

1. Let  $X$  be a random variable from some population with mean  $\mu_X = 12$  cm and standard deviation  $\sigma_X = 0.04$ .

- (a) If  $\bar{X}$  is the sample mean from a sample of size  $n = 16$  then we know that the **mean** and **standard error** for the sampling distribution of  $\bar{X}$  are

$$\mu_{\bar{X}} = \mu_X = 12 \text{ cm} \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.04}{\sqrt{16}} = \frac{0.04}{4} = 0.01 \text{ cm}$$

Therefore, regardless of the underlying distribution, the sampling distribution of  $\bar{X}$  is centered at 12 cm with standard deviation 0.01 cm.

- (b) Similarly, if  $\bar{X}$  is the sample mean from a sample of size  $n = 64$ , then we know that the **mean** and **standard error** for the sampling distribution of  $\bar{X}$  are

$$\mu_{\bar{X}} = \mu_X = 12 \text{ cm} \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.04}{\sqrt{64}} = \frac{0.04}{8} = 0.005 \text{ cm}$$

Notice that the standard deviation is reduced by 50% if we increase the sample size from 16 to 64.

- (c) Both distributions are centered on 12 cm, but the sample mean based on the sample of size  $n=64$  will be more concentrated at the mean since its standard deviation is half the standard deviation of the one based on  $n = 16$ . This is true even if the population from which the samples were drawn isn't normally distributed.
- (d) If the population is  $N(\mu = 12, \sigma = 0.04)$  then we can compute the probabilities by standardizing to the standard normal distribution. First we note that

$$P(\bar{X} < 11.96 \quad \text{or} \quad \bar{X} > 12.04) = 1 - P(11.96 < \bar{X} < 12.04)$$

and

$$P(11.96 < \bar{X} < 12.04) = P\left(\frac{11.96 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < Z < \frac{12.04 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right).$$

For  $n = 1$ , this probability is

$$P(11.96 < \bar{X} < 12.04) = P\left(\frac{11.96 - 12}{0.04} < Z < \frac{12.04 - 12}{0.04}\right) = P(-1 < Z < 1) = 0.6826$$

For  $n = 16$ , this probability is

$$P(11.96 < \bar{X} < 12.04) = P\left(\frac{11.96 - 12}{0.01} < Z < \frac{12.04 - 12}{0.01}\right) = P(-4 < Z < 4) \approx 1$$

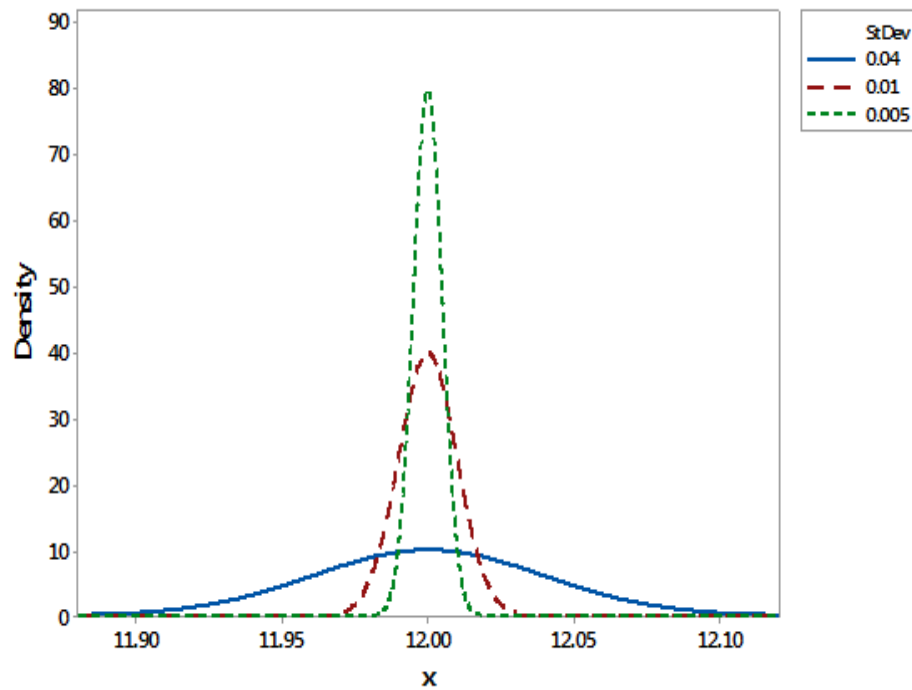
For  $n = 64$ , this probability is

$$P(11.96 < \bar{X} < 12.04) = P\left(\frac{11.96 - 12}{0.005} < Z < \frac{12.04 - 12}{0.005}\right) = P(-8 < Z < 8) \approx 1$$

So, for  $n = 1, 16$  and  $64$ , respectively,

$$P(\bar{X} < 11.96 \quad \text{or} \quad \bar{X} > 12.04) \approx 0.3174, 1, \text{ and } 1.$$

- (e) Below is the graph of all three sampling distributions,  $n = 1, 16$  and  $64$ , using the probability distributions (with varying parameters) feature in MINITAB.



2. We have a sample of size 277 of 18-year-old American males waist measurements with sample mean circumference of 86.3 cm, i.e.,  $n = 277$  and  $\bar{x} = 86.3$ .

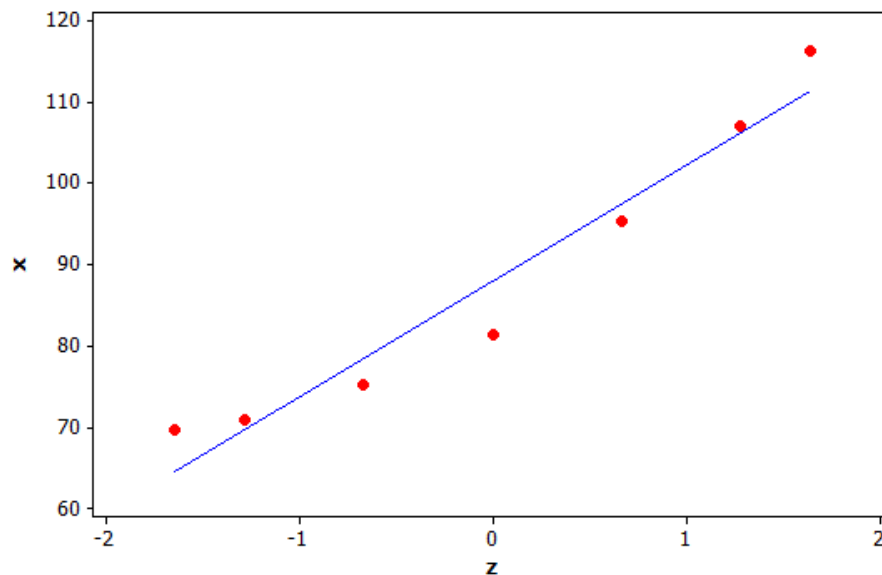
(a) The following sample percentiles were given in the problem,

$5^{th}$	$10^{th}$	$25^{th}$	$50^{th}$	$75^{th}$	$90^{th}$	$95^{th}$
69.6 cm	70.9 cm	75.2 cm	81.3 cm	95.4 cm	107.1 cm	116.4 cm

We must find the corresponding percentiles for the standard normal distribution (either using the table from the book or minitab, i used the book), to find the following percentiles

$5^{th}$	$10^{th}$	$25^{th}$	$50^{th}$	$75^{th}$	$90^{th}$	$95^{th}$
-1.645	-1.28	-0.67	0	0.67	1.28	1.645

We plot the sample percentiles against the standard normal percentiles and determine if the points fit on a straight line. If they do, then there is no obvious violation of normality.



It's not a perfect fit and the 5th and 10th sample percentiles rising above the line is indicative of heavier tails than a normal population, but there is no clear violation of the normality assumption. Also, note that the 5th, 10th, 25 percentiles are closer to the median than are the 75th, 90th and 95th percentiles and the sample mean is to the right of the sample median, which are indications of a potential positively skewed distribution.

- (b) if the population mean waist size is  $\mu = 85$  cm and the population standard deviation is  $\sigma = 15$  cm, then, from the central limit theorem (CLT), the sampling distribution of  $\bar{X}$  is approximately Normal with mean  $\mu_{\bar{X}} = 85$  and standard error  $\sigma_{\bar{X}} = 15/\sqrt{277} = 0.9013$ . Therefore, the probability that the sample mean waist size is at least 86.3 is

$$P(\bar{X} \geq 86.3) \approx P\left(Z \geq \frac{86.3 - 85}{(15/\sqrt{277})}\right) = P(Z \geq 1.44) = 1 - \Psi(1.44) = 1 - 0.9251 = 0.0749.$$

where  $Z$  is a standard normal random variable.

- (c) If the true mean waist size were  $\mu = 82$  instead then,

$$P(\bar{X} \geq 86.3) \approx P(Z \geq 4.77) = 1 - \Psi(4.77) \approx 0$$

So, it seems that the sample mean of at least 86.3 would more likely come from a population with mean 85 cm than a mean of 82 cm.

3. Let  $X$  be the breaking strength of a rivet from some population which has a mean of  $\mu_X = 10000$  psi and standard deviation of  $\sigma_X = 500$  psi.

- (a) If we take a random sample of size  $n = 40$  from the population of rivets, we know that the distribution of the sample mean, by the central limit theorem, is approximately normal with mean  $\mu_{\bar{X}} = 10000$  and standard deviation  $\sigma_{\bar{X}} = 500/\sqrt{40} \approx 79.06$ . and

$$P(9900 < \bar{X} < 10200) \approx P\left(\frac{9900 - 10000}{(500/\sqrt{40})} < Z < \frac{10200 - 10000}{(500/\sqrt{40})}\right) = P(-1.26 < Z < 2.53)$$

From the cumulative normal table,

$$P(-1.26 < Z < 2.53) = \Psi(2.53) - \Psi(-1.26) = 0.9943 - 0.1038 = 0.8860$$

so,

$$P(9900 < \bar{X} < 10200) \approx 0.8860.$$

- (b) When the sample size was  $n = 40 > 30$  we could use the normal approximation based on the central limit theorem. However, with a sample of size 15 and no further information about the shape of the underlying distribution, we can't assess how accurate the CLT approximation would be.

The sample mean

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{336.2}{27} = 12.45$$

The sample variance is

$$s^2 = \frac{\sum x_i^2 - (\sum x)^2/n}{n-1} = \frac{4314.99 - (336.2)^2/27}{26} = 4.95$$

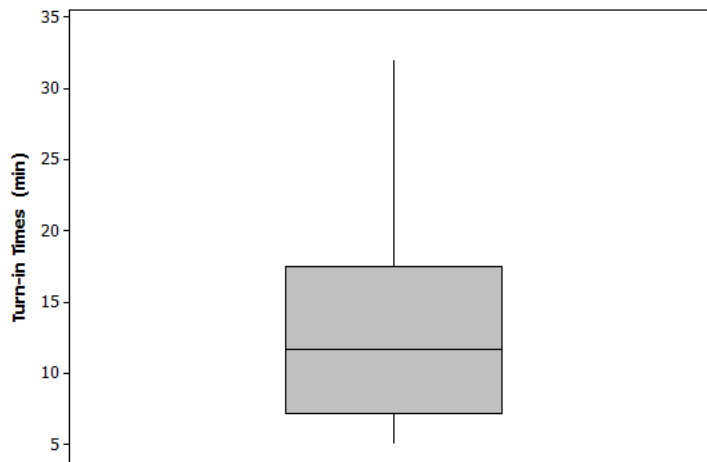
and, the sample standard deviation

$$s = \sqrt{s^2} = \sqrt{4.95} = 2.22$$

From MINITAB, the sample percentiles are

5 <sup>th</sup>	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	95 <sup>th</sup>
5.4	5.6	7.2	11.7	17.5	23.6	23.7

Also from MINITAB, the boxplot given below

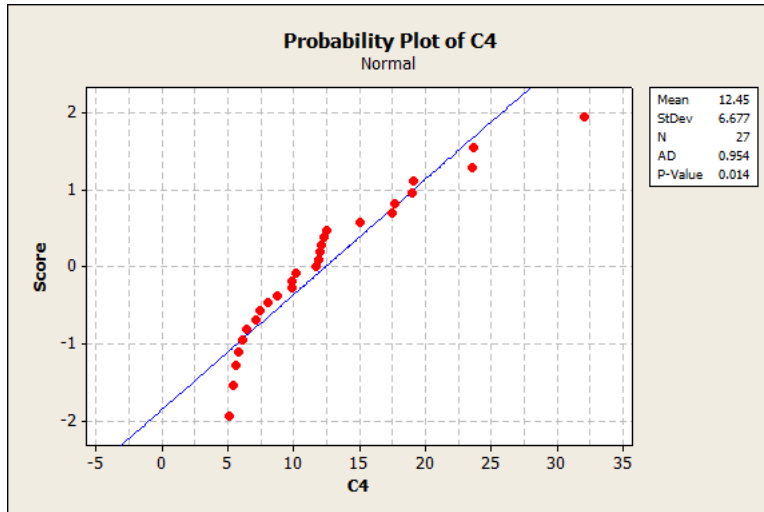


Notice that the sample mean  $\bar{x} = 12.45$  is greater than the sample median  $\tilde{x} = 11.7$ , indicative that the sample distribution may be positively skewed. The boxplot, above, confirms this the range for the first half of the data is much shorter than the second half. The sample median is closer to the 25th percentile than it is to the 75th percentile, and the distance between the third quartile to the upper whisker is much larger than the lower counterparts. This is definitely a positively skewed distribution. Take a look at the following stem-and-leaf plot.

Stem-and-leaf of C4 N = 27 Leaf Unit = 1.

freq	stem	leaves
12	0	5555666778899
(7)	1	0112222
8	1	57799
3	2	33
1	2	
1	3	2

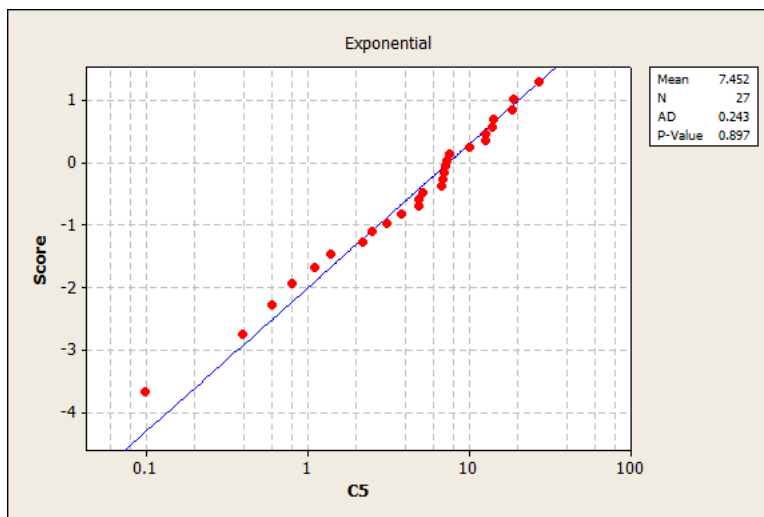
**Assessment of Normality:** There are a number of ways to do this, but we already know from the above discussion that the data tend to exhibit non-normality behavior. However, can calculate all sample percentiles, and compute the corresponding percentiles from the standard normal distribution and plot these as pairs of points in a scatterplot.



**Assessment of Exponentiality:** Recall for exponential ( $\beta$ ), the cdf is, for some  $x > 0$ ,  $F(x) = 1 - e^{-x/\beta}$ . For any  $p \in (0, 1)$ , set the cdf equal to  $p$  and solve for  $X_p$  to get

$$X_p = -\beta \ln(1 - p)$$

Find all of these for the standard exponential and plot against the sample percentiles,



Easy to see, since the data points fit tightly on a line, these data exhibit the properties of an exponential distribution.

**Sampling distribution of the sample mean for given scale:** if the population is exponentially distributed with scale  $\beta = 15$  mins, the each distribution of the sample mean from a sample of size  $n = 27$  is Gamma with shape parameter 27 and scale parameter  $15/27$ , i.e.,  $\bar{X}$  is Gamma(27,  $15/27$ ).