Data analysis (PHY231) — HS 2020

Test November 3, 2020

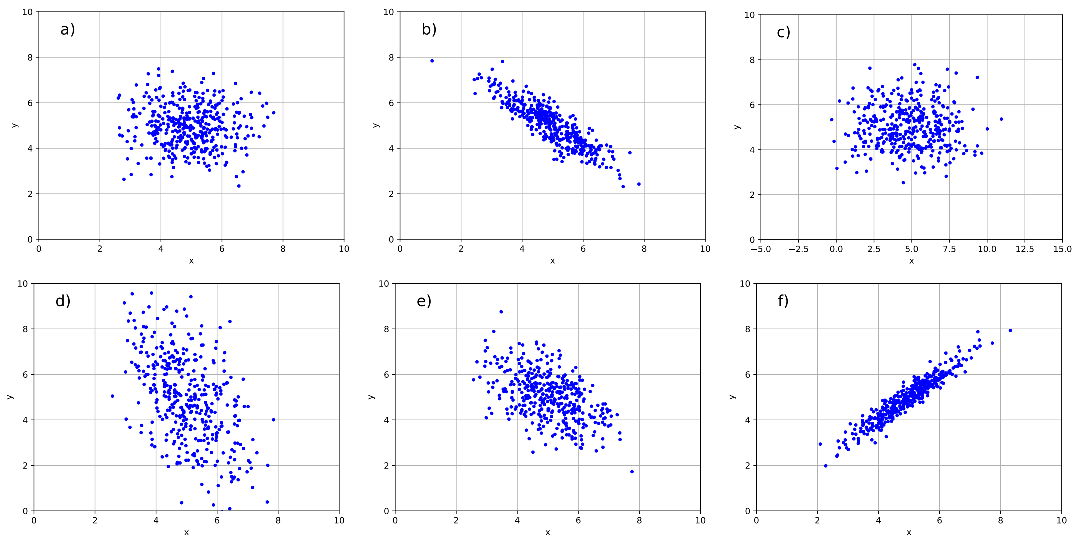
Time:	120 minutes
Tools:	Pen, paper & calculator
Number of pages:	3 (pages 2,3 and 4)
Number of problems:	5
Max. number of points:	21

Instructions:

- The solutions should be written down on paper in blue or black pen.
- Write your full name and your Matrikelnummer on all your paper sheets.
- Keep your hands in vision of your webcam.
- Only one hand is allowed on your laptop at one time.
- You should be prepared to share your screen, show your room or test the audio if the assistant asks.
- If you have a question, please use the raise hand feature. Do not use the chat function.
- A formula sheet is at the bottom of the exam.
- After the end of the exam, please scan/take a photo using your smartphone/laptop and email da@physik.uzh.ch using your university email address.
- **Stay connected until everyone has sent their solutions**.

Problem 1 *(4 points)*

Below you see data which represent measurements of x and y as a scatter plot.



The following scenarios denote the variance (V(x) and V(y)) and correlation coefficient ρ(x,y).

Scenarios	
1: V(x) = 1, V(y) = 1, ρ(x,y) = -0.90	**2:** V(x) = 1, V(y) = 4, ρ(x,y) = -0.5
3: V(x) = 1, V(y) = 1, ρ(x,y) = 0.0	**4:** V(x) = 4, V(y) = 1, ρ(x,y) = 0.0
5: V(x) = 1, V(y) = 1, ρ(x,y) = -0.5	**6:** V(x) = 1, V(y) = 1, ρ(x,y) = 0.95

(a) Match the scenarios with the scatter plots above by marking the correct letter from the plot with the right scenario number. [3 points]

(b) Which scenario would give the lowest uncertainty on the ratio $z = y/x$? [1 point]

Problem 2 *(5 points)*

You have obtained the following set of measurements for the acceleration due to gravity on the earth, g.

1) $g = (9.5228 \pm 0.5576)$ m/s^2
2) $g = (9.6432 \pm 0.2334)$ m/s^2
3) $g = (9.7089 \pm 0.0823)$ m/s^2

(a) Round the measurements to the appropriate number of significant figures.

(b) Will the weighted arithmetic mean (shown as 'weighted mean' in the lectures) be lower, higher or the same as the simple mean?

(c) How would the **simple** mean qualitatively change if there was a common systematic uncertainty for all measurements?

(d) How would the **weighted** mean qualitatively change if there was a common systematic uncertainty for all measurements?

(e) Do you think measurement number 3) is in statistical agreement with the known value of $g = 9.81\,\text{m/s}^2$? Explain your reasoning.

Problem 3 *(4 points)*

You measure the length of two desks in order to comply with social distancing guidelines. The measurements of the desks are $l_1 = 1.20 \pm 0.03$ m and $l_2 = 1.18 \pm 0.04$ m. Assume that the two uncertainties are independent.

(a) Calculate the value and uncertainty on the difference in lengths of the two desks $d = l_1 - l_2$.

(b) Calculate the value and uncertainty on the sum of the lengths of the two desks $s = l_1 + l_2$.

(c) If there was an additive systematic uncertainty common to both measurements, how would that qualitatively change the uncertainties for the difference in part a) and the sum in part b)?

Problem 4 *(5 points)*

You setup an experiment with an electron source and detector for 16 days. The electron source produces electrons and the detector detects them. On average, 100 electrons per day are detected.

(a) What is the statistical uncertainty on this average?

(b) Assuming the average detection rate is constant, what is the statistical uncertainty on the number of detections for the 17th day?

(c) Assuming the average detection rate is constant, what is the probability that there are more than 120 detections on the 17th day?

(d) Recalculate your answer to part b), assuming that you know that exactly 150 electrons are produced from the source.

Problem 5 *(3 points)*

Imagine that on average, one in every 10 people in Switzerland is infected with COVID-19. The COVID-19 test has a 100% probability to be negative if you are not infected and a 30% probability to be negative if you are infected (false-negative rate). *These numbers are completely made up just for the sake of this question.*

(a) What is the prior probability that a random person is infected with COVID-19?

(b) Use Bayes' theorem to calculate the posterior probability that a random person is infected, given a negative test.

(c) If the person also has symptoms (e.g. coughing), how do the prior and posterior probabilities qualitatively change? Explain your reasoning.

Formula sheet

Bayes' theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Binomial distribution: $P(r|p,n) = \frac{n!}{r!(n-r)!}p^r(1-p)^{n-r}$, $\bar{r} = n \cdot p$, $V(r) = n \cdot p \cdot (1-p)$

Poisson distribution: $P(r|\lambda) = \frac{e^{-\lambda}\lambda^r}{r!}$, $\bar{r} = \lambda$, $V(r) = \lambda$.

Weighted arithmetic mean: $\bar{x} = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$, $\sigma(\bar{x}) = \frac{1}{\sqrt{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}}$.

Variance: $V_x = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$

Covariance: $\text{cov}(x,y) = \frac{1}{N} \cdot \sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\bar{y}$

Correlation coefficient: $\rho(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$

Additive systematic correlation: $\text{cov}(x,y) = \sigma_{syst}^2$

Multiplicative systematic correlation: $\text{cov}(x,y) = \sigma_f^2 \cdot x \cdot y$

Error propagation with one variable: $\sigma(f) = \frac{\partial f(x)}{\partial x}|_{x=\bar{x}}\sigma(x)$

Error propagation with two variables: $\sigma(f) = \sqrt{|\frac{\partial f}{\partial x}|^2 \cdot \sigma_x^2 + |\frac{\partial f}{\partial y}|^2 \cdot \sigma_y^2 + 2\frac{\partial f}{\partial x}\frac{\partial f}{\partial y} \cdot cov(x,y)}$