Re-examination – Statistics for the courses: Research process, Research overview and design.

Ignacio Rangel. Clinical Medicine. Örebro University

Bengt Löfstrand

Date: 7/6 2014

Score: pass with 62.5 % of overall score, e.g. 10p

It is allowed to bring in: a book in statistics and a pocket calculator

1.- indicate if the following statements are true or false –no further comments to the answers are required (**3p)**

a) $-1 \geq r \geq 1$

(0.5p)

b) r = 0.5 predicts better the effect of *x* on *y* than r = - 0.5

(0.5p)

c) SEM is a measurement of uncertainty of the mean

(0.5p)

d) nominal categorical data and ordinal categorical data can be summarised with the same statistical measures

(0.5p)

e) as long as the sample size is large, categorical data can be analysed with normally distributed based procedures

(0.5p)

f) the z-test has a restrictive requirement: the value of the true standard deviation of the response is to be known

(0.5p)

2.- to study the risk of smoking on miscarriage a group of researchers chose from the Swedish inhabitants, 500 women. They received a questionnaire where they were asked about their smoking habits and whether they have had a miscarriage or not. **(3p)**

a) Define which is the population in this study? (0.5p)

b) Define which is the randomly selected sample in the study? (0.5p)

c) The women were divided in smokers and non-smokers. 280 women were smokers while the rest were non-smokers. What is the probability to randomly select a non-smoker among all the women in the study? And, what is the probability to randomly select 3 non-smokers in 3 different selections? (1p)

d) Formulate the different steps for a hypothesis test to answer the question: is there a higher risk for smoking women than for non-smoking women to suffer miscarriage? (1p)

3.- In a study where researchers wanted to investigate if there was a correlation between exposure for passive smoking and miscarriage, a total of 250 women were asked if they were exposed for passive smoking. The answer alternatives were “never”, “sometimes” and “often”. Results were as follows: 40 answered “never”, 120 answered “sometimes” and the rest answered “often”.

a) What kind of data is this? Make a figure to summarise the results, present an “average” point as well as a distribution measurement of the data **(2p)**

4.- Two pain relieving drugs, drug Z and drug M, were compared in a clinical trial. Two independent groups of patients were chosen. Drug Z was given to 150 patients while drug M was given to 180 patients. Results were as follow: 80 patients felt relief after taking drug Z; 95 patients felt better after taking drug M. Perform a suitable test with significance level 5% to determine if the effect of both drugs was similar or not. **(3p)**

5.- Assume that the weight among 10-year old girls is normally distributed and that the average weight is 33,5 kg. From a free-school in a town in southern Sweden, where children are served only vegetarian food, the average weight among a randomly selected sample of 41 students (same age and sex) was 32,3 kg (standard deviation = 3,75 kg). Is it possible to conclude that the average weight in the sampled free-school was different from the standard average weight of 10-year old girls? Base your conclusion on the application of hypothesis testing development.

(3p)

Questions from: Bengt Löfstrand

(2p)

Five diabetic patients had plasma glucose levels (mmol/l) measured before and after oral administration of 100 g glucose with the following results.

Patient no	Change in plasma glucose (mmol/l)
1	0.47
2	3.12
3	4.22
4	2.15
5	5.03

a Calculate the mean and standard deviation (SD) of change in plasma glucose.

b On the basis of these data, how many diabetic patients would need to be studied so that the width of the 95 % confidence interval for the mean change in plasma glucose level was
0.5 mmol/l ? (Assume that the Normal distribution is the appropriate distribution for the change in plasma glucose.)