Solution for the re-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

Solution to question 1

1. The probability of receiving a SMS in a given week is $P(X < 0.75 \cdot 12) + P(X > 1.25 \cdot 12) = 2 \cdot P(X < 0.75 \cdot 12) = 2 \cdot P(X > 1.25 \cdot 12)$. Last equalities is due to symmetry and any can be used. Let $p = 2 \cdot P(X < 0.75 \cdot 12) = 0.1336$. The text errorneous says a 'day' instead of 'week', hence any reply that makes sense should be given full credit. This has no influence on subsequent Q&A.

2. Let Y be the number of SMS received during eight weeks then $Y \sim Bin(8; 0.1336)$ and $P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - 0.1336)^8 = 0.682497208$.

3. A single household. Let $X^{treatment} \sim N(11.8, 4)$ and $X^{control} \sim N(12, 4)$. Let $D = X^{treatment} - X^{control}$. $D \sim N(-0.2, 8)$. $P(D < 0) = P(Z \leq \frac{0.2}{\sqrt{8}}) = 0.5279$. ($Z$ is standard normal).

   Average of 100 households: Let $\overline{X}^{treatment} = \frac{1}{100}\sum_{i=1}^{100} X_i^{treatment}$ and $\overline{X}^{control} = \frac{1}{100}\sum_{i=1}^{100} X_i^{control}$. $E[\overline{X}^{treatment}] = 11.8$ and $E[\overline{X}^{control}] = 12$ from linearity of expectations. Also $Var(\overline{X}^{treatment}) = \frac{4}{100}$ and $Var(\overline{X}^{control}) = \frac{4}{100}$ from i.i.d. Hence $\overline{X}^{treatment} \sim N(11.8, \frac{4}{100})$ and $\overline{X}^{control} \sim N(12, \frac{4}{100})$. Define $\overline{D} = \overline{X}^{treatment} - \overline{X}^{control}$. $\overline{D} \sim N(-0.2, \frac{8}{100})$. $P(\overline{D} < 0) = P(Z \leq \frac{0.2}{\sqrt{\frac{8}{100}}}) = 0.7611$. A single household has much larger variance than the mean of a group of households. That is uncertainty is reduced as the sample increases and collective behaviour is centered at the mean (law of large numbers).

Solution to question 2

1. In case of hiring $P(survival|A) = P(N = 0|A) + P(N = 1|A) \cdot P(D = 1|N = 1, A) = 0.8 + 0.2 \cdot 0.7 = 0.94$. $P(survival|B) = 0.95$. $P(survival|C) = 0.97$. In case of not hiring the probability is the last row in table 1.

2. Use Bayes theorem

$$P(A|N = 1, D = 1) = \frac{P(N = 1, D = 1|A) \cdot P(A)}{P(N = 1, D = 1|A) \cdot P(A) + P(N = 1, D = 1|B) \cdot P(B) + P(N = 1, D = 1|C) \cdot P(C)} = \frac{0.2 \cdot 0.7 \cdot \frac{1}{3}}{0.2 \cdot 0.7 \cdot \frac{1}{3} + 0.1 \cdot 0.5 \cdot \frac{1}{3} + 0.05 \cdot 0.4 \cdot \frac{1}{3}} = \frac{2}{3}.$$

   This is also expected as the evidence is in favour of country A.

3. Use Bayes theorem once more

$$P(A|N=0) =$$
$$\frac{P(N=0|A)\cdot P(A)}{P(N=0|A)\cdot P(A) + P(N=0|B)\cdot P(B) + P(N=0|C)\cdot P(C)} =$$
$$\frac{0.8\cdot\frac{2}{3}}{0.8\cdot\frac{2}{3} + 0.9\cdot\frac{5}{21} + 0.95\cdot\frac{2}{21}} = 0.63636363$$

   The idea here is that as new information arrives the Bayes theorem can be used again and again, using the posterior probabililty as prior beliefs.

Solution to question 3

1. $\overline{X}_1 \pm 1.96\sqrt{\frac{\sigma_1^2}{n}} = 480.4 \pm 1.96\frac{260.7}{\sqrt{2886}} = 480.4 \pm 9.3$. The number of observations is very large (2886).

2. $u = \frac{\overline{X}_2 - 500}{\sqrt{\frac{\sigma_2^2}{n}}} = \frac{509.2-500}{\frac{248.1}{\sqrt{3038}}} = 2.04$ which is significant. $P(u \geq 2.04) = 2\%$, so we reject the hypothesis that the girls have an average of 500 and conclude that there are above 500 (whcih is the average for all the countries).

3. $u = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{509.2-480.4}{\sqrt{\frac{248.1^2}{3038} + \frac{260.7^2}{2886}}} = 4.35$ which is very significant. $2\cdot P(u \geq 4.35)$ very small.It makes sense the boys are below the girls in reading (this is also true in all other countries).

4. You could argue that this is a hyper geometric distribution. But it can be approximated to a binomial distribution. Otherwise you can consider two outcomes, independence and (almost) the same probability.

5. The estimator for p will be: $\frac{1070}{5924} = 0.18$. It is an unbiased estimator, where the variance will go to zero as the numbers (of n) will go to infinity. So it is a consistent estimator.

6.

	boys	girls	total
Language			
Danish	2381	2464	4845
Other	505	574	1079
Total	2886	3038	5924
	2360.3	2484.7	4845
	525.7	553.3	1079
	20.7	-20.7	
	-20.7	20.7	
	0.180789	0.171744	0.352534
	0.811793	0.771177	1.58297
			1.935504
			0.164158

We calculate the expected values. Then calculate the difference of the observed and expected value.And finally calcualte the test statistics, which is 1.93. This is χ^2 distributed with one degree of freedom. $P(\chi^2 \geq 1.93) = 16.4\%$ It is not significant. So we will not reject the assumption that gender is independent of spoken language at home. You can also test if two indpendent binomial distributions have the same p.