Institute and Faculty of Actuaries

CS1 Specimen Questions and Solutions

July 2020

Q1

A survey showed that 40% of investors invest in at least two companies in order to diversify their risk. Let *X* be the random variable denoting the number of investors who have invested in more than one company in a random sample of 300 investors.

(i) Determine the distribution of *X*. [2]

(ii) Calculate an approximate probability that *X* is greater than 100. [2]

[Total 4]

Solution

(i) *X follows a binomial (300, 0.4) distribution* *[0.5]*

So, approximately X follows a normal distribution *[0.5]*

*with E(X) = 300*0.4 = 120 and V(X) = 120*0.6 = 72* *[1]*

(ii) *Using continuity correction* *[0.5]*

P(X>100) = P(X>=100.5) = 1- Phi((100.5-120)/sqrt(72)) = 1-Phi(-2.298)

= Phi(2.298) = 0.989 *[1.5]*

Q2

Let $X_1, X_2, \dots, X_n$ be a random sample consisting of independent random variables with mean μ and variance σ^2. Consider the sample mean:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

In your answer you may denote μ by mu, σ^2 by sigma2 and $\bar{X}$ by Xbar.

(i) Write down the expected value of $\bar{X}$. [1]

(ii) Derive the variance of $\bar{X}$. [2]

(iii) Comment on the variance of variable $\bar{X}$ as compared to the variance of X_i. [1]

An actuary is interested in exploring the difference in the size of claim losses from two insurance portfolios, and can take samples of claims from these portfolios.

(iv) Explain how the answer to part (iii) can affect the precision of the actuary's comparison. [2]

[Total 6]

Solution

(i) *E(Xbar) = mu* *[1]*

(ii) *V(Xbar) = V(sum_over_1_to_n(X_i)/n) = sum_over_1_to_n(V(X_i))/ n^2 because of independence* *[1]*

*= n*sigma2/n^2 = sigma2/n* *[1]*

(iii) *The variance of the sample mean is smaller compared to the variance of individual variables by a factor of 1/n.* *[1]*

(iv) *Individual values are less precise than the average of a sample. The actuary should take large samples and compare the means to improve the precision of the comparison.* *[2]*

Q3

X and Y are discrete random variables with joint distribution as follows:

	$X = 0$	$X = 1$	$X = 3$
$Y = -1$	0.08	0.03	0.00
$Y = 0$	0.03	0.12	0.20
$Y = 3$	0.11	0.11	0.06
$Y = 4.5$	0.04	0.20	0.02

(i) (a) Identify which **one** of the following options gives the correct value of the expectation $E(Y \mid X = 1)$: [2]

(A1) 1.3789

(A2) 2.6087

(A3) 3.1398

(A4) 4.0945

(b) Identify which **one** of the following options gives the correct value of the variance $\mathrm{var}(X \mid Y = 3)$: [2]

(A1) 1.2487

(A2) 0.9832

(A3) 1.9388

(A4) 2.2235

(ii) Calculate the probability functions of the marginal distributions for X and Y.

[2]

(iii) Determine whether X and Y are independent.

[2]

[Total 8]

Solution

(i)(a) $E(Y \mid X = 1)$

$$= \sum_y y\, P(Y = y \mid X = 1)$$

$$= \sum_y y\, \frac{P(Y = y,\ X = 1)}{P(X = 1)}$$

$$= \left(-1 \times \frac{0.03}{0.46}\right) + \left(3 \times \frac{0.11}{0.46}\right) + \left(4.5 \times \frac{0.2}{0.46}\right)$$

$$= 2.6087$$

Ans: (A2) *[2]*

[Details are for information. Candidates will not be required to show working.]

(i)(b) $var(X \mid Y = 3) = E(X^2 \mid Y = 3) - (E(X \mid Y = 3))^2$

$$= \left(1 \times \frac{0.11}{0.28}\right) + \left(9 \times \frac{0.06}{0.28}\right) - \left(\left(1 \times \frac{0.11}{0.28}\right) + \left(3 \times \frac{0.06}{0.28}\right)\right)^2$$

$$= 2.3214 - (1.0357)^2$$
$$= 1.2487$$

Ans: (A1) *[2]*

[Details are for information. Candidates will not be required to show working.]

(ii) *Summing columns gives:*

$P(X = 0) = 0.26$, $P(X = 1) = 0.46$, $P(X=3) = 0.28$ *[1]*

Summing rows gives:

$P(Y = -1) = 0.11$, $P(Y = 0) = 0.35$, $P(Y = 3) = 0.28$,

$P(Y = 4.5) = 0.26$ *[1]*

(iii) *Test whether* $P(X = x, Y = y) = P(X = x)\, P(Y = y)$ *for all pairs* *[1]*

Show that this result does not hold for one pair, for example:

$P(X = 0, Y = -1) = 0.08$, *not equal to* $P(X = 0) * P(Y = -1)$

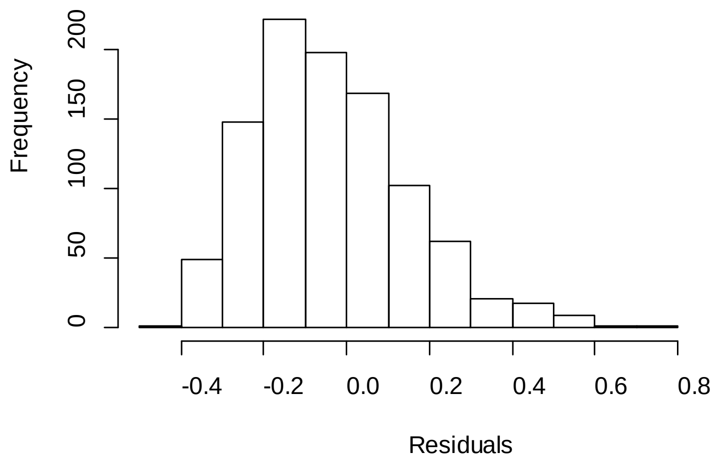
So X and Y are not independent. *[1]*

Q4

An actuary is asked to check a linear regression calculation performed by a trainee. The trainee reports a least squares slope parameter estimate of $\hat{b} = 13.7$ and a sample correlation coefficient $r = -0.89$.

(i) Justify why this suggests that the trainee has made an error. [2]

In a different simple linear regression model, a histogram of the residuals is shown below.



(ii) Comment on the validity of the assumptions of the linear model. [2]

The following pairs of data are available:

x:	0	1	2	3	4	5	6	7	8	9
y:	−1.35	− 4.96	− 9.20	−13.15	−16.70	− 21.23	− 25.14	− 28.44	− 33.68	−37.39

for which

$$\bar{y} = -19.124, \quad \sum_{i=1}^{10}(y_i - \bar{y})^2 = 1{,}329.523, \quad \sum_{i=1}^{10}(x_i - \bar{x})^2 = 82.5$$

$$\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y}) = -331.05$$

A linear model of the form $y = a + bx + e$ is fitted to the data, where the error terms (e) independently follow a $N(0, \sigma^2)$ distribution, and where a, b and σ^2 are unknown parameters.

(iii) Determine the fitted line of the regression model. [3]

(iv) (a) Identify which **one** of the following options gives the correct estimate of the variance σ^2 of the model. [1]

(A1) 0.612

(A2) 1.098

(A3) 0.971

(A4) 0.139

(b) Identify which **one** of the following options gives the correct estimate of the variance of the predicted mean response if $x = 11$. [2]

(A1) 0.161

(A2) 0.085

(A3) 0.287

(A4) 0.309

(c) Calculate a 95% confidence interval for the predicted mean response if $x = 11$. [2]

(v) Comment on the width of a 95% confidence interval for the predicted mean response if $x = 3.5$, as compared to the width of the interval in part (iv), without calculating the new interval. [2]

[Total 14]

Solution

(i) The regression slope suggests a positive relationship between the two variables, while the correlation coefficient shows a strong negative relationship. *[2]*

(ii) The histogram suggests a non-symmetric distribution for the residuals and therefore the assumption that the errors follow a N(0,sigma^2) distribution does not seem valid. *[2]*

(iii) bhat = Sxy/Sxx = -331.05/82.5 = -4.013 *[1]*
*ahat = ybar – bhat * xbar = -19.24 + 4.013 * (45/10) = -1.066* *[1]*
Line given as: yhat = -1.066 – 4.013 x *[1]*

(iv) *(a)*

$$\hat{\sigma}^2 = \frac{\left(S_{yy} - \frac{{S_{xy}}^2}{S_{xx}}\right)}{n-2} = \frac{\left(1329.523 \ - \frac{(-331.05)^2}{82.5}\right)}{8} = \ 0.139$$

Ans: (A4) *[1]*

[Details are for information. Candidates will not be required to show working.]

(b)

$$V(\hat{y}) = \left(\frac{1}{n} + \frac{(xnew - \bar{x})^2}{S_{xx}}\right) \times \hat{\sigma}^2 = \ \left(\frac{1}{10} + \frac{(11 - \frac{45}{10})^2}{82.5}\right) \times 0.139 = \ \ 0.085$$

Ans: (A2) *[2]*

[Details are for information. Candidates will not be required to show working.]

(c)

*Predicted value is: yhat = -1.066 – 4.013 * 11 = - 45.209* *[0.5]*

Critical value (from tables) is $t_{8,0.025} = 2.306$ *[0.5]*

95% CI for mean yhat is given by:
*- 45.209 - 2.306 * sqrt(0.085), - 45.209 + 2.306 * sqrt(0.085),* *[0.5]*

i.e. i.e. $(-45.881, -44.537)$*.* *[0.5]*

(v) *The width of the interval is only affected by V(yhat), which depends on the new x value through the term (x_new – xbar)^2. This term will now be smaller as the new x_new = 3.5 value is closer to xbar than x = 11. Therefore the interval will be narrower.*

[2]