AP Statistics – Lesson Notes - Chapter 16: Random Variables

Expected Value = weighted mean or 'center'

In a small raffle at a high school football game, there are 100 tickets. There is 1 winning ticket which will pay $50 if selected. There are 3 runner-up tickets which will pay $5 each if selected. The other 96 tickets pay nothing if selected.

If you play this raffle each week for many years, what would you 'expect' your winnings to be?

$P(\text{winning ticket}) = \frac{1}{100}$ $\quad$ *In the long run, you win* $= \$50\left(\frac{1}{100}\right) + \$5\left(\frac{3}{100}\right) + \$0\left(\frac{96}{100}\right)$

$P(\text{runner up ticket}) = \frac{3}{100}$ $\quad$ *In the long run, you win* $= \$0.65$

$P(\text{losing ticket}) = \frac{96}{100}$

The 'expected value' of your winnings, over time, would be $0.65. per week, on average

What if the raffle were arranged as follows: There were 99 tickets and 33 of them were 'winning' (paying $50), 33 were 'runner-up' (paying $5), and 33 were 'losing' (paying $0)?

In the long run, you win $= \$50\left(\frac{1}{3}\right) + \$5\left(\frac{1}{3}\right) + \$0\left(\frac{1}{3}\right)$

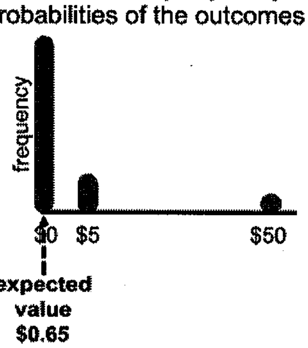
In the long run, you win $= \$18.33$

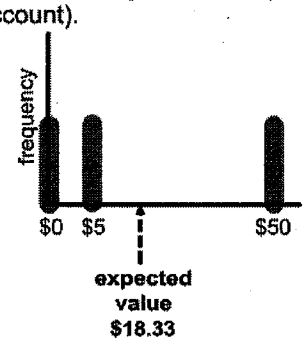
This would just be the mean value of the possible winnings (50, 5, 0):

$$\$50\left(\frac{1}{3}\right) + \$5\left(\frac{1}{3}\right) + \$0\left(\frac{1}{3}\right)$$
$$\frac{50}{3} + \frac{5}{3} + \frac{0}{3}$$
$$\frac{50+5+0}{3}$$
$$18.33$$

If you think of all the individual outcomes of playing the raffle as data values, they would form a distribution. The expected value is the mean (or center) of this distribution.

If the outcomes are all equally likely this is simply the mean of the values. If the outcomes are not equally likely, expected value is the weighted mean (the mean with the probabilities of the outcomes taken into account).





Expected value is the weighted mean (the mean with the probabilities of the outcomes taken into account)...the sum of each outcome multiplied by the probability of that outcome occurring.

$$\mu = E(X) = \sum x_i \cdot P(x_i)$$

X is called a **random variable**. It is a variable that represents the possible values of the outcome of a random process. In algebra, a variable is a placeholder for a particular value. In statistics, a random variable is a placeholder for a value which is expected to vary randomly.

Expected Value is denoted by μ because it represents the 'center' or mean of a distribution of possible outcome values of the random variable.

Assigning probabilities to each possible value of a random variable produces a **probability model** for the random variable.

Probability Model - a listing of all outcomes along with their probabilities

probability distribution

(# heads) X	0	1	2	3
P	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

cumulative probability distribution

(# heads) X	0	1	2	3
P	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
cumulative P	$\frac{1}{64}$	$\frac{10}{64}$	$\frac{37}{64}$	$\frac{64}{64}$

(like an ogive)

Expected Value distributions also have spread

We can find the variance and standard deviation of the spread of the values of a random variable:

$$\sigma^2 = Var(X) = \sum (x_i - \mu)^2 \cdot P(x_i)$$

$$\sigma = SD(X) = \sqrt{\sum (x_i - \mu)^2 \cdot P(x_i)}$$

Random variables can be discrete or continuous

The random variable in the practice problem represented the number of boys. This can only take on specific (discrete) values (0, 1, 2, 3, not 1.4 boys) so this random variable was a **discrete random variable.**

A random variable might be modeling quantities which can take any value (for example, heights of men) and that would be called a **continuous random variable.**

Note that even in the case of discrete random variables, the values of parameters such as μ and σ do not have to be obtainable discrete values.

Transforming *a single distribution*: Effect on means and variances

As with means and variances of data, if we add or subtract a constant, only measures of center (the mean) is shifted:

$$\mu_{x+c} = \mu_x + c$$

If we multiply by a constant, measures of center are both multiplied by the constant, but the variance is multiplied by the square of the constant:

$$\mu_{cx} = c\mu_x$$

$$\sigma_{cx}{}^2 = c^2 \sigma_x{}^2$$

(Intuitively: the reason the variance is multiplied by the square of the constant is that variance is not in the original units...it is units squared, because variance is computed by squaring the distances. So if every individual data point is multiplied by *c* then the variance is multiplied by the square of *c*.)

Combining means and variances of ***different distributions***

If you have two random variables, X and Y, if you make a new variable, Z, by adding or subtracting these variables, the new random variable also has an expected value and variance.

#4b) Speed Dating (use rules for combining means and variances): To save time and money, many single people have decided to try speed dating. At a speed dating event, women sit in a circle and men spend about five minutes getting to know a woman before moving on to the next one. Suppose that the height M of male speed daters follows a Normal distribution with a mean of 68 inches and a standard deviation of 4 inches and the height F of female speed daters follows a Normal distribution with a mean of 64.5 inches and a standard deviation of 3 inches. What is the probability that a randomly selected female speed dater is taller than the randomly selected male speed dater she is paired with?

If the typical male height is 68 inches and the typical female height is 64.5 inches, it makes sense that the typical height difference would be 68 - 64.5 = 3.5 inches. This is always true: the combined mean is the sum (or difference) of the individual means:

$$\mu_{x\pm y} = \mu_x \pm \mu_y$$

But it <u>isn't true that the standard deviations just add or subtract</u>. Each value (male and female height) varies randomly around their mean values. In some couples, you may have a taller man and much shorter woman, some a taller women with a shorter man. So how can we combine these?

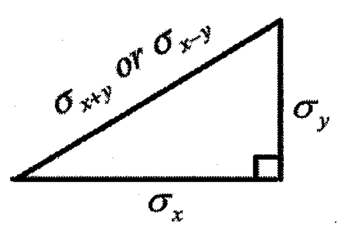
There is no way to intuitively explain the correct result. It can be proven that the standard deviations do not add, but instead that the ***variances always add***:

$$\sigma^2_{x\pm y} = \sigma_x^{\,2} + \sigma_y^{\,2}$$

(See 'more about why' on mrfelling.com site for a proof of this result)

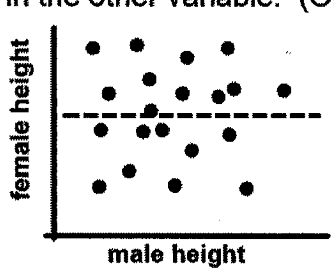
There <u>**is**</u> an intuitive way, however, to remember this result. It is sometimes called the 'Pythagorean Theorem of Statistics':

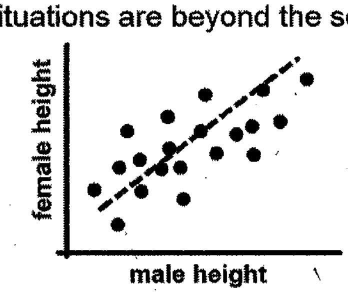
$$\sigma^2_{x\pm y} = \sigma_x^{\,2} + \sigma_y^{\,2}$$



- **Very important**: Variances (squares of standard deviation) can only be combined in this way if the two variables are **independent**. Variation in one variable does not cause variation in the other variable. (Other situations are beyond the scope of this class)

can use these methods

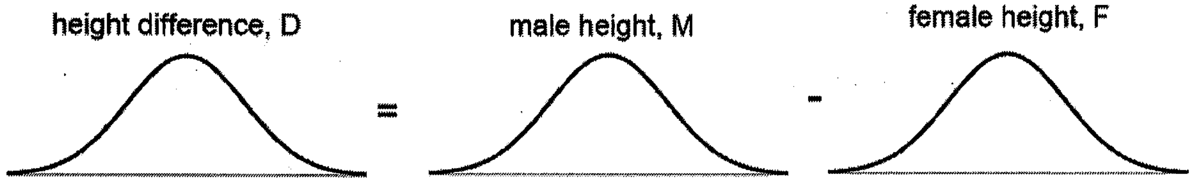




cannot use these methods

- It doesn't matter if you add or subtract, variation is still combining, so the individual variances are always added.

#4b) **Speed Dating (use rules for combining means and variances):** To save time and money, many single people have decided to try speed dating. At a speed dating event, women sit in a circle and men spend about five minutes getting to know a woman before moving on to the next one. Suppose that the height *M* of male speed daters follows a Normal distribution with a mean of 68 inches and a standard deviation of 4 inches and the height *F* of female speed daters follows a Normal distribution with a mean of 64.5 inches and a standard deviation of 3 inches. What is the probability that a randomly selected female speed dater is taller than the randomly selected male speed dater she is paired with?



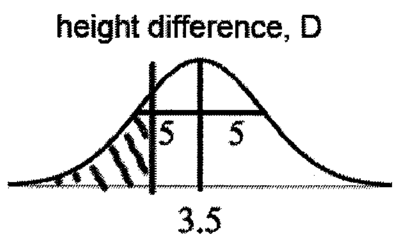
the algebra: $D = M - F$

means match the algebra: $\mu_D = \mu_M - \mu_F$ $\qquad \mu_D = 68 - 64.5 = 3.5\text{ in}$

variances always add: $\sigma_D^{\,2} = \sigma_M^{\,2} + \sigma_F^{\,2}$ $\qquad \sigma_D^2 = 4^2 + 3^2$

$\sigma_D = \sqrt{4^2 + 3^2} = 5\text{ in}$

Now we can use normalcdf to find the probability:



$$P(D < 0) = normalcdf(-999, 0, 3.5, 5) = 0.242$$

Summary: Transforming or combining random variables

Transforming <u>a single distribution</u>:

1) mean (or any individual data value) are affected by multiply and add/sub
2) SD, variance (or any measure of spread) is only affected by multiply

Combining <u>multiple distributions</u>:

1) Write out the algebra
2) Means match the algebra
3) Variances always add