ALLAMA IQBAL OPEN UNIVERSITY

USMAN NOTES
www.usmannotes.com

Name:

Registration No.:

Course Code: 794

Program: MSc Forestry Extension

Semester: Autumn 2026

ALLAMA IQBAL OPEN UNIVERSITY, ISLAMABAD

Course: Statistics (794)

Level: MSc Forestry Extension

Semester: Autumn 2026

Assignment No. 2 (Units 6–9)

Total Marks: 100 | Pass Marks: 50

SOLVED ASSIGNMENT GUIDE

(For Study & Reference)

USMAN NOTES www.usmannotes.com

Note: This guide is prepared for study and reference purposes only. Students are advised to read the relevant units of the course book carefully and to write the assignment in their own words, with their own examples and presentation, rather than copying this guide word for word.

Question 1: Null and Alternative Hypotheses, Steps in Hypothesis Testing; Test of a Claimed Mean (10 + 10 Marks)

Question as printed: (a) Define Null Hypothesis and Alternative Hypothesis. Explain the steps involved in hypothesis testing. (b) A manufacturer claims that the average weight of rice bags produced by a machine is 50 kg. A sample of 49 bags gave a mean weight of 48.5 kg. The population standard deviation is known to be 4 kg. Test at 5% level of significance whether the machine is producing bags with the claimed average weight.

Introduction

Units 6–9 of this course move from describing data to drawing inferences from samples, and the central instrument of that inference is the test of significance. A researcher in forestry or agriculture rarely observes a whole population: he weighs a sample of bags, measures a sample of trees, or harvests a sample of plots, and must then judge a claim about the population from incomplete and variable evidence. Because samples vary, the difference between a sample mean and a claimed population mean is, by itself, evidence of nothing; the difference must be measured against the variability that sampling alone would produce. Hypothesis testing is the formal procedure that makes this judgement with a stated, controlled risk of error. This answer defines the two hypotheses between which the procedure adjudicates, lays out its steps in full, and then works the procedure through on the rice-bag claim exactly as an examiner would require: hypotheses, significance level, statistic, computation, critical value, decision and interpretation.

USMAN NOTES www.usmannotes.com

Part (a): The Two Hypotheses and the Steps of the Test

Null hypothesis (H_0). The null hypothesis, symbolised H_0, is a definite statement about a population parameter which asserts that there is no real difference or effect: the population mean equals the claimed value, the two population means are equal, the treatment has done nothing. Whatever discrepancy the sample shows is attributed entirely to the chance variation of sampling. Three features of H_0 deserve emphasis. First, it is always framed as an equality (H_0: μ = 50 kg in part (b)), because the sampling distribution of the test statistic can only be worked out for one specified parameter value. Second, it is the hypothesis actually tested: the data are examined to see whether they are compatible with it. Third, the outcome regarding H_0 is asymmetric — it is either rejected, or not rejected ('retained for want of evidence against it'); a

test never proves H_0, just as failure to detect a disease in one examination does not prove perfect health. The phrase 'no difference' gives the hypothesis its name: null, that is, of no effect.

Alternative hypothesis (H_1). The alternative hypothesis, symbolised H_1, is the statement that is accepted when H_0 is rejected: there is a real difference, the parameter does not equal the claimed value. H_1 may be two-sided — H_1: $\mu \neq 50$, difference in either direction — which is appropriate when, as in part (b), the question is simply whether the machine is producing bags 'with the claimed average weight', so that overfilling would falsify the claim just as underfilling would. Or H_1 may be one-sided — H_1: $\mu < 50$ (systematic underfilling, the consumer's concern) or H_1: $\mu > 50$ — appropriate only when the enquiry is genuinely directional and fixed in advance. The choice is consequential: a two-sided test divides the significance level between the two tails of the distribution (2½% each at the 5% level, giving the critical values ±1.96), whereas a one-sided test concentrates it in one tail (critical value 1.645). Choosing a one-sided alternative after seeing the data, in order to reach significance, is a misuse of the method.

Errors of the test. Every test balances two possible mistakes. Rejecting H_0 when it is true is the Type I error; the investigator guards against it by fixing its probability, the level of significance α, in advance — customarily 5%, or 1% when a false rejection would be costly. Failing to reject H_0 when it is false is the Type II error, of probability β; $1 - \beta$, the probability of detecting a real difference of a given size, is the power of the test, which rises with sample size and falls as α is made smaller. With these ideas in place, the standard procedure, due essentially to Neyman and Pearson, runs in the following steps.

- **1. Formulate the hypotheses:** write H_0 as an equality about the parameter and H_1 in the two-sided or one-sided form the problem demands.

- **2. Fix the level of significance:** choose α (here 5%) before computing anything.
- **3. Select the test statistic and its distribution:** z for a mean when σ is known (or n is large), t when σ is estimated from the sample, χ^2 for a variance or for independence in a contingency table, F for several means at once; record the degrees of freedom where the distribution needs them.
- **4. State the decision rule:** obtain the critical value from the tables at the chosen α and identify the critical (rejection) region; or plan to compare the p-value — the probability, under H_0, of a result at least as extreme as the one observed — with α.
- **5. Compute the statistic** from the sample, using the value asserted by H_0.
- **6. Decide and interpret:** reject H_0 if the statistic lies in the critical region (equivalently $p < \alpha$), otherwise do not reject it; then translate the statistical decision back into the subject-matter conclusion — what it means for the machine, the fertiliser, or the method — including the direction of any difference found.

Part (b): The Rice-Bag Claim Worked Through (z-Test)

Data: $\mu_0 = 50$ kg (claim), $n = 49$ bags, $\bar{x} = 48.5$ kg, $\sigma = 4$ kg known, $\alpha = 0.05$. Following the steps: (1) H_0: $\mu = 50$ kg; H_1: $\mu \neq 50$ kg (two-sided — the question asks whether the bags have the claimed weight, in either direction). (2) $\alpha = 5\%$. (3) Because σ is known, the statistic $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$ is standard normal under H_0. (4) Decision rule: reject H_0 if $|z| > 1.96$. (5) Computation: the standard error of the mean is $\sigma/\sqrt{n} = 4/7 = 0.5714$ kg; $z = (48.5 - 50)/0.5714 = -1.5/0.5714 = -2.63$ (more precisely -2.625). The p-value of this result is $2 \times P(Z < -2.625) \approx 0.0087$. (6) Decision: $|z| = 2.625 > 1.96$, and $p \approx 0.0087 < 0.05$, so H_0 is rejected at the 5% level of significance — and since $|z| = 2.625$ also exceeds the 1% two-sided critical value 2.576 (equivalently, the p-value $0.0087 < 0.01$), the claim is rejected at the 1% level as well.

Interpretation in plain terms: if the machine truly delivered bags averaging 50 kg, a random sample of 49 bags would show a mean as far from 50 as 48.5 kg (in either direction) less than once in a hundred samples. That is too rare to credit to luck at the 5% standard the test set itself, so the claim is rejected: the machine is not producing bags with the claimed average weight. The direction of the discrepancy is downward — the bags average about 1.5 kg (3%) light — so the practical recommendation is inspection and recalibration of the filling machine, and the manufacturer should note that sustained underfilling of this size is also a consumer-protection and legal-metrology problem, not merely a statistical one. Finally, the logic should be stated correctly in the examination: we have not 'proved' that $\mu = 48.5$ (the sample mean is only an estimate, with standard error 0.57 kg; a 95% confidence interval for μ is $48.5 \pm 1.96 \times 0.5714 = 48.5 \pm 1.12$, i.e. 47.38 to 49.62 kg — note that 50 lies outside it, which is exactly the same conclusion as the test). What the test establishes is that the claimed value 50 kg is incompatible with the evidence at the 5% level.

USMAN NOTES www.usmannotes.com

Conclusion

Hypothesis testing adjudicates between a null hypothesis of no difference (stated as an equality and tested directly) and an alternative hypothesis (one- or two-sided), through the fixed sequence: hypotheses, significance level, statistic, decision rule, computation, interpretation. For the rice bags, $z = -2.625$ crosses both the 5% and 1% two-sided critical values, so the manufacturer's 50 kg claim is rejected and the machine is judged to be underfilling.

Question 2: Analysis of a 5 × 5 Latin Square — Five Fertilisers (20 Marks)

Question as printed: Five fertilizers, A, B, C, D, and E are tested in a Latin Square Design. The layout and yields are as given in the booklet's table (rows and columns 1–5; e.g. row 1: B 5.6, D

5.2, E 3.3, A 9.3, C 13.4; the full layout is reproduced in the solution below). Analyse the data for evidence at the 5% level of significance that the mean yields are not all equal for the five fertilizers. Also, pick out the best fertilizer.

Introduction

Field experiments are plagued by soil heterogeneity: fertility often changes across a field in one direction (down a slope) and in another (along an old watercourse) at the same time. The Latin square design is the classical answer when variation must be controlled in two directions simultaneously. The five treatments each appear exactly once in every row and exactly once in every column, so row differences and column differences can be estimated and removed from the experimental error, leaving a cleaner comparison of the treatments than either a completely randomised design or a one-way blocked design could give on the same land. This question presents a 5 × 5 Latin square comparing five fertilisers, A–E, with plot yields recorded in each cell. The task is a complete analysis of variance for the Latin square, an F-test at the 5% level of the hypothesis that all five fertiliser means are equal, and — because a significant F by itself does not say which fertiliser to recommend — an identification, supported by a critical difference, of the best fertiliser.

The Layout and the Data

Rows \ Columns	**1**	**2**	**3**	**4**	**5**	**Row total**
1	B 5.6	D 5.2	E 3.3	A 9.3	C 13.4	36.8
2	C 6.8	A 3.0	B 6.0	E 5.0	D 5.0	25.8
3	D 8.2	C 15.2	A 8.4	B 6.1	E 4.6	42.5
4	E 5.2	B 7.2	C 14.1	D 8.4	A 6.9	41.8
5	A 9.2	E 6.3	D 11.8	C 15.6	B 7.6	50.5
Column total	35.0	36.9	43.6	44.4	37.5	G = 197.4

The reader may verify the Latin property from the table: each letter A–E occurs once in every row and once in every column. Treatment totals (each over 5 plots) are obtained by collecting the yields of each letter wherever it stands: A = 9.3 + 3.0 + 8.4 + 6.9 + 9.2 = 36.8; B = 5.6 + 6.0 + 6.1 + 7.2 + 7.6 = 32.5; C = 13.4 + 6.8 + 15.2 + 14.1 + 15.6 = 65.1; D = 5.2 + 5.0 + 8.2 + 8.4 + 11.8 = 38.6; E = 3.3 + 5.0 + 4.6 + 5.2 + 6.3 = 24.4. Check: 36.8 + 32.5 + 65.1 + 38.6 + 24.4 = 197.4 = G, agreeing with the sum of the row totals (36.8 + 25.8 + 42.5 + 41.8 + 50.5 = 197.4) and of the column totals (35.0 + 36.9 + 43.6 + 44.4 + 37.5 = 197.4). The corresponding treatment means are A 7.36, B 6.50, C 13.02, D 7.72, E 4.88 (overall mean 197.4/25 = 7.896). Already the eye sees that C stands far above the rest — its five yields (13.4, 6.8, 15.2, 14.1, 15.6) include the four largest values in the experiment — but the analysis of variance is needed to judge whether differences of this size exceed what plot-to-plot error could produce.

Analysis of Variance

With N = 25 observations, the correction factor is CF = G^2/N = $(197.4)^2/25$ = 38,966.76/25 = 1,558.6704. The total sum of squares is Σy^2 − CF: the sum of the squares of all 25 yields is 1,864.34, so Total SS = 1,864.34 − 1,558.67 = 305.67. Rows SS = (Σ row totals^2)/5 − CF = ($36.8^2 + 25.8^2 + 42.5^2 + 41.8^2 + 50.5^2$)/5 − CF = 8,123.62/5 − 1,558.67 = 1,624.72 − 1,558.67 = 66.05. Columns SS = ($35.0^2 + 36.9^2 + 43.6^2 + 44.4^2 + 37.5^2$)/5 − CF = 7,865.18/5 − 1,558.67 = 1,573.04 − 1,558.67 = 14.37. Treatments (fertilisers) SS = ($36.8^2 + 32.5^2 + 65.1^2 + 38.6^2 + 24.4^2$)/5 − CF = 8,733.82/5 − 1,558.67 = 1,746.76 − 1,558.67 = 188.09. Error SS follows by subtraction: 305.67 − 66.05 − 14.37 − 188.09 = 37.16. Degrees of freedom: total 24; rows, columns and treatments 4 each; error (5−1)(5−2) = 12. The ANOVA table is:

Source	d.f.	SS	MS	F	F (5%)
Rows	4	66.05	16.51	5.33	3.26
Columns	4	14.37	3.59	1.16	3.26

Fertilisers (treatments)	4	188.09	47.02	15.19	3.26
Error	12	37.16	3.096	—	—
Total	24	305.67	—	—	—

Test and Identification of the Best Fertiliser

For the fertilisers, F = 47.02/3.096 = 15.19, which far exceeds the tabulated $F_{0.05}(4, 12) = 3.26$. The null hypothesis that the five fertiliser mean yields are all equal is therefore rejected at the 5% level (indeed at the 1% level as well, the 1% point being 5.41): there is strong evidence that the mean yields are not all equal — the fertilisers genuinely differ in their effect on yield. The blocking was also worth doing in the row direction (Rows F = 5.33 > 3.26, significant: fertility varied appreciably from row to row, and the design has removed that variation from the error), while column differences are not significant (F = 1.16), though their removal costs nothing in interpretation and the design remains valid.

To pick out the best fertiliser, compare the treatment means using the critical difference (least significant difference) at 5%: $CD = t_{0.05}(12) \times \sqrt{(2 \times MSE/5)} = 2.179 \times \sqrt{(2 \times 3.096/5)} = 2.179 \times 1.113 = 2.42$. Any two means differing by more than 2.42 differ significantly. Fertiliser C's mean, 13.02, exceeds the next best, D (7.72), by 5.30 — more than twice the critical difference — and exceeds A (7.36), B (6.50) and E (4.88) by still larger margins. C is therefore significantly better than every other fertiliser tested. Among the remainder, D, A and B (7.72, 7.36, 6.50) do not differ significantly from one another (gaps below 2.42 within the group), B and E differ by only 1.62 (not significant), though D and A do significantly exceed E (gaps 2.84 and 2.48). The recommendation is unambiguous: fertiliser C is the best of the five, with a mean yield of 13.02 units against an overall mean of 7.90 — roughly 65% above average and about 69% above the second-placed fertiliser D.

Conclusion

The Latin square analysis separates row, column and treatment variation cleanly. The fertiliser F-ratio of 15.19 on (4, 12) degrees of freedom is far beyond the 5% critical value 3.26, so the hypothesis of equal fertiliser means is decisively rejected; and the comparison of means against the critical difference 2.42 identifies fertiliser C (mean yield 13.02) as significantly the best. The significant row effect confirms that the double blocking of the Latin square earned its keep.

Question 3: Factorial Designs — Merits, Demerits and Uses; Analysis of a Three-Factor Factorial Experiment (20 Marks)

Question as printed: Elaborate on the merits, demerits, and uses of factorial designs. How would you analyse the data by using a three-factor factorial experiment?

Introduction

The one-factor-at-a-time method — vary nitrogen while holding everything else fixed, then vary seed rate, then spacing — is the intuitive way to experiment, and it is statistically wasteful and sometimes actively misleading. It is wasteful because each run informs us about only one factor; it is misleading because factors in nature interact: the best level of nitrogen depends on the seed rate, and a method that never varies two factors together can never discover this. Factorial designs, developed by R. A. Fisher and F. Yates at Rothamsted, meet both objections: every factor is varied at two or more levels, all combinations of levels are run, and from the same set of plots the experimenter estimates every main effect and every interaction. This answer first sets out what a factorial design is, then weighs its merits and demerits honestly, lists its main uses, and finally describes in operational detail how the data of a three-factor factorial experiment are analysed — the same machinery that Question 4 then applies to real numbers.

What a Factorial Design Is

A factorial experiment is one in which the treatments are all the combinations of the levels of two or more factors. If factor A has a levels, B has b levels and C has c levels, the design has a × b × c treatment combinations; with two levels each it is a 2^3 design with 8 combinations, written in Yates' notation as (1), a, b, ab, c, ac, bc, abc, where a letter present means that factor is at its high level and (1) means all factors at their low levels. The effect of a factor averaged over all levels of the others is its main effect; the failure of one factor's effect to be the same at all levels of another is an interaction (two-factor, e.g. AB; three-factor, ABC). The design may be laid out in randomised blocks, in a Latin square, or completely randomised, and it may be replicated; in large designs, confounding and fractional replication are used to keep block sizes manageable, at the price of sacrificing information on selected (usually high-order) interactions.

Merits of Factorial Designs

- **Efficiency (hidden replication):** every observation serves in the estimation of every effect. In a 2^3 with r replicates, each main effect is a difference between two means each based on 4r observations, although only 8r plots exist — the experiment does the work of several one-factor experiments for the plots of one.
- **Interactions are estimable:** the design alone reveals whether factors act independently or modify each other's effects — often the most valuable scientific finding, e.g. that a fertiliser responds only in the presence of irrigation.
- **Wider inductive basis:** conclusions about a factor are drawn across the range of conditions of the other factors, so they generalise better than conclusions drawn at one fixed background.

- **No wasted runs:** even if some factor turns out to have no effect, its plots still contribute fully to estimating the other effects and the error; in one-factor-at-a-time work, a redundant factor wastes its whole experiment.
- **Optimality:** for estimating all effects with given precision, the factorial arrangement in orthogonal designs gives each effect the same, minimal variance, and the effects are estimated independently of one another.

Demerits of Factorial Designs

- **Size grows explosively:** the number of combinations multiplies — $2^6 = 64$, $3^4 = 81$ — so with several factors or levels the experiment can become too large for the land, material or budget available, and large blocks reintroduce the heterogeneity blocking was meant to remove (met by confounding or fractional replication, which sacrifice some information).
- **Complexity of execution:** many treatment combinations are harder to apply correctly in the field; a single labelling or application mistake can corrupt several effect estimates.
- **Interpretation becomes difficult:** when high-order interactions are significant, main effects lose their simple meaning, and explaining a three-factor interaction to a non-statistician is genuinely hard.
- **Some combinations may be useless or impossible:** certain level combinations may be agronomically absurd, physically impossible, or of no practical interest, yet the full factorial insists on running them.
- **Assumptions multiply:** the standard analysis presumes a common error variance across all combinations; with very diverse treatments this may fail, and transformations or special analyses are then needed.

Uses of Factorial Designs

Factorial designs are used wherever several inputs jointly determine a response: in agriculture, the classical N × P × K fertiliser trials, variety × spacing × sowing-date trials and irrigation × fertiliser studies; in forestry, species × spacing × thinning and nursery trials combining seed treatment, sowing medium and watering regime; in industry, optimising processes whose yield depends on temperature, pressure and catalyst together; in medicine, testing combinations of drugs or of drug and dosage; in animal husbandry, feed × breed × housing studies; and in social and extension research, combining training method, media and follow-up. In each case the decisive advantage is the same: interactions, which one-factor experiments are structurally blind to, are exactly what real systems exhibit.

Analysing a Three-Factor Factorial Experiment

Suppose factors A, B, C are each at two levels, giving the 8 combinations (1), a, b, ab, c, ac, bc, abc, laid out in r randomised blocks (replications), 8r plots in all. The analysis proceeds as follows. First, total the yield of each combination over the r replications, giving 8 treatment totals. Second, obtain the seven effect contrasts from these totals, most conveniently by Yates' algorithm: write the totals in standard order, then repeatedly form pair-sums and pair-differences (three cycles for three factors); the final column gives, in order, the grand total and the contrasts for A, B, AB, C, AC, BC, ABC. Each effect (the average effect of the factor or interaction) equals its contrast divided by $r \times 2^2 = 4r$, and each sum of squares equals $(\text{contrast})^2/(r \times 2^3) = (\text{contrast})^2/8r$, with 1 degree of freedom. Third, compute the ordinary block (replication) sum of squares from the block totals and the total sum of squares from all 8r observations, and obtain the error sum of squares by subtraction: Total SS − Blocks SS − (A + B + AB + C + AC + BC + ABC) SS, with $8r - 1 - (r - 1) - 7 = 7(r - 1)$ degrees of freedom. Fourth, form the ANOVA

table with mean squares and test each effect by F = MS(effect)/MS(error) against $F_{0.05}(1, 7(r-1))$ (equivalently a t-test, since each effect has 1 d.f. and $F = t^2$).

Source	d.f.	How obtained
Replications (blocks)	r − 1	From block totals
A, B, C (main effects)	1 each	Contrast² / 8r (Yates)
AB, AC, BC (two-factor interactions)	1 each	Contrast² / 8r (Yates)
ABC (three-factor interaction)	1	Contrast² / 8r (Yates)
Error	7(r − 1)	By subtraction
Total	8r − 1	From all observations

Interpretation follows a fixed etiquette. Examine the interactions first: if AB is significant, the effect of A depends on the level of B, and the two factors must be discussed together from their two-way table of means; the main effects of A and B are then reported only as averages, with the qualification stated. If the interactions are negligible, the main effects are interpreted directly, each as an average over all combinations of the other factors — this pooling of all the data into every comparison is the 'hidden replication' that makes the design so efficient. Significant effects are quantified by their effect estimates and standard error $\sqrt{(MSE/2r)}$ for an effect … more precisely the standard error of an effect is $\sqrt{(MSE \times 4/(8r))} = \sqrt{(MSE/(2r))}$; treatment means can also be compared with a critical difference based on the error mean square. Where some high-order interaction is assumed negligible a priori, its sum of squares is sometimes pooled with error to gain degrees of freedom — a legitimate economy, provided the assumption is declared. For factors at more than two levels the same structure generalises: each main effect carries (levels − 1) degrees of freedom and interactions the products of these, computed from the corresponding marginal tables instead of Yates' algorithm, which is specific to the 2^n series.

Conclusion

Factorial designs trade size and complexity for a decisive gain: all main effects and all interactions estimated efficiently from one experiment, with every plot contributing to every comparison. Their demerits — combinatorial growth, harder management and interpretation — are real but manageable through confounding and fractional replication. The analysis of the three-factor case, by Yates' algorithm and the analysis of variance, partitions the treatment variation into seven single-degree-of-freedom effects and tests each against the experimental error, as the next question demonstrates numerically.

Question 4: Analysis of Variance of a 2^3 Factorial Experiment in Three Replications (20 Marks)

Question as printed: The following data were obtained from a 2^3-factorial experiment repeated three times. Replication 1: (1) 12, a 15, b 24, ab 23, c 17, ac 16, bc 24, abc 28. Replication 2: (1) 19, a 20, b 16, ab 17, c 25, ac 19, bc 23, abc 25. Replication 3: (1) 10, a 16, b 17, ab 27, c 21, ac 19, bc 29, abc 20. Set up an analysis of variance to test the significance of main effects and interactions.

Introduction

Question 3 described the analysis of a three-factor factorial experiment in principle; this question requires it in practice. Three factors, A, B and C, each at two levels, give $2^3 = 8$ treatment combinations, and the whole set is repeated in three replications, giving 24 observations. The analysis must partition the total variation into replications, the three main effects A, B, C, the three two-factor interactions AB, AC, BC, the three-factor interaction ABC, and experimental error, and must test each of the seven factorial effects against the error. Because each effect has a

single degree of freedom, the whole treatment structure is extracted from eight treatment totals by Yates' algorithm. The workings below display every step — totals, Yates' table, sums of squares, the ANOVA table and the F-tests — so that the arithmetic can be followed and checked, and the interpretation at the end states precisely which factors matter and which do not.

Treatment Totals and Yates' Algorithm

Treatment combination	Rep 1	Rep 2	Rep 3	Total
(1)	12	19	10	41
a	15	20	16	51
b	24	16	17	57
ab	23	17	27	67
c	17	25	21	63
ac	16	19	19	54
bc	24	23	29	76
abc	28	25	20	73
Replication total	159	164	159	G = 482

Yates' algorithm on the totals in standard order (pair-sums down the upper half, pair-differences down the lower half, three cycles) yields the contrasts: grand total 482; A = 8; B = 64; AB = 6; C = 50; AC = −32; BC = 0; ABC = 6. Each effect = contrast/(r × 2²) = contrast/12, and each sum of squares = (contrast)²/(r × 2³) = (contrast)²/24. Thus: A effect = 8/12 = 0.67, SS = 64/24 = 2.67; B effect = 64/12 = 5.33, SS = 4,096/24 = 170.67; AB effect = 6/12 = 0.50, SS = 36/24 = 1.50; C effect = 50/12 = 4.17, SS = 2,500/24 = 104.17; AC effect = −32/12 = −2.67, SS = 1,024/24 = 42.67; BC effect = 0/12 = 0.00, SS = 0.00 (the BC contrast is exactly zero — B behaves identically at both levels of C in these totals); ABC effect = 6/12 = 0.50, SS = 36/24 = 1.50. As a check, the seven effect sums of squares add to 2.67 + 170.67 + 1.50 + 104.17 + 42.67 + 0.00 + 1.50 = 323.17, which equals the treatment sum of squares computed independently below.

The Analysis of Variance

Grand total G = 482 over N = 24 observations; correction factor CF = 482²/24 = 232,324/24 = 9,680.17. The sum of squares of all 24 observations is 10,242, so Total SS = 10,242 − 9,680.17 = 561.83. Replication SS: the replication totals are 159, 164, 159, so Rep SS = (159² + 164² + 159²)/8 − CF = 77,458/8 − 9,680.17 = 9,682.25 − 9,680.17 = 2.08. Treatment SS = (Σ treatment totals²)/3 − CF = (41² + 51² + 57² + 67² + 63² + 54² + 76² + 73²)/3 − CF = 30,010/3 − 9,680.17 = 10,003.33 − 9,680.17 = 323.17, agreeing with the sum of the Yates components, as it must. Error SS = 561.83 − 2.08 − 323.17 = 236.58, with 7(r − 1) = 14 degrees of freedom, giving the error mean square 236.58/14 = 16.90. The tabulated F at 5% for (1, 14) degrees of freedom is 4.60 (at 1% it is 8.86).

Source	d.f.	SS	MS	F	Verdict at 5%
Replications	2	2.08	1.04	—	—
A	1	2.67	2.67	0.16	Not significant
B	1	170.67	170.67	10.10	Significant (also at 1%)
AB	1	1.50	1.50	0.09	Not significant
C	1	104.17	104.17	6.16	Significant
AC	1	42.67	42.67	2.53	Not significant
BC	1	0.00	0.00	0.00	Not significant
ABC	1	1.50	1.50	0.09	Not significant
Error	14	236.58	16.90	—	—
Total	23	561.83	—	—	—

Interpretation

Two effects, and only two, are significant. The main effect of B is highly significant (F = 10.10, exceeding even the 1% point 8.86): raising B from its low to its high level increases the response by an average of 5.33 units, averaged over all combinations of A and C. The main effect of C is significant at the 5% level (F = 6.16 > 4.60): raising C increases the response by an average of 4.17 units. Factor A has no detectable effect (effect 0.67, F = 0.16), and none of the interactions

approaches significance — AB (0.09), AC (2.53, the largest of them, its effect of −2.67 suggesting at most a hint that A's trivial effect reverses between levels of C), BC (exactly 0.00) and ABC (0.09). Because the interactions are negligible, the main effects may be interpreted directly and additively: the predicted response for any combination is the grand mean 482/24 = 20.08, plus half the B effect (±2.67) if B is high/low, plus half the C effect (±2.08) if C is high/low, with A contributing nothing worth counting. The best combination on this evidence is therefore (high B, high C) with A at either level — the bc and abc treatments, whose observed means are indeed the two largest but one (76/3 = 25.33 for bc and 73/3 = 24.33 for abc, statistically indistinguishable). The replications contribute almost nothing (SS = 2.08), indicating a homogeneous experimental area. Note also the economy of the design: 24 plots have separately tested seven effects, each main effect being based on all 24 observations (12 at each level) — the hidden replication described in Question 3.

USMAN NOTES www.usmannotes.com

Conclusion

The analysis of variance of the 2^3 experiment shows a clear-cut result: factors B (F = 10.10) and C (F = 6.16) significantly increase the response, by 5.33 and 4.17 units respectively; factor A and all two- and three-factor interactions are non-significant. The factors act independently and additively, and the recommended combination is high levels of B and C, with A immaterial.

Question 5: The Covariance Model and Its Assumptions; Analysis of Covariance for Grain and Straw Yields (5 + 15 Marks)

Question as printed: (a) Explain the covariance model and its assumptions. (b) Each of 4 blocks was divided into 3 plots, and 3 different treatments A, B, and C were distributed at random among the plots of each block. The rows correspond to blocks, and the yields of the grain and

straw were denoted by X and Y respectively. Examine the covariance between the yields of grain and straw to show whether the yield of straw after correction for the yield of grain varies significantly with treatment. Data (X = grain, Y = straw): Block 1: A (65, 32), B (75, 38), C (72, 33); Block 2: A (68, 26), B (54, 20), C (69, 30); Block 3: A (71, 26), B (71, 28), C (69, 38); Block 4: A (62, 33), B (64, 29), C (61, 30).

Introduction

In many experiments the response we care about is dragged up and down by a second quantity that the treatments do not control but which we can measure on every plot: the final yield of a crop by the initial soil fertility or the plant stand, the weight gain of an animal by its initial weight, the straw yield of a cereal plot by its grain yield. If that accompanying variable — the covariate — is ignored, its variation inflates the experimental error and can even masquerade as a treatment effect, because randomisation never balances it exactly. The analysis of covariance (ANCOVA), Fisher's elegant combination of the analysis of variance with regression, removes the part of the response that is linearly predictable from the covariate and then compares the treatments on what remains. Part (a) states the covariance model and its assumptions; part (b) carries out the full analysis for a randomised block experiment with three treatments in four blocks, where straw yield Y is to be compared between treatments after correction for the concomitant grain yield X.

Part (a): The Covariance Model and Its Assumptions

For a randomised block design with t treatments, b blocks and one covariate, the covariance model is $Y_{ij} = \mu + \tau_i + \beta_j + \gamma(X_{ij} - \bar{X}) + \varepsilon_{ij}$, where Y_{ij} is the response of the plot receiving treatment i in block j, μ is the general mean, τ_i the treatment effect, β_j the block effect, X_{ij} the covariate measured on the same plot, $\bar{X}$ the overall mean of the covariate, γ the common

regression coefficient of Y on X, and ε_{ij} the random error. The term $\gamma(X_{ij} - \bar{X})$ is the adjustment: each observation is corrected to what it would be expected to be had its plot carried the average value of the covariate. Treatment means adjusted in this way, $\bar{Y}_i(adj) = \bar{Y}_i - \hat{\gamma}(\bar{X}_i - \bar{X})$, are then compared, and the test of treatments uses an F-ratio whose error mean square has been reduced by the regression on X, at the cost of one degree of freedom for estimating γ. The model rests on the following assumptions.

- **The covariate is unaffected by the treatments:** X must not itself be changed by the treatments; if it is (e.g. adjusting final weight for a mid-trial weight that the feeding treatments influenced), the adjustment removes part of the very effect under study and the analysis is invalid. Ideally X is measured before the treatments act.
- **Linearity:** the regression of Y on X within treatment groups is adequately described by a straight line.
- **Homogeneity of regression:** the slope γ is the same for all treatments (parallel regression lines); if slopes differ, a single adjustment is meaningless and separate slopes or a different analysis are required.
- **The usual ANOVA assumptions hold for the errors:** ε_{ij} are independent, normally distributed, with mean zero and constant variance σ^2.
- **The covariate is measured without (appreciable) error,** or at least with error small relative to its range; errors in X bias the estimated slope towards zero and under-adjust the means.
- **Independence of covariate and treatment allocation:** secured by randomisation, so that differences in X between treatment groups are purely random and safely adjustable.

When these assumptions hold, ANCOVA delivers two benefits: greater precision (the error variance shrinks from σ^2 to roughly $\sigma^2(1 - \rho^2)$, where ρ is the within-group correlation of X and

Y) and fairer comparisons (treatment means corrected to a common covariate level). Its price is one error degree of freedom and the loss of validity altogether if the treatments have themselves moved the covariate.

Part (b): The Analysis of Covariance Worked Out

The data consist of 12 plots: 4 blocks × 3 treatments, with X = grain yield and Y = straw yield on each plot. Grand totals: ΣX = 801 ($\bar{X}$ = 66.75) and ΣY = 363 ($\bar{Y}$ = 30.25). Following the standard ANCOVA scheme, sums of squares and sums of products are computed for X, for Y, and for the cross-products XY, each partitioned into blocks, treatments and error exactly as in a two-way analysis of variance, using correction terms $CF_x = 801^2/12 = 53{,}466.75$, $CF_y = 363^2/12 = 10{,}980.75$ and $CF_{xy} = 801 \times 363/12 = 24{,}230.25$.

Source	d.f.	SS of X (Sxx)	Sum of products (Sxy)	SS of Y (Syy)
Blocks	3	171.58	92.42	123.58
Treatments	2	6.50	15.50	38.00
Error	6	194.17	73.83	124.67
Total	11	372.25	181.75	286.25

From the error line, the within-experiment regression coefficient of straw on grain is b = Sxy(error)/Sxx(error) = 73.83/194.17 = 0.3803: within blocks and treatments, a plot yielding one unit more grain yields on average about 0.38 units more straw — a sensible positive association for two products of the same crop. The error sum of squares for Y, adjusted for this regression, is Syy(error) − [Sxy(error)]²/Sxx(error) = 124.67 − (73.83)²/194.17 = 124.67 − 28.07 = 96.59, on 6 − 1 = 5 degrees of freedom (one lost for the regression), giving the adjusted error mean square 96.59/5 = 19.32. The test of the regression itself gives F = 28.07/19.32 = 1.45 on (1, 5) d.f. (critical value 6.61), so in this small experiment the regression, though positive and of plausible

size, is not itself statistically significant — the covariate removes only a modest amount of error (124.67 → 96.59). Next, pool treatments with error: Sxx = 6.50 + 194.17 = 200.67, Sxy = 15.50 + 73.83 = 89.33, Syy = 38.00 + 124.67 = 162.67, whose adjusted sum of squares is $162.67 - (89.33)^2/200.67 = 162.67 - 39.77 = 122.90$ on 8 − 1 = 7 degrees of freedom. The adjusted treatment sum of squares is the difference, 122.90 − 96.59 = 26.31 on 2 degrees of freedom, mean square 13.15. The adjusted ANOVA is therefore:

Source	d.f.	Adjusted SS	MS	F
Treatments (adjusted for X)	2	26.31	13.15	0.68
Error (adjusted)	5	96.59	19.32	—
Total (adjusted)	7	122.90	—	—

The F-ratio for treatments is F = 13.15/19.32 = 0.68, far below the tabulated $F_{0.05}(2, 5) = 5.79$. The adjusted treatment means tell the same story. The treatment means of X are A 66.50, B 66.00, C 67.75, and of Y are A 29.25, B 28.75, C 32.75; adjusting each Y-mean to the common grain level $\bar{X} = 66.75$ by $\bar{Y}(adj) = \bar{Y} - 0.3803(\bar{X}_i - 66.75)$ gives A: 29.25 − 0.3803(−0.25) = 29.35, B: 28.75 − 0.3803(−0.75) = 29.04, C: 32.75 − 0.3803(1.00) = 32.37. Treatment C's straw advantage survives the correction — it is not explained away by C's slightly higher grain yields — but with only four blocks the experiment is too imprecise to establish it: the adjusted means 29.35, 29.04 and 32.37 are well within sampling variation of one another, as the F-test confirms. (For reference, the ordinary unadjusted analysis of Y alone gives treatment F = (38.00/2)/(124.67/6) = 0.91, also non-significant; covariance adjustment reduced the error mean square from 20.78 to 19.32 but simultaneously changed the treatment sum of squares, and the conclusion is unchanged.)

Conclusion

Examining the covariance between grain and straw yields: the error regression of straw on grain is b = 0.380 (positive, as expected biologically), and after every plot's straw yield is corrected for its grain yield, the adjusted treatment sum of squares is 26.31 on 2 degrees of freedom against an adjusted error mean square of 19.32, giving F = 0.68, which is not significant at the 5% level (critical value 5.79). Hence the yield of straw, after correction for the yield of grain, does NOT vary significantly with treatment: on this evidence, treatments A, B and C produce the same straw yield once their grain yields are allowed for, C's higher raw straw mean (32.75) notwithstanding.

USMAN NOTES www.usmannotes.com