TMA4245 Statistics
Exam August 10 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

A person delivering blood sees that his levels of homoglobin (X, in g/dl) in the period 1993 to 2012 appear independent and normally distributed with mean 16.14 and standard deviation 0.63.

a) Athletes are not allowed to compete if their level of hemoglobin exceeds 17.5 (the blood is too thick). What is the probability that our person is not allowed to compete?

Before larger competitions, like the Tour de France bike race, there are many tests to prevent doping. What is the probability that our person gets at least one of five observations above the limit at 17.5? Describe briefly the assumptions needed to compute this probability.

After 2000 they discovered that the method used has too much uncertainty, with variance of 0.63^2. A new method was introduced from 2007. Fifteen blood samples made with the new method on our person gave $\sum_{i=1}^{15}(x_i - \bar{x})^2 = 2.2$.

b) Formulate a hypothesis test to check if the new method has smaller variance than the old one.

What assumptions are done to perform this test?

What is the conclusion when data are as above, and the significance level is 0.05?

Oppgave 2

A problem with windmills for energy production is that birds can collide and die. As a trial project, a windmill (A) is set along the Norwegian coast, and one registers the number of birds that collide and die. Experience from Denmark indicates that the expected number of birds that die per week is $\lambda = 1$. Assume that the random variable Y, the number of birds that collide and die with mill A per week, is Poisson distributed with this parameter λ.

a) Assuming the same conditions in Norway as in Denmark, and independence between weeks, compute the probability that more than 10 birds collide and die during the first five weeks.

Given that less than 5 birds collide, what is the probability that no birds collided?

b) During the first four years, 261 birds die by mill A. Estimate the parameter λ by maximum likelihood from this data.

Another mill (B) is placed further inland, with less birds. Assume X, the number of birds that collide and die with mill B per week, is Poisson distributed with parameter ν. Assume further independence between mills A and B. Let $Z = X + Y$ be the total number of birds that die by the two mills.

c) Compute the moment generating function $M_Z(t)$ of the random variable Z.

Use the moment generating function to find the distribution of Z.

Oppgave 3

Teodor works in the ice crem bar by the camping. He wonders whether the income of sales changes with temperature. Let x_i be temperature at 2pm day i. We consider this temperature and the income Y_i for $i = 1, \ldots, n$ days.

Assume a regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, for $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are independent normal error terms with mean 0 and variance τ^2.

It can be shown that an unbiased estimator for β_1 is $\hat{\beta}_1 = \frac{\sum_{i=1}^n Y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

a) Show that the variance of $\hat{\beta}_1$ is $\frac{\tau^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

Let further $\hat{\beta}_0$ be an estimator for β_0. Explain why $s^2 = \frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$ is a useful estimator for τ^2. Why do we divide by $n - 2$?

The income can be split in sales of cream and limonade ice cream. Define these by y_i^f and y_i^s, respectively. Teodor thinks the income may increase differently with temperature, and suggest a model as follows:

$$Y_i^f = \beta_0^f + \beta_1^f x_i + \epsilon_i^f, \quad Y_i^s = \beta_0^s + \beta_1^s x_i + \epsilon_i^s \quad i = 1, \ldots, n.$$

Here, $\epsilon_1^f, \ldots, \epsilon_n^f, \epsilon_1^s, \ldots, \epsilon_n^s$ are independent normally distributed noise terms with mean 0 and variance σ^2.

b) Derive a 90% confidence interval for the differance in the slopes, $\beta_1^f - \beta_1^s$.

Compute the interval when there are $n = 14$ days with data, $\hat{\beta}_1^f = 1720$ and $\hat{\beta}_1^s = 1442$. Moreover, $\sum_{i=1}^n (x_i - \bar{x})^2 = 260.4$, $\sum_{i=1}^{14} (Y_i^f - \hat{\beta}_0^f - \hat{\beta}_1^f x_i)^2 = 7046^2$ and $\sum_{i=1}^{14} (Y_i^s - \hat{\beta}_0^s - \hat{\beta}_1^s x_i)^2 = 7300^2$.

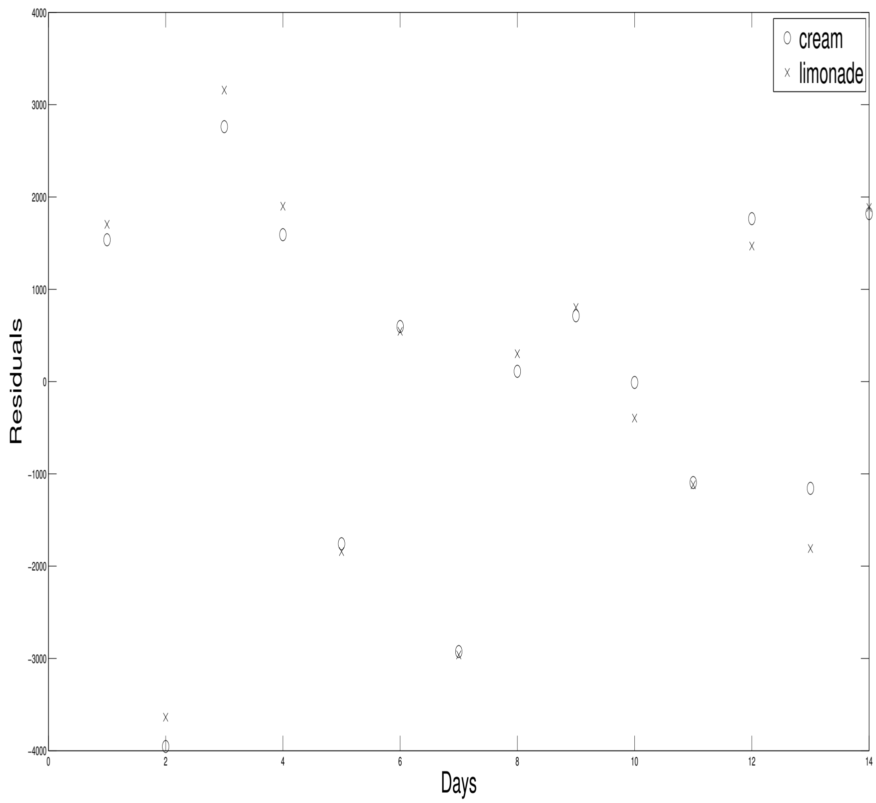
Note: In total 4 parameters are estimated in the regression lines here.

Figure 1 shows residuals after fitting regression lines for the income of cream and limonade ice cream these 14 days.

c) Use the plot in Figure 1 to discuss whether the assumptions about the error terms in b) seem reasonable.

Oppgave 4

In a simple version of the game Craps a player throws two dices. The sum of dices are registered: a number between 2 and 12. There are two options for winning: i) *Direct win:* get 7 or 11 in



Figur 1: Estimated residuals after fitting regression lines for the income of cream and limonade ice cream these 14 days.

the first round. ii) *Delayed win:* Get 4, 5, 6, 8, 9 or 10 in the first round and then get the same number in a later round, but before getting 7.

Compute the probability of a *direct win*.

Let A_i = win a *delayed win* by a sum i, where $i = 4, 5, 6, 8, 9, 10$. Compute $P(A_4)$.

Compute finally the probability of winning in Craps. Hint: $\sum_{k=0}^{\infty}(\frac{q}{36})^k = \frac{1}{1-q/36}$, for $q \in \{1, \ldots, 35\}$.

Fasit

1. **a**) 0.154,0.075 **b**) Reject H_0

2. **a**) 0.014,0.015 **b**) 1.25

3. **b**) $[-32, 588]$

4. 8/36,1/36,0.493