

CHAPTER 9 Appendix

Essential Probability Rules

This chapter introduces you to the basics of probability, its underpinnings in “chance” events, and the four basic rules:

1. $P(S) = 1$, where S is the sample space (the collection of all possible outcomes)
2. $0 \leq P(A) \leq 1$, where A is any event
3. $P(A \text{ or } B) = P(A) + P(B)$, when events A and B have no outcomes in common (are *disjoint*)
4. $P(\text{not } A) = 1 - P(A)$

You are also introduced to the idea of a density curve—that is, a probability model for a continuous random variable. In the case of a uniform random variable, finding the probability that random variable X is between a and b is found by multiplying the length of the interval from a to b that is under the non-zero portion of the density by the height of the density, because the density forms a rectangle, and the area of a rectangle is $A = lw$.

Most calculations here require only a simple calculator (and can even be done mentally). Software can help you find the mean, variance, and standard deviation for discrete random variables. Those procedures are described here. In all cases, we assume the variable values are entered in one column of a data sheet and the probability of each value is entered in a separate column.



Excel

1. Place the cursor in a blank column adjacent to one of your data columns.
2. Enter a formula to multiply the probability and the value. For example, in cell C1 enter

=A1*B1

3. Drag the cursor to copy the formula down the rows.
4. Place the cursor in an empty cell and enter a formula to sum the entries just created to find the mean. For example:

=sum (C1 . . Cn)

5. Calculate the variance as $\sum_{i=1}^n (x_i - \mu)^2 p(x_i)$ by once again placing the cursor in an empty cell (for example, D1) and entering the formula (assuming the mean is in cell F1):

= (A1 - F\$1) ^ 2 * B1

Drag this formula down the column to replicate it into each row of the probability distribution. Sum the entries just created as was done in step 4.

6. Click in an empty cell and calculate the standard deviation as the square root of the variance. Here, we assume the variance has been stored in cell F2:

=sqrt (F2)



1. **Analyze → Distribution**
2. Click to enter the column with the variable values into **Y, Columns.**
3. Click to enter the column with the probabilities into **Weight.**
4. **OK**

TA9-1

In the results, the mean given is μ . To calculate the variance and standard deviation, we must recognize that JMP uses a formula that is different from ours,

$$\sigma^2 = \frac{\sum (x - \mu)^2 p(x)}{n - 1}.$$

We need to adjust for the $(n - 1)$ factor.

5. Click the red triangle next to **Summary Statistics** → **Customize Summary Statistics**. Check the box next to **Variance**.
6. **OK**
7. Multiply the variance shown by $(n - 1)$. This is the correct variance.
8. Take the square root of the variance for the standard deviation.

Note: Alternatively, you can use commands like those in the Excel steps shown previously to create the variance after first obtaining the mean.

1. **Cols** → **New Col**
2. Give the new column a name.
3. **Column Properties** → **Formula** → **Edit Formula**
4. Enter the formula using the operator icons x^y and $\times$ (times). In the following example, substitute the numeric value of the mean for “mean”:

(Variable-mean)²×Prob
5. Click **OK** twice.
6. **Analyze** → **Distribution**
7. Click the Red Triangle next to the variable name and select **Sum**. This sum is the variance—ignore all other entries in the column.



Minitab

We assume the variable values are in C1 and the probabilities in C2. This is easiest done with commands, so if you do not have the MTB> prompt, click in the Session window and then click **Editor** → **Enable Commands**.

```
MTB> Let C3=C1*C2
MTB> Sum C3                                     (this displays the mean)
MTB> Let C4=(c1-mean)**2*c2                   (use the value above for
                                                "mean")
MTB> Sum C4                                     (this displays the
                                                variance)
MTB> Let k1=sqrt(variance)                     (use the value above for
                                                "variance")
MTB> print k1
```

SPSS

1. Enter the values of the variable in one column and the probabilities in a second column—say, **X** and **P**.
2. **Transform** → **Compute Variable**
3. Enter a new variable name (say, **XP**) and the formula **X*P**.
4. **OK**
5. **Analyze** → **Descriptive Statistics** → **Descriptives**
6. Click to enter the variable just created (**XP** in this example) as the variable.
7. **Options**
8. Check the box next to **Sum**.
9. **Continue** → **OK** (The entry for the sum is the mean of the random variable.)
10. **Transform** → **Compute Variable**
11. **Reset** (to blank out the formula area).

12. Enter a new variable name (say, **XMUSQ**) and the formula

(X-mean) **2 *P

where “mean” is the value just found.

13. **OK**

14. **Analyze** → **Descriptive Statistics** → **Descriptives**

15. Click to enter the variable just created (**XMUSQ** in this example). The sum displayed is the variance.

CRUNCH IT!

CrunchIt! cannot perform these calculations. At this time, you cannot use this software to obtain the sum of a column (variable) or find weighted statistics.



TI-83/-84

1. Enter the variable values in one list (say, **L1**) and the probabilities in another (say, **L2**).

2. **STAT** **►** **1** (for 1-VarStats)

3. Enter the two list names separated by a comma. The command should look like this:

1-VarStats L1,L2

Note: Newer TI-84 calculators will have an intermediate screen on 1-VarStats that asks for List (the variable) and FreqList (the probabilities).

4. In the displayed results, $\bar{x}$ is μ (the calculator has only one symbol for a mean), and σ_x is the population standard deviation.

For the variance (without rounding error), press **VAR** **5** (for Statistics) **4** (for σ_x) **x^2** .



Calculate the mean as shown in the example commands:

```
> x<-c(1,2,3,4,5)
> p<-c(.1,.2,.2,.1,.4)
> xp=x*p
> sum(xp)
[1] 3.5
```

Calculate the variance as shown in the example commands:

```
> xmmusq=(x-3.5)**2*p
> Var=sum(xmmusq)
```

Calculate the standard deviation (if asked):

```
> sqrt(Var)
```