**IN-SEMESRER 2 PRACTICE PROBLEMS**

**Q-1** The average working life of a certain IC chip is claimed to be 10000 hours, with standard deviation of 500 hours. We test a sample of size 100 of the chips, and calculate the sample mean. Find the probability that the sample mean is less than (10000 – 10\*N) hours.

**Standard normal distribution -- cumulative**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Index | z | F(z) | Index | z | F(z) |
| 1 | 0.00 | 0.5000 | 11 | 1.00 | 0.8413 |
| 2 | 0.10 | 0.5398 | 12 | 1.10 | 0.8643 |
| 3 | 0.20 | 0.5793 | 13 | 1.20 | 0.8849 |
| 4 | 0.30 | 0.6179 | 14 | 1.30 | 0.9032 |
| 5 | 0.40 | 0.6554 | 15 | 1.40 | 0.9192 |
| 6 | 0.50 | 0.6915 | 16 | 1.50 | 0.9332 |
| 7 | 0.60 | 0.7257 | 17 | 1.60 | 0.9452 |
| 8 | 0.70 | 0.7580 | 18 | 1.70 | 0.9554 |
| 9 | 0.80 | 0.7881 | 19 | 1.80 | 0.9641 |
| 10 | 0.90 | 0.8159 | 20 | 1.90 | 0.9713 |
|  |  |  | 21 | 2.00 | 0.9772 |

**🡪 S.D. of sample mean = 500/10 = 50. So 10\*N is equivalent to N/5 standard deviations. If N = 7, say, then that is 1.4 standard deviations. So answer = 1 – 0.9192.**

**Q-2** Random variable X has unknown probability distribution, but it is known that mX = 1000 and sX = 10\*N. Using Tchebycheff inequality, find a lower bound on the probability that the value of the random variable is within the interval 900 < X < 1100.

**We know that: Prob[ |X–m| > c ] < s2/c2, where c > 0.**

**Say N = 6. Then s = 60. Take c = 100. Then Prob[ |X–m| > c ] < 3600/10000 = 0.36. So answer = 0.64.**

**Note: Upper bound probability of being outside the interval = 1 - lower bound probability of being inside the interval.**

**Q-3** Five pairs of values of random variables X and Y are tabulated below. Find the linear regression of Y on X. The relevant formulas are provided below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | -2 | -1 | 0 | 1 | 2 |
| Y | -3M | -M | 0 | M | 3M |

**To be calculated exactly as shown in the book example, discussed in class.**

**FOR THE FOLLOWING FIVE PROBLEMS, SOLUTIONS ARE IN THE EXCEL SHEET.**

Q-1. Consider the following observed values of random variables X and Y, consisting of FIVE values each. For this pair of RVs, (a) Find the covariance Cov(X,Y) and (b) Find the regression coefficient rXY.

X values -2 -1 0 1 2   
Y values -3M -2M 1 2M 3M  
  
2. (a) For random variable X, it is known that the mean is 50 and the standard deviation is M. Using only Tchebycheff inequality, find an upper bound on the probability that the value of X will be outside the interval [48 ... 52].

(b) Further, it is given that the RV X of part (a) is normally distributed. Find the value DX such that Prob( |X-50| < DX ) = 0.98. Use the standard normal values provided.

3. Let X represent hourly traffic at a railway station, which is assumed to be Poisson distributed with parameter l = M2. Samples of 100 such traffic readings are taken, and their average Xmean calculated.

(a) Using CLT, find the standard deviation of Xmean.

(b) Find the range [Xmin ... Xmax], centred on l = M2, such that the value of X is within that range with a probability of 90%. Use the standard normal values provided.

4. (a) Random variable X has mean value 100 and standard deviation 5. For a left-tailed test of hypothesis, the level of significance (LOS) a is taken as M (in percent). Find the critical region of X, using the standard normal values provided.

(b) The average fatality rate for a specific category of hospitalized patients is 20%. In a given hospital, out of 100 patients of this category, 16 have died. At LOS = M (in percent), can this hospital be considered significantly better than average, for this category of patients? Use the standard normal values provided.

5. (a) Discrete random variable X has N = 5\*M equi-probable outcomes. A partition U is defined on the sample space S such that each event in U corresponds to exactly one outcome. Find the entropy H(U).

(b) The partition U of part (a) is modified by merging (that is, combining) ANY THREE of the events of U. Let the resulting partition be V. Find the entropy H(V).

Values of the standard normal variable, for Q-2, Q-3 and Q-4:

z: 1.645 1.751 1.881 2.054 2.326

CDF F(z): 0.95 0.96 0.97 0.98 0.99