then assume than the population is normally distributed.

Then, by CLT part 1, we know that our sampling distribution of the sample mean is also normally distributed.

**(c) Find the chance that, for a sample of size 25, the average diameter is within 0.01 cm of the mean.**

- **Use correct probability notation,**
- **draw a picture that represents the probability you are finding,**
- **and justify your answer.**

$\mu_{\overline{x}} = \mu = 12$ cm

$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.04cm}{\sqrt{25}} = 0.08$

$\overline{x} \sim N(12, 0.08)$

$P(11.99 \leq x \leq 12.01)$

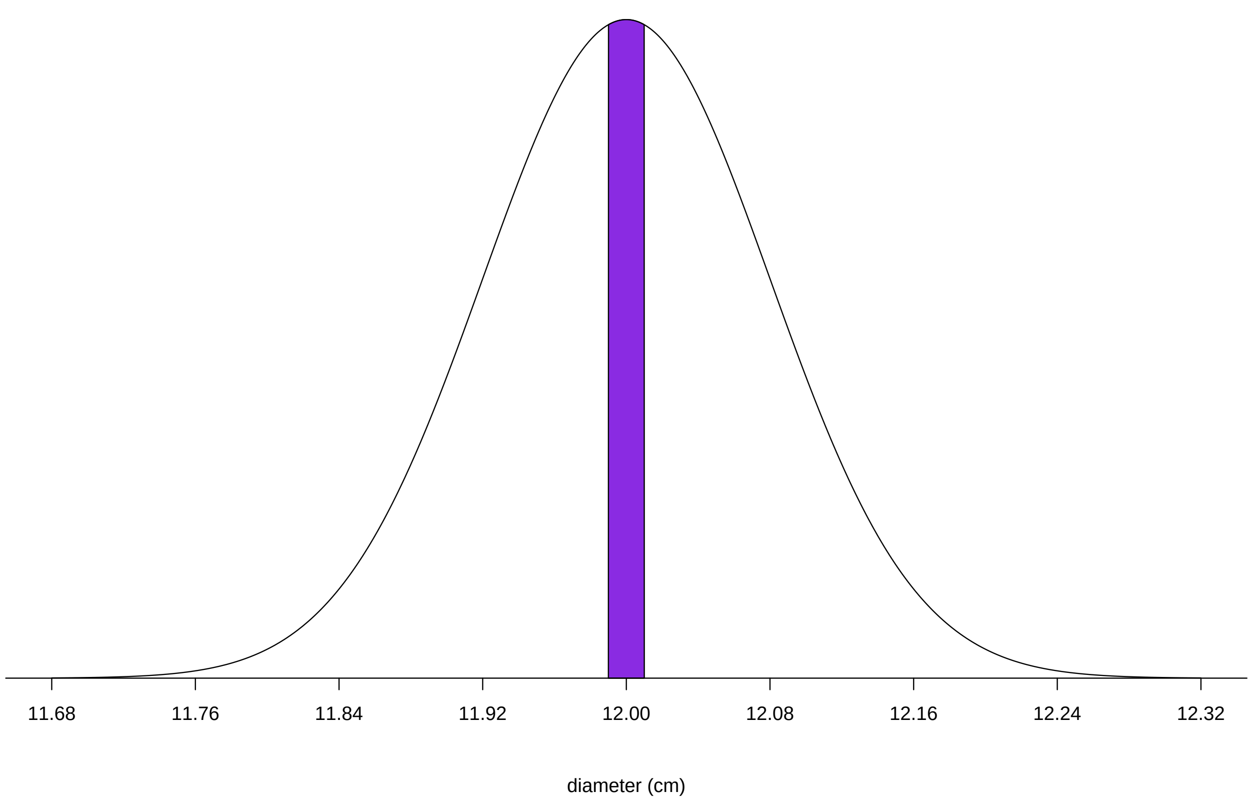
```
pnorm(12.01,mean=12,sd=0.08) - pnorm(11.99,mean=12,sd=0.08)
```

```
## [1] 0.09947645
```

Because our population is described by a mean and standard deviation, we can then assume than the population is normally distributed.

Then, by CLT part 1, we know that our sampling distribution of the sample mean is also normally distributed.

**Sampling Distribution of the Sample Mean of the inside diameter of a randomly selected piston ring N(12, 0.008) n=25**



**(d) Find the chance that, for a sample of size 50, the average diameter is within 0.01 cm of the mean.**

- **Use correct probability notation,**
- **draw a picture that represents the probability you are finding,**
- **and justify your answer.**

$\mu_{\overline{x}} = \mu = 12$ cm

$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.04cm}{\sqrt{50}} = 0.0056568542 \approx 0.00566$

$\overline{x} \sim N(12, 0.08)$

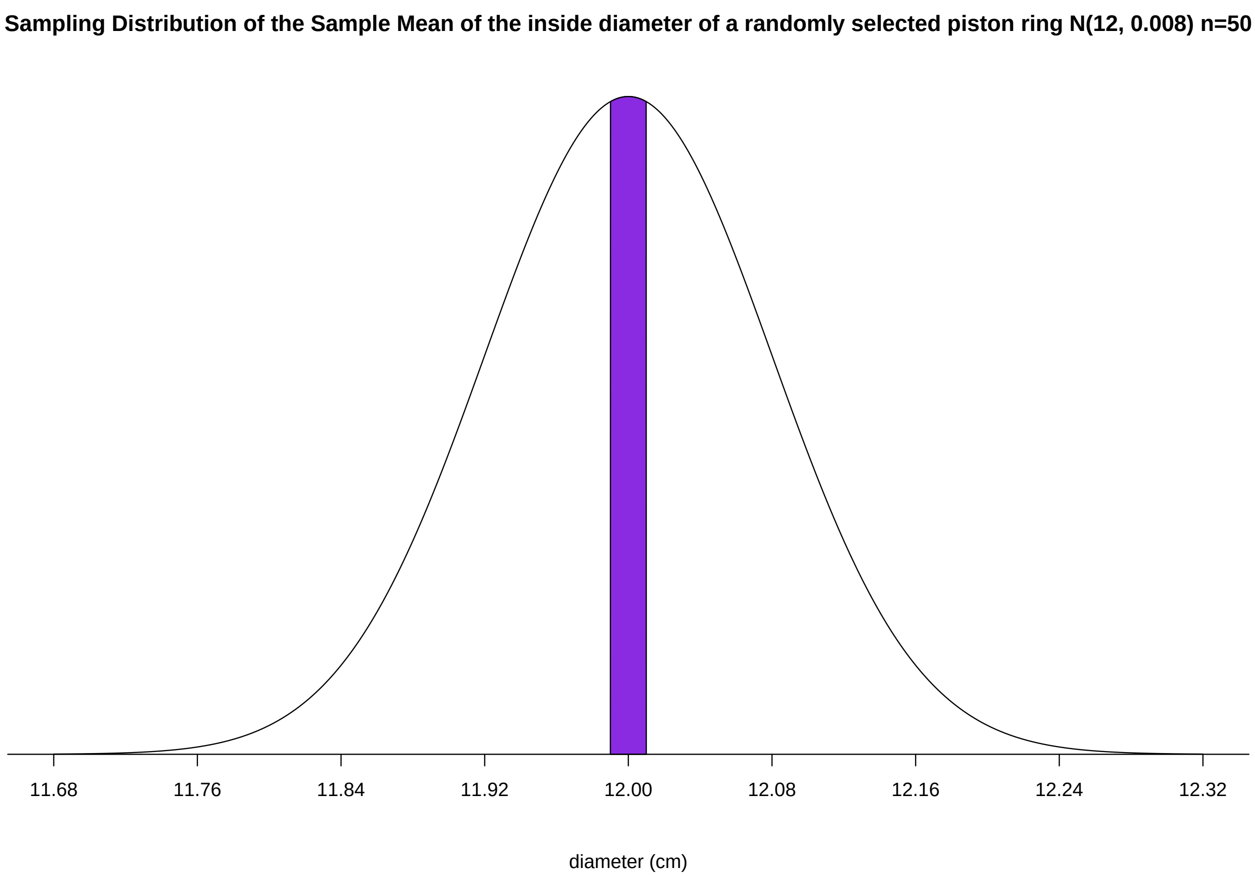
$P(11.99 \leq x \leq 12.01)$

```
pnorm(12.01,mean=12,sd=0.0056568542) - pnorm(11.99,mean=12,sd=0.0056568542)
```

```
## [1] 0.9229001
```

Because our population is described by a mean and standard deviation, we can then assume than the population is normally distributed.

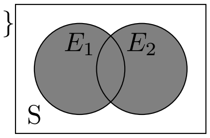
Then, by CLT part 1, we know that our sampling distribution of the sample mean is also normally distributed.



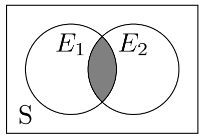
**5. This homework is intentionally short to give you more time for Lab 4 and to study for the test. You will, again, be permitted one 1-sided, hand-written sheet of notes on 8"x11" paper, in your own handwriting. To get you started early in studying for the test hand in with your homework:**

- **definition of union, intersection, complement, and empty set (and associated Venn diagrams)**

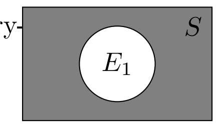
Union: $E_1 \cup E_2 = \{x \in S : x \in E_1 \vee x \in E_2, \text{(or both)}\}$




Intersection: $E_1 \cap E_2 = \{x \in S : x \in E_1 \wedge x \in E_2\}$



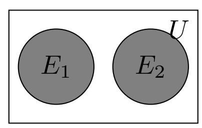

Complement: The complement of an event E is everything in S than is not in E. (E').




empty set: The empty set or null set is the event containing no outcomes.

- **axioms of probability & the special addition rule (and associated Venn diagram)**

Axioms of Probability and the Special Addition Rule (SAR)




- Let $E \subseteq S$ where S is the sample space.

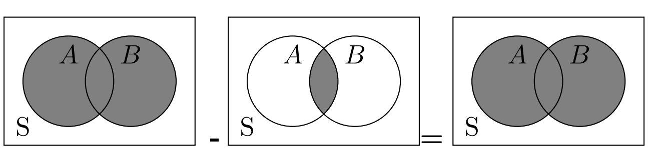
1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. Special Addition Rule (SAR)
   - (Can only be applied to unions when events are disjoint)
   - If $E_1$ and $E_2$ are mutually exclusive (i.e. 'disjoint') ($E_1 \cap E_2 = \emptyset$),
   - then $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

- **the general addition rule (and associated Venn diagram)**

General Addition Rule:

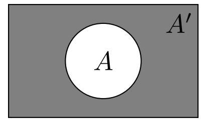
- For arbitrary (might overlap) events (not mutually exclusive)
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(First venn diagram has duplicate of $A \cap B$)



- **the complement rule (and associated Venn diagram)**

  - $P(A') = 1 - P(A)$




- **multiplication rule for independent events (MRFIE)**

If A and B are independent, then $P(A \cap B) = P(A)P(B)$

Only applies to intersection of independent events.

- **conditional probability formula**

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

(Probability of A given B).

- **Bayes' theorem (p121)**

If $P(A), P(B), P(B') > 0$, then $P(B|A) = \frac{P(B) \cdot P(A|B)}{P(B) \cdot P(A|B) + P(B') \cdot (P(A|B')}$

- **The quantity being minimized in a basic least-squares regression. (p73)**

Sum of Squared Errors (SSE) $= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

- **The meaning of the slope in a basic least-square regression and an example with a sentence.**

If we increase x by 1 unit, we expect y to increase/decrease by b units on average.

$b = r \cdot \frac{S_y}{S_x}$

Example: If y is the number of cars, and x is the time (sec), and the slope is 5x.

If we increase the time by 1 second, we expect the number of cars to increase by 5 on average.

- **The meaning of the coefficient of determination in a least-squares regression (for basic or more complex models) and an example with a sentence. (p 79)**

Coefficient of Determination $= R^2 = 1 - \frac{SSResid}{SSTot}$

Meaning of $R^2$ (general): In the context of the model $\hat{y} = a + bx, R^2$ is the proportion of variation in Y that is explained by a linear relationship (model) with x.

Meaning of $R^2$ (specific): If the model is $\hat{y} = a + b_1 x + b_2 x^2, R^2$ is the proportion of variation of Y explained by a quadratic relationship of x with y.

- **What a residual plot is, what it's used for, and how to interpret it (for basic and more complex least-square regression models).**

A residual plot is a plot of the residuals $(y - \hat{y})$ vs the predicted (fitted) values $(\hat{y})$.

It is used to verify that the model accurately fits the data, and that all of the variation in the data is accounted for by the model.

In all cases, we want the residual plot to resemble a amorphous cloud. If this is the case, we can state that the model accounts for all of the variation. However, if the residual plot does not resemble an amorphous cloud, we can say that the model does not account for all of the variation of the data. This can occur even if we have a good $R^2$ value, so we always need to check the residual plot as verification of a good model.