Solution for the exam in Economic program 2011-II Econometrics A, 2. year

Academic aim:

The aim of the course is to introduce the students to probability theory and statistics. The aim is for the student to be able to:

- understand the most important basic concepts of probability theory such as: probability, simultaneous-, marginal- and conditional probabilities, distribution, density function, independence, means, variance and covariance and apply these ideas on specific problems.
- know the result from the central limit theory.
- know and recognize the most commonly applied discrete and continuous distributions such as: Bernoulli, binomial, Poisson, multinomial, negative binomial, hypergeometric, geometric, uniform, normal, Chi-squared, exponential, gamma, t-, F-distribution and work with these distributions in relation to specific problems.
- understand the most important statistical concepts such as: random sampling, likelihood function, sufficient statistics, the properties and distributions of statistics, estimation, and maximum likelihood estimation and moment estimation, consistency, confidence interval, hypotheses, test statistics, test probability, level of significance, type I and II errors, power functions.
- perform a simple statistical analysis involving estimation, inference and hypothesis test e.g. the comparison of the means in two populations or test of independence for discrete stochastic variable.
- describe the result of his or her own analysis and considerations in a clear and distinct manner

In order for the student to obtain the highest grade possible, the student must demonstrate the mastery of the above-mentioned skills

1 Solution to question 1

1. Let X_t be the number of costumers arriving within t minutes. Then $X_t \sim Pois(1t)$. In this case, ten-minute interval, we have $X \sim Pois(10)$. The expected number of arrivals is 10. $P(X > 15) = 1 - P(X \leq 15) = 1 - 0.951 = 0.049$

2. The sum of two Poisson variables is also Poisson $\lambda_A + \lambda_B = \lambda$. In this case $\lambda_B = 0.8$ and $\lambda = 1$ and therefore $\lambda_A = 0.2$. We expect 0.2 type A customers within a minute out of 1 of any type. Hence the probability of type A is 0.2. Let Y be the number of type A given n arrivals then $Y \sim bin(n, 0.2)$. if $n = 4$ then $P(Y = 1) = 4 \cdot 0.2 \cdot 0.8^3 = 0.4096$.

3. The number of costumers waiting depends on the number of arrivals in ten-minute interval. Let $Z(= X - 15, X > 15)$ be the number of persons waiting. $Z \sim Pois(10|X > 15)$. The probability of more than 22 arrivals is very close to zero and therefore we can safely ignore more than 7 is waiting. The number of persons waiting of type A can be described by $W \sim bin(Z, 0.2)$.

Z	$P(Z)$	$P(W=1)$	$P(Z) \cdot P(W=1)$
1	$0.973 - 0.951 = 0.022$	0.20	0.0044
2	$0.986 - 0.973 = 0.013$	0.32	0.00416
3	0.007	0.384	0.002688
4	0.004	0.4096	0.0016384
5	0.001	0.4096	0.0004096
6	0.001	0.3932	0.0003932
7	0.001	0.3670	0.0003670

The probability of exactly one type A is waiting is the sum of probabilities in the last column 0.014056.

2 Solution to question 2

1.

	no children	1 child	2 children	3 or more children
probability of participation	0.9	0.8	0.7	0.6

2. We have $P(W|A) = \frac{P(W,A)}{P(A)} = \frac{\sum_C P(W,A,C)}{P(A)} = \frac{\sum_C P(W|C)P(C|A)P(A)}{P(A)} = \sum_C P(W|C)P(C|A)$. From this we find that $P(W = participate|age\ is\ 40 - 49) = 0.7$

3. The negative correlation between the probability of participation and age is due to the positive correlation between children and age, and the negative correlation between children and participation.

3 Solution to question 3

1. identical probabilities and assuming independence.

2. $\widehat{p} = \frac{109}{633} = 0,17$. This estimate is unbiased and the variance will approach zero then n goes to infinity. Hence it is consistent estimator.

3. $\widehat{p} \pm \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0,17 \pm \sqrt{\frac{0,17(1-0,17)}{633}} = 0,17 \pm 0,03$

4. $z = \frac{\widehat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0,17-0,25}{\sqrt{\frac{0,25(1-0,25)}{633}}} = -4,5; \Phi(z) = 0,0$, so the significans probability is very small. We reject null hypothesis ($H_0 : p = 0.25$) and accepts the alternative $p < 0,25$.

5.

	Occupation	Occupation	Occupation	total
	A	B	C	
Obs.	124	248	261	633
f. probability	$0,2$	$0,4$	$0,4$	1
f values	$126,6$	$253,2$	$253,2$	633
$O-F$	$-2,6$	$-5,2$	$7,8$	0
$\frac{(O-F)^2}{F}$	$0,053397$	$0,106793$	$0,240284$	$0,400474$
				$0,818537$

The test is χ^2 distributed with 2 deg. free. (all expected values are larger than 5). $P(\chi^2(2) > 0,4) = 0,82$.This is a very large significans probability and the null hypothesis cannot be rejected.

6.

	Occupation	Occupation	Occupation	Total
Headache	A	B	C	
yes	19	52	38	109
no	105	196	223	425
total	124	248	261	633
	$21,4$	$42,7$	$44,9$	109
	$102,6$	$205,3$	$216,1$	524
	$-2,4$	$9,3$	$-6,9$	0
	$2,4$	$-9,3$	$6,9$	0
	$0,26$	$2,02$	$1,07$	$3,36$
	$0,05$	$0,42$	$0,22$	$0,70$
				$4,05$
				$0,131797$

The test size is $4,05$ which is distributed $\chi^2(2)$, Deg. free. $=(3-1)(2-1)$.The expected values are large enough.Siginificans probability is $13,2\% > 5\%$ and the null hypothesis is accepted, that is there is independence between occupations and headache.