

## MS-A0503 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

Exam 12.3.2014/ Aro

Write clearly *on every sheet of paper* :

- MS-A0503, tentti/exam, date
- Your student identification number
- IN BLOCK LETTERS in the following order: your family name, your first name(s)
- Your degree programme
- Signature

A calculator plus the collection of formulae and statistical tables by Mellin are allowed.

Justify all your answers adequately; e.g. a mere number without a justification as an answer will not give points.

1. (a) Let  $\Pr(A) = 0.5$  and  $\Pr(B) = 0.1$ . Find  $\Pr(A \cup B)$  if
  - Task 1: A and B are mutually exclusive
  - Task 2:  $\Pr(A|B) = 0.5$
  - Task 3: A and B are independent
- (b) 1 % of the employees of a firm are drug-addicts. A quick-test exposes 95 % of the drug-addicts, but on the other hand 2 % of those who do not abuse drugs will be erroneously classified to be drug-addicts by the quick-test. Compute the probability that an employee that is classified to be a drug-addict by the quick test is in reality not abusing drugs.
2. A factory manufactures products using very strict quality standards. On average, only 60% of the products meet these quality standards. Products are picked for inspection one at a time and at random.
  - Task 1: What is the probability that we need to inspect at least 4 products in order to find the *first* defective product?
  - Task 2: What is the expected number of products that we need to check in order to find the *first* defective product?
  - Task 3: What is the probability that we need to inspect at least 5 products to find the *third* defective product?
3. The Sparks Electric Institute has published figures on the number of kilowatt hours used annually by various home appliances. It claims that a vacuum cleaner uses an average of 46 kilowatt hours per year.

A random sample of 12 vacuum cleaners is picked and the kilowatt hours per year recorded. A sample mean of  $\bar{x} = 42$  and sample variance of  $s^2 = 141.61$  are obtained. Is there sufficient evidence at the 5% significance level to suggest that vacuum cleaners use, on average, less than 46 kilowatt hours annually. Assume that the distribution of the number of kilowatt hours used annually by a vacuum cleaner is normal.

4. A survey was commissioned to compare the distribution of support for three political parties, Vasemmistoliitto (Vas), Sosialidemokraattinen puolue (SDP) and Vihreä liitto (Vih), among the electorate of three municipalities A, B and C.

For the comparison, from the electorate of the three municipalities A, B and C independent random samples were picked whose sizes were: 300 (municipality A), 180 (municipality B) and 340 (municipality C) and those chosen were asked which political party they were going to vote for in the next election. The results of the survey are given in the table below.

Test the null hypothesis that support for the three parties is distributed in the same way in municipalities A, B and C, taking as the alternative hypothesis that support is not distributed similarly. Use the 5% significance level in the test.

	party	Vas	Sdp	Vih	Sample size
municipality	A	60	160	80	300
	B	40	80	60	180
	C	60	160	120	340

5. An experiment involved studying the dependence of mileage (variable  $y$ , miles per gallon) on vehicle weight (variable  $x$ , tonnes) for seven lorries. The results obtained are given in the table below.

$i$	1	2	3	4	5	6	7
$x_i$	8.00	24.50	27.00	14.50	28.50	12.75	21.25
$y_i$	7.69	4.97	4.56	6.49	4.34	6.24	4.45

The following statistics were calculated from the data:

$$\bar{x} = 19.5 \qquad \bar{y} = 5.5343$$

$$s_x^2 = 61.354 \qquad s_y^2 = 1.655$$

$$s_x = 7.833 \qquad s_y = 1.286$$

$$s_{xy} = -9.518 \qquad r_{xy} = -0.945$$

- (a) Determine the least-squares estimates for the parameters  $\beta_0$  and  $\beta_1$  in the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i : N(0, \sigma^2), i = 1, 2, \dots, n$$

- (b) Researchers calculated a value of  $s^2 = 0.214$  for the unbiased estimator of  $\sigma^2$ . What is the predicted value and 95% confidence interval for the *expected value* of  $y$  at  $x = 40$ ?