# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

| concepts | list of questions |
|---|---|
| 01 introduction to statistics and data analysis | ch 1: n/a |
| 02 probability | ch 2: 95, 115, 120 |
| 03 random variable’s probability distributions | ch 3: 26, 56, 68 |
| 04 mathematical expectation | ch 4: 47, 70 |
| 05 discrete probability distributions | ch 5: 16, 71,87 |
| 06 continuous probability distributions | ch 6: 8, 32, 66, example 6.17 |
| 08 fundamental sampling distributions data descriptions | ch 8: 12, 23, 27, 29, 48, 49 |
| 09 one and two sample estimation | ch 9: 4, 10, 14, 38, 46, 55, 66 |
| 10 one and two sample test hyptheses | ch 10: 23 (add p-value approach), 24, 42, 46, 59, 62 |

# probability and statistics for engineers and scientists

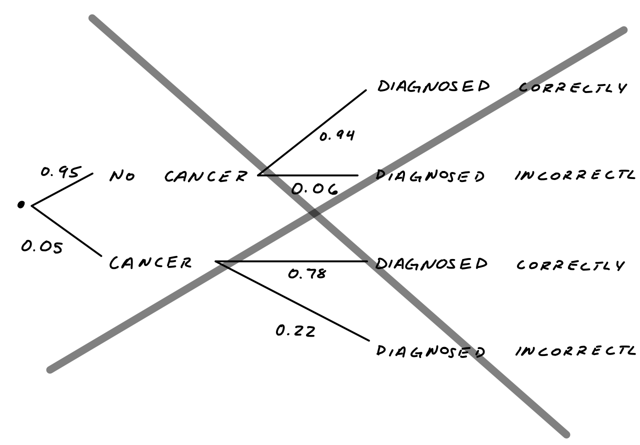
morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

2.95 in a certain region of the country it is known from past experience that the probability of selecting an adult over 40 years of age with cancer is 0.05. if the probability of a doctor correctly diagnosing a person with cancer as having the disease is 0.78 and the probability of incorrectly diagnosing a person without cancer as having the disease is 0.06, what is the probability that an adult over 40 years of age is diagnosed as having cancer?

C DENOTES THE EVENT THAT A PERSON HAS CANCER
C' DENOTES THE EVENT THAT A PERSON DOES NOT HAVE CANCER
D DENOTE THE EVENT THAT A PERSON HAS BEEN DIAGNOSED WITH CANCER




$P(C) = 0.05$ HAS CANCER
$P(C') = 1 - P(C) = 0.95$ NO CANCER
$P(D \mid C) = 0.78$ HAS CANCER DIAGNOSED CORRECTLY
$P(D \mid C') = 0.06$ NO CANCER DIAGNOSED INCORRECTLY
$P(D) = ?$ DIAGNOSED WITH CANCER

ACCORDING TO THEOREM OF TOTAL PROBABILITY

$$P(D) = P(D \mid C) \cdot P(C) + P(D \mid C') \cdot P(C')$$
$$= 0.78 \cdot 0.05 + 0.06 \cdot 0.95$$
$$= 0.039 + 0.057$$
$$\boxed{P(D) = 0.096}$$

2.95 in a certain region of the country it is known from past experience that the probability of selecting an adult over 40 years of age with cancer is 0.05. if the probability of a doctor correctly diagnosing a person with cancer as having the disease is 0.78 and the probability of incorrectly diagnosing a person without cancer as having the disease is 0.06, what is the probability that an adult over 40 years of age is diagnosed as having cancer?

$P(C) = 0.05$ CANCER

$P(C') = 1 - P(C) = 0.95$ NO CANCER

$P(C|D) = 0.78$ CORRECT DIAGNOSIS CANCER

$P(C'|D) = 0.06$ INCORRECT DIAGNOSIS NO CANCER

$P(D) = ?$

ACCORDING TO THE TOTAL LAW OF PROBABILITY

$P(D) = P(C|D) \cdot P(C) + P(C'|D) \cdot P(C')$

$P(D) = 0.78 \cdot 0.05 + 0.06 \cdot 0.95$

$\boxed{P(D) = 0.096}$

# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

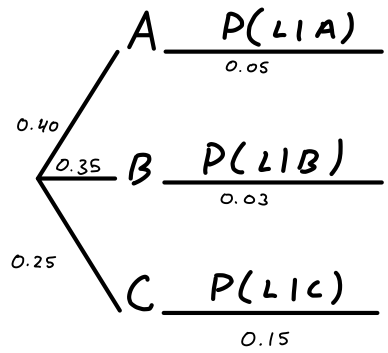
may 03 2023 - may 09 2023

2.115 a certain federal agency employs three consulting firms (A, B, and C) with probabilities 0.40, 0.35, and 0.25, respectively. from past experience it is known that the probability of cost overruns for the firms are 0.05, 0.03, and 0.15, respectively. suppose a cost overrun is experienced by the agency.

a. what is the probability that the consulting firm involved is company C?

b. what is the probability that it is company A?

$P(A) = 0.40$ COMPANY GETS CHOSEN
$P(B) = 0.35$ COMPANY GETS CHOSEN
$P(C) = 0.25$ COMPANY GETS CHOSEN
$P(L|A) = 0.05$ COMPAY A EXPERIENCES COST OVERRUNS
$P(L|B) = 0.03$ COMPANY B EXPERIENCES COST OVERRUNS
$P(L|C) = 0.15$ COMPANY C EXPERIENCES COST OVERRUNS




$P(A) \cdot P(L|A) = 0.40 \cdot 0.05 = 0.02$

$P(B) \cdot P(L|B) =$

$P(C) \cdot P(L|C)$

$P(C|L) = ?$
$P(A|L) = ?$

USING BAYES THEOREM

$$P(C|L) = \frac{P(C)P(L|C)}{P(C)P(L|C) + P(B)P(L|B) + P(A)P(L|A)} = \frac{(0.40)(0.05)}{(0.40)(0.05) + (0.35)(0.03) + (0.25)(0.15)} = 0\ 2941$$

$P(C|L) = 2.941$

$$P(A|L) = \frac{P(A)P(L|A)}{P(A)P(L|A) + P(B)P(L|B) + P(C)P(L|C)} = \frac{(0.40)(0.05)}{(0.40)(0.05) + (0.35)(0.03) + (0.25)(0.15)} = 0.294$$

$P(A|L) = 0.294$

# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

2.120 a rare disease exists with which only 1 in 500 is affected. a test for the disease exists, but of course it is not infallible. a correct positive result (patient actually has the disease) occurs 95% of the time, while a false positive result (patient does not have the disease) occurs 1% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

$P(D) = 1/500 = 0.002$ D DENOTES POSITIVE

$P(D') = 1 - P(D) = 0.998$ D' DENOTES NEGATIVE

$P(T) = ?$ T DENOTES CORRECT / INCORRECT POSITIVE RESULT

$P(T \mid D) = 0.95$ POSITIVE RESULT GIVEN NO DISEASE

$P(T \mid D') = 0.05$ POSITIVE RESULT GIVEN DISEASE

$$P(D \mid T) = \frac{P(T \mid D)\, P(D)}{P(T \mid D) P(D) + P(T \mid P(T \mid D')} = \frac{0.002\,(0.95)}{(0.95)(0.002) + (0.01)(0.998)} = 0.16$$

$P(D \mid T) = 0.16$

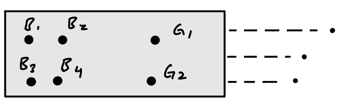
# probability and statistics for engineers and scientists

morgan bergen

## final date may 09 2023

may 03 2023 - may 09 2023

3.26 from a box containing 4 black balls and 2 green balls, 3 balls are drawn in succession, each ball being replaced in the box before the next draw is made. Find the probability distribution for the number of green balls.




IF X REPRESENTS A RANDOM VARIABLE THAT REPRESENTS THE NUMBER OF GREEN BALLS DRAWN IN 3 ATTEMPTS

SAMPLING WITH REPLACEMENT

$N = 6$      $X$   $G_1$ $G_2$ $B_1$

GREEN DRAW $= 2/6$    $G_1$ $G_2$ $B_2$

BLUE DRAW $= 4/6$    $G_1$ $G_2$ $B_3$

$G_N = 2$

$B_N = 4$

IN ORDER TO OBTAIN THE PROBABILITY DISTRIBUTION FOR RV $X$

$$P(X = x) = \left(\frac{2}{6}\right)^x \left(\frac{4}{6}\right)^{3-x} = \frac{2^{3-x}}{27}$$

$$\boxed{P(X = x) = \frac{2^{3-x}}{27}}$$

# probability and statistics for engineers and scientists

morgan bergen

## final date may 09 2023

may 03 2023 - may 09 2023

**3.56** the joint density function of the random variables X and Y is

f(x, y) = { 6x | 0 < x < 1, 0 < y < 1 - x; 0 | elsewhere }

a. Show that X and Y are not independent

b. Find P(X > 0.3 | Y = 0.5)

PDF JOINT FOR RV $X$ & $Y$

$$f(x,y) = \begin{cases} 6x, & 0 < x < 1,\ 0 < y < 1-x \\ 0, & \text{ELSEWHERE} \end{cases}$$

THE RANDOM VARIABLES $X$ & $Y$ WILL BE INDEPENDENT IFF THE EQUALITY HOLDS TRUE FOR ALL VALUES $x$ & $y$ WHICH ARE THE VARIABLES WE ASSUME

$$f(x,y) = g(x)\,h(y)$$

MARGINAL DENSITY FUNCTION $g(x)$ IS,

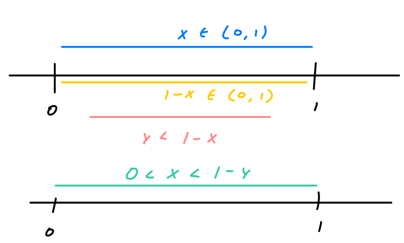
$$g(x) = \int_y f(x,y)\,dy = \int_0^{1-x} 6x\,dy = 6xy\Big|_0^{1-x} = 6x(1-x) - 6x(0) = 6x - 6x^2 \quad \text{FOR } 0 < x < 1$$

MARGINAL DENSITY FUNCTION $h(x)$ IS,

$$h(y) = \int_x f(x,y) = \int_0^{1-y} 6x\,dx = \frac{6x^2}{2}\Big|_0^{1-y} = 3x^2\Big|_0^{1-y} = 3(1-y)^2 - 3(0)^2 = 3(1-y)(1-y) = 3(1-y-y+y^2) = 3(1-2y+y^2) = 3 - 6y + 3y^2$$

$6x,\quad 0 < x < 1 \Rightarrow \int_x = \int_0^{1-y}$ $\qquad y \in \langle 0, 1 \rangle$ BECAUSE $y < 1-x$

$0 < y < 1-x \Rightarrow \int_y = \int_0^{1-x}$ $\qquad x < 1-y \Rightarrow 0 < x < 1-y$




$$f(x,y) = g(x)\,h(y) = (6x - 6x^2)(3 - 6y + 3y^2)$$

$$6x = 12x - 36xy + 18xy - 18x^2 + 36x^2y - 18y^2x^2$$

$$6x \neq -18y^2y^2 - 18xy + 36x^2y + 12x$$

$\boxed{\text{THEREFORE } X \text{ \& } Y \text{ ARE NOT INDEPENDENT}}$

FIND $P(X > 0.3 \mid y = 0.5)$

$$P(x > 0.3 \mid y = 0.5) = P(0.3 < x < 1-y \mid y = 0.5)$$
$$= P(0.3 < x < 1 - 0.5)$$
$$= P(0.3 < x < 0.5 \mid y = 0.5)$$

$$P(0.3 < x < 0.5 \mid y = 0.5) = \int_x f(x \mid 0.5)\,dx$$
$$= \int_{0.3}^{0.5} \frac{f(x, 0.5)}{h(0.5)}\,dx$$
$$= \int_{0.3}^{0.5} \frac{6x}{3(1-0.5)^2}\,dx$$
$$= \int_{0.3}^{0.5} 8x\,dx = \frac{8x^2}{2}\Big|_{0.3}^{0.5}$$
$$= 4(0.5)^2 - 4(0.3)^2$$
$$= 0.64$$

$\boxed{\therefore\ P(x > 0.3 \mid y = 0.5) = 0.64}$

# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

3.68 consider the following joint probability density function of the random variables X and Y

f(x, y) = { (3x - y)/9 | 1 < x < 3, 1 < y < 2; 0 | elsewhere }

a. find the marginal density functions of X and Y

b. are X and Y independent?

c. find P(X > 2)

$$f(x,y) = \begin{cases} 3x-y/9, & 1<x<3,\ 1<y<2 \\ 0, & \text{ELSEWHERE} \end{cases}$$

$$f(x,y) = g(x)\,h(y)$$

$$g(x) = \int_y f(x,y)\,dy = \int_1^2 \frac{3x-y}{9}\,dy = \int_1^2 \frac{3x}{9} - \frac{y}{9}\,dy = \frac{3xy}{18} - \frac{y^2}{18}\Big|_1^2 = \left(\frac{3x(2)}{9} - \frac{2^2}{18}\right) - \left(\frac{3x(1)}{9} - \frac{1^2}{18}\right) = \frac{x}{3} - \frac{1}{6} \quad \text{FOR } 1<x<3$$

$$h(y) = \int_x f(x,y)\,dx = \int_1^3 \frac{3x-y}{9}\,dx = \int_1^3 \frac{3x}{9} - \frac{y}{9}\,dx = \frac{1}{3}\frac{x^2}{2} - \frac{xy}{9}\Big|_1^3 = \left(\frac{3^2}{6} - \frac{1^2}{6}\right) - \left(\frac{3y}{9} - \frac{y}{9}\right) = \frac{4}{3} - \frac{2y}{9} \quad \text{FOR } 1<y<2$$

THE MARGINAL DENSITY FUNCTION

$$g(x) = \frac{x}{3} - \frac{1}{6} \quad \text{FOR } 1<x<3$$

$$h(y) = \frac{4}{3} - \frac{2y}{9} \quad \text{FOR } 1<y<2$$

ARE X & Y INDEPENDENT?

$$f(x,y) = g(x)\,h(y) = \left(\frac{x}{3} - \frac{1}{6}\right)\left(\frac{4}{3} - \frac{2y}{9}\right) = \left(\frac{2x-1}{6}\right)\left(\frac{12-2y}{9}\right) = \frac{(2x-1)(6-y)}{27}$$

$$f(x,y) = \frac{3x-y}{9} = g(x)\,h(y) = \frac{(2x-1)(6-y)}{27}$$

$$\boxed{f(x,y) \neq g(x)\,h(y) \ \therefore \ \text{ARE NOT INDEPENDENT}}$$

FIND $P(X>2)$

$$P(X>2) = P(2<X<3,\ 1<Y<2)$$

$$= \int_x\int_y \frac{3x-y}{9}\,dy\,dx = \int_2^3\int_1^2 \frac{3x-y}{9}\,dy\,dx = \int_2^3 \left(\frac{3xy}{9} - \frac{y^2}{18}\right)\Big|_1^2 dx = \int_2^3 \frac{6x-2}{9}\,dx = \left(\frac{6x^2}{18} - \frac{2x}{9}\right)\Big|_2^3 = \frac{2}{3}$$

$$\boxed{P(2<x<3,\ 1<y<2) = \frac{2}{3}}$$

# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

**4.47** for the random variables X and Y whose joint density function is given in exercise 3.40 page 105, find the covariance

**3.40** suppose the joint density function for the random variables X and Y is

f(x, y) = { 2/3(x + 2y), 0 ≤ x ≤ 1, 0 ≤ y ≤ 1; 0, elsewhere }

find the marginal density function of X and Y

find the probability that the drive through facility is busy less than one half of the time

FIND COVARIANCE

JOINT DENSITY FUNCTION FOR X & Y

$$f(x,y) = \begin{cases} 2/3(x+2y) & , 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & , \text{ELSEWHERE} \end{cases}$$

IN ORDER TO FIND THE COVARIANCE

$\sigma_{XY} = E(XY) - \mu_X \mu_Y$

FIRST MARGINAL DENSITY FUNCTION

$f(x,y) = g(x)h(y)$

$$g(x) = \int_y f(x,y)\,dy = \int_0^1 \frac{2}{3}(x+2y)\,dy = \frac{2}{3}\int_0^1 x+2y\,dy = \frac{2}{3}\left(xy + y^2\right)\Big|_{y=0}^{y=1} = \frac{2}{3}(x(1)+1^2) - \frac{2}{3}(x(0)+0^2) = \frac{2x+2}{3}, \quad 0 \le x \le 1$$

$$h(y) = \int_x f(x,y)\,dx = \int_0^1 \frac{2}{3}(x+2y)\,dx = \frac{2}{3}\int_0^1 x+2y\,dx = \frac{2}{3}\left(\frac{x^2}{2} + 2xy\right)\Big|_{x=0}^{x=1} = \frac{1}{3} + \frac{4y}{3} = \frac{4y+1}{3}, \quad 0 \le y \le 1$$

$$g(x) = \frac{2x+2}{3}, \quad 0 \le x \le 1$$

$$h(y) = \frac{4y+1}{3}, \quad 0 \le y \le 1$$

$\sigma_{XY} = E(XY) - \mu_X \mu_Y$

$$\mu_X = \int_x x\,g(x)\,dx = \int_0^1 x\left(\frac{2x+2}{3}\right)dx = \int_0^1 \frac{2x^2}{3} + \frac{2x}{3}\,dx = \frac{2x^3}{9} + \frac{2x^2}{6}\Big|_0^1 = \frac{2}{9} + \frac{2}{6} = \frac{8}{36} + \frac{12}{36} = \frac{5}{9}$$

$$\mu_Y = \int_y y\,h(y)\,dy = \int_0^1 y\left(\frac{4y+1}{3}\right)dy = \int_0^1 \frac{4y^2+y}{3}\,dy = \frac{4y^3}{9} + \frac{y^2}{6}\Big|_0^1 = \frac{4}{9} + \frac{1}{6} = \frac{24}{36} + \frac{6}{36} = \frac{11}{18}$$

$$E(XY) = \int_x\int_y xy\,f(x,y)\,dy\,dx = \frac{2}{3}\int_0^1\int_0^1 x^2y + 2xy^2\,dy\,dx = \frac{2}{3}\left(\frac{1}{3}\cdot\frac{1}{2}x^3 + \frac{1}{2}\cdot\frac{2}{3}x^2\right)\Big|_0^1 = \frac{1}{3}$$

$\sigma_{XY} = E(XY) - \mu_X \mu_Y$

$\sigma_{XY} = 1/3 - (5/9)(11/18) = -1/162$

$\boxed{\text{COVARIANCE IS } \sigma_{XY} = -1/162}$

# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

**4.70** consider review exercise 3.46 on page 107. there are two service lines. the random variables X and Y are the proportions of time that line 1 and line 2 are in use respectively. the joint probability density function for (X, Y) is given by

f(x, y) = { 3/2(x^2 + y^2), 0 ≤ x, y ≤ 1; 0, elsewhere }

**a.** determine whether or not X and Y are independent

**b.** it is of interest to know something about the proportion of Z = X + Y, the sum of the two proportions. find E(X + Y), also find E(XY)

**c.** find Var(X), Var(Y), Cov(X, Y)

**d.** find Var(X + Y)

$$f(x,y) = \begin{cases} 3/2(x^2+y^2), & 0 \le x,\ y \le 1 \\ 0, & \text{ELSEWHERE} \end{cases}$$

$$g(x) = \int_y f(x,y)\,dy = \int_0^1 \frac{3}{2}(x^2+y^2)\,dy = \frac{3}{2}\left(x^2y + y^3/3\right)\Big|_0^1 = \frac{3x^2y}{2} + \frac{y^3}{3}\Big|_0^1 = \frac{3}{2}x^2 + \frac{1}{2}, \quad \text{FOR } y \le 1$$

$$h(y) = \int_x f(x,y)\,dx = \frac{3}{2}\int_0^1 x^2 + y^2\,dx = \frac{3}{2}\left(\frac{x^3}{3} + xy^2\right)\Big|_0^1 = \frac{3}{2}\left(y^2 + \frac{1}{3}\right) = \frac{3}{2}y^2 + \frac{1}{2}$$

$$f(x,y) = g(x)h(y)$$

$$\frac{3x^2+3y^2}{2} \ne \left(\frac{3x^2+1}{2}\right)\left(\frac{3y^2+1}{2}\right) \quad \therefore \text{NOT INDEPENDENT}$$

$$E(X+Y) = E(X) + E(Y)$$

$$E(X) = \int_x x\,g(x)\,dx = \int_0^1 x\,\frac{3x^2+1}{2}\,dx = \frac{5}{8}$$

$$E(Y) = \int_Y y\,h(y)\,dy = \int_0^1 y\,\frac{3y^2+1}{2}\,dy = \frac{5}{8}$$

$$E(Z) = E(X+Y) = E(X) + E(Y) = 5/4$$

$$E(XY) = \int_x\int_y xy\,f(x,y) \int_0^1\int_0^1 xy \cdot \frac{3}{2}(x^2+y^2)\,dy\,dx = \frac{3}{8}$$

FIND VAR(X), VAR(Y), COV(X, Y)

$$VAR(X) = E(X^2) - (E(X))^2$$

$$VAR(Y) = E(Y^2)\ (E(Y))^2$$

$$COV(X,Y) = E(XY) - E(X)E(Y)$$

$$E(X^2) = \int_y\int_x x^2 f(x,y)\,dx\,dy$$

$$E(Y^2) = \int_x\int_y y^2 f(x,y)\,dy\,dx$$

$$E(X^2) = \int_x\int_y x^2 f(x,y)\,dy\,dx = \int_x x^2 \int_y f(x,y)\,dy\,dx = \int_0^1 x^2 g(x)\,dx = \int_0^1 x^2\ \frac{3}{2}x^2 + \frac{1}{2}\,dx = \frac{7}{15}$$

$$E(Y^2) = \int_y\int_x y^2 f(x,y)\,dy = \int_y y^2 \int_x f(x,y)\,dy = \int_0^1 y^2 h(y)\,dy = \int_0^1 y^2\ \frac{3}{2}y^2 + \frac{1}{2}\,dy = \frac{7}{15}$$

$$VAR(X) = E(X^2) - E(X)^2 = 7/15 - (5/8)^2 = 7/15 - 25/64 = 73/960$$

$$VAR(Y) = E(Y^2) - E(Y)^2 = 7/15 - (5/8)^2 = 7/15 - 25/64 = 73/960$$

$$COV(X,Y) = E(XY) - E(X)E(Y) = 3/8 - 5/8 \cdot 5/8 = -1/64$$

FIND $VAR(X+Y) = VAR(X) + VAR(Y) + 2COV(X,Y)$

$$VAR(X+Y) = VAR(X) + VAR(Y) + 2COV(X,Y) = 73/960 + 73/960 + 2(-1/64) = 29/240$$

**5.16** suppose that airplane engines operate independently and fail with probability equal to 0.4. assuming that a plane makes a safe flight if at least one-half of its engines run, determine whether a 4-engine plane or a 2- engine plane has the higher probability for a successful flight.

PROBABILITY 0.4 OF FAILED ENGINES FOR BOTH 4 & 2

PLANE SUCCEEDS IF 1/2 OF ENGINES RUN

OBJECTIVE IS TO FIND WHETHER A 4 ENGINE OR 2 ENGINE PLANE SUCCEEDS

FAIL $p' = 0.4$

SUCCESS $p = 1 - p' = 0.6$

NUMBER OF ENGINES IS $N$

$X$ RANDOM VARIABLE REPRESENTS THAT ENGINE RUNS PROPERLY

$X$ BINOMIAL RV WITH $N = 4$, $p = 0.6$

IF MORE THAN HALF OF THE ENGINES SUCCEED

$P(X \geq N/2) = P(X \geq 2)$

4 ENGINE PLANE $N = 4$, $p = 0.6$, $q = 0.4$

$$P(X = x) = b(x; N, p) = \binom{N}{x} p^x q^{N-x} = \binom{N}{x} p^x (1-p)^{N-x}$$

$$P(X = x) = b(x; N, p) = \binom{N}{x} p^x (q)^{N-x} = \binom{4}{x} 0.6^x (0.4)^{4-x}, \quad x = 0, 1, 2, 3, 4$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$$
$$= 1 - \sum_{i=0}^{1} P(X = i) = 1 - \sum_{i=0}^{1} b(i; 4, 0.6)$$

$$P(X \geq 2) = 1 - \sum_{i=0}^{1} b(i; 4, 0.6) = 1 - 0.1792 = 0.8208$$

$P(X \geq 2) = 0.8208$ FOR 4 ENGINE PLANE

FOR A 2 PLANE ENGINE, $N = 2$ $p = 0.6$, $q = 0.4$

$$P(X = x) = b(x; 2, 0.6) = \binom{N}{x} p^x q^{N-x} = \binom{2}{x} 0.6^x \, 0.4^{2-x}$$

$$P(X \geq 1) = 1 - P(X > 1) = 1 - P(X = 0)$$
$$= 1 - b(x; 2, 0.6) = 1 - \binom{2}{0} 0.6^0 \, 0.4^{2-0} = 1 - 0.16 = 0.84$$

$P(X \geq 1) = 0.84$ FOR 2 ENGINE PLANE

$$\boxed{\begin{array}{l} P(X \geq 1) = 0.84 > P(X \geq 2) = 0.8208 \\ \text{THUS 2 ENGINE PLANE HAS HIGHER PROB} \end{array}}$$

**5.71** for a certain type of copper wire, it is known that, on the average, 1.5 flaws occur per millimeter. assuming that the number of flaws is a Poisson random variable, what is the probability that no flaws occur in a certain portion of wire of length 5 millimeters? what is the mean number of flaws in a portion of length 5 millimeters?

$X$ REPRESENTS THE # OF FLAWS IN 5 mm WIRE

$\lambda = 1.5$ REPRESENTS AVERAGE FLAW PER 1 mm

$t = 5$ REPRESENTS INTERVAL OF LENGTH IS 5 mm

$X$ HAS A POISSON DISTRO

$\lambda t = 7.5$ REPRESENTS POISSON PARAMETER

PROBABILITY MASS FUNCTION

$P(X=x) = p(x; \lambda t)$ WHERE $x = 0, 1, 2, \ldots$

$$P(X=x) = p(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

NO FLAWS OCCUR $P(X=0)$

$$P(X=0) = p(0; \lambda t) = p(0; 7.5) = \frac{e^{-7.5}(7.5)^0}{0!} = e^{-7.5} = 0.000553$$

MEAN NUMBER OF FLAWS $t = 5$

$$\mu = E(X) = \lambda t = 5 \cdot 1.5 = 7.5$$

5.87 imperfections in computer circuit boards and computer chips lend themselves to statistical treatment.

for a particular type of board, the probability of a diode failure is 0.03 and the board contains 200 diodes.

a. what is the mean number of failures among the diodes?

b. what is the variance?

c. the board will work if there are no defective diodes. what is the probability that a board will work?

P OF FAILURE 0.03 DIODE IN BOARD

BOARD CONTAINS N = 200 DIODES

$p = 0.03, \quad q = 1 - 0.03 = 0.97, \quad n = 200$

TRIALS ARE INDEPENDENT

MEAN NUMBER OF FAILURES

$\mu = E(x) = np = 200 \cdot 0.03 = 6$

VARIANCE

$\sigma^2 = np(1-p) = 200(0.03)(q) = 200 \cdot 0.03(0.97) = 5.82$

PROBABILITY THERE ARE $x=0$ DEFECTS

$$P(X=0) = p(x;\mu) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, 3, \dots$$

$$P(X=0) = \frac{e^{-6}(6)^0}{0!} = 0.00248$$

6.8 given a normal distribution with $\mu = 30$ and $\sigma = 6$, find

a. the normal curve area to the right of $x = 17$;

b. the normal curve area to the left of $x = 22$;

c. the normal curve area between $x = 32$ and $x = 41$;

d. the value of $x$ that has 80% of the normal curve area to the left;

e. the two values of $x$ that contain the middle 75% of the normal curve area.

NORMAL DISTRIBUTION $\mu = 30,\ \sigma = 6$

$z = \frac{x - \mu}{\sigma}$ , $x = \sigma z + \mu$

FIND THE NORMAL CURVE TO THE RIGHT OF $x = 17$

$P(X \geq 17) = P(Z \geq z)$

$z = \frac{17 - 30}{6} = -2.17$

$P(X \geq 17) = P(Z > -2.17) = 1 - P(Z \leq -2.17) = 1 - 0.0150 = 0.985$

$P(X \geq 17) = 0.985$

NORMAL CURVE TO THE LEFT OF $x = 22$

$P(X \leq 22) = P\left(Z < \left(z = \frac{x - \mu}{\sigma}\right)\right) = P\left(Z < z = \frac{22 - 30}{6}\right) = P(Z \leq -1.\bar{3}) = 0.0918$

$\boxed{P(X \leq 22) = 0.918}$

NORMAL CURVE $x = 32$ & $x = 41$

$P(32 \leq X \leq 41) = P\left(\frac{32 - 30}{6} \leq Z \leq \frac{41 - 30}{6}\right)$

$= P(0.\bar{3} \leq Z \leq 1.8\bar{3})$

$= P(Z < 1.8\bar{3}) - P(Z < 0.\bar{3})$

$= 0.9664 - 0.6293$

$\boxed{P(32 \leq X \leq 41) = 0.3371}$

FIND THE VALUE OF X HAS 80%.

$z = \frac{x - \mu}{\sigma} \Rightarrow x = \sigma z + \mu = 6 \cdot 0.842 + 30 = 35.05$

LETS FIND X VALUE IN THE MIDDLE 75% NORMAL CURVE

$P(-z \leq Z \leq z) = 0.75$

$P(Z \leq z) - (1 - P(Z \leq -z)) = 0.75$

$2P(Z \leq z) - 1 = 0.75$

$2P(Z \leq z) = 1.75$

$P(Z \leq z) = 0.875$

$z_1 = -1.15$ & $z_2 = 1.15$

$z_1 = \frac{x - \mu}{\sigma} \Rightarrow x_1 = \sigma z_1 + \mu = 6 \cdot (-1.15) + 30 = 23.1$

$z_2 = \frac{x - \mu}{\sigma} \Rightarrow x_2 = \sigma z_2 + \mu = 6 \cdot (1.15) + 30 = 36.9$

$\boxed{x_1 = 23.1 \quad \& \quad x_2 = 36.9}$

**6.66** a certain type of device has an advertised failure rate of 0.01 per hour. the failure rate is constant and the exponential distribution applies.

a. what is the mean time to failure?

b. what is the probability that 200 hours will pass before a failure is observed?

FAILURE RATE $\lambda = 0.01$

LET RANDOM VARIABLE $X$ REPRESENT THE FAILURE RATE OF A CERTAIN DEVICE

$X$ IS AN EXPONENTIAL DISTRIBUTION WITH PARAMETER $\lambda = 0.01$

MEAN TIME TO FAILURE

$\mu = 1/\lambda = 1/0.01 = 100$

PROBABILITY OF FAILURE AFTER 200 HRS

$P(x \geq 200) = 1 - P(x \leq 200) = 1 - (1 - e^{-0.01 \cdot 200}) = 0.1353$

$P(x \geq 200) = 0.1353$

**Example 6.17:** Suppose that a system contains a certain type of component whose time, in years, to failure is given by $T$. The random variable $T$ is modeled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

***Solution:*** The probability that a given component is still functioning after 8 years is given by

$$P(T > 8) = \frac{1}{5}\int_8^\infty e^{-t/5}\, dt = e^{-8/5} \approx 0.2.$$

Let $X$ represent the number of components functioning after 8 years. Then using the binomial distribution, we have

$$P(X \geq 2) = \sum_{x=2}^{5} b(x; 5, 0.2) = 1 - \sum_{x=0}^{1} b(x; 5, 0.2) = 1 - 0.7373 = 0.2627.$$

There are exercises and examples in Chapter 3 where the reader has already encountered the exponential distribution. Others involving waiting time and reliability include Example 6.24 and some of the exercises and review exercises at the end of this chapter.

$T$ RANDOM VARIABLE THAT FOLLOWS EXPONENTIAL DISTRIBUTION

$\beta = 5$ MEAN TIME FOR FAILURE

$N = 5$ COMPONENTS

$t = 8$ TIME IN YEARS

$$P(T > t) = \frac{1}{\beta}\int_t^\infty e^{-t/N}\, dt$$

$$P(T > 8) = \frac{1}{5}\int_8^\infty e^{-t/5}\, dt = \frac{1}{5} e^{-t/5} x = e^{-8/5} \simeq 0.2$$

$$P(X \geq 2) = \sum_{X=2}^{5} b(X; N, P) = \sum_{X=2}^{5} b(x; 5, 0.2) =$$

$$P(X = x) = \sum_{X=0}^{r} b(x; N, P)$$ BINOMIAL PROBABILITY SUMS

$$P(X \geq 2) = \sum_{X=2}^{5} b(x; 5, 0.2) = 1 - \sum_{X=0}^{1} b(x; 5, 0.2) = 1 - 0.7373 = 0.2627$$

**8.12** the tar contents of 8 brands of cigarettes selected at random from the latest list released by the federal trade are as follows:

7.3, 8.6, 10.4, 16.1, 12.2, 15.1, 14.5, and 9.3 milligrams. calculate

**a.** the mean

**b.** the variance

THE MEAN $\bar{x} = \sum_{i=0}^{n} x_i / 8$

$$\boxed{\bar{x} = \frac{7.3 + 8.6 + 10.4 + \ldots + 9.3}{8} = \frac{93.5}{8} = 11.6875}$$

THE VARIANCE $S^2$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

$$= \frac{((7.3 - 11.8675)^2 + (8.6 - 11.8675)^2 + (10.4 - 11.8675)^2 + (16.1 - 11.8675)^2 + (12.2 - 11.8675)^2 + (15.1 - 11.8675)^2 + (14.5 - 11.8675)^2 + (9.3 - 11.8675)^2}{7}$$

$\boxed{S^2 = 10.776}$ VARIANCE

# probability and statistics for engineers and scientists

morgan bergen

# final date may 09 2023

may 03 2023 - may 09 2023

**8.23** the random variable X, represents the number of cherries in a cherry puff, has the following probability distribution

| x | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| P(X = x) | 0.2 | 0.4 | 0.3 | 0.1 |

**a.** find the mean $\mu$ and variance $\sigma^2$ of X

**b.** find the mean $\mu_{\bar{X}}$ and variance $\sigma^2_{\bar{X}}$ of the mean $\bar{X}$ for random samples of 36 cherry puffs

**c.** find the probability that the average number of cherries in 36 cherry puffs will be less than 5.5

$\mu_x = 4 \cdot P(X=4) + 5 \cdot P(X=5) + 6 \cdot P(X=6) + 7 \cdot P(X=7) = 4 \cdot 0.2 + 5 \cdot 0.4 + 6 \cdot 0.3 + 7 \cdot 0.1 = 5.3$

$\sigma_x = (4-\mu_x)^2 \cdot P(X=4) + (5-\mu_x)^2 \cdot P(X=5) + (6-\mu_x)^2 \cdot P(X=6) + (7-\mu_x)^2 \cdot P(X=7) = (4-5.3)^2 \cdot 0.2 + \dots + (7-5.3)^2 \cdot 0.1 = 0.81$

THE MEAN $\mu_{\bar{x}}$ & VARIANCE $\sigma_{\bar{x}}$

$$\boxed{\begin{aligned} &\mu_{\bar{x}} = \mu_x = 5.3 \\ &\sigma^2_{\bar{x}} = \sigma^2_x / N = 0.81/36 = 0.0225 \end{aligned}}$$

FIND THE PROBABILITY THAT AVG IS LESS THAN 5.5

$$Z = \frac{\bar{X} - \mu_x}{\sigma_x / \sqrt{N}} = \frac{\bar{x} - \mu_x}{\sqrt{\sigma_{\bar{x}}}}$$

$$\boxed{P(\bar{x} < 5.5) = P(Z < \bar{x} - \mu / \sigma_{\bar{x}}) = P(Z < 5.5 - 5.3 / \sqrt{0.0225}) = P(Z < 1.33) = 0.9082}$$

8.27 in a chemical process, the amount of a certain type of impurity in the output is difficult to control and is thus a random variable.

speculation is that the **population mean amount of the impurity is 0.20** gram per gram of output. it is known that the **standard deviation is 0.1** gram per gram.

an experiment is conducted to gain more insight regarding the speculation that **μ = 0.2**. the process is run on a **lab scale 50 times** and the **sample average $\bar{x}$** turns out to be 0.23 gram per gram. comment on the speculation that the **mean amount of impurity is 0.20 gram per gram**. make use of the central limit theorem in your work.

$\mu_x = 0.20$ POPULATION MEAN

$\sigma_x = 0.1$ STANDARD DEVIATION POPULATION

$N = 50$ SAMPLE SIZE

$\bar{x} = 0.23$ SAMPLING MEAN

$X$ IS THE RANDOM VARIABLE FOR IMPURITY IN GRAM PER GRAM OUTPUT

WE WANT TO VERIFY IF $\mu_x = 0.20$ OR NOT

SINCE SAMPLING AVG IS $\bar{x} = 0.23$ WE WILL DETERMINE WHETHER THE SAMPLING MEAN IS GREATER THAN OR EQUAL TO 0.23

SAMPLE MEAN & STANDARD DEVIATION

$\mu_{\bar{x}} = \mu_x = 0.20$

$\sigma_{\bar{x}} = \sigma_x / \sqrt{N} = 0.1/\sqrt{50} = 0.01414$

$$Z = \frac{\bar{X} - \mu_x}{\sigma_x / \sqrt{N}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$P(\bar{x} \geq 0.23) = P\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \leq Z\right) = P\left(\frac{0.23 - 0.20}{0.01414} \leq Z\right) = P(Z \geq 2.122) = 1 - P(Z < 2.122) = 0.0169$$

$P(\bar{x} \geq 0.23) = 1.69\%$ CHANCE & IMPURITY ISNT 0.20

**8.29** the distribution of heights of a certain breed of terrier has a mean of 72 centimeters and a standard deviation of 10 centimeters, whereas the distribution of heights of a certain breed of poodle has a mean of 28 centimeters with a standard deviation of 5 centimeters. assuming that the sample means can be measured to any degree of accuracy, find the probability that the sample mean for a random sample of heights of 64 terriers exceeds the sample mean for a random sample of heights of 100 poodles by at most 44.2 centimeters.

$X$ IS RANDOM VARIABLE FOR TERRIERS HEIGHT

$\mu_X = 72$ MEAN

$\sigma_X = 10$ STANDARD DEVIATION

$N_X = 64$ SAMPLE SIZE

$Y$ RANDOM VARIABLE FOR POODLES HEIGHT

$\mu_Y = 28$ MEAN

$\sigma_Y = 5$ STANDARD DEVIATION

$N_Y = 100$ SAMPLE SIZE

$\bar{X} - \bar{Y}$ TERRIERS EXCEED POODLES HEIGHT

$P(\bar{X} - \bar{Y} \leq 44.2)$ PROBABILITY OF $X$ EXCEEDING $Y$ BY 44.2 CM

$\bar{X} - \bar{Y} = Z \Rightarrow P(Z \leq 44.2)$

$$Z = \frac{\bar{X} - \bar{Y} - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{44.2 - \mu_{\bar{X}-\bar{Y}}}{\sigma_{\bar{X}-\bar{Y}}} = \frac{44.2 - 44}{1.3463} = 0.1486$$

$$\mu_{\bar{X}-\bar{Y}} = \mu_X - \mu_Y = 72 - 28 = 44$$

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}} = \sqrt{\frac{10^2}{64} + \frac{5^2}{100}} = 1.3463$$

$$P(Z \leq 0.1486) = 0.5591$$

$$\boxed{P(\bar{X} - \bar{Y} \leq 44.2) = 0.5591}$$

**8.48** a manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. to maintain this average, 16 batteries are tested each month. if the computed t-value falls between –t_0.025 and t_0.025, the firm is satisfied with its claim. what conclusion should the firm draw from a sample that has a mean of x = 27.5 hours and a standard deviation of s = 5 hours? assume the distribution of battery lives to be approximately normal.

$\mu = 30$, $N = 16$

$T = \dfrac{\bar{X} - \mu}{S/\sqrt{N}}$ HAS A T DISTRIBUTION

$V = N - 1 = 15$ DEGREES OF FREEDOM

$t = \dfrac{\bar{x} - \mu}{s/\sqrt{N}} = \dfrac{27.5 - 30}{5/\sqrt{16}} = -2$

$t_{0.025} = 2.131 \Rightarrow -t_{0.025} = -2.131$

THE FIRM IS SATISFIED

# probability and statistics for engineers and scientists

morgan bergen

## final date may 09 2023

may 03 2023 - may 09 2023

**9.4** the heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters.

**a.** construct a 98% confidence interval for the mean height of all college students.

**b.** what can we assert with 98% confidence about the possible size of our error if we estimate the mean height of all college students to be 174.5 centimeters?

$N = 30$ STUDENTS

$\bar{x} = 174.5$

$\sigma_x^2 = 6.9$

$\alpha = 0.02$

$98\% \Rightarrow 0.98 = (1-\alpha) \Rightarrow \alpha = 0.02$

IF $\bar{x}$ MEAN OF A RANDOM SAMPLE SIZE $N$ FROM A POPULATION WITH $\sigma^2$ VARIANCE, $100(1-\alpha)\%$ CONFIDENCE INTERVAL IS

$$\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{N}} < \mu < \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{N}}$$

WHERE $Z_{\alpha/2}$ Z VALUE AN AREA OF $\alpha/2$ TO THE RIGHT

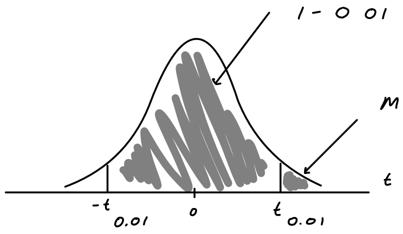
$$174.5 - Z_{0.02/2}\frac{6.9}{\sqrt{50}} < \mu < 174.5 + Z_{0.02/2}\frac{6.9}{\sqrt{50}} \Rightarrow \boxed{172.2 < \mu < 176.8}$$

CONFIDENCE INTERVAL

$Z_{0.02/2} = Z_{0.01} = 2.33$

$1 - 0.01 = 0.99$

MUST FIND $Z_{0.01}$ IN TABLE




$$Z_{\alpha/2}\frac{\sigma}{\sqrt{N}} = Z_{0.01}\frac{6.9}{\sqrt{50}} = 2.33 \cdot 0.976 = 2.274$$

WE CAN ASSERT WITH 98% CONFIDENCE NOT EXCEED 2.274cm

## probability and statistics for engineers and scientists

morgan bergen

## final date may 09 2023

may 03 2023 - may 09 2023

**9.10** a random sample of 12 graduates of a certain secretarial school typed an average of 79.3 words per minute with a standard deviation of 7.8 words per minute.

assuming a normal distribution for the number of words typed per minute,

find a 95% confidence interval for the average number of words typed by all graduates of this school.

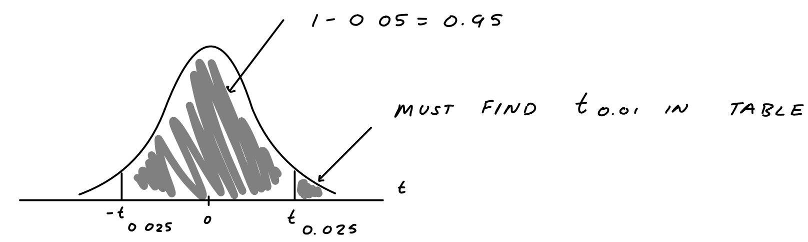
$N = 12$ , $\mu = 79.3$ , $\sigma^2 = 7.8$

$v = 12 - 1$ , $v = 11$ DEGREES OF FREEDOM

$95\% = (1 - \alpha) 100\%$

$0.95 = 1 - \alpha$

$\alpha = 0.05$




$$\bar{x} - t_{\alpha/2} \frac{\sigma^2}{\sqrt{N}} \leq \mu \leq \bar{x} - t_{\alpha/2} \frac{\sigma^2}{\sqrt{N}}$$

$t_{0.05/2} = t_{0.025} = t_{0.025}$ & $v = 11$ DEGREES OF FREEDOM

$t_{0.025} = 2.201$

$$79.3 - 2.201 \cdot \frac{7.8}{\sqrt{12}} \leq \mu \leq 79.3 + 2.201 \cdot \frac{7.8}{\sqrt{12}}$$

$$\boxed{74.34 < \mu < 84.26}$$

**9.14** the following measurements were recorded for the drying time, in hours, of a certain brand of latex paint:

| | | | | |
|---|---|---|---|---|
| 3.4 | 2.5 | 4.8 | 2.9 | 3.6 |
| 2.8 | 3.3 | 5.6 | 3.7 | 2.8 |
| 4.4 | 4.0 | 5.2 | 3.0 | 4.8 |

assuming that the measurements represent a random sample from a normal population,

find a 95% prediction interval for the drying time for the next trial of the paint.

$N = 15$

$$\bar{x} = \frac{1}{N}\sum_{i=0}^{N} x_i = \frac{3.4 + 2.5 + 4.8 + \dots + 4.8}{15} = 3.8$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2 = 0.9427$$

$$s = \sqrt{0.9427} \simeq 0.97$$

$$0.95 = 1 - \alpha \Rightarrow \alpha = 0.05$$

$v = 15 - 1 = 14$ DEGREES OF FREEDOM

$$\bar{x} - t_{\alpha/2} \cdot s\sqrt{1 + 1/N} < x_0 < \bar{x} + t_{\alpha/2} \cdot s\sqrt{1 + 1/N}$$

$$3.8 - t_{0.025} \cdot s\sqrt{1 + 1/15} < x_0 < 3.8 + t_{0.025} \cdot s\sqrt{1 + 1/15}$$

$$3.8 - 2.145 \cdot 0.97\sqrt{1 + 1/15} < x_0 < 3.8 + 2.145 \cdot 0.97\sqrt{1 + 1/15}$$

$$1.65 < x_0 < 5.95$$

**9.38** two catalysts in a batch chemical process, are being compared for their effect on the output of the process reaction.
a sample of 12 batches was prepared using catalyst 1, and a sample of 10 batches was prepared using catalyst 2.
the 12 batches for which catalyst 1 was used in the reaction gave an average yield of 85 with a sample standard deviation of 4, and
the 10 batches for which catalyst 2 was used gave an average yield of 81 and a sample standard deviation of 5.
find a 90% confidence interval for the difference between the population means,
assuming that the populations are approximately normally distributed with equal variances.

$\bar{x}_1 = 85, \quad s_1 = 4, \quad n_1 = 12$

$\bar{x}_2 = 81, \quad s_2 = 5, \quad n_2 = 10$

90% CONFIDENCE INTERVAL FOR MEANS

$\alpha = 0.1$

FROM THE THEOREM FOR CONFIDENCE INTERVAL FOR $\mu_1 - \mu_2$ WITH UNKNOWN VARIANCE $\sigma_1$ & $\sigma_2$ WHERE $\sigma_1 = \sigma_2$ REQ

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

POOLED ESTIMATE FOR STANDARD DEVIATION $s_p$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(12-1)4^2 + (10-1)5^2}{12 + 10 - 2} = 20.05$$

$s_p = \sqrt{20.05} = 4.4777$

$$(85 - 81) - 1.725 \cdot 4.4777 \sqrt{\frac{1}{10} + \frac{1}{12}} < \mu_1 - \mu_2 < (85 - 81) + 1.725 \cdot 4.4777 \sqrt{\frac{1}{10} + \frac{1}{12}}$$

$t = 0.05 \qquad v = (n_1 + n_2) - 2 = 12 + 10 - 2 = 20$

$v = 20 \qquad t_{0.05} = 1.725$

$$\boxed{0.692 < \mu_1 - \mu_2 < 7.307}$$

**9.46** the following data represent the running times of films produced by two motion picture companies

| company | time(minutes) |
|---|---|
| I | 103 94 110 87 98 |
| II | 97 82 123 92 175 88 118 |

compute a 90% confidence interval for the difference between the average running times of films produced by the two companies.

assume that the running time differences are approximately normally distributed with unequal variances.

$$\bar{x}_1 = \frac{1}{N_1}\sum_{i=0}^{N} x_i = \frac{103+94+110+87+98}{5} \simeq 98.4$$

$$\bar{x}_2 = \frac{1}{N_2}\sum_{i=0}^{N} x_i = \frac{97+82+123+92+175+88+118}{7} \simeq 110.7$$

$$S_2 = \sqrt{\frac{1}{N_1-1}\sum_{i=0}^{N_1}(x_i-\bar{x}_1)} = \sqrt{\frac{1}{5}\left((103-98.4)^2+\dots+(98-98.4)^2\right)} = 8.73$$

$$S_2 = \sqrt{\frac{1}{N_2-1}\sum_{i=0}^{N_2}(x_i-\bar{x}_2)} = \sqrt{\frac{1}{6}\left((97-110.7)^2+\dots+(118-110.7)^2\right.} = 32.2$$

$N_1 = 5$
$N_2 = 7$
$\bar{X}_1 = 98.4$
$\bar{X}_2 = 110.71$
$S_1 = 8.73$
$S_2 = 32.2$
$\alpha = 0.1$
$\alpha/2 = 0.5$
$t_{0.05} = 1.894$
$v = 7$

$0.90 = 1-\alpha \Rightarrow \alpha = 0.1$ & $\alpha/2 = 0.5$

$$v = \frac{(S_1^2/N_1 + S_2^2/N_2)^2}{\left[(S_1^2/N_1)/(N_1-1)\right] - \left[(S_2^2/N_2)^2/(N_2-1)\right]} = \frac{\left((8.73^2/5)+(32.2^2/7)\right)^2}{\left[(8.73^2/5)^2\right]+\left[(32.2^2/7)^2\right]} \simeq 7$$

$$(\bar{x}_1-\bar{x}_2) - t_{\alpha/2}\cdot\sqrt{\frac{S_1^2}{N_1}+\frac{S_2^2}{N_1}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t_{\alpha/2}\sqrt{\frac{S_1^2}{N_1}+\frac{S_2^2}{N_2}}$$

$$(98.4-110.71) - 1.894\sqrt{\frac{8.73^2}{5}+\frac{32.2^2}{7}} < \mu_1-\mu_2 < (98.4-110.71) + 1.894\sqrt{\frac{8.73^2}{5}+\frac{32.2^2}{7}}$$

$$-36.5 < \mu_1-\mu_2 < 11.9$$

# probability and statistics for engineers and scientists

morgan bergen

## final date may 09 2023

may 03 2023 - may 09 2023

**9.55** a new rocket launching system is being considered for deployment of small, short-range rockets.
the existing system has p = 0.8 as the probability of a successful launch.
a sample of 40 experimental launches is made with the new system, and 34 are successful.

a. construct a 95% confidence interval for p.

b. would you conclude that the new system is better?

N = 40 EXPERIMENTAL LAUNCHES

34/40 SUCCESSFUL LAUNCHES

$\hat{p} = Y/N = 0.85$ SUCCESS

$\hat{q} = 1 - 0.85 = 0.15$ FAILURE

0.95 CONFIDENCE INTERVAL

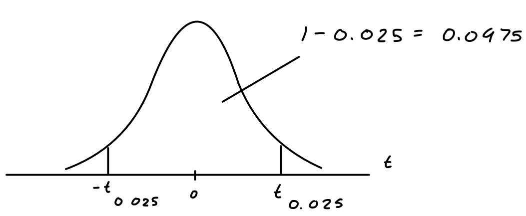
$0.95 = 1 - \alpha \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

THE PROPORTION OF SUCCESSES IN A RANDOM SAMPLE SIZE N & $\hat{q} = 1 - \hat{p}$

AN APPROXIMATE $100(1-\alpha)\%$ CONFIDENCE INTERVAL FOR THE BINOMIAL PARAMETER P IS

WHERE $Z_{\alpha/2}$ IS THE Z VALUE LEAVING AN AREA OF $\alpha/2$ TO THE RIGHT

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}} < p < \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}}$$




$Z_{0.025} = 1.96$

THE $Z_{0.025}$ IS THE Z VALUE LEAVING AN AREA OF 0.025

TO THE RIGHT / AREA OF 0.975 TO THE LEFT

USING NORMAL PROBABILITY TABLE WE SEE THE CLOSEST Z VALUE

WHICH IS LEAVING AN AREA OF 0.025 TO THE RIGHT, AN AREA

OF 0.075 TO THE LEFT IS $Z_{0.025} = 1.96$

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}} < p < \hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

$$0.85 - 1.96\sqrt{\frac{0.85 \cdot 0.15}{40}} < p < 0.85 - 1.96\sqrt{\frac{0.85 \cdot 0.15}{40}}$$

$0.7393 < p < 0.9607$

SINCE P = 0.8 & IS IN THE MIDDLE WE THUS CANT CONCLUDE THE NEW SYSTEM IS BETTER

**9.66** ten engineering schools in the united states were surveyed.

the sample contained 250 electrical engineers, 80 being women; 175 chemical engineers, 40 being women.

compute a 90% confidence interval for the difference between the proportions of women in these two fields of engineering.

is there a significant difference between the two proportions?

$n_1 = 250$ IF $\hat{p}_1$ & $\hat{p}_2$ ARE THE PROPORTIONS OF SUCCESSES IN RANDOM SAMPLES

$n_2 = 175$ OF SIZES $n_1$ & $n_2$ RESPECTIVELY $\hat{q}_1 = 1 - \hat{p}_1$ & $\hat{q}_2 = 1 - \hat{p}_2$ AN

$x_1 = 80$ APPROXIMATE $100(1-\alpha)\%$ CONFIDENCE INTERVAL FOR THE DIFFERENCES

$x_2 = 40$ OF THE 2 BINOMIAL PARAMETERS $p_1 - p_2$ ARE BELOW

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

WHERE $z_{\alpha/2}$ IS THE Z VALUE LEAVING AN AREA OF $\alpha/2$ TO THE RIGHT

$\hat{p}_1 = x_1/n_1 = 80/250 = 0.32$

$\hat{p}_2 = x_2/n_2 = 40/175 = 0.22857$

$\hat{q}_1 = 1 - \hat{p} = 1 - 0.32 = 0.68$

$\hat{q}_2 = 1 - \hat{q} = 1 - 0.23 = 0.77$

$0.90 = 1 - \alpha \Rightarrow \alpha = 0.10 \Rightarrow \alpha/2 = 0.05$

TABLE A.3 0.95

| z | .00 | ... | .04 |
|---|---|---|---|
| 0.0 | | | |
| ⋮ | | | ⋮ |
| 1.6 | ..... | | 0.9495 |

$z_{\alpha/2} = z_{0.05}$ IS THE Z VALUE LEAVING AN AREA OF 0.05 TO THE RIGHT

WE WANT TO FIND THE CLOSEST Z VALUE OF THE AREA 1-0.05, 0.95 TO THE LEFT

$z_{0.05} = 1.645$

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(0.32 - 0.23) - 1.645\sqrt{\frac{0.32 \cdot 0.68}{250} + \frac{0.2286 \cdot 0.7714}{175}} < p_1 - p_2 < (0.32 - 0.23) + 1.645\sqrt{\frac{0.32 \cdot 0.68}{250} + \frac{0.2286 \cdot 0.7714}{175}}$$

$$\boxed{0.0201 < p_1 - p_2 < 0.1627}$$

**10.23** add p-value approach

test the hypothesis that the average content of containers of a particular lubricant is 10 liters if the contents of a random sample of 10 containers are 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3, and 9.8 liters.

use a 0.01 level of significance and assume that the distribution of contents is normal.

$$\bar{x} = \frac{1}{N}\sum_{i=0}^{N} x_i = \frac{10.2 + 9.7 + 10.1 + 9.8 + 9.9 + 10.4 + 9.8}{10} = 10.06$$

$$s = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N}(x_i - \bar{x})^2} = \sqrt{\frac{(10.2 - 10.06)^2 + \dots + (9.8 - 10.6)^2}{10-1}} = 0.246$$

TEST THE HYPOTHESIS THAT THE AVG IS 10

$H_0 : \mu = 10$ NULL HYPOTHESIS

$H_1 : \mu \neq 10$ ALTERNATIVE HYPOTHESIS

$\alpha = 0.01$

TEST STATISTIC $t = \frac{\bar{x} - \mu_0}{S/\sqrt{N}} \sim t_{(N-1)}$

CRITICAL REGION $t > t_{\alpha/2,(N-1)}$ OR $t < -t_{\alpha/2}$

$$t = \frac{10.06 - 10}{0.246/\sqrt{10}} = 0.77$$

$t_{\alpha/2} = t_{0.01/2} = t_{0.05,9} = 3.250$

$t < 3.250$ WE DO NOT REJECT $H_0$

10.24 the average height of females in the freshman class of a certain college has historically been 162.5 centimeters with a standard deviation of 6.9 centimeters.
is there reason to believe that there has been a change in the average height
if a random sample of 50 females in the present freshman class has an average height of 165.2 centimeters?
use a P-value in your conclusion. assume the standard deviation remains the same.

$\mu_0 = 162.5$ $H_0: \mu = 162.5$ AVE HEIGHT HISTORICALLY

$\sigma = 6.9$ $H_1: \mu \neq 162.5$ IS THERE REASON TO BELIEVE THERE'S A CHANGE IN AVERAGE HEIGHT

$\alpha = 0.05$ $\bar{x} = 165.2$

$\alpha/2 = 0.025$ $N = 50$

$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{N}} = \dfrac{165 - 162.5}{6.9/\sqrt{50}} = 2.77$

$P(|Z| > 2.77) = 2P(Z < -2.77) = 2 \cdot 0.0028 = 0.0056$

$P(|Z| > 2.77) = 0.0056 < \alpha = 0.05$

SINCE P VALUE IS LESS THAN THE LEVEL OF SIGNIFICANCE WE REJECT $H_0$

& CONCLUDE THAT AVG HEIGHT IS NOT 162.5

**10.42** five samples of a ferrous-type substance were used to determine if there is a difference between a laboratory chemical analysis and an X-ray fluorescence analysis of the iron content.

each sample was split into two subsamples and the two types of analysis were applied. following are the coded data showing the iron content analysis:

assuming that the populations are normal, test at the 0.05 level of significance whether the two methods of analysis give, on the average, the same result.

| | sample | | | | |
|---|---|---|---|---|---|
| analysis | 1 | 2 | 3 | 4 | 5 |
| x-ray | 2.0 | 2.0 | 2.3 | 2.1 | 2.4 |
| chemical | 2.2 | 1.9 | 2.5 | 2.3 | 2.4 |

$X$ REPRESENTS RANDOM VARIABLE FOR XRAY

$Y$ REPRESENTS RANDOM VARIABLE FOR CHEMICAL

$N = 5$ SAMPLE SIZE

$\alpha = 0.05$ LEVEL OF SIGNIFICANCE

$H_0 : \mu_x - \mu_y = 0$ NULL HYPOTHESIS

$H_1 : \mu_x - \mu_y \neq 0$ ALTERNATIVE HYPOTHESIS

$d_i = x_i - y_i$ SAMPLE MEAN

$d_0 = \mu_x - \mu_y = 0$ DIFFERENCE BETWEEN POPULATIONS UNDER NULL HYPOTHESIS

$s_d^2 = \frac{1}{N-1} \sum_{i=0}^{N} (d_i - \bar{d})^2$ SAMPLE STANDARD DEVIATION OF THE DIFFERENCES BETWEEN THE PAIRED OBSERVATIONS

$\bar{d} = \frac{1}{N} \sum_{i=0}^{N} d_i$ SAMPLE MEAN

$d_1 = 2.0 - 2.2 = -0.2$ $\quad (d_1 - \bar{d})^2 = (0.2 - 0.1)^2 = 0.01$

$d_2 = 1.9 - 2.0 = -0.1$ $\quad (d_2 - d)^2 = (-0.1 - 0.1)^2 = 0.04$

$d_3 = 2.3 - 2.5 = -0.2$ $\quad (d_3 - d)^2 = (0.2 - 0.1)^2 = 0\ 01$

$d_4 = 2.1 - 2.3 = -0.2$ $\quad (d_4 - d)^2 = (0.2 - 0.1)^2 = 0.01$

$d_5 = 2.4 - 2.4 = 0$ $\quad (d_5 - d)^2 = (0 - 0.1)^2 = 0.01$

$$s_d^2 = \frac{1}{N-1} \sum_{i=0}^{N} (d_i - \bar{d})^2 = \frac{0.01 + 0.04 + 0.01 + 0.01 + 0.01}{4} = 0.02$$

$$\bar{d} = \frac{1}{N} \sum_{i=0}^{N} d_i = \frac{0.2 + -0.1 + 0.2 + 0.2 + 0}{5} = 0.1$$

$$s_d = \sqrt{s_d^2} = \sqrt{0.02} = 0.1414$$

TEST STATISTIC $t = \frac{\bar{d} - \bar{d}_0}{s_d / \sqrt{N}} = \frac{0.1 - 0}{0.1414/\sqrt{5}} \approx 1.58$

$t_{\alpha/2,(N-1)} = t_{0.05/2,(N-1)} = t_{0.025}$ WITH $v = 5 - 1 = 4$ DEGREES OF FREEDOM

$t_{0.025}$ , $\alpha = 0.025$ & $v = 4$ CAN BE FOUND IN TABLE A.4

$t_{0.025} = \pm 2.7764$

CRITICAL REGION IS THUS $(-\infty, -2.7764) \cup (2.7764, \infty)$

$-2.7764 < 1.58 < 2.7764$ $\quad H_0 : \mu_x - \mu_y = 0$ IS TRUE

10.59 A fuel oil company claims that one-fifth of the homes in a certain city are heated by oil.

Do we have reason to believe that fewer than one-fifth are heated by oil if, in a random sample of 1000 homes in this city,

136 are heated by oil? Use a P-value in your conclusion.

$N = 1000$ SAMPLE SIZE

$x = 136$ NUMBER OF HEATED HOMES

$\alpha = 0.05$ LEVEL OF SIGNIFICANCE

$H_0 : p = 1/5$ NULL HYPOTHESIS

"OIL COMPANY CLAIMS THAT 1/5 OF HOMES ARE HEATED"

$H_1 : p < 1/5$ ALTERNATIVE HYPOTHESIS

"FEWER THAN 1/5 OF HOMES ARE HEATED"

$\hat{p} = x/N$ SAMPLE PROPORTION

$\hat{p} = 136/1000$

$\hat{p} = 0.136$

TEST STATISTIC

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} = \frac{0.136 - 1/5}{\sqrt{\frac{1/5\,(1-1/5)}{1000}}} = \frac{-0.064}{0.0126} = -5.06$$

$P(Z < -5.06) = 0$

0 < 0.05, SO WE CAN REJECT THE NULL HYPOTHESIS

& CONCLUDE THAT $H_1$ IS CORRECT THAT FEWER THAN 1/5 ARE HEATED IN OIL

10.62 In a controlled laboratory experiment, scientists at the University of Minnesota discovered that 25% of a certain strain of rats subjected to a 20% coffee bean diet and then force-fed a powerful cancer-causing chemical later developed cancerous tumors. Would we have reason to believe that the proportion of rats developing tumors when subjected to this diet has increased if the experiment were repeated and 16 of 48 rats developed tumors? Use a 0.05 level of significance.

$N = 48$ SAMPLE SIZE

$H_0 : p = 0.25$ NULL HYPOTHESIS

$H_1 : p > 0.25$ ALTERNATIVE HYPOTHESIS

$\alpha = 0.05$ LEVEL OF SIGNIFICANCE

$x = 16$ TOTAL RATS WITH TUMOR

SAMPLE PROPORTION

$\hat{p} = x/n = 16/48 = 0.333$

TEST STATISTIC

$$Z = \frac{0.333 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{48}}} = 1.33 \qquad Z_{0.05} = 1.645$$

$\therefore$ THE # OF RATS THAT HAVE A TUMOR INCREASED