STOCKHOLM UNIVERSITY
Department of Statistics
Spring 2020
Ulf Högnäs (examiner)

EXAM – BASIC STATISTICS FOR ECONOMISTS

2020-08-17

Time:	15.00-21.00
Approved aid:	Any books or notes. You are not allowed to communicate with anyone else during the exam. You are allowed to use Excel or other software to check your work, but for the complete written solutions, you have to show calculations on paper.

NOTE: You will only be required to solve 7 out of 28 problems. Which 7 problems that you are asked to solve is specified on the next pages.

- **Problems 1 – 20 MULTIPLE CHOICE QUESTIONS – max 50 points**
  - A total of 12 multiple choice questions with five alternative answers per question one of which is the correct answer. Mark your answers on the attached answer form. If you prefer, you can make a handwritten version, but please make it clear.
  - Mark exactly one answer and do not provide written solution.

- **Problems 21 – 28: COMPLETE WRITTEN SOLUTIONS – max 50 points**
  - For full marks, clear, comprehensive and well-motivated solutions are required. Unclear and unexplained solutions may result in point deductions even if the final answer is correct.
  - Check your calculations and solutions before submitting. Careless mistakes may result in unnecessary point deductions.

- The maximum number of points is stated for each question. The maximum total number of points is 50 + 50 = 100. At least 50 points is required to pass (grades A-E). The grading scale may be adjusted toward more generous grades:

  A: 90 – 100 points

  B: 80 – 89 points

  C: 70 – 79 points

  D: 60 – 69 points

  E: 50 – 59 points

  Fx: 40 – 49 points

  F: 0 – 40 points

  NOTE! Fx and F are failing grades that require re-examination. Students who receive the grade Fx or F cannot supplement for a higher grade.

- Solutions will be posted on Athena after the exam. **GOOD LUCK!**

Find your anonymous code in the table. Solve the problems listed on that row. Be careful to answer those and only those problems.

code	1-4	5-8	9-12	13-16	17-20	21-24	25-28
0001-AKT	1	5	9	13	17	21	25
0004-EUZ	2	6	10	14	18	22	26
0005-DCY	3	7	11	15	19	23	27
0006-XTL	4	8	12	16	20	24	28
0007-CHE	1	5	9	13	17	21	25
0008-SDL	2	6	10	14	18	22	26
0009-NNM	3	7	11	15	19	23	27
0010-PHR	4	8	12	16	20	24	28
0011-ALP	1	5	9	13	17	21	25
0012-OMH	2	6	10	14	18	22	26
0013-HKY	3	7	11	15	19	23	27
0014-CPP	4	8	12	16	20	24	28
0016-EGD	1	5	9	13	17	21	25
0017-YJM	2	6	10	14	18	22	26
0018-YLG	3	7	11	15	19	23	27
0019-FRF	4	8	12	16	20	24	28
0020-ANK	1	5	9	13	17	21	25
0022-SGG	2	6	10	14	18	22	26
0023-CAY	3	7	11	15	19	23	27
0024-XBF	4	8	12	16	20	24	28
0025-OEX	1	5	9	13	17	21	25
0026-WRZ	2	6	10	14	18	22	26
0027-NPE	3	7	11	15	19	23	27
0028-XOH	4	8	12	16	20	24	28
0029-KBT	1	5	9	13	17	21	25
0030-GFY	2	6	10	14	18	22	26
0031-GUR	3	7	11	15	19	23	27
0032-SAW	4	8	12	16	20	24	28
0033-GLZ	1	5	9	13	17	21	25
0034-MGF	2	6	10	14	18	22	26
0035-FBY	3	7	11	15	19	23	27
0036-CKC	4	8	12	16	20	24	28
0037-BOZ	1	5	9	13	17	21	25
0038-YFE	2	6	10	14	18	22	26
0039-EZC	3	7	11	15	19	23	27
0040-XPH	4	8	12	16	20	24	28
0041-SZY	1	5	9	13	17	21	25
0042-NEP	2	6	10	14	18	22	26
0043-DXE	3	7	11	15	19	23	27
0046-XRH	4	8	12	16	20	24	28
0047-EKS	1	5	9	13	17	21	25
0048-UNN	2	6	10	14	18	22	26
0049-EBC	3	7	11	15	19	23	27
0050-ZBS	4	8	12	16	20	24	28
0051-MWE	1	5	9	13	17	21	25
0052-DPF	2	6	10	14	18	22	26
0053-DUK	3	7	11	15	19	23	27

0054-JCS	4	8	12	16	20	24	28
0056-PGT	1	5	9	13	17	21	25
0057-PTL	2	6	10	14	18	22	26
0058-MWR	3	7	11	15	19	23	27
0059-MRO	4	8	12	16	20	24	28
0060-OFG	1	5	9	13	17	21	25
0061-TYN	2	6	10	14	18	22	26
0062-TTW	3	7	11	15	19	23	27
0063-DSK	4	8	12	16	20	24	28
0065-DLB	1	5	9	13	17	21	25
0066-DOJ	2	6	10	14	18	22	26
0067-CAA	3	7	11	15	19	23	27
0068-MMU	4	8	12	16	20	24	28
0069-RFL	1	5	9	13	17	21	25
0071-OML	2	6	10	14	18	22	26
0072-OKK	3	7	11	15	19	23	27
0073-AZJ	4	8	12	16	20	24	28
0075-KCK	1	5	9	13	17	21	25
0076-MXT	2	6	10	14	18	22	26
0077-WJJ	3	7	11	15	19	23	27
0078-GSL	4	8	12	16	20	24	28
0079-SKG	1	5	9	13	17	21	25
0080-PMU	2	6	10	14	18	22	26
0081-NOW	3	7	11	15	19	23	27
0082-MYB	4	8	12	16	20	24	28
0083-NJN	1	5	9	13	17	21	25
0085-KGJ	2	6	10	14	18	22	26
0086-BNP	3	7	11	15	19	23	27
0087-UJH	4	8	12	16	20	24	28
0088-BWA	1	5	9	13	17	21	25
0089-LNL	2	6	10	14	18	22	26
0091-OJY	3	7	11	15	19	23	27
0092-GPH	4	8	12	16	20	24	28
0093-OOG	1	5	9	13	17	21	25
0094-LWC	2	6	10	14	18	22	26
0095-YCL	3	7	11	15	19	23	27
0096-BJO	4	8	12	16	20	24	28
0097-AGD	1	5	9	13	17	21	25
0098-UFA	2	6	10	14	18	22	26
0099-JJM	3	7	11	15	19	23	27
0100-OYZ	4	8	12	16	20	24	28
0101-DMB	1	5	9	13	17	21	25
0102-YLH	2	6	10	14	18	22	26
0104-HFM	3	7	11	15	19	23	27
0105-UHH	4	8	12	16	20	24	28
0106-RNU	1	5	9	13	17	21	25
0107-CDK	2	6	10	14	18	22	26
0108-LRN	3	7	11	15	19	23	27
0109-BNB	4	8	12	16	20	24	28
0110-YJW	1	5	9	13	17	21	25
0111-SXG	2	6	10	14	18	22	26
0112-XBM	3	7	11	15	19	23	27

0113-BRZ	4	8	12	16	20	24	28
0114-HSZ	1	5	9	13	17	21	25
0115-XUE	2	6	10	14	18	22	26
0116-AHT	3	7	11	15	19	23	27
0117-YKK	4	8	12	16	20	24	28
0118-WNF	1	5	9	13	17	21	25
0119-KHX	2	6	10	14	18	22	26
0120-JBM	3	7	11	15	19	23	27
0121-ZYT	4	8	12	16	20	24	28
0122-LUC	1	5	9	13	17	21	25
0123-DWK	2	6	10	14	18	22	26
0125-ZWF	3	7	11	15	19	23	27
0126-BJN	4	8	12	16	20	24	28
0127-EWU	1	5	9	13	17	21	25
0128-HSY	2	6	10	14	18	22	26
0129-AWW	3	7	11	15	19	23	27
0130-MXE	4	8	12	16	20	24	28
0131-UWO	1	5	9	13	17	21	25
0132-MDG	2	6	10	14	18	22	26
0133-LGG	3	7	11	15	19	23	27
0134-ONU	4	8	12	16	20	24	28
0135-UAS	1	5	9	13	17	21	25
0136-OAY	2	6	10	14	18	22	26
0138-XPD	3	7	11	15	19	23	27
0139-RMP	4	8	12	16	20	24	28
0140-YRD	1	5	9	13	17	21	25
0142-FER	2	6	10	14	18	22	26
0143-KTW	3	7	11	15	19	23	27
0144-HBF	4	8	12	16	20	24	28
0145-CXX	1	5	9	13	17	21	25
0146-KFF	2	6	10	14	18	22	26
0147-BNN	3	7	11	15	19	23	27
0148-XOM	4	8	12	16	20	24	28
0149-CPG	1	5	9	13	17	21	25
0150-BKE	2	6	10	14	18	22	26
0151-ERR	3	7	11	15	19	23	27
0152-OBB	4	8	12	16	20	24	28
0154-JUX	1	5	9	13	17	21	25
0155-BNM	2	6	10	14	18	22	26
0156-AKA	3	7	11	15	19	23	27
0157-JNH	4	8	12	16	20	24	28
0158-OPF	1	5	9	13	17	21	25
0159-LUN	2	6	10	14	18	22	26
0160-JNO	3	7	11	15	19	23	27
0161-JWB	4	8	12	16	20	24	28
0162-DMM	1	5	9	13	17	21	25
0163-KCD	2	6	10	14	18	22	26
0164-LAF	3	7	11	15	19	23	27
0165-HWT	4	8	12	16	20	24	28
0166-YGH	1	5	9	13	17	21	25
0167-FRM	2	6	10	14	18	22	26
0169-AZN	3	7	11	15	19	23	27

0170-NJO	4	8	12	16	20	24	28
0171-FAX	1	5	9	13	17	21	25
0172-AXL	2	6	10	14	18	22	26
0173-AWP	3	7	11	15	19	23	27
0174-LEF	4	8	12	16	20	24	28
0175-BRS	1	5	9	13	17	21	25
0176-ZNN	2	6	10	14	18	22	26
0177-ZLK	3	7	11	15	19	23	27
0178-XFZ	4	8	12	16	20	24	28
0179-AXH	1	5	9	13	17	21	25
0180-JWX	2	6	10	14	18	22	26
0181-KOD	3	7	11	15	19	23	27
0182-PYU	4	8	12	16	20	24	28
0183-NLZ	1	5	9	13	17	21	25
0184-MAE	2	6	10	14	18	22	26
0185-NZH	3	7	11	15	19	23	27
0186-PMB	4	8	12	16	20	24	28
0187-UZK	1	5	9	13	17	21	25
0188-DWR	2	6	10	14	18	22	26
0189-ACH	3	7	11	15	19	23	27
0190-LSR	4	8	12	16	20	24	28
0191-DMA	1	5	9	13	17	21	25
0192-PRU	2	6	10	14	18	22	26
0193-ZYC	3	7	11	15	19	23	27
0195-NBY	4	8	12	16	20	24	28
0196-KXZ	1	5	9	13	17	21	25
0197-ONK	2	6	10	14	18	22	26
0198-PPS	3	7	11	15	19	23	27
0199-PGH	4	8	12	16	20	24	28
0200-TYF	1	5	9	13	17	21	25
0201-RLC	2	6	10	14	18	22	26
0202-LXW	3	7	11	15	19	23	27
0203-RRA	4	8	12	16	20	24	28
0204-BCP	1	5	9	13	17	21	25
0205-YXE	2	6	10	14	18	22	26
0206-UPE	3	7	11	15	19	23	27
0207-ANA	4	8	12	16	20	24	28
0208-OOH	1	5	9	13	17	21	25
0209-UOR	2	6	10	14	18	22	26
0210-CXF	3	7	11	15	19	23	27
0211-BRO	4	8	12	16	20	24	28

Answer form for multiple choice. You can make your own form, put please be clear.

Number	Part	A	B	C	D	E
	a.					
	b.					
	a.					
	b.					
	a.					
	b.					
	a.					
	b.					
	a.					
	b.					

QUESTION 1

The tables below show the percentage of eligible voters who voted in the election for Swedish parliament 2014, by age category and sex. It also shows the total number of eligible voters, in thousands, for each category.

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
men	18-24	79,3	420
	25-29	78,9	290
	30-34	82,5	269
	35-39	85,3	276
	40-44	87,0	307
	45-49	85,5	321
	50-54	86,7	300
	55-59	86,9	281
	60-64	88,2	277
	65-69	91,6	272
	70-74	91,3	250
	75-79	87,4	152
	80+	80,8	199
	total	**85,2**	**3614**

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
women	18-24	83,3	393
	25-29	84,0	286
	30-34	85,4	265
	35-39	86,8	267
	40-44	89,0	302
	45-49	88,2	320
	50-54	88,8	292
	55-59	91,0	279
	60-64	90,8	277
	65-69	92,2	281
	70-74	90,5	259
	75-79	86,7	178
	80+	69,5	316
	total	**86,4**	**3715**

a. What percentage of eligible male voters age 18-39 voted in the election? Choose the alternative closest to your own answer. (5p)

(A) 79,9%
(B) 81,2%
(C) 81,5%
(D) 81,9%
(E) 82,1%

b. Find the interval that contains the 25th percentile for age among women who voted. *Tip: First calculate the number of voters in each age category. Then find the cumulative frequencies and the relative cumulative frequencies.* (5p)

(A) 18-24

(B) 25-29

(C) 30-34

(D) 35-39

(E) 40-44

QUESTION 2

The tables below show the percentage of eligible voters who voted in the election for Swedish parliament 2014, by age category and sex. It also shows the total number of eligible voters, in thousands, for each category.

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
men	18-24	79,3	420
	25-29	78,9	290
	30-34	82,5	269
	35-39	85,3	276
	40-44	87,0	307
	45-49	85,5	321
	50-54	86,7	300
	55-59	86,9	281
	60-64	88,2	277
	65-69	91,6	272
	70-74	91,3	250
	75-79	87,4	152
	80+	80,8	199
	total	**85,2**	**3614**

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
women	18-24	83,3	393
	25-29	84,0	286
	30-34	85,4	265
	35-39	86,8	267
	40-44	89,0	302
	45-49	88,2	320
	50-54	88,8	292
	55-59	91,0	279
	60-64	90,8	277
	65-69	92,2	281
	70-74	90,5	259
	75-79	86,7	178
	80+	69,5	316
	total	**86,4**	**3715**

a. What percentage of eligible female voters age 70 and older voted in the election? Choose the alternative closest to your own answer. (5p)
   (A) 80,8%
   (B) 81,0%
   (C) 81,5%
   (D) 82,2%
   (E) 82,4%

b. Find the interval that contains the 50th percentile for age among men who voted. *Tip: First calculate the number of voters in each age category. Then find the cumulative frequencies and the relative cumulative frequencies.* (5p)

(A) 35-39

(B) 40-44

(C) 45-49

(D) 50-54

(E) 55-59

QUESTION 3

The tables below show the percentage of eligible voters who voted in the election for Swedish parliament 2014, by age category and sex. It also shows the total number of eligible voters, in thousands, for each category.

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
men	18-24	79,3	420
	25-29	78,9	290
	30-34	82,5	269
	35-39	85,3	276
	40-44	87,0	307
	45-49	85,5	321
	50-54	86,7	300
	55-59	86,9	281
	60-64	88,2	277
	65-69	91,6	272
	70-74	91,3	250
	75-79	87,4	152
	80+	80,8	199
	total	**85,2**	**3614**

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
women	18-24	83,3	393
	25-29	84,0	286
	30-34	85,4	265
	35-39	86,8	267
	40-44	89,0	302
	45-49	88,2	320
	50-54	88,8	292
	55-59	91,0	279
	60-64	90,8	277
	65-69	92,2	281
	70-74	90,5	259
	75-79	86,7	178
	80+	69,5	316
	total	**86,4**	**3715**

a. What percentage of eligible male voters age 70 and older voted in the election? Choose the alternative closest to your own answer. (5p)
   (A) 86,1%
   (B) 86,3%
   (C) 86,5%
   (D) 86,8%
   (E) 87,0%

b. Find the interval that contains the 50th percentile for age among women who voted. *Tip: First calculate the number of voters in each age category. Then find the cumulative frequencies and the relative cumulative frequencies.* (5p)
   (A) 35-39
   (B) 40-44
   (C) 45-49
   (D) 50-54
   (E) 50-59 *<- I meant to write 55-59, but 50-59 is correct since 50-54 is correct*

QUESTION 4

The tables below show the percentage of eligible voters who voted in the election for Swedish parliament 2014, by age category and sex. It also shows the total number of <u>eligible</u> voters, in thousands, for each category.

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
men	18-24	79,3	420
	25-29	78,9	290
	30-34	82,5	269
	35-39	85,3	276
	40-44	87,0	307
	45-49	85,5	321
	50-54	86,7	300
	55-59	86,9	281
	60-64	88,2	277
	65-69	91,6	272
	70-74	91,3	250
	75-79	87,4	152
	80+	80,8	199
	total	**85,2**	**3614**

		Voted, percent	Number of eligible voters, 1000s
		2014	2014
women	18-24	83,3	393
	25-29	84,0	286
	30-34	85,4	265
	35-39	86,8	267
	40-44	89,0	302
	45-49	88,2	320
	50-54	88,8	292
	55-59	91,0	279
	60-64	90,8	277
	65-69	92,2	281
	70-74	90,5	259
	75-79	86,7	178
	80+	69,5	316
	total	**86,4**	**3715**

a. What percentage of eligible <u>female voters age 18-39</u> voted in the election? Choose the alternative closest to your own answer. (5p)

(A) 84,0%
(B) 84,2%
(C) 84,7%
(D) 84,9%
(E) 85,1%

b. Find the interval that contains the <u>25^{th} percentile by age</u> among men who voted. *Tip: First caculate the number of voters in each age category. Then find the cumulative frequencies and the relative cumulative frequencies.* (5p)

   (A) 18-24
   (B) 25-29
   (C) 30-34
   (D) 35-39
   (E) 40-44

QUESTION 5

A pizzeria sells pizzas for €10 each. Suppose that the number of pizzas sold per week is approximately normally distributed with mean 1500 and standard deviation 180.

The pizzeria also sells cans of soda for €2 each. Suppose that the number of sodas sold per week is normally distributed with mean 1200 and standard deviation 150. Suppose that the correlation between number of pizzas sold and number of sodas sold is 0.9.

a. Find the standard deviation of the total revenue from pizza and soda (total sales in Euros). Choose the alternative closest to your answer. (5p)
   (A) 1825
   (B) 1884
   (C) 1941
   (D) 1953
   (E) 2074

The pizzeria sells only two variations of pizza: *Margarita* and *Marinara*. The Marinara pizza comes without cheese, so only 25% of customers prefer Marinara, while 75% of customers prefer Margarita.

b. Suppose that we randomly select 20 customers. What is the probability that more than 5 out of these 20 customers prefer the Marinara pizza? (5p)
   (A) 33%
   (B) 38%
   (C) 41%
   (D) 59%
   (E) 62%

QUESTION 6

A pizzeria sells pizzas for €20 each. Suppose that the number of pizzas sold per week is approximately normally distributed with mean 1500 and standard deviation 180.

The pizzeria also sells cans of soda for €2 each. Suppose that the number of sodas sold per week is normally distributed with mean 1200 and standard deviation 150. Suppose that the correlation between number of pizzas sold and number of sodas sold is 0.8.

a. Find the standard deviation of the total revenue from pizza and soda (total sales in Euros). Choose the alternative closest to your answer. (5p)
   (A) 3649
   (B) 3666
   (C) 3730
   (D) 3844
   (E) 4389

The pizzeria sells only two variations of pizza: *Margarita* and *Marinara*. The Marinara pizza comes without cheese, so only 25% of customers prefer Marinara, while 75% of customers prefer Margarita.

b. Suppose that we randomly select 16 customers. What is the probability that at least 4 out of these 16 customers prefer the Marinara pizza? Choose the alternative closest to your answer. (5p)
   (A) 37%
   (B) 40%
   (C) 60%
   (D) 63%
   (E) 66%

QUESTION 7

A pizzeria sells pizzas for €20 each. Suppose that the number of pizzas sold per week is approximately normally distributed with mean 1500 and standard deviation 180.

The pizzeria also sells cans of soda for €2 each. Suppose that the number of sodas sold per week is normally distributed with mean 1200 and standard deviation 150. Suppose that the correlation between number of pizzas sold and number of sodas sold is 0.9.

a. Find the standard deviation of the total revenue from pizza and soda (total sales in Euros). Choose the alternative closest to your answer. (5p)
   (A) 3650
   (B) 3673
   (C) 3745
   (D) 3872
   (E) 4174

The pizzeria sells only two variations of pizza: *Margarita* and *Marinara*. The Marinara pizza comes without cheese, so only 20% of customers prefer Marinara, while 80% of customers prefer Margarita.

b. Suppose that we randomly select 16 customers. What is the probability that at least 5 out of these 16 customers prefer the Marinara pizza? Choose the alternative closest to your answer. (5p)
   (A) 8%
   (B) 20%
   (C) 25%
   (D) 80%
   (E) 92%

QUESTION 8

A pizzeria sells pizzas for €15 each. Suppose that the number of pizzas sold per week is approximately normally distributed with mean 1500 and standard deviation 180.

The pizzeria also sells cans of soda for €2 each. Suppose that the number of sodas sold per week is normally distributed with mean 1200 and standard deviation 150. Suppose that the correlation between number of pizzas sold and number of sodas sold is 0.9.

a. Find the standard deviation of the total revenue from pizza and soda (total sales in Euros). Choose the alternative closest to your answer. (5p)
   (A) 2766
   (B) 2776
   (C) 2835
   (D) 2848
   (E) 2973

The pizzeria sells only two variations of pizza: *Margarita* and *Marinara*. The Marinara pizza comes without cheese, so only 20% of customers prefer Marinara, while 80% of customers prefer Margarita.

b. Suppose that we randomly select 18 customers. What is the probability that at least 3 out of these 18 customers prefer the Marinara pizza? Choose the alternative closest to your answer. (5p)
   (A) 27%
   (B) 45%
   (C) 50%
   (D) 73%
   (E) 80%

QUESTION 9

A hospital collected data on newborn babies over a two-year period. After analyzing the data, a medical doctor created a statistical model for the babies' weights. She found that the birthweights were approximately normally distributed with mean 3200 grams and standard deviation 500 grams.

a. A newborn baby is said to have *low birth weight* if their weight is less than 2500 g at birth. Use the medical doctor's model to find the probability that a randomly selected baby is born with low birth weight. (5p)
   (A) 4%
   (B) 6%
   (C) 8%
   (D) 10%
   (E) 12%

b. Use the medical doctor's model to find the probability that mean birthweight of 10 randomly selected babies is between 3100 and 3300 grams. (5p)
   (A) 27%
   (B) 32%
   (C) 37%
   (D) 42%
   (E) 47%

QUESTION 10

A hospital collected data on newborn babies over a two-year period. After analyzing the data, a medical doctor created a statistical model for the babies' weights. She found that the birthweights were approximately normally distributed with mean 3300 grams and standard deviation 600 grams.

a. A newborn baby is said to have *low birth weight* if their weight is less than 2500 g at birth. Use the medical doctor's model to find the probability that a randomly selected baby is born with low birth weight. (5p)
   (A) 5%
   (B) 7%
   (C) 9%
   (D) 11%
   (E) 13%

b. Use the medical doctor's model to find the probability that mean birthweight of 10 randomly selected babies is between 3100 and 3500 grams. (5p)
   (A) 63%
   (B) 67%
   (C) 71%
   (D) 75%
   (E) 79%

QUESTION 11

A hospital collected data on newborn babies over a two-year period. After analyzing the data, a medical doctor created a statistical model for the babies' weights. She found that the birthweights were approximately normally distributed with mean 3100 grams and standard deviation 600 grams.

a. A newborn baby is said to have *low birth weight* if their weight is less than 2500 g at birth. Use the medical doctor's model to find the probability that a randomly selected baby is born with low birth weight. (5p)
   (A) 8%
   (B) 10%
   (C) 12%
   (D) 14%
   (E) 16%

b. Use the medical doctor's model to find the probability that mean birthweight of 10 randomly selected babies is between 3000 and 3200 grams. (5p)
   (A) 40%
   (B) 42%
   (C) 44%
   (D) 46%
   (E) 48%

QUESTION 12

A hospital collected data on newborn babies over a two-year period. After analyzing the data, a medical doctor created a statistical model for the babies' weights. She found that the birthweights were approximately normally distributed with mean 3100 grams and standard deviation 400 grams.

a. A newborn baby is said to have *low birth weight* if their weight is less than 2500 g at birth. Use the medical doctor's model to find the probability that a randomly selected baby is born with low birth weight. (5p)
   (A) 5%
   (B) 7%
   (C) 9%
   (D) 11%
   (E) 13%

b. Use the medical doctor's model to find the probability that mean birthweight of 10 randomly selected babies is between 2800 and 3400 grams. (5p)
   (A) 90%
   (B) 92%
   (C) 94%
   (D) 96%
   (E) 98%

QUESTION 13

An insurance company wants to estimate average future claim in SEK per burglary in two cities, city A and city B. A claim is a sum of money requested from the insurance company by the policy holder. The insurance company used all burglaries on file from last year as sample and you can regard this as a random sample. The data is summarized below:

	City A	City B
mean	21300	25000
sample standard dev.	11000	12000
n	150	210

a. Find a 90% confidence interval for the estimated mean claim amount in City A. Choose the alternative closest to your answer. (5p)
   (A) (19974; 22625)
   (B) (19823; 22777)
   (C) (19679; 22920)
   (D) (19532; 23068)
   (E) (19385; 23214)

b. Find a 99% confidence interval for the difference in mean between the two groups, $\mu_B - \mu_A$. Choose the alternative closest to your answer. (5p)
   (A) (1497, 5903)
   (B) (1183, 6217)
   (C) (868, 6532)
   (D) (553, 6847)
   (E) (239, 7161)

QUESTION 14

An insurance company wants to estimate average future claim in SEK per burglary in two cities, city A and city B. A claim is a sum of money requested from the insurance company by the policy holder. The insurance company used all burglaries on file from last year as sample and you can regard this as a random sample. The data is summarized below:

	City A	City B
mean	24100	25000
sample standard dev.	9000	10500
n	300	250

a. Find a 99% confidence interval for the estimated mean claim amount in City A. Choose the alternative closest to your answer. (5p)
(A) (23029; 25170)
(B) (22895; 25304)
(C) (22762; 25438)
(D) (22627; 25572)
(E) (22494; 25706)

b. Find a 90% confidence interval for the difference in mean between the two groups, $\mu_B - \mu_A$. Choose the alternative closest to your answer. (5p)
(A) (-206; 2006)
(B) (-345; 2145)
(C) (-487, 2287)
(D) (-621; 2421)
(E) (-759; 2559)

QUESTION 15

An insurance company wants to estimate average future claim in SEK per burglary in two cities, city A and city B. A claim is a sum of money requested from the insurance company by the policy holder. The insurance company used all burglaries on file from last year as sample and you can regard this as a random sample. The data is summarized below:

	City A	City B
mean	19500	24000
sample standard dev.	9000	10500
n	85	100

a. Find a 90% confidence interval for the estimated mean claim amount in City A. Choose the alternative closest to your answer. (5p)
   (A) (18059; 20941)
   (B) (17894; 21106)
   (C) (17739; 21261)
   (D) (17578; 21421)
   (E) (17419; 21581)

b. Find a 99% confidence interval for the difference in mean between the two groups, $\mu_B - \mu_A$. Choose the alternative closest to your answer. (5p)
   (A) (2284; 6715)
   (B) (1915; 7085)
   (C) (1546; 7454)
   (D) (1176; 7824)
   (E) (807; 8193)

QUESTION 16

An insurance company wants to estimate average future claim in SEK per burglary in two cities, city A and city B. A claim is a sum of money requested from the insurance company by the policy holder. The insurance company used all burglaries on file from last year as sample and you can regard this as a random sample. The data is summarized below:

	City A	City B
mean	19500	21000
sample standard dev.	9000	9500
n	160	100

a. Find a 90% confidence interval for the estimated mean claim amount in City A. Choose the alternative closest to your answer. (5p)
   (A) (18450; 20550)
   (B) (18330; 20670)
   (C) (18216; 20783)
   (D) (18100; 20900)
   (E) (17983; 21017)

b. Find a 99% confidence interval for the difference in mean between the two groups, $\mu_B - \mu_A$. Choose the alternative closest to your answer. (5p)
   (A) (-640; 3640)
   (B) (-946; 3946)
   (C) (-1251; 4252)
   (D) (-1557; 4557)
   (E) (-1863; 4863)

QUESTION 17

A mobile phone manufacturer is about to launch a new mobile phone model, the M10. It will come in four colors: black, silver, grey, and gold. The table below shows the relative popularity of the same colors for their last model, the M9, as measured by early sales:

Color	Black	Silver	Grey	Gold
Choice	32%	20%	30%	18%

The manufacturer wants to investigate whether color preferences have changed, before the new model is released. As part of a survey of 200 randomly selected potential customers were asked "which of the four colors: black, silver, grey, and gold, would you prefer?

Color	Black	Silver	Grey	Gold	Total
Choice	69	39	55	37	200

Test at the 1% level if the relative popularity of colors is different from the popularity of colors for the previous model (listed in the first table).

a. What is the critical value?
   (A) 5.991
   (B) 9.210
   (C) 11.345
   (D) 13.277
   (E) 15.086

b. What is the value of the test variable? (5p)
   (A) 0.42
   (B) 0.77
   (C) 0.86
   (D) 8.11
   (E) 16.65

QUESTION 18

A mobile phone manufacturer is about to launch a new mobile phone model, the M10. It will come in four colors: black, silver, grey, and gold. The table below shows the relative popularity of the same colors for their last model, the M9, as measured by early sales:

Color	Black	Silver	Grey	Gold
Choice	25%	20%	35%	20%

The manufacturer wants to investigate whether color preferences have changed, before the new model is released. As part of a survey of 200 randomly selected potential customers were asked “which of the four colors: black, silver, grey, and gold, would you prefer?

Color	Black	Silver	Grey	Gold	Total
Choice	69	39	55	37	200

Test at the 5% level if the relative popularity of colors is different from the popularity of colors for the previous model (listed in the first table).

a. What is the critical value?
   (A) 5.991
   (B) 7.815
   (C) 9.488
   (D) 11.070
   (E) 12.592

b. What is the value of the test variable? (5p)
   (A) 0.97
   (B) 1.12
   (C) 10.68
   (D) 14.01
   (E) 15.34

QUESTION 19

A mobile phone manufacturer is about to launch a new mobile phone model, the M10. It will come in four colors: black, silver, grey, and gold. The table below shows the relative popularity of the same colors for their last model, the M9, as measured by early sales:

Color	Black	Silver	Grey	Gold
Choice	25%	20%	35%	20%

The manufacturer wants to investigate whether color preferences have changed, before the new model is released. As part of a survey of 200 randomly selected potential customers were asked "which of the four colors: black, silver, grey, and gold, would you prefer?

Color	Black	Silver	Grey	Gold	Total
Choice	60	35	55	50	200

Test at the 5% level if the relative popularity of colors is different from the popularity of colors for the previous model (listed in the first table).

a. What is the critical value?
   (A) 5.991
   (B) 7.815
   (C) 9.488
   (D) 11.070
   (E) 12.592

b. What is the value of the test variable? (5p)
   (A) 0.63
   (B) 2.27
   (C) 7.11
   (D) 8.34
   (E) 10.97

QUESTION 20

A mobile phone manufacturer is about to launch a new mobile phone model, the M10. It will come in four colors: black, silver, grey, and gold. The table below shows the relative popularity of the same colors for their last model, the M9, as measured by early sales:

Color	Black	Silver	Grey	Gold
Choice	35%	20%	25%	20%

The manufacturer wants to investigate whether color preferences have changed, before the new model is released. As part of a survey of 200 randomly selected potential customers were asked "which of the four colors: black, silver, grey, and gold, would you prefer?

Color	Black	Silver	Grey	Gold	Total
Choice	71	39	72	18	200

Test at the 1% level if the relative popularity of colors is different from the popularity of colors for the previous model (listed in the first table).

a. What is the critical value?
   (A) 5.991
   (B) 9.210
   (C) 11.345
   (D) 13.277
   (E) 15.086

b. What is the value of the test variable? (5p)
   (A) 1.42
   (B) 3.97
   (C) 8.11
   (D) 11.32
   (E) 21.82

QUESTION 21

A factory produces cans of crushed tomato. If the machine is working correctly, the cans weigh 415 grams each, on average, and assume that the weights are normally distributed. A manager collects a random sample of 10 cans; the weights are listed in the table below:

weights	390	405	402	393	417	403	420	431	409	410

Test at the 5% level whether the population mean weight of the cans is <u>different</u> from 415 grams.

a. State the necessary assumptions and your hypotheses. (5p)
b. State the test variable, the decision rule, and the critical value. (5p)
c. Calculate the test statistic, draw conclusions and interpret the test. (5p)

The factory has another, similar machine and the manager decides to collect a random sample of 10 cans from this machine as well. Again, she wants to test whether the population mean weight of the cans produced is <u>different from 415 grams</u>. **This time, assume that the true population mean is 405 grams per can and that the standard deviation is 12 grams (these figures are unknown to the manager). Assume that the sample standard deviation will be 12 grams as well.**

d. Find the probability that the manager correctly rejects the null hypothesis. (5p)
e. What is the probability in (d) called? Briefly explain what the manager could have done to increase this probability. (5p)

QUESTION 22

A factory produces cans of crushed tomato. If the machine is working correctly, the cans weigh 415 grams each, on average, and assume that the weights are normally distributed. A manager collects a random sample of 10 cans; the weights are listed in the table below:

weights	389	404	401	392	417	402	421	432	409	411

Test at the 5% level whether the population mean weight of the cans is <u>different</u> from 425 grams.

a. State the necessary assumptions and your hypotheses. (5p)
b. State the test variable, the decision rule, and the critical value. (5p)
c. Calculate the test statistic, draw conclusions and interpret the test. (5p)

The factory has another, similar machine and the manager decides to collect a random sample of 10 cans from this machine as well. Again, she wants to test whether the population mean weight of the cans produced is <u>different from 425 grams</u>. **This time, assume that the true population mean is 405 grams per can and that the standard deviation is 16 grams (these figures are unknown to the manager). Assume that the sample standard deviation will be 16 grams as well.**

d. Find the probability that the manager correctly rejects the null hypothesis. (5p)
e. What is the probability in (d) called? Briefly explain what the manager could have done to increase this probability. (5p)

QUESTION 23

A factory produces cans of crushed tomato. If the machine is working correctly, the cans weigh 415 grams each, on average, and assume that the weights are normally distributed. A manager collects a random sample of 10 cans; the weights are listed in the table below:

weights	410	419	418	412	428	430	437	423	424	418

Test at the 5% level whether the population mean weight of the cans is <u>different</u> from 415 grams.

a. State the necessary assumptions and your hypotheses. (5p)
b. State the test variable, the decision rule, and the critical value. (5p)
c. Calculate the test statistic, draw conclusions and interpret the test. (5p)

The factory has another, similar machine and the manager decides to collect a random sample of 10 cans from this machine as well. Again, she wants to test whether the population mean weight of the cans produced is <u>different from 415 grams</u>. **This time, assume that the true population mean is 420 grams per can and that the standard deviation is 10 grams (these figures are unknown to the manager). Assume that the sample standard deviation will be 10 grams as well.**

d. Find the probability that the manager correctly rejects the null hypothesis. (5p)
e. What is the probability in (d) called? Briefly explain what the manager could have done to increase this probability. (5p)

QUESTION 24

A factory produces cans of crushed tomato. If the machine is working correctly, the cans weigh 415 grams each, on average, and assume that the weights are normally distributed. A manager collects a random sample of 10 cans; the weights are listed in the table below:

weights	388	399	397	390	409	398	412	421	403	404

Test at the 5% level whether the population mean weight of the cans is <u>different</u> from 415 grams.

a. State the necessary assumptions and your hypotheses. (5p)
b. State the test variable, the decision rule, and the critical value. (5p)
c. Calculate the test statistic, draw conclusions and interpret the test. (5p)

The factory has another, similar machine and the manager decides to collect a random sample of 10 cans from this machine as well. Again, she wants to test whether the population mean weight of the cans produced is <u>different from 415 grams</u>. **This time, assume that the true population mean is 400 grams per can and that the standard deviation is 12 grams (these figures are unknown to the manager). Assume that the sample standard deviation will be 12 grams as well.**

d. Find the probability that the manager correctly rejects the null hypothesis. (5p)
e. What is the probability in (d) called? Briefly explain what the manager could have done to increase this probability. (5p)

QUESTION 25

A finance student has to write a paper for a class. He decides to estimate the *Beta coefficient* of gold. To this end, he uses 11 months of data. He calculates the return in percent of gold and the return in percent of the S&P 500-index. You can find the data in the table below:

month	Gold return (y)	sp500 return (x)	Residuals
Feb	1	-3	0,31
Mar	-1	0	-1,00
Apr	-1	2	-0,54
May	-4	0	-4,00
Jun	-2	4	-1,07
Jul	-2	3	-1,31
Aug	-1	0	-1,00
Sep	2	-7	0,38
Oct	1	2	1,46
Nov	5	-10	2,68
Dec	2	9	4,08

He decides to calculate the beta by estimating the following (simplified) regression model:

$$Y = \alpha + \beta X + \varepsilon$$

Where *Y* is the percentage return of the Gold and *X* is the percentage return of the S&P500. The Beta coefficient is the estimated slope coefficient from this regression model.

a. Estimate the variance of *X* (S&P 500 return) and estimate the covariance between *X* and *Y*. (5p)
b. Use your answers in part a to estimate the model parameters and interpret the numerical value of the slope coefficient. Clearly state the estimated model. (5p)
c. Calculate the residual variance and the coefficient of determination. To make your calculations easier, you can use the rounded residuals from the "Residuals" column in the table. (5p)

For c and d you are asked to test if the slope coefficient is significantly <u>smaller than zero</u> at 5% level of significance.

d. State the hypotheses, test statistic, critical value and decision rule. *Tip: you can answer this one even if you did not solve part a-c.* (5p)
e. Calculate the test variable, and state the conclusion of the test. (5p)

QUESTION 26

A finance student has to write a paper for a class. He decides to estimate the *Beta coefficient* of The Russell 2000 index (RUT). To this end, he uses 11 months of data. He calculates the return in percent of RUT and the return in percent of the S&P 500-index. You can find the data in the table below:

month	RUT return (y)	sp500 return (x)	Residuals
Feb	1	-3	2,72
Mar	1	0	1,00
Apr	6	2	4,85
May	1	0	1,00
Jun	2	4	-0,30
Jul	4	3	2,28
Aug	-3	0	-3,00
Sep	-10	-6	-6,55
Oct	1	2	-0,15
Nov	-4	-10	1,74
Dec	1	8	-3,60

He decides to calculate the beta by estimating the following (simplified) regression model:

$$Y = \alpha + \beta X + \varepsilon$$

Where Y is the percentage return of RUT and X is the percentage return of the S&P500. The Beta coefficient is the estimated slope coefficient from this regression model.

a. Estimate the variance of X (S&P 500 return) and estimate the covariance between X and Y. (5p)
b. Use your answers in part a to estimate the model parameters and interpret the numerical value of the slope coefficient. Clearly state the estimated model. (5p)
c. Calculate the residual variance and the coefficient of determination. To make your calculations easier, you can use the rounded residuals from the "Residuals" column in the table. (5p)

For c and d you are asked to test if the slope coefficient is significantly different from <u>one</u> at 5% level of significance.

d. State the hypotheses, test statistic, critical value and decision rule. *Tip: you can answer this one even if you did not solve part a-c.* (5p)
e. Calculate the test variable, and state the conclusion of the test. (5p)

QUESTION 27

A finance student has to write a paper for a class. He decides to estimate the *Beta coefficient* of the stock MercadoLibre, Inc. (MELI). To this end, he uses 11 months of data. He calculates the return in percent of MELI and the return in percent of the S&P 500-index. You can find the data in the table below:

month	MELI return (y)	sp500 return (x)	Residuals
Feb	13	4	-0,81
Mar	0	0	-0,44
Apr	8	1	4,22
May	18	1	14,22
Jun	-9	0	-9,44
Jul	14	2	6,87
Aug	-11	0	-11,44
Sep	0	2	-7,13
Oct	-7	2	-14,13
Nov	14	0	13,56
Dec	15	3	4,53

He decides to calculate the beta by estimating the following (simplified) regression model:

$$Y = \alpha + \beta X + \varepsilon$$

Where *Y* is the percentage return of MELI and *X* is the percentage return of the S&P500. The Beta coefficient is the estimated slope coefficient from this regression model.

a. Estimate the variance of *X* (S&P 500 return) and estimate the covariance between *X* and *Y*. (5p)
b. Use your answers in part a to estimate the model parameters and interpret the numerical value of the slope coefficient. Clearly state the estimated model. (5p)
c. Calculate the residual variance and the coefficient of determination. To make your calculations easier, you can use the rounded residuals from the "Residuals" column in the table. (5p)

For c and d you are asked to test if the slope coefficient is significantly different from <u>one</u> at 5% level of significance.

d. State the hypotheses, test statistic, critical value and decision rule. *Tip: you can answer this one even if you did not solve part a-c.* (5p)
e. Calculate the test variable, and state the conclusion of the test. (5p)

QUESTION 28

A finance student has to write a paper for a class. He wants to examine how well the Stockholm exchange follows the S&P 500. He decides to estimate something similar to the *Beta coefficient* of OMX Stockholm 30 Index (OMX). To this end, he uses 11 months of data. He calculates the return in percent of OMX and the return in percent of the S&P 500-index. You can find the data in the table below:

month	OMX return (y)	sp500 return (x)	Residuals
Feb	-3	-3	-0,06
Mar	2	0	2,36
Apr	-1	2	-2,36
May	1	0	1,36
Jun	4	4	0,92
Jul	3	3	0,78
Aug	0	0	0,36
Sep	-8	-7	-1,62
Oct	-2	2	-3,36
Nov	-7	-9	1,11
Dec	7	8	0,48

He decides to calculate the beta by estimating the following (simplified) regression model:

$$Y = \alpha + \beta X + \varepsilon$$

Where *Y* is the percentage return of OMX and *X* is the percentage return of the S&P500. The Beta coefficient is the estimated slope coefficient from this regression model.

a. Estimate the variance of *X* (S&P 500 return) and estimate the covariance between *X* and *Y*. (5p)
b. Use your answers in part a to estimate the model parameters and interpret the numerical value of the slope coefficient. Clearly state the estimated model. (5p)
c. Calculate the residual variance and the coefficient of determination. To make your calculations easier, you can use the rounded residuals from the "Residuals" column in the table. (5p)

For c and d you are asked to test if the slope coefficient is significantly different from <u>one</u> at 5% level of significance.

d. State the hypotheses, test statistic, critical value and decision rule. *Tip: you can answer this one even if you did not solve part a-c.* (5p)
e. Calculate the test variable, and state the conclusion of the test. (5p)

SOLUTIONS

This is solutions to question 1, 5, 9, 13, 17, 21 and 25. The other problems from 1-20 are almost identical to these, except for the numbers. I think you can use these seven problems to figure out how to solve the rest. If you have any questions, please email me. Please contact me if you find any typos or other problems. Answers to problems 22-24 and 26-28 can be found at the end of this document. Follow the methods shown in 21 and 25.

/Ulf

1a.

The percentage of eligible voters who did vote is

$$\frac{number\ of\ voters}{number\ of\ eligible\ voters}$$

Men 18-39 is made up of four age categories in the table. First find the number of voters in each of the four categories and add them up:

$$\frac{79{,}3}{100}\cdot 420+\frac{78{,}9}{100}\cdot 290+\frac{82{,}5}{100}\cdot 269+\frac{85{,}3}{100}\cdot 276=1019{,}223$$

Then find the number of eligible voters in the same categories:

$$420+290+269+276=1255$$

Divide to find the answer:

$$\frac{1019{,}223}{1255}=0{,}81213\ \approx 81{,}2\%$$

- -

1b.

Follow the tip:

age	voted	eligible	freq	cum freq	cum rel freq
18-24	83,3	393	327,369	327,369	0,102021615
25-29	84	286	240,24	567,609	0,176890259
30-34	85,4	265	226,31	793,919	0,247417742
35-39	86,8	267	231,756	1025,675	0,319642423
40-44	89	302	268,78	1294,455	0,403405302
45-49	88,2	320	282,24	…	…
50-54	88,8	292	259,296		
55-59	91	279	253,89		
60-64	90,8	277	251,516		
65-69	92,2	281	259,082		
70-74	90,5	259	234,395		
75-79	86,7	178	154,326		
80+	69,5	316	219,62		
		sum:	3208,82		

The goal is to find the age such that 25% of female voters were younger than that age.

First find the frequency (number) of voters in each age category. Add all the frequencies up to get the total number of voters: 3208,82 thousand female voters.

Then find the cumulative frequency of the first few categories. This will tell you how many female votes that were in that category or younger. For example, we can se that 327 thousand female voters were 18-24 and that 568 thousand voters were 29 or younger.

Then divide the cumulative frequencies by the total frequency to get the cumulative relative frequencies. We see that:

10,2% were 18-24

17,7% were 29 or younger

24,7% were 34 or younger

31,9% were 39 or younger

This means that the 25th percentile must be in the age category 35-39.

5a.

If the number of pizzas sold is *X* and the number of sodas sold is *Y*, then the total revenue from pizza and soda is

$$10\,X + 2\,Y$$

The variance is given by the formula on page three:

$$\begin{aligned} &Var(aX + bY + c) \\ &\quad = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X,Y) \\ &\qquad = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY} \end{aligned}$$

First, we need to convert correlation to covariance. This formula is also on page 3:

Correlation between X and Y:

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

We have:

$$\begin{aligned} &Corr(X,Y) = 0{,}9 \\ &Var(X) = 180^2 \\ &Var(Y) = 150^2 \\ &a = 10 \\ &b = 2 \\ &c = 0 \end{aligned}$$

Find the covariance:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sqrt{180^2 150^2}} = 0{,}9$$

…solve for covariance:

$$Cov(X,Y) = 180 \cdot 150 \cdot 0{,}9$$

Then we plug everything into the formula for variance:

$$Var(10X + 2Y) = 10^2 \cdot 180^2 + 2^2 \cdot 150^2 + 2 \cdot 10 \cdot 2 \cdot 180 \cdot 150 \cdot 0{,}9 = 4\,392\,000$$

Then take the square root to get the standard deviation:

$$\sqrt{4\,392\,000} \approx 2096.$$

5b.

The question reads*: Suppose that we randomly select 20 customers. What is the probability that <u>more than 5</u> out of these 20 customers prefer the Marinara pizza?*

The number of customers out of the 20 who prefer Marinara pizza is binomially distributed with

$$n = 20$$
$$P = 0{,}25$$

If X is the number of successes, we seek:

$$\mathrm{P}(X > 5) = 1 - \mathrm{P}(X \leq 5)$$

Illustration:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<u>not</u> more than 5						**more than 5**														

The probability $\mathrm{P}(X \leq 5)$ can be found in the table:

TABLE 7 Continued Binomial Distribution

n	*x*	*p* = 0.05	0.1	0.15	0.2	0.25
19	9	1.00000	1.00000	0.99986	0.99842	0.99110
	10			0.99998	0.99969	0.99771
	11			1.00000	0.99995	0.99952
	12				0.99999	0.99992
	13				1.00000	0.99999
	14					1.00000
	15					
	16					
	17					
	18					
20	0	0.35849	0.12158	0.03876	0.01153	0.00317
	1	0.73584	0.39175	0.17556	0.06918	0.02431
	2	0.92452	0.67693	0.40490	0.20608	0.09126
	3	0.98410	0.86705	0.64773	0.41145	0.22516
	4	0.99743	0.95683	0.82985	0.62965	0.41484
	5	0.99967	0.98875	0.93269	0.80421	0.61717

So $\mathrm{P}(X > 5) = 1 - 0{,}61717 \approx 0{,}38 =$ 38%

9a.

If X is the birth weight of a randomly chosen baby, then:

$$X \sim N(3200; 500^2)$$

We seek

$$\mathrm{P}(X < 2500) = \mathrm{P}\left(\frac{X-3200}{500} < \frac{2500-3200}{500}\right) = \mathrm{P}\left(Z < \frac{-700}{500}\right)$$
$$= \mathrm{P}(Z < -1{,}4) = 1 - \mathrm{P}(Z < 1{,}4)$$

We find $\mathrm{P}(Z < 1{,}4)$ in table 1:

$$1 - \mathrm{P}(Z < 1{,}4) = 1 - 0{,}91924 \approx 8\%$$

9b.

First, we find the distribution of the sample mean. From the formula sheet:

SAMPLING THEORY

Assume the observations $X_1, X_2, \dots, X_n$ in a sample are independent and have the same (finite) mean μ_X, and the same (finite) variance σ_X^2. Define $\bar{X} = \sum_i X_i/n$

Mean and variance:	$E(\bar{X}) = \mu_{\bar{X}} = \mu_X$	$Var(\bar{X}) = \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$	$SE(\bar{X}) = \frac{\sigma_X}{\sqrt{n}}$

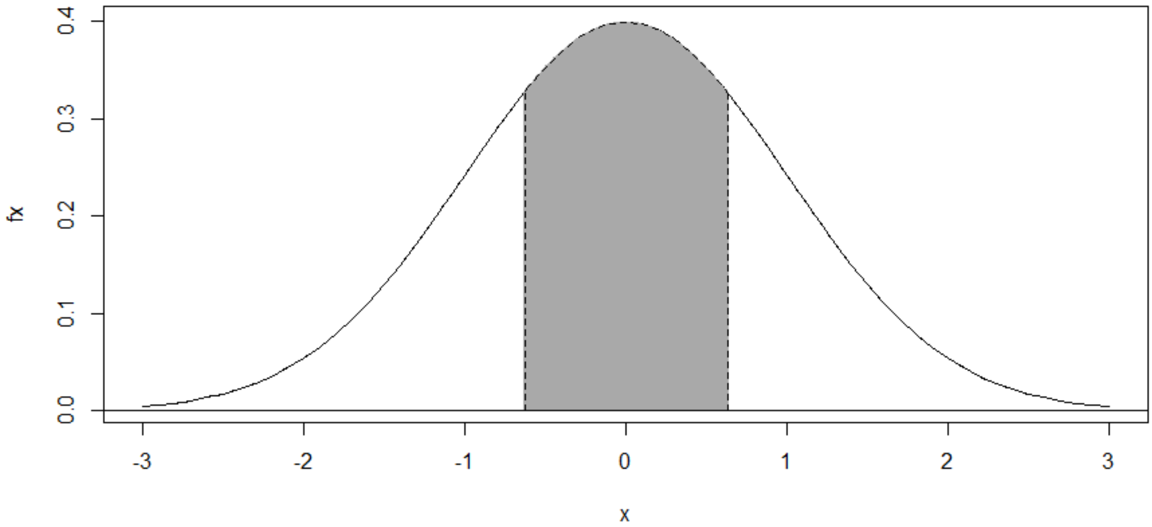
Since the weighs are normally distributed, the sample mean will also be normally distributed:

$\bar{X} \sim N\left(3200; \frac{500^2}{10}\right)$

We seek

$$\mathrm{P}(3100 < \bar{X} < 3000) = \mathrm{P}\left(\frac{3100-3200}{\frac{500}{\sqrt{10}}} < Z < \frac{3300-3200}{\frac{500}{\sqrt{10}}}\right) \approx \mathrm{P}(-0{,}63 < \mathrm{Z} < 0{,}63)$$

It is a good idea to draw a picture:



We seek the probability of the grey area, so we need to subtract the area of the tails (they have the same area).

The area of each tail is:

$$1 - \mathrm{F}(0{,}63)$$

So to get our probability, we subtract each tail from 1:

$$1 - \big(1 - \mathrm{F}(0{,}63)\big) - \big(1 - \mathrm{F}(0{,}63)\big) = 1 - 1 + \mathrm{F}(0{,}63) - 1 + \mathrm{F}(0{,}63)$$
$$= 2 \cdot \mathrm{F}(0{,}63) - 1.$$

We get $\mathrm{F}(0{,}63)$ from table 1:

$$\mathrm{F}(0{,}63) = 0{,}73565$$

And finally:

$$2 \cdot \mathrm{F}(0{,}63) - 1 = 0{,}4713 \ \approx 47\%$$

13a.

- for the mean μ_X

$n \geq 30$, regardless of the distribution of X_i:	σ_X^2 known:	$\bar{x} \pm z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$
	σ_X^2 unknown:	$\bar{x} \pm z_{\alpha/2} \frac{s_X}{\sqrt{n}}$

We have:

$$\bar{x} = 21300$$
$$\alpha = 1 - 0{,}9 = 0{,}1$$
$$\frac{\alpha}{2} = 0{,}05$$
$$z_{0{,}05} = 1{,}6499$$
$$s = 11000$$
$$n = 150$$

Plug it into the formula

$$21300 - 1{,}6449 \frac{11000}{\sqrt{150}} = 19823$$
$$21300 + 1{,}6449 \frac{11000}{\sqrt{150}} = 22777$$

13b.

– for the difference $\mu_X - \mu_Y$ (two independent samples)

$n_X, n_Y \geq 30$ regardless of distribution:	σ_X^2, σ_Y^2 known:	$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_x} + \frac{\sigma_Y^2}{n_y}}$
	σ_X^2, σ_Y^2 unknown:	$\bar{x} - \bar{y} \pm z_{\alpha/2}\sqrt{\frac{s_X^2}{n_x} + \frac{s_Y^2}{n_y}}$

We have:

$$\bar{y} = 25000$$
$$\bar{x} = 21300$$
$$\alpha = 1 - 0{,}99 = 0{,}01$$
$$\frac{\alpha}{2} = 0{,}005$$
$$z_{0{,}05} = 2{,}5758$$
$$s_y = 12000$$
$$s_x = 11000$$
$$n_y = 210$$
$$n_x = 150$$

Plug it into the formula

$$(25000 - 21300) - 2{,}5758\sqrt{\frac{11000^2}{150} + \frac{12000^2}{210}} = 553$$
$$(25000 - 21300) + 2{,}5758\sqrt{\frac{11000^2}{150} + \frac{12000^2}{210}} = 6847$$

17a.

This is a goodness-of-fit test, so one of the Chi-square tests covered in the course. The degrees of freedom is number of categories minus one, so 4-1 = 3. We find the value in table 4:

TABLE 4. χ^2-distribution

$Q \in \chi^2(v)$ where v = degrees of freedom.

The value of q_α if $P(Q > q_\alpha) = \alpha$ where α is a given probability.

v	α = 0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01
1	0.000	0.000	0.000	0.001	0.004	3.841	5.024	6.635
2	0.002	0.010	0.020	0.051	0.103	5.991	7.378	9.210
3	0.024	0.072	0.115	0.216	0.352	7.815	9.348	11.345
4	0.091	0.207	0.297	0.484	0.711	9.488	11.143	13.277
5	0.210	0.412	0.554	0.831	1.145	11.070	12.833	15.086

17b.

The test variable is best calculated in table form:

	black	silver	grey	gold	sum
O	69	39	55	37	200
P	0,32	0,2	0,3	0,18	1
E	64	40	60	36	200
O-E	5	-1	-5	1	0
$(O-E)^2$	25	1	25	1	52
$(O-E)^2/E$	0,390625	0,025	0,416667	0,027778	0,860069

QUESTION 21:

Test at the 5% level whether the population mean weight of the cans is <u>different</u> from 415 grams.

a. *State the necessary assumptions and your hypotheses. (5p)*

Assumptions: that the weights are normally distributed is included in the problem, so this does not need to be mentioned again. We need the random sample to be **Independent and Identically distributed**.

Hypotheses:

$$H_0 : \mu_X = 415$$
$$H_1 : \mu_X \neq 415$$

b. *State the test variable, the decision rule, and the critical value. (5p)*

<u>Test variable:</u>

σ_X^2 unknown: $$t_{n-1} = \frac{\bar{X}-\mu_0}{s_X/\sqrt{n}}$$

<u>Critical value:</u>

The test variable follows a *t*-distribution and this is a two-sided test:

$$\alpha = 0{,}05$$
$$\frac{\alpha}{2} = 0{,}025$$
$$n = 10$$
$$n - 1 = 9$$

Now we can get the critical value from table 3:

$$t_{0{,}05;\,9} = 2{,}262$$

<u>Decision rule:</u>

Reject the null hypothesis if $|t_{obs}| > 2{,}262$.

c. *Calculate the test statistic, draw conclusions and interpret the test. (5p)*

We do not know the population variance, so we will need to calculate s_x, the sample standard deviation, from the sample. We sum the weights and find the sample mean

$$\bar{x} = \frac{4080}{10} = 408$$

Then we make a table to help us find the sample variance:

weights (x)	390	405	402	393	417	403	420	431	409	410	**4080**
$x - \bar{x}$	-18	-3	-6	-15	9	-5	12	23	1	2	**0**
$(x - \bar{x})^2$	324	9	36	225	81	25	144	529	1	4	**1378**

Sample variance:

$$s_x^2 = \frac{1378}{9}$$

Sample standard deviation:

$$s_x = \sqrt{\frac{1378}{9}}$$

We also note that the population mean in the null hypothesis is $\mu_0 = 415$

<u>The observed value of the test statistic:</u>

$$t_{obs} = \frac{408 - 415}{\sqrt{\frac{1378}{9}} \Big/ \sqrt{10}} \approx -1{,}789$$

<u>Conclusion:</u>

Since $|t_{obs}| = 1{,}789 \ngtr 2{,}262$ (not greater than), we fail to reject the null.

<u>Interpretation:</u>

We have not found statistically significant evidence that the mean weight of the cans is different from 415 g, at the 5% level of significance.

The factory has another, similar machine and the manager decides to collect a random sample of 10 cans from this machine as well. Again, she wants to test whether the population mean weight of the cans produced is <u>different from 415 grams</u>. ***This time, assume that the true population mean is 405 grams per can and that the standard deviation is 12 grams (these figures are unknown to the manager). Assume that the sample standard deviation will be 12 grams as well.***

d. Find the probability that the manager correctly rejects the null hypothesis. (5p)

Note: This problem was indented to be one of the harder problems of the exam. If you have shown that you get the general idea, you will get partial credit.

We know the true distribution of the weights:

$$X \sim N(405; 12^2)$$

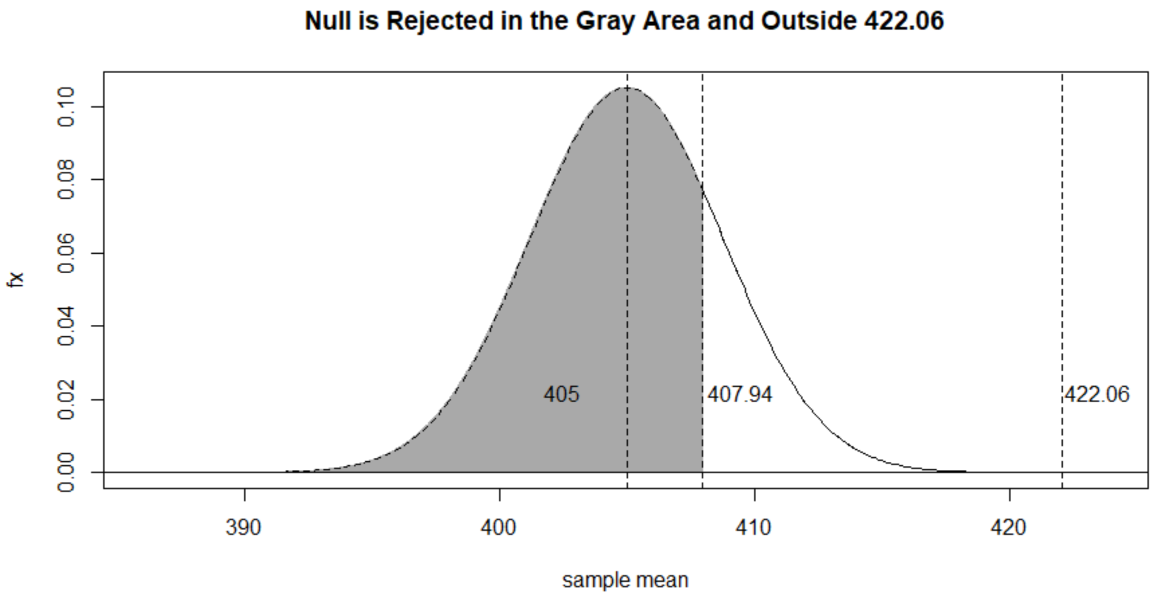
…but the manager does not. First, we need to find the manager's decision rule in grams. At the 5% level of significance, the manager will reject the null if the test variable is 1,96 standard deviations or more from $\mu_0 = 415$, in either direction. The sample standard deviation is 12 and $n = 10$, so she will reject if

$$\bar{X} < 415 - 1{,}96 \cdot \frac{12}{\sqrt{10}} = 407{,}94$$

…or if

$$\bar{X} > 415 + 1{,}96 \cdot \frac{12}{\sqrt{10}} = 422{,}06$$

Draw a picture:



So, the probability that she rejects the null is

$$\mathrm{P}(\bar{X} < 407{,}94) + \mathrm{P}(\bar{X} > 422{,}06)$$

We calculate each part separately:

$$\mathrm{P}(\bar{X} < 391{,}48) = \mathrm{P}\left(Z < \frac{407{,}94 - 405}{12/\sqrt{10}}\right) \approx \mathrm{P}(Z < 0{,}77)$$

and

$$\mathrm{P}(\bar{X} > 422{,}06) = \mathrm{P}\left(Z < \frac{422{,}06 - 405}{12/\sqrt{10}}\right) \approx \mathrm{P}(Z > 4{,}50)$$

We find the values in table 1:

$$\mathrm{P}(Z < 0{,}77) = \mathrm{F}(0{,}77) = 0{,}77935$$

and

$$\mathrm{P}(Z > 4{,}50) = 1 - \mathrm{F}(4{,}50)$$

The table one only goes to 4.0; we will round our value for $\mathrm{F}(4{,}50)$ to one since $\mathrm{F}(4{,}00) = 0.99998 \approx 1$ and $\mathrm{F}(4{,}50)$ is even closer to one.

$$\mathrm{P}(Z > 4{,}50) = 1 - \mathrm{F}(4{,}50) \approx 1 - 1 = 0$$

Answer:

$$0{,}77935 + 0 = 0{,}77935$$

The probability to correctly reject the false null hypothesis μ = 415 is 78%.

e. *What is the probability in (d) called? Briefly explain what the manager could have done to increase this probability. (5p)*

The probability of correctly rejecting a false null hypothesis is called **power**. She could increase the power by increasing the sample size. She could also change the level of significance (to 10%, for example).

QUESTION 25:

a.
Estimate the variance of X (S&P 500 return) and estimate the covariance between X and Y. (5p)

month	y	x	Residuals	$(x-\overline{x})^2$	$(x-\overline{x})(y-\overline{y})$
Feb	1	-3	0,31	9	-3
Mar	-1	0	-1	0	0
Apr	-1	2	-0,54	4	-2
May	-4	0	-4	0	0
Jun	-2	4	-1,07	16	-8
Jul	-2	3	-1,31	9	-6
Aug	-1	0	-1	0	0
Sep	2	-7	0,38	49	-14
Oct	1	2	1,46	4	2
Nov	5	-10	2,68	100	-50
Dec	2	9	4,08	81	18
sum	**0**	**0**	**-0,01**	**272**	**-63**

We start by finding $\overline{x}$ and $\overline{y}$ – and we find that they are both zero. We then create the table above.

Now we are ready to estimate the variance of *X* and the covariance between *X* and *Y*:

$$s_x^2 = \frac{272}{10} = 27{,}2$$
$$s_{xy} = \frac{-63}{10} = -6{,}3$$

b.
Use your answers in part a to estimate the model parameters and interpret the numerical value of the slope coefficient. Clearly state the estimated model. (5p)

We will use the formula from page 7:

$$b_1 = \frac{Cov(x,y)}{s_x^2} = r_{xy}\frac{s_y}{s_x} \qquad b_0 = \bar{y} - b_1\bar{x}$$

We did all the hard work in (a), so

$$b_1 = \frac{27{,}2}{-6{,}3} \approx -0{,}2316$$

and

$$b_0 = 0 - (-0{,}2316)\cdot 0 = 0$$

The estimated model is (we can leave out the intercept since it is zero):

$$\hat{y} = -0{,}2316\,x$$

c.
Calculate the residual variance and the coefficient of determination. To make your calculations easier, you can use the rounded residuals from the "Residuals" column in the table. (5p)

We find the relevant formulas from the formula sheet:

Residual variance: $$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-K-1}$$

and

Sum-of-squares: $$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)s_y^2$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

Coefficient of determination: $$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

We need to make two more columns:

month	y	x	Residuals	e^2	$(y-\bar{y})^2$
Feb	1	-3	0,31	0,0961	1
Mar	-1	0	-1	1	1
Apr	-1	2	-0,54	0,2916	1
May	-4	0	-4	16	16
Jun	-2	4	-1,07	1,1449	4
Jul	-2	3	-1,31	1,7161	4
Aug	-1	0	-1	1	1
Sep	2	-7	0,38	0,1444	4
Oct	1	2	1,46	2,1316	1
Nov	5	-10	2,68	7,1824	25
Dec	2	9	4,08	16,6464	4
sum	**0**	**0**	**-0,01**	**47,3535**	**62**

We take the sum of the squared residuals (*SSE*) from the table and divide by $11 - 1 - 1 = 9$

$$s_e^2 = \frac{47{,}3535}{9} = 5{,}2615$$

To find Coefficient of determination, we take *SSE* and *SST* straight from the table

$$R^2 = 1 - \frac{47{,}3535}{62} \approx 0{,}236 = 23{,}6\%$$

For d and e you are asked to test if the slope coefficient is significantly <u>smaller than zero</u> at 5% level of significance.

d.

State the hypotheses, test statistic, critical value and decision rule. Tip: you can answer this one even if you did not solve part a-c. (5p)

<u>Hypotheses:</u>

$$H_0 : \beta = 0$$
$$H_1 : \beta < 0$$

<u>Test statistic:</u>

$$t_{n-K-1} = \frac{b_j - \beta_j^*}{s_{b_j}}$$

<u>Critical value:</u>

This is a one-sided test at $\alpha = 0{,}05$ and the degrees of freedom are $11 - 1 - 1 = 9$. Since we will reject is the test variable is smaller than zero, we need the critical value to the left:

$$t_{0{,}05;9} = -1{,}833$$

<u>Decision Rule:</u>

We reject the null if the observed test variable is smaller than $-1{,}833$.

e.

Calculate the test variable, and state the conclusion of the test. (5p)

Good news: we calculated b_1 in (b) and $\beta_1^* = 0$. We still need s_{b_1}:

$$s_{b_1}^2 = \frac{s_e^2}{(n-1)s_x^2} = \frac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

More good news: we calculated the numerator of this in (c) and we can get the denominator from the table in (a):

$$s_{b_1}^2 = \frac{5,2615}{272}$$

So,

$$t_{obs} = \frac{-0,2316 - 0}{\sqrt{\frac{5,2615}{272}}} \approx -1,67$$

Conclusion:

Since $-1,67 \nless -1,833$ (not less than), we fail to reject the null at 5% level of significance. We cannot conclude that β is less than zero.

Answers to Questions 22, 23, 24, 26, 27, and 28

22a, 23a, 24a. Compare to 21a.

22b, 23b ,24b. Compare to 21b.

22c. test statistic: $t_{obs} = -4,14$

23c. test statistic: $t_{obs} = 2,64$

24c. test statistic: $t_{obs} = -4,07$

22d. The probability is 98%

23d. The probability is 35%

24d. The probability is 98%

22e, 23e, 24e. Compare to 21e.

26a.

$$s_x^2 = 24,2$$
$$s_{xy} = 13,9$$

27a.

$$s_x^2 = 1,85$$
$$s_{xy} = 6,2$$

28a.

$$s_x^2 = 23,6$$
$$s_{xy} = 20,3$$

26b. Estimated model:

$$\hat{y} = 0 + 0{,}574\,x$$

27b. Estimated model:

$$\hat{y} = 0{,}441 + 3{,}343\,x$$

28b. Estimated model:

$$\hat{y} = -0{,}364 + 0{,}860\,x$$

26c.

$$s_e^2 = 11{,}79$$
$$R^2 = 42{,}9\%$$

27c.

$$s_e^2 = 104{,}8$$
$$R^2 = 18{,}0\%$$

28c.

$$s_e^2 = 3{,}328$$
$$R^2 = 85{,}4\%$$

26d.

Hypotheses:

$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

Critical value:

$$t_{0,025;9} = 2.262$$

Decision Rule:

We reject the null if $|t_{obs}| = 2.262$

27d.

Hypotheses:

$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

Critical value:

$$t_{0,025;9} = 2.262$$

Decision Rule:

We reject the null if $|t_{obs}| = 2.262$

28d.

Hypotheses:

$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

Critical value:

$$t_{0,025;9} = 2.262$$

Decision Rule:

We reject the null if $|t_{obs}| = 2.262$

26e.

$$s_{b_1}^2 = \frac{11,79}{242}$$

$$t_{obs} = \frac{0,574 - 1}{\sqrt{\frac{11,79}{242}}} \approx -1,93$$

27e.

$$s_{b_1}^2 = \frac{104,7}{18,55}$$

$$t_{obs} = \frac{3,34 - 1}{\sqrt{\frac{104,7}{18,55}}} \approx 0,98$$

28e.

$$s_{b_1}^2 = \frac{3,328}{236}$$

$$t_{obs} = \frac{0,86 - 1}{\sqrt{\frac{3,328}{236}}} \approx -1,18$$