STATISTICAL INFERENCE (MS-C1620). 8.1.–17.4.2025. Aalto University. University Lecturer Pekka Pere.

Exam 17.4.2024

You may use a scientific calculator, i.e. a calculator that has operations for trigonometry, power, exponential function, logarithm, and binomial coefficients. No other type of calculator is allowed. In particular, you may not use programmable calculators which can run program code, symbolic calculators which can manipulate symbolic expressions, graphical calculators which are able to plot functions, or calculators with Internet connection. You may use your handwritten double-sided A4 note at the exam. **Present the intermediate steps of your calculations, and justify your answers in detail. A correct answer alone is worth zero points.** Some quantiles of the Normal distribution are at the end of the exam sheets. Check your answers. Next double-check them. Best of success to the exam!

1. Item a) is 4.5 points. Item b) is 1.5 points.

a) In an animation a ball with eyes(!) pushed another ball, a green cube watched(!), entered, and stopped the aggression of the ball with eyes. In another animation a red cube looked at only or did not stop the aggression. Six-month-old infants ($n = 20$) watched the animations. Next the infants were let to choose a green or red cube. 17 out of 20 infants chose the green cube.

Let us assume that six-month-olds choose a green or a red cube with probability π. An experiment is organised in which 20 six-month-olds are let to choose between a green and red cube independently.[1]

- Calculate the 95 % Wald confidence interval for the proportion of green cubes chosen in experiments of the kind just described.
- Calculate a plus four confidence interval for the proportion of green cubes chosen in experiments of the kind just described.
- Compare the intervals. Which is the more meaningful one and why?

b) Newspapers have given accounts of nonopening parachutes:[2]

- Helsingin Sanomat 19.8.2022 tells how Mariette Hägglund jumped from an aeroplane at a height of about four kilometers but her parachute did not open! Fortunately she had a backup parachute which opened. Hägglund estimated that a parachute does not open on every 500th jump.
- The Guardian 8.3.2024: "Five killed and 10 injured in Gaza aid airdrop when parachute fails to open – – they were hit by a pallet of aid parachuted into the territory as part of a humanitarian airdrop."

A factory manufactures 300 parachutes with a new method. All of them are tested if they open or not. All open. Calculate a 95% confidence interval for parachutes manufactured with the new method which fail to open.

---

[1] The question is based on the article Y. Kanakogi, Y. Inoue, G. Matsuda, D. Butler, K. Hiraki, and M. Myowa–Yamakoshi (2017): Preverbal Infants Affirm Third-Party Interventions That Protect Victims from Aggressors. *Nature. Human Behaviour.* doi:10.1038/s41562-016-0037.

[2] `https://www.hs.fi/kotimaa/art-2000009012531.html` and `https://www.theguardian.com/world/2024/mar/08/gaza-aid-airdrop-kills-civilians-when-parachute-fails-to-open-witness-says` (retrieved 15.4.2024).

2. The histograms in the figure display how 1 244 persons (Finns) evaluate their own attractiveness on scale 0–100 (the larger the figure the more attractive the person), how outsiders evaluate attractiveness of the same person, and the difference of the evaluations.[3] Outsiders appear to evaluate attractiveness of a person more favourably than the persons themselves do: The sample mean of attractiveness is 53.68167 or 58.65675 when evaluated by the person him-/herself or by an outsider, respectively. The sample mean of the difference of the evaluations is -4.97508. Sample standard deviations are 401.0748 (own evaluation), 291.1234 (evaluation by an outsider), and 221.5834 (difference of the evaluations).

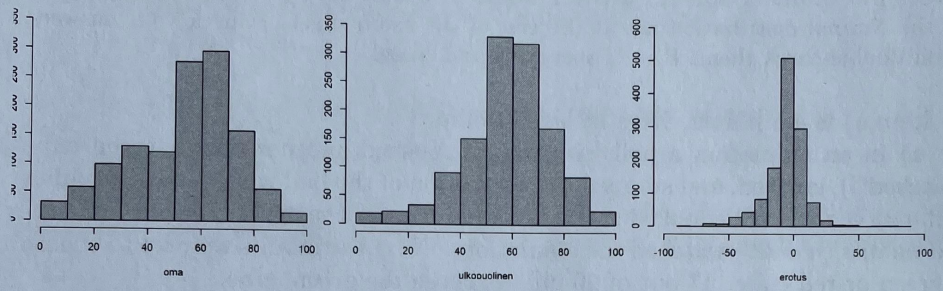


Figure 1: Attractiveness evaluation by test subjects and by outsiders, and the difference of the evaluations.

Let us assume that the evaluations and the difference of them follow Normal distributions. Let us investigate if the difference between the evaluations is statistically significant.

a) Explain carefully the statistic which is fit to assess if the difference between the evaluations is statistically significant. Calculate the statistic.

b) What distribution does the statistic follow exactly?

c) Test the null hypothesis of no difference between the evaluations at significance level 0.001. (You can use an approximate critical value.)

d) Is the assumption of normality of the evaluations and their difference necessary for the test you have conducted? Explain carefully.

[3] J. Antfolk and A. Sjölund (2018): High Parental Investment in Childhood is Associated with Increased Mate Value in Adulthood. *Personality and Individual Differences*, 127, 144–150.

3. Items a–d): 6 points. Item e): An extra 1.5 points. (It is possible to get full points 24/24 from the exam even if you answered wrongly at item e).)

Answer A, B, C, D, or E to the following questions.[4] The data is composed of the four observations (the small circles) in the figure. One of the observations is located in the origin. "The regression line" refers below to a simple linear regression with an intercept unless stated otherwise. The coefficient of determination is calculated with formula $1 - \mathsf{SSE}/\mathsf{SST} = 1 - \mathsf{RSS}/\mathsf{SYY}$ (in the notation of the lecture notes or the theoretical exercises, respectively).

a) Which of the lines describes best the average value of Y?

A. Ya B. Yb C. Yc D. Yd E. Ye

b) Which of the lines describes best the regression line?

A. Ya B. Yb C. Yc D. Yd E. Ye

c) The coefficient of determination for the regression of item b) is

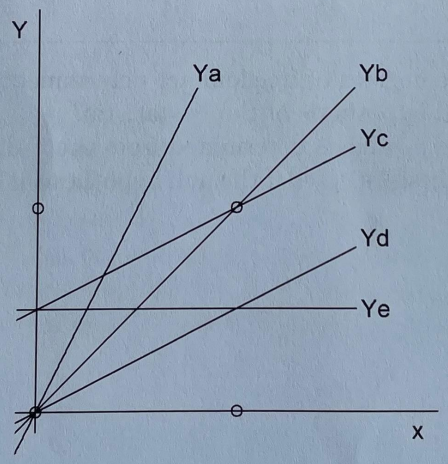
A. < 0 B. 0 C. between 0 and 0.5 D. 0.5 E. > 0.5

d) Which of the lines describes best the regression line when there is no intercept in the model (the regression line is forced to go through the origin)?

A. Ya B. Yb C. Yc D. Yd E. Ye

e) The coefficient of determination for the regression of item d) is

A. < 0 B. 0 C. välillä $(0, 0.5)$ D. 0.5 E. > 0.5



[4]The question is from the article W. Becker and P. Kennedy (1992): A Lesson in Least Squares and R Squared. *American Statistician*, 46, 282–283.

4. First impressions count! Dallas *et al.* (2018, 2020) tested if even word order matters.[5] In the USA chain restaurants with twenty or more locations have had to provide calorie information on their menus since 2018. The intention is to encourage consumers to choose food with less calories. Research suggests that the mandate has had little effect. Dallas *et al.* (*op. cit.*) think that the reason is that calorie labels are typically displayed to the right of food items on menus. They hypothesize that if calorie information was moved from the right to the left of food items, so that the calorie information was processed earlier, consumers would order lower calorie foods. The researchers conducted an experiment to test the hypothesis.

Students in a student canteen were asked to participate in the experiment. A menu was given by random assignment to those who agreed to participate. There were three kinds of menus: the menu either had calorie information to the left of (before) each food item, had calorie information to the right of (after) each food item, or had no calorie information.

The means and standard deviations of the calories of the food choices the three groups made are in the table. The researchers carried out a one-way ANOVA. The F-statistic was 3.60, its p-value was 0.03, and the degrees of freedoms of the null distribution (F) of the statistic were 2 and 146.

	calories (kcal)		
	to the left	to the right	not reported
mean	654.53	865.41	914.34
standard deviation	390.45	517.26	560.94
sample size	45	54	50

a) Explain how the degrees of freedom are determined.

b) What is the null hypothesis of the F-statistic?

c) Explain how the p-value is determined from the F-distribution.

c) Is the null hypothesis rejected if the null hypothesis is tested at significance level 0.05?

A hint to any of the questions: The 0.925th, 0.95th, 0.975th, 0.99th, 0.995th, and 0.9995th quantiles of the Standard Normal distribution are 1.440, 1.645, 1.960, 2.326, 2.576, and 3.29, respectively.

[5] Dallas, S.K., P.J. Liu, and P.A. Ubel (2018): Don't Count Calorie Labeling Out: Calorie Counts on the Left Side Items Lead to Lower Calorie Food Choices. *Journal of Consumer Psychology*, 29, 60–69. S.K. Dallas *et al.* (2020): Corrigendum. *Journal of Consumer Psychology*, 30, 571.