STAT 230

July 25 Tutorial Solutions

Preview

In this tutorial, we will look at an example where covariance and correlation are calculated. We will interpret the result and comment on independence.

We will also do an example with indicator variables, where expectation and variance are calculated.

We will conclude the tutorial with two questions involving the Central Limit Theorem.

Problem 1

You are given the following joint probability function, f(x, y), for the random variables X and Y in the table below:

y

x

f(x,y)	0	1	2
1	0.2	0.1	0.1
2	0.1	0	0.1
3	0.1	0.2	0.1

a. Determine the correlation coefficient, ρ.

b. Are X and Y dependent or independent?

Problem 1 Solution

You are given the following joint probability function, f(x, y), for the random variables X and Y in the table below:

y

x

f(x,y)	0	1	2
1	0.2	0.1	0.1
2	0.1	0	0.1
3	0.1	0.2	0.1

a. Calculate the correlation coefficient, ρ and interpret its value.

Cov(X, Y) = E(XY) – E(X)*E(Y)
E(XY) = 0 + 0 + 0 + (1)(1)0.1 + 0 + (3)(1)0.2 + (1)(2)0.1 + (2)(2)0.1 + (3)(2)0.1 = 1.9
E(X) = 1*0.4 + 2*0.2 + 3*0.4 = 2
$E(X^2) = 1^2(0.4) + 2^2(0.2) + 3^2(0.4) = 4.8$
$Var(X) = 4.8 - (2)^2 = 0.8$
E(Y) = 0*0.4 + 1*0.3 + 2*0.3 = 0.9
$E(Y^2) = 0^2(0.4) + 1^2(0.3) + 2^2(0.3) = 1.5$
$Var(Y) = 1.5 - (0.9)^2 = 0.69$

So, Cov(X, Y) = 1.9 – (2)(0.9) = 0.1

The correlation coefficient, ρ, is given by: $\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)*Var(Y)}} = \frac{0.1}{\sqrt{0.8*0.69}} = 0.135$

This suggests that there is a very weak positive correlation between X and Y.

Problem 1 Solution (Cont'd)

You are given the following joint probability function, f(x, y), for the random variables X and Y in the table below:

y

x

f(x,y)	0	1	2
1	0.2	0.1	0.1
2	0.1	0	0.1
3	0.1	0.2	0.1

b. Are X and Y dependent or independent?

We see that Cov(X, Y) does not equal 0, so X and Y are not independent.

So, we can also say that X and Y are dependent.

Problem 2

Suppose that 6 players are seated at a round table. They are identified as Player 1, Player 2, …., Player 6. Each player will flip a fair coin. A player will score a point when their flip matches the flip of both the player on their left AND on their right. Let X represent the number of points scored during one round.

a. Calculate E(X).
b. Calculate Var(X)

Problem 2 Solution

Suppose that 6 players are seated at a round table. They are identified as Player 1, Player 2, …., Player 6. Each player will flip a fair coin. A player will score a point when their flip matches the flip of both the player on their left AND on their right. Let X represent the number of points scored during one round.

a. Calculate E(X)

Note to TAs:

It may be a good idea to draw a picture to help visualize this question. Draw a circle to represent the table, and then place the 6 players on the circle. Thus students can visualize them sitting at a round table.

We will use an indicator variable to solve for E(X) and Var(X).

To start, for i = 1, 2, …, 6.

Let X_i = $\begin{cases} \mathbf{1}, & \textbf{if player i scores a point} \\ \mathbf{0}, & \textbf{if player i does not score a point} \end{cases}$

Player i scores a point when they and the two adjacent players all flip heads or tails.

So, p = P(X_i = 1) = $\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^3 = \frac{1}{4}$. **Thus, E($X_i$) =** $1 * \left(\frac{1}{4}\right) + 0 * \left(\frac{3}{4}\right) = \frac{1}{4}$.

So, E(X) = $\sum_{i=1}^{6} E(X_i) = 6p = \frac{3}{2}$.

Problem 2 Solution (Cont'd)

Suppose that 6 players are seated at a round table. They are identified as Player 1, Player 2, …., Player 6. Each player will flip a fair coin. A player will score a point when their flip matches the flip of both the player on their left AND on their right. Let X represent the number of points scored during one round.

b. Calculate Var(X)

In part a., we found E(X_i) = p = $\frac{1}{4}$ and E(X) = 6p = $\frac{3}{2}$.

To find Var(X), we start by determining Var(X_i).

$\boldsymbol{E\left(X_i^2\right) = 0^2 * (1-p) + 1^2(p) = p = \frac{1}{4}}$

So, Var(X_i) = p - $p^2 = p(1-p) = \frac{3}{16}$

Now, to get Var(X), we will need to determine the covariance of all the X_i's. Let's first look at the scores of two adjacent players, say X_1 and X_2. We also need to consider the case where we have X_1 and X_6 (not "numerically adjacent", but they are sitting next to each other).

Problem 2 Solution (Cont'd)

In order to determine E(X_1X_2), we notice that X_1X_2 = 1 iff Players 6, 1, 2, and 3 either flip all heads or all tails. Otherwise, this will be 0.

(Take a minute to think about this).

E(X_1X_2) = P(X_1X_2 = 1) = P(all 4 flip heads or all 4 flip tails) = $\left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^4 = \frac{1}{8}$**.**

So, Cov(X_1, X_2) = E(X_1X_2) – E(X_1)E(X_2) = $\frac{1}{8} - \left(\frac{1}{4}\right)^2 = \frac{1}{16}$**.**

This result will be the same for all other "numerically adjacent" pairings, as well as Cov(X_1, X_6).

So, for i = 1, 2, 3, 4, 5.

Cov(X_i, X_{i+1}) = Cov$(X_1, X_6) = \frac{1}{16}$.

Problem 2 Solution (Cont'd)

Now, let's consider the scores of two players that are not sitting next to each other. For example, Players 1 and 3.

In this case, X_1X_3 = 1 iff Players 6, 1, 2, 3, and 4 either flip all heads or all tails. Otherwise, X_1X_3 = 0.

Note to TAs: In this case, Player 5's result does not affect the calculation.

So, $P(X_1X_3 = 1) = \left(\frac{1}{2}\right)^5 + \left(\frac{1}{2}\right)^5 = \frac{1}{16}$.

So, $Cov(X_1, X_3) = E(X_1X_3) - E(X_1)E(X_3) = \frac{1}{16} - \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = 0$.

By the same argument, the covariance of the scores of all situations where 2 players are not sitting next to each other is 0.

Problem 2 Solution (Cont'd)

Now, we can calculate Var(X).

Var(X) = $\sum_{i=1}^{6} Var(X_i) + 2\sum_{i=1}^{5}\sum_{j=i+1}^{6} Cov(X_i, X_j)$

Var(X) = $6 * \left(\frac{3}{16}\right) + 2 * \left(\sum_{i=1}^{5} \text{Cov}(X_i, X_{i+1}) + \text{Cov}(X_1, X_6)\right)$

Var(X) = $\frac{18}{16} + 2 * 6\left(\frac{1}{16}\right) = \frac{30}{16} = \frac{15}{8}$ **= 1.875**

Problem 3

Suppose that the weights of bananas (in kg) in your favourite grocery store follow a Uniform(0.1, 0.15) distribution.

a. What is the approximate probability that a bin of 100 bananas at this grocery store weighs between 12.35 and 12.65 kg?

b. Suppose that this grocery store is selling bananas in 20 kg boxes (for those real banana lovers!). What is the minimum number of bananas needed in each box so that the probability that the box has less than 20 kg of bananas is no more than 5%.

Problem 3 Solution

Suppose that the weights of bananas in your favourite grocery store (in kg) follow a Uniform(0.1, 0.15) distribution.

a. What is the approximate probability that a bin of 100 bananas at this grocery store weighs between 12.35 and 12.65 kg?

Let X_i represent the weight of the i^{th} banana.

X_i~Uniform(0.1, 0.15), for i = 1, 2, …, n; independent.

$E(X_i) = \mu = \frac{0.1+0.15}{2} = 0.125$ and $Var(X_i) = \sigma^2 = \frac{(0.15-0.1)^2}{12} = \frac{(0.05)^2}{12}$.

Let X represent the total weight of the 100 bananas.

So, $X = \sum_{i=1}^{100} X_i$.

Now, let $Y = \frac{X-100\mu}{\sqrt{100}\sigma} = \frac{X-12.5}{0.5/\sqrt{12}}$.

By the Central Limit Theorem (CLT), the cdf of Y is approximately the cdf of a standard Normal random variable (Z ~ N(0, 1)).

Problem 3 Solution (Cont'd)

So, P(12.35 < X < 12.65) = P(X < 12.65) – P(X < 12.35)

= $\boldsymbol{P(\frac{X-12.5}{0.5/\sqrt{12}} < \frac{12.65-12.5}{0.5/\sqrt{12}})}$ **- P**$\boldsymbol{(\frac{X-12.5}{0.5/\sqrt{12}} < \frac{12.35-12.5}{0.5/\sqrt{12}})}$

= P(Y < 1.04) – P(Y < -1.04)

We stated that the cdf of Y is approximately the cdf of Z ~ N(0, 1).

So, by the CLT, we approximate this probability using:

P(Z < 1.04) – P(Z < -1.04) = 2*P(Z < 1.04) – 1
= 2*(0.85083) – 1 = 0.70166

Problem 3 Solution (Cont'd)

Suppose that the weights of bananas in your favourite grocery store (in kg) follow a Uniform(0.1, 0.15) distribution.

b. Suppose that this grocery store is selling bananas in 20 kg boxes (for those real banana lovers!). What is the minimum number of bananas needed in each box so that the probability that the box has less than 20 kg of bananas is no more than 5%?

Let W represent the total weight of the n bananas in the box. W = $\sum_{i=1}^{n} X_i$.

Let's define V = $\frac{W-n\mu}{\sqrt{n}\sigma} = \frac{W-0.125n}{0.05\sqrt{n}/\sqrt{12}}$

By the CLT, the cdf of V is approximately the cdf of Z ~ N(0, 1).

We want to find the minimum value, n, such that P(W < 20) ≤ 0.05

$P\left(V < \frac{20-0.125n}{0.05\sqrt{n}/\sqrt{12}}\right) \approx P\left(Z < \frac{20-0.125n}{0.05\sqrt{n}/\sqrt{12}}\right)$ **(by the CLT)**

We want this probability to be no more than 0.05. As such, the corresponding quantile must be less than 0.

Problem 3 Solution (Cont'd)

Using symmetry, from the Normal quantiles table in the Course Notes, we see that z = -1.6449 in this case (as the 95th percentile of the distribution of Z is given by z = 1.6449).

So, $\frac{20-0.125n}{0.05\sqrt{n}/\sqrt{12}} \leq -1.6449$

=> $20\sqrt{12} - \frac{\sqrt{12}n}{8} \leq -0.082245\sqrt{n}$.

Now divide both sides by $\sqrt{12}$ **and let x =** $\sqrt{n}$.

Then organize the terms, and we have the following quadratic equation:

$$0.125x^2 - 0.0237421x - 20 \geq 0$$

Solving this quadratic, we get: $x \geq \frac{-(-0.237421)\pm\sqrt{(-0.0237421)^2-4(0.125)(-20)}}{2(0.125)}$.

We know that there are two roots, but the negative solution, $x \geq \frac{-(-0.237421)-\sqrt{(-0.0237421)^2-4(0.125)(-20)}}{2(0.125)}$ **is inadmissible.**

So, $x \geq \frac{-(-0.237421)+\sqrt{(-0.0237421)^2-4(0.125)(-20)}}{2(0.125)}$ **= 12.744436**

Remember that x = $\sqrt{n}$. **So, we have** $n \geq (12.744436)^2 = 162.421$

Therefore, a minimum of 163 bananas are needed.

Problem 4

Suppose that a fair coin is to be flipped 50 times.

a. What is the probability that exactly 25 heads will be flipped?
b. Approximate the probability in part a. using a Normal approximation.

Problem 4 Solution

Suppose that a fair coin is to be flipped 50 times.

a. What is the probability that exactly 25 heads will be flipped.

Let X represent the number of heads in 50 flips of a fair coin.

We know that X ~ Binomial(n = 50, p = 0.5)

P(X = x) = $\binom{50}{25}\left(\frac{1}{2}\right)^{50}$**, for x = 0, 1, …, 50; 0, otherwise.**

So, P(X = 25) = $\binom{50}{25}\left(\frac{1}{2}\right)^{50}$ **= 0.1122752**

Note: Using R: P(X = 25) = dbinom(25,50,0.5)

Problem 4 Solution (Cont'd)

Suppose that a fair coin is to be flipped 50 times.

b. Approximate the probability in part a. using a Normal approximation.

Let X represent the number of heads in 50 flips of a fair coin.

We know that X ~ Binomial(n = 50, p = 0.5)

Here, we will use the Normal Approximation to the Binomial and we'll see how well the approximation worked.

Note: We have to apply a continuity correction here. If you don't, your approximate probability will be 0!

So, we know that $P(X = 25)$ can also be written as $P(24 < X < 26)$.

So, applying the c.c., we have $P(24.5 < X < 25.5)$.

Problem 4 Solution (Cont'd)

As we saw in the previous problem, by the CLT, the cdf of this distribution will be approximately the cdf of a standard Normal distribution (Z ~ N(0,1)).

Here, we state that X is approximately Normal with a mean of np = 25 and a variance of np(1-p) = 50*0.5*0.5 = 12.5

So, the approximate probability is given by:

P(X < 25.5) – P(X < 24.5) = $P\left(\frac{X-np}{\sqrt{np(1-p)}} < \frac{25.5-25}{\sqrt{12.5}}\right) - P\left(\frac{X-np}{\sqrt{np(1-p)}} < \frac{24.5-25}{\sqrt{12.5}}\right)$ **= P(Z < 0.14) – P(Z < -0.14)**

= 2*P(Z < 0.14) – 1 = 2*(0.55567) – 1 = 0.11134

(Recall that the exact answer from part a. was 0.1122752).

The approximation worked quite well!

Note: When np and n(1-p) are both greater than 5, the Normal Approximation to the Binomial will work well.

This was the case here: np = 25 and n(1-p) = 25. Both are much greater than 5.