STOCKHOLM UNIVERSITY
Department of Statistics

2018-03-10, revised 2019-10-15

Basic Statistics for Economists
Plenary Exercise Booklet

These exercises were collected from H. Nyquist's "Statistikens grunder" by Marcus Berg and Annika Tillander, then edited and additional exercises added by M. Carlson and translated by Emma Pettersson.

Plenary Exercise 1

EXERCISE 1

Define the following experiment: “throw three dice and note whether each die was even or odd”. We could define the sample space so that it consists of the following outcomes: “only even numbers”, “only odd numbers”, “two even and one odd”, “one even and two odd”. However, this classification is not appropriate for the use of the classical probability definition, as the different outcomes are not equally likely (why?).

If, on the other hand, we take into account the outcome of each individual die roll, we can describe the sample space with “EEE”, “EEO”, “EOE”,…”OOO” where E = Even and O = Odd on the corresponding die. The probability of each of these composite outcomes are the same. This is only true if we assume that the probability of getting an even or an odd number are the same all the time (is it reasonable this assumption?).

Using the classical probability definition, calculate the probability of the following events:

a. At least two dice are odd,
b. At most two dice are odd,
c. Exactly one dice is odd,
d. Exactly two dice are odd.

EXERCISE 2

My sister has two children. For the sake of simplifying the exercise, let’s say that there are only two genders, girls and boys, let’s also assume that the chances of having a boy and a girl are the same, and finally assume that the children’s genders are independent of one another.

a. What is the probability that both children are girls?
b. If I said that at least one of the children is a girl, what is the conditional probability that both are girls?
c. If I say that the oldest child is a girl, what is the conditional probability that both children are girls?

EXERCISE 3

According to experts, the probability that a certain football team wins their next match is 10%. The probability that the team wins the match after that is 30% and the probability that the team loses both matches is 65%. What is the probability that the team wins exactly one match?

EXERCISE 4

Assume that the figures in the contingency table on the following page show the probabilities for the different outcomes in the experiment: “a person is chosen randomly, and we record whether or not they get cancer as well as whether or not they smoke”.

	Gets cancer	Doesn’t get cancer	Margin
Smoker	0.0054	0.1546	
Non-Smoker	0.0006	0.8394	
Margin			

Calculate the probability that a person gets lung cancer given that the person is (a) a smoker, (b) a non-smoker, and (c) how much bigger the risk of developing lung cancer is for a smoker in comparison to a non-smoker.

EXERCISE 5

Let X be a random variable that can assume the values 0 and 1 with the probabilities 0.3 and 0.7. Calculate the expected value $E(X)$ and the variance $Var(X)$.

EXERCISE 6

Let X be a random variable that can assume the values 0,1,2,3,4 and the probability for each outcome can be calculated with the following:

$$p(x) = \frac{3 - |2 - x|}{9}$$

Verify that this a probability function and present the probability distribution and the cumulative probability distribution. Calculate the expected value and variance.

EXERCISE 7 (MONTY HALL PROBLEM)

On a TV-gameshow, players can choose between 3 doors. Behind one of the doors there is a car, and behind the other two are goats. The game begins when a player selects a door. The object behind the selected door isn’t revealed to the player, however, the host of the show (who knows the location of the car) opens one of the two remaining two doors and reveals one of the goats. The host then gives the player the chance to either change doors or stick with the door he selected at the beginning. Which is the best strategy, should the player change doors or not? Does knowing what is behind one of the doors change anything?

Plenary Exercise 2

EXERCISE 1

The probability that a certain type of light bulb will stay lit for more than 1000 hours is 0.9. Three of those lamps are selected, the probability that the different lamps go out are independent of each other. Calculate the probability that after 1000 hours:

a. All the lamps have gone out,
b. Exactly one lamp stayed lit,
c. At least two lamps stayed lit.

EXERCISE 2

A company does research on two goods A and B. The probability distribution for the demand of the goods are given in the tables below where X and Y represent the quantity demanded of the different goods.

	Quantity demanded of good A						
x	0	1	2	3	4	5	6
$P_x(x)$	0.05	0.07	0.20	0.25	0.30	0.10	0.03

	Quantity demanded of good B						
y	0	1	2	3	4	5	6
$P_y(y)$	0.10	0.12	0.15	0.20	0.25	0.10	0.08

a. Calculate the probability than no products are demanded on a certain day. State the assumptions you need to make in order to do your calculations.
b. Calculate the probability that at least three of good A <u>and</u> at most four of good B are demanded on a certain day.
c. Calculate the probability that at most 3 of good A <u>and</u> at least three of good B are demanded on a certain day.
d. Calculate the probability the exactly three goods are demanded on a certain day <u>and</u> that they are the same type of good.
e. Calculate the expected value $\mu_x = E(X)$, the variance $\sigma_x^2 = Var(X)$ and the standard deviation σ_x for X.
f. If the demand for A and B are independent of each other, what is the correlation between X and Y. If they aren't independent, can the correlation be 0?

EXERCISE 3

A construction company is considering taking on a job that will give them 180 000 SEK in profit if the job is executed in time but where the penalty for missing the deadline is 60 000 SEK.

a. Assume that the probability that the job is completed before deadline is 0.8. Should the company take the job? (Tip: What is the expected profit?)
b. What should the probability that the job is completed in time be so that the expected profit is zero?

EXERCISE 4

A course has two exams; the first is give midcourse, the second at the end of the course. Both exams consist of ten multiple choice questions and with each question four alternatives are given of which only one is correct. A correct answer gives one point. To pass the course a student must have at least 4 points from each of the two exams and at least a total score of 10 points combined.

a. If a student simply guesses the correct answers, what is the probability that the student will pass the course?
b. Given that a student has passed the first exam, what is the probability that the student will pass the course (by guessing the right answers in both exams)?

EXERCISE 5

Assume that the value for small houses in a certain area are normally distributed with mean of 1 400 000 SEK and standard deviation 800 000 SEK. What proportion of the houses in the area will have a value over 2 000 000 SEK?

EXERCISE 6

The weight in grams of a certain type of canned food is normally distributed with a standard deviation of $\sigma = 10$. The mean μ can be regulated by adjusting the settings on the filling machine. Which is the smallest value that μ can take such that only 5% of the cans have a weight under 300g?

Plenary Exercise 3

EXERCISE 1

A company produces and sells paper in packages of 5000 sheets. The weight of a 1 m^2 paper sheet is normally distributed with an expected value of 80 g and a standard deviation of 4 grams. If the weight of the package is less than 399.5 kg, the buyer will receive a discount on the package.

a. What is the probability that the package gets returned?
b. The company has recently sold 1000 packages. How many packages are expected to be returned?
c. A sheet of paper costs 0.3 SEK to produce. One sheet of paper sells for 0.4 SEK. If a package is returned, it is sold at a discounted price of 1200 SEK. What is the expected profit per package?
d. If the company reduces the average weight of a sheet of paper to 79.8 g, the cost of producing one sheet of paper will be reduced to 0.28 kr. Is this a good decision for the company to make?

EXERCISE 2

A pharmaceutical company produces a large variety of tablets and pills. Assume that the weight of a randomly selected tablet (in grams) is a normally distributed random variable with a mean μ and a standard deviation $\sigma = 0.02\ g$. During an inspection, several tablets are randomly selected and weighed. Assume that $\mu = 0.65\ g$.

a. Calculate the probability that the weight of a randomly selected tablet lies in the interval (0.64, 0.66).
b. Calculate the probability that the mean weight of 30 randomly selected tablets lies outside of the interval (0.64, 0.66).
c. How many tablets should be weighed (what sample size is needed) so that the probability that the mean weight of the tablets lies outside the interval (0.64, 0.66) is at most 0.05?

EXERCISE 3

An advertising firm held a campaign to introduce a new product. When the campaign was over, the firm claimed that 20% of the target group knew about the product. To test this claim, the producers of the product took a random sample of 1000 individuals from the target group and found that 150 people knew about the product.

a. Assume the advertising firm's claim is correct. What is the expected value and standard deviation of the number of people in the sample that know about the product?
b. What is the probability that 150 individuals or less in a sample of 1000 individuals know about the product?
c. Is it reasonable for the producers to believe the advertising firm's claims?

Plenary Exercise 4

EXERCISE 1

To measure reaction times, a psychologist uses a measuring instrument that produces measurements which are normally distributed with a standard deviation of $\sigma = 0.05$ seconds. Assume that the psychologist measures the following 5 reaction times during an experiment:

$$1.05 \quad 0.97 \quad 1.02 \quad 1.01 \quad 0.96$$

Calculate confidence intervals for the population mean with $95\%, 99\%$ and 99.9% confidence levels.

EXERCISE 2

Over the last fiscal year, a company had a gross profit which, in relation to its size, was significantly lower than the industry average. It is decided that an audit should be conducted and that the company's sales should be reviewed. A random sample of 100 invoices were drawn and inspected. For 11 of the invoices in the sample, the reported amount was entered incorrectly. In total, the book value was 22 000 SEK less than the invoiced amounts. The accountant, who has an interest in statistics, calculated that the standard deviation for the accounting errors (the difference between the invoiced and book value) was 700 SEK in the sample.

a. Estimate the proportion of invoices which were incorrectly reported. Calculate both a point estimate and a 95% confidence interval.
b. Estimate the average accounting error for the invoices. Give both a point estimate and a 95% confidence interval.

EXERCISE 3

The time delay between the end of the financial year and the publication of audit reports for industrial companies and companies in the financial sector were studied. A random sample of 250 industrial companies had a mean time delay of 65.04 days and a standard deviation of 35.72 days. A random sample of 238 companies in the financial sector had a mean time delay of 56.74 days and a standard deviation of 34.87 days.

a. Calculate a 95% confidence interval for the mean time delay of all industrial companies.
b. Calculate a 95% confidence interval for the mean time delay of companies in the financial sector.
c. Is there evidence to support the claim that the time delays for industrial and financial companies are different? Investigate this by comparing the confidence intervals in questions a and b, as well as calculating a confidence interval for the difference.

EXERCISE 4

In a study about living and well-being, there were questions concerning the fears and behaviours of individuals. One question was about feeling uncomfortable outdoors in the evenings and at night time. A group of individuals were asked this question in February, when 42% of 320 respondents answered “yes”. Another group of individuals were asked the same question in August, then 30% of 200 answered “yes”.

a. Calculate 95% confidence interval for the population proportion of Yes-answers for both groups (P_{feb} and P_{aug}). State the assumptions you must make.
b. How large of a sample is needed for the August group confidence interval to be the same length as the February group confidence interval.

Plenary Exercise 5

EXERCISE 1

When inspecting the weight of twelve 2 kg packages of flour, the average weight in the sample was 1.95 kg. At other inspections, the standard deviation was found to be $\sigma = 0.10\ kg$. The weight of a package of flour is a normally distributed random variable. Is the result from the inspection consistent with the hypothesis that all packages of flour have an average weight of 2 kg? Test this hypothesis using a 5% significance level.

EXERCISE 2

According to previous investigations, 8% of the items produced during a certain production process are defective. After some alterations to the production process, you want to investigate if there has been a change in the proportion of defective items. A random sample of 160 items were chosen. 10 of these items were found to be defective. Is it reasonable to believe that the new production process has had any effect (or did the process's average error rate remain at 8%)? The risk of wrongly concluding that the new process has had an effect on the error rate should not exceed 5%.

EXERCISE 3

A company wants to test the 'average life span' of a cutting tool. Specifically, they want to test whether the tool can be used for cutting a certain type of object 3000 times, or if it can only be used for less than 3000 cuts. Six observations were recorded: 2970, 3020, 3005, 2900, 2940, and 2925. Which conclusions can be drawn at the 5% significance level? Assume that life spans are independent and normally distributed.

EXERCISE 4

The head of a Bakery company called Rokak has as an objective to get at least 40% of the customers in the market to be aware of their product "Rokak's Cookies".

a. The head of the company feels that the company's marketing is unsatisfactory. If he/she finds evidence that less than 40% are aware of "Rokak's Cookies", a drastic reorganisation of the marketing department will take place. Market research is being conducted to investigate the markets awareness of the cookies. From a sample of 500 randomly selected individuals in the market, 180 knew about "Rokak's Cookies". Use a statistical test to help the head of the company make their decision. State and justify the assumptions you use in order to perform the test.
b. Soon after the investigation, a new marketing campaign for the product is launched. Later on, a new market survey was conducted where 195 of 500 randomly drawn individuals knew about the products name. Use a statistical hypothesis test to see if the campaign had an effect and improved customer awareness of the product.

Plenary Exercise 6

EXERCISE 1

The solutions to the exercises a) – h) should be calculated by hand, the summary output from Excel on the following page should just be used to double check your calculations.

A nationwide retail chain sells stereo packages. For a certain type of stereo, retailers are allowed to set their own prices. The company's management team wants to get an idea of how the demand for that stereo package is affected by price. They take a random sample of eight retail stores and collect information about prices and quantities sold in the previous month. The table below shows the results of the investigation.

No.	***x* = price (1000kr)**	***y = quantity sold***
1	5.9	41
2	6.2	38
3	7.2	34
4	6.3	40
5	5.2	44
6	7.1	36
7	4.9	45
8	5.2	42

	***x* = price (1000kr)**	***y = quantity sold***
Sample size	8	8
Mean	6	40
Standard Deviation	0.868496237	3.817254062
Minimum	4.9	34
Maximum	7.2	45
Sum	48	320
Sum of squares	293.28	12902

Assume that the relationship between quantity demanded and price can be described with the model:

$$y = \beta_0 + \beta_1 x + \varepsilon \,, \varepsilon \sim N(0, \sigma^2)$$

a. Present the data in a suitable diagram.
b. Explain what the parameter β_1 stands for in the model.

c. Estimate the parameters β_0 and β_1 with the ordinary least squares method and interpret the estimate of β_1.
d. Calculate SSE.
e. Calculate and interpret the coefficient of determination.
f. Test $H_0: \beta_1 = 0$ against $H_A: \beta_1 < 0$. Use the 0.05 significance level.
g. Assume that another retailer sets their price at 6000 kr. Estimate a 90% prediction interval for the retailer's sales using your estimated model.
h. Calculate the correlation between x and y. What would happen if we measured price in kronor (instead of 1000 kr)? What would happen to the estimate of β_1?

Summary Output:

Regression Statistics	*Column1*
Multiple R	0.965230355
R Square	0.931669638
Adjusted R Square	0.920281244
Standard Error	1.077782984
Observations	8

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	95.03030303	95.0303	81.808696	0.000102364
Residual	6	6.96969697	1.161616		
Total	7	102			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	65.45454545	2.839949333	23.04779	4.372E-07	58.50543978	72.40365113
x=price (1000kr)	-4.24242424	0.469044839	-9.04482	0.0001024	-5.390135617	-3.09471286

EXERCISE 2

A delivery company has measured the travel time (y, number of hours), the driving distance (X_1, miles) and the number of deliveries (X_2) for $n = 10$ drivers in a day. On the following couple of pages you will find a data set, descriptive statistics and regression estimates for the following 2 models:

Model 1: $y = \beta_0 + \beta_1 X_1 + \varepsilon$

Model 2: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

a. Estimate a 95% confidence interval for the average change in travel time when the driving distance increases by 1 mile.
b. Same as a. but under the condition that all the number of deliveries is held constant.
c. Test whether model 2 can help explain the variation in y.
d. Test whether model 2 is better than model 1 at explaining the variation in y.
e. Calculate the adjusted coefficient of determination for model 2.
f. Estimate the expected travel time for a driver that has the task of driving 15 miles and making 3 deliveries.

No.	$\boldsymbol{y}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$
1	9.3	20	4
2	4.8	10	3
3	8.9	20	4
4	6.5	20	2
5	4.2	10	2
6	6.2	16	2
7	7.4	15	3
8	6	13	4
9	7.6	18	3
10	6.1	18	2

	$\boldsymbol{y}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$
Size	10	10	10
Mean	6.7	16	2.9
Standard Deviation	1.629587542	3.915780041	0.875595
Minimum	4.2	10	2
Maximum	9.3	20	4
Sum	67	160	29
Sum of squares	472.8	2698	91

Summary Output Model 1:

Regression Statistics	*Column1*
Multiple R	0.81490571
R Square	0.66407131
Adjusted R Square	0.62208023
Standard Error	1.00179187
Observations	10

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	1	15.8713043	15.8713043	15.8145781	0.004080177
Residual	8	8.02869565	1.00358696		
Total	9	23.9			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.27391304	1.40074452	0.90945424	0.38968736	-1.956209623	4.50403571
x_1	0.33913043	0.08527819	3.97675473	0.00408018	0.142478582	0.53578229

Summary Output Model 2:

Regression Statistics	*Column1*
Multiple R	0.950678166
R Square	0.903788975
Adjusted R Square	-
Standard Error	0.573142152
Observations	10

ANOVA	*df*	*SS*	*MS*	*F*	*Significance F*
Regression	2	21.60055651	10.80028	32.878367	0.00027624
Residual	7	2.299443486	0.328492		
Total	9	23.9			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.868701467	0.951547725	-0.91294	0.3916343	-3.118754293	1.38135136
x_1	0.305672994	0.049442473	6.182397	0.000453	0.188760124	0.422585864
x_2	0.923425367	0.221113461	4.176251	0.0041566	0.400575115	1.446275618

Plenary Exercise 7

EXERCISE 1

A psychologist performed the following experiment. The speed of some cars on a road (with hidden measuring instruments) was recorded in two places a few kilometres apart. At first, the psychologist found that the cars were travelling at around the same speed at both the locations. Later on, a police car was parked in between the two locations. The speed (in km/h) of 12 cars at the two locations was then measured, the results are shown in the table below:

Car no.	1	2	3	4	5	6	7	8	9	10	11	12
Speed before police car	83	88	106	131	84	87	129	97	92	91	124	99
Speed after police car	61	84	95	121	83	79	92	88	69	90	115	100

The psychologist's hypothesis was that the presence of the police car would make drivers slow down. Propose a suitable statistical test, state the assumptions for the test and perform the test at the 5% significance level.

EXERCISE 2

The braking power of two different car models was compared. A random sample of 64 cars was drawn for each type of car. The braking distance (in meters) required for the cars to stop was measured (the cars were driving at 65 km/h). After the test, the following statistics were calculated:

	Car 1	Car 2
Sample Size	64	64
Mean	29.5	27.25
Variance	6.37	5.44

Can the test provide sufficient statistical evidence that there is a difference in the average breaking power of the two car models? Set up an appropriate hypothesis test to answer the question. State the necessary assumptions and perform the test at the 5% significance level. What conclusions can you draw?

EXERCISE 3

According to a hypothesis, half of all employees who work at a certain company drive to and from work. Of the others who don't take their cars, 2/3 use public transportation and 1/3 cycle. To test this hypothesis, a random sample of $n = 50$ workers who worked at the company were selected. The results showed that 22 employees took their cars and 14 used public transport. Test the hypothesis at the 0.05 significance level.

EXERCISE 4

An experiment was performed on an interview survey conducted annually, researchers wanted to investigate whether or not non-response was affected by the data collection method. The large sample was divided into 3 smaller groups where different interview methods were used. The methods were: CAPI (computer assisted telephone interview), PAPI (paper assisted personal interview), and MM (mixed mode, respondents could choose between CAPI and PAPI). Of the 1208 in the sample, the following was observed:

$\boldsymbol{O_{ij}}$	MM	CATI	PAPI	Sum
Response	264	297	320	881
Non-Response	110	72	145	327
Sum	374	369	465	1208

Is non-response affected by the data collection method? Test this using the 0.05 significance level.

EXERCISE 5

To investigate how users of different phones experienced a new app, a sample of $n = 200$ individuals was selected. Respondents were asked to answer the following two questions:

1. Did you like the app (Yes/No)?
2. Do you use an iPhone (yes/No)?

The following results were found:

- 80 answered “Yes” to both questions
- 15 answered “No” to both questions
- 80 answered "No" to the first question and “Yes” to the second.
- 25 answered "Yes" to the first question and “No” to the second

We want to see if the answer to the first question depends on whether a person is an iPhone user or not. Test the hypothesis using the 5% significance level.

EXERCISE 6

The table below, quarterly data for employment in Sweden (for the age group 15-74, in 1000s, 2014-2016) are shown in the table below.

	Q1	Q2	Q3	Q4
2014	4 119.8	4 239.7	4 329.6	4 224.5
2015	4 194.5	4 296.7	4 378.5	4 316.5
2016	4 284.8	4 412.3	4 443.4	4 389.2

a. verify that the seasonal indices when using a multiplicative model are

$$S_1 = 0.98261 \qquad S_2 = 1.00515 \qquad S_3 = 1.01836 \qquad S_4 = 0.99388$$

(see NCT p. 692-697).

b. Calculate the seasonally adjusted series using the seasonal indices in a)

c. Illustrate both the original time series and the seasonally adjusted series in the same chart and comment. What do you see?

ANSWERS

PLENARY EXERCISE 1

Exercise 1: a) 1/2 b) 7/8 c) 3/8 d) 3/8

Exercise 2: a) 1/4 b) 1/3 c) 1/2

Exercise 3: 0,30

Exercise 4: a) 0.03375 b) 0.000714 c) 47.25

Exercise 5: $E(X) = 0.7$ $V(X) = 0.21$

Exercise 6: $P(x) > 0$ for all x and $\sum_{x \in S} P(x) = 1$ $E(X) = 2$ $V(X) = 1.333$

Exercise 7: See https://en.wikipedia.org/wiki/Monty_Hall_problem

PLENARY EXERCISE 2

Exercise 1: a) 0.001 b) 0.027 c) 0.972

Exercise 2: a) 0.005 b) 0.5576 c) 0.3591 d) 0.035

e) $E(X) = 3.1$ $Var(X) = 1{,}89$ f) 0 and yes

Exercise 3: a) Yes b) 1/4

Exercise 4: a) 0.01186 b) 0.05294

Exercise65: 0.227

Exercise 7: $\mu \geq 316{,}449 \approx 316$

PLENARY EXERCISE 3

Exercise 1: a) 0.0384 b) ≈ 39 c) 469.28

d) $E(profit) = -169.28 \Rightarrow$ Don't change

Exercise 2: a) $P(-0.5 < Z < 0.5) = 0.3829$

b) $1 - P(-2.7 < Z < 2.7) = 0.007$

Exercise 3: a) $E(X) = 200$ $V(X) = 160 \Rightarrow$ standard deviation $\sigma \approx 12.65$

b) $P(X \leq 150) \approx 0$ c) No

PLENARY EXERCISE 4

Exercise 1: Confidence intervals for reaction times (in seconds)

95 %: $(0{,}958, 1{,}046)$

99 %: $(0{,}944, 1{,}060)$

99,9 %: $(0{,}928, 1{,}076)$

Exercise 2: a) Point estimate: 0.11 — 95 % confidence interval: $(0.049, 0.171)$

b) Point estimate: 220 SEK — 95 % confidence interval $(82.8\,, 357.2)$

Exercise 3: a) CI för μ_{Ind}: $(60.61, 69.47)$ — b) CI för μ_{Fin}: $(52.31;\ 61.17)$

c) The confidence intervals in a) and b) overlap, this indicates that it is possible that $\mu_{Ind} = \mu_{Fin}$ However, the confidence interval for the difference $\mu_{Ind} - \mu_{Fin}$ is $(2.036, 14.56)$, this indicates that the difference is greater than zero. Which one is right? Why?

Exercise 4: a) p_{feb}: $(36.6\%\,.\,47.4\%)$ — p_{aug}: $(23.6\%\,.\,36.4\%)$

b) $n_{aug} = 276$

PLENARY EXERCISE 5

Exercise 1: $z_{obs} = -1.73 \Rightarrow$ Do not reject H_0

Exercise 2: $z_{obs} = -0.815 \Rightarrow$ Do not reject H_0

Exercise 3: $t = -2.094 \Rightarrow$ Reject H_0

Exercise 4: a) $z_{obs} = -1.83 \Rightarrow$ Reject H_0

b) $z_{obs} = 0.98 \Rightarrow$ Do not reject H_0 (or $z_{obs} = -0.98$)

PLENARY EXERCISE 6

Exercise 1: a) scatterplot; there is a clear negative linear relationship, no deviating observations

b) β_1 is the expected increase in sales (y) when the price (x) increases by $1000\ kr$

c) $\widehat{\beta_0} = 65.455$ and $\widehat{\beta_1} = -4.2424$ — d) 6.97

e) $R^2 = 0.932$ — f) $t_{obs} = -9.04 \Rightarrow$ Reject H_0

g) $(37.78, 42.22)$

h) $r_{xy} = -\sqrt{R^2} = -0.9654$. The correlation is independent of scale, so it does change. However, the regression coefficient is scale dependant so:

$$\widehat{\beta_1}^{(kr)} = \frac{\widehat{\beta_1}^{(1000kr)}}{1000} = -0.00424$$

Exercise 2: a) $(0.142, 536)$ b) $(0.189, 0.423)$

c) $F_{obs} = 32.88 \Rightarrow$ Reject H_0 d) $|T_{obs}| = 4.18 \Rightarrow$ Reject H_0

e) $R^2_{adj} = 0.877$ f) $(6.037, 6.936)$

PLENARY EXERCISE 7

Exercise 1: Depending on how you calculated the differences, $t_{obs} = 3.49$ or $t_{obs} = -3.49 \Rightarrow$ reject H_0

Exercise 2: $z_{obs} = 5.24 \Rightarrow$ reject H_0

Exercise 3: $\chi^2_{obs} = 4.64 \Rightarrow$ cannot reject H_0

Exercise 4: $\chi^2_{obs} = 15.7 \Rightarrow$ reject H_0

Exercise 5: $\chi^2_{obs} = 2.005$ or $|z_{obs}| = 1.416 \Rightarrow$ cannot Reject H_0

Exercise 6: a) See solutions

b) Seasonal adjusted series:

	Q1	Q2	Q3	Q4
2014	4 192.70	4 218.00	4 251.53	4 250.52
2015	4 268.72	4 274.70	4 299.55	4 343.09
2016	4 360.62	4 389.71	4 363.28	4 416.23

c) Times series chart: