MAS 170 Elementary Statistics Spring 2022
Essential Clarifications to Freedman's Text
updated 2 February 2023

Rounding in calculations

In general, do not round off any numbers in a multi-step calculation until the final step. If you must round, keep lots of digits. A rule of thumb is to round to the nearest 0.001, that is, to the nearest 3rd digit to the right of the decimal point.

Exceptions: When calculating a z value, you may round to the nearest 0.01, because that's the level of accuracy in the z column on the normal table. When calculating an area under the normal curve, you may usually round the final answer to the nearest whole percent.

"percentile" and "percentile rank"
(pp.90--91)

The percentile rank of a data value A is the percent of all data whose value is less than or equal to A. If the percent of data less than or equal to A is p, then A has p-th percentile. In picture form, the percentile rank of a data value A is the area under the histogram to the left of A. The text introduces percentile and percentile rank in the context of normally distributed data, but the terms percentile and percentile rank apply to all types of data, whether it is normally distributed or not.

"football-shaped" or "shaped like a football"
(p.120 and throughout Ch 8--12)

The text uses the term "football-shaped" very loosely to describe a scatter diagram that has linear association. Worse, the term "football-shaped" is sometimes used to imply homoscedasticity. Even worse, "football-shaped" seems to sometimes imply that both variables involved in the scatter diagram are normally distributed (see the box on p.197). Because of this vagueness and ambiguity, we will instead use the terms "shows linear association", "homoscedastic", and "normally distributed" for these various attributes that may or may not apply to 2-variable data. In particular, in the box on p.197, and also in item 7 of the Ch 11 summary on p.201, replace the first sentence with "Suppose that a scatter diagram shows linear association, is homoscedastic, and that both variables are distributed normally."

outcome, event, and "thing"

In a box model, an ***outcome*** is a sequence of tickets obtained by random draws, either with or without replacement. For example, here are the 12 possible outcomes for 2 draws, without replacement, from the box {A,K,Q,J}. (This is a box model for "two cards are dealt from a 4-card deck containing an ace, a king, a queen, and a jack".)

set of all outcomes = {AK,AQ,AJ,KA,KQ,KJ,QA,QK,QJ,JA,JK,JQ}

In this notation, "AK" means the ace is drawn first, and the king second, while "KA" means the king is first and the ace is second.

An ***event*** is a set of outcomes. For example, the events "get a Q" and "get K on the first draw" in the example above are

"get a Q" = {QA,QK,QJ,AQ,KQ,JQ}
"get K on draw 1" = {KA,KQ,KJ}.

Another example is the event "get the hand A-K". In a "hand", it does not matter what order the cards are dealt in, so we have

"get the hand A-K" = {AK,KA}.

To say that two events E and F "both happen" can be rephrased "E AND F". To say that "at least one of the events E,F happens" can be rephrased "E OR F". For example, for the events

E = "get an ace" = {AK,AQ,AJ,KA,QA,JA}
F = "get a king" = {KA,KQ,KJ,AK,QK,JK}

we have

E AND F = "get an ace and get a king" = {AK,KA}.
E OR F = "get an ace or get a king" = {AK,AQ,AJ,KA,QA,JA,KQ,KJ,QK,JK}.

[Comment for students who know about intersection and union of sets: the event "E AND F" is the same thing as the intersection of the sets E,F, and the event "E OR F" is the same thing as the union of the sets E,F.]

We will use the correct terms "outcome" and "event" instead of the vague term "thing" used in the text in Ch 13.

some probability symbols and the terms "conditional probability" and "independent" and "dependent" events
===

[Note: The symbols given in this section appear in Friedman's text in a note on p.227 at the end of Exercise Set B.]

We will write P(E) to denote the probability for an event E. For example, in the game of 2 draws, with replacement, from the box {A,K,Q,J}, the probability of the single outcome AK is

P(AK) = 1/12 = approx 8.3%

because there are 12 equally likely outcomes for the 2 draws. The probability for getting the hand A-K is

P(get the hand A-K) = 2/12 = approx 16.7%

because there are 2 equally likely outcomes in the event "get the hand A-K", and there are 12 equally likely outcomes in all for the 2 draws. The probability of getting a Q is

P(get a Q) = 6/12 = 50%

because there are 6 equally likely outcomes in the event "get a Q". The probability of getting K on the first draw is

P(get K on draw 1) = 3/12 = 25%

because there are 3 equally likely outcomes in the event "get K on draw 1".

The textbook does not give a definition of the term *conditional probability*; instead, they explain by example in Ch 13 Sec. 2 (p.226). Here is the definition: The conditional probability of event E given that event F has happened, denoted P(E|F), is

P(E|F) = P(E AND F) / P(F)

where the event "E AND F" is the set of all outcomes that belong to both events E and F. For example, in 2 draws with replacement from the box {A,K,Q,J}, we have

P(get a Q | get K on draw 1) = 1/3 = approx 33.3%

because there is 1 outcome, namely KQ, that has a Q, among the 3 outcomes that have K on draw 1. We also have

P(get K on draw 1 | get a Q) = 1/6 = approx 16.7%

because there is 1 outcome, namely KQ, for which draw 1 is a K, among the 6 outcomes that have a Q.

The definitions of independent and dependent events (box on p.230) should be replaced by the following: "Two events E,F are *independent* if P(F|E) = P(F). Otherwise, the events E,F are *dependent*." Or in words, "Two events E,F are *independent* if the probability of F given that E has happened is equal to the probability of F. Otherwise, the events E,F are *dependent*."