

# Chapter 3 (p129)

**3.39 Grade distributions.** Each row in the table below is a proposed grade distribution for a class. Identify each as a valid or invalid probability distribution, and explain your reasoning.

	<i>Grades</i>				
	A	B	C	D	F
(a)	0.3	0.3	0.3	0.2	0.1
(b)	0	0	1	0	0
(c)	0.3	0.3	0.3	0	0
(d)	0.3	0.5	0.2	0.1	-0.1
(e)	0.2	0.4	0.2	0.1	0.1
(f)	0	-0.1	1.1	0	0

**3.39 Grade distributions.** Each row in the table below is a proposed grade distribution for a class. Identify each as a valid or invalid probability distribution, and explain your reasoning.

	<i>Grades</i>				
	A	B	C	D	F
(a)	0.3	0.3	0.3	0.2	0.1
(b)	0	0	1	0	0
(c)	0.3	0.3	0.3	0	0
(d)	0.3	0.5	0.2	0.1	-0.1
(e)	0.2	0.4	0.2	0.1	0.1
(f)	0	-0.1	1.1	0	0

Label	Valid	Reason
a	N	Sum $\neq 1$
b	Y	Sum=1
c	N	Sum $\neq 1$
d	N	Sum $\neq 1$ , Negative p
e	Y	Sum=1
f	N	Negative p

**3.40 Health coverage, frequencies.** The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table summarizes two variables for the respondents: health status and health coverage, which describes whether each respondent had health insurance.<sup>64</sup>

		<i>Health Status</i>					Total
		Excellent	Very good	Good	Fair	Poor	
<i>Health Coverage</i>	No	459	727	854	385	99	2,524
	Yes	4,198	6,245	4,821	1,634	578	17,476
	Total	4,657	6,972	5,675	2,019	677	20,000

- If we draw one individual at random, what is the probability that the respondent has excellent health and doesn't have health coverage?
- If we draw one individual at random, what is the probability that the respondent has excellent health or doesn't have health coverage?

<sup>64</sup>Office of Surveillance, Epidemiology, and Laboratory Services Behavioral Risk Factor Surveillance System, BRFSS 2010 Survey Data.

<sup>65</sup>Source: CIA Factbook, Country Comparison: HIV/AIDS - Adult Prevalence Rate.

**3.40 Health coverage, frequencies.** The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table summarizes two variables for the respondents: health status and health coverage, which describes whether each respondent had health insurance.<sup>64</sup>

		<i>Health Status</i>					Total
		Excellent	Very good	Good	Fair	Poor	
<i>Health Coverage</i>	No	459	727	854	385	99	2,524
	Yes	4,198	6,245	4,821	1,634	578	17,476
	Total	4,657	6,972	5,675	2,019	677	20,000

- (a) If we draw one individual at random, what is the probability that the respondent has excellent health and doesn't have health coverage?
- (b) If we draw one individual at random, what is the probability that the respondent has excellent health or doesn't have health coverage?

<sup>64</sup>Office of Surveillance, Epidemiology, and Laboratory Services Behavioral Risk Factor Surveillance System, BRFSS 2010 Survey Data.

<sup>65</sup>Source: CIA Factbook, Country Comparison: HIV/AIDS - Adult Prevalence Rate.

a)

X=Health Coverage

Y=Health Status

$$P(X=\text{No}, Y=\text{Excellent}) = 459/20000 =$$

$$\text{b) } P(X=\text{No OR } Y=\text{Excellent}) = P(X=\text{No}) + P(Y=\text{Excellent}) - P(\text{No And Excellent}) = 2524/20000 + 4657/20000 = .126 + .232 = .36$$

**3.41 HIV in Swaziland.** Swaziland has the highest HIV prevalence in the world: 25.9% of this country's population is infected with HIV.<sup>65</sup> The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?

---

<sup>64</sup>Office of Surveillance, Epidemiology, and Laboratory Services Behavioral Risk Factor Surveillance System, BRFSS 2010 Survey Data.

<sup>65</sup>Source: CIA Factbook, Country Comparison: HIV/AIDS - Adult Prevalence Rate.

explanation of inverting Sens and Spec to PPV and NPV based on prev.

<https://journals.sagepub.com/doi/pdf/10.1177/201010581102000411>

demo of prevalence adjusted ppv:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636062/>

# Confusion Matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

**3.41 HIV in Swaziland.** Swaziland has the highest HIV prevalence in the world: 25.9% of this country's population is infected with HIV.<sup>65</sup> The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?

D=Disease (HIV)

T=Test (ELISA)

$P(D=1) = .259$

$P(T=1 | D=1) = .997$  (Sensitivity)

$P(T=0 | D=0) = .926$  (Specificity)

Question:  $P(D=1 | T=1) = ?$  = (Positive Predictive Value)

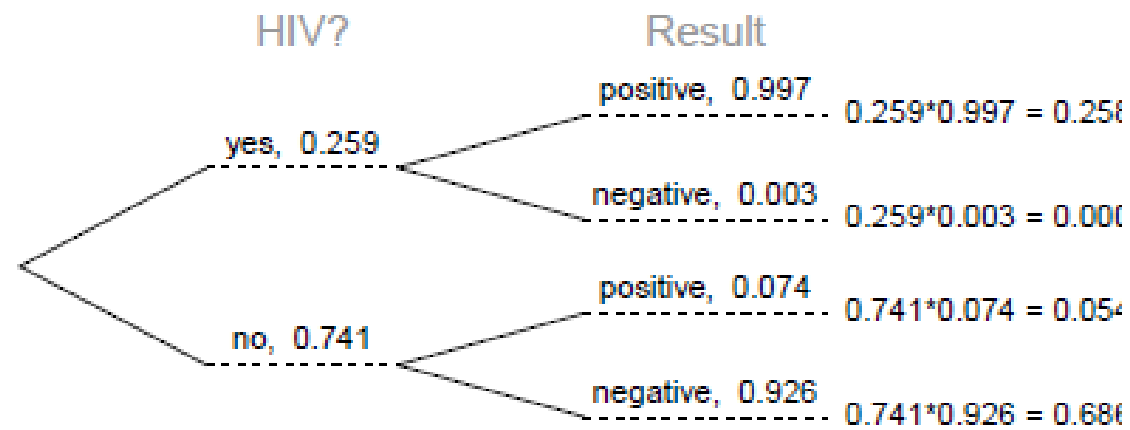
$P(D=1 | T=1) = P(D=1, T=1) / P(T=1) = P(T=1 | D=1) P(D=1) / P(T=1) = .997 * .259 / P(T=1)$

$P(T=1) = P(T=1 | D=1) P(D=1) + P(T=1 | D=0) P(D=0) = .997 * .259 + (1 - .926) * (1 - .259) = .258 + .054 = .313$

$P(D=1 | T=1) = .997 * .259 / .313 = .825$



3.41 0.8247.



At the end of 2019, an estimated 1,189,700 million people aged 13 and older had HIV in the United States.

$$P(D=1) = 1189700/350000000 = .003$$

$$P(D=1 | T=1) = ?$$

$$P(D=1 | T=1) = P(T=1 | D=1)P(D=1)/P(T=1)$$

$$P(T=1) = P(T=1 | D=1)P(D=1) + P(T=1 | D=0)P(D=0)$$

$$P(D=1 | T=1) = P(T=1 | D=1)P(D=1)/P(T=1)$$

$$P(T=1) = P(T=1 | D=1)P(D=1) + P(T=1 | D=0)P(D=0)$$

**3.42 Twins.** About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex – half are males and the other half are females. One-quarter of fraternal twins are both male, one-quarter both female, and one-half are mixes: one male, one female. You have just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

$$P(T=i)=.3$$

$$P(T=f)=.7$$

$$P(S=mm | T=f)=.25$$

$$P(S=ff | T=f)=.25$$

$$P(S=mf | T=f)=.5$$

$$P(T=i | S=ff) = P(S=ff | T=i)P(T=i)/P(S=ff) = .5 * .3 / .5 = .3$$

**3.43 Cost of breakfast.** Sally gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of 30¢ (\$0.30), the average price of a muffin is \$2.50 with a standard deviation of 15¢, and the two prices are independent of each other.

- (a) What is the mean and standard deviation of the amount she spends on breakfast daily?
- (b) What is the mean and standard deviation of the amount she spends on breakfast weekly (7 days)?

**3.43 Cost of breakfast.** Sally gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of 30¢ (\$0.30), the average price of a muffin is \$2.50 with a standard deviation of 15¢, and the two prices are independent of each other.

- (a) What is the mean and standard deviation of the amount she spends on breakfast daily?
- (b) What is the mean and standard deviation of the amount she spends on breakfast weekly (7 days)?

**3.43 Cost of breakfast.** Sally gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of 30¢ (\$0.30), the average price of a muffin is \$2.50 with a standard deviation of 15¢, and the two prices are independent of each other.

- (a) What is the mean and standard deviation of the amount she spends on breakfast daily?
- (b) What is the mean and standard deviation of the amount she spends on breakfast weekly (7 days)?

**3.43** (a)  $E = \$3.90$ .  $SD = \$0.34$ .

(b)  $E = \$27.30$ .  $SD = \$0.89$ .

**3.45 Variance of a mean, Part I.** Suppose we have independent observations  $X_1$  and  $X_2$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of the two values:  $\frac{X_1+X_2}{2}$ ?



**3.45 Variance of a mean, Part I.** Suppose we have independent observations  $X_1$  and  $X_2$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of the two values:  $\frac{X_1+X_2}{2}$ ?

$$\begin{aligned} \mathbf{3.45} \quad & \text{Var} \left( \frac{X_1+X_2}{2} \right) \\ &= \text{Var} \left( \frac{X_1}{2} + \frac{X_2}{2} \right) \\ &= \frac{\text{Var}(X_1)}{2^2} + \frac{\text{Var}(X_2)}{2^2} \\ &= \frac{\sigma^2}{4} + \frac{\sigma^2}{4} \\ &= \sigma^2/2 \end{aligned}$$

**3.47 Variance of a mean, Part III.** Suppose we have  $n$  independent observations  $X_1, X_2, \dots, X_n$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of these  $n$  values:  $\frac{X_1 + X_2 + \dots + X_n}{n}$ ?

**3.47 Variance of a mean, Part III.** Suppose we have  $n$  independent observations  $X_1, X_2, \dots, X_n$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of these  $n$  values:  
 $\frac{X_1 + X_2 + \dots + X_n}{n}$ ?

$$\begin{aligned}
 & \mathbf{3.47} \quad \text{Var} \left( \frac{X_1 + X_2 + \dots + X_n}{n} \right) \\
 &= \text{Var} \left( \frac{X_1}{n} + \frac{X_2}{n} + \dots + \frac{X_n}{n} \right) \\
 &= \frac{\text{Var}(X_1)}{n^2} + \frac{\text{Var}(X_2)}{n^2} + \dots + \frac{\text{Var}(X_n)}{n^2} \\
 &= \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \dots + \frac{\sigma^2}{n^2} \quad (\text{there are } n \text{ of these terms}) \\
 &= n \frac{\sigma^2}{n^2} \\
 &= \sigma^2 / n
 \end{aligned}$$

**3.44 Scooping ice cream.** Ice cream usually comes in 1.5 quart boxes (48 fluid ounces), and ice cream scoops hold about 2 ounces. However, there is some variability in the amount of ice cream in a box as well as the amount of ice cream scooped out. We represent the amount of ice cream in the box as  $X$  and the amount scooped out as  $Y$ . Suppose these random variables have the following means, standard deviations, and variances:

	mean	SD	variance
$X$	48	1	1
$Y$	2	0.25	0.0625

- An entire box of ice cream, plus 3 scoops from a second box is served at a party. How much ice cream do you expect to have been served at this party? What is the standard deviation of the amount of ice cream served?
- How much ice cream would you expect to be left in the box after scooping out one scoop of ice cream? That is, find the expected value of  $X - Y$ . What is the standard deviation of the amount left in the box?
- Using the context of this exercise, explain why we add variances when we subtract one random variable from another.

**3.45 Variance of a mean, Part I.** Suppose we have independent observations  $X_1$  and  $X_2$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of the two values:  $\frac{X_1+X_2}{2}$ ?

**3.46 Variance of a mean, Part II.** Suppose we have 3 independent observations  $X_1, X_2, X_3$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of these 3 values:  $\frac{X_1+X_2+X_3}{3}$ ?

**3.47 Variance of a mean, Part III.** Suppose we have  $n$  independent observations  $X_1, X_2, \dots, X_n$  from a distribution with mean  $\mu$  and standard deviation  $\sigma$ . What is the variance of the mean of these  $n$  values:  $\frac{X_1+X_2+\dots+X_n}{n}$ ?