

# DEMO Applied Statistics Exam - **First part**

--> Each question may have zero, one, or more correct answers. Indicate **all** correct answers.  
--> The response to each of the 10 questions is considered valid if and only if all correct answers are indicated correctly.

---

The use of R and RStudio is allowed.

Consulting the R help (using the ? command) and course materials is allowed.

The use of other sources is not allowed, particularly searching for solutions to questions on the internet or communicating with others during the exam.

Hi, Filippo. When you submit this form, the owner will see your name and email address.

1

The following figure depicts the summary of a linear model explaining the total daily entries into Area C in Milan (variable "**Total**"), based on the number of diesel car entries (variable "**Diesel**"), petrol car entries (variable "**Petrol**"), and the "**Weekend**" factor (dummy variable, where Weekend=1 if Saturday/Sunday, 0 otherwise).

Identify the true statement(s):



Call:

```
lm(formula = Total ~ Petrol + Diesel + Weekend, data
```

Residuals:

Min	1Q	Median	3Q	Max
-4722.3	-1057.1	-451.3	644.5	5361.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	6450.5088	2665.4669	2.420	0.022524	*
Petrol	1.1911	0.2753	4.326	0.000186	**
Diesel	1.2203	0.1717	7.109	1.21e-07	**
Weekend	-3097.2144	1402.2830	-2.209	0.035869	*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.'

Residual standard error: 2149 on 27 degrees of freedom

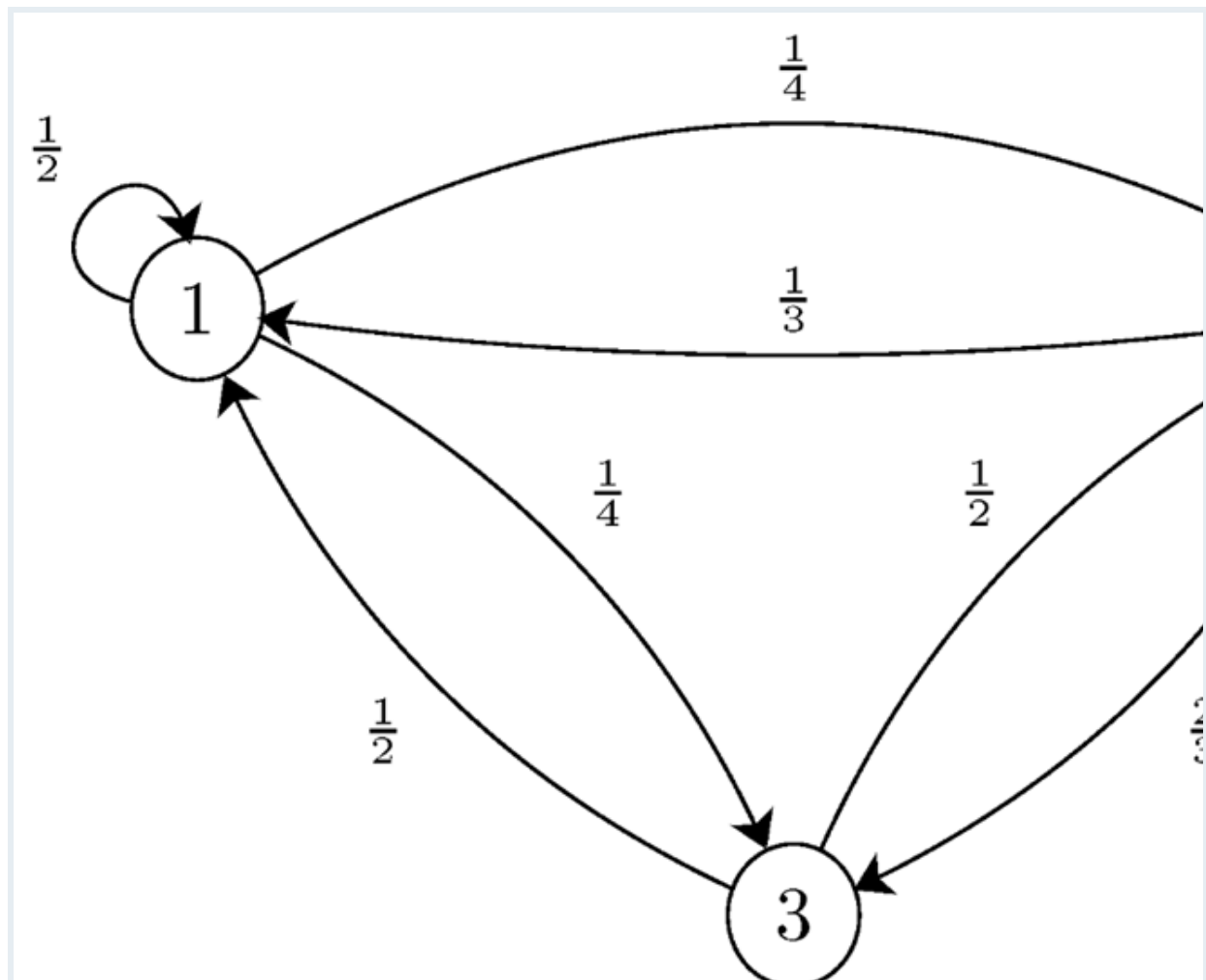
Multiple R-squared: 0.9131, Adjusted R-squared:

F-statistic: 94.58 on 3 and 27 DF, p-value: 1.936e-1

- ☐ Based on the F-statistic, there is at least one predictor with a coefficient significantly different from zero in the model.
- ☐ The R2 index indicates poor fit of the model to the data.
- ☐ The "Weekend" factor is significant at the 5% level.
- ☐ All other factors being equal, during the weekend, the average total daily number of entries is higher.

2

Consider the following Markov Chain and indicate the true statement(s): 



- ☐ The chain has period 2
- ☐ The chain is irreducible and aperiodic

3

In a hypothesis test:



- ☐ The p-value can take any value between -1 and 1.
- ☐ The p-value is the smallest level of significance for which the null hypothesis is rejected.

4

Consider two models, Model A and Model B, fitted to the same dataset. The AIC values for Model A and Model B are 150 and 160, respectively.

The BIC values for Model A and Model B are 155 and 165, respectively.

Indicate the true statement(s): 

- ☐ Model A is preferred to B according to BIC but not AIC
- ☐ Model A is preferred to B according to both AIC and BIC
- ☐ Model A is preferred to B according to AIC but not BIC
- ☐ Model B is preferred to A according to both AIC and BIC

5

The median is 

- ☐ The arithmetic average of all values
- ☐ The value that occurs most frequently
- ☐ The middle value when the data is arranged in ascending order
- ☐ The difference between the largest and smallest values

6

Which statement accurately describes the bias-variance trade-off? 

- ☐ Increasing model complexity tends to reduce bias but increase variance.
- ☐ Decreasing model complexity tends to increase bias but reduce variance.
- ☐ Balancing bias and variance is not crucial for model performance.

☐ Bias and variance are independent and do not influence each other.

7

The following figure presents the summary of a linear model explaining the rental price of a villa in Greece (variable '**price**,' expressed in euros per day), based on the distance from the sea (variable '**distance**') and the presence of a swimming pool (dummy variable **D1**, D1=1 if the pool is present).

Identify the true statements. 

```
Call:
lm(formula = price ~ distance + D1, data = villas)

Residuals:
    Min       1Q   Median       3Q      Max
-42.219 -14.428  -3.318   13.104   55.403

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  297.285329   9.546230   31.14  <2e-16 **
distance     -0.081217   0.005759  -14.10  <2e-16 **
D1           74.138656   6.056087   12.24  <2e-16 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'

Residual standard error: 22.31 on 52 degrees of freedom
Multiple R-squared:  0.8782,    Adjusted R-squared:
F-statistic: 187.5 on 2 and 52 DF,  p-value: < 2.2e-16
```

- ☐ The R2 index suggests a poor fit of the model to the data.
- ☐ The factor 'presence of a swimming pool' is significant
- ☐ As the distance from the sea increases, on average, a significant increase in the rental price of a villa is observed, holding other factors constant.
- ☐ The distance from the sea significantly influences the average rental price of a villa.



This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

**Microsoft Forms** | AI-Powered surveys, quizzes and polls [Create my own form](#)

[Privacy and cookies](#) | [Terms of use](#)