Statistical inference (MS-C1620). 8.1.–17.4.2024. Aalto University. University Lecturer Pekka Pere.

2nd re-examination 27.8.2024

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 have a personally handwritten double-sided A4 at the exam. **Present the intermediate steps of all your calculations, and justify all your answers in detail. A correct answer alone is worth zero points. All assignments are worth six points.** Check your answers. Next double-check them. Best of success to the exam! Note1: Some quantiles of the Standard normal distribution and formulae are at the end of the exam sheet. Note2: You can have a handwritten two-sided A4 memory sheet with you at the exam. You do not have to return it with your answers.

1. Telecommunication corporation DNA conducts annually a school survey about phone use and purchases. The survey 2.–13.2.2023 was answered by 704 parents of children between 5 and 12 years of age.[1] For the age group 10–12 years all or 325 parents answered that his or her child owns a phone (smartphone, ordinary cell phone, or wrist phone).

Let us assume that the responses are a random sample. Calculate an approximate 95% confidence interval for the proportion of children between 10–12 years of age who do not own a phone. Use a meaningful way of calculating the interval.

2. John Arbuthnot (1667–1735) observed that during the years 1629–1710 more boys than girls were born each year in London. He calculated a tiny probability for the event that more boys are born than girls for 82 years if both boys and girls are born with probability 1/2. He reasoned that the probability of a newborn to be a boy must be greater than 1/2. Arbuthnot regarded it as evidence of God's good will: more men die prematurely than women, so there will be one of opposite sex for everybody (figure). Arbuthnot's calculation and reasoning are regarded as the first statistical test ever.[2]

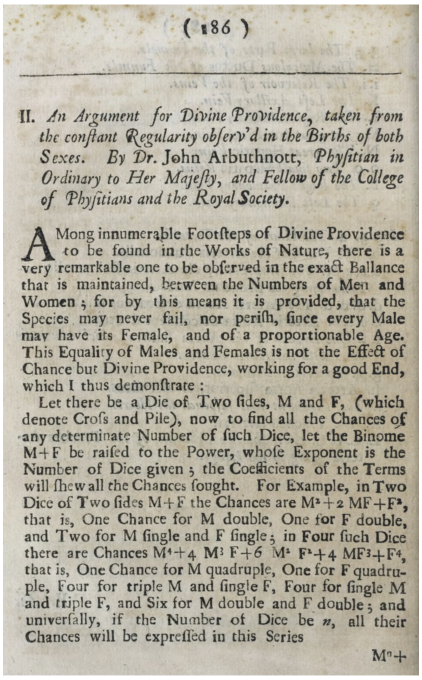
The histogram depicts the distribution of the ratio of number of newborn boys to girls in London 1629–1710. The ratio would be 1 if the same number of boys and girls were born during a year. The ratio was larger than 1 for all years in the data.

a) Formulate the sign test statistic and calculate the value of it for the null hypothesis that the probability of a newborn boy or girl is the same 1/2. Can you reject the null hypothesis at significance level 0.002 (two-sided test)?

[1] https://www.dna.fi/documents/94506/11594975/DNA_Koululaistutkimus_2023_medialle_final_pdf.pdf (read 5.6.2024). Nepa conducted the poll by drawing a random sample from an online panel. I thank DNA Oyj and Nepa Insight Oy for the detailed information about the poll (personal notification 6.–7.6.2024).

[2] J. Arbuthnot (1710): II. An Argument for Divine Providence, Taken from the Constant Regularity Observ'd in the Births of Both Sexes. By Dr. John Arbuthnott, Physitian in Ordinary to Her Majesty, and Fellow of the College of Physitians and the Royal Society. *Philosophical Transactions of the Royal Society*, 27, 186–190. The ratio figures are taken from the HistData package.

[3] Figure: Royal Society (https://royalsocietypublishing.org/doi/10.1098/rstl.1710.0011). Editing of the figure: Pekka Pere (2024).



(186)

II. *An Argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes. By Dr.* John Arbuthnott, *Physitian in Ordinary to Her Majesty, and Fellow of the College of Physitians and the Royal Society.*

AMong innumerable Footsteps of Divine Providence to be found in the Works of Nature, there is a very remarkable one to be observed in the exact Ballance that is maintained, between the Numbers of Men and Women; for by this means it is provided, that the Species may never fail, nor perish, since every Male may have its Female, and of a proportionable Age. This Equality of Males and Females is not the Effect of Chance but Divine Providence, working for a good End, which I thus demonstrate:

Let there be a Die of Two sides, M and F, (which denote Cross and Pile), now to find all the Chances of any determinate Number of such Dice, let the Binome M+F be raised to the Power, whose Exponent is the Number of Dice given; the Coefficients of the Terms will shew all the Chances sought. For Example, in Two Dice of Two sides M+F the Chances are $M^2+2\,MF+F^2$, that is, One Chance for M double, One for F double, and Two for M single and F single; in Four such Dice there are Chances $M^4+4\,M^3F+6\,M^2F^2+4\,MF^3+F^4$, that is, One Chance for M quadruple, One for F quadruple, Four for triple M and single F, Four for single M and triple F, and Six for M double and F double; and universally, if the Number of Dice be *n*, all their Chances will be expressed in this Series

M^n+

Figure 1: The 1st page of Arbuthnot's article (1710).[3]

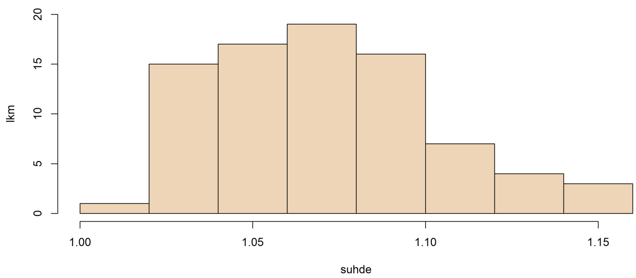


Figure 2: The distribution of the ratio of number of born boys to the number of born girls in London 1629–1710 (suhde = ratio, lkm = frequency).

b) The ratio of number of born boys to the number of born girls is saved in dataframe `Arbuthnot` under name `Ratio`. Command `shapiro.test(Arbuthnot $Ratio)` returns ("p-value" refers to a two-sided test):

```
##          Shapiro-Wilk normality test
## data:  Arbuthnot$Ratio
## W = 0.95956, p-value = 0.01117
```

What is the null hypothesis associated with the R output above? Should the null hypothesis be rejected at significance level 0.01 if the test were conducted at significance level 0.01 (two-sided test)? Explain.

3. Fire resistance of cross-laminated timber panels (figure) was studied with experiments. The panels were composed of lumber boards glued together in perpendicular layers. Charring rates were measured for the bottom 1st and 2nd layers when fire was directed to the panel from below. The number of such charring rate experiments was 53 for both layers.[4] Let us assume that the data are composed of two independent samples of charring rate measurements from the two layers and that the measurements are Normally distributed ($\mathsf{N}(\mu_1, \sigma_1^2)$ and $\mathsf{N}(\mu_2, \sigma_2^2)$) and independent also within the samples.

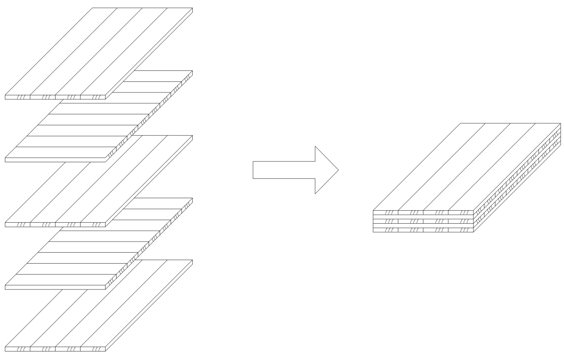


Figure 3: Structure of a cross-laminated timber panel.

a) Sample variances of the charring rates are $s_1^2 = 0.002843024$ (1st layer) and $s_2^2 = 0.1669577$ (2nd layer). Present a test statistic and calculate its value for the null hypothesis that the variances are the same ($\sigma_1^2 = \sigma_2^2$). What distribution (perhaps approximately) does the statistic follow under the null hypothesis? What are the degrees of freedom of the distribution if such a concept is associated with the distribution? Explain concisely how could the p-value of the test statistic be calculated from the left or right tail of the distribution (one-sided test with alternative hypothesis $\sigma_1^2 < \sigma_2^2$).

[4]Mika Alanen, Mikko Malaska, Mikko Salminen, Pyry Paavola, and Sami Pajunen (2024): Experimental Determination of the Charring Rate of Cross-Laminated Timber Panels (manuscript). I thank Mika Alanen and Mikko Malaska for the data and explanation of the experiments, and Mika Alanen for the figure.

b) Sample means of the charring rates are $\hat{\mu}_1 = 0.6818273$ (1st layer) and $\hat{\mu}_2 = 1.116031$ (2nd layer). Present a test statistic and calculate its value for the null hypothesis that the means are the same ($\mu_1 = \mu_2$). What distribution (perhaps approximately) does the statistic follow under the null hypothesis? What are the degrees of freedom of the distribution if such a concept is associated with the distribution? Should the null hypothesis be rejected at significance level 0.001 (two-sided test)?

4. Let us continue analyses of the charring rate experiments. The charring rate data in the table is categorised according to the positioning of sensors (distance from gab between two lumber boards) for measuring the charring rates and specimen orientation (horizontal or vertical). The charring rates for the 1st and 2nd layers are in the upper and lower part of the table, respectively. The data are categorised further by distance of sensor from gab between two lumber boards (small, at most 6 measurements, moderate, at most 2 measurements, or large, at most 6 measurements) and whether the specimen were in vertical (P1, P2, or P3) or horizontal (V1, V2, or V3) position. Each figure in a cell refers to a particular positioning of sensor and specimen. An outlier (7.20 in red) in the 2nd layer and its counterpart (0.65) in the 1st layer have been removed from the data before the statistical analyses. After the removals the data consists of 53 measurements in both layers.

charring rates in the 1st layer (mm/min)

distance of sensor from vertical gap between two lumber boards

	small						moderate		large					
P1	0.63	0.80			0.73	0.67	0.66		0.63	0.60			0.65	0.61
P2	0.65	0.70			0.66	0.67			0.75	0.72	0.61	0.67	0.73	0.66
P3	0.65	0.64	0.62	0.64	0.63	0.59			0.72	0.62			0.63	0.69
V1	0.67	0.67			0.77	0.68	0.68	0.77	0.65	0.66			0.66	0.68
V2	0.72	0.78			0.70	0.72	0.70	0.67	0.67	0.67			0.63	0.75
V3	0.71				0.82		0.77		0.69				0.70	

charring rates in the 2nd layer (mm/min)

distance of sensor from vertical gap between two lumber boards

	small						moderate		large					
P1	2.67	0.67			0.95	0.75	1.71		1.67	1.07			2.06	0.94
P2	7.20	1.47			1.47	0.87			0.73	1.01	1.53	1.41	2.32	1.26
P3	0.80	0.73	0.97	0.85	0.73	0.69			0.58	0.90			1.13	0.91
V1	0.77	1.54			0.71	1.24	1.12	1.11	1.27	1.17			1.17	1.15
V2	0.77	1.16			1.14	1.02	1.09	1.26	0.96	1.05			1.20	1.12
V3	0.84				0.88		0.76		0.85				0.92	

The measurements in the two layers are paired: Charring rate measurements have been done under exactly the same circumstances in corresponding cells in the two layers (e.g. the first observations of the P1 rows 0.63 and 2.67 in groups “small” are a pair). The experimental design suggests that the measurements from the 1st and 2nd layers are pairwise correlated. Let us thus relax the previous assumption that the measurements of the 1st and 2nd layers are independent and assume instead that the measurements are pairwise independent. Let us also assume that the pairwise differences ($0.63 - 2.67 = -2.04$, $0.80 - 0.67 = -2.04$, $0.73 - 0.95 = -0.22$, $\ldots$) are Normally distributed ($\mathsf{N}(\mu_D, \sigma_D^2)$). The sample size is 53.

The sensor distances or specimen orientation are not explicitly taken account of in the assignment. Pairwise differencing is considered to eliminate such effects from the data. Let us challenge the null hypothesis of equal charring rates in the two layers with a pairwise t test.

a) In a previous assignment it was found that the variance of charring rate is much larger in the 2nd layer than in the 1st layer. Is the large difference between the variances a barrier for applying the pairwise t test, or does the pairwise t test presume identical variances in the groups to be compared? (2 p)

b) Sample mean and variance of the pairwise differences are -0.434204 and 0.1789338, respectively. Present the pairwise t test statistic and calculate its value for the null hypothesis of equal charring rates in the two layers. What distribution (perhaps approximately) does the statistic follow under the null hypothesis? What are the degrees of freedom of the distribution if such a concept is associated with the distribution? Should the null hypothesis be rejected at significance level 0.001 (two-sided test)? (4 p)

A few quantiles and indicative formulae

The 0.925th, 0.95th, 0.975th, 0.99th, 0.9925th, 0.995th, 0.999th, and 0.9995th quantiles of the Standard Normal distribution are 1.440, 1.645, 1.960, 2.326, 2.432, 2.576, 3.090, and 3.291, respectively.

- $\mathsf{P}(A \cup B) = \mathsf{P}(A) + P(B) - \mathsf{P}(A \cap B)$, $\quad \mathsf{P}(A \cap B) = \mathsf{P}(A|B)P(B)$, $\mathsf{P}(A|B) = \mathsf{P}(A \cap B)/\mathsf{P}(B)$, $\quad \mathsf{P}(A^C) = 1 - \mathsf{P}(A)$
- Under independence $\mathsf{P}(A\ |B) = \mathsf{P}(A)$ and $\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$
- $\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(A\ |B_i)P(B_i)$, $\quad \mathsf{P}(B_i|A) = \dfrac{\mathsf{P}(A|B_i)P(B_i)}{\sum_{i=1}^{n} \mathsf{P}(A|B_i)P(B_i)}$
- $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$, $\quad \dbinom{n}{0} = \dbinom{n}{n} = 1$
- $\mathsf{P}(Y = y) = \dbinom{n}{y} \pi^y (1-\pi)^{n-y}$, $\quad \mathsf{E}(Y) = n\pi$, $\quad \mathsf{V}(Y) = n\pi(1-\pi)$
- $\mathsf{P}(Y = y) = \dfrac{\dbinom{l}{y}\dbinom{m}{n-y}}{\dbinom{l+m}{n}}$, $\quad \mathsf{E}(Y) = n\pi$, $\quad \mathsf{V}(Y) = n\pi(1-\pi) \times \dfrac{l+m-n}{l+m-1}$
- $\mathsf{P}(Y = y) = e^{-\mu}\dfrac{\mu^y}{y!}$, $\quad \mathsf{E}(Y) = \mathsf{V}(Y) = \mu$
- $\dfrac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \overset{n \text{ large}}{\sim} \mathsf{N}(0, 1)$ eli $\hat{\mu} \overset{n \text{ large}}{\sim} \mathsf{N}(\mu, \sigma^2/n)$, $\quad \hat{\mu} = \bar{x}$
- $\hat{\pi} \pm \mathsf{z}_{1-\alpha/2} \sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n}}$
- Plus four confidence interval:

yes	no	Σ
$y + 2$	$n - y + 2$	$n + 4$

- Rule of three: $\left[0,\ \dfrac{3}{n}\right] \approx \left[0,\ \dfrac{3}{n+1}\right]$
- $\hat{\pi}_1 - \hat{\pi}_2 \pm \mathsf{z}_{1-\alpha/2} \sqrt{\dfrac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \dfrac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$
- Agresti–Caffo confidence interval:

	yes	no	Σ
group 1	$n_{11} + 1$	$n_{12} + 1$	$n_1 + 2$
group 2	$n_{21} + 1$	$n_{22} + 1$	$n_2 + 2$

- $\hat{\pi}_{1+} - \hat{\pi}_{+1} \pm \mathsf{z}_{1-\alpha/2} \dfrac{\sqrt{n_{12} + n_{21} - (n_{12} - n_{21})^2/n}}{n}$, $\quad \hat{\pi}_{1+} - \hat{\pi}_{+1} = \dfrac{n_{12} - n_{21}}{n}$
- $\hat{\mu} \pm \mathsf{t}_{1-\alpha/2}(n-1) \times \sqrt{\dfrac{s^2}{n}}$
- $\hat{\mu}_1 - \hat{\mu}_2 \pm \mathsf{t}_{1-\alpha/2}(n_1 + n_2 - 2) s \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$

- $\hat{\mu}_1 - \hat{\mu}_2 \pm \mathsf{t}_{1-\alpha/2}(\nu)\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad \nu = \mathsf{int}\left[\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}\right]$
- $z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$
- $z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})}}, \quad \hat{\pi} = \frac{n_1\hat{\pi}_1 + n_2\hat{\pi}_2}{n_1 + n_2} = \frac{y_1 + y_2}{n_1 + n_2}$
- $z = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$
- $X^2 = \sum_{i=1}^{c} \frac{(N_i - e_i)^2}{e_i} \overset{n \text{ large}}{\sim} \chi^2(c - 1 - s)$
- $X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(N_{ij} - e_{ij})^2}{e_{ij}} \overset{n \text{ large}}{\sim} \chi^2((I-1)(J-1))$
- $t = \frac{\hat{\mu} - \mu_0}{s/\sqrt{n}} \sim \mathsf{t}(n-1)$
- $t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathsf{t}(n_1 + n_2 - 2), \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$
- $t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim \mathsf{t}(\nu), \quad \nu = \mathsf{int}\left[\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}\right].$
- $\frac{\hat{\mu}_D - \mu_{D0}}{s_D/\sqrt{n}} \sim \mathsf{t}(n-1)$
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$
- $F = \frac{s_1^2}{s_2^2} \sim \mathsf{F}(n_1 - 1, n_2 - 1)$