MATH40005 Spring 2022 Blackboard Quiz 2

Question 1

Suppose the correlation between two samples $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots y_n$ is computed to be r_{xy}. Which of the following values CANNOT be the value of r_{xy}?

(a) 0.0

(b) 0.5

(c) -0.9

(d) 1.1

(e) 1.0

Question 2

Suppose the random variables $X_1, X_2, \ldots, X_n$ are assumed to be independent and follow a normal distribution with both the mean μ and the variance σ^2 known. What is the distribution of $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$?

(a) $N(\mu, \sigma^2)$

(b) $N(0, 1)$

(c) t_n

(d) t_{n-1}

(e) $U[0, 1]$

Question 3

Suppose the random variables $X_1, X_2, \ldots, X_n$ are assumed to be independent and follow a normal distribution with the mean μ known and the variance σ^2 unknown. Suppose that the sample variance is S^2. What is the distribution of $\frac{\overline{X}-\mu}{S/\sqrt{n}}$?

(a) $N(\mu, \sigma^2)$

(b) $N(0, 1)$

(c) t_n

(d) t_{n-1}

(e) $U[0, 1]$

Question 4

We wish to maximise the likelihood $L(\theta|x)$, but the algebraic expression for $L(\theta|x)$ is very complicated. However, the algebraic expression for the log-likelihood $\log L(\theta|x)$ is much simpler. Suppose we find the value of θ that maximises the log-likelihood to be $\theta = f(x)$. Which of the following statements is correct?

(a) $\theta = f(x)$ does not maximise the likelihood

(b) $\theta = f(x)$ maximises the likelihood

(c) $\theta = f(x)$ might maximise the likelihood in certain cases, but not in all cases.

Question 5

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are all distributed according to a distribution F_X with unknown mean μ_1 and unknown variance σ_1^2, and the random variables $Y_1, Y_2, \ldots, Y_m$ are all distributed according to a distribution F_Y with unknown mean μ_2 and unknown variance σ_2^2. Suppose data are observed for these random variables and we wish to use Student's two-sample t test to test if μ_1 and μ_2 are different. What is the appropriate null hypothesis in this case?

(a) $\mu_1 \neq \mu_2$

(b) $\mu_1 = \mu_2$

(c) $\mu_1 < \mu_2$

(d) $\mu_1 > \mu_2$

(e) $\mu_1 = 0$

Question 6

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are all distributed according to a distribution F_X with unknown mean μ_1 and unknown variance σ_1^2, and the random variables $Y_1, Y_2, \ldots, Y_m$ are all distributed according to a distribution F_Y with unknown mean μ_2 and unknown variance σ_2^2. Suppose data are observed for these random variables and we wish to use Student's two-sample t test to test if μ_1 and μ_2 are different. Which of the following assumptions is NOT necessary for the appropriate test statistic T to follow a t-distribution?

(a) F_X and F_Y are both normal distributions.

(b) The random variables $X_1, X_2, ..., X_n$ are independent.

(c) The random variables $X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_m$ are independent.

(d) $\sigma_1^2 = \sigma_2^2$

(e) $m = n$

Question 7

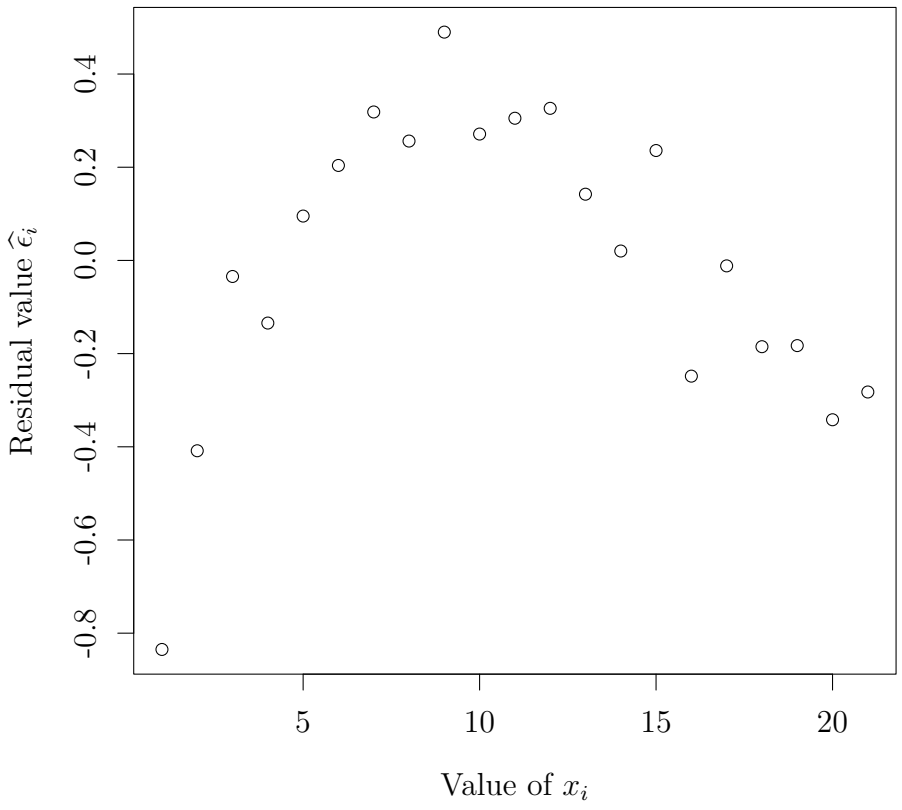
Suppose we have data $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, and we wish to fit a linear regression model of the form

$$Y_i = \beta_0 + \beta_1 x_1 + \epsilon_i$$

where each $\epsilon_i \sim N(0, \sigma^2)$, for some unknown value σ^2. Suppose we decide to choose β_0 and β_1 to be $\widehat{\beta}_0$ and $\widehat{\beta}_1$. If we plot the residuals

$$\widehat{\epsilon}_i = y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i),$$

we get the figure below.



Which one of the following statements is correct?

(a) The data fit the linear model well.

(b) The data do not fit the linear model well.

(c) We cannot tell if the data fit the linear model well or not.

Question 8

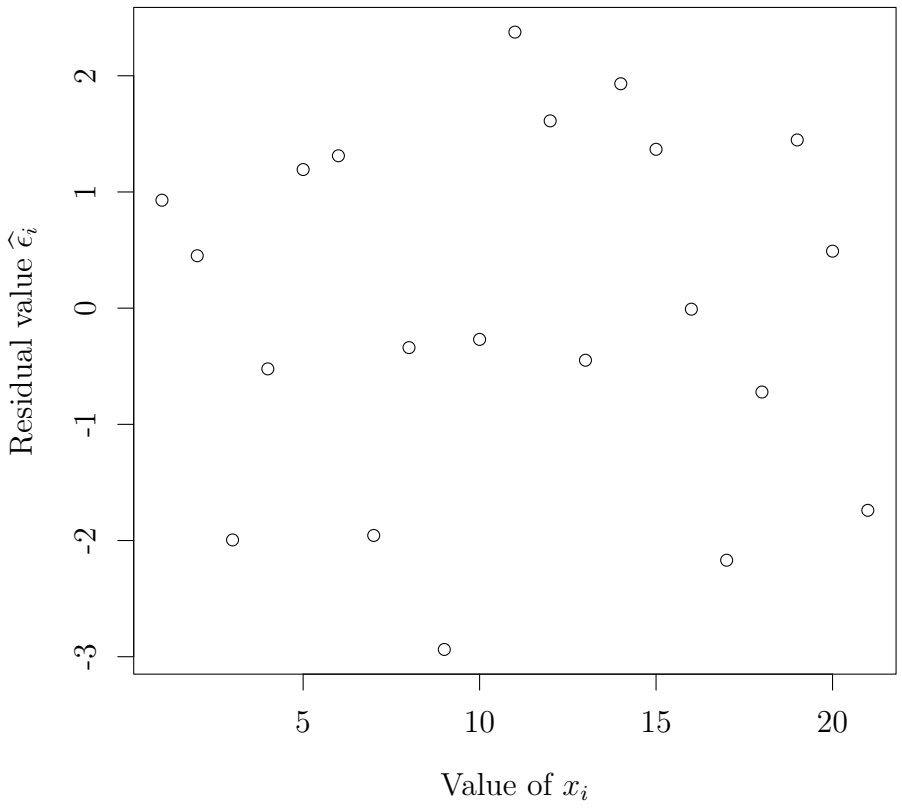
Suppose we have data $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, and we wish to fit a linear regression model of the form

$$Y_i = \beta_0 + \beta_1 x_1 + \epsilon_i$$

where each $\epsilon_i \sim N(0, \sigma^2)$, for some unknown value σ^2. Suppose we decide to choose β_0 and β_1 to be $\widehat{\beta}_0$ and $\widehat{\beta}_1$. If we plot the residuals

$$\widehat{\epsilon}_i = y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i),$$

we get the figure below.



Which one of the following statements is correct?

(a) The data fit the linear model well.

(b) The data do not fit the linear model well.

(c) We cannot tell if the data fit the linear model well or not.

Question 9

Suppose the random variables $X_1, X_2, \ldots, X_n$ are assumed to be independently and identically distributed according to an exponential distribution with unknown parameter θ. If the prior distribution $\pi(\theta)$ for θ is chosen to be a $\Gamma(\alpha, \beta)$ distribution, what type of distribution is the posterior distribution $\pi(\theta|\mathbf{x})$?

(a) Exponential distribution.

(b) Gamma distribution.

(c) Unknown distribution, depends on the data.

(d) Normal distribution.

Question 10

Suppose an experiment results in the following dataset:

$$\mathbf{x} = \{45, 89, 32, 67\}.$$

Select **all** of the following which are bootstrap samples of the above data set $\mathbf{x}$.

(a) $\{32, 45, 67, 89\}$

(b) $\{45, 32, 32, 89\}$

(c) $\{32, 67, 45, 89, 45\}$

(d) $\{45, 89, 11, 67\}$

Solutions

1. **(d)**; because $r_{xy} \in [0, 1]$, and $1.1 \notin [0, 1]$.

2. **(b)**; see Corollary 3.1.3 in the notes, and then compute the expectation and variance.

3. **(d)**; see Proposition 3.4.1 and Theorem 3.4.4 in the notes.

4. **(b)**; because log is a monotonic function, see Section 8.5.2 in the notes.

5. **(b)**; because we wish to test for any difference, i.e. our alternative is $\mu_1 \neq \mu_2$, so our null hypothesis is the complement of this.

6. **(e)**; because it is not required that the two samples have the same size, but the other four conditions are required.

7. **(b)**; because there is clearly trend in the residuals, and so they are unlikely to be independent, which violates one of the model assumptions.

8. **(a)**; because the residuals are centred around 0 and appear to be independent. However, if it is argued that it is hard to tell if the residuals are **normally distributed**, then option (c) would also be acceptable, given this reasoning.

9. **(b)**; because the gamma distribution is the conjugate prior of the exponential distribution. See Exercise 10.3.6 in the notes.

10. **(a) and (b)**; (a): all of the element in the sample are in $\mathbf{x}$, and it has the same number of elements as $\mathbf{x}$. (b): for the same reason as (a), and it does not matter if there are duplicates, since bootstrap samples are sampled with replacement. (c): not a bootstrap sample because it has more elements than $\mathbf{x}$. (d): not a bootstrap sample because $11 \notin \mathbf{x}$.