PO91Q - Fundamentals in Quantitative Research Methods

Mock Exam 2025/26

Dr Florian Linke
Florian.Linke@warwick.ac.uk

Table of contents

UNIVERSITY OF WARWICK

Preface

- Part I contains the questions for the exam
  - These are divided into four blocks: A, B, C, and D.
  - Block A is compulsory and must be answered by all candidates.
  - You must then select ONE block out of B, C, and D.
  - Should you change your mind during the exam, indicate clearly which block you wish to have marked.
  - You can achieve a maximum of 120 points.
  - Information on the maximum number of points obtainable is outlined at the beginning of each block
- Part II contains the case study to which all questions in Block A of Part I refer
- Part III contains the relevant statistical tables
  - Normal curve tail probabilities are provided in Table 4, page 9.
  - t-distribution critical values are provided in Table 5, page 10.
- Part IV contains the formula collection and notation of the module
  - If you use notation other than in the formula collection (page 11) or notation table (page 12), give an explanation.
- Allowed time: 2 hours.
- A university-approved calculator is allowed.

Part I: Exam Questions

1 | Block A – Compulsory

80 points, points for each question according to level

All questions in this section relate to the case study in Part II.

Level 1 (each question 4 points)

1. Subsection 7.1: what does `header=TRUE` achieve?
2. Which package is necessary for the recoding presented in Subsection 7.4?
3. Verbalise the code in Subsection 7.2.

Level 2 (each question 6 points)

4. Rewrite the recoding of `drive_cat` in Subsection 7.4 with the option `right = TRUE`.
5. Interpret the output of Subsection 7.5
6. Interpret the slope coefficient in Model 1 in Subsection 7.8.
7. Interpret the intercept of Model 2 in Subsection 7.8.
8. Subsection 7.8: Why is the goodness of fit in Models 1 and 2 so different?

Level 3 (each question 8 points)

9. Subsection 7.6: What is the sample mean of `simd$drive_primary`?
10. Given the information provided in Subsection 7.3, would a non-linear transformation be necessary for including the variable `cime_rate` in a regression model? Why / why not?

2 | Block B – True / False Statements

40 points, 5 points for each statement

State whether each of the following statements is true or false. Explain why the statement is true or false.

1. With a smaller effect size of interest, a higher sample size is required to achieve the same level of statistical power.
2. The p-value and the α-level are the same thing.
3. Adjusted R^2 ($\bar{R}^2$) can be negative.

3 | Block C – Calculations

40 points, points for each question in parentheses

1. Fit a linear regression model of the type $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ to the data in Table 1. Specify the full sample regression function. **(20)**

x	y
4	10
7	13
8	9
11	5
19	0

Table 1: Regression Data

4 | Block D – Explanations

40 points, 8 points for each method / concept

Explain each of the following methods or concepts in your own words, and illustrate your explanation with a worked example of your choice.

1. Non-Linear Transformation of Variables
2. Statistical Power

Part II: Case Study

5 | Preface

This case study resembles a shorter version of one you will receive for the exam.

Today you will be working with the "Scottish Index of Multiple Deprivation". The data are taken from Scottish Government (2020).

> The Scottish Index of Multiple Deprivation is a relative measure of deprivation across 6,976 small areas (called data zones). If an area is identified as 'deprived', this can relate to people having a low income but it can also mean fewer resources or opportunities. SIMD looks at the extent to which an area is deprived across seven domains: income, employment, education, health, access to services, crime and housing. It ranks data zones from most deprived (ranked 1) to least deprived (ranked 6,976).
>
> SIMD is the Scottish Government's standard approach to identify areas of multiple deprivation in Scotland. It can help improve understanding about the outcomes and circumstances of people living in the most deprived areas in Scotland. It can also allow effective targeting of policies and funding where the aim is to wholly or partly tackle or take account of area concentrations of multiple deprivation.
>
> Data zones in rural areas tend to cover a large land area and reflect a more mixed picture of people experiencing different levels of deprivation. This means that SIMD is less helpful at identifying the smaller pockets of deprivation found in more rural areas, compared to the larger pockets found in urban areas. SIMD domain indicators can still be useful in rural areas if analysed separately from urban data zones or combined with other data.

6 | Codebook

Variable	Type	Description
`Data_Zone`	Code	2011 Data Zone
`Intermediate_Zone`	Name	2011 Intermediate Zone name
`Council_area`	Name	Council area name
`ALCOHOL`	Standardised ratio	Hospital stays related to alcohol use: standardised ratio
`drive_primary`	Time (minutes)	Average drive time to a primary school in minutes
`crime_rate`	Rate per 10,000 population	Recorded crimes of violence, sexual offences, domestic housebreaking, vandalism, drugs offences, and common assault per 10,000 people
`sim_rank`	Rank	Rank of data zones from most deprived (ranked 1) to least deprived (ranked 6,976)
`SMR`	Standardised ratio	Standardised mortality ratio

Table 2: Codebook for `simd_exam` Data Set

7 | Output

7.1 Import Data

```
simd <- read.csv('simd.csv', header=TRUE)
```

7.2 Filter

```
simd2 <- filter(simd, crime_rate<400)
```

7.3 Summary

```
summary(simd2$crime_rate)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.83   94.43  161.48  177.65  250.05  399.95
```

7.4 Recoding

```
simd2 %>%
  mutate(drive_cat = cut(drive_primary,
                         breaks = c(0, 2.79, 3.572, 30),
                         labels = c("Short", "Average", "Long"),
                         right = FALSE,
                         include.lowest = TRUE)) -> simd2

simd2 %>%
  mutate(alc_cat = cut(ALCOHOL,
                         breaks = c(0, 31.83, 644),
                         labels = c("Low", "High"),
                         right = FALSE,
                         include.lowest = TRUE)) -> simd2
```

7.5 Inference 1

```
t.test(simd2$crime_rate,
       mu = 200,
       alternative = "two.sided")
##
##  One Sample t-test
##
## data:  simd2$crime_rate
## t = -16.103, df = 4969, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 200
## 95 percent confidence interval:
##  174.9249 180.3677
## sample estimates:
## mean of x
##  177.6463
```

7.6 Inference 2

```
sd(simd2$drive_primary)
## [1] 1.34018

t.test(simd2$drive_primary, alternative = "less", mu = 5)
##
##  One Sample t-test
##
## data:  simd2$drive_primary
## t = -97.62, df = 4499, p-value < 2.2e-16
## alternative hypothesis: true mean is less than 5
```

7.7 Regression Models

```
model1 <- lm(SMR ~ ALCOHOL, data=simd2)

model2 <- lm(SMR ~ alc_cat, data=simd2)
```

7.8 Results Table

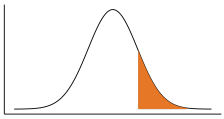
	Dependent Variable: Standardised Mortality Ratio	
	(1)	(2)
Alcohol-related admissions to hospital	0.215***	
	(0.007)	
Alcohol related admissions to hospital (high)		25.388***
		(1.200)
(Intercept)	75.386***	74.014***
	(0.755)	(1.039)
Num.Obs.	4970	4970
R2	0.162	0.083
R2 Adj.	0.162	0.083

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

Table 3: Regression Models

Part III: Statistical Tables

8 | Normal-Distribution



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002

Table 4: Right-Tail Probabilities under the Normal Distribution

9 | t-Distribution

	Confidence Level									
	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	One-Tail Probability									
df	$t_{0.25}$	$t_{0.20}$	$t_{0.15}$	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.22	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Table 5: Probabilities under the t-Distribution

Part IV: Formulae and Notation

10 | Formulae

Statistic	Formula
Confidence Interval	$Pr(\bar{y} - t_{\alpha/2} \cdot se \leq \mu \leq \bar{y} + t_{\alpha/2} \cdot se) = 1 - \alpha$
Deviation	$d = y_i - \bar{y}$
Mean	$\bar{y} = \frac{\Sigma y_i}{n}$ $\mu = \Sigma y P(y) = E[y]$
Position of pth percentile	$P = (n + 1) \cdot \frac{p}{100}$
Range	$y_{range} = y_{max} - y_{min}$
Standard Deviation	$s = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n - 1}}$
Standard Error	$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$ $se = \frac{s}{\sqrt{n}}$
t-test	$t = \frac{\bar{y} - \mu_0}{se}$
Variance	$s^2 = \frac{\Sigma(y_i - \bar{y})^2}{n - 1}$
z-score	$z = \frac{y - \mu}{\sigma}$

Table 6: Formulae for PO11Q

11 | Notation

Symbol	Explanation
d	Deviation
n	Sample Size
s	Standard Deviation
s^2	Variance
$\bar{y}$	Mean
y_i	Observation i
f	(Absolute) Frequency
cf	Cumulative (Absolute) Frequency
rf	Relative Frequency
crf	Cumulative Relative Frequency
$E[x]$	The expected value of x
μ	Mean of the Population
se	Standard Error (with s of sample)
σ	Standard Deviation of the Population
$\sigma_{\bar{y}}$	Standard Error (with σ of population)
t	t-value
z	z-value

Table 7: Notation

Part V: References

Scottish Government. (2020). Scottish Index of Multiple Deprivation 2020 (SIMD 2020). https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/