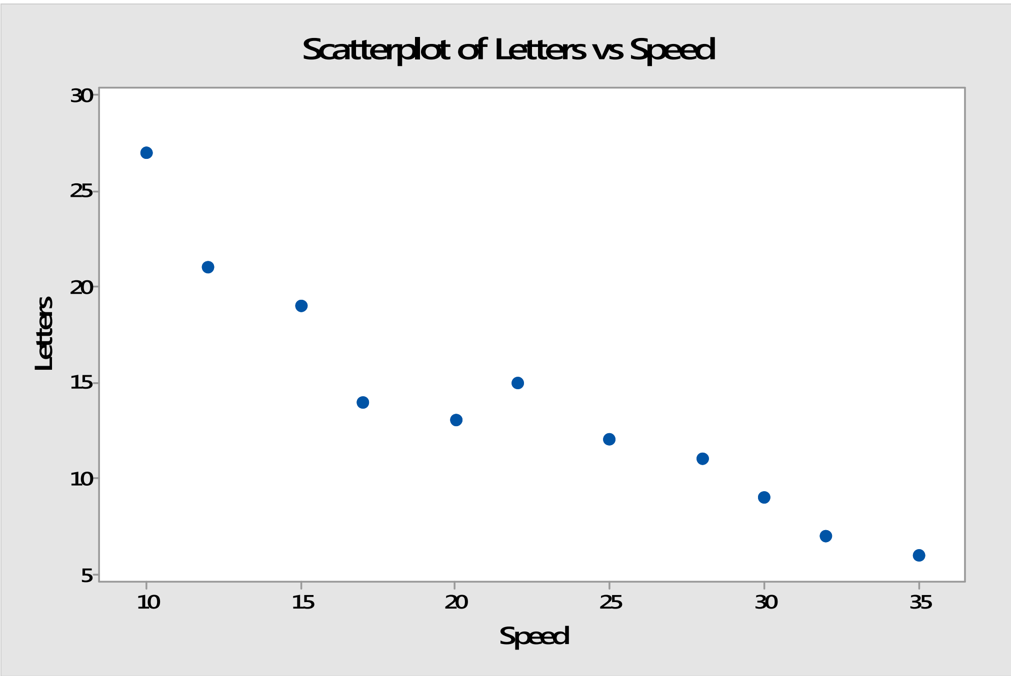
Solutions to Assignment #4 (19 Marks)

Question 1. (4 points)

a) (1 point) The scatterplot using MINITAB is



b) (1 point)

The scatterplot indicates that as the conveyor belt speed increases, the number of misclassified letters decreases in a linear fashion. Thus a linear association between conveyor speed and number of misclassified letters is reasonable. In particular, a strong, negative (inverse) relationship between the two variables is visible.

c) (2 points) The MINITAB results are:

Correlation: Speed, Letters

```
Pearson correlation of Speed and Letters = -0.945
P-Value = 0.000
```

Since $r = -0.945$ then there is a strong, inverse relationship between conveyor belt speed and number of misclassified letters. As the line speed increases, the number of misclassified letters found decreases, thus the manager's concerns seem to be justified.

Question 2. (4 points)

a) (1 point) We have $\hat{p} = \frac{x}{n} = \frac{41}{100} = 0.41$.

b) (3 points) Since the manager believed the true proportion was 52%, then $p = 0.52$.

To find $P(\hat{p} \leq 0.41)$, we use the sampling distribution of $\hat{p}$. Since $np = 100(0.52) = 52 > 10$ and $n(1-p) = 100(1-0.52) = 48 > 10$ then the sampling distribution of $\hat{p}$ is approximately Normal with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

So $P(\hat{p} \leq 0.41) = P\left(Z \leq \frac{0.41-p}{\sqrt{\frac{p(1-p)}{n}}}\right) = P\left(Z \leq \frac{0.41-0.52}{\sqrt{\frac{0.52(1-0.52)}{100}}}\right) = P(Z \leq -2.20) = 0.0139.$

Question 3. (6 Points)

a) (2 points) X = inorganic mercury content in a single cigarette, where $\mu_X = 25$ and $\sigma_X = 12$.

Since $n = 100 > 30$ then by the Central Limit Theorem, the sampling distribution of the sample mean is approximately normally distributed as $\bar{X} \sim approx\ N(\mu_X, \frac{\sigma_X}{\sqrt{n}})$.

In this example, $\bar{X} \sim approx\ \ N(25, 12)$ since $\frac{\sigma_X}{\sqrt{n}} = \frac{12}{\sqrt{100}} = 1.2$

b) (2 points) Since σ_X is estimated reliably (i.e. is known) then we have:

$$P(\bar{X} > 29) = P\left(Z > \frac{29 - \mu_{\bar{X}}}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z > \frac{29-25}{1.2}\right) = P(Z > 3.33) = P(Z < -3.33) = 0.0004$$

c) (2 points) In this case since the sample standard deviation is used, we employ the *t*-distribution to calculate the probability as:

$P(\bar{X} > 29) = P\left(T > \frac{29-\mu_{\bar{X}}}{\frac{s}{\sqrt{n}}}\right) = P\left(T > \frac{29-25}{2}\right) = P(T > 2) \approx 0.025$ with df = 24.

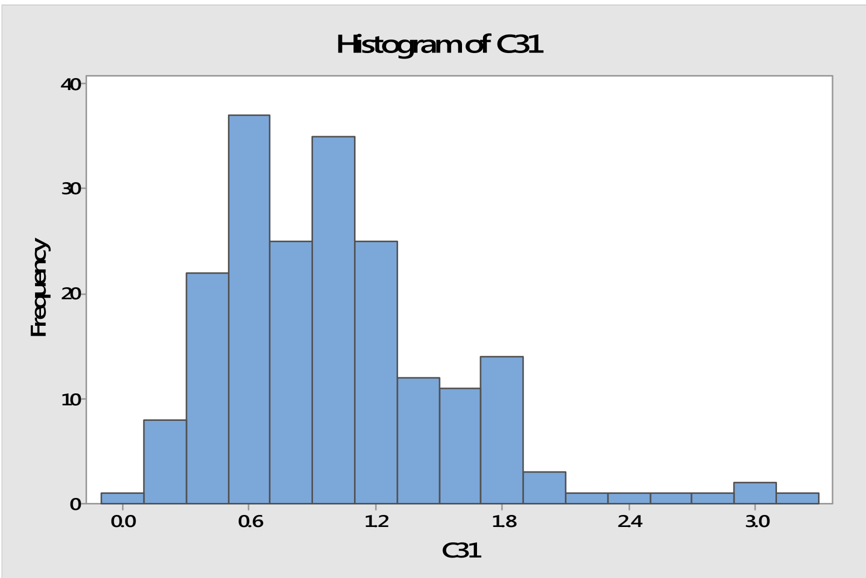
(Equivalently it is correct to state that 0.05 < $P(T > 2)$ < 0.025 since t=2.064 with 24 degrees of freedom).

Question 4. (5 Points)

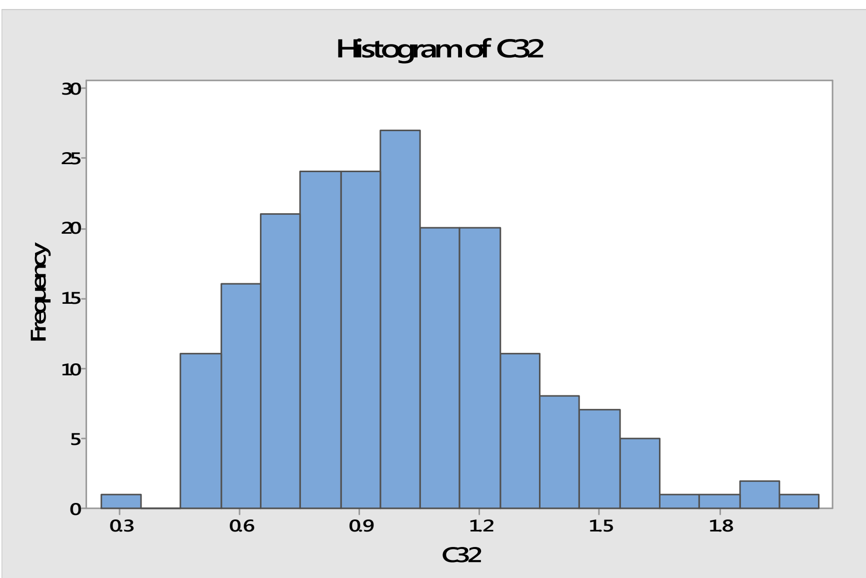
a) (4 points – 1 mark for each histogram, 1 mark for correct interpretation)

Note that answers will vary due to randomization

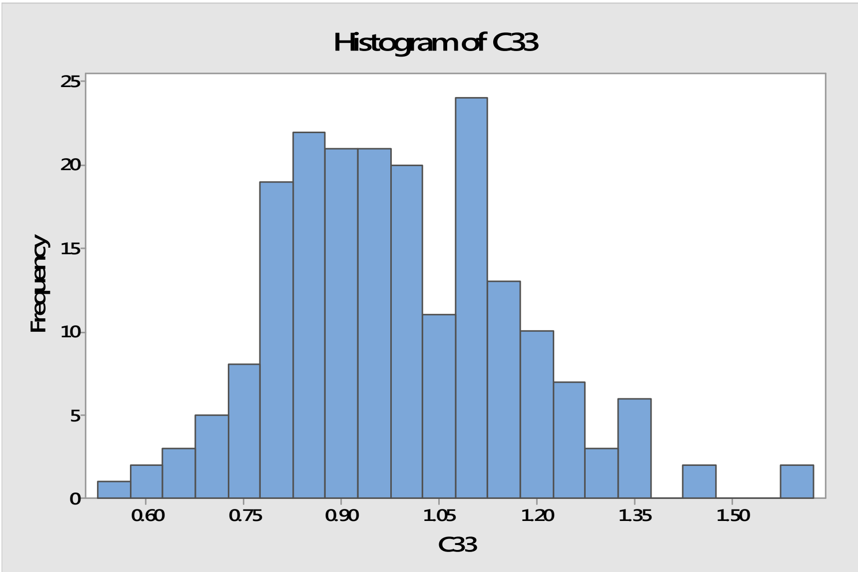
Histogram of sample means with n = 3:



Histogram of sample means with n = 10:



Histogram of sample means with n = 30:



Answers will vary, but the histograms appear more normally distributed (bell shaped and symmetric) as the sample size increases from n = 3 to n = 30.

b) (1 point)

The theoretical result outlined in this procedure is the Central Limit Theorem. For any distribution, as the sample size increases, the sampling distribution of the sample mean is approximately normally distributed. The approximation improves as the sample size increases.