Statistics. Exam-demo

17 March 2016

Problem 1

The distribution of random variable X is given by the table:

value of X	1	4	-3
probability	0.2	0.3	?

1. What is masked by "?" in the table?
2. Find $E(X)$, $E(X^2)$, $E(X^3)$
3. Find $Var(X)$ and σ_X

Problem 2

The distribution of random variables X and Y is given by the table:

joint probabilities	$X = 0$	$X = -1$	$X = 2$
$Y = 0$	0.2	0.1	0.1
$Y = 2$	0.2	0.3	0.1

1. Find $Var(X)$, $Var(Y)$
2. Find $Cov(X, Y)$
3. Find $Corr(X, Y)$

Problem 3

The density function of a random variable X is given by

$$f(x) = \begin{cases} \frac{2}{3}x, x \in [0; 1.5] \\ 4 - 2x, x \in [1.5; 2] \\ 0, \text{otherwise} \end{cases}$$

1. Draw the density function
2. Find $P(X > 1)$
3. Find the theoretical median of X
4. What is bigger $E(X)$ or 1?

Problem 4

Random variables X and Y are independent, $X \sim N(5; 10)$, $Y \sim N(-3; 5)$.

1. Find $E(X - 2Y)$, $Var(X - 2Y)$
2. Find $Corr(X - 2Y, X + 3Y)$

Problem 5

Consider a small random sample of 4 observations, $X_1 = 2$, $X_2 = -3$, $X_3 = 4$, $X_4 = 1$.

Calculate sample mean, sample variance and sample median.

Problem 6

Researcher Eleonora has sampled 900 small firms. The sample mean revenue was equal to 3000 dollars per week with sample standard deviation of 1000 dollars per week.

1. At 5% significance level test the hypothesis that the expected revenue is equal to 3100.
2. Build a 95% confidence interval for expected revenue.

Problem 7

Researcher Veniamin wonders whether the type of pies he eats influences the number of solved problems. Here y_i is the number of solved problems during the evening, x_i — number of apple-pies eaten, z_i — number of meat-pies eaten:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$$

Table 3: Fitting linear model: y ~ x + z

	Estimate	Std. Error	t value	Pr(>|t|)
x	3.056	0.06083	50.24	3.001e-71
z	1.917	0.085	22.55	2.344e-40
(Intercept)	4.658	0.4869	9.568	1.15e-15

1. Interpret the coefficients before x_i
2. Test the significance of every coefficient at $\alpha = 0.1$ significance level
3. Calculate 95% confidence interval for coefficient β_2
4. Forecast y for $x = 3$ and $z = 7$

Problem 8

Consider two regression models:

Model A:
$$\hat{y}_i = 90 + 2.8x_i$$

Model B:
$$\hat{y}_i = -80 + 3.1x_i - 2.6z_i + 2.9w_i$$

The F-test for these two models gave P-value of 0.0002016.

1. Describe the F-test for nested model comparison: What is the H_0 for the test? The alternative hypothesis H_a?
2. Which model do you prefer in this particular case and why?

Problem 9

Researcher Veniamin would like to predict whether it will rain tomorrow (y_i) given the number of times his cat has meowed today (x_i). Veniamin has estimated the logit model:

$$P(y_i = 1) = F(\beta_1 + \beta_2 x_i)$$

Coefficient	Estimate	Std. error
$\hat{\beta}_1$	5	2.8
$\hat{\beta}_2$	-2	0.5

1. What is the probability of rain if the cat has meowed 3 times?
2. Is the coefficient $\hat{\beta}_2$ significantly different from zero at 5% significance level?

Problem 10

You have 4 observations on weight, x_i (kg), and ice-cream preference (y_i):

x_i	60	80	90	70
y_i	chocolate	vanilla	chocolate	chocolate

1. What is the Gini index?
2. Build a classification tree for y using x as predictor