Stockholm University

STOCKHOLM UNIVERSITY
Department of Statistics
Fall 2021

Ulf Högnäs (examiner)

EXAM – BASIC STATISTICS FOR ECONOMISTS

2021-01-14

Time:	10.00-16.00 (Start submitting early; no late submissions will be accepted.)
Approved aid:	Any books, notes, or digital resources. You are not allowed to communicate with anyone during the exam. This includes chats, messages, and internet forums.

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 or on one page. 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).

  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.

  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.

0001-LBO	1	5	9	13	17	21	25
0002-AFB	2	6	10	14	18	22	26
0003-LJK	3	7	11	15	19	23	27
0004-HLP	4	8	12	16	20	24	28
0005-FOT	1	5	9	13	17	21	25
0006-NLC	2	6	10	14	18	22	26
0007-TMK	3	7	11	15	19	23	27
0008-TOW	4	8	12	16	20	24	28
0009-TAP	1	5	9	13	17	21	25
0010-ESJ	2	6	10	14	18	22	26
0011-HMN	3	7	11	15	19	23	27
0012-HXO	4	8	12	16	20	24	28
0013-CWR	1	5	9	13	17	21	25
0014-GMN	2	6	10	14	18	22	26
0015-FAX	3	7	11	15	19	23	27
0016-KUW	4	8	12	16	20	24	28
0017-EOU	1	5	9	13	17	21	25
0018-UJL	2	6	10	14	18	22	26
0019-SWZ	3	7	11	15	19	23	27
0020-MXS	4	8	12	16	20	24	28
0021-ZTD	1	5	9	13	17	21	25
0022-KNX	2	6	10	14	18	22	26
0023-ZDD	3	7	11	15	19	23	27
0024-JUP	4	8	12	16	20	24	28
0026-PGM	1	5	9	13	17	21	25
0027-USC	2	6	10	14	18	22	26
0028-HPN	3	7	11	15	19	23	27
0029-AUT	4	8	12	16	20	24	28
0030-LNM	1	5	9	13	17	21	25
0031-YYA	2	6	10	14	18	22	26
0032-SHE	3	7	11	15	19	23	27
0033-SYH	4	8	12	16	20	24	28
0034-GKN	1	5	9	13	17	21	25
0035-ABX	2	6	10	14	18	22	26
0036-SLO	3	7	11	15	19	23	27
0037-RGT	4	8	12	16	20	24	28
0038-WSF	1	5	9	13	17	21	25
0041-BWS	2	6	10	14	18	22	26
0042-OEN	3	7	11	15	19	23	27

0043-ONB	4	8	12	16	20	24	28
0044-KFN	1	5	9	13	17	21	25
0045-TTH	2	6	10	14	18	22	26
0047-CXT	3	7	11	15	19	23	27
0048-BBX	4	8	12	16	20	24	28
0050-WYX	1	5	9	13	17	21	25
0051-UJS	2	6	10	14	18	22	26
0052-GZN	3	7	11	15	19	23	27
0053-YEJ	4	8	12	16	20	24	28
0054-YDX	1	5	9	13	17	21	25
0055-WKS	2	6	10	14	18	22	26
0056-JTC	3	7	11	15	19	23	27
0057-WNK	4	8	12	16	20	24	28
0058-FPU	1	5	9	13	17	21	25
0059-KGG	2	6	10	14	18	22	26
0061-TKH	3	7	11	15	19	23	27
0062-LCN	4	8	12	16	20	24	28
0063-WHE	1	5	9	13	17	21	25
0064-HRK	2	6	10	14	18	22	26
0065-TAD	3	7	11	15	19	23	27
0067-KSY	4	8	12	16	20	24	28
0068-KUS	1	5	9	13	17	21	25
0069-MOU	2	6	10	14	18	22	26
0070-DZS	3	7	11	15	19	23	27
0071-XXS	4	8	12	16	20	24	28
0072-EEA	1	5	9	13	17	21	25
0073-CEJ	2	6	10	14	18	22	26
0074-ZTS	3	7	11	15	19	23	27
0075-ZNO	4	8	12	16	20	24	28
0076-RFM	1	5	9	13	17	21	25
0078-XHZ	2	6	10	14	18	22	26
0079-PZO	3	7	11	15	19	23	27
0081-GYY	4	8	12	16	20	24	28
0082-LHA	1	5	9	13	17	21	25
0083-JMU	2	6	10	14	18	22	26
0084-SXU	3	7	11	15	19	23	27
0085-EUG	4	8	12	16	20	24	28
0086-ORZ	1	5	9	13	17	21	25
0087-CDG	2	6	10	14	18	22	26
0089-DUY	3	7	11	15	19	23	27
0090-OOE	4	8	12	16	20	24	28

0092-CLH	1	5	9	13	17	21	25
0093-DPJ	2	6	10	14	18	22	26
0094-HKF	3	7	11	15	19	23	27
0095-XRR	4	8	12	16	20	24	28
0096-LUH	1	5	9	13	17	21	25
0097-FTC	2	6	10	14	18	22	26
0099-WAO	3	7	11	15	19	23	27
0100-YWX	4	8	12	16	20	24	28
0102-ELX	1	5	9	13	17	21	25
0103-LDY	2	6	10	14	18	22	26
0104-WXM	3	7	11	15	19	23	27
0105-XZN	4	8	12	16	20	24	28
0107-RJF	1	5	9	13	17	21	25
0108-FZL	2	6	10	14	18	22	26
0109-CGO	3	7	11	15	19	23	27
0110-BKF	4	8	12	16	20	24	28
0111-DOZ	1	5	9	13	17	21	25
0112-ZBJ	2	6	10	14	18	22	26
0113-BWG	3	7	11	15	19	23	27
0114-GGF	4	8	12	16	20	24	28
0115-PPX	1	5	9	13	17	21	25
0116-JLE	2	6	10	14	18	22	26
0117-BZH	3	7	11	15	19	23	27
0118-BFB	4	8	12	16	20	24	28
0119-YLW	1	5	9	13	17	21	25
0120-NMK	2	6	10	14	18	22	26
0121-KPY	3	7	11	15	19	23	27
0122-XBO	4	8	12	16	20	24	28
0123-XFB	1	5	9	13	17	21	25
0124-BFW	2	6	10	14	18	22	26
0125-XPA	3	7	11	15	19	23	27
0127-EDS	4	8	12	16	20	24	28
0128-PDA	1	5	9	13	17	21	25
0129-CRH	2	6	10	14	18	22	26
0130-TWF	3	7	11	15	19	23	27
0131-JGG	4	8	12	16	20	24	28
0132-MZP	1	5	9	13	17	21	25
0133-YXL	2	6	10	14	18	22	26
0134-YJP	3	7	11	15	19	23	27
0135-XWZ	4	8	12	16	20	24	28
0136-PEJ	1	5	9	13	17	21	25

0137-OBY	2	6	10	14	18	22	26
0138-UWU	3	7	11	15	19	23	27
0139-TNF	4	8	12	16	20	24	28
0140-JUL	1	5	9	13	17	21	25
0141-JGK	2	6	10	14	18	22	26
0142-HMZ	3	7	11	15	19	23	27
0143-SCU	4	8	12	16	20	24	28
0144-OWZ	1	5	9	13	17	21	25
0145-AEB	2	6	10	14	18	22	26
0146-KAR	3	7	11	15	19	23	27
0147-DGF	4	8	12	16	20	24	28
0148-KMZ	1	5	9	13	17	21	25
0150-ZOM	2	6	10	14	18	22	26
0151-TXK	3	7	11	15	19	23	27
0152-UYZ	4	8	12	16	20	24	28
0153-WBN	1	5	9	13	17	21	25
0154-CSA	2	6	10	14	18	22	26
0155-ZXJ	3	7	11	15	19	23	27
0156-WTL	4	8	12	16	20	24	28
0157-CCN	1	5	9	13	17	21	25
0158-AKP	2	6	10	14	18	22	26
0160-BYD	3	7	11	15	19	23	27
0161-RNX	4	8	12	16	20	24	28
0162-RCB	1	5	9	13	17	21	25
0163-LUH	2	6	10	14	18	22	26
0164-ZWG	3	7	11	15	19	23	27
0165-MUZ	4	8	12	16	20	24	28
0166-WOD	1	5	9	13	17	21	25
0167-FLX	2	6	10	14	18	22	26
0168-GZW	3	7	11	15	19	23	27
0169-MLO	4	8	12	16	20	24	28
0170-HFS	1	5	9	13	17	21	25
0171-CWU	2	6	10	14	18	22	26
0173-CZT	3	7	11	15	19	23	27
0174-DXG	4	8	12	16	20	24	28
0175-WRO	1	5	9	13	17	21	25
0176-RTR	2	6	10	14	18	22	26
0177-RPO	3	7	11	15	19	23	27
0178-DHO	4	8	12	16	20	24	28
0179-OBN	1	5	9	13	17	21	25
0181-NXW	2	6	10	14	18	22	26

0182-AKS	3	7	11	15	19	23	27
0183-SFK	4	8	12	16	20	24	28
0184-SZB	1	5	9	13	17	21	25
0185-OON	2	6	10	14	18	22	26
0186-GNZ	3	7	11	15	19	23	27
0187-FZP	4	8	12	16	20	24	28
0188-DLK	1	5	9	13	17	21	25
0189-ZBH	2	6	10	14	18	22	26
0190-NGS	3	7	11	15	19	23	27
0191-OEW	4	8	12	16	20	24	28
0192-AXA	1	5	9	13	17	21	25
0193-YKK	2	6	10	14	18	22	26
0194-PLZ	3	7	11	15	19	23	27
0195-JZC	4	8	12	16	20	24	28
0197-KXL	1	5	9	13	17	21	25
0198-MAT	2	6	10	14	18	22	26
0199-FFL	3	7	11	15	19	23	27
0200-LBB	4	8	12	16	20	24	28
0201-ERM	1	5	9	13	17	21	25
0202-NTW	2	6	10	14	18	22	26
0203-DLB	3	7	11	15	19	23	27
0204-HOB	4	8	12	16	20	24	28
0205-PPU	1	5	9	13	17	21	25
0208-DZW	2	6	10	14	18	22	26
0209-HOG	3	7	11	15	19	23	27
0210-ETJ	4	8	12	16	20	24	28
0211-LZZ	1	5	9	13	17	21	25
0212-XDZ	2	6	10	14	18	22	26
0213-XNO	3	7	11	15	19	23	27
0214-BNG	4	8	12	16	20	24	28
0216-KKJ	1	5	9	13	17	21	25
0217-RWH	2	6	10	14	18	22	26
0218-WAK	3	7	11	15	19	23	27
0219-ANH	4	8	12	16	20	24	28
0220-RJZ	1	5	9	13	17	21	25
0221-JYO	2	6	10	14	18	22	26
0222-NCL	3	7	11	15	19	23	27
0223-MEO	4	8	12	16	20	24	28
0224-OGJ	1	5	9	13	17	21	25
0225-SZO	2	6	10	14	18	22	26
0226-FUN	3	7	11	15	19	23	27

0227-XHX	4	8	12	16	20	24	28
0229-DTH	1	5	9	13	17	21	25
0231-HMC	2	6	10	14	18	22	26
0232-WHB	3	7	11	15	19	23	27
0233-XTJ	4	8	12	16	20	24	28
0234-CSC	1	5	9	13	17	21	25
0235-BGU	2	6	10	14	18	22	26
0236-JBR	3	7	11	15	19	23	27
0237-MWS	4	8	12	16	20	24	28
0238-ZKB	1	5	9	13	17	21	25
0239-UJO	2	6	10	14	18	22	26
0240-SGH	3	7	11	15	19	23	27
0241-WNT	4	8	12	16	20	24	28
0242-BCO	1	5	9	13	17	21	25
0243-BUK	2	6	10	14	18	22	26
0244-GEL	3	7	11	15	19	23	27
0245-HEP	4	8	12	16	20	24	28
0246-LXY	1	5	9	13	17	21	25
0247-ZHL	2	6	10	14	18	22	26
0248-GGS	3	7	11	15	19	23	27
0249-HLY	4	8	12	16	20	24	28
0250-GOE	1	5	9	13	17	21	25
0252-FKS	2	6	10	14	18	22	26
0253-CKW	3	7	11	15	19	23	27
0254-HDJ	4	8	12	16	20	24	28
0255-OSN	1	5	9	13	17	21	25
0256-TDU	2	6	10	14	18	22	26
0257-ZUY	3	7	11	15	19	23	27
0258-MMH	4	8	12	16	20	24	28
0259-GUN	1	5	9	13	17	21	25
0260-SPU	2	6	10	14	18	22	26
0262-RFO	3	7	11	15	19	23	27
0263-SYX	4	8	12	16	20	24	28
0264-UCR	1	5	9	13	17	21	25
0265-MRX	2	6	10	14	18	22	26
0266-GZN	3	7	11	15	19	23	27
0267-GKZ	4	8	12	16	20	24	28
0268-AFR	1	5	9	13	17	21	25
0269-NRO	2	6	10	14	18	22	26
0270-ZTF	3	7	11	15	19	23	27
0271-GJF	4	8	12	16	20	24	28

0272-XRG	1	5	9	13	17	21	25
0273-JMS	2	6	10	14	18	22	26
0274-GCE	3	7	11	15	19	23	27
0275-CWK	4	8	12	16	20	24	28
0276-OGG	1	5	9	13	17	21	25
0277-LPX	2	6	10	14	18	22	26
0278-JEU	3	7	11	15	19	23	27
0279-RRT	4	8	12	16	20	24	28
0280-CHS	1	5	9	13	17	21	25
0281-JNR	2	6	10	14	18	22	26
0282-MEM	3	7	11	15	19	23	27
0283-NPN	4	8	12	16	20	24	28
0284-NDO	1	5	9	13	17	21	25
0285-BJG	2	6	10	14	18	22	26
0286-MMY	3	7	11	15	19	23	27
0287-HOC	4	8	12	16	20	24	28
0288-URG	1	5	9	13	17	21	25
0289-LSW	2	6	10	14	18	22	26
0290-KDC	3	7	11	15	19	23	27
0291-EUD	4	8	12	16	20	24	28
0292-OOA	1	5	9	13	17	21	25
0293-KMP	2	6	10	14	18	22	26
0294-ZPO	3	7	11	15	19	23	27
0295-JYO	4	8	12	16	20	24	28
0296-GTO	1	5	9	13	17	21	25
0297-GFA	2	6	10	14	18	22	26
0298-EZK	3	7	11	15	19	23	27
0299-NKH	4	8	12	16	20	24	28
0300-MBW	1	5	9	13	17	21	25
0301-JRE	2	6	10	14	18	22	26
0302-YZZ	3	7	11	15	19	23	27
0303-ESJ	4	8	12	16	20	24	28
0304-SYO	1	5	9	13	17	21	25
0306-KPA	2	6	10	14	18	22	26
0307-OAT	3	7	11	15	19	23	27
0308-TNP	4	8	12	16	20	24	28
0309-KBD	1	5	9	13	17	21	25
0310-LUZ	2	6	10	14	18	22	26
0311-DRX	3	7	11	15	19	23	27
0312-XTP	4	8	12	16	20	24	28
0313-JMX	1	5	9	13	17	21	25

0315-EOZ	2	6	10	14	18	22	26
0316-BYG	3	7	11	15	19	23	27
0317-WUM	4	8	12	16	20	24	28
0318-GJW	1	5	9	13	17	21	25
0320-COP	2	6	10	14	18	22	26
0321-SSF	3	7	11	15	19	23	27
0322-PXK	4	8	12	16	20	24	28
0323-PSJ	1	5	9	13	17	21	25
0324-COX	2	6	10	14	18	22	26
0325-ALD	3	7	11	15	19	23	27
0326-RLM	4	8	12	16	20	24	28
0327-XSY	1	5	9	13	17	21	25
0328-PKH	2	6	10	14	18	22	26
0329-WYW	3	7	11	15	19	23	27
0330-MPG	4	8	12	16	20	24	28
0331-AOO	1	5	9	13	17	21	25
0332-NLN	2	6	10	14	18	22	26
0333-JYD	3	7	11	15	19	23	27
0334-SGD	4	8	12	16	20	24	28
0335-WLE	1	5	9	13	17	21	25
0336-APD	2	6	10	14	18	22	26
0337-XHB	3	7	11	15	19	23	27
0338-ZXP	4	8	12	16	20	24	28
0339-SJB	1	5	9	13	17	21	25
0340-XPK	2	6	10	14	18	22	26
0341-BUX	3	7	11	15	19	23	27
0342-UNL	4	8	12	16	20	24	28
0343-ESG	1	5	9	13	17	21	25
0344-RHF	2	6	10	14	18	22	26
0345-UDM	3	7	11	15	19	23	27
0346-OME	4	8	12	16	20	24	28
0347-RDY	1	5	9	13	17	21	25
0348-LDK	2	6	10	14	18	22	26
0349-KCE	3	7	11	15	19	23	27
0350-YDJ	4	8	12	16	20	24	28

Answer form for multiple choice. You can make your own form, put please be clear and answer on one page. Do <u>not</u> submit solutions to the multiple-choice problems.

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

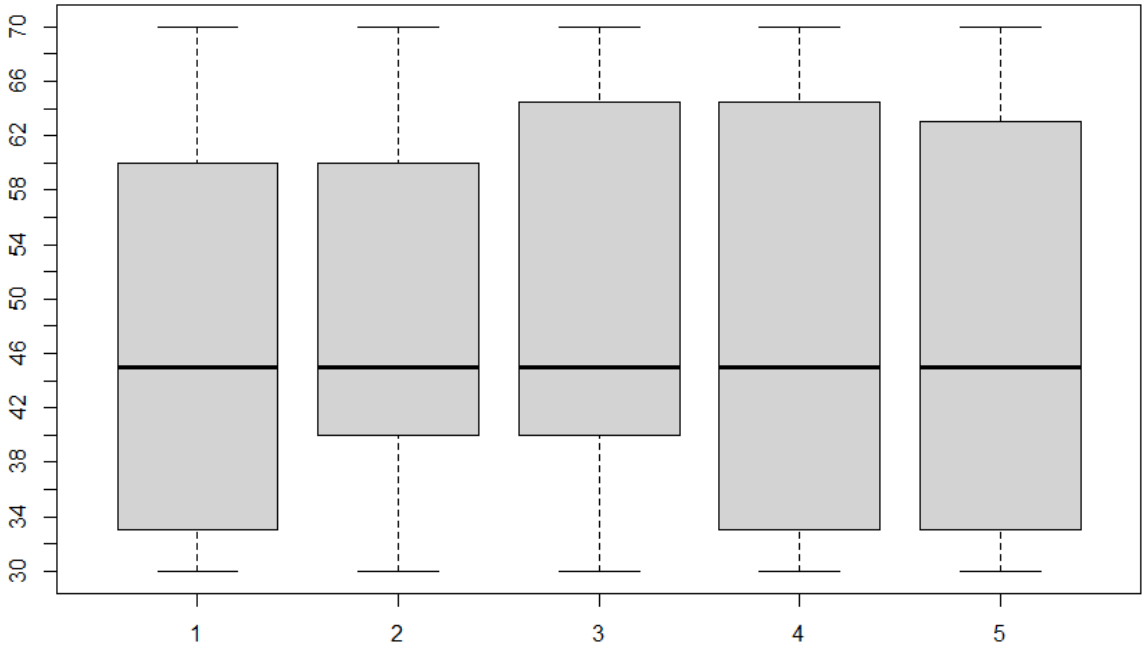
Problem 1

The numbers shown below is a sample of measured levels of precipitation in millimeters (rain) in a region of *Brazil*.

65, 60, 65, 43, 40, 50, 58, 33, 30, 42, 64, 33, 40, 31, 45, 60, 70, 58, 45, 44

a. **Which of the box plots in figure 1 represents the precipitation data?** (5p)
   (A) Boxplot 1
   (B) Boxplot 2
   (C) Boxplot 3
   (D) Boxplot 4
   (E) Boxplot 5

Figure 1.



Note: only the even numbers have tick-marks on the vertical axis.

As part of a study, one thousand 30-year-old adults were asked about their monthly income. The table below shows the number of respondents, by gender and income category. For example, 83 respondents were women who had an income between 15001 and 20000 SEK/month:

	Income category, SEK/month					
Gender	**≤15000**	**15001-20000**	**20001-25000**	**25001-30000**	**30001-35000**	**>35000**
Women	199	83	92	59	29	24
Men	144	51	83	97	67	72
All	343	134	175	156	96	96

	Income category					
	≤15000	≤20000	≤25000	≤30000	≤35000	>35000
Women			x			
Men						

b. **Create a table of <u>cumulative relative frequencies, conditional on gender.</u> What should the value in the cell maked "x" be?** Choose the alternative closest to your answer. You do not have to fill out the rest of the table. (5p)

(A) 16%
(B) 19%
(C) 47%
(D) 54%
(E) 77%

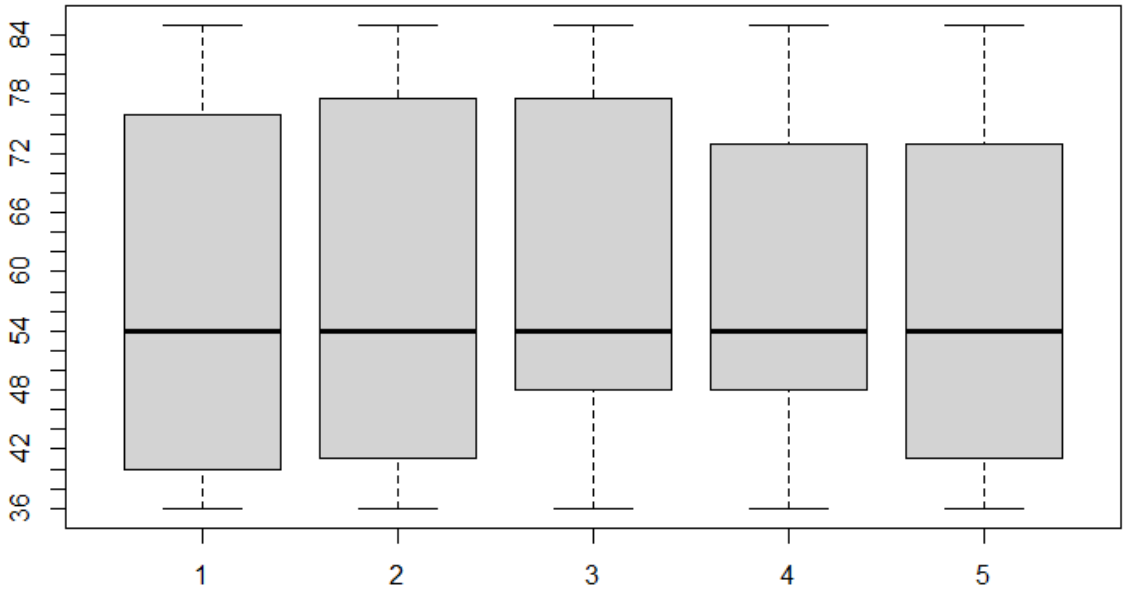
Problem 2

The numbers shown below is a sample of measured levels of precipitation in millimeters (rain) in a region of *Brazil*.

79, 36, 53, 70, 77, 85, 70, 79, 40, 54, 48, 48, 51, 54, 61, 73, 73, 52, 40, 38

a. **Which of the box plots in Figure 2 represents the precipitation data?** (5p)
   (A) Boxplot 1
   (B) Boxplot 2
   (C) Boxplot 3
   (D) Boxplot 4
   (E) Boxplot 5

Figure 2.



Note: only the even numbers have tick-marks on the vertical axis

As part of a study, one thousand 30-year-old adults were asked about their monthly income. The table below shows the number of respondents, by gender and income category. For example, 83 respondents were women who had an income between 15001 and 20000 SEK/month:

	Income category, SEK/month					
Gender	**≤15000**	**15001-20000**	**20001-25000**	**25001-30000**	**30001-35000**	**>35000**
Women	199	83	92	59	29	24
Men	144	51	83	97	67	72
All	343	134	175	156	96	96

	Income category					
	≤15000	≤20000	≤25000	≤30000	≤35000	>35000
Women						
Men					x	

a. **Create a table of <u>cumulative relative frequencies, conditional on gender.</u> What should the value in the cell maked "x" be?** Choose the alternative closest to your answer. You do not have to fill out the rest of the table. (5p)
   (A) 6%
   (B) 13%
   (C) 60%
   (D) 86%
   (E) 90%

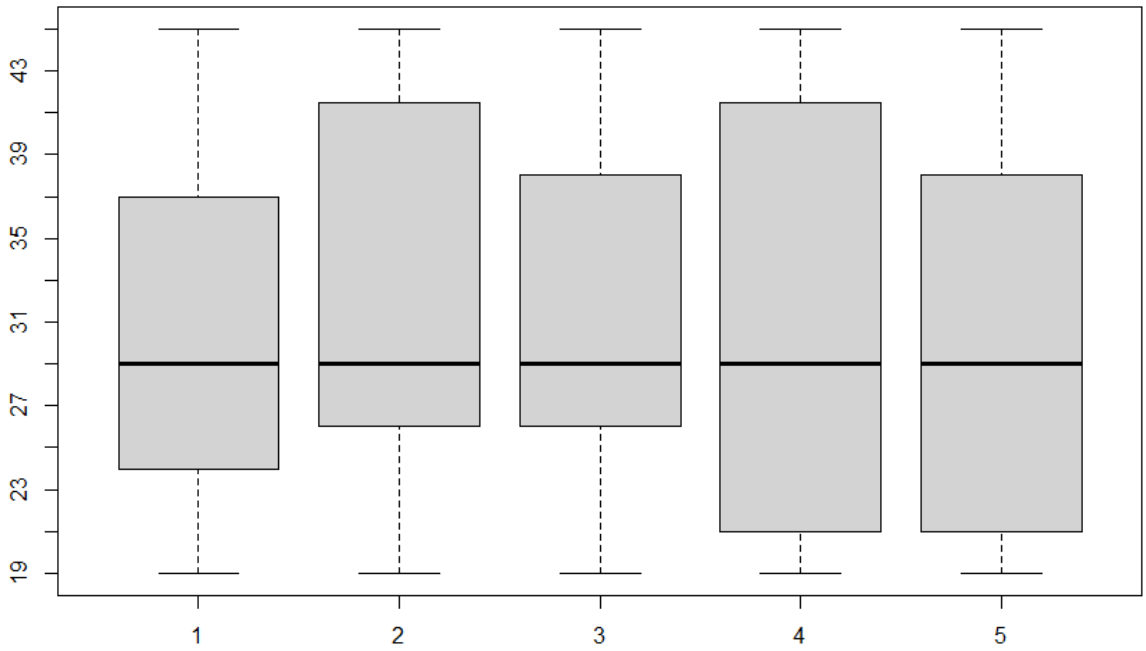
Problem 3

The numbers shown below is a sample of measured levels of precipitation in millimeters (rain) in a region of *Brazil*.

26, 29, 21, 37, 20, 32, 28, 38, 27, 29, 42, 45, 28, 41, 19, 37, 42, 21, 38, 26

a. **Which of the box plots in figure 1 represents the precipitation data?** (5p)
   (A) Boxplot 1
   (B) Boxplot 2
   (C) Boxplot 3
   (D) Boxplot 4
   (E) Boxplot 5

Figure 3.



Note: only the odd numbers have a tick-marks on the vertical axis.

As part of a study, one thousand 30-year-old adults were asked about their monthly income. The table below shows the number of respondents, by gender and income category. For example, 83 respondents were women who had an income between 15001 and 20000 SEK/month:

	Income category, SEK/month					
Gender	**≤15000**	**15001-20000**	**20001-25000**	**25001-30000**	**30001-35000**	**>35000**
Women	199	83	92	59	29	24
Men	144	51	83	97	67	72
All	343	134	175	156	96	96

	Income category					
	≤15000	≤20000	≤25000	≤30000	≤35000	>35000
Women				x		
Men						

b. **Create a table of <u>cumulative relative frequencies, conditional on gender.</u> What should the value in the cell maked "x" be?** Choose the alternative closest to your answer. You do not have to fill out the rest of the table. (5p)

(A) 12%
(B) 19%
(C) 47%
(D) 73%
(E) 89%

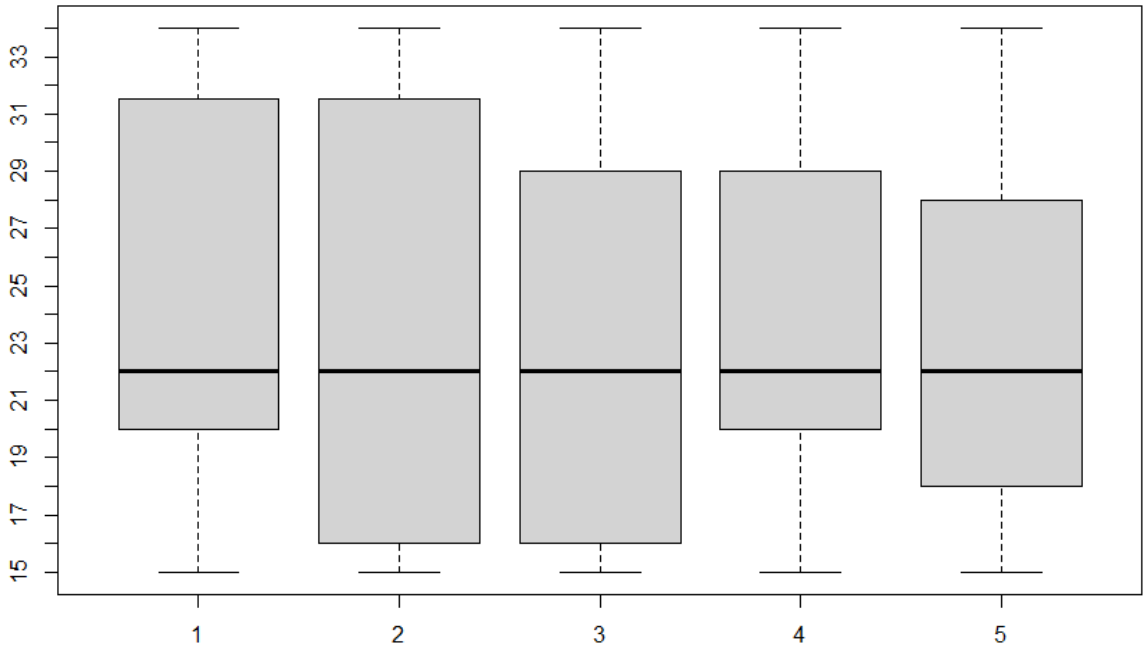
Problem 4

The numbers shown below is a sample of measured levels of precipitation in millimeters (rain) in a region of *Brazil*.

22, 34, 22, 16, 32, 21, 15, 20, 29, 24, 29, 32, 16, 21, 28, 20, 15, 31, 28, 22

a. **Which of the box plots in Figure 2 represents the precipitation data?** (5p)
   (A) Boxplot 1
   (B) Boxplot 2
   (C) Boxplot 3
   (D) Boxplot 4
   (E) Boxplot 5

Figure 4.



As part of a study, one thousand 30-year-old adults were asked about their monthly income. The table below shows the number of respondents, by gender and income category. For example, 83 respondents were women who had an income between 15001 and 20000 SEK/month:

	Income category, SEK/month					
Gender	**≤15000**	**15001-20000**	**20001-25000**	**25001-30000**	**30001-35000**	**>35000**
Women	199	83	92	59	29	24
Men	144	51	83	97	67	72
All	343	134	175	156	96	96

	Income category					
	≤15000	≤20000	≤25000	≤30000	≤35000	>35000
Women						
Men				x		

b. **Create a table of <u>cumulative relative frequencies, conditional on gender.</u> What should the value in the cell maked "x" be?** Choose the alternative closest to your answer. You do not have to fill out the table. (5p)

(A) 12%
(B) 19%
(C) 47%
(D) 53%
(E) 73%

Problem 5

Judith plays the first round of a chess tournament. She can win, draw (neither player wins), or lose. The probabilities are given below:

1st game	win	draw	loss
Probability	0.52	0.15	0.33

The next table shows the probabilities of winning the second game, conditional on the result of the first game:

		2nd			
		win	**draw**	**loss**	**sum**
1st	**win**	0.5	0.2	0.3	1
	draw	0.4	0.1	0.5	1
	loss	0.25	0.05	0.7	1

a. **Find the total probability that Judith wins the second game (without knowing the outcome of the first game).** Choose the alternative closest to your answer (5p)
   (A) 0.38
   (B) 0.39
   (C) 0.40
   (D) 0.41
   (E) 0.42

In the last round, she faces the current tournament leader. If she wins the game, she will win first price worth 1000 SEK. If she draws the game, she will win second price worth 500 SEK. If she loses, she will win third price worth 250 SEK. The probabilities can be found in the table below:

	win	draw	loss
Probability	0.2	0.4	0.4

b. **Find the expected value and standard deviation of her winnings in SEK.** Choose the alternative closest to your answer. (5p)
   (A) Expected value 500, standard deviation 274
   (B) Expected value 550, standard deviation 274
   (C) Expected value 500, standard deviation 559
   (D) Expected value 583, standard deviation 274
   (E) Expected value 583, standard deviation 559

Problem 6

Judith plays the first round of a chess tournament. She can win, draw (neither player wins), or lose. The probabilities are given below:

1st game	win	draw	loss
Probability	0.4	0.1	0.5

The next table shows the probabilities of winning the second game, conditional on the result of the first game:

		2nd			
		win	**draw**	**loss**	**sum**
1st	**win**	0.5	0.2	0.3	1
	draw	0.4	0.1	0.5	1
	loss	0.25	0.05	0.7	1

a. **Find the total probability that Judith wins the second game (without knowing the outcome of the first game).** Choose the alternative closest to your answer (5p)
   (A) 0.37
   (B) 0.38
   (C) 0.39
   (D) 0.40
   (E) 0.41

In the last round, she faces the current tournament leader. If she wins the game, she will win first price worth 1000 SEK. If she draws the game, she will win second price worth 500 SEK. If she loses, she will win third price worth 300 SEK. The probabilities can be found in the table below:

	win	draw	loss
Probability	0.25	0.15	0.6

b. **Find the expected value and standard deviation of her winnings in SEK.** Choose the alternative closest to your answer. (5p)
   (A) Expected value 505, standard deviation 294
   (B) Expected value 505, standard deviation 426
   (C) Expected value 600, standard deviation 294
   (D) Expected value 600, standard deviation 426
   (E) Expected value 550, standard deviation 294

Problem 7

Judith plays the first round of a chess tournament. She can win, draw (neither player wins), or lose. The probabilities are given below:

1st game	win	draw	loss
Probability	0.7	0.1	0.2

The next table shows the probabilities of winning the second game, conditional on the result of the first game:

		2nd			
		win	**draw**	**loss**	**sum**
1st	**win**	0.5	0.2	0.3	1
	draw	0.4	0.1	0.5	1
	loss	0.25	0.05	0.7	1

a. **Find the total probability that Judith wins the second game (without knowing the outcome of the first game).** Choose the alternative closest to your answer (5p)
   (A) 0.41
   (B) 0.42
   (C) 0.43
   (D) 0.44
   (E) 0.45

In the last round, she faces the current tournament leader. If she wins the game, she will win first price worth 1000 SEK. If she draws the game, she will win second price worth 500 SEK. If she loses, she will win third price worth 250 SEK. The probabilities can be found in the table below:

	win	draw	loss
Probability	0.1	0.3	0.6

b. **Find the expected value and standard deviation of her winnings in SEK.** Choose the alternative closest to your answer. (5p)
   (A) Expected value 430, standard deviation 229
   (B) Expected value 400, standard deviation 210
   (C) Expected value 430, standard deviation 210
   (D) Expected value 400, standard deviation 229
   (E) Expected value 500, standard deviation 210

Problem 8

Judith plays the first round of a chess tournament. She can win, draw (neither player wins), or lose. The probabilities are given below:

1st game	win	draw	loss
Probability	0.5	0.1	0.4

The next table shows the probabilities of winning the second game, conditional on the result of the first game:

		2nd			
		win	**draw**	**loss**	**sum**
1st	**win**	0.5	0.2	0.3	1
	draw	0.4	0.1	0.5	1
	loss	0.25	0.05	0.7	1

a. **Find the total probability that Judith wins the second game (without knowing the outcome of the first game).** Choose the alternative closest to your answer (5p)
   (A) 0.39
   (B) 0.40
   (C) 0.41
   (D) 0.42
   (E) 0.43

In the last round, she faces the current tournament leader. If she wins the game, she will win first price worth 1000 SEK. If she draws the game, she will win second price worth 500 SEK. If she loses, she will win third price worth 300 SEK. The probabilities can be found in the table below:

	win	draw	loss
Probability	0.2	0.4	0.4

b. **Find the expected value and standard deviation of her winnings in SEK.** Choose the alternative closest to your answer. (5p)
   (A) Expected value 500, standard deviation 273
   (B) Expected value 520, standard deviation 273
   (C) Expected value 500, standard deviation 256
   (D) Expected value 600, standard deviation 273
   (E) Expected value 520, standard deviation 256

Problem 9

A manager a textile factory estimates that the number of t-shirts produced per work day is approximately normally distributed with mean 2300 and standard deviation 200. Each t-shirt brings in $3 in revenue.

a. **Find the probability that the revenue from t-shirts will be at least $6000 tomorrow, according to the manager's model.** Choose the alternative closest to your answer. (5p)
   (A) 7%
   (B) 25%
   (C) 55%
   (D) 75%
   (E) 93%

Assume that 95% of all Swedes own at least one computer. We draw a simple random sample of 20 Swedes.

b. **What is the probability that 18 or fewer of these 20 own at least one computer?** Choose the alterative closest to your answer. (5p)
   (A) 8%
   (B) 22%
   (C) 26%
   (D) 64%
   (E) 74%

Problem 10

A manager a textile factory estimates that the number of t-shirts produced per work day is approximately normally distributed with mean 3500 and standard deviation 400. Each t-shirt brings in $4 in revenue.

a. **Find the probability that the revenue from t-shirts will be at least $10000 tomorrow, according to the manager's model.** Choose the alternative closest to your answer. (5p)
   (A) 5%
   (B) 38%
   (C) 51%
   (D) 62%
   (E) 99%

Assume that 95% of all Swedes own at least one computer. We draw a simple random sample of 20 Swedes.

b. **What is the probability that 18 or fewer of these 20 own at least one computer?** Choose the alterative closest to your answer. (5p)
   (A) 8%
   (B) 22%
   (C) 26%
   (D) 64%
   (E) 74%

Problem 11

A manager a textile factory estimates that the number of t-shirts produced per work day is approximately normally distributed with mean 2300 and standard deviation 400. Each t-shirt brings in $5 in revenue.

a. **Find the probability that the revenue from t-shirts will be at least $10000 tomorrow, according to the manager's model.** Choose the alternative closest to your answer. (5p)
   (A) 7%
   (B) 33%
   (C) 58%
   (D) 77%
   (E) 93%

Assume that 95% of all Swedes own at least one computer. We draw a simple random sample of 20 Swedes.

b. **What is the probability that 18 or fewer of these 20 own at least one computer?** Choose the alterative closest to your answer. (5p)
   (A) 8%
   (B) 22%
   (C) 26%
   (D) 64%
   (E) 74%

Problem 12

A manager a textile factory estimates that the number of t-shirts produced per work day is approximately normally distributed with mean 3000 and standard deviation 200. Each t-shirt brings in $3 in revenue.

a. **Find the probability that the revenue from t-shirts will be at least $8250 tomorrow, according to the manager's model.** Choose the alternative closest to your answer. (5p)
   (A) 5%
   (B) 20%
   (C) 63%
   (D) 89%
   (E) 95%

Assume that 95% of all Swedes own at least one computer. We draw a simple random sample of 20 Swedes.

b. **What is the probability that 18 or fewer of these 20 own at least one computer?** Choose the alterative closest to your answer. (5p)
   (A) 8%
   (B) 22%
   (C) 26%
   (D) 89%
   (E) 74%

Problem 13

A farmer wants to compare two different kinds of corn seed, "type A" and "type B." She uses type A for half her field and type B for the other half. The two halves are otherwise almost identical. In the table below, you can find three years of crops, in bushels per acre:

	2018	2019	2020
type A	150	140	89
type B	130	134	83

This is paired data. The reason is that the yield of any crop, like corn, depends heavily on the weather, which is the same for the two parts of the field each year.

You can assume that all three years are independent of each other. Also, assume that the crops are normally distributed.

a. **Find a 95% confidence interval for the difference in bushels per acre,** $\mu_A - \mu_B$. Choose the alternative closest to your answer. (5p)
   (A) (1.5, 19.8)
   (B) (-3, 24.3)
   (C) (-3.1, 25.5)
   (D) (-3.3, 26.7)
   (E) (-9.4, 30.7)

Assume that the natural lifespan of a lab mouse kept in a cage has a known population standard deviation of 100 days and that the lifespan is normally distributed. A researcher wants to estimate the population lifespan of such a mouse, with a 95% confidence interval.

b. **Find the minimum sample size that the researcher needs to use to guarantee that the margin of error is at most 10 days**. Choose the alternative closest to your answer. (5p)
   (A) 40
   (B) 246
   (C) 385
   (D) 984
   (E) 1537

Problem 14

A farmer wants to compare two different kinds of corn seed, "type A" and "type B." She uses type A for half her field and type B for the other half. The two halves are otherwise almost identical. In the table below, you can find three years of crops, in bushels per acre:

	2018	2019	2020
type A	140	89	150
type B	134	83	130

This is paired data. The reason is that the yield of any crop, like corn, depends heavily on the weather, which is the same for the two parts of the field each year.

You can assume that all three years are independent of each other. Also, assume that the crops are normally distributed.

a. **Find a 95% confidence interval for the difference in bushels per acre,** $\mu_A - \mu_B$. Choose the alternative closest to your answer. (5p)
   (A) (1.5, 19.8)
   (B) (-3, 24.3)
   (C) (-3.1, 25.5)
   (D) (-3.3, 26.7)
   (E) (-9.4, 30.7)

Assume that the natural lifespan of a lab mouse kept in a cage has a known population standard deviation of 80 days and that the lifespan is normally distributed. A researcher wants to estimate the population lifespan of such a mouse, with a 95% confidence interval.

b. **Find the minimum sample size that the researcher needs to use to guarantee that the margin of error is at most 10 days**. Choose the alternative closest to your answer. (5p)
   (A) 40
   (B) 246
   (C) 385
   (D) 984
   (E) 1537

Problem 15

A farmer wants to compare two different kinds of corn seed, "type A" and "type B." She uses type A for half her field and type B for the other half. The two halves are otherwise almost identical. In the table below, you can find three years of crops, in bushels per acre:

	2018	2019	2020
type A	140	150	89
type B	134	130	83

This is paired data. The reason is that the yield of any crop, like corn, depends heavily on the weather, which is the same for the two parts of the field each year.

You can assume that all three years are independent of each other. Also, assume that the crops are normally distributed.

a. **Find a 95% confidence interval for the difference in bushels per acre,** $\mu_A - \mu_B$. Choose the alternative closest to your answer. (5p)
   (A) (1.5, 19.8)
   (B) (-3, 24.3)
   (C) (-3.1, 25.5)
   (D) (-3.3, 26.7)
   (E) (-9.4, 30.7)

Assume that the natural lifespan of a lab mouse kept in a cage has a known population standard deviation of 100 days and that the lifespan is normally distributed. A researcher wants to estimate the population lifespan of such a mouse, with a 95% confidence interval.

b. **Find the minimum sample size that the researcher needs to use to guarantee that the margin of error is at most 5 days**. Choose the alternative closest to your answer. (5p)
   (A) 40
   (B) 246
   (C) 385
   (D) 984
   (E) 1537

Problem 16

A farmer wants to compare two different kinds of corn seed, "type A" and "type B." She uses type A for half her field and type B for the other half. The two halves are otherwise almost identical. In the table below, you can find three years of crops, in bushels per acre:

	2018	2019	2020
type A	89	150	140
type B	83	130	134

This is paired data. The reason is that the yield of any crop, like corn, depends heavily on the weather, which is the same for the two parts of the field each year.

You can assume that all three years are independent of each other. Also, assume that the crops are normally distributed.

a. **Find a 95% confidence interval for the difference in bushels per acre,** $\mu_A - \mu_B$. Choose the alternative closest to your answer. (5p)
   (A) (1.5, 19.8)
   (B) (-3, 24.3)
   (C) (-3.1, 25.5)
   (D) (-3.3, 26.7)
   (E) (-9.4, 30.7)

Assume that the natural lifespan of a lab mouse kept in a cage has a known population standard deviation of 80 days and that the lifespan is normally distributed. A researcher wants to estimate the population lifespan of such a mouse, with a 95% confidence interval.

b. **Find the minimum sample size that the researcher needs to use to guarantee that the margin of error is at most 5 days**. Choose the alternative closest to your answer. (5p)
   (A) 40
   (B) 246
   (C) 385
   (D) 984
   (E) 1537

Problem 17

A European roulette wheel has 37 pockets numbered 0 to 36 (see figure 5). A self-proclaimed roulette expert claims to be good at predicting which number the ball will land on, by watching the ball spin around the wheel before it lands.

You decide to test this claim at the 10% level of significance. Let P be the true probability that the expert guesses correctly. Your hypotheses are

$$H_0: P = \frac{1}{37}$$
$$H_1: P > \frac{1}{37}$$

You patiently observe the expert guess the outcome by watching the ball before it lands 1000 times. He guesses correctly 33 times out of the 1000.

a. **What is the critical value of the test?** Choose the value closest to your answer. (5p)
   (A) 1.28
   (B) 1.64
   (C) 1.96
   (D) 2.32
   (E) 2.58

b. **What is the value of the test variable?** Choose the value closest to your answer. (5p)
   (A) 0.39
   (B) 0.55
   (C) 0.80
   (D) 1.02
   (E) 1.16

Figure 5

Picture by Ralf Roletschek
Source: https://commons.wikimedia.org/wiki/File:13-02-27-spielbank-wiesbaden-by-RalfR-094.jpg

Problem 18

In the casino game craps, one of the players roll two dice by throwing them onto a special "craps table," see figure 6. A self-proclaimed craps expert claims to be able to roll a seven (the sum of the dice) more often than other people, by holding and throwing the dice a particular way.

You decide to test this claim at the 10% level of significance. Let P be the true probability that the expert rolls a seven. Your hypotheses are

$$H_0: P = \frac{1}{6}$$
$$H_1: P > \frac{1}{6}$$

You patiently observe the expert attempt to roll seven, 500 times. The expert succeeds 90 times.

a. **What is the critical value of the test?** Choose the value closest to your answer. (5p)
   (A) 1.28
   (B) 1.64
   (C) 1.96
   (D) 2.32
   (E) 2.58

b. **What is the value of the test variable?** Choose the value closest to your answer. (5p)
   (A) 0.61
   (B) 0.80
   (C) 1.44
   (D) 2.01
   (E) 2.21

Figure 6

Picture by Dalton Precht
Source: https://www.dvidshub.net/image/1028190

Problem 19

A European roulette wheel has 37 pockets numbered 0 to 36 (see figure 5). A self-proclaimed roulette expert claims to be good at predicting which number the ball will land on, by watching the ball spin around the wheel before it lands.

You decide to test this claim at the 10% level of significance. Let P be the true probability that the expert guesses correctly. Your hypotheses are

$$H_0: P = \frac{1}{37}$$
$$H_1: P > \frac{1}{37}$$

You patiently observe the expert guess the outcome by watching the ball before it lands 500 times. He guesses correctly 15 times out of the 500.

a. **What is the critical value of the test?** Choose the value closest to your answer. (5p)
   (A) 1.28
   (B) 1.64
   (C) 1.96
   (D) 2.32
   (E) 2.58

b. **What is the value of the test variable?** Choose the value closest to your answer. (5p)
   (A) 0.41
   (B) 0.91
   (C) 1.26
   (D) 1.54
   (E) 1.74

Figure 5

Picture by Ralf Roletschek
Source: https://commons.wikimedia.org/wiki/File:13-02-27-spielbank-wiesbaden-by-RalfR-094.jpg

Problem 20

In the casino game craps, one of the players roll two dice by throwing them onto a special "craps table," see figure 6. A self-proclaimed craps expert claims to be able to roll a seven (the sum of the dice) more often than other people, by holding and throwing the dice a particular way.

You decide to test this claim at the 10% level of significance. Let P be the true probability that the expert rolls a seven. Your hypotheses are

$$H_0: P = \frac{1}{6}$$
$$H_1: P > \frac{1}{6}$$

You patiently observe the expert attempt to roll seven, 1000 times. The expert succeeds 175 times.

a. **What is the critical value of the test?** Choose the value closest to your answer. (5p)
   (A) 1.28
   (B) 1.64
   (C) 1.96
   (D) 2.32
   (E) 2.58

b. **What is the value of the test variable?** Choose the value closest to your answer. (5p)
   (A) 0.62
   (B) 0.71
   (C) 1.44
   (D) 2.01
   (E) 2.21

Figure 6

Picture by Dalton Precht
Source: https://www.dvidshub.net/image/1028190

Problem 21

A researcher studies the exercise habits of Swedish high-school students (gymnasiet). She collects a simple random sample of students for six of Sweden's biggest cities. As part of the survey, she asks the students if they are part of an organized sports club. The results can be found in the table below:

Organized Sports	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Part of	6	5	38	12	6	8	75
Not Part	14	15	52	16	14	14	125
	20	20	90	28	20	22	200

Test at the 5% level of significance whether being part of an organized sports club is independent of home city.

a. **State assumption, hypotheses, and test variable.** (5p)
b. **Find the critical value, and state the decision rule.** (5p)
c. **Calculate the observed test variable and state the conclusion of the test.** (5p)

The survey also includes the question "Have you used cannabis in the last 12 months?" The answers can be found in the table below.

Cannabis Use	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Yes	3	2	15	6	6	3	35
No	17	18	75	22	14	19	165
	20	20	90	28	20	22	200

d. **Is it possible to use this data to test whether cannabis use is independent of city? Be as precise as you can.** Either way, you do not need to perform the test. Hint: you might have to do some calculation to answer this question precisely. (5p)
e. As part of the study, the researcher wants to carry out long in-person interviews with three of the students. She picks three students from the total sample of 200 at random. **Find the probability that none of the three are from Stockholm.** Show your work. (5p)

Problem 22

A researcher studies the exercise habits of Swedish high-school students (gymnasiet). She collects a simple random sample of students for six of Sweden's biggest cities. As part of the survey, she asks the students if they are part of an organized sports club. The results can be found in the table below:

Organized Sports	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Part	2	8	40	9	8	6	73
Not Part	18	12	60	17	8	12	127
	20	20	100	26	16	18	200

Test at the 5% level of significance whether being part of an organized sports club is independent of home city.

a. **State assumption, hypotheses, and test variable.** (5p)
b. **Find the critical value, and state the decision rule.** (5p)
c. **Calculate the observed test variable and state the conclusion of the test.** (5p)

The survey also includes the question "Have you used cannabis in the last 12 months?" The answers can be found in the table below.

Cannabis Use	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Yes	2	3	12	2	3	1	23
No	18	17	88	24	13	17	177
	20	20	100	26	16	18	200

d. **Is it possible to use this data to test whether cannabis use is independent of city? Be as precise as you can.** Either way, you do not need to perform the test. Hint: you might have to do some calculation to answer this question precisely. (5p)
e. As part of the study, the researcher wants to carry out long in-person interviews with three of the students. She picks three students from the total sample of 200 at random. **Find the probability that none of the three are from Stockholm.** Show your work. (5p)

Problem 23

A researcher studies the exercise habits of Swedish high-school students (gymnasiet). She collects a simple random sample of students for six of Sweden's biggest cities. As part of the survey, she asks the students if they are part of an organized sports club. The results can be found in the table below:

Organized Sports	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Part	2	15	27	10	5	8	67
Not Part	18	25	53	18	11	8	133
	20	40	80	28	16	16	200

Test at the 5% level of significance whether being part of an organized sports club is independent of home city.

a. **State assumption, hypotheses, and test variable.** (5p)
b. **Find the critical value, and state the decision rule.** (5p)
c. **Calculate the observed test variable and state the conclusion of the test.** (5p)

The survey also includes the question "Have you used cannabis in the last 12 months?" The answers can be found in the table below.

Cannabis use	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Yes	2	5	10	2	3	1	23
No	18	35	70	26	13	15	177
	20	40	80	28	16	16	200

d. **Is it possible to use this data to test whether cannabis use is independent of city? Be as precise as you can.** Either way, you do not need to perform the test. Hint: you might have to do some calculation to answer this question precisely. (5p)
e. As part of the study, the researcher wants to carry out long in-person interviews with three of the students. She picks three students from the total sample of 200 at random. **Find the probability that all of three are from Göteborg.** Show your work. (5p)

Problem 24

A researcher studies the exercise habits of Swedish high-school students (gymnasiet). She collects a simple random sample of students for six of Sweden's biggest cities. As part of the survey, she asks the students if they are part of an organized sports club. The results can be found in the table below:

Organized Sports	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Part	4	14	30	2	11	7	68
Not Part	21	21	50	22	9	9	132
	25	35	80	24	20	16	200

Test at the 5% level of significance whether being part of an organized sports club is independent of home city.

a. **State assumption, hypotheses, and test variable.** (5p)
b. **Find the critical value, and state the decision rule.** (5p)
c. **Calculate the observed test variable and state the conclusion of the test.** (5p)

The survey also includes the question "Have you used cannabis in the last 12 months?" The answers can be found in the table below.

Cannabis use	Malmö	Göteborg	Stockholm	Uppsala	Västerås	Örebro	
Yes	2	5	9	2	2	3	23
No	23	30	71	22	18	13	177
	25	35	80	24	20	16	200

d. **Is it possible to use this data to test whether cannabis use is independent of city? Be as precise as you can.** Either way, you do not need to perform the test. Hint: you might have to do some calculation to answer this question precisely. (5p)
e. As part of the study, the researcher wants to carry out long in-person interviews with three of the students. She picks three students from the total sample of 200 at random. **Find the probability that none of the three are from Göteborg.** Show your work. (5p)

Problem 25

David, a computer scientist, wants to predict the price of the cryptocurrency Bitcoin.

He writes a programming script that collects and categorizes tweets from Twitter about Bitcoin, every day for 11 days; he uses this data to calculate a value called *Bullishness* (x). High Bullishness values (more than 0.7) means that there were more twitter recommendations to buy Bitcoin than there were recommendations to sell, and vice versa. David believes that the Bullishness value can help him predict the following day's returns (one-day percentage change in the price of the cryptocurrency).

You can find 11 days of data below:

Return (y)	-1	0	1	-2	0	1	2	1	8	1	0
Bullishness, previous day (x)	0.6	1.8	1.2	1.6	1.1	1.3	1.7	1.7	2.6	0.7	2.2

David decides on the following model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where *Y* is the daily percentage return of Bitcoin and *X* is the Bullishness from the previous day. Remember: you need to show your work!

a. **Estimate the sample variance of *X* 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)

To make the following calculations easier, you are given the residual variance: $s_e^2 = 4.96$.

c. **Calculate a 95% prediction interval for tomorrow's daily return of Bitcoin, given that the Bullishness was (x) was 2.0 today. Interpret the result.** (5p)

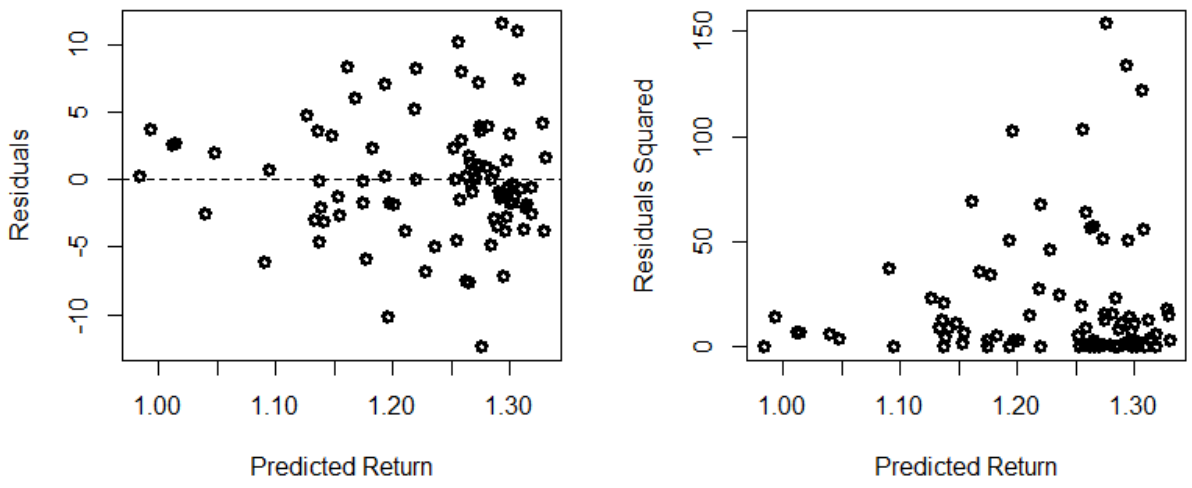
David wants to test whether the slope is significantly different from zero, at the 5% level of significance.

d. **Calculate the value of the test variable. Use this value to find the p-value of the test. Use the p-value to determine if David can reject the null.** You do not need to perform the rest of the hypothesis test. (5p)

David collects more data. After estimating the same model to the new data, he creates some plots. In Figure 9, you can see the residuals plotted against the predicted returns and the squared residuals plotted against the percentage returns.

e. **Interpret the graphs. Briefly explain why this is a sign of a problem with the model.** (5p)

Figure 9.



Note: Crypto currencies happened to show extremely high returns during this period. Past returns are no guarantee of future returns. This problem is loosely inspired by the paper Olivier Kraaijeveld, Johannes De Smedt, *The predictive power of public Twitter sentiment for forecasting cryptocurrency prices*, Journal of International Financial Markets, Institutions and Money, Volume 65,2020,101188,ISSN 1042-4431, https://doi.org/10.1016/j.intfin.2020.101188.
However, the bullishness data for this problem is made up. Reading the paper will not help you to solve the problem.

Problem 26

David, a computer scientist, wants to predict the price of the cryptocurrency Ethereum.

He writes a programming script that collects and categorizes tweets from Twitter about Ethereum, every day for 11 days; he uses this data to calculate a value called *Bullishness* (x). High Bullishness values (more than 0.7) means that there were more twitter recommendations to buy Ethereum than there were recommendations to sell, and vice versa. David believes that the Bullishness value can help him predict the following day's returns (one-day percentage change in the price of the cryptocurrency).

You can find 11 days of data below:

Return (y)	-2	0	-1	-3	1	3	0	-3	6	6	-1
Bullishness, previous day (x)	1.2	0.6	1.3	1	1.3	1.8	0.3	1	1.5	0.7	0.3

David decides on the following model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where *Y* is the daily percentage return of Bitcoin and *X* is the Bullishness from the previous day. Remember: you need to show your work!

a. **Estimate the sample variance of *X* 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)

To make the following calculations easier, you are given the residual variance: $s_e^2 = 10.7$.

c. **Calculate a 95% prediction interval for tomorrow's daily return of Ethereum, given that the Bullishness was (x) was 1.5 today. Interpret the result.** (5p)

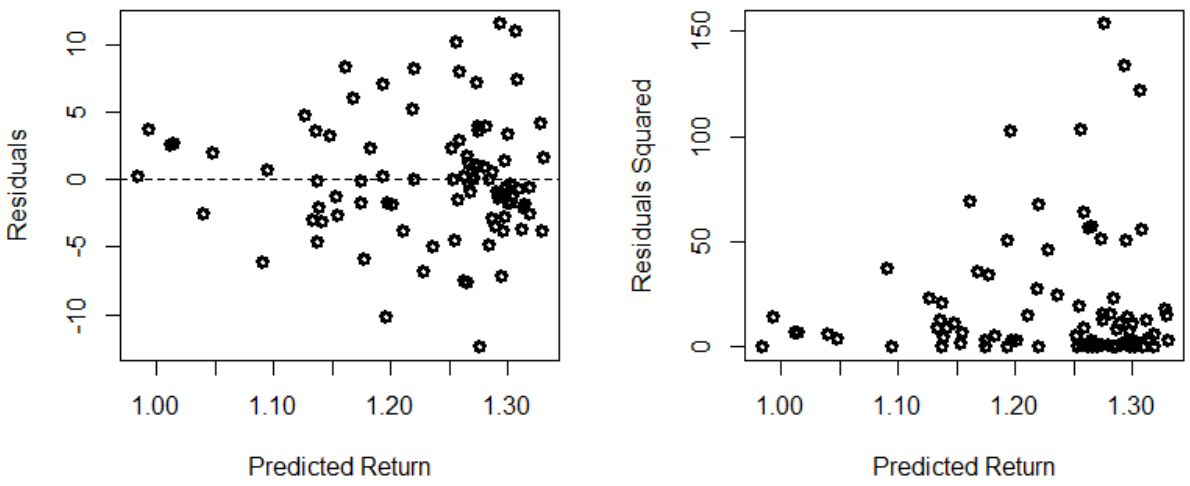
David wants to test whether the slope is significantly different from zero, at the 5% level of significance.

d. **Calculate the value of the test variable. Use this value to find the p-value of the test. Use the p-value to determine if David can reject the null.** You do not need to perform the rest of the hypothesis test. (5p)

David collects more data. After estimating the same model to the new data, he creates some plots. In Figure 10, you can see the residuals plotted against the predicted returns and the squared residuals plotted against the percentage returns.

e. **Interpret the graphs. Briefly explain why this is a sign of a problem with the model.** (5p)

Figure 10.



Note: Crypto currencies happened to show extremely high returns during this period. Past returns are no guarantee of future returns. This problem is loosely inspired by the paper Olivier Kraaijeveld, Johannes De Smedt, *The predictive power of public Twitter sentiment for forecasting cryptocurrency prices*, Journal of International Financial Markets, Institutions and Money, Volume 65,2020,101188,ISSN 1042-4431, https://doi.org/10.1016/j.intfin.2020.101188.
However, the bullishness data for this problem is made up. Reading the paper will not help you to solve the problem.

Problem 27

David, a computer scientist, wants to predict the price of the cryptocurrency Bitcoin Cash.

He writes a programming script that collects and categorizes tweets from Twitter about Bitcoin Cash, every day for 11 days; he uses this data to calculate a value called *Bullishness* (x). High Bullishness values (more than 0.7) means that there were more twitter recommendations to buy Bitcoin Cash than there were recommendations to sell, and vice versa. David believes that the Bullishness value can help him predict the following day's returns (one-day percentage change in the price of the cryptocurrency).

You can find 11 days of data below:

Return (y)	6	2	1	-4	-2	1	0	-3	6	3	1
Bullishness, previous day (x)	1.2	0.6	1.3	0.5	1.5	1.8	0.5	0.2	1.5	2.7	0.3

David decides on the following model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where *Y* is the daily percentage return of Bitcoin Cash and *X* is the Bullishness from the previous day. Remember: you need to show your work!

a. **Estimate the sample variance of *X* 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)

To make the following calculations easier, you are given the residual variance: $s_e^2 = 10.4$.

c. **Calculate a 95% prediction interval for tomorrow's daily return of Bitcoin Cash, given that the Bullishness was (x) was 2.0 today. Interpret the result.** (5p)

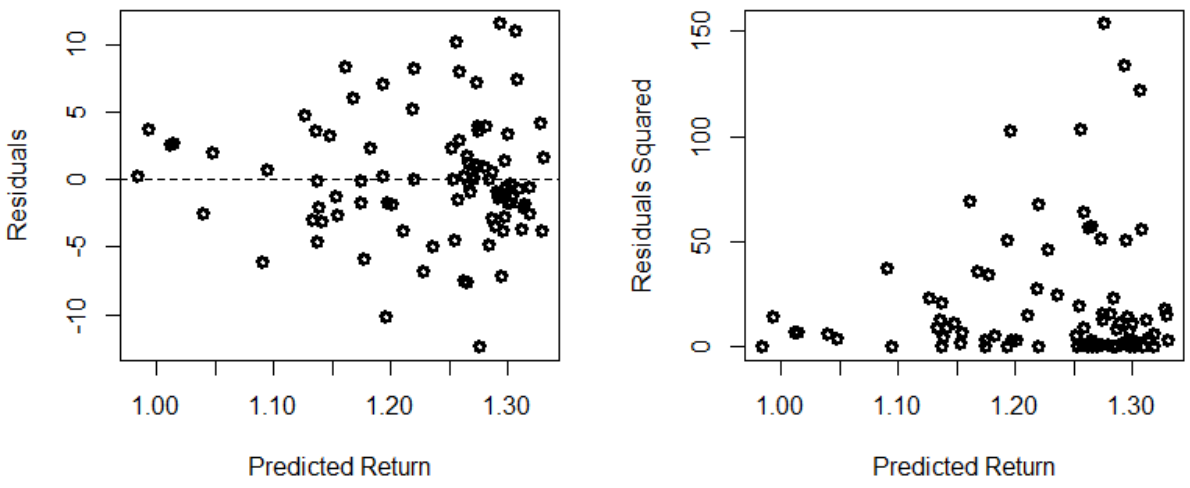
David wants to test whether the slope is significantly different from zero, at the 5% level of significance.

d. **Calculate the value of the test variable. Use this value to find and approximate p-value of the test. Use the p-value to determine if David can reject the null.** You do not need to perform the rest of the hypothesis test. (5p)

David collects more data. After estimating the same model to the new data, he creates some plots. In Figure 11, you can see the residuals plotted against the predicted returns and the squared residuals plotted against the percentage returns.

e. **Interpret the graphs. Briefly explain why this is a sign of a problem with the model.** (5p)

Figure 11.



Note: Crypto currencies happened to show extremely high returns during this period. Past returns are no guarantee of future returns. This problem is loosely inspired by the paper Olivier Kraaijeveld, Johannes De Smedt, *The predictive power of public Twitter sentiment for forecasting cryptocurrency prices*, Journal of International Financial Markets, Institutions and Money, Volume 65,2020,101188,ISSN 1042-4431, https://doi.org/10.1016/j.intfin.2020.101188.
However, the bullishness data for this problem is made up. Reading the paper will not help you to solve the problem.

Problem 28

David, a computer scientist, wants to predict the price of the cryptocurrency Litecoin.

He writes a programming script that collects and categorizes tweets from Twitter about Litecoin, every day for 11 days; he uses this data to calculate a value called *Bullishness* (x). High Bullishness values (more than 0.7) means that there were more twitter recommendations to buy Litecoin than there were recommendations to sell, and vice versa. David believes that the Bullishness value can help him predict the following day's returns (one-day percentage change in the price of the cryptocurrency).

You can find 11 days of data below:

Daily Return (y)	-2	0	-1	-4	-1	1	1	-2	13	2	2
Bullishness, previous day (x)	0.6	1.8	1.2	1.6	1.1	1.3	1.7	1.7	2.6	0.7	2.2

David decides on the following model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where *Y* is the daily percentage return of Bitcoin and *X* is the Bullishness from the previous day. Remember: you need to show your work!

a. **Estimate the sample variance of *X* 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)

To make the following calculations easier, you are given the residual variance: $s_e^2 = 14.3$.

c. **Calculate a 95% prediction interval for tomorrow's daily return of Litecoin, given that the Bullishness was (x) was 0.5 today. Interpret the result.** (5p)

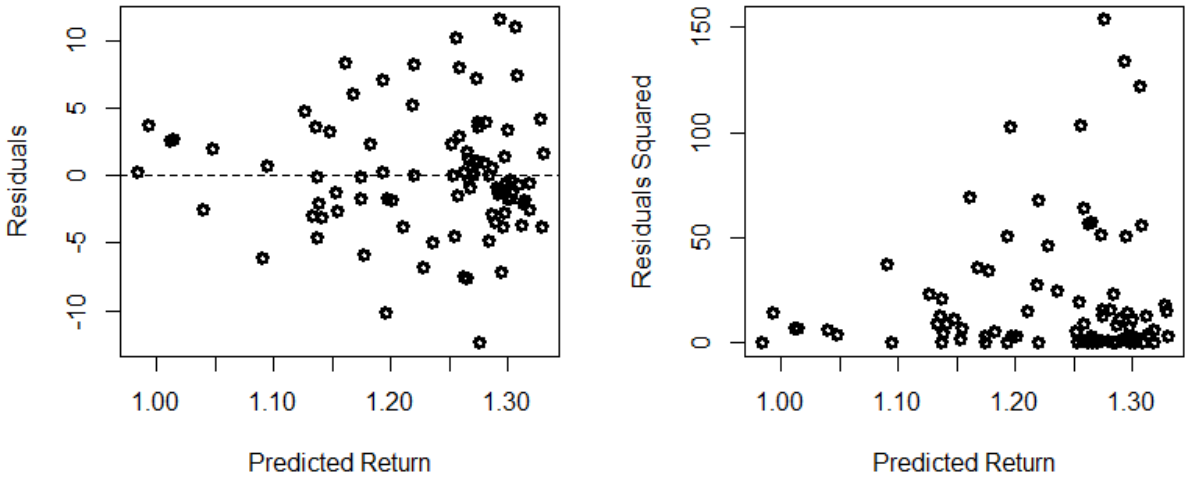
David wants to test whether the slope is significantly different from zero, at the 5% level of significance.

d. **Calculate the value of the test variable. Use this value to find the p-value of the test. Use the p-value to determine if David can reject the null.** You do not need to perform the rest of the hypothesis test. (5p)

David collects more data. After estimating the same model to the new data, he creates some plots. In Figure 12, you can see the residuals plotted against the predicted returns and the squared residuals plotted against the percentage returns.

e. **Interpret the graphs. Briefly explain why this is a sign of a problem with the model.** (5p)

Figure 12.



Note: Crypto currencies happened to show extremely high returns during this period. Past returns are no guarantee of future returns. This problem is loosely inspired by the paper Olivier Kraaijeveld, Johannes De Smedt, *The predictive power of public Twitter sentiment for forecasting cryptocurrency prices*, Journal of International Financial Markets, Institutions and Money, Volume 65,2020,101188,ISSN 1042-4431, https://doi.org/10.1016/j.intfin.2020.101188.
However, the bullishness data for this problem is made up. Reading the paper will not help you to solve the problem.