1 Graded part of the homework

You should hand in the printed version of the homework 10 december 2015. Each late day will be penalized by 20%.

1. The law of distribution of a random variable X is given in the table

Value of X	0	1	-2
Probability	0.2	0.3	0.5

   (a) Find $\mathbb{P}(X > 0), \mathbb{P}(X^2 > 3)$

   (b) Find $\mathbb{E}(X), \mathbb{E}(X^2), \mathbb{E}(5X + 10), \mathbb{V}\text{ar}(X), \mathbb{E}(X^3)$

   (c) Find the mode of the random variable X

2. The height of a randomly selected person is normally distributed with expected value of 170 cm, and standard deviation of 5 cm.

   (a) Find the median height, the modal height.

   (b) What is the probability that a randomly chosen person will be taller than 180 cm? Lower than 165 cm? Between 165 cm and 190 cm?

   (c) What is threshold height for the top 5% tallest people? The top 10% lowest people?

   Hint: you may use `pnorm` and `qnorm` functions in `R`

3. Download the dataset on Moscow flat prices, `goo.gl/zxwS7u`. Open it in `R`.

   (a) Plot the histogram of of living span. What is the most frequent living span?

   (b) Plot the scatter-plot of kitchen span against living span. Are they positively or negatively related?

   (c) Provide summary statistics for span variables. What is the size of the biggest kitchen? The smallest living span?

   (d) What is the sample correlation between living span and kitchen span? Interpret this value.

4. Use the Moscow flat dataset. Estimate the linear regression model where the price of a flat depends on living span, kitchen span and dummy-variable for brick buildings.

   (a) Check the significance of each coefficient on the 5% level.

   (b) Check the significance of the regression.

   (c) Are brick buildings priced higher of lower than non-brick ones?

5. Use the Moscow flat dataset. Compare two models:

   Model A:

$$price_i = \beta_0 + \beta_{livesp} \cdot livesp_i + \beta_{kitsp} \cdot kitsp_i + \beta_{brick} \cdot brick_i + \varepsilon_i$$

   Model B:

$$price_i = \beta_0 + \beta_{brick} \cdot brick_i + \varepsilon_i$$

   (a) Carefully specify H_0 and H_a and significance level.

   (b) Which model is restricted one and which one is unrestricted?

   (c) What is the p-value of the F-test?

   (d) What is your conclusion?

2 Optional part. Just for fun!

1. You have a sample of 5 observations for variables X and Y:

X	1	2	3	2	-1
Y	5	-2	7	10	1

   (a) For each variable calculate sample mean, sample variance, sample standard deviation

   (b) Calculate sample correlation between X and Y

2. The random variable X is normally distributed $N(10; 4)$, and random variable Y is uniformly distributed on the segment $[5; 7]$.

   (a) Find $\mathbb{E}(X)$, $\mathbb{V}\text{ar}(X)$, $\mathbb{E}(Y)$

   (b) Find $\mathbb{P}(X > 10)$, $\mathbb{P}(Y < 6)$, $\mathbb{P}(Y > 6.5)$

   (c) Sketch the density functions of X and Y

   (d) Using `R` function `qnorm` find $\mathbb{P}(X < 14)$, $\mathbb{P}(X > 12)$

3. Use `R` to generate a random sample of 1000 observations from normal distribution $N(10, 4)$ (let's denote this vector by `x`) and a random sample of 1000 observations from distribution uniform between 5 and 7 (let's denote this vector by `y`).

   (a) Show first 10 values of `x` and `y`

   (b) Plot histogram of `x` and `y`

   (c) For each variable calculate sample mean, sample variance, sample standard deviation

   (d) Create a new vector `z` equal to `x` plus `y`

   (e) Show the scatter plot of `x` and `z`

   (f) Calculate sample correlation between `x` and `z`

   Hint: you may use functions like `rnorm`, `runif`, `var`, `mean`, `sd`, `qplot`. Do not forget to attach the package `ggplot2`.

4. You have a random sample of 100 persons. The mean wage is equal to 42 th. roubles with sample stardard deviation of 5 th. roubles. At 5% significance level test the null hypothesis that the mean wage in the population is equal to 40 th. roubles against the alternative that it is higher than 40 th. roubles.

5. James Bond tests a top secret hypothesis H_0 against even more secret alternative hypothesis H_a at $\alpha = 0.1$ significance level. The P-value is equal to 0.085. Should James reject the null-hypothesis?

Figure 1: Russian bears love Statistics!