TMA4255 Applied Statistics
Solution to Exercise 9

Problem 1

$\bar{X}$- and S-diagram for the given data are in Figure 1.

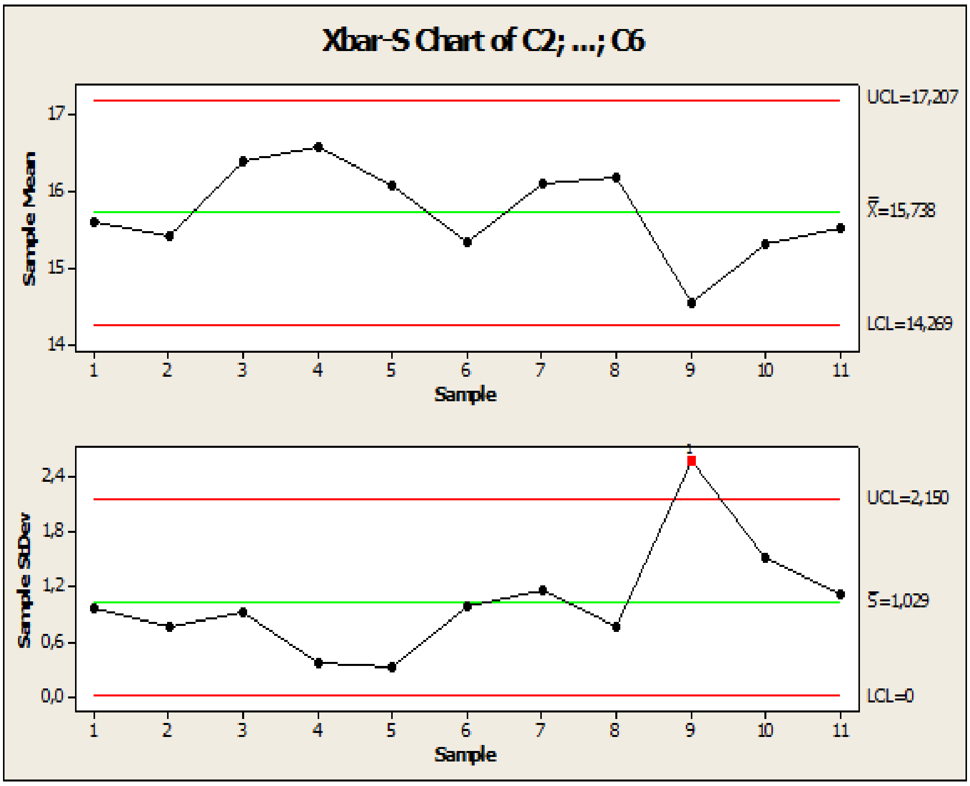

Figure 1: $\bar{X}-$ and S-diagram for the data in problem 1.

The process is said to be in control if all the sample averages $\bar{x}_i$ are within the limits $\mu \pm 3\sigma/\sqrt{n}$ (expect ± 3 stdev.)

μ is estimated by the average of all the samples: $\overline{\overline{X}} = \sum_{i=1}^{k} \bar{X}_i$, k = number of samples.

σ: We ukse $\bar{S} = \frac{1}{k}\sum_{i=1}^{k} S_i$ where $S_i = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2$. Then σ is estimated by $\hat{\sigma} = \bar{S}/c_4$.

X-chart

$$UCL = \hat{\mu} + 3\frac{\hat{\sigma}}{\sqrt{n}} = \overline{\overline{X}} + \frac{3}{\sqrt{n}}\frac{\bar{S}}{c_4} = 15.74 + \frac{3}{\sqrt{5}} \cdot \frac{1.029}{0.94} = 15.74 + 1.47 = 17.21$$

$$LCL = \hat{\mu} - 3\frac{\hat{\sigma}}{\sqrt{n}} = 15.74 - 1.47 = 14.27$$

S-chart uses the variation in the data as a measure of quality. If the lower limit is below 0 it is set to 0 (done here). $c_4 = 0.94$ for $n = 5$.

$$UCL_S = 1.029 + 3 \cdot \frac{1.029}{c_4}\sqrt{1-c_4^2} = 1.029 + 1.120 = 2.15$$

$$LCL_S = \max(1.029 - 2.15, 0) = 0$$

Conclusion
X-chart: All observations lie within the control limits.
S-chart: Observation from sample 9 falls outside the control limits. $\Rightarrow$ the process needs to be reestimated.

Calculate new control limits omitting sample 9.

If all the observations are now within the new limits, we can assume that the new estimate gives valid control limits for a process under control.

The new estimate for $\hat{\sigma} = 0.875$, and $\overline{\overline{X}} = 15.856$.

$$UCL_{\text{ny}} = 15.865 + \frac{3}{\sqrt{5}} \cdot \frac{0.875}{0.94} = 15.865 + 1.249 = 17.1$$

$$LCL_{\text{ny}} = 15.856 - 1.249 = 14.6$$

$$UCL_{S\,\text{ny}} = UCL_S = 0.875 + 3 \cdot \frac{0.875}{c_4}\sqrt{1-c_4^2} = 0.875 + 0.953 = 1.828$$

$$LCL_{S\,\text{ny}} = \max(0.875 - 0.953, 0) = 0$$

$\Rightarrow$ all observations lie within the new limits shown in Figure 2.

Problem 2

P-chart: Quality is measured by attributes (defective/not defective). Cheaper to do, but gives less information about the process.

Assumes that the number of defective $D \sim \text{bin}(n, p)$ in sample of size n.

$$\Rightarrow E(D) = np, \text{Var}(D) = np(1-p)$$

$$\Rightarrow E(D/n) = p, \text{Var}(D/n) = p(1-p)/n.$$

We want to find control limits for the proportion p of defective.

$$E(D/n) \pm 3\sqrt{\text{Var}(D/n)} = p \pm 3\sqrt{p(1-p)/n}$$

Estimate p by $\tilde{p} = \frac{1}{k}\sum_{i=1}^{k}\frac{D_i}{n}$, k: number of samples, n: sample size.

$$\tilde{p} = \frac{1}{30}\sum_{i=1}^{30}\frac{D_i}{1000} = 0.0106$$

$$UCL_p = \tilde{p} + 3\sqrt{\tilde{p}(1-\tilde{p})/n} = 0.0203$$

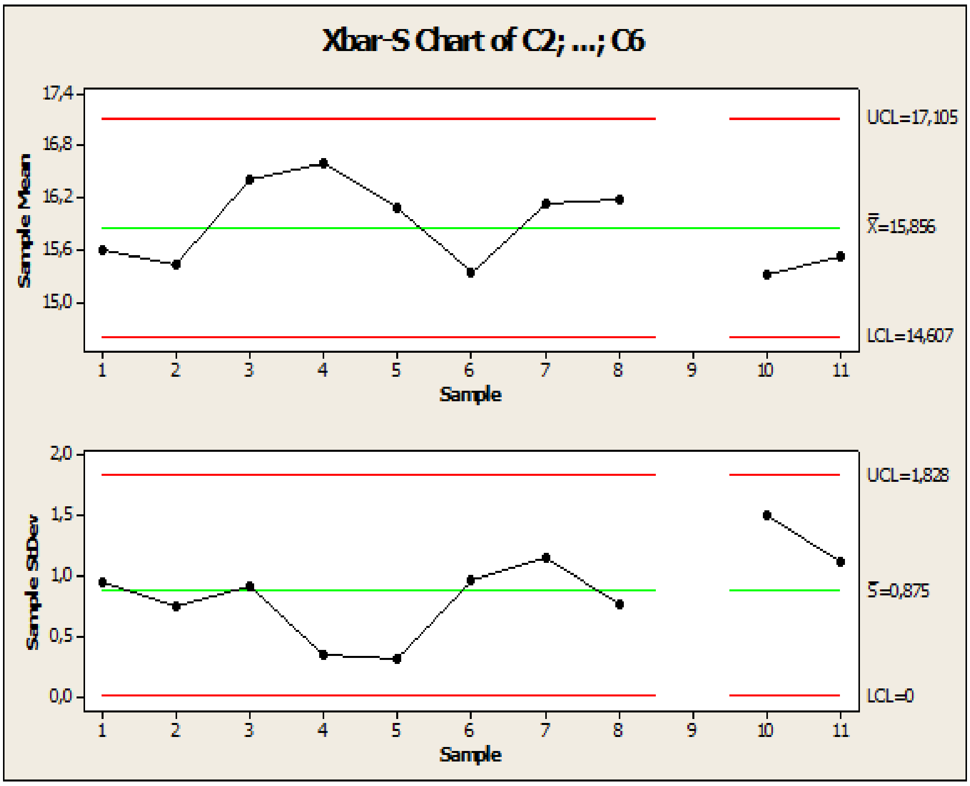

Figure 2: $\bar{X}-$ and S-diagram for the data in problem 1 when observation 9 is omitted.

$$LCL_p = \tilde{p} + 3\sqrt{\tilde{p}(1-\tilde{p})/n} = 0.000885$$

Minitab gives $p-$diagram as shown in Figure 3.

The figure shows that $\Rightarrow$ All observations lie within the limits.

$\Rightarrow$ The process is in control.

Problem 3

Investigates C = the number of defects in each unit, where we assume $C \sim \text{Po}(\lambda)$ so that $E(C) = \text{Var}(C) = \lambda$.

The control limits: $\mu \pm 3\delta = \lambda \pm 3\sqrt{\lambda}$

Estimate λ by $\bar{C} = \frac{1}{k}\sum_{i=1}^{k} C_i$, k = the number of elements.

$$\bar{C} = \frac{1}{13}\sum_{i=1}^{13} C_i = 7.577$$

$$UCL_C = \bar{C} + 3\sqrt{\bar{C}} = 15.83$$

$$LCL_C = \bar{C} - 3\sqrt{\bar{C}} = -0.68$$

When the estimate of $LCL_C < 0$ it is set equal to 0 by definition.

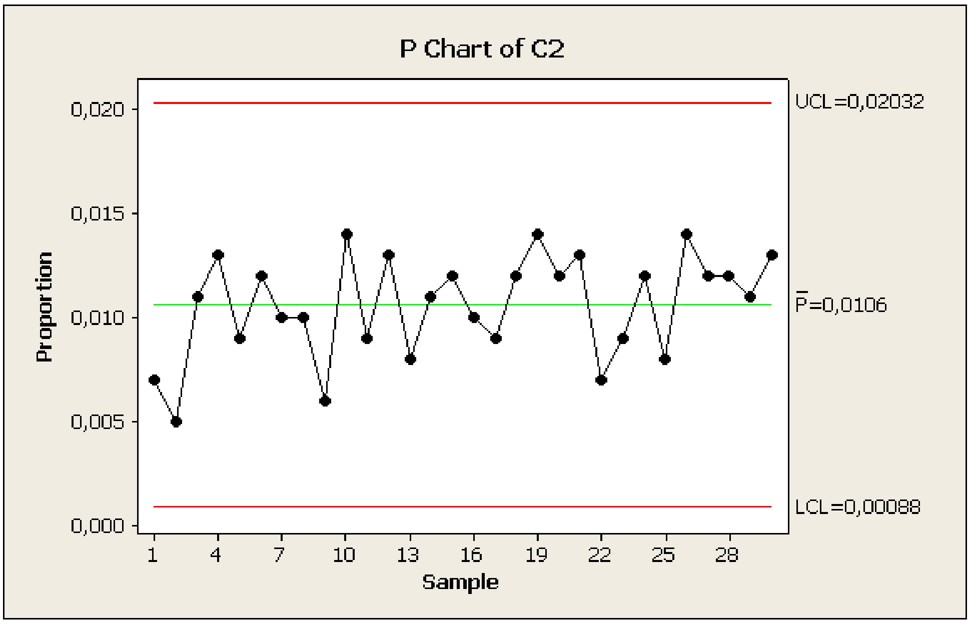

Figure 3: $p-$diagram for the data in problem 2.

From the $c-$diagram in Figure 4 we see that observation 5 falls outside the upper limit $\Rightarrow$ the process is out of control.

Find new limits by omitting observation 5:

$$\bar{C}_{\text{ny}} = 7.2$$

$$UCL_C = \bar{C}_{\text{ny}} + 3\sqrt{\bar{C}_{\text{ny}}} = 15.25$$

$$LCL_C = \bar{C}_{\text{ny}} - 3\sqrt{\bar{C}_{\text{ny}}} < 0$$

$\Rightarrow LCL_C = 0$

From Figure 5 we see that all observations are now inside the new limits.

Problem 4

We will test

$$H_0: \text{ the model is correct, i.e. } T \sim f(t) \text{ against } H_1: \text{ the model is wrong}$$

It is reasonable to use a Goodness-of-Fit test to do this hypothesis test. We use the test statistic

$$W = \sum_{i=1}^{6} \frac{(X_i - e_i)^2}{e_i},$$

Where X_i and e_i are observed and expected value for interval i respectively. This test statistic is chi-squared distributed with 4 degrees of freedom. Because this is based on the maximum likelihood estimate of the unknown parameter θ we get 4 instead of 5 degrees of freedom.

To find the estimated values, e_i, we first need to find the probability of observing times in each interval. We may either use the gamma distribution directly, or use $Z = \frac{2T}{\theta}$, which is chi-squared

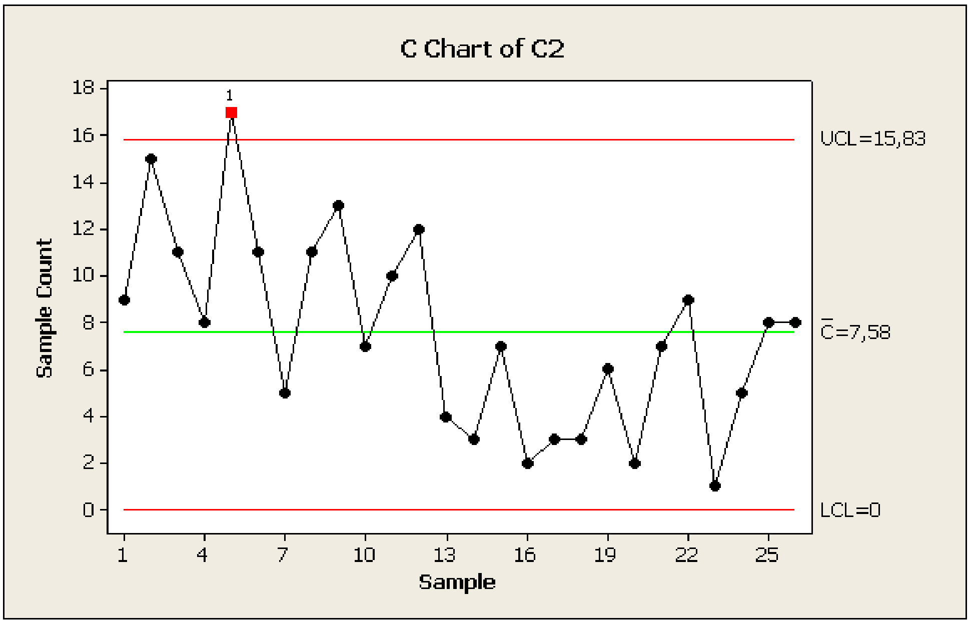

Figure 4: $c-$diagram for the data in problem 3.

distributed with 4 degrees of freedom. We need to insert the estimate for θ, $\bar{T} = 1800/60 = 30$.

Interval nr.	Interval	Interval for Z	$P(Z < z_{int})$	$P(Z \in int)$
1	(0, 25]	(0, 1.66]	0.20	0.20
2	(25, 40]	(1.66, 2.66]	0.38	0.18
3	(40, 55]	(2.66, 3.66]	0.54	0.16
4	(55, 70]	(3.66, 4.66]	0.67	0.13
5	(70, 90]	(4.66, 6.00]	0.80	0.12
6	(90, $\rightarrow$)	(6.00, $\rightarrow$)	1.00	0.20

Interval nr.	Estimated value	Observed value	$\frac{(X_i - e_i)^2}{e_i}$
1	12.19	12	0.003
2	10.90	7	1.397
3	9.72	11	0.167
4	7.78	12	2.281
5	7.44	8	0.041
6	11.94	10	0.317

If we add the numbers in the last column, we get the value $W = 4.208$ of our test statistic. This corresponds to a p-value of 0.38 which does not give us enough reason to reject H_0.

Problem 5

We will investigate the milk drinking habits among young and old people. We state this as a hypothesis test with

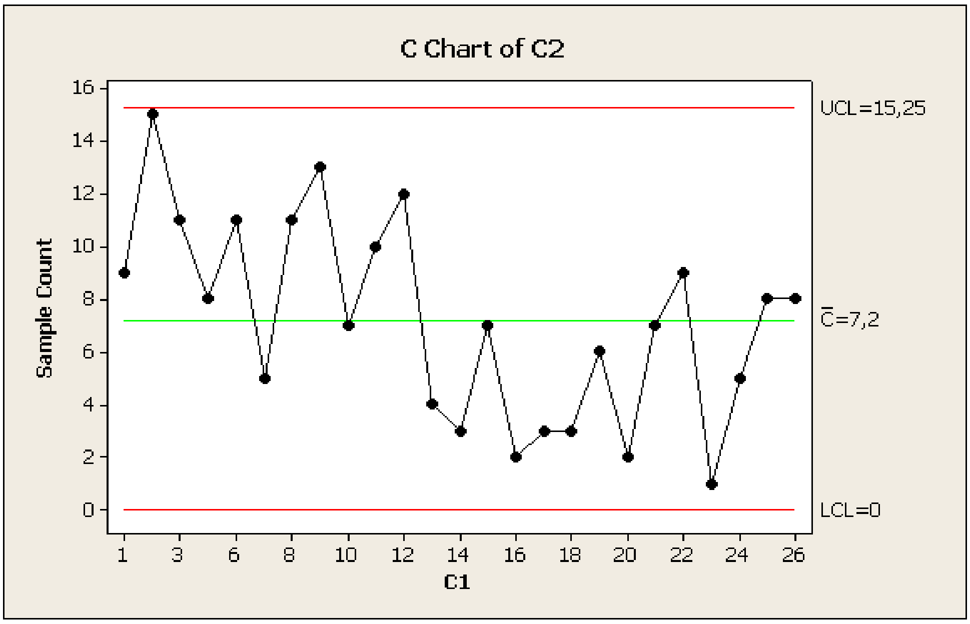

Figure 5: $c-$diagram for the data in problem 4 when observation 5 is omitted.

$$H_0: \text{Habits are independent of age} \quad \text{against}$$
$$H_1: \text{Habits depend on age.}$$

The way we have stated the null hypothesis, we can use an independence test as given in the text book. We base the test on the statistic

$$W = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}},$$

which is chi-square distributed with $(2-1)(3-1) = 2$ degrees of freedom. In this expression o_{ij} and e_{ij} are observed and epected value respectively of the random element in row i and column j. The expected value is given by

$$e_{ij} = \frac{(\sum_{j=1}^{3} o_{ij}) \cdot (\sum_{i=1}^{2} o_{ij})}{\sum_{j=1}^{3} \sum_{i=1}^{2} o_{ij}}.$$

I.e. row sums multiplied with column sums and divided with the total number of observations. With the observations put into Minitab with people under 45 in row 1 and people over 45 in row 2, we get the following results:

```
Chi-Square Test: Skummet; Lett; Hel

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

        Skummet   Lett    Hel  Total
```

```
    1       138     83     64    285
         115,14  85,50  84,36
          4,539  0,073  4,914

    2        64     67     84    215
          86,86  64,50  63,64
          6,016  0,097  6,514

Total       202    150    148    500

Chi-Sq = 22,152; DF = 2; P-Value = 0,000
```

Sadly, MINITAB does not output the exact p-value, but R gives $1.5 \cdot 10^{-5}$. Because the p-value is very small, and in particular below 0.05, we reject H_0 and conclude that there is a difference between the milk drinking habits of young people and that of old people.

Problem 6

See course WWW-page under Exam for solution to August 2012, Problem 1 on homogeniety.