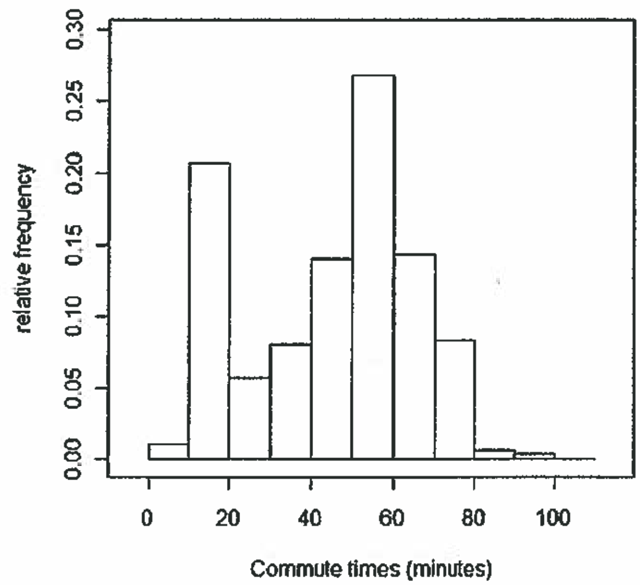
Solution key   Version 1

1. A university has two campuses – a downtown campus and a suburban campus. A total of 300 students from both campuses are surveyed. The following is a histogram of the commute times (in minutes) from home to campus the students spend on a typical weekday.



a) Describe the shape of the distribution of the commute times. [1 mark]

Bimodal (one peak at ~15 and another at ~55) and asymmetrical.

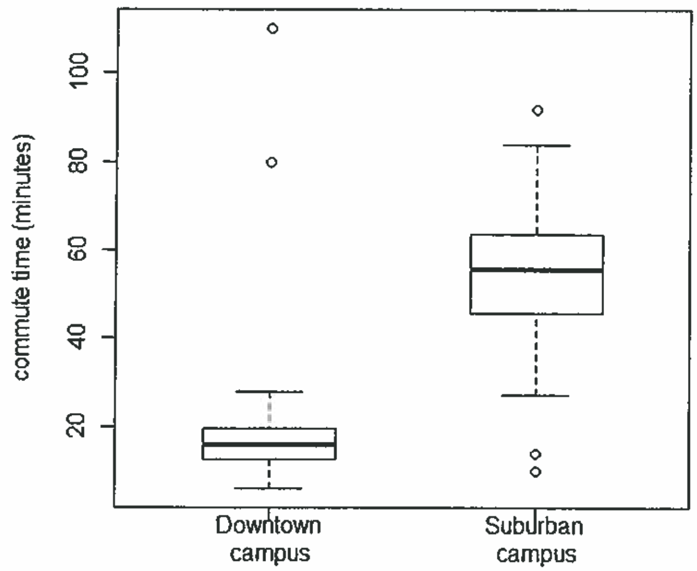
b) Where does the mean commute time fall? Circle the most appropriate answer. [2 marks]

A. Below 50 minutes.
B. Between 50 and 60 minutes.
C. Above 60 minutes.

c) Where does the 40*th* percentile of the commute times fall? Circle the most appropriate answer. [2 marks]

A. Between 20 and 30 minutes.
B. Between 30 and 40 minutes.
C. Between 40 and 50 minutes.
D. Between 50 and 60 minutes.

d) We now look at the commute times for students going to the two campuses separately. The following are side-by-side boxplots of the commute times.



Which of the following statements is / are correct? Check all that apply. [4 marks]

____ For students going to the downtown campus, the maximum commute time is between 20 and 40 minutes.

____ For students going to the downtown campus, the distribution of the commute times is roughly symmteric.

_✓_ The commute times of students going to the suburban campus have a larger IQR than those of students going to the downtown campus.

_✓_ Over 75% of students going to the suburban campus spend more than 40 minutes on commuting.

____ None of the above.

e) The sample of 300 students who are surveyed consists of a simple random sample of 80 students drawn from the downtown campus and a simple random sample of 220 drawn from the suburban campus. Identify the population and the sampling method used. [4 marks]

Population: All students of the university (from both campuses)

Sampling method: stratified random sampling (campus is a stratum)

2. In the same survey, the 300 students are also asked if they take bus or drive to campus. Among the 80 students going to the downtown campus, 48 take bus and 32 drive. Among the 220 students going to the suburban campus, 132 take bus and 88 drive.

a) A student is randomly chosen from the 300 students.
Define two events:
A = The student goes to the downtown campus.
B = The student drives.

Are A, B independent events? Justify your answer. [4 marks]

$P(A) = \frac{80}{300}$ $P(B) = \frac{32+88}{300} = \frac{120}{300}$

$P(A \text{ and } B) = \frac{32}{300} = 0.1067$

$P(A) \times P(B) = \frac{80}{300} \times \frac{120}{300}$

$= 0.1067$

$= P(A \text{ and } B)$

$\boxed{A, B \text{ are independent.}}$

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

$= \frac{32/300}{120/300}$

$= 32/120$

$= 0.2667$

$= P(A) = \frac{80}{300}$

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

$= \frac{32/300}{80/300}$

$= 32/80$

$= 0.4$

$= P(B) = \frac{120}{300}$

b) Two students are randomly chosen from the 300 students without replacement. Given that both students drive, what is the probability that they go to different campuses? Circle your answer. You need not show any calculation. [2.5 marks]

A. 0.0314 B. 0.0628

C. 0.1972 D. 0.3944 ← $\frac{32}{120} \times \frac{88}{119} + \frac{88}{120} \times \frac{32}{119}$

c) Five students are randomly chosen from the 300 students with replacement. What is the probability that at least one drives to campus? Circle your answer. You need not show any calculation. [2.5 marks]

A. $5(\frac{120}{300})^4(\frac{180}{300})$ B. $1 - (\frac{120}{300})^5$

C. $1 - (\frac{180}{300})^5$ D. $(1 - \frac{120}{300})^5$

P(at least 1 drives)

$= 1 - P(\text{none drives})$

$= 1 - P(\text{all take bus})$

$= 1 - \left(\frac{48+132}{300}\right)^5$

3. CleanSmile Corporation makes electric toothbrushes. The lifetimes of a model of electric toothbrush manufactured by the company are normally distributed with mean 34 months and standard deviation 7 months. Any toothbrush that fails during the warranty period will be replaced free by the company.

a) How long should the warranty period be if the company does not want to replace more than 2.5% of the toothbrushes? [4 marks]

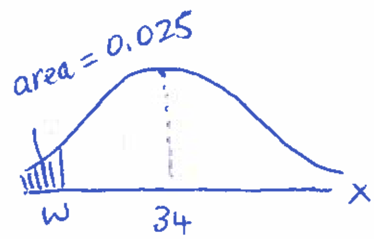
Answer: The warranty period is 20.28 months.

Show your calculation:

X = lifetime of a toothbrush $\sim N(\mu = 34, \sigma = 7)$

Toothbrush will be replaced if its lifetime X is shorter than the warranty period w

i.e. $P(X < w) = 0.025$



z-score for w is -1.96

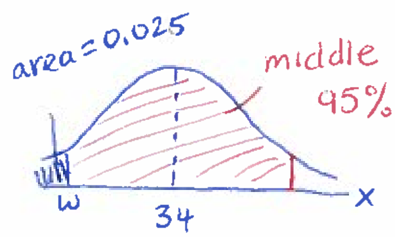
$$-1.96 = \frac{w - \mu}{\sigma} = \frac{w - 34}{7}$$

$$w = 34 - 1.96(7) = 20.28$$

OR use 68-95-99.7 rule

w is 2 SD below μ

$$w = 34 - 2(7) = 20$$



b) Fill in the blanks. Give your answers in two decimal places. You need not show any calculation. [2 marks]

The lifetimes of the electric toothbrushes in years will have

mean $\underset{\uparrow \ \frac{34}{12}}{2.83 \text{ yr}}$ and variance $\underset{\uparrow \ \frac{7^2}{12^2}}{0.34 \text{ yr}^2}$.

4. In a factory, trainees are trained to operate a device to assemble a product. By the end of a one-month training period, they will be hired if they successfully assemble 5 or more products within an hour. Based on the existing training method, the number of products that are assembled within an hour by a trainee after training, X, has this probability distribution:

x	2	3	4	5	6
$P(X = x)$	0.01	0.01	0.16	0.67	0.15

a) $E(X)$ is computed to be 4.94. Interpret this value in the context of this question. [2 marks]

If we repeatedly observe the trainees, the long run average of # products assembled within an hour by these trainees is 4.94.

b) What is the probability that in a random sample of 130 trainees, more than 80% are hired at the end of the training period? You must use an approximation method to solve this problem. State any assumptions you make in your calculation. [6 marks]

Y = # trainees who are hired $\sim \text{Bin}(n = 130,\ p = 0.67 + 0.15 = 0.82)$

↑ probability of assembling ≥5 products

$\hat{p}$ = sample proportion of 130 trainees who will be hired

Check conditions:

(1) The sample is random

(2) $n = 130 < 5\%$ of all trainees (this is an assumption)

(3) $np = 130(0.82) = 106.6 > 10$; $n(1-p) = 130(0.18) = 23.4 > 10$

Normal approx. to $\hat{p}$

$$\hat{p} \overset{\text{approx}}{\sim} N\left(\mu(\hat{p}) = 0.82,\ \sigma(\hat{p}) = \sqrt{\frac{0.82(0.18)}{130}} = 0.0337\right)$$

$P(\hat{p} > 0.8)$

$= P\left(\dfrac{\hat{p} - 0.82}{0.0337} > \dfrac{0.8 - 0.82}{0.0337}\right)$

$= P(Z > -0.59)$

$= 1 - 0.2776$

$= 0.7224$

Normal approx to Binomial

$$Y \overset{\text{approx.}}{\sim} N\left(\mu = np = 106.6,\ \sigma = \sqrt{np(1-p)} = 4.3804\right)$$

$P(Y > 0.8 \times 130)$

$= P(Y > 104) = P(Y \geq 105)$

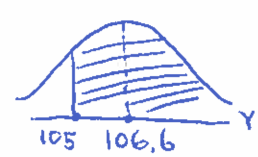
$= P(Y > 104.5)$ ← continuity correction

$= P\left(\dfrac{Y - 106.6}{4.3804} > \dfrac{104.5 - 106.6}{4.3804}\right)$

$= P(Z > -0.48)$

$= 1 - 0.3156$

$= 0.6844$



c) There is a new method for training workers to operate the device. The training manager plans to conduct a randomized block design experiment to compare the new method with the existing method. He will block the experiment by the handedness of the trainees. Since left handers operate the device differently from right handers when assembling products, handedness may affect productivity. The incoming group of 100 trainees will participate in the experiment.

i. Explain the purpose of blocking by handedness of the trainees. [2 marks]

To eliminate the effect of handedness on productivity of the trainees when comparing the 2 training methods.

ii. What are the treatments? [3 marks]

How many? 2

List them all: (1) existing training method
(2) new training method

iii. What is the response variable? Limit your answer to 10 words. [2 marks]

The number of products that are assembled within an hour by a trainee

iv. Describe how the 100 trainees are assigned to the different treatments. [3 marks]

The trainees are first divided into 2 groups by their handedness.

Within each group (block), the trainees are randomized to one of the 2 treatments.

5. In a city with over 500,000 registered vehicles, the costs of auto insurance of all registered vehicles have a mean of \$1423 and a standard deviation of \$533.2.
Sixty-five percent of all registered vehicles have auto insurance that costs over \$1423.

a) True or false? The distribution of auto insurance costs of all registered vehicles is approximately Normal. [3 marks]

Circle your answer: *True* *False*

Justify your answer:

The distribution is not symmetric - more than 50% of insurance costs are greater than the mean. Hence the distribution cannot be Normal.

b) Sixteen registered vehicles are randomly chosen. What is the standard deviation of the total auto insurance costs of the 16 vehicles? Circle your answer. [2 marks]

A. \$33.325 B. \$133.3

C. \$2132.8 D. \$8531.2

c) Sixteen registered vehicles are randomly chosen. The number of vehicles (out of the 16 vehicles) that have auto insurance costing over \$1423 will follow (circle only one answer) [2.5 marks]

A. approximately the Normal distribution with standard deviation $\frac{533.2}{\sqrt{16}}$.
B. approximately the Normal distribution with standard deviation $\sqrt{16(0.65)(1-0.65)}$.
C. approximately the Normal distribution with standard deviation $\sqrt{\frac{(0.65)(1-0.65)}{16}}$.
D. the Binomial distribution with standard deviation $\sqrt{16(0.65)(1-0.65)}$.
E. both (B) and (D).
F. both (C) and (D).

d) Consider repeated random samples of 100 registered vehicles. The average auto insurance cost of the 100 registered vehicles over these repeated samples will follow (circle only one answer) [2.5 marks]

A. a non-Normal distribution with standard deviation 533.2.
B. approximately the Normal distribution with standard deviation $\frac{533.2}{\sqrt{100}}$.
C. approximately the Normal distribution with standard deviation $\sqrt{100(0.65)(1-0.65)}$.
D. approximately the Normal distribution with standard deviation $\sqrt{\frac{(0.65)(1-0.65)}{100}}$.
E. the Binomial distribution with standard deviation $\sqrt{100(0.65)(1-0.65)}$.
F. Both (C) and (E).