

MATH 204: Principles of Statistics 2

Practice Final Examination: Solutions

Winter 2026

MATH 204 Specific Instructions.

Please write your answers on the exam in the boxes provided for each question; only answers written inside the correct boxes will be marked. Extra space for rough work can be found at the end of the exam paper. Please write clearly. Please do not use highlighter.

This paper contains five questions. Each question carries 20 marks. Credit will be given for all questions attempted. The total mark available is 100 but rescaling of the final mark may occur.

Calculators may be used. Dictionaries and Translation dictionaries are permitted. A formula sheet is provided. Standard notation from the course will be used, and you may use standard notation in your answers without definition. For full marks, you must give justification for your answers.

1. In an experimental study designed to compare the efficacy of five pain relief treatments (A,B,C,D,E), a balanced completely randomized design involving a total of 20 patients was used. The measured outcome is a pain reduction score measured on a continuous scale from -10 to 10.

The data were analyzed as a Completely Randomized Design (CRD) which produced the following ANOVA table from which some values have been deleted, and labelled (i)-(vi).

Source	DF	SS	MS	F
Treatment	(i)	(iii)	(iv)	(vi)
ERROR	(ii)	67.938	(v)	
Total	19	163.274		

(a) Identify the values (i)-(vi)

6 MARKS

Q1a

Using the usual computation rules, we find the table to be

Source	DF	SS	MS	F
Treatments	4	95.336	23.834	5.263
ERROR	15	67.938	4.529	
Total	19	163.274		

so

(i) = 4 (ii) = 15 (iii) = 95.336 (iv) = 23.834 (v) = 4.529 (vi) = 5.263

(b) Carry out the ANOVA F-test and report the conclusion.

4 MARKS

Q1b

We compare the F-statistic with the $F_{4,15}$ distribution upper $\alpha = 0.05$ point, which is from tables found to be 3.06. The test statistic lies in the rejection region, and we therefore reject H_0 and conclude that the treatments do have pain relief results.

Unless otherwise directed, you may choose the default $\alpha = 0.05$ level in all tests.

- (c) Taking drug treatment A as the baseline treatment, the following table reports the parameter summaries using the usual contrast parameterization.

	Estimate	Std. Error	t value	p-value
(Intercept)	-0.1273	1.0641	-0.1196	0.9064
TreatmentB	0.3875	1.5049	0.2575	0.8003
TreatmentC	4.7498	1.5049	3.1564	0.0065
TreatmentD	3.1467	1.5049	2.0910	0.0540
TreatmentE	5.3071	1.5049	3.5266	0.0031

Construct a 95% confidence interval for the difference in means between Treatment C and Treatment A. 6 MARKS

Q1c

In this case we may construct the confidence interval by the standard formula as given in the analysis of the linear model which underpins the ANOVA analysis

$$\text{Estimate} \pm t_{\alpha/2, n-k} \times \text{e.s.e}(\text{Estimate})$$

where the $n - k$ is the error degrees of freedom. Here $t_{\alpha/2, n-k} = t_{0.025, 15} = 2.1314$, so the interval is

$$4.7498 \pm 2.1314 \times 1.5049 = (1.5422, 7.9574).$$

The other acceptable approach is to take the post-hoc two sample test route, and use the interval based on the estimate plus or minus

$$t_{\alpha/2, n_1+n_2-2} \times s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = t_{0.025, 6} \times \sqrt{4.529} \times \sqrt{\frac{1}{4} + \frac{1}{4}}$$

which yields (1.0676, 8.4320). Here we do not have access to the pooled estimate s_P ; if we did, we would use that in place of s .

- (d) Explain what steps you would need to take to verify the statistical conclusions of the analysis based on these data. 4 MARKS

Q1d

The assumptions would need to be checked, specifically Normality of the residual errors (which should be checked using a histogram or series of histograms by treatment group), and equality of variances across the treatment groups, which should be checked using Levene's test. Here, independence of the residual errors is guaranteed by the CRD setting.

2. A randomized block design (RBD) was used to compare the standardized reading test scores of children in three age groups who underwent four different literacy preparations. Based on the observed data, the following ANOVA table was constructed.

SOURCE	DF	SS	MS	F	p
Preparation Group	3	43.896	14.632	4.048	0.0141
Age Group	2	27.956	13.978	3.867	0.0301
Interaction	6	40.631	6.772	1.873	0.1124
ERROR	36	130.140	3.615		
TOTAL	47	242.622			

The sums of squares quantities are reported to 3 decimal places, and the p-values are reported to 4 decimal places. The study is balanced.

- (a) Explain briefly why an RBD would be used in this setting.

4 MARKS

Q2a

Because we believe that there would be a systematic difference between the reading abilities of children of different ages, irrespective of the preparation type.

- (b) Identify all the conclusions that can be drawn from the ANOVA table. Justify these conclusions.

6 MARKS

Q2b

We can conclude that there is a significant treatment effect, a significant block effect, but no significant interaction. In this balanced design, we can interpret the p-values directly to make these conclusions. The effect of changing the preparation from one level to another is the same within each Age group.

- (c) Compute the estimate of the residual error variance, σ^2 , if the interaction were to be omitted and the data re-analyzed 4 MARKS

Q2c

The estimate in the new analysis is

$$\frac{SSE}{n - b - k + 1} = \frac{40.631 + 130.140}{48 - 3 - 4 + 1} = \frac{170.771}{42} = 4.066$$

- (d) Construct the ANOVA table if the data were to be analyzed as a Completely Randomized Design investigation into the literacy preparation, and identify whether this analysis would indicate that there was a significant difference between literacy preparation groups. 6 MARKS

Q2d

Here we simply pool the blocking and interaction rows into the error row.

SOURCE	DF	SS	MS	F
Preparation Group	3	43.896	14.632	3.240
ERROR	44	198.727	4.517	
TOTAL	47	242.622		

We compare this with the $F_{0.05,3,44}$ critical value, which is 2.82, so the result is still significant.

3. In a study of the relationship between dietary salt intake and increased systolic blood pressure, $n = 78$ healthy human subjects were recruited for a study. Using dietary diaries, the amount of salt (in grammes), denoted x_1 , consumed per week was recorded, along with the subject's age (x_2) The response, Y , a rescaled measure of excess (above normal) systolic blood pressure, was recorded at the end of the study.

The first model fitted was the main effects only model $X_1 + X_2$, and this yielded the following ANOVA table:

	DF	SS	MS	F
X1	1	564.3428	564.3428	5.7563
X2	1	257.5689	257.5689	2.6272
ERROR	75	7352.8946	98.0386	
TOTAL	77	8174.8063		

The coefficients table for the first analysis was

	Estimate	Std. Error	t value	p-value
(Intercept)	27.2112	4.1348	6.5811	0.0000
X1	-3.1812	2.2691	-1.4020	0.1650
X2	1.8402	1.1353	1.6209	0.1092

The second model fitted was $X_2 + X_1$: this yield the following ANOVA table

	DF	SS	MS	F
X2	1	629.2117	629.2117	6.4180
X1	1	192.7000	192.7000	1.9656
ERROR	75	7352.8946	98.0386	
TOTAL	77	8174.8063		

and the following table of coefficients:

	Estimate	Std. Error	t value	p-value
(Intercept)	27.2112	4.1348	6.5811	0.0000
X2	1.8402	1.1353	1.6209	0.1092
X1	-3.1812	2.2691	-1.4020	0.1650

- (a) Write out, in terms of coefficients and predictors, the exact forms of the models fitted in these two analyses. 6 MARKS

Q3a

First analysis : $E[Y|x] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
Second analysis : $E[Y|x] = \beta_0 + \beta_1 x_2 + \beta_2 x_1$

that is, these are in fact the **same** model, but with the terms relabelled. This is why the parameter estimates are the same. Note, however, that the ANOVA tables are different; this is because the ANOVA tables cannot be constructed uniquely when we have combinations of continuous predictors.

- (b) A third analysis fitted the model $X_1 + X_2 + X_1 : X_2$, and yielded the SSE value of 7208.224. On the evidence of this result, does the inclusion of the interaction term yield a significantly improved fit compared to either the fit of the first analysis or the second analysis ? Justify your answer. 6 MARKS

Q3b

In this case

$$F_{obs} = \frac{(SSE_R - SSE_C)/(k - g)}{SSE_C/(n - k - 1)}$$

where $k = 3, g = 2$ so that

$$k - g = (n - g - 1) - (n - k - 1) = 1.$$

Therefore

$$F_{obs} = \frac{(7352.895 - 7208.224)}{7208.224/74} = 1.485$$

We compare this with the 0.05 upper α point of the $F_{k-g, n-k-1}$ distribution, which in this case is

$$F_{0.05, 1, 74} = 3.97$$

and we do not reject this as an adequate simplification. Therefore, there is no evidence that the interaction should be included.

(c) A fourth analysis led to the following ANOVA table

	DF	SS	MS	F
X1	1	564.3428	564.3428	5.6357
ERROR	76	7610.4635	100.1377	
TOTAL	77	8174.8063		

and the following table of coefficients:

	Estimate	Std. Error	t value	p-value
(Intercept)	27.3947	4.1772	6.5581	0.0000
X1	0.4822	0.2031	2.3740	0.0201

Compute the correlation coefficient between X_1 and Y .

5 MARKS

Q3c

We have that

$$r^2 = R^2 = \frac{SSR}{SS} = \frac{564.343}{8174.806} = 0.069$$

and as the coefficient attached to X_1 is positive, we have that

$$r = \sqrt{0.069} = 0.263$$

- (d) Assuming that the Normality, independence and equal variance assumptions for the residual errors are met for this problem, can we state from the analysis in (c) that in reality, the expected value of Y increases by an amount equal to 0.4822 for a unit increase in X_1 ? Justify your answer.

3 MARKS

Q3c

No. This would only be the case if the model is correctly specified, and there is some evidence that this model is not correctly specified, as X_2 seems to be influential.

4. The effect of different wage incentives on productivity of workers in a factory was studied. Workers on two shifts were categorized into three skill levels, and each worker was given a personal incentive scheme involving a financial bonus per item produced. Productivity was measured by recording the average number of items produced per day over a three month period after the scheme was introduced. Data recorded are

- Skill: a factor predictor with three levels corresponding to the skill level of the worker (1, 2, 3).
- Shift: a factor predictor with two levels corresponding to the shift to which the worker was assigned (Shift 1 or Shift 2)
- Inc: a continuous covariate measuring the amount of financial incentive (in cents per item)
- Y: the daily average number of items produced.

The cross-categorization of workers by levels of the factor predictors is recorded below:

		Skill			Total
		1	2	3	
Shift	1	68	49	37	154
	2	12	90	52	154
Total		80	139	89	308

(a) Is there any evidence that workers possessing different Skill levels have different Shift assignments ? Justify your answer. 8 MARKS

Q4a

This question can be answered by a Chi-squared test. The fitted values are

		Skill			Total
		1	2	3	
Shift	1	40	69.5	44.5	154
	2	40	69.5	44.5	154
Total		80	139	89	308

which yields a chi-squared statistic of 53.822. By comparing with the χ^2_2 upper 0.05 point which is 5.991, we see that this is a strongly significant result, so we reject the null hypothesis of 'No difference in Shift assignments'.

- (b) Several analyses of the outcome were carried out using linear model analysis. The table below contains the SSE quantities and total numbers of parameters (denoted p) for several models, with standard model notation used.

	Model	SSE	p
0	1	11716.425	1
1	Skill	11567.411	3
2	Shift	10226.327	2
3	Inc	7407.532	2
4	Skill+Inc	7180.273	4
5	Shift+Inc	5640.334	3
6	Skill+Shift+Inc	5617.966	5
7	Skill+Shift+Inc+Shift:Inc	5586.372	6

On the basis of this information, select the most appropriate model from among those fitted to explain the variation in the response. 6 MARKS

Q4b

Before performing formal tests, we note from the changes in the SSE quantities that Inc certainly seems to have an effect, that Shift seems to have an effect, but that Skill has at most a minimal effect. Using the notation adopted in the course, we therefore consider the following comparisons

Reduced	Complete	SSE_R	SSE_C	k	g	$n - k - 1$	F
5	7	5640.334	5586.372	5	2	302	0.97
4	7	7180.273	5586.372	5	3	302	43.08
2	5	10226.327	5640.334	2	1	305	247.99
3	5	7407.532	5640.334	2	1	305	95.56

From these calculations, after comparing with the critical values from the F-table, we can conclude that the Model 5

Shift + Inc

seems to be the most appropriate model.

(c) Compute the R^2 (R-squared) statistic for the model Skill+Shift+Inc. 6 MARKS

Q4c

The R-squared statistic for any model is

$$R^2 = \frac{SSR}{SS}$$

where SSR is the aggregated contribution of every non-intercept term. We can compute SSR for any model using the identity

$$SS = SSR + SSE$$

where SS is equal to SSE from the null model (Model 0). Hence here

$$R^2 = \frac{11716.425 - 5617.966}{11716.425} = 0.521$$

5. (a) To investigate the potential relationship between a vaccination against measles and the presence of a specific microbiotic organism in the gut of children under twelve years old, a small case-control study was carried out. A total of 18 children for whom the organism was present, and 24 controls, were recruited and their vaccination status (Yes/No) was recorded. Count data for the study are recorded in the table below:

Vaccinated	Present		Total
	Yes	No	
Yes	17	22	39
No	1	2	3
Total	18	24	42

Using a large sample test based on the odds ratio, state whether there is any statistical evidence for an association. 6 MARKS

Q5a

Here the estimate is

$$\hat{\psi} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = 1.545$$

The statistic is

$$Z = \frac{\log \hat{\psi}}{\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}}$$

which in this case is equal to $Z = 0.344$ which we compare against the standard Normal distribution: this does not lead to a rejection of the null hypothesis, and we can conclude that there is no evidence of an association.

- (b) Is the large sample test appropriate here? Justify your answer. If a large sample test could not be used, identify one other procedure that could be used to test the null hypothesis of independence. 4 MARKS

Q5b

Here the estimated common incidence rate $39/42$ is quite near to 1, and so the Normal approximation is questionable. 2 MARKS

Fisher's exact test could be used as a test of independence, and randomization/permutation procedures could also be used. Both construct the exact null distribution of a suitably chosen test statistic (one of the counts in the table), assuming that the row and column labels are essentially assigned independently, conditional on the row and column totals. 2 MARKS

- (c) Use the Wilcoxon (Rank Sum) Test procedure to test for equality of medians for the following data: the data are measurements of lung capacity (in thousands of cm^3) for smokers (Pop 1.) and non-smokers (Pop 2.).

Smokers : 4.524, 4.777, 4.878, 4.885, 5.083, 5.148, 5.197, 5.210, 5.520, 5.612

Non-smokers : 5.012, 5.150, 5.155, 5.464, 5.465, 5.595, 5.668, 5.856, 6.096

- (i) Test at $\alpha = 0.05$ using a alternative hypothesis that specifies that non-smokers have a larger lung capacity. 8 MARKS

Q5ci

Here we have the ranks for sample 2 as 5, 8, 9, 12, 13, 15, 17, 18, 19, so $R_2 = 116$. We observe from the tables that the critical values for the sample sizes of $n_1 = 10$ and $n_2 = 9$ are 66 and 114, and 69 and 111. We need the latter pair for a one-tailed test at level 0.05. Hence we reject the null hypothesis in favour of the alternative, as $R_2 = 116 > 111$.

For a large sample version of the test, we have that

$$U = R_2 - \frac{n_2(n_2 + 1)}{2} = 116 - 9 \times 10/2 = 71$$

and thus using the formula sheet formula

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = 2.123$$

This lies in the one-sided rejection region $\{Z \geq 1.64\}$, so we reject the null hypothesis of equal medians in favour of the chosen alternative.

The sum of the ranks is always equal to $n(n + 1)/2 = 190$, so $R_1 = n(n + 1)/2 - R_2 = 74$. Hence $R_2 > R_1$. This supports the conclusion that median for non-smokers is greater than for smokers, as the non-smokers get higher ranks by a notable amount.

- (ii) Identify the smallest and largest possible values that the R_2 statistic could take for samples of this size. 2 MARKS

Q5cii

$$\text{Smallest value : } 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 = \frac{9 \times 10}{2} = 45$$

$$\text{Largest value : } 11 + 12 + 13 + 14 + 15 + 16 + 17 + 18 + 19 = \frac{19 \times 20}{2} - \frac{10 \times 11}{2} = 135.$$