Exercise 4A.1

There are three components of the GRE general test - Analytical, Verbal, Quantitative. Suppose the scores of prospective graduate students from Ohio have the following distributions:

Analytical is $N(580, 2500)$
Verbal is $N(670, 900)$
Quantitative is $N(650, 1600)$

(a) What is the expected total score for a randomly selected student from Ohio?

ANSWER: 1900

Let A = analytical score, V = verbal score, Q = quantitative score, T = total score
$E(T) = E(A + V + Q) = E(A) + E(V) + E(Q) = 580 + 670 + 650 = 1900$

(b) Suppose 3 prospective students are chosen at random. Let Y be the sum of their three <u>verbal</u> scores. What is the expected value and standard deviation of Y?

ANSWER: $E(Y) = 2010$, $V(Y) = 2700$

Let V_1, V_2, V_3 be the scores of the three students. Each has a $N(670, 900)$ distribution.
$E(Y) = E(V_1 + V_2 + V_3) = E(V_1) + E(V_2) + E(V_3) = 3 \times E(V) = 3 \times 670 = 2010$
Assuming the three students' scores are independent:
$V(Y) = V(V_1 + V_2 + V_3) = V(V_1) + V(V_2) + V(V_3) = 3 \times Var(V) = 3 \times 900 = 2700$
$SD(Y) = \sqrt{V(Y)} = \sqrt{2700} = 51.96$

Exercise 4A.2

The number of people wishing to check out at a supermarket averages 3.4 people per 10 minutes at peak time. Let X be a random variable indicating the number of people checking out in 10 minutes.

(a) What is the distribution of X? (Be specific.)

ANSWER: $X \sim Poisson(\lambda = 3.4)$

Since we are counting something (people checking out) in a fixed time period, this is the Poisson distribution. The parameter is the rate – 3.4 people per 10 minutes. So $X \sim Poisson(\lambda = 3.4)$.

(b) What is the expected number of people who will check out in a 20 minute interval? (Hint: think of this as two independent 10-minute intervals.)

ANSWER: 6.8

X_1 = number of people checking out in the first 10 minutes
X_2 = number of people checking out in the next 10 minutes
$X_1 + X_2$ = total number in 20 minutes
Both X_1 and X_2 have Poisson(3.4) distributions → $E(X_1) = E(X_2) = 3.4$
$E(X_1 + X_2) = E(X_1) + E(X_2) = 3.4 + 3.4 = 6.8$

Exercise 4A.3

Let X be a random variable indicating the minutes after 8am that an instructor starts class. Suppose it is known that $E(X) = 2.3$ and $\text{Var}(X) = 1.44$. Further assume that the start time of different days are independent of each other.

(a) Let Y be the total number of minutes late over a random sample of 20 class days. What are the expected value and variance of Y?

ANSWER: $E(Y) = 46$, $V(Y) = 28.8$

$$Y = \sum_{i=1}^{20} X_i$$

$E(Y) = E\left(\sum_{i=1}^{20} X_i\right) = 20 \times E(X) = 20 \times 2.3 = 46$

$V(Y) = V\left(\sum_{i=1}^{20} X_i\right) = 20 \times V(X) = 20 \times 1.44 = 28.8$

(b) What are the expected value and variance of $Y/20$? (This is the average minutes late over the 20 days.)

ANSWER: $E(Y/20) = 2.3$, $V(Y/20) = 0.72$

$E(Y/20) = E(Y)/20 = 46/20 = 2.3$
$V(Y/20) = V(Y)/(20^2) = 28.8/400 = 0.072$

Exercise 4A.4

Suppose that among all OSU undergrads, the mean age is 20.5 years with a standard deviation of 2.2 years. If you randomly select 25 OSU undergrads, what is the probability that the mean of your sample will be between 20 and 21?

ANSWER: 0.7458

X_i = *age of randomly selected OSU undergrad*

$E(X_i) = \mu = 20.5$ *and* $SD(X_i) = \sigma = 2.2$

$n = 25$

Mean of the sample $= \overline{X} = \dfrac{1}{25} \sum_{i=1}^{25} X_i$

By Central Limit Theorem, $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$, *thus* $\overline{X} \sim N\left(20.5, \dfrac{2.2^2}{25}\right) = N(20.5, 0.1936)$

$P(20 < \overline{X} < 21) = P\left(\dfrac{20 - 20.5}{\sqrt{0.1936}} < Z < \dfrac{21 - 20.5}{\sqrt{0.1936}}\right)$

$\qquad\qquad\qquad = P(-1.14 < Z < 1.14)$

$\qquad\qquad\qquad = 1 - 2 \times P(Z > 1.14) = 1 - 2 \times 0.1271 = 0.7458$

$\qquad\qquad\qquad$ *(draw a picture to help with that last step. . .)*

Exercise 4A.5

A hospital administrator believes that for a particular hospital the average ER wait time is 3 hours and that the standard deviation is 1.5 hours.

(a) If this is true and you sample 20 patients, how likely is it that your sample will have a mean wait time of greater than 3.5 hours?

ANSWER: 0.0681

$E(X_i) = \mu = 3$ *and* $SD(X_i) = \sigma = 1.5$

$n = 20$

$E(\overline{X}) = \mu = 3$ *and* $SE(\overline{X}) = \sigma/\sqrt{n} = 1.5/\sqrt{20} = 0.3354$

By Central Limit Theorem, $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$, *thus* $\overline{X} \sim N\left(3, \dfrac{1.5^2}{20}\right) = N(3, 0.1125)$

$P(\overline{X} > 3.5) = P\left(Z > \dfrac{3.5 - 3}{\sqrt{0.1125}}\right) = P(Z > 1.49) = 0.0681$

(b) What is this probability if you sample 50 patients instead?

ANSWER: 0.0091

μ *and* σ *are same as part (a), but now* $n = 50$

By Central Limit Theorem, $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$, *thus* $\overline{X} \sim N\left(3, \dfrac{1.5^2}{50}\right) = N(3, 0.045)$

$P(\overline{X} > 3.5) = P\left(Z > \dfrac{3.5 - 3}{\sqrt{0.045}}\right) = P(Z > 2.36) = 0.00091$

(c) (Challenge Problem) If this is true and you sample 20 patients, the middle 95% of all sample means would be expected to fall between what two numbers?

ANSWER: (2.34, 3.66)

From part (a), $\overline{X} \sim N(3, 0.1125)$
Start with Z − want to know the value of a where $P(-a < Z < a) = 0.95$
So $P(Z > a) = 0.025$
From normal table, $P(Z > 1.96) = 0.025 \rightarrow a = 1.96$
Back-transform to $\overline{X}$:

If $Z = \dfrac{\overline{X} - \mu}{\sqrt{V(\overline{X})}}$ then $\overline{X} = \sqrt{V(\overline{X})} \times Z + \mu$

Plug in $+a$ and $-a$ to find the bounds on $\overline{X}$
$\sqrt{0.1125} \times a + 3 = 0.3354 \times 1.96 + 3 = 3.66$
$\sqrt{0.1125} \times (-a) + 3 = 0.3354 \times -1.96 + 3 = 2.34$
$(2.34, 3.66)$