Basisprüfung B. Sc.
Statistics and Probability Theory
WS 2008

Prof. Dr. M.H. Faber

ETH Zürich

Friday February 08th 2008
14:00 – 16:00

Surname: ....

Name: ....

Stud. Nr.: ....

Course of studies: ....

Basisprüfung B. Sc.: Statistics and Probability Theory

Civil, Environmental and Geomatic Engineering

Date and duration:

Friday February 08th 2008
Start: 14:00
Duration: 120 minutes

Aids:

- All documentation and manuals allowed (Lecture notes, Exercise tutorials, other books and print-outs etc.)
- Calculators are allowed, but no communication medium (e.g. cell phones, calculators with Bluetooth etc.)

Administration:

- During the 15 minutes reading, it is not allowed to write on the solution sheets.
- Please place your **Legi-**card on your desk.
- Control first if you have received all the materials:
    - General information and exercises (16 pages)
    - 6 sheets of paper (checkered and stamped)
- Write your **name on every sheet** of paper.
- Use **only** the provided sheets of paper (6 checkered and officially stamped sheets) and use a **new sheet for every exercise**.
- Other sheets will **not** be considered in the corrections!
- When you have finished, place **all** materials in the envelope and leave it on your desk.
- You are allowed to leave the exam at any time until 15:45. After that time, you need to wait until the end of the exam.

Content of the exam:

Content	Description	Page	Points
Data	Data on half-cell potential measurements		
Exercise 1	Descriptive statistics	4	20
Exercise 2	Bayes' theorem	6	20
Exercise 3	Probability paper	7	20
Exercise 4	Confidence interval and hypothesis testing	10	20
Exercise 5	Goodness of fit test for a distribution	11	25
Exercise 6	Estimation of parameters for a distribution	13	15
Annexes	Table A: Cumulative distribution function of the Standard Normal distribution	14	**-**
	Table B: Critical values of the Kolmogorov-Smirnov test	15	**-**
Glossary	English-German	16	**-**
			120

Remarks:

- All exercises 1 to 6 have to be solved.
- If you are having difficulties in a certain question but need a value/number in order to continue, then make an assumption, mark it as an assumption, and continue the calculations with that value.

<u>Data</u>

Note: The following data will be the basis for the following exercises. Note although, that the exercises should be solved independently one from the other.

The half-cell potential measurement test is an inspection technique commonly used to assess the presence of corrosion in reinforced concrete structures by measuring the electric potential [mV].

Table 1 contains the electric potentials measured by a half-cell potential measurement test for concrete elements with and without presence of corrosion in the reinforcement. The actual state of corrosion has been assessed by exposing the reinforcement after the potential measurements.

Table 1: Measured potential values for concrete elements with and without the presence of corrosion in reinforcement (sorted).

	No corrosion	Corrosion
	Potential [mV]	Potential [mV]
1	-221	-532
2	-201	-455
3	-166	-432
4	-146	-373
5	-120	-360
6	-118	-337
7	-113	-312
8	-86	-280
9	-50	-279
10	-5	-260
11		-253
12		-250
13		-233
14		-210
15		-165
16		-163
17		-160

<u>Exercise 1:</u>

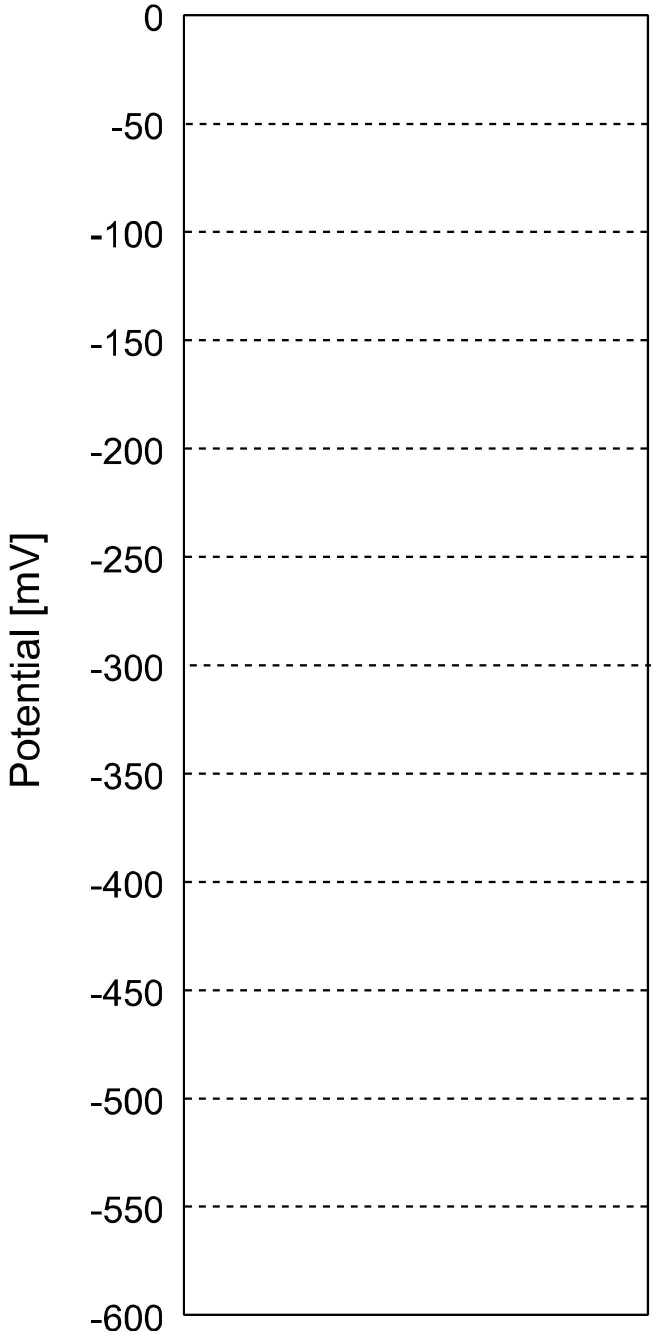
Descriptive statistics **(20 Points)**

A) Assess the sample size, the sample mean, the sample median and the sample standard deviation for the measured potential values for the concrete elements with the presence of corrosion in the reinforcement shown in Table 1 (right column) on page 3. Write your results in the following boxes.

Sample size:		Sample mean:	
Sample median:		Sample standard deviation:	

B) Draw the Tukey box plot in Figure 1 for the case when corrosion is present in the reinforcement, using the data given in Table 1 (right column) on page 3. Calculate the following values and indicate them in the plot: median, upper and lower quantiles, interquartile range, upper and lower adjacent values, and outside values (if there are no outside values, please remark it, too).

Figure 1: Tukey box plot.



Exercise 2:

Bayes' theorem **(20 Points)**

The testing institute that carried out the half-cell potential measurements claims that a measurement of less than -200 [mV] indicates a state of corrosion of the reinforcement.

Let C be the event that the reinforcement is in a true state of corrosion. Let I represent the event that the half cell potential measurement indicates a state of corrosion. $\overline{I}$ represents the complementary set.

Table 2: Probability of indication from the half-cell potential measurements.

True state of corrosion	**Indication from potential measurement**	
	I	$\overline{I}$
C	$P(I \mid C) =$	$P(\overline{I} \mid C) =$
$\overline{C}$	$P(I \mid \overline{C}) =$	$P(\overline{I} \mid \overline{C}) =$

A) Fill out Table 2 by using the data given in Table 1 (on page 3) and using the statement made by the testing institute.

B) The half-cell potential measurement is now carried out to assess the probability of corrosion in another structure. From past experience, it is known that the probability of the reinforcement being in a state of corrosion is 5%.

i) What is the probability that an indication from the half-cell potential measurement **correctly** corresponds to the true state of corrosion?

ii) The half cell potential at a location in the structure is measured as -274 [mV]. What is the probability that the reinforcement is in a true state of corrosion?

(**<u>Note</u>**: Part **B ii)** is independent of part **B i)**)

Exercise 3:

Probability paper **(20 Points)**

The cumulative distribution function for a shifted exponential distribution can be expressed as:

$$F_X(x) = 1 - \exp\left(-\lambda\left(x - \varepsilon\right)\right), \qquad x \geq \varepsilon,\ \ \lambda > 0$$

where λ and ε are the parameters of the distribution.

The suitability of this distribution for the data in the left column of Table 1 (on page 3) is being investigated. A probability paper for the shifted exponential distribution needs to be constructed for this purpose.

A) Define the vertical axis of the probability paper in Figure 2.

B) The data from the left column of Table 1 (on page 3) has already been copied into the right column of Table 3. Fill out the second and third columns in Table 3.

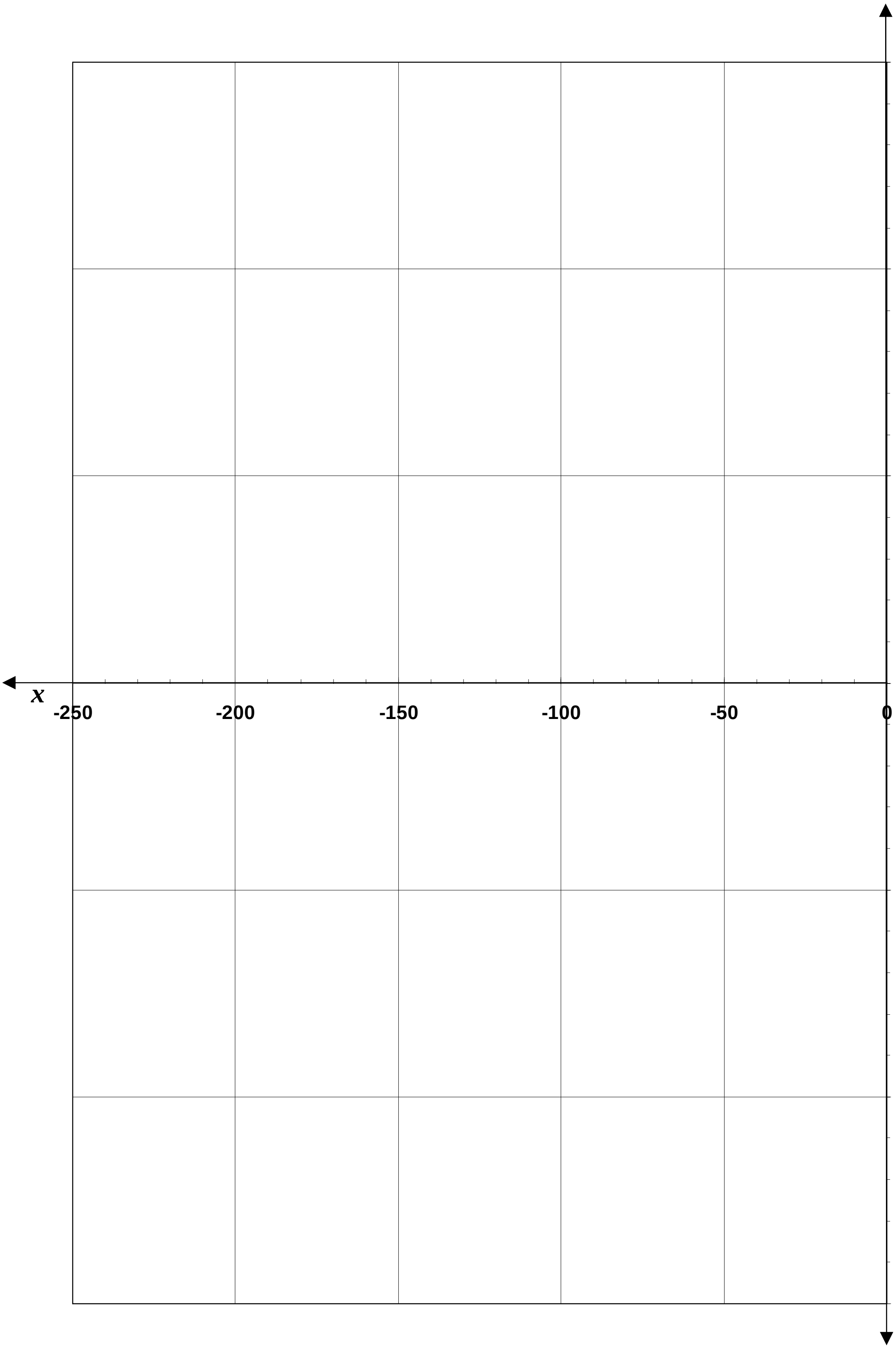
C) Scale the vertical axis of the probability paper in Figure 2 suitably corresponding to the number of observations in Table 3.

D) Check if the data given in the left column of Table 1 can be represented by a shifted exponential distribution. Justify your answer suitably.

Table 3: Calculations for the construction of probability paper.

Observation i	$i/(10+1)$	Vertical axis in Figure 2	Potential x_i in [mV] (sorted)
1			-221
2			-201
3			-166
4			-146
5			-120
6			-118
7			-113
8			-86
9			-50
10			-5

Figure 2: Probability paper for the shifted exponential distribution.



Exercise 4:

Confidence interval and hypothesis testing (20 Points)

A) It is known that the half-cell potential of not corroded samples can be modeled as a Normal distributed random variable X with standard deviation σ_X = 60 [mV]. Based on the known standard deviation σ_X and using the samples in the left column of Table 1, calculate the 90% confidence interval of the mean value of the random variable X.

B) From the first testing institute, you have the measurement results given in Table 1. A second testing institute has been contracted for testing the same concrete elements with the half-cell potential measurements. The objective is to check whether the different devices used by the different institutes lead to a similar, or to a significantly different sample mean.

The measurements from the first testing institute lead to a sample mean $\overline{x_1}$ = -122 [mV] with standard deviation s_{x_1} = 65 [mV] for the not corroded elements. The measurements from the second testing institute lead to a sample mean $\overline{x_2}$ = -130 [mV] with standard deviation s_{x_2} = 65 [mV] for the same not corroded elements.

Can it be said, based on this data, that the mean of the half-cell potential for not corroded elements measured by the two devices is the same at the 5% significance level? Assume a Normal distribution for the data.

C) Assuming the half-cell potential measurements of not corroded elements to be Normal distributed with mean value μ_X = -125 [mV] and standard deviation σ_X = 60 [mV], what is the probability that the potential measurement of a not corroded reinforced concrete element is less than -200 [mV]?

Exercise 5:

Goodness of fit test for a distribution (25 Points)

For the data in the left column of Table 1 (represented here in the first column of Table 4) showing the half-cell potential values measured for concrete elements without the presence of corrosion, the goodness of fit for a Normal distribution with parameters $\mu_X =$ -150 [mV] and $\sigma_X =$40 [mV] shall be checked with a test.

A) Fill out Table 4 completely.

B) Carry out a Kolmogorov-Smirnov goodness of fit test to check at the 10% significance level, if the data of the potential values indicating no presence of corrosion can be represented by the above given Normal distribution.

Table 4: Calculation sheet.

i	x_i : Potential [mV]	$F_o(x_i^o)=\frac{i}{n}$	$F_p(x_i^o)$	$\lvert F_o(x_i^o)-F_p(x_i^o)\rvert$
1	-221			
2	-201			
3	-166			
4	-146			
5	-120			
6	-118			
7	-113			
8	-86			
9	-50			
10	-5			

C) An engineer has established two different models, model I and model II, for the half-cell potential measurement values of not corroded elements shown in Table 1 (left column). She now carries out a Kolmogorov-Smirnov test at the 10% significance level. She also calculates the sample likelihood for both models (see Table 5).

Table 5: Sample likelihoods for the two models I and II.

Model	Sample statistic	Sample likelihood
I	0.331	0.402
II	0.356	0.275

1. Can the engineer accept at the 10% significance level the distribution described by model I and by model II being representative for the half cell potential measurement data in the left column of Table 1? Justify your answer.

2. Which one of the two models is more suitable for modeling the half cell potential measurement data in the left column of Table 1? Justify your answer.

Exercise 6:

Estimation of parameters for a distribution (15 Points)

To estimate the probability p_c of corroded pillars of a reinforced structure, an engineering consulting company is contracted to carry out half-cell potential measurements. Out of 27 samples, 17 pillars are found to be corroded, and 10 pillars are found to be not corroded.

What is the maximum likelihood estimate for p_c ?

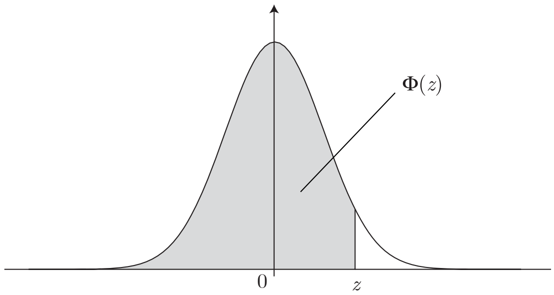
A) Write down the likelihood function to be maximised.

B) Estimate the parameter p_c with the Maximum Likelihood Method, writing down the log-likelihood function, and maximizing the function for the parameter p_c .

Hint: Use the Binomial distribution.

Annex: Table A

Cumulative distribution function of the Standard Normal distribution $\Phi(z)$.



Probability density function of the standard normal random variable.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.10	0.9821356
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.20	0.9860966
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.30	0.9892759
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.40	0.9918025
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.50	0.9937903
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.60	0.9953388
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.70	0.9965330
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.80	0.9974449
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.90	0.9981342
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	3.00	0.9986501
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	3.10	0.9990324
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	3.20	0.9993129
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	3.30	0.9995166
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	3.40	0.9996631
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	3.50	0.9997674
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	3.60	0.9998409
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	3.70	0.9998922
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	3.80	0.9999277
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	3.90	0.9999519
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	4.00	0.9999683
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	4.10	0.9999793
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	4.20	0.9999867
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	4.30	0.9999915
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	4.40	0.9999946
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	4.50	0.9999966
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	4.60	0.9999979
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	4.70	0.9999987
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	4.80	0.9999992
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	4.90	0.9999995
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	5.00	0.9999997
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649		
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656		
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664		
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671		
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678		
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686		
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693		
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699		
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706		
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713		
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719		
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726		
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732		
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738		
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744		
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750		
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756		
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761		
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767		

Annex: Table B

Critical values of the Kolmogorov-Smirnov test.

n	α=0.01	0.02	0.05	0.1	0.2
1	0.995	0.990	0.975	0.950	0.900
2	0.929	0.900	0.842	0.776	0.684
3	0.829	0.785	0.708	0.636	0.565
4	0.734	0.689	0.624	0.565	0.493
5	0.669	0.627	0.563	0.509	0.447
6	0.617	0.577	0.519	0.468	0.410
7	0.576	0.538	0.483	0.436	0.381
8	0.542	0.507	0.454	0.410	0.358
9	0.513	0.480	0.430	0.387	0.339
10	0.489	0.457	0.409	0.369	0.323
11	0.468	0.437	0.391	0.352	0.308
12	0.449	0.419	0.375	0.338	0.296
13	0.432	0.404	0.361	0.325	0.285
14	0.418	0.390	0.349	0.314	0.275
15	0.404	0.377	0.338	0.304	0.266
16	0.392	0.366	0.327	0.295	0.258
17	0.381	0.355	0.318	0.286	0.250
18	0.371	0.346	0.309	0.279	0.244
19	0.361	0.337	0.301	0.271	0.237
20	0.352	0.329	0.294	0.265	0.232
21	0.344	0.321	0.287	0.259	0.226
22	0.337	0.314	0.281	0.253	0.221
23	0.330	0.307	0.275	0.248	0.217
24	0.323	0.301	0.269	0.242	0.212
25	0.317	0.295	0.264	0.238	0.208
26	0.311	0.290	0.259	0.233	0.204
27	0.305	0.284	0.254	0.229	0.200
28	0.300	0.279	0.250	0.225	0.197
29	0.295	0.275	0.246	0.221	0.194
30	0.290	0.270	0.242	0.218	0.190
31	0.285	0.266	0.238	0.214	0.187
32	0.281	0.262	0.234	0.211	0.185
33	0.277	0.258	0.231	0.208	0.182
34	0.273	0.254	0.227	0.205	0.179
35	0.269	0.251	0.224	0.202	0.177
36	0.265	0.247	0.221	0.199	0.174
37	0.262	0.244	0.218	0.196	0.172
38	0.258	0.241	0.215	0.194	0.170
39	0.255	0.238	0.213	0.192	0.168
40	0.252	0.235	0.210	0.189	0.166
$n > 40$	$1.63/\sqrt{n}$	$1.52/\sqrt{n}$	$1.36/\sqrt{n}$	$1.22/\sqrt{n}$	$1.07/\sqrt{n}$

α : Significance level.

n : Sample size.

Glossary

Assumption	Annahme
Appropriate	passend
Column	Spalte
Contract	Vertrag abschliessen
Corrosion	Korrosion
Derive	Herleiten aus
Establish	Aufstellen, errichten
Exceed	Überschreiten, grösser sein als
Experience	Erfahrung
Half-cell potential measurement	Potentialfeldmessung
Pillar	Pfeiler
Postulate	Postulieren, Vorschlagen
Reinforced concrete	Bewehrter Beton/Armierter Beton
State	Zustand
Testing institute	Testinstitut