1. An investor is rating two potential additions to a stock portfolio on the basis of the likely value of the stock after holding it one year. For each dollar invested, the person models

**Stock A** price after one year as normal with mean \( \mu = 1.15 \) and \( \sigma = .05 \)

**Stock B** price after one year as normal with mean \( \mu = 1.25 \) and \( \sigma = .10 \)

a) Which of the two stocks does this person judge most likely to be worth at least $1.10 per dollar invested after one year? Show your work.

\[
Z_A = \frac{1.10 - 1.15}{.05} = -1 \\
Z_B = \frac{1.10 - 1.25}{.10} = -1.5 \\
\]

\[
P(Z > -1.5) > P(Z > -1.0)
\]

Circle one: **Stock A** or **Stock B**

b) The investor is 90\% sure that Stock A will be worth at least how much per dollar invested after holding it one year?

\[
z = Z + Z \cdot \sigma \\
= 1.15 + (-1.282)(.05) \\
= 1.086
\]

2. The book *Data, Statistics and Decision Models with Excel* by Harnett and Horrell contains information from a real bill sent by the IRS to a real company demanding payment of back excise taxes. A (random) sample of \( n = 160 \) company sales invoices (from a much larger pool of 22,794 invoices that were in dispute) produced \( \bar{x} = \$11.21 \) owed per invoice with \( s = \$20.59 \).

a) We will assume that $0 is the minimum that can be owed on any particular invoice. From this and the sketchy information about the sample supplied above, is it possible that the data set had a roughly normal/bell-shaped histogram? Why or why not?

*These data could not have been approximately normal. All the observations are \( \geq 0 \) and therefore no more than \( .544 \) standard deviations below the mean. A normal-looking distribution will be symmetric about a mean extending roughly 3 standard deviations on each side.*
b) Give approximate 90% confidence limits for the mean excise tax owed per invoice.

\[ \text{Use } \bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad \text{i.e.} \quad 11.21 \pm 1.655 \frac{20.59}{\sqrt{160}} \]
\[ \text{i.e.} \quad 11.21 \pm 2.69 \]

205

c) Does the validity of your answer in b) depend upon the 22,794 invoices having a normal
relative frequency distribution of excise tax owed? Explain.

No. The sample size is large and the t-intervals
are "robust" against non-normality for large n.

See Moore page 380.

523

d) Apparently, the bill sent to the company was for $194,647, or about $8.54 per invoice. Put
yourself in the position of an IRS supervisor. The sample mean ($11.21) is substantially larger
than the amount billed. Should you be convinced that the $8.54 was clearly too low (and that
your agents should clearly have sent a larger bill)? (Show appropriate null and alternative
hypotheses and report an appropriate p-value.)

\[ H_0 : \mu = 8.54 \]
\[ H_a : \mu > 8.54 \]
\[ T = \frac{11.21 - 8.54}{\frac{20.59}{\sqrt{160}}} = 1.64 \]
\[ p\text{-value} = P(T > 1.64) \approx 0.0518 \]

This is fairly strong, but not overwhelming evidence
that the bill was too small (note that by the standards
3. Attached to this exam is a JMP report for the analysis of some data from \( n = 65 \) case files
of tax preparation company. Represented are values of

\[ y = \text{deductions claimed} \]

and also \( \sqrt{y} \).

a) Describe the shape of the \( y \) distribution in 10 words or less.

The distribution is skewed right.
b) Using the scale below, make a boxplot for the deduction data.

![Boxplot Image]

\[
\begin{align*}
0 & \quad 5000 \quad 10,000 \quad 15,000 \quad 20,000 \\
\end{align*}
\]

c) Ignore the graduated nature of the income tax and suppose that each deduction dollar produces $.28 worth of reduction in tax owed. That is, consider the quantity

\[x = .28y\]

What are the mean and standard deviation of \(x\)?

\[
\bar{x} = .28 \bar{y} \quad \text{and} \quad s_x = .28 s_y
\]

\[
\bar{x} = 1621 \quad s_x = 867
\]

d) The first histogram suggests that using the raw deduction data, \(y\), one should probably NOT use the formula from class to make a prediction interval for \(y_{\text{new}}\). But in this regard, the second histogram looks "better" than the first. Use the second part of the JMP report and make a 95\% prediction interval for \((\sqrt{y})_{\text{new}}\). Then square the end points of your interval to get a prediction interval for a single additional deduction value.

Interval for an additional \(\sqrt{y}\):

\[
\bar{x} \pm t s \sqrt{1 + \frac{1}{n}}
\]

\[
78.87 \pm (2.01842) \sqrt{1 + \frac{1}{6}}
\]

\[
78.87 \pm 37.12
\]

\[
(397.5, 110.99)
\]

Interval for an additional \(y\) from the above:

\[
12.35 \quad 12.315
\]