EXPECTATION FORMULAS

Maurice J. Dupré
Department of Mathematics
New Orleans, LA 70118
email: mdupre@tulane.edu
November 2010

In the formulas given below, U, W, X, Y are any unknowns, and A, B, C, D, K, are any statements, unless otherwise specified. $E(X|A)$ is the **Expected Value** of X given A. Always I_A denotes the **Indicator** unknown for the statement A, and this means I_A is one or zero according to whether A is true or false. Definite real numbers are denoted by lower case symbols $a, b, c, d, u, v, w, x, y, z$ and are also referred to as constants. Unknowns can be added and multiplied and all constants are also considered as unknowns. Think of any description of a numerical quantity as being an unknown, so since 3 or π describe numerical quantities they are unknowns. The unknowns form a set $\mathcal{A}$ containing the set of all real numbers $\mathbb{R}$. We call $\mathcal{A}$ the **Algebra of Unknowns.**

1. FUNDAMENTAL AXIOMS

INFORMATION CANNOT BE IGNORED

$$E(X|(X = a)\&B) = a$$

CONSEQUENCE (since a=a)

$$E(a|B) = a$$

ORDER AXIOM

$$[B \Rightarrow (X \leq Y)] \Rightarrow E(X|B) \leq E(Y|B)$$

2. ADDITION AXIOM

$$E(a \cdot X \pm b \cdot Y|B) = a \cdot E(X|B) \pm b \cdot E(Y|B)$$

3. MULTIPLICATION RULE (follows from the assumption that $E(X \cdot I_A|B)$ as a function of X depends only on the number $E(X|A \ \& \ B)$ for any fixed A, B, by then taking $X = a$ and evaluating both $E(X \cdot I_B|C)$ and $E(X|B\&C)$)

$$E(X \cdot I_B|C) = E(X|B \ \& \ C) \cdot E(I_B|C)$$

4. DEFINITION OF PROBABILITY

$$P(A|B) = E(I_A|B)$$

5. NOTATION (when the given A is understood or is the same throughout)

$$\mu_X = E(X) = E(X|A), \ E(X|B) = E(X|B\&A), \ P(B|A) = P(B)$$

MULTIPLICATION RULE AND PROBABILITY

$$E(X \cdot I_B) = E(X|B) \cdot P(B)$$

6. LOGIC CONVERTED TO ALGEBRA

$$I_{A\&B} = I_A \cdot I_B$$

$$I_{A\ or\ B} = I_A + I_B - I_{A\&B}$$

$$I_{not\ A} = 1 - I_A$$

7. MULTIPLICATION RULE FOR PROBABILITY (use $X = I_A$ in Multiplication Rule and definition of probability)

$$P(A\ \&\ B) = P(A|B) \cdot P(B)$$

8. RULES OF PROBABILITY (follow from $0 \leq I_A \leq 1$, the conversion of logic to algebra, the axioms, and the definition of probability)

$$0 \leq P(A) \leq 1$$

$$P(A\ or\ B) = P(A) + P(B) - P(A\&B)$$

$$P(not\ A) = 1 - P(A)$$

9. PARTITION PRINCIPLE (consequence of axioms and Multiplication Rule)
If $B_1, B_2, B_3, ..., B_n$ are statements and exactly one is true (called a **Partition**), then

$$\sum_{k=1}^{n} I_{B_k} = 1, \ \sum_{k=1}^{n} P(B_k) = 1$$

$$X = \sum_{k=1}^{n} X \cdot I_{B_k}$$

$$E(X) = \sum_{k=1}^{n} E(X|B_k) \cdot P(B_k)$$

$$P(A) = \sum_{k=1}^{n} P(A|B_k) \cdot P(B_k)$$

10. BAYES' RULES

$$P(A|B) \cdot P(B) = P(A\&B) = P(B\&A) = P(B|A) \cdot P(A)$$

$$P(A|B) = \frac{P(A\&B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}$$

11. DEFINITION OF COVARIANCE

$$Cov(X,Y) = E([X-\mu_X]\cdot[Y-\mu_Y])$$

12. DEFINITION OF VARIANCE

$$Var(X) = Cov(X,X)$$

13. DEFINITION OF STANDARD DEVIATION

$$SD(X) = \sigma_X = \sqrt{Var(X)}$$

14. DEFINITION OF CORRELLATION

$$\rho = \rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

15. DEFINITION OF STANDARDIZATION

$$Z_X = \frac{X-\mu_X}{\sigma_X},\ X = \mu_X + \sigma_X \cdot Z_X$$

16. CALCULATION FORMULAS

$$Cov(X,Y) = E(X\cdot Y) - \mu_X\cdot\mu_Y = \rho\cdot\sigma_X\cdot\sigma_Y$$

$$E(X\cdot Y) = \mu_X\cdot\mu_Y + Cov(X,Y)$$

$$Var(X\pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$$

$$\rho = Cov(Z_X, Z_Y) = E(Z_X\cdot Z_Y)$$

$$Cov(X,c) = 0,\ Cov(X,Y) = Cov(Y,X),\ Var(X\pm c) = Var(X)$$

$$Cov(W, a\cdot X \pm b\cdot Y) = a\cdot Cov(W,X) \pm b\cdot Cov(W,Y),\ Var(c\cdot X) = c^2\cdot Var(X)$$

17. LINEAR REGRESSION (of Y on X, with $W = \beta_0 + \beta_1\cdot X$ and $R_Y = W_X - Y$)

$$E(Y|X=x) = \beta_0 + \beta_1\cdot x$$

$$\beta_1 = \rho\cdot\frac{\sigma_Y}{\sigma_X},\ \beta_0 = \mu_Y - \beta_1\cdot\mu_X$$

$$E(R_Y^2) = \sigma_Y^2\cdot(1-\rho^2)$$

For $W = a + bX$, and $R = W - Y$,

$$E(R^2) = [E(R)]^2 + Var(R) = [(a+b\mu_X)-\mu_Y]^2 + Var(W) + Var(Y) - 2Cov(W,Y)$$

$$= [(a+b\mu_X)-\mu_Y]^2 + b^2\sigma_X^2 + \sigma_Y^2 - 2b\rho\sigma_X\sigma_Y$$

$$= [(a+b\mu_X)-\mu_Y]^2 + (1-\rho^2)\sigma_Y^2 + (\rho\sigma_Y)^2 + (b\sigma_X)^2 - 2(\rho\sigma_Y)(b\sigma_X)$$

$$= [(a+b\mu_X)-\mu_Y]^2 + (1-\rho^2)\sigma_Y^2 + [(\rho\sigma_Y)-(b\sigma_X)]^2$$

18. CUMULATIVE DISTRIBUTION FUNCTION

$$F_X(x) = P(X \leq x),\ x \in \mathbb{R}$$

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

19. PROBABILITY DISTRIBUTION FUNCTION (for discrete unknown)

$$p_X(k) = P(X = k),\ k \in \mathbb{R}$$

20. PROBABILITY DENSITY FUNCTION

$$f_X(x) = F'_X(x) = \left(\frac{d}{dx}F_X\right)(x),\ x \in \mathbb{R}$$

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$$

21. TRANSFORM

$$[\mathcal{M}(h)](t) = \int_{-\infty}^{\infty} e^{tx}h(x)dx = \int_{-\infty}^{\infty} \exp(tx) \cdot h(x)dx$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$\mathcal{M}(a \cdot g \pm b \cdot h) = a \cdot \mathcal{M}(g) \pm b \cdot \mathcal{M}(h)$$

$$[\mathcal{M}(x \cdot h)](t) = [\frac{d}{dt}\mathcal{M}(h)](t)$$

For $A(x)$ the statement that x is in the interval $[a, b] \subset \mathbb{R}$, we denote $I_{A(x)}$ by $I_{[a,b]}(x)$

$$\int_{-\infty}^{\infty} I_{[a,b]}(x) \cdot h(x)dx = \int_a^b h(x)dx$$

$$[\mathcal{M}(I_{[a,b]})](t) = \frac{e^{bt} - e^{at}}{t} = \sum_{n=0}^{\infty}[b^{n+1} - a^{n+1}]\frac{t^n}{(n+1)!} = \sum_{n=0}^{\infty} \frac{[b^{n+1} - a^{n+1}]}{n+1}\frac{t^n}{n!}$$

$$f(t) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(t-c)^n = \sum_{n=0}^{\infty} f^{(n)}(c)\frac{(t-c)^n}{n!},\ f^{(n)} = \frac{d^n}{dx^n}f$$

22. MOMENT GENERATING FUNCTION

$$m_X = E(e^{tX}) = \mathcal{M}(f_X)$$

$$m_X^{(n)}(t) = E(X^n e^{tX}),\ n = 0, 1, 2, 3, ...$$

$$m_X^{(n)}(0) = E(X^n),\ n = 0, 1, 2, 3, ...$$

$$m_{X \pm c}(t) = e^{\pm ct} \cdot m_X(t)$$

$$m_{cX}(t) = m_X(ct)$$

23. DIRAC DELTA FUNCTION

Not really a function but it is denoted δ with the property that for any smooth function h with compact support

$$h(0) = \int_{-\infty}^{\infty} h(x)\delta(x)dx$$

$$\delta_c(x) = \delta(x - c)$$

$$h(c) = \int_{-\infty}^{\infty} h(x)\delta_c(x)dx$$

$$\mathcal{M}(\delta_c)(t) = e^{ct}$$

If $A_1, A_2, A_3, ..., A_n$ is a partition, $v_1, v_2, v_3, ..., v_n \in \mathbb{R}$, then

$$X = \sum_{k=1}^{n} v_k \cdot I_{A_k}$$

is a **Simple Unknown** and

$$m_X(t) = \sum_{k=1}^{n} P(A_k) \cdot e^{v_k t} = \mathcal{M}(f_X),$$

$$f_X(x) = \sum_{k=1}^{n} P(A_k) \cdot \delta_{v_k}$$

24. UNIFORM DISTRIBUTION

If X is uniformly distributed on $[a, b]$ then

$$f_X = \frac{1}{b-a} \cdot I_{[a,b]}$$

$$m_X(t) = \frac{1}{b-a}\mathcal{M}(I_{[a,b]}) = \frac{e^{bt} - e^{at}}{(b-a)t} = \sum_{n=0}^{\infty} \left[\frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}\right] \frac{t^n}{n!}$$

$$E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}$$

$$E(X) = \frac{a+b}{2}$$

$$\sigma_X = \frac{(b-a)/2}{\sqrt{3}}$$

25. NORMAL DISTRIBUTION

$$\mu = E(X),\ \sigma = SD(X)$$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)$$

For Z standard normal ($\mu = 0,\ \sigma = 1$)

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$$

$$m_Z(t) = e^{t^2/2} = \sum_{n=0}^{\infty} \frac{(2n)!}{2^n \cdot n!} \frac{t^{2n}}{(2n)!},\ E(Z^n) = \frac{(2n)!}{2^n \cdot n!} = (2n-1)!!$$

26. SAMPLING DISTRIBUTIONS

Start with any random variable X and let $X_1, X_2, X_3, ..., X_n$ be observations of X.

$$E(X_k) = \mu_X,\ SD(X_k) = \sigma_X,\ k = 1, 2, 3, ..., n$$

$$\bar{X}_n = \frac{1}{n} T_n,\ T_n = \sum_{k=1}^{n} X_k$$

$$\overline{X - \bar{X}_n} = \sum_{k=1}^{n} (X_k - \bar{X}_n) = 0,\ \overline{(X - \bar{X}_n)^2} = \frac{1}{n} \sum_{k=1}^{n} (X_k - \bar{X})^2$$

$$E(T_n) = n\mu_X,\ E(\bar{X}_n) = \mu_X$$

$$\overline{(X - \bar{X})^2} = \overline{X^2} - [\bar{X}]^2$$

FAIR SAMPLING CONSTANT/CONDITION

$$c = \frac{Cov(X_k, T_n)}{\sigma_X^2},\ k = 1, 2, 3, ..., n$$

Then

$$Var(T_n) = c \cdot n\sigma_X^2,\ SD(T_n) = \sqrt{c} \cdot \sqrt{n} \cdot \sigma_X,\ SD(\bar{X}_n) = \sqrt{c} \frac{\sigma_X}{\sqrt{n}},\ Var(\bar{X}_n) = c \frac{\sigma_X^2}{n}.$$

$$S^2 = \frac{n}{n - c} \overline{(X - \bar{X})^2},\ E(S^2) = \sigma_X^2$$

Denote **Independent Random Sampling** by IRS and **Simple Random Sampling** by SRS. Here N is population size, n is sample size.

$$c_{IRS} = 1,\ c_{SRS} = \frac{N - n}{N - 1}$$

$$SD(\bar{X}_n)_{IRS} = \frac{\sigma_X}{\sqrt{n}},\ SD(T_n)_{IRS} = \sqrt{n}\sigma_X$$

$$SD(\bar{X}_n)_{SRS} = \sqrt{c_{SRS}} \cdot SD(\bar{X}_n)_{IRS} = \sqrt{\frac{N - n}{N - 1}} \cdot \frac{\sigma_X}{\sqrt{n}}$$

$$SD(T_n)_{SRS} = \sqrt{c_{SRS}} \cdot SD(T_n)_{IRS} = \sqrt{\frac{N - n}{N - 1}} \cdot \sqrt{n}\sigma_X$$

$$\frac{n}{n - c_{IRS}} = \frac{n}{n - 1}$$

$$\frac{n}{n - c_{SRS}} = \frac{N - 1}{N} \cdot \frac{n}{n - 1} = \frac{N - 1}{N} \cdot \frac{n}{n - c_{IRS}}$$

$$(S^2)_{SRS} = \frac{N - 1}{N} (S^2)_{IRS}$$

27. CHI-SQUARE DISTRIBUTION

Begin with standard normal Z and let $Z_1, Z_2, Z_3, ..., Z_d$ be an independent random sample of size d. Define W_d and χ^2_d by

$$\chi^2_d = f_{W^2},\ W^2_d = \sum_{k=1}^{d} Z^2_k,\ m_{W^2}(t) = (1-2t)^{-d/2}$$

If X is normal and IRS is used, then

$$f_U = \chi^2_{n-1},\ \text{where } U = \frac{(n-1)S^2}{\sigma^2_X},\ \text{and } S^2 = \frac{n}{n-1}\overline{(X-\bar{X}_n)^2}$$

The random variable W^2_d is said to have the **Chi-Square Distribution** with $d-$**Degrees of Freedom**.

28. TWO RANDOM VARIABLE SAMPLING

Start with two random variables X and Y on the same population with paired observations $(X_1, Y_1), (X_2, Y_2), X_3, Y_3), ..., (X_n, Y_n)$.

$$Cov(X, Y) = \rho \cdot \sigma_X \cdot \sigma_Y$$

$$F_{X_k} = F_X,\ E(X_k) = \mu_X,\ SD(X_k) = \sigma_X,\ k = 1, 2, 3, ..., n$$

$$F_{Y_k} = F_Y,\ E(Y_k) = \mu_Y,\ SD(Y_k) = \sigma_Y,\ k = 1, 2, 3, ..., n$$

$$T_X = \sum_{k=1}^{n} X_k,\ T_Y = \sum_{k=1}^{n} Y_k$$

$$Cov(X_k, Y_k) = \rho \cdot \sigma_X \cdot \sigma_Y,\ k = 1, 2, 3, ..., n$$

$$\overline{(X-\bar{X}_n)(Y_k-\bar{Y}_n)} = \frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X}_n)(Y_k-\bar{Y}_n)$$

$$\overline{(X-\bar{X})(Y-\bar{Y})} = \overline{XY} - [\bar{X}][\bar{Y}]$$

FAIR PAIRING CONDITIONS/CONSTANTS

$$c_X = \frac{Cov(X_k, T_X)}{\sigma^2_X},\ c_Y = \frac{Cov(Y_k, T_Y)}{\sigma^2_Y},\ k = 1, 2, 3, ..., n$$

$$d = \frac{Cov(X_k, T_Y)}{(\sqrt{c_X}\sigma_X) \cdot (\sqrt{c_Y}\sigma_Y)},\ k = 1, 2, 3, ..., n$$

$$e = \frac{Cov(T_X, Y_k)}{(\sqrt{c_X}\sigma_X) \cdot (\sqrt{c_Y}\sigma_Y)},\ k = 1, 2, 3, ..., n$$

Define

$$c = \sqrt{c_X \cdot c_Y},$$

It follows that $d = e$ and in fact $\rho(T_X, T_Y) = d = e = \rho(\bar{X}_n, \bar{Y}_n)$

$$Cov(T_X, T_Y) = ncd\sigma_X\sigma_Y = \rho(T_X, T_Y)(\sqrt{c_X n}\sigma_X)(\sqrt{c_Y n}\sigma_Y)$$

$$Cov(\bar{X}_n, \bar{Y}_n) = \frac{1}{n}c \cdot d \cdot \sigma_X \cdot \sigma_Y = d \cdot (\sqrt{c_X}\frac{\sigma_X}{\sqrt{n}}) \cdot (\sqrt{c_Y}\frac{\sigma_Y}{\sqrt{n}})$$

For SRS it is reasonable that $Cov(X_k, Y_l)$ for $k \neq l$ is otherwise independent of k, l and this implies

$$d_{IRS} = \rho = d_{SRS}.$$

Moreover, for SRS,

$$c_X = \frac{N-n}{N-1} = c_Y = c_{SRS},$$

On the other hand, if $\rho = 0$, then it is reasonable that $Cov(X_k, Y_l) = 0$ for any k, l, and thus in general, we can write

$$d = a\rho.$$

Then

$$E\left(\overline{(X-\bar{X})(Y-\bar{Y})}\right) = \left(1 - \frac{ac}{n}\right) Cov(X, Y),$$

so in general,

$$E\left(\left[\frac{n}{n-ac}\right] \overline{(X-\bar{X})(Y-\bar{Y}))}\right) = Cov(X, Y),$$

and

$$a_{IRS} = 1 = a_{SRS},$$

so

$$E\left(\left[\frac{n}{n-c}\right] \overline{(X-\bar{X})(Y-\bar{Y}))}\right) = Cov(X, Y), \text{ for } IRS \text{ or } SRS.$$

29. The t-DISTRIBUTION

Suppose that Z_0 and W_d^2 are independent random variables, that Z_0 is standard normal and that W_d^2 has chi-square distribution for d degrees of freedom. Then

$$t_d = \frac{Z_0}{\sqrt{\frac{W_d^2}{d}}}$$

has what is called the **Student $t-$Distribution** for d **Degrees of Freedom**.

If $Z_0, Z_1, Z_2, ..., Z_d$ are all mutually independent standard normal random variables, then with

$$W_d^2 = \sum_{k=1}^{d} Z_k^2,$$

the random variable the random variable W_d^2 has the chi-square distribution for d degrees of freedom and is indepedent of Z_0, so t_d can be defined as

$$t_d = \frac{Z_0}{\sqrt{\frac{W_d^2}{d}}},$$

with this specific choice of W_d^2.

As $d \to \infty$ the distribution of t_d becomes standard normal.

30. SAMPLING TO ESTIMATE THE MEAN

If X is a random variable with known standard deviation σ_X but with unknown mean, not a common circumstance, then in case $\bar{X}_n$ is normal, a confidence interval for the mean can easily be given for any **Level of Confidence** usually denoted C. The confidence level is usually specified in advance and we seek a **Margin of Error** denoted M so that

$$P(|\bar{X}_n - \mu_X| \leq M) = C.$$

We can standardize $\bar{X}_n$ and denote the result simply by Z, so Z is standard normal and using the inverse normal we choose z_C so that

$$P(|Z| \leq z_C) = C.$$

Then

$$M = z_C \cdot SD(\bar{X}_n).$$

Suppose X is a random variable and the sample unknowns $X_1, X_2, X_3, ..., X_n$ forming a sample of size n for X are pairwise uncorrelated, which is true in case of IRS. Then the random variables $X_2 - \bar{X}_1,\ X_3 - \bar{X}_2, ...,\ X_n - \bar{X}_{n-1},\ \bar{X}_n$ are all pairwise uncorrelated. Let Z_k denoting the standardization of $X_{k+1} - \bar{X}_k$, for $k < n$, and let Z_0 be the standardization of $\bar{X}_n$. We then have

$$\frac{(n-1)S^2}{\sigma_X^2} = \sum_{k=1}^{n-1} Z_k^2 = W_{n-1}^2$$

and

$$\frac{\bar{X}_n - \mu_X}{S/\sqrt{n}} = \frac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}} \cdot \frac{1}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma_X^2}}} = \frac{Z_0}{\sqrt{\frac{W_{n-1}^2}{n-1}}} = t_{n-1}$$

If X is normal and we have IRS, then as $Z_0, Z_1, Z_2, ..., Z_{n-1}$ are all pairwise uncorrelated and formed as linear combinations of the independent normal random variables $X_1, X_2, X_3, ..., X_n$, it follows from linear algebra that in fact $Z_0, Z_1, ..., Z_{n-1}$ are all independent standard normal random variables, and therfore t_{n-1} has the $t-$distribution for $n-1$ degrees of freedom. We can therefore use the inverse t-distribution in place of the inverse standard normal when we need to use s in place of σ_X, so denoting t_C the number chosen so that

$$P(|t_{n-1}| \leq t_C) = C,$$

we have now

$$M = t_C \cdot s,$$

with s the value of the sample standard deviation from the sample data.

31. SAMPLING TO ESTIMATE VARIANCE

If X is a normal random variable and we need to estimate both μ_X and σ_X, then the problem of estimating σ_X actually comes first, since it is used in estimating μ_X. As noted in (27. Chi-Square Distribution), the distribution of $(n-1)S^2/\sigma_X^2$ is chi-square with $n-1$ degrees of freedom. Thus, to make a confidence interval with confidence level C, we use χ_d^2 with $d = n-1$. If we set

$$(n-1)\frac{S^2}{\sigma_X^2} = W_d^2, \ d = n-1,$$

then denoting by $(W_d^2)_A$ the value of W_d^2 for which

$$P(W_d^2 \leq (W_d^2)_A) = A,$$

we have

$$A = \int_0^{(W_d^2)_A} \chi_d^2(x)dx,$$

is the area under the graph of χ_d^2 to the left of $(W_d^2)_A$.

So we want a region of area C in the middle under the graph of χ_d^2 and that means we symmetrically want the range of values of W_d^2 from $(W_d^2)_A$, to $(W_d^2)_B$, where $A = (1-C)/2$, and $B = 1-(1-C)/2 = (1+C)/2$.

Therefore,

$$P\left[(W_d^2)_A \leq \frac{(n-1)S^2}{\sigma_X^2} \leq (W_d^2)_B\right] = C.$$

Since taking reciprocals reverses the order for positive numbers, the inequality in the brackets is equivalent to

$$\frac{(n-1)S^2}{(W_d^2)_B} \leq \sigma_X^2 \leq \frac{(n-1)S^2}{(W_d^2)_A},$$

so this is the confidence interval for σ_X^2 with confidence level C.

32. SAMPLING TWO INDEPENDENT UNKNOWNS

When trying to determine the difference in means of two exclusive populations, we can form the population of all pairs which can be formed by taking the first member of the pair from the first population and the second member of the pair from the second population. Specifically, if A and B are sets and X is a random variable on A and B is a random variable on B, then we can form the **Cartesian Product**, denoted $A \times B$, and defined by

$$A \times B = \{(a,b) : a \in A \ \& \ b \in B\}.$$

A random variable X on A is a real-valued function which when the outcome is $a \in A$ the value of X is the number $X(a)$. Likewise, if Y is a random variable on B, then for $b \in B$, the value $Y(b)$ is the value of Y for the outcome b. The random variable X can then be viewed as a random variable on the set $A \times B$ by defining

$$X(a,b) = X(a),$$

and likewise, Y can be viewed as a random variable on $A \times B$ by defining

$$Y(a,b) = Y(b).$$

Thus, on $A \times B$, the variable X ignores the second entry of a pair and Y ignores the first entry of a pair. The result is that X and Y become independent random variables on the common population $A \times B$.

Suppose now we have $X_1, X_2, X_3, ..., X_{n_X}$ and $Y_1, Y_2, Y-3, ..., Y_{n_Y}$ are independent random samples for each random variable, so $\bar{X}_{n_X}$ or simply $\bar{X}$ is the sample mean random variable for the population A, and likewise $\bar{Y}_{n_Y}$ or simply $\bar{Y}$ is the sample mean random variable for the population B. These two sample mean random variables then become independent random variables on $A \times B$, and

$$E(\bar{X} - \bar{Y}) = \mu_X - \mu_Y.$$

But, since they are indepedent,

$$SD(\bar{X} - \bar{Y}) = \sqrt{Var(\bar{X}) + Var(\bar{Y})}.$$

On the other hand, since the samples are IRS's,

$$Var(\bar{X}) = \frac{\sigma_X^2}{n_x} \text{ and } Var(\bar{Y}) = \frac{\sigma_Y^2}{n_Y}.$$

Therefore

$$SD(\bar{X} - \bar{Y}) = \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}.$$

Thus, if σ_X and σ_Y are known, then for the case of normal random variables, the margin of error with confidence C in a confidence level is simply

$$z_C \cdot SD(\bar{X} - \bar{Y}),$$

where

$$z_C = invNorm\left(\frac{1+C}{2}, 0, 1\right).$$

If you do not know σ_X^2 and σ_Y^2, then you must estimate them from the sample data using sample variances S_X^2 and S_Y^2. Of course then, you are dealing with the distribution of

$$t_d = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}},$$

and in certain cases, this can be shown to be a $t-$distribution for d degrees of freedom. How to find d then is the next problem, and the result depends on the type of situation.

CASE $\sigma_X = \sigma_Y$.

Then with $\sigma_X = \sigma = \sigma_Y$, we have

$$SD(\bar{X} - \bar{Y}) = \sqrt{\sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}.$$

In this case, $(n_X - 1)S_X^2/\sigma^2$ and $(n_Y - 1)S_Y^2/\sigma^2$, are independent chi-square distributed variables with degrees of freedom $df_X = n_X - 1$ and $df_Y = n_Y - 1$, respectively, so their sum,

$$\frac{(n_X - 1)S_X^2}{\sigma^2} + \frac{(n_Y - 1)S_Y^2}{\sigma^2} = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{\sigma^2}$$

also has chi-square distribution with degrees of freedom the total for each sample. Here S_{pool}^2, called the **Pooled Variance** defined by

$$S_{pool}^2 = \frac{df_X \cdot S_X^2 + df_Y \cdot S_Y^2}{df_X + df_Y}$$

then has the property

$$\frac{S_{pool}^2}{\sigma^2} = \frac{W_d^2}{d},\ d = df_X + df_Y,$$

with W_d^2 having the chi-square distribution for d degrees of freedom found by simply adding degrees of freedom for each sample,

$$d = df_X + df_Y.$$

Case where standard deviations are unknown and there relationship to each other is unkown.

Here, the distribution is not really known, but there seems to be consensus among statisticians that

$$\frac{W_d^2}{d} = \frac{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}{SD(\bar{X} - \bar{Y})}$$

where W_d should be chi-square distributed for some number of degrees of freedom d, and the two terms in the numerator on the right hand side are independent of each other. The expected value of any chi-square distribution is the number of degrees of freedom, but our equation does not give us anything other than $d/d = 1$, a triviality. Using independence of the two terms in the numerator allows us to compute the variance of the numerator and equate that to the variance of W_d^2 to get an equation for d. The result is (for instance, see EXPECTATION PRIMER)

$$d = \frac{\left(\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)^2}{\frac{1}{df_X}\left(\frac{\sigma_X^2}{n_X}\right)^2 + \frac{1}{df_Y}\left(\frac{\sigma_Y^2}{n_Y}\right)^2}.$$

Notice now that d may not even be a whole number and that we need the values of the standard deviations to know the degrees of freedom. Accepted practice here is to use the sample standard deviations in place of the actual standard deviations to determine the degrees of freedom d, and then the confidence interval is calculated with the $t-$distribution for d degrees of freedom.

DEPARTMENT OF MATHEMATICS, TULANE UNIVERSTIY, NEW ORLEANS, LA 70118
E-mail address: mdupre@tulane.edu