# Module III review

## Statistics

Anton Afanasev

Higher School of Economics

DSBA 211

March 22, 2023

# Seminar Overview

# Quiz

The random variables $X_1$ and $X_2$ are each normally distributed with mean 1 and variance 1. Their correlation coefficient is 0.25. Find

$$\mathsf{P}\left(2X_1 > X_2\right).$$

# Quiz

# Problem 1

Let $X_1, X_2, \ldots, X_n$ be the random sample from a random variable $X$ with the distribution with p.d.f.:

$$f(x) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & x < \theta \end{cases}$$

1. Find $\widehat{\theta}_{ML}$.
2. Find the probability $P\left(\widehat{\theta}_{ML} > \mathsf{E}(X)\right)$
3. Find the median $m$ of the distribution of $\widehat{\theta}_{ML}$.

# Problem 1

# Problem 1

# Problem 1

# Problem 2

The data $(x_i, y_i)$, $i = 1, \ldots, m$ are generated by the model $y_i = \beta x_i + x_i \varepsilon_i, \quad \varepsilon_i \sim$ i.i.d. $\mathcal{N}\left(0, \sigma^2\right)$. Two students Meghen and Harry are going to estimate the parameter. Meghen uses the Least Square estimator for the equation $y_i = \beta x_i + u_i$, and have got the estimator $\widehat{\beta}_M$. Harry divide the model equation by $x_i$ and have got the estimator $\widehat{\beta}_H$ using the Least Square estimation for the equation $z_i = \dfrac{y_i}{x_i} = \beta + \varepsilon_i$.

1. Find $\widehat{\beta}_M$, $\widehat{\beta}_H$. Are these estimators unbiased?
2. Which estimator you would prefer? Why?

# Problem 2

# Problem 2

# Problem 2

# Problem 3

There are $m$ white and $n$ black balls in the box. Players take out one ball one after another and return it back to the box. A player wins if he takes out a white ball first. Find probabilities to win for each player (first, second). Is this game fair?

# Problem 3

# Problem 4

1000 father-son pairs were randomly selected. The survey showed that 510 out of 654 light-eyed fathers have light-eyed sons, and 213 out of 346 dark-eyed fathers have dark-eyed sons.

1. Construct 95% confidence intervals
   - for the proportion of fathers that have a son with the same eye colour,
   - for the proportion of light-eyed fathers that have a light-eyed son,
   - for the proportion of dark-eyed fathers that have a dark-eyed son.
2. Is it possible to say (at the 5% significance level) that proportion of fathers that have a son with the same eye colour is higher among dark-eyed fathers? Justify your answer.
3. Fill in the contingency table and test a hypothesis that eye colour of a father and eye colour of a son are statistically independent. Let the significance level be 5%.

# Problem 4

# Problem 4

# Problem 4

# Problem 5

Students A, B and C are discussing the solution of a Problem from their home assignment in "Statistics". The solution is given by one of their colleagues. Student A will say the solution is false if and only if either B or C say they have found a mistake in the solution. Neither of B and C have had time to go through the solution, but B will randomly with a 50% probability say that he has found a mistake. C will wait to hear what B says, but if B does not say he has found a mistake, then C will randomly with a 20% probability say that he has found a mistake in the solution. What is the probability that B said he found a mistake if A says the solution is false?

# Problem 5

# Problem 6

Four types of cars were tested for the gas consumption per 100 km. 12, 8, 7, and 9 cars were tested of types $A$, $B$, $C$, and $D$, respectively. Average gas consumption in these tests were 9.75 , 11.25, 9.00, and 11.0 liters per 100 km for the types $A$, $B$, $C$, and $D$, respectively. To test null hypothesis that all cars have equal gas consumption results were arranged in ANOVA table, but some entries in the table are missed.

| Source | Degrees of Freedom | Sum of Squares | Mean Square | $F$-value |
|---|---|---|---|---|
| Between | | | | |
| Within | | | 3.0241 | |
| Total | | | | |

1. Complete the table.
2. Is there significant difference in gas consumption between 4 types of cars? (use 5% significance level)
3. Construct simultaneous 95% confidence intervals for the difference of gas consumption between cars of types $A$ and $B$ and between types $A$ and $C$.

# Problem 6

# Problem 6

# Problem 6

# Problem 7

In the following tasks $\mathrm{CI}_{1-\alpha}$ and $\mathrm{PI}_{1-\alpha}$ are confidence and prediction intervals respectively with significance level $\alpha$.

1. $y_1 = 2\alpha + \varepsilon_1, y_2 = 4\alpha + \varepsilon_2$. Assume $\varepsilon_i \sim$ i.i.d. $\mathcal{N}(0,1)$. Derive OLS estimator $\widehat{\alpha}$. Find $\mathsf{E}\left(\widehat{\alpha}\right)$. Find $\mathsf{V}\left(\widehat{\alpha}\right)$.
   Let $y_3 = 3\alpha + \varepsilon_3$. Sample $y_1 = 4, y_2 = 6$. $y_3$ is unknown. Derive $\mathrm{CI}_{95\%}$ for $\mathsf{E}(y_3)$. Derive $\mathrm{PI}_{95\%}$ for $y_3$.
2. $y_1 = -\alpha + \varepsilon_1, y_2 = 2\alpha + \varepsilon_2$. Assume $\varepsilon_i \sim$ i.i.d. $\mathcal{N}(0,1)$. Derive OLS estimator $\widehat{\alpha}$. Find $\mathsf{E}\left(\widehat{\alpha}\right)$. Find $\mathsf{V}\left(\widehat{\alpha}\right)$.
   Let $y_3 = 5\alpha + \varepsilon_3$. Sample $y_1 = -1, y_2 = 1$. $y_3$ is unknown. Derive $\mathrm{CI}_{95\%}$ for $\mathsf{E}(y_3)$. Derive $\mathrm{PI}_{95\%}$ for $y_3$.
3. $y_1 = \alpha + \beta + \varepsilon_1, y_2 = -\alpha + \varepsilon_2, y_3 = \alpha + 2\beta + \varepsilon_3$. Assume $\varepsilon_i \sim$ i.i.d. $\mathcal{N}(0,1)$. Derive OLS estimators $\widehat{\alpha}, \widehat{\beta}$. Find $\mathsf{E}\left(\widehat{\alpha}\right), \mathsf{E}\left(\widehat{\beta}\right)$. Find $\mathsf{V}\left(\widehat{\alpha}\right), \mathsf{V}\left(\widehat{\beta}\right), \mathsf{Cov}\left(\widehat{\alpha}, \widehat{\beta}\right), \mathsf{V}\left(2\widehat{\alpha} + \widehat{\beta}\right)$.
   Let $y_4 = 2\alpha + \beta + \varepsilon_4$. Sample $y_1 = -1, y_2 = 1, y_3 = 3$. $y_4$ is unknown. Derive $\mathrm{CI}_{95\%}$ for $\mathsf{E}(y_4)$. Derive $\mathrm{PI}_{95\%}$ for $y_4$.

# Problem 7

# Problem 7

# Problem 7

# Problem 7

Look at the time!