Inferential Statistics for Business
Step-by-Step Solutions

by **Jovan Samardzic**

KU Leuven • Bachelor of Business Administration (Brussels)

Hello, I am Jovan Samke!

Your go-to tutor for making studying easy!

+32 490 10 08 21

jovan@samkedemia.com

samkedemia

My Story

- I am 25-year old guy from Serbia, and I went to **the best math high school in Europe.**
- I did my BSc in Mathematics, had **a 93% GPA**, and stacked up a bunch of awards.
- I graduated with an MSc in Statistics and Data Science from KU Leuven — **as the only recipient of a full scholarship** — and earned **magna cum laude** honors.
- I have done hundreds of tutoring sessions, helping dozens of BBA students succeed, **nearly all of them passed their exams.**

Courses I Will Help You Pass

Math A | **Math B** | **Stats 1** | **Stats 2**

Why This Works for You?

Guaranteed Success

- Experienced tutor
- Top study resources
- Proven strategies

Designed for You

- Personalized support
- Clear and simple
- Straight to the point

Relax & Rise

- Friendly approach
- Build confidence
- Grow skills, not stress

Let’s maximize your results! DM me now!

Contents

Chapter 1 - Introduction

Chapter 1 - Question 1

Question

A portfolio consists of a fraction p of investments in real estate and a fraction $1-p$ of investments in stocks. The return on the investments in real estate is X and the return on the investments in stocks is Y. The return of the portfolio is $R = pX + (1-p)Y$.

Suppose that $\mu_X = 5$%, $\sigma_X = 2.8$%, $\mu_Y = 11.7$%, $\sigma_Y = 8.6$%, and $\rho_{X,Y} = -0.16$.

(a) Give an economic explanation for the negative correlation.

(b) Calculate the expectation and standard deviation of R in the case $p = 0.7$.

(c) Determine the value of p at which Var$[R]$ attains its minimum. (In mathematical finance, Var$[R]$ is a measure for the risk of the portfolio.)

Solution

Framework:

- **Random experiment:** Observe the annual return of a real estate investment and a stock investment.
- **Random variables:**
  - X: Return on the real estate investment.
  - Y: Return on the stock investment.
  - $R = pX + (1-p)Y$: Return on the combined portfolio, where p is the fraction invested in real estate.
- **Given:** The parameters of the return distributions (e.g., 5% = 0.05).
  - $\mu_X = 0.05$ and $\sigma_X = 0.028$
  - $\mu_Y = 0.117$ and $\sigma_Y = 0.086$

- Population correlation $\rho_{XY} = -0.16$

(a) Provide a brief economic explanation for the negative correlation.

Answer: When stock returns are high, investors may move money from real estate to stocks, causing real estate returns to fall slightly. This inverse relationship results in a negative correlation.

(b) Calculate the expectation and standard deviation of the portfolio return R for $p = 0.7$.

Tip : Mean and Variance of a Linear Combination

If $R = aX + bY$, where X and Y are two random variables, then the mean and variance of R are given by:

$$\mu_R = E[R] = a\mu_X + b\mu_Y$$
$$\sigma_R^2 = \text{Var}(R) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho_{XY}\sigma_X\sigma_Y$$

We use the formulas for a linear combination of random variables, with weights $a = p = 0.7$ and $b = 1 - p = 0.3$.

Expected Return (μ_R)

$$\mu_R = a\mu_X + b\mu_Y = (0.7 \cdot 0.05) + (0.3 \cdot 0.117) = 0.035 + 0.0351 = 0.0701$$

Standard Deviation (σ_R)

First, we calculate the variance, σ_R^2:

$$\begin{aligned}\sigma_R^2 &= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho_{XY}\sigma_X\sigma_Y \\ &= (0.7)^2(0.028)^2 + (0.3)^2(0.086)^2 + 2(0.7)(0.3)(-0.16)(0.028)(0.086) \\ &= 0.0008879824\end{aligned}$$

The standard deviation is the square root of the variance:

$$\sigma_R = \sqrt{0.0008879824} \approx 0.0298$$

Answer: The expected return is 0.0701 (or 7.01%), and the standard deviation is approximately 0.0298 (or 2.98%).

(c) Find the fraction p that minimizes the portfolio risk (variance).
The portfolio variance, $\text{Var}(R)$, as a function of p is:

$$\text{Var}(R) = p^2\sigma_X^2 + (1-p)^2\sigma_Y^2 + 2p(1-p)\rho_{XY}\sigma_X\sigma_Y$$

Substituting the given values and simplifying gives the quadratic function:

$$\text{Var}(R) = 0.00895056p^2 - 0.01556256p + 0.007396$$

Method 1: Vertex of a Parabola

This is a quadratic function of the form $Ap^2 + Bp + C$. Since $A > 0$, the parabola opens upwards and its minimum is at the vertex. The p-coordinate of the vertex is given by $p = -B/(2A)$.

$$p = \frac{-(-0.01556256)}{2 \cdot 0.00895056} = \frac{0.01556256}{0.01790112} \approx 0.8694$$

Method 2: Using Derivatives

To find the minimum, we take the first derivative of the variance with respect to p and set it to zero.

$$\frac{d}{dp}\text{Var}(R) = 2(0.00895056)p - 0.01556256 = 0.01790112p - 0.01556256$$

Setting the derivative to zero and solving for p:

$$0.01790112p = 0.01556256 \quad \Rightarrow \quad p = \frac{0.01556256}{0.01790112} \approx 0.8694$$

The second derivative is positive (0.01790112), confirming this value of p corresponds to a minimum.

Answer: The portfolio risk is minimized when approximately 86.94% ($p \approx 0.8694$) of the portfolio is invested in real estate.

Chapter 1 - Question 2

Question

Let X be the taxable income of the wife in a randomly selected heterosexual married couple and Y the taxable income of the husband in the same married couple. Assume that (X, Y) has a bivariate normal distribution with

$$\mu_X = €1734, \quad \sigma_X = €366, \quad \mu_Y = €1921, \quad \sigma_Y = €453, \quad \rho_{X,Y} = 0.45.$$

(a) What is the probability distribution of the total taxable income of a randomly chosen married couple?

(b) Women have to pay a tax of 30% and men a tax of 40%. What is the probability that the net income of the wife is larger than the net income of the husband in a randomly selected married couple?

Solution

Framework:

- **Random experiment:** Pick a random heterosexual married couple and record the taxable incomes of the wife and husband.
- **Random variables:**
  - $X =$ taxable income of the wife (in €)
  - $Y =$ taxable income of the husband (in €)
- **Given:** The income variables follow a bivariate normal distribution with the following population parameters:
  - $\mu_X = 1734$, $\sigma_X = 366$
  - $\mu_Y = 1921$, $\sigma_Y = 453$
  - Population correlation $\rho_{XY} = 0.45$

Tip : Bivariate Normal Distribution

A **bivariate normal distribution** is a model for two random variables, X and Y, that are each normally distributed and are linearly related to each other. Their relationship is described by the correlation coefficient, ρ_{XY}.

(a) What is the probability distribution of the total taxable income of a randomly chosen married couple?

We define the total taxable income as the random variable $T = X + Y$. Since X and Y are from a bivariate normal distribution, their sum T is also normally distributed. We just need to find its mean and standard deviation.

Mean of the total income (μ_T) is:

$$\mu_T = \mu_X + \mu_Y = 1734 + 1921 = 3655$$

Variance of the total income (σ_T) is:

$$\sigma_T^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y = 366^2 + 453^2 + 2(0.45)(366)(453) = 488383.2$$

The standard deviation is the square root of the variance:

$$\sigma_T = \sqrt{488383.2} \approx 698.84$$

Answer: The total taxable income T follows a normal distribution $T \sim N(\mu = 3655, \sigma = 698.84)$.

(b) What is the probability that the net income of the wife is larger than the net income of the husband?

Tip : Distribution of a Linear Combination of Normal Variables

If X and Y follow a bivariate normal distribution, then any linear combination is also normally distributed:

$$aX + bY + c \sim N\left(a\mu_X + b\mu_Y + c, \quad \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab \cdot \sigma_X \cdot \sigma_Y \cdot \rho_{XY}}\right)$$

The wife pays 30% tax, so her net income is $X_{net} = 0.7X$. The husband pays 40% tax, so his net income is $Y_{net} = 0.6Y$. We want to find the probability that the wife's net income is greater than the husband's, which can be written as:

$$P(0.7X > 0.6Y) \quad \Rightarrow \quad P(0.7X - 0.6Y > 0)$$

To solve this, we define a new random variable D representing the difference in net incomes, $D = 0.7X - 0.6Y$. Since D is a linear combination of X and Y, it will also be normally distributed. We now find its parameters.

The mean of this difference, μ_D, is:

$$\mu_D = 0.7\mu_X - 0.6\mu_Y = 0.7(1734) - 0.6(1921) = 61.2$$

The variance of the difference, σ_D^2, is:

$$\begin{aligned}\sigma_D^2 &= (0.7)^2\sigma_X^2 + (-0.6)^2\sigma_Y^2 + 2(0.7)(-0.6)\sigma_X\sigma_Y\rho_{XY} \\ &= (0.49)(366^2) + (0.36)(453^2) - 0.84(0.45)(366)(453) = 76842.04\end{aligned}$$

The standard deviation, σ_D, is the square root of the variance:

$$\sigma_D = \sqrt{76842.04} \approx 277.20$$

So, the difference in net incomes follows the distribution $D \sim N(61.2,\ 277.20)$. We can now calculate the probability $P(D > 0)$ using a TI-84 calculator:

$$P(D > 0) = \texttt{normalcdf(0, 1E99, 61.2, 277.20)} \approx 0.4126$$

Answer: The probability that the wife's net income is larger than the husband's is approximately 0.4126.

Chapter 1 - Question 3

Question

The weights of butter dishes that leave some factory are normally distributed with mean 250 g and standard deviation 4 g. Let X and Y measure the weights of two independently and randomly chosen butter dishes from the production.

(a) What is the value of $P(X > Y)$?

(b) Calculate $P(|X - Y| > 4)$.

(c) Determine $P(\max(X, Y) < 250)$.

Solution

Framework:

- **Random experiment:** Pick two random butter dishes from the factory and record their weights.

- **Random variables:**
  - X = Weight of the first butter dish (in g)
  - Y = Weight of the second butter dish (in g)
- **Given:**
  - The distributions are $X \sim N(\mu = 250, \sigma = 4)$ and $Y \sim N(\mu = 250, \sigma = 4)$.
  - The two random variables X and Y are independent.

Tip : Linear Combination of Independent Normal Variables

If $X \sim N(\mu_X, \sigma_X)$ and $Y \sim N(\mu_Y, \sigma_Y)$ are independent, then any linear combination is also normally distributed:

$$aX + bY + c \sim N\left(a\mu_X + b\mu_Y + c,\ \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}\right)$$

(a) What is the value of $P(X > Y)$?

To evaluate this probability, we can rearrange the inequality to be $P(X - Y > 0)$. This allows us to define a new random variable for the difference, $D = X - Y$. Since D is a linear combination of two independent normal variables, it is also normally distributed.

The mean of this difference, μ_D, is:

$$\mu_D = \mu_X - \mu_Y = 250 - 250 = 0$$

The variance of the difference, σ_D^2, is found by adding the individual variances (since they are independent):

$$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 = 4^2 + 4^2 = 16 + 16 = 32$$

The standard deviation, σ_D, is the square root of the variance:

$$\sigma_D = \sqrt{32} \approx 5.657$$

So, the difference follows the distribution $D \sim N(0,\ \sqrt{32})$. Since this distribution is symmetric around a mean of 0, the probability of obtaining a value greater than

0 is exactly half.

$$P(X > Y) = P(D > 0) = \texttt{normalcdf(0, 1E99, 0, } \sqrt{32}\texttt{)} = 0.5$$

Answer: The probability that the weight of the first dish is greater than the second is 0.5.

(b) Calculate $P(|X - Y| > 4)$.
We find this probability using the complement rule. First, we define the difference as the random variable $D = X - Y \sim N(0, \sqrt{32})$. Then:

$$\begin{aligned} P(|X - Y| > 4) = P(|D| > 4) = 1 - P(|D| < 4) = 1 - P(-4 \leq D \leq 4) \\ = 1 - \texttt{normalcdf(-4, 4, 0,} \sqrt{32}\texttt{)} \\ \approx 1 - 0.5205 = 0.4795 \end{aligned}$$

Answer: The probability that the absolute difference in weights is greater than 4 g is approximately 0.4795.

(c) Determine $P(\max(X, Y) < 250)$.

Tip : Probability of a Maximum

The statement "the maximum of two values is less than a number" is a concise way of saying that **both** of the individual values must be less than that number.

$$P(\max(X, Y) < a) = P(X < a \text{ and } Y < a)$$

Furthermore, if X and Y are independent events, this simplifies to the product of their individual probabilities:

$$P(\max(X, Y) < a) = P(X < a) \cdot P(Y < a)$$

Since the choices of butter dishes are independent, we can multiply their probabilities:

$$P(\max(X, Y) < 250) = P(X < 250 \text{ and } Y < 250) = P(X < 250) \cdot P(Y < 250)$$

First, we find the probability for a single dish, $P(X < 250)$. Since the distribution $X \sim N(250,\ 4)$ is symmetric around its mean of 250, the probability of being less than the mean is exactly 0.5.

$$P(X < 250) = 0.5$$

Now we can calculate the final probability:

$$P(\max(X, Y) < 250) = 0.5 \cdot 0.5 = 0.25$$

Answer: The probability that the maximum weight of the two dishes is less than 250 g is 0.25.

Chapter 1 - Question 4

Question

The width W (in meters) and length L (in meters) of a rectangular piece of a building site are modeled using a bivariate normal distribution with the following parameters:

- Mean and standard deviation of the width: $\mu_W = 30$, $\sigma_W = 2$
- Mean and standard deviation of the length: $\mu_L = 50$, $\sigma_L = 3$
- The correlation between width and length is denoted by ρ_{WL}.

The perimeter of the building site is given by the formula $P = 2(W + L)$.

(a) Calculate $P(P \leq 165)$ in the case that the correlation is $\rho_{WL} = 0$.

(b) Calculate $P(P \leq 165)$ in the case that the correlation is $\rho_{WL} = 0.5$.

(c) Calculate $P(P \leq 165)$ in the case that the correlation is $\rho_{WL} = -0.5$.

Solution

Framework:

- **Random experiment:** Pick a random rectangular piece of a building site and record its width and length.

- **Random variables:**
  - $W =$ Width of the site (in m)
  - $L =$ Length of the site (in m)
- **Given:**
  - $W \sim N(\mu_W = 30, \sigma_W = 2)$ and $L \sim N(\mu_L = 50, \sigma_L = 3)$.
  - W and L follow a bivariate normal distribution.
  - The perimeter is the random variable $P = 2(W + L) = 2W + 2L$.

(a) Calculate $P(P \leq 165)$ in case $\rho_{WL} = 0$.
Since W and L are normally distributed, their linear combination $P = 2W + 2L$ is also normally distributed. We need to find its mean and standard deviation.
The mean of the perimeter, μ_P, is:

$$\mu_P = 2\mu_W + 2\mu_L = 2(30) + 2(50) = 160$$

The variance of the perimeter, σ_P^2, when $\rho_{WL} = 0$ is:

$$\sigma_P^2 = 2^2\sigma_W^2 + 2^2\sigma_L^2 = 4(2^2) + 4(3^2) = 16 + 36 = 52$$

The standard deviation, σ_P, is therefore $\sqrt{52}$.

So, the perimeter follows the distribution $P \sim N(160, \sqrt{52})$. We can now find the required probability.

$$P(P \leq 165) = \texttt{normalcdf}(-1\text{E}99, 165, 160, \sqrt{52}) \approx 0.7560$$

Answer: The probability that the perimeter is less than or equal to 165 m is approximately 0.7560.

(b) Calculate $P(P \leq 165)$ in case $\rho_{WL} = 0.5$.
The mean of the perimeter remains $\mu_P = 160$. However, the variance now includes the covariance term.
The variance of the perimeter, σ_P^2, when $\rho_{WL} = 0.5$ is:

$$\begin{aligned}\sigma_P^2 &= 2^2\sigma_W^2 + 2^2\sigma_L^2 + 2(2)(2)\sigma_W\sigma_L\rho_{WL} \\ &= 4(2^2) + 4(3^2) + 8(0.5)(2)(3) = 16 + 36 + 24 = 76\end{aligned}$$

The standard deviation is $\sigma_P = \sqrt{76}$. The distribution is $P \sim N(160, \sqrt{76})$.

$$P(P \leq 165) = \texttt{normalcdf}(-1E99, 165, 160, \sqrt{76}) \approx 0.7169$$

Answer: The probability that the perimeter is less than or equal to 165 m is approximately 0.7169.

(c) Calculate $P(P \leq 165)$ in case $\rho_{WL} = -0.5$.
The mean of the perimeter is still $\mu_P = 160$.
The variance of the perimeter, σ_P^2, when $\rho_{WL} = -0.5$ is:

$$\begin{aligned}\sigma_P^2 &= 2^2\sigma_W^2 + 2^2\sigma_L^2 + 2(2)(2)\sigma_W\sigma_L\rho_{WL} \\ &= 4(2^2) + 4(3^2) + 8(-0.5)(2)(3) = 16 + 36 - 24 = 28\end{aligned}$$

The standard deviation is $\sigma_P = \sqrt{28}$. The distribution is $P \sim N(160, \sqrt{28})$.

$$P(P \leq 165) = \texttt{normalcdf(1E99, 165, 160,}\sqrt{28}\texttt{)} \approx 0.8276$$

Answer: The probability that the perimeter is less than or equal to 165 m is approximately 0.8276.

Chapter 1 - Question 5

Question

In an elevator, there is a warning: "maximum 500 kg or 6 persons." Suppose that the body weights are normally distributed with mean 75 kg and standard deviation 15 kg. The body weight of a random person can be considered as independent of the body weights of other randomly chosen people.
What is the probability that 6 randomly chosen persons weigh more than 500 kg?

Solution

Framework:

- **Random experiment:** Pick 6 random people and record their body weights.

- **Random variables:**
  - X_i = The body weight of the i-th person (in kg), for $i = 1, \ldots, 6$.
  - $S_6 = X_1 + X_2 + \cdots + X_6$ = The total body weight of 6 people (in kg).
- **Given:**
  - The distribution for a single person's weight is $X_i \sim N(\mu = 75, \sigma = 15)$.
  - The weights X_i are independent.
  - The sample size is $n = 6$.

Tip : Sum of Independent Normal Random Variables

If $X_1, \ldots, X_n$ are independent random variables and each one follows a normal distribution $X_i \sim N(\mu,\ \sigma)$, then their sum, $S_n = X_1 + \cdots + X_n$, also follows an exact normal distribution:

$$S_n \sim N\left(n \cdot \mu,\ \sqrt{n} \cdot \sigma\right)$$

We need to find the probability that the total weight of 6 people, S_6, is greater than 500 kg. Since each person's weight is an independent normal variable, their sum is also normally distributed. We find the parameters of its distribution as follows:

$$S_6 \sim N(n \cdot \mu,\ \sqrt{n} \cdot \sigma) = N(6 \cdot 75,\ \sqrt{6} \cdot 15) = N(450,\ \sqrt{1350})$$

We can now find the probability using a TI-84 calculator:

$$P(S_6 > 500) = \texttt{normalcdf(500, 1E99, 450, }\sqrt{\texttt{1350}}\texttt{)} \approx 0.0868$$

Answer: The probability that 6 randomly chosen people weigh more than 500 kg is approximately 0.0868.

Chapter 1 - Question 6

Question

In a barbershop, the durations of the haircuts of men are normally distributed with mean 21 minutes and standard deviation 6 minutes, and the durations of the haircuts of women are normally distributed with mean 48 minutes and standard deviation 12 minutes. The durations of different haircuts can be assumed to be independent.

(a) There are 3 male customers in this barbershop. What is the probability that the total duration of their haircuts exceeds 1.5 hours?

(b) What is the probability that the average duration of the haircuts of 3 men is larger than 30 minutes?

(c) What is the probability that the total duration of the haircuts of 5 men is smaller than the total duration of the haircuts of 2 women?

Solution

Framework:

- **Random experiment:** Pick random male and female customers and record their haircut durations.
- **Random variables:**
  - M = Duration of a haircut for a man (in minutes)
  - W = Duration of a haircut for a woman (in minutes)
- **Given:**
  - The distributions are $M \sim N(\mu_M = 21, \sigma_M = 6)$ and $W \sim N(\mu_W = 48, \sigma_W = 12)$.
  - All haircut durations are independent.

(a) There are 3 male customers in this barbershop. What is the probability that the total duration of their haircuts exceeds 1.5 hours?

Let S_M be the total time for 3 men's haircuts. Since the duration of each haircut is an independent normal variable, their sum S_M is also normally distributed. We find the parameters of its distribution as follows:

$$S_M \sim N(n \cdot \mu_M,\ \sqrt{n} \cdot \sigma_M) = N(3 \cdot 21,\ \sqrt{3} \cdot 6) = N(63,\ \sqrt{108})$$

We want to find the probability that this total exceeds 1.5 hours, which is 90 minutes:

$$P(S_M > 90) = \texttt{normalcdf(90, 1E99, 63, }\sqrt{\texttt{108}}\texttt{)} \approx 0.0047$$

Answer: The probability that the total haircut time of 3 men exceeds 1.5 hours is approximately 0.0047.

(b) What is the probability that the average duration of the haircuts of 3 men is larger than 30 minutes?

Tip : Average of Independent Normal Variables

If $X_1, \ldots, X_n$ are independent and each $X_i \sim N(\mu, \sigma)$, their average $\overline{X_n} = (X_1 + \cdots + X_n)/n$ is also exactly normally distributed:

$$\overline{X_n} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right)$$

Let $\overline{M_3}$ be the average time for 3 men's haircuts. Since the individual haircut times are normally distributed, their average is also normally distributed. We find the parameters of its distribution using the formula:

$$\overline{M_3} \sim N\left(\mu_M,\ \frac{\sigma_M}{\sqrt{n}}\right) = N\left(21,\ \frac{6}{\sqrt{3}}\right) = N(21,\ \sqrt{12})$$

Now we find the probability that this average exceeds 30 minutes.

$$P(\overline{M_3} > 30) = \texttt{normalcdf(30, 1E99, 21, }\sqrt{\texttt{12}}\texttt{)} \approx 0.0047$$

Answer: The probability that the average haircut time for 3 men exceeds 30 minutes is approximately 0.0047.

(c) What is the probability that the total duration of the haircuts of 5 men is smaller than the total duration of the haircuts of 2 women?

Let S_{M5} be the total time for 5 men, and S_{W2} be the total time for 2 women. Since individual haircut times are independent and normally distributed, their sums are also normally distributed.

First, we find the distribution for the sum of 5 men's haircuts:

$$S_{M5} \sim N(n_M \cdot \mu_M,\ \sqrt{n_M} \cdot \sigma_M) = N(5 \cdot 21,\ \sqrt{5} \cdot 6) = N(105,\ \sqrt{180})$$

Next, we find the distribution for the sum of 2 women's haircuts:

$$S_{W2} \sim N(n_W \cdot \mu_W,\ \sqrt{n_W} \cdot \sigma_W) = N(2 \cdot 48,\ \sqrt{2} \cdot 12) = N(96,\ \sqrt{288})$$

We want to find $P(S_{M5} < S_{W2})$, which is equivalent to finding the probability that their difference, $D = S_{M5} - S_{W2}$, is less than 0. The distribution of this difference is found as follows:

$$\begin{aligned} D = S_{M5} - S_{W2} \ &\sim\ N(\mu_{S_{M5}} - \mu_{S_{W2}},\ \sqrt{\sigma^2_{S_{M5}} + \sigma^2_{S_{W2}}}) \\ &=\ N(105 - 96,\ \sqrt{180 + 288}) = N(9,\ \sqrt{468}) \end{aligned}$$

We can now calculate the probability $P(D < 0)$ using a TI-84 calculator.

$$P(D < 0) = \texttt{normalcdf(-1E99, 0, 9, } \sqrt{468}\texttt{)} \approx 0.3387$$

Answer: The probability that the total haircut time of 5 men is smaller than that of 2 women is approximately 0.3387.

Chapter 1 - Question 7

Question

38% of the Belgian people have blood group O^+. Let S_{250} be the number of persons with blood group O^+ in a group of 250 randomly selected Belgian people. What is the probability that there are at least 90 and less than 100 persons with the blood group O^+ in a group of 250 randomly chosen Belgian people?

Give both an exact and an approximate calculation of this probability.

Solution

Framework:

- **Random experiment:** Pick 250 random Belgian people and record the number of individuals with blood group O^+.
- **Random variable:**
  - S_{250} = The number of people with blood group O^+ in a sample of 250.
- **Given:**
  - The population proportion of people with blood group O^+ is $\pi = 0.38$.
  - The sample size is $n = 250$.
  - The random variable follows a Binomial distribution, $S_{250} \sim \text{Bin}(n = 250, \pi = 0.38)$, as there is a fixed number of independent trials with two outcomes and a constant probability of success.

(a) Exact Calculation

Tip : Binomial Probability Calculation

To calculate the probability for a range of outcomes in a binomial distribution, such as $P(a \le X \le b)$, we use the cumulative distribution function (CDF). This is calculated as:

$$P(a \le X \le b) = P(X \le b) - P(X \le a - 1)$$

On a TI-84 calculator, $P(X \le k)$ is found using `binomcdf(n, p, k)`.

We need to find the probability $P(90 \le S_{250} < 100)$. Since the number of people must be an integer, this is equivalent to $P(90 \le S_{250} \le 99)$. Using the rule for calculating probabilities for a range:

$$\begin{aligned} P(90 \le S_{250} \le 99) &= P(S_{250} \le 99) - P(S_{250} \le 89) \\ &= \texttt{binomcdf(250, 0.38, 99)} - \texttt{binomcdf(250, 0.38, 89)} \\ &\approx 0.7770 - 0.2923 = 0.4847 \end{aligned}$$

Answer: The exact probability is approximately 0.4847.

(b) Approximate Calculation using the Normal Distribution

Tip : Normal Approximation to the Binomial

A **binomial distribution** can be approximated by a normal one when the sample size n is large ($n > 30$) and the distribution is not too skewed, which is checked by ($n\pi \geq 10$ and $n(1-\pi) \geq 10$).

The approximation is:

$$\text{Bin}(n, \pi) \;\approx\; N\left(n\pi,\; \sqrt{n\pi(1-\pi)}\right)$$

First, we check if the conditions for the normal approximation are met:

- $n = 250 > 30$
- $n\pi = 250 \cdot 0.38 = 95 \geq 10$
- $n(1-\pi) = 250 \cdot 0.62 = 155 \geq 10$

Since all conditions are met, we can approximate the Binomial distribution with a Normal distribution, which we'll call Y:

$$Y \approx N\left(n\pi,\; \sqrt{n\pi(1-\pi)}\right) = N(250 \cdot 0.38,\; \sqrt{250 \cdot 0.38 \cdot 0.62}) = N(95,\; \sqrt{58.9})$$

Tip : Applying the Continuity Correction

When approximating a discrete distribution with a continuous one, convert the discrete integer range into a continuous interval by expanding it by 0.5 on each side. This helps capture the full probability of the discrete values.

Let X be the discrete variable and Y be the continuous approximation. The three main conversions are:

- $P(X \leq k) \quad \Rightarrow \quad P(Y \leq k + 0.5)$
- $P(X \geq k) \quad \Rightarrow \quad P(Y \geq k - 0.5)$
- $P(a \leq X \leq b) \quad \Rightarrow \quad P(a - 0.5 \leq Y \leq b + 0.5)$

To calculate $P(90 \leq S_{250} \leq 99)$, we apply a continuity correction. The discrete integer range from 90 to 99 is represented by the continuous interval from 89.5 to

99.5.

$$P(89.5 \leq Y \leq 99.5) = \texttt{normalcdf(89.5, 99.5, 95, }\sqrt{\texttt{58.9}}\texttt{)} \approx 0.4844$$

Answer: The approximate probability using the normal distribution with continuity correction is 0.4844.

Chapter 1 - Question 8

Question

The daily demand for a product in a shop can be modelled with the probability distribution:

k	0	1	2	3	4	5
$P(D=k)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

The demand during a day is independent of the demands during other days.

(a) What is the probability that the total demand during 365 days is at least 950?

(b) How many days do we have to wait in order to have at least 95% chance of observing a total demand that is at least 1800 units?

(c) What is the probability that we have to observe more than 40 days until the fifth day without demand for this product?

Solution

Framework:

- **Random experiment:** Pick a random day and record the product demand.
- **Random variable:**
  - $D =$ The daily demand for a product (in units).
- **Given:** The probability distribution for the daily demand D:

k	0	1	2	3	4	5
$P(D=k)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

  The daily demands are independent.

From the distribution table, we can calculate the mean and standard deviation for a single day's demand.
The mean of the daily demand, μ, is:

$$\mu = E[D] = 0\left(\frac{1}{8}\right) + 1\left(\frac{1}{8}\right) + \cdots + 5\left(\frac{1}{8}\right) = \frac{20}{8} = 2.5$$

The variance of the daily demand, σ^2, is:

$$\sigma^2 = E[D^2] - \mu^2 \;=\; \left(0^2\left(\frac{1}{8}\right) + \cdots + 5^2\left(\frac{1}{8}\right)\right) - 2.5^2 = 2.25$$

The standard deviation, σ, is therefore $\sqrt{2.25} = 1.5$.

(a) What is the probability that the total demand during 365 days is at least 950?

Tip : Central Limit Theorem for Sums

The Central Limit Theorem (CLT) states that if $X_1, \ldots, X_n$ are independent random variables from **any** distribution with mean μ and standard deviation σ, and if the sample size is large enough ($n > 30$), then:

$$S_n = \sum_{i=1}^{n} X_i \approx N(n \cdot \mu,\ \sqrt{n} \cdot \sigma)$$

Let S_{365} be the total demand over 365 days. Since $n = 365$ is large, and we can assume that demands on different days are independent, we can apply the Central Limit Theorem:

$$S_{365} \approx N(n \cdot \mu,\ \sqrt{n} \cdot \sigma) = N(365 \cdot 2.5,\ \sqrt{365} \cdot 1.5) \approx N(912.5,\ 28.658)$$

Hence:

$$P(S_{365} \geq 950) = \texttt{normalcdf(950, 1E99, 912.5, 28.658)} \approx 0.0953$$

Answer: The probability that the total demand is at least 950 is approximately 0.0953.

(b) How many days do we have to wait in order to have at least 95% chance of observing a total demand that is at least 1800 units?

Tip : Finding Sample Size for a Target Probability

This type of problem asks for the minimum sample size n needed for a sum (S_n) or average ($\overline{X_n}$) to meet a certain probability threshold, like $P(S_n > k) \geq p$.

A practical way to solve this is to use a calculator's table feature. Define a function in the `Y=` editor that calculates the probability for a given n (using the calculator's `X` variable). Then, by exploring the table of values for `X` and `Y`, you can find the first integer n that satisfies the probability condition.

We need to find the smallest integer n that satisfies $P(S_n \geq 1800) \geq 0.95$, where the total demand is approximately $S_n \approx N(2.5n,\ 1.5\sqrt{n})$.

Since the daily demand can only take integer values $\{0, 1, 2, 3, 4, 5\}$, the total demand S_n is also a discrete variable. We therefore apply a continuity correction, approximating the discrete condition $S_n \geq 1800$ with the continuous interval from 1799.5 to infinity.

We enter the probability calculation into the `Y=` editor, using the calculator's `X` variable to represent the number of days, n:

```
Y1 = normalcdf(1799.5, 1E99, 2.5*X, 1.5*√(X))
```

By setting a starting point in `TBLSET` (e.g., `TblStart=745`) and viewing the `TABLE`, we can scroll to find the first integer `X` for which the corresponding Y_1 value exceeds 0.95. This process shows that $n = 747$ is the first integer number of days to meet the condition.

Answer: We have to wait at least 747 days.

Note: *In a real scenario, you wouldn't know to start the table at 745. A practical approach is to make an initial guess (e.g., n=700), check the resulting probability, and if it's too low, jump to a significantly larger n (e.g., n=750). This allows you to quickly narrow down the range where the solution lies.*

(c) What is the probability that we have to observe more than 40 days until the fifth day without demand for this product?

Tip : Negative Binomial Distribution

The **negative binomial distribution**, $NB(r, \pi)$, models the number of trials required to achieve a fixed number of successes, r.

This is a negative binomial problem where a "success" is a day with zero demand ($\pi = 1/8$) and we are waiting for the 5th success ($r = 5$). We want to find the probability of waiting more than 40 days, $P(Y > 40)$.

Method 1: Exact Calculation (via Binomial)

Tip : Negative Binomial to Binomial Link

Let $Y \sim NB(r, \pi)$ be the trial number of the r-th success, and let $X_n \sim \text{Bin}(n, \pi)$ be the number of successes in the first n trials.

The probability of needing **more than n** trials for the r-th success is equivalent to having **at most $r - 1$** successes in those first n trials:

$$P(Y > n) = P(X_n \leq r - 1)$$

Using this link, this problem is equivalent to finding the probability of 4 or fewer successes in the first 40 days for a variable $X_{40} \sim \text{Bin}(40, 1/8)$.

$$P(Y > 40) = P(X_{40} \leq 4) = \texttt{binomcdf(40, 1/8, 4)} \approx 0.4287$$

Method 2: Approximate Calculation (via Normal)

Alternatively, we can approximate the binomial probability. Since the conditions $n \geq 30$, $n\pi = 5$ and $n(1 - \pi) = 35$ are met, we can use a normal distribution as an approximation:

$$X_{40} \approx N(n\pi,\ \sqrt{n\pi(1-\pi)})\ =\ N(5,\ \sqrt{4.375})$$

Applying the continuity correction for $P(X_{40} \leq 4)$, we calculate $P(W \leq 4.5)$:

$$P(W \leq 4.5)\ =\ \texttt{normalcdf(-1E99, 4.5, 5, }\sqrt{\texttt{4.375}}\texttt{)} \approx 0.4055$$

Answer: The exact probability is approximately 0.4287. The normal approximation gives 0.4055.

Chapter 1 - Question 9

Question

The duration of a random pregnancy is a left-skewed continuous random variable with expectation 280 days and standard deviation 12 days. What is the probability that the average duration of 50 pregnancies is at most 275 days?

Solution

Framework:

- **Random experiment:** Pick 50 random pregnancies and record their durations.
- **Random variables:**
  - X = The duration of a random pregnancy (in days).
  - $\overline{X_{50}}$ = The average duration of 50 pregnancies (in days).
- **Given:**
  - The population mean is $\mu = 280$ days.
  - The population standard deviation is $\sigma = 12$ days.
  - The population distribution of X is left-skewed.
  - The sample size is $n = 50$.

Tip : Central Limit Theorem for the Sample Mean

The Central Limit Theorem (CLT) states that if $X_1, \ldots, X_n$ are independent random variables from **any** distribution with mean μ and standard deviation σ, and if the sample size is large enough ($n > 30$), then their average $\overline{X_n}$ is approximately normally distributed:

$$\overline{X_n} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Even though the underlying distribution of a single pregnancy's duration is left-skewed, we can apply the Central Limit Theorem because our sample size of $n = 50$ is large enough:

$$\overline{X_{50}} \approx N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(280, \frac{12}{\sqrt{50}}\right)$$

Now we can calculate the desired probability:

$$P(\overline{X_{50}} \leq 275) = \texttt{normalcdf(-1E99, 275, 280, 12/}\sqrt{\texttt{50}}\texttt{)} \approx 0.0016$$

Answer: The probability that the average duration of 50 pregnancies is at most 275 days is approximately 0.0016.

Chapter 1 - Question 10

Question

The body lengths of students are normally distributed with (population) mean 178.6 cm and (population) standard deviation 7.6 cm.

(a) What is the probability that a randomly chosen student is taller than 1.9 m? Interpret this probability in the context of the population.

(b) What is the probability that the average body length of two randomly chosen students is larger than 1.9 m?

(c) What is the probability that the average body length of ten randomly chosen students is larger than 1.9 m?

(d) Graphically display the probabilities that you have calculated in a-c.

Solution

Framework:

- **Random experiment:** Pick a random student and record their body length.
- **Random variable:**
  - $X =$ The body length of a random student (in cm).
- **Given:**
  - The population mean is $\mu = 178.6$ cm.
  - The population standard deviation is $\sigma = 7.6$ cm.
  - The population distribution is normal, so $X \sim N(178.6,\ 7.6)$.
  - The height of interest is 1.9 m, which we convert to cm: 1.9 m $= 190$ cm.

(a) What is the probability that a randomly chosen student is taller than 1.9 m? Interpret this probability.

We need to find the probability $P(X > 190)$ using the population distribution $X \sim N(178.6,\ 7.6)$.

$$P(X > 190) = \texttt{normalcdf(190, 1E99, 178.6, 7.6)} \approx 0.0668$$

Answer: The probability is approximately 0.0668. We expect about 6.68% of the student population to be taller than 1.9 m.

(b) What is the probability that the average body length of two randomly chosen students is larger than 1.9 m?

Tip : Distribution of the Sample Mean

If the population is normally distributed with $X \sim N(\mu, \sigma)$, then the sample mean $\overline{X_n}$ for a sample of size n is also **exactly** normally distributed:

$$\overline{X_n} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right)$$

First, we find the distribution of the sample mean $\overline{X_2}$.

$$\overline{X_2} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right) = N\left(178.6,\ \frac{7.6}{\sqrt{2}}\right)$$

Now we calculate the probability using this new distribution.

$$P(\overline{X_2} > 190) = \texttt{normalcdf(190, 1E99, 178.6, 7.6/}\sqrt{2}\texttt{)} \approx 0.01695$$

Answer: The probability that the average body length of two students is larger than 1.9 m is approximately 0.01695.

(c) What is the probability that the average body length of ten randomly chosen students is larger than 1.9 m?

The logic is the same as in part (b), but now with a sample size of $n = 10$. We first find the distribution of the sample mean $\overline{X_{10}}$.

$$\overline{X_{10}} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right) = N\left(178.6,\ \frac{7.6}{\sqrt{10}}\right)$$

Now we calculate the probability $P(\overline{X_{10}} > 190)$.

$$P(\overline{X_{10}} > 190) = \texttt{normalcdf(190, 1E99, 178.6, 7.6/}\sqrt{\texttt{10}}\texttt{)} \approx 0.0000001052$$

Answer: The probability that the average body length of ten students is larger than 1.9 m is extremely small, approximately 0.0000001052.

(d) Graphically display the probabilities that you have calculated in a-c.
To visualize the results, we plot the three probability density functions for the distributions corresponding to $n = 1, n = 2$, and $n = 10$.

- For $n = 1$, the distribution is $X \sim N(178.6,\ 7.6)$, entered in the TI-84 as `Y1 = normalpdf(X, 178.6, 7.6)`.
- For $n = 2$, the distribution is $\overline{X_2} \sim N(178.6,\ 7.6/\sqrt{2})$, entered as `Y2 = normalpdf(X, 178.6, 7.6/√2)`.
- For $n = 10$, the distribution is $\overline{X_{10}} \sim N(178.6,\ 7.6/\sqrt{10})$, entered as `Y3 = normalpdf(X, 178.6, 7.6/√10)`.

We set the viewing window. A good range is three standard deviations around the mean:

$$[\mu - 3\sigma,\ \mu + 3\sigma] = [178.6 - 3(7.6),\ 178.6 + 3(7.6)] = [155.8,\ 201.4]$$

Finally, we shade the area for $x > 190$ under each curve using the calculator's integral function, found by pressing `2nd` + `TRACE` $\rightarrow$ `7:∫f(x)dx`.

Answer: The graph shows that as the sample size n increases, the distribution of the sample mean becomes narrower, and the probability of an extreme average decreases.