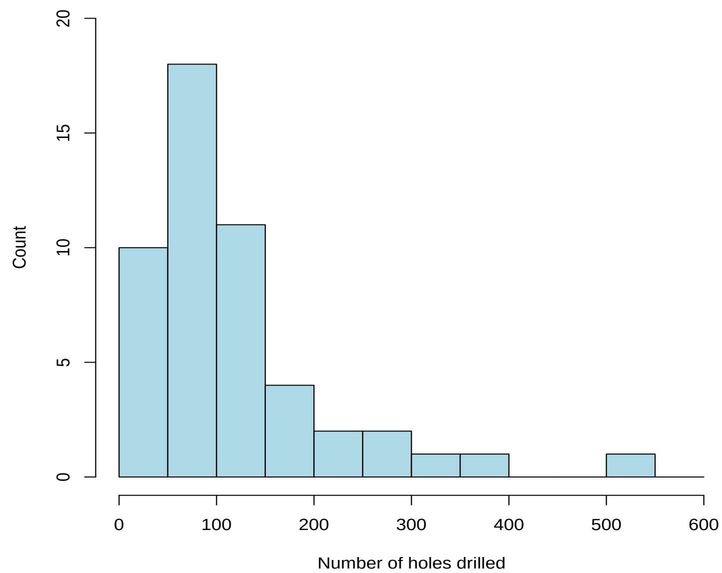
1. (a) We can think of the population as all microdrills of this specific type produced by its manufacturer. The data in the problem are for a sample of $n = 50$ microdrills. This is probably a small part of the entire population.
(b) I used R to produce the histogram below:



Insofar as selecting a population-level model (i.e., for the population of all microdrills), we would want to select one that matches the shape we see in the histogram. The gamma, the Weibull, and the lognormal all come to mind as potential choices. All of these distributions are skewed to the right, which is the shape we see in the histogram.
(c) An estimate of the population mean μ is the sample mean:

$$\overline{x} \approx 119.2 \text{ holes.}$$

An estimate of the population standard deviation σ is the sample standard deviation:

$$s \approx 97.4 \text{ holes.}$$

I used R to calculate both of these estimates:

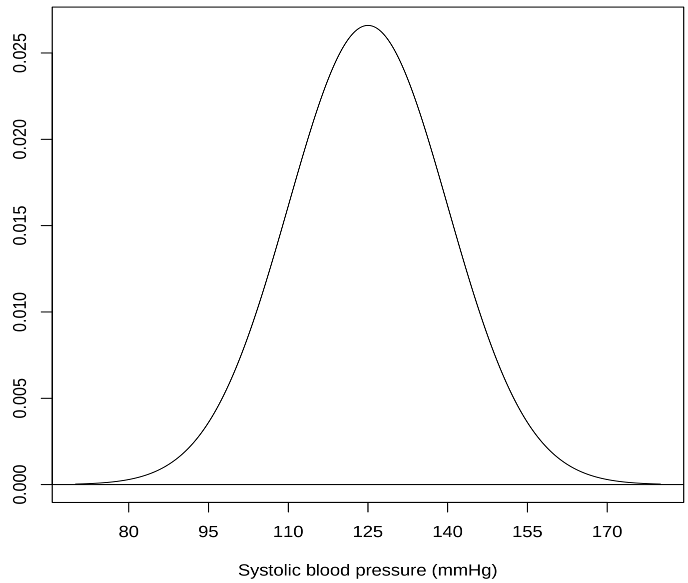
```
> options(digits=4)
> mean(holes)
[1] 119.2
> sd(holes)
[1] 97.4
```

2. (a) From Google's AI Overview,

> There were approximately 101 million American males aged 18 and over according to 2021 data from InfoPlease. This number comes from the 2020 Census data, where the total number of adults 18 and over was 209,128,094, with males making up 100,994,367 of that population.

The $\mathcal{N}(125, 15^2)$ distribution is the population-level model for SBP for all 101 million males in this population.

(b) Here is a graph of the population-level pdf:



From the 68-95-99.7% Rule,

- about 68% of American males will have SBP between 110 and 140 mmHg.
- about 95% of American males will have SBP between 95 and 155 mmHg.
- about 99.7% of American males will have SBP between 80 and 170 mmHg.

Here is where these percentages come from:

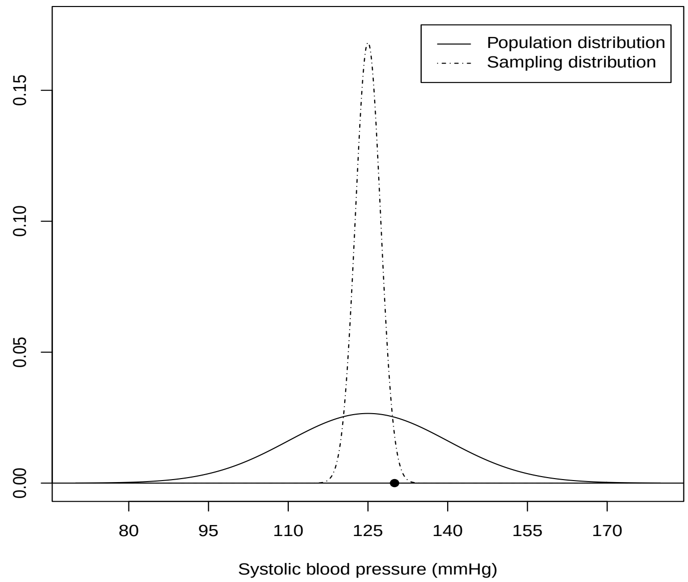
```
> options(digits=3)
> pnorm(140,125,15)-pnorm(110,125,15) # within one standard deviation of the mean
[1] 0.683
> pnorm(155,125,15)-pnorm(95,125,15) # within two standard deviations of the mean
[1] 0.954
> pnorm(170,125,15)-pnorm(80,125,15) # within three standard deviations of the mean
[1] 0.997
```

(c) We want $P(X > 130)$, the proportion of all American males aged 18 and over whose SBP is larger than 130 mmHg. From R,

```
> 1-pnorm(130,125,15) # 1-P(X<=130)
[1] 0.369
```

(d) In this part, we want $P(\overline{X} > 130)$, where $\overline{X}$ is the sample mean SBP of a random sample of $n = 40$ American males from this population. The sampling distribution of $\overline{X}$ is

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \implies \overline{X} \sim \mathcal{N}\left(125, \frac{15^2}{40}\right).$$



Therefore,

$$P(\overline{X} > 130) = P\left(\frac{\overline{X} - 125}{15/\sqrt{40}} > \frac{130 - 125}{15/\sqrt{40}}\right) \approx P(Z > 2.11) \approx 0.0174.$$

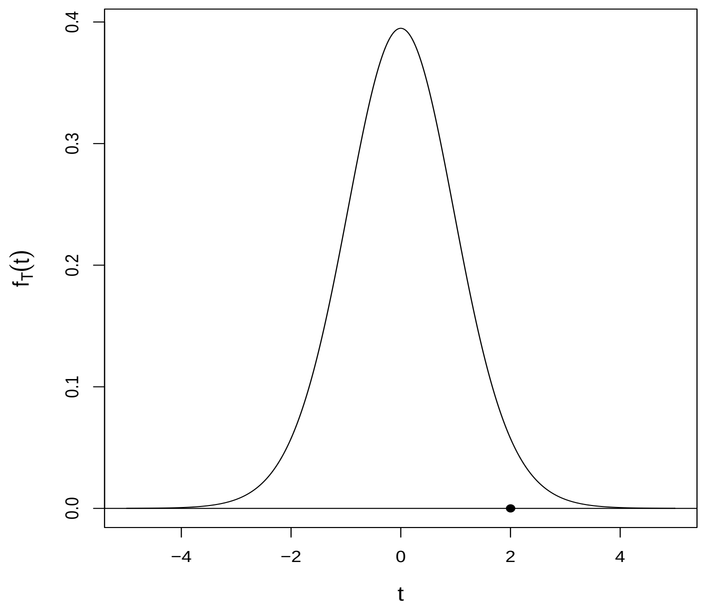
```
> 1-pnorm(2.11,0,1) # P(Z>2.11)
[1] 0.0174
```

Note that $P(X > 130)$ and $P(\overline{X} > 130)$ are different probabilities! $P(X > 130)$ is the probability a single American male (aged 18 and over) has SBP above 130 mmHg. $P(\overline{X} > 130)$ is the probability the sample mean SBP of 40 American males (aged 18 and over) is larger than 130 mmHg. The first probability uses the population distribution for all individuals. The second probability uses the sampling distribution of the sample mean; see the figure above.

3. (a) The sample mean and sample standard deviation are

$$\begin{aligned} \overline{x} &\approx 1.314 \text{ inches} \\ s &\approx 0.03497 \text{ inches.} \end{aligned}$$

```
> options(digits=4)
> mean(pipes)
[1] 1.314
> sd(pipes)
[1] 0.03497
```



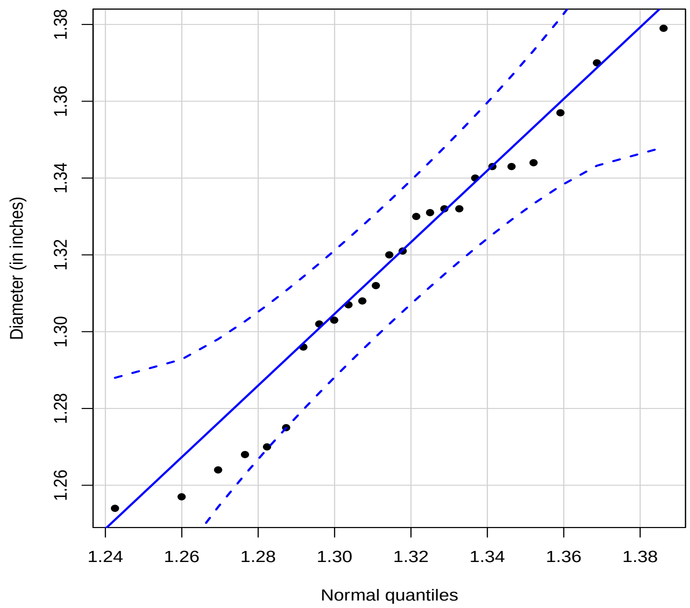
(b) If the manufacturer's claim is correct, then the population mean outside diameter is $\mu = 1.300$ inches. Under this assumption,

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{1.314 - 1.300}{0.03497/\sqrt{25}} \approx 2.002.$$

(c) See the $t(24)$ pdf above. The value of t is shown using a solid circle.
(d) If the manufacturer's claim was correct, t should hover around 0; see the figure above. The further away from 0 it is, the more evidence we have the manufacturer's claim is incorrect. This value of t falls in the upper right tail, which would be consistent with $\mu > 1.300$ inches (contradicting the manufacturer's claim). It doesn't fall "far out" in the upper tail, but it is certainly approaching what I would consider to be "unusual" if the manufacturer's claim was correct.
(e) A normal qq plot of the pipe diameter data is on the next page (top). There is general agreement between the observed data and the normal quantiles. We conclude the data are well represented by a normal distribution.

ADDITIONAL R CODE:

```
# Problem 1(b)
holes = c(11,14,20,23,31,36,39,44,47,50,59,61,65,67,68,71,74,76,78,79,
    81,84,85,89,91,93,96,99,101,104,105,105,112,118,123,136,139,141,148,158,
    161,168,184,206,248,263,289,322,388,512)
bins = seq(0,600,50)
hist(holes,breaks=bins,xlab="Number of holes drilled",ylab="Count",ylim=c(0,20),
    main="",col="lightblue")
```




```
# Problem 2(b)
x = seq(70,180,0.1)
plot(x,dnorm(x,125,15),type="l",lty=1,xlab="Systolic blood pressure (mmHg)",
    ylab="",xaxp=c(80,170,6))
abline(h=0)


# Problem 2(c)
x = seq(70,180,0.1)
plot(x,dnorm(x,125,15),type="l",lty=1,xlab="Systolic blood pressure (mmHg)",
    ylab="",xaxp=c(80,170,6),ylim=c(0,0.175))
abline(h=0)
# Add sampling distribution
lines(x,dnorm(x,125,15/sqrt(40)),lty=4)
points(x=130,y=0,pch=19,cex=1)
# Add legend
legend(135,0.175,lty = c(1,4),
    c(
    expression(paste("Population distribution")),
    expression(paste("Sampling distribution"))
    ))


# Problem 3
pipes = c(1.312,1.270,1.331,1.370,1.330,1.254,1.340,1.320,1.343,1.332,
    1.379,1.357,1.264,1.332,1.307,1.257,1.275,1.302,1.296,1.321,
    1.344,1.308,1.303,1.268,1.343)
```

```
# 3(c)
t = seq(-5,5,0.01)
plot(t,dt(t,24),type="l",lty=1,xlab="t",ylab=expression(f[T](t)),cex.lab=1.25)
abline(h=0)
points(x=2.002,y=0,pch=19,cex=1)

# 3(e)
qqPlot(pipes,distribution="norm",mean=mean(pipes),sd=sd(pipes),
    xlab="Normal quantiles",ylab="Diameter (in inches)",pch=16,
    envelope=list(border=TRUE,style="lines"),id=FALSE)
```