Chapter 8

Analysis of variance

Exercises

Contents

8 Analysis of variance

Exercises **1**
8.1 Environment action plans 3
8.2 Environment action plans (part 2) 7
8.3 Plastic film 12
8.4 Brass alloys 14
8.5 Plastic tubes 17
8.6 Joining methods 21
8.7 Remoulade 26
8.8 Transport times 29

8.1 Environment action plans

|||| Exercise 8.1 Environment action plans

To assess whether the two Danish aquatic environment action plans affected nitrogen levels in a river, nitrogen concentration (measured in $\mathrm{g/m^3}$) was recorded at three time points:

- 1998: immediately before the first action plan,
- 2003: immediately before the second action plan,
- 2011: after both action plans had been implemented.

At each time point, the nitrogen concentration was measured six times. The observed concentrations are shown below:

	N_{1998}	N_{2003}	N_{2011}
	5.01	5.59	3.02
	6.23	5.13	4.76
	5.98	5.33	3.46
	5.31	4.65	4.12
	5.13	5.52	4.51
	5.65	4.92	4.42
Mean	5.552	5.190	4.048

The total sum of squares is given by: $SST = 11.4944$.

A one-way ANOVA was performed to compare the mean nitrogen concentrations across the three years. The resulting ANOVA table is shown below, where some values have been substituted with letters:

```
fit = smf.ols('nitrogen ~ year', data=D).fit()
print(sm.stats.anova_lm(fit))

            df   sums_q  mean_sq          F    PR(>F)
Year         A        B        C          U         V
Residuals    D   4.1060        E        NaN       NaN
```

a) Determine the values represented by the letters A–D.

Solution

With $k = 3$ groups and $n = 18$ observations:

$$A = k - 1 = 2, \quad D = n - k = 15.$$

$$B = SS(Year) = SST - SSE = 11.4944 - 4.1060 = 7.3884,$$

$$C = MS(Year) = \frac{SS(Year)}{k-1} = B/2 = 3.6942.$$

b) At $\alpha = 0.05$, what critical value is used for the hypothesis test (represented by the letters U and V)?

Solution

The F-test uses the $F(k-1,\ n-k) = F(2, 15)$ distribution.

$$F_{0.05}(2, 15) = 3.682.$$

```
print(stats.f.ppf(0.95, dfn=2, dfd=15))
3.682
```

c) Using a significance level of $\alpha = 0.05$, test whether the mean nitrogen concentration differs between years. State the null hypothesis, calculate the test statistic and p-value, and draw an appropriate conclusion.

Solution

To test whether the mean nitrogen concentration is the same in all three years, we use the ANOVA test statistic

$$F_{\text{obs}} = \frac{\text{MS(Year)}}{\text{MSE}} = \frac{7.3884/2}{4.1060/15} = 13.50.$$

Under the null hypothesis

$$H_0 : \mu_{1998} = \mu_{2003} = \mu_{2011},$$

the test statistic follows an F distribution with $(2, 15)$ degrees of freedom. The corresponding p-value is

$$p = P(F_{2,15} > 13.50) = 0.00044.$$

```
Fobs = (7.3884/2) / (4.1060/15)
print(1 - stats.f.cdf(Fobs, dfn=2, dfd=15))
```

Since

$$p = 0.00044 < \alpha$$

we reject H_0. There is strong evidence that the mean nitrogen concentration is not the same in all three years. Thus, the data indicate statistically significant differences in mean nitrogen concentration between years.

d) Compute the 90% confidence interval for the mean difference between 2011 and 1998.

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

|||| Solution

Use Method 8.9:

$$\mu_{2011} - \mu_{1998} = 4.048 - 5.552 = -1.5032011, \quad MSE = \frac{4.1060}{15} = 0.2737$$

Degrees of freedom: $n - k = 15$.

$$\begin{aligned} & -1.503 \pm t_{0.05,15}\sqrt{MSE\left(\tfrac{1}{6} + \tfrac{1}{6}\right)} \\ =& -1.503 \pm 1.753\sqrt{0.2737 \times \tfrac{1}{3}} \\ =& [-1.503 \pm 0.528] \\ =& [-2.03;\ -0.97] \end{aligned}$$

```
diff = 4.048 - 5.552
ME = stats.t.ppf(0.95, 15) * np.sqrt(4.1060/15 * (1/3))
print(diff - ME, diff + ME)
-2.03 -0.97
```

Exam-style questions

e) A different one-way ANOVA with $k = 4$ groups and $n = 24$ total observations gives $SST = 85.6$ and $SSE = 52.3$. What are $SS(Tr)$, $MS(Tr)$, and MSE?

1) $SS(Tr) = 137.9, MS(Tr) = 45.97, MSE = 2.615$
2) $SS(Tr) = 33.3, MS(Tr) = 11.1, MSE = 2.615$
3) $SS(Tr) = 33.3, MS(Tr) = 8.325, MSE = 2.615$
4) $SS(Tr) = 33.3, MS(Tr) = 11.1, MSE = 2.18$
5) $SS(Tr) = 33.3, MS(Tr) = 11.1, MSE = 17.43$

|||| Solution

Correct answer: **2)**. Using the same decomposition as in the first question of this exercise: $SS(Tr) = SST - SSE = 85.6 - 52.3 = 33.3$; $MS(Tr) = SS(Tr)/(k-1) = 33.3/3 = 11.1$; and $MSE = SSE/(n-k) = 52.3/20 = 2.615$.

8.2 Environment action plans (part 2)

|||| Exercise 8.2 Environment action plans (part 2)

This exercise uses the same nitrogen concentration data as Exercise 8.1 (measurements in g/m^3 at three time points: 1998, 2003, 2011; 6 replicates each).

The group variances are added to the table:

	N_{1998}	N_{2003}	N_{2011}
	5.01	5.59	3.02
	6.23	5.13	4.76
	5.98	5.33	3.46
	5.31	4.65	4.12
	5.13	5.52	4.51
	5.65	4.92	4.42
Mean	5.5517	5.1900	4.0483
Variance	0.2366	0.1313	0.4533

Further, the overall mean is: $\bar{y} = 4.93$.

a) Compute the three sums of squares (SST, $SS(Tr)$, and SSE) using the group means, group variances, and overall mean. Show the formulas explicitly.

|||| Solution

Using Theorem 8.2:

$$\begin{aligned} SS(Tr) &= \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 \\ &= 6(5.5517 - 4.93)^2 + 6(5.1900 - 4.93)^2 + 6(4.0483 - 4.93)^2 = 7.388 \end{aligned}$$

For SSE we note that:

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

and therefore we can compute:

$$SSE = 5 \times 0.2366 + 5 \times 0.1313 + 5 \times 0.4533 = 4.106$$

Finally for SST:

$$SST = SS(Tr) + SSE = 7.388 + 4.106 = 11.494$$

The raw data can be read into Python by the following code:

```
nitrogen = np.array([
5.01, 5.59, 3.02,
6.23, 5.13, 4.76,
5.98, 5.33, 3.46,
5.31, 4.65, 4.12,
5.13, 5.52, 4.51,
5.65, 4.92, 4.42
])
year = pd.Categorical(np.tile(["1998", "2003", "2011"], 6))
D = pd.DataFrame({"nitrogen": nitrogen, "year": year})
```

b) Find the SST value in Python using the sample variance function ("`np.var(ddof=1)`").

|||| Solution

From eqn. (8-10) we see that SST can be calculates from the sample variance of the data:

$$SST = (n-1)s_y^2$$

```
print(17 * D["nitrogen"].var(ddof=1))
11.494
```

c) Run the ANOVA in Python and produce the ANOVA table.

Solution

It may be done as follows:

```
fit = smf.ols("nitrogen ~ year", data=D).fit()
anova = sm.stats.anova_lm(fit)
print(anova)
             df      sum_sq     mean_sq             F      PR(>F)
year        2.0    7.388433     3.694217    13.495787    0.000444
Residual   15.0    4.105967     0.273731          NaN         NaN
```

d) Do a complete post hoc analysis, comparing all 3 years pairwise.

Solution

We want to make three pairwise comparisons: compare 1998 with 2003, 1998 with 2011 and 2003 with 2011 (number of comparisons $M = 3$). Since there is an equal number of observations in all three groups ($n_i = 6$), we can compare the groups by computing the "Least Significant Difference" (the LSD-value, see eqn. (8-28)). And since there are 3 multiple comparisons we will use $\alpha_{Bonferroni} = 0.05/3 = 0.01667$.

$$LSD_{0.01667} = t_{1-0.01667/2}(15)\sqrt{2 \times 0.2737/6}$$

For this we need the quantile $t_{10.01667/2}$ from a t-distribution with $nk = 15$ degrees of freedom:

```
print(stats.t.ppf(1-0.01667/2, 15))
2.6936395462653473
```

which gives

$$LSD_{0.01667} = 0.814$$

Comparing the three means (5.552, 5.190, 4.048):

- $|5.552 - 5.190| = 0.362 < 0.814$: 1998 and 2003 are *not* significantly different.
- $|5.552 - 4.048| = 1.504 > 0.814$: 1998 and 2011 *are* significantly different.
- $|5.190 - 4.048| = 1.142 > 0.814$: 2003 and 2011 *are* significantly different.

The nitrogen level in 2011 is significantly lower than in both 1998 and 2003.

e) Use Python to do model validation by residual analysis.

Solution

Check two plots:

1. Normal Q-Q plot of the residuals — should be approximately linear, indicating normality.
2. Plot residuals per group (scatterplot or boxplot) — should show no systematic pattern and approximately equal spread across groups.

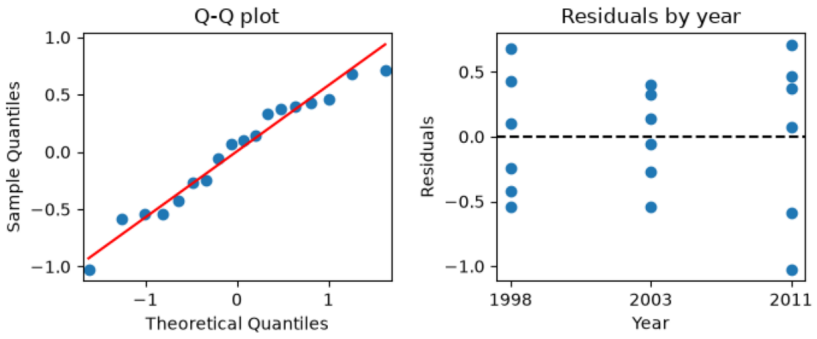
```
# Add residuals to dataframe
D["residuals"] = fit.resid

# Two diagnostic plots
fig, axes = plt.subplots(1, 2, figsize=(7, 3))

# Q-Q plot
sm.qqplot(fit.resid, line="q", ax=axes[0])
axes[0].set_title("Q-Q plot")

# Residuals by year
axes[1].plot(D["year"], D["residuals"], "o")
axes[1].axhline(0, color="black", linestyle="--")
axes[1].set_xlabel("Year")
axes[1].set_ylabel("Residuals")
axes[1].set_title("Residuals by year")

plt.tight_layout()
plt.show()
```



There is no important deviation from normality in the Q-Q plot. The scatter plot of residuals does not indicate obvious variance differences between years.

8.3 Plastic film

|||| Exercise 8.3 Plastic film

A company is starting a production of a new type of patch. For the product a thin plastic film is to beused. Samples of products were received from 5 possible suppliers. Each sample consisted of 20 measurements of the film thickness and the following data were found:

	Average film thickness ($\bar{x}$ in μm)	Sample standard deviation (s in μm)
Supplier 1	31.4	1.9
Supplier 2	30.6	1.6
Supplier 3	30.5	2.2
Supplier 4	31.3	1.8
Supplier 5	29.2	2.2

From the usual calculations for a one-way analysis of variance the following is obtained:

Source	Degrees of freedom	Sum of Squares
Supplier	4	$SS(Tr) = 62$
Error	95	$SSE = 362.71$

a) Is there a significant ($\alpha = 5\%$) difference between the mean film thicknesses for the suppliers?

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

|||| Solution

The F-test statistics for one-way ANOVA is

$$F_{\text{obs}} = \frac{MS(Tr)}{MSE} = \frac{62/4}{362.71/95} = \frac{15.5}{3.818} = 4.06.$$

and the relevant critical value is $F_{0.05}(4, 95)$ to be found in Python by:

```
print(stats.f.ppf(0.95, 4, 95))
2.467
```

Conclusion: $F_{\text{obs}} = 4.06 > 2.47$, so H_0 is rejected — there are significant differences.

b) Compute a 95% confidence interval for the difference in mean film thickness between Supplier 1 and Supplier 4 (as a single pre-planned comparison).

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

|||| Solution

We use Method 8.9. We need the quantile $t_{1-\alpha/2}(95)$ to be found in Python by:

```
print(stats.t.ppf(0.975, 95))
1.985
```

We also have:

$$\bar{y}_1 - \bar{y}_4 = 31.4 - 31.3 = 0.1, \quad MSE = 362.71/95 = 3.818.$$

which we insert in the formula in Method 8.9:

$$0.1 \pm 1.985\sqrt{3.818\left(\frac{1}{20}+\frac{1}{20}\right)} = 0.1 \pm 1.985\sqrt{0.3818} = 0.1 \pm 1.227 = [-1.127;\ 1.327].$$

8.4 Brass alloys

|||| Exercise 8.4 Brass alloys

When brass is used in a production, the modulus of elasticity, E, of the material is often important for the functionality. The modulus of elasticity for 6 different brass alloys is measured. 5 samples from each alloy are tested. The results are shown in the table below where the measured modulus of elasticity is given in GPa:

Brass alloys					
M1	M2	M3	M4	M5	M6
82.5	82.7	92.2	96.5	88.9	75.6
83.7	81.9	106.8	93.8	89.2	78.1
80.9	78.9	104.6	92.1	94.2	92.2
95.2	83.6	94.5	87.4	91.4	87.3
80.8	78.6	100.7	89.6	90.1	83.8

A one-way ANOVA gives the following output (some values replaced by A, B, C):

```
           Df  Sum Sq  Mean Sq  F value     Pr(>F)
alloy       A 1192.51  238.501  9.9446  3.007e-05
Residuals   B      C   23.983      NaN       NaN
```

a) What are the values of A, B, and C?

|||| Solution

In a one-way ANOVA with $k = 6$ groups and $n = 30$ observations:

$$A = k - 1 = 5,$$
$$B = n - k = 24,$$
$$C = SSE = MSE \times (n - k) = 23.983 \times 24 = 575.59.$$

b) Which of the following gives the complete set of assumptions for a one-way ANOVA?

1) Observations are independent, the data in each group are normally distributed, and the group variances are equal.
2) Observations are independent and the data in each group are normally distributed, and the group means are equal.
3) The sample sizes are independent and normally distributed.
4) The sample size in each group is neither too large nor too small.
5) Observations are independent, the data in each group are normally distributed, and the groups have approximately the same interquartile range (IQR).

|||| Solution

Correct answer: **1)**.

The assumptions of a one-way ANOVA are:

- Independence of observations,
- Approximate normality within each group,
- Equal population variances across groups (homoscedasticity).

c) Compute a 95% confidence interval for the single pre-planned difference between brass alloys M1 and M2.

Use Python to compute quantiles and probabilities from relevant probability distribution(s).

|||| Solution

We use Method 8.9. We need the quantile $t_{1-\alpha/2}(24)$ to be found in Python by:

```
print(stats.t.ppf(0.975, 24))
2.064
```

The group means are $\bar{y}_1 = 84.62$ and $\bar{y}_2 = 81.14$, giving $\bar{y}_1 - \bar{y}_2 = 3.48$. With $MSE = 23.983$ and $n_1 = n_2 = 5$:

$$3.48 \pm t_{0.975,24}\sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 3.48 \pm 2.064\sqrt{23.983 \times \frac{2}{5}} = [-2.91;\ 9.87].$$

8.5 Plastic tubes

|||| Exercise 8.5 Plastic tubes

Plastic tubes are produced using two different types of granules (base materials), from four different suppliers. The tensile strength of the tubes is measured in MPa and recorded for each combination of supplier and granule type. The measurements from this trial are shown below:

	Granule	
	g1	g2
Supplier a	34.2	33.1
Supplier b	34.8	31.2
Supplier c	31.3	30.2
Supplier d	31.9	31.6

The following Python code fits a two-way ANOVA model with supplier and granule type as categorical factors, then prints the ANOVA table. (Optionally, try to tun the code in your Python notebook to reproduce the results)

```
D = pd.DataFrame({
    "strength": [34.2, 34.8, 31.3, 31.9,
                 33.1, 31.2, 30.2, 31.6],
    "supplier": pd.Categorical(["a", "b", "c", "d",
                                "a", "b", "c", "d"]),
    "granule": pd.Categorical([1, 1, 1, 1, 2, 2, 2, 2])
})

fit = smf.ols("strength ~ supplier + granule", data=D).fit()
print(sm.stats.anova_lm(fit))
```

The output of the two-way ANOVA is:

```
             df     sum_sq    mean_sq          F    PR(>F)
supplier   3.0   10.03375   3.344583   3.253749  0.179225
granule    1.0    4.65125   4.651250   4.524929  0.123339
Residual   3.0    3.08375   1.027917        NaN       NaN
```

a) Which distribution was used to find the p-value 0.1792?

|||| Solution

The p-value for the supplier factor comes from an F-test in a two-way ANOVA (see "F-test for group mean differences" in the formula compendium, Section 9, p. 41). To identify which F-distribution is used, we need its two degrees of freedom.

The first corresponds to the supplier factor. Since there are $k = 4$ suppliers, we have

$$\nu_1 = k - 1 = 4 - 1 = 3.$$

The second corresponds to the residuals. With $k = 4$ suppliers and $l = 2$ granule types, we have

$$\nu_2 = (k-1)(l-1) = (4-1)(2-1) = 3.$$

Under the null hypothesis of no supplier effect, the test statistic follows an F-distribution with these degrees of freedom (Theorem 8.22 in the textbook, p. 332).

The distribution used is therefore the F-distribution with degrees of freedom $\nu_1 = 3$ and $\nu_2 = 3$.

b) Which conclusion is best supported by the two-way ANOVA results, at a significance level of $\alpha = 0.05$?

1) A significant difference has been found between the tensile-strength variances.
2) A significant difference in mean tensile strength has been found between granule types, but not between suppliers.
3) No significant difference in mean tensile strength has been found between suppliers or between granule types.
4) Significant differences in mean tensile strength have been found for both suppliers and granule types.
5) A significant difference in mean tensile strength has been found between suppliers, but not between granule types.

|||| Solution

For each factor, the null hypothesis is that the population mean tensile strength is the same across its levels: across all four suppliers for the supplier test, and across the two granule types for the granule test.

We reject a null hypothesis if its p-value is below $\alpha = 0.05$ (see "P-value" under "Group hypothesis testing" and "Block hypothesis testing" in the formula compendium, Section 9, p. 41). From the ANOVA table,

$$p_{\text{supplier}} = 0.1792 > 0.05, \qquad p_{\text{granule}} = 0.1233 > 0.05.$$

Both p-values exceed 0.05, so neither null hypothesis is rejected. The correct answer is therefore **3)**: no significant difference in mean tensile strength has been found between suppliers or between granule types.

This does not establish that the population means are equal; the data do not provide sufficient evidence of a difference at the 5% significance level (see Remark 3.26 in the textbook, p. 137).

Exam-style questions

c) Supplier c has a mean tensile strength of 30.75 MPa, which is evidently lower than the mean tensile strength of 33.00 MPa for supplier b. However, the supplier effect is not statistically significant ($p = 0.1792$). Given that there is only one observation per combination of supplier-granule, which statement about this result is correct?

1) The non-significant result shows that the observed difference between suppliers b and c is due to sampling variation rather than a real supplier effect.
2) Differences between granule types remain in the residual variation, making differences between suppliers harder to distinguish from random error.
3) Without repeated measurements for each combination, the error variance cannot be estimated, so the supplier test has no residual degrees of freedom.
4) The error variance is estimated with only three residual degrees of freedom, which can limit the power to detect supplier differences.
5) The supplier test requires significant differences between every pair of suppliers, making an overall significant supplier effect harder to establish.

Solution

Correct answer: **4)**.

With $k = 4$ suppliers and $l = 2$ granule types, the residual degrees of freedom are

$$(k-1)(l-1) = (4-1)(2-1) = 3$$

(see the two-way ANOVA table in the formula compendium, Section 9, p. 39).

With so few residual degrees of freedom, there is considerable uncertainty in the estimated error variance. This can limit statistical power, meaning that real supplier differences may go undetected (see "Type II error" in the formula compendium, p. 20). The non-significant result therefore does not establish that the suppliers have equal population mean tensile strengths.

8.6 Joining methods

Exercise 8.6 Joining methods

A series of experiments is carried out to investigate how joining method and material affect the strength of a joint. Three joining methods (A, B, C) are tested with four materials (1, 2, 3, 4). One strength measurement is recorded for each combination of joining method and material. The measurements are shown below, together with the row and column averages:

	Material				Row
Joining method	1	2	3	4	average
A	242	214	254	248	239.50
B	248	214	248	247	239.25
C	236	211	245	243	233.75
Column average	242	213	249	246	

A two-way ANOVA model is fitted to the data. The Python code used for this analysis produces the output shown below. Six numerical values have been replaced by the symbols A-F:

```
                df   sum_sq   mean_sq          F    PR(>F)
Joiningmethod    A    84.5         B          C  0.050408
Material         D       E   825.00          F  0.000016
Residual       6.0    49.5      8.25        NaN       NaN
```

a) Calculate the missing values A, B and C in the ANOVA table.

Solution

The missing entries A, B and C are, respectively, the degrees of freedom, mean square and F-statistic for joining method (see the two-way ANOVA table in the textbook, p. 334, and in the formula compendium, Section 9, p. 39).

There are $k = 3$ joining methods, so the degrees of freedom for this factor are

$$A = k - 1 = 3 - 1 = 2.$$

The mean square is the sum of squares divided by the corresponding degrees of freedom. For joining method, this gives

$$B = \frac{84.5}{2} = 42.25.$$

Finally, the F-statistic is the mean square for the factor divided by the residual mean square, MSE. Therefore,

$$C = \frac{B}{MSE} = \frac{42.25}{8.25} \approx 5.12.$$

b) What are the conclusions regarding the importance of the two factors ($\alpha = 5\%$)?

|||| Solution

The two-way ANOVA provides a separate F-test for each factor: material and joining method. For each test, the null hypothesis is that the population mean strength is the same across the levels of that factor.

We reject this hypothesis if the p-value is below $\alpha = 0.05$ (see "P-value" under "Group hypothesis testing" and "Block hypothesis testing" in the formula compendium, Section 9, p. 41).

From the ANOVA table,

$$p_{\text{material}} \approx 0.000016 < 0.05, \qquad p_{\text{joining method}} \approx 0.050408 > 0.05.$$

For material, we reject the null hypothesis: there is evidence of a difference in mean strength between at least two materials.

For joining method, the p-value is slightly above 0.05, so we do not reject the null hypothesis. No significant difference in mean strength has been found between joining methods at the 5% significance level. This however, does not establish that their population means are equal.

c) Perform a post hoc analysis for materials, and for joining methods. Make pairwise comparisons of mean strength using confidence intervals, hypothesis tests, or both. Apply the Bonferroni correction with an overall significance level of $\alpha = 0.05$ for each factor.

Solution

We compare mean strengths in pairs within each factor, using $MSE = 8.25$ and 6 residual degrees of freedom from the two-way ANOVA (see Section 8.3.3 in the textbook, p. 334).

For M pairwise comparisons, the Bonferroni correction uses $\alpha_B = 0.05/M$ for each comparison. This keeps the probability of incorrectly declaring at least one difference at or below 5% within each factor.

For each pair, we test whether the two population means are equal. We reject this hypothesis if the absolute difference between their sample means exceeds the Bonferroni-corrected least significant difference (LSD) (see "Least significant difference (LSD)" in the formula compendium, Section 9, p. 42). Equivalently, the confidence interval $\bar{y}_i - \bar{y}_j \pm LSD_B$ then excludes zero.

Joining methods

There are 3 pairwise comparisons, so $\alpha_B = 0.05/3$. Each joining-method mean is based on 4 observations, giving

$$LSD_B = t_{1-0.05/(2\times 3),6}\sqrt{\frac{2 \times 8.25}{4}} \approx 6.677.$$

The numerical value can be calculated in Python:

```
LSD_join = stats.t.ppf(1-0.05/6, 6) * np.sqrt(2*8.25/4)
print(f'LSD joining methods: {LSD_join:.3f}')
```

The largest absolute difference between joining-method means is $239.50 - 233.75 = 5.75 < 6.677$. All three comparisons therefore fall below the threshold.

No significant pairwise differences in mean strength are found between joining methods.

Materials

There are 6 pairwise comparisons, so $\alpha_B = 0.05/6$. Each material mean is based on 3 observations, giving

$$LSD_B = t_{1-0.05/(2\times 6),6}\sqrt{\frac{2 \times 8.25}{3}} \approx 9.060.$$

Similarly, in Python:

```
LSD_mat = stats.t.ppf(1-0.05/12, 6) * np.sqrt(2*8.25/3)
print(f'LSD materials: {LSD_mat:.3f}')
```

Material 2 has the lowest sample mean, 213. Its absolute differences from materials 1, 3 and 4 are 29, 36 and 33, respectively, all exceeding 9.060. Among materials 1, 3 and 4, the largest absolute difference is $249 - 242 = 7 < 9.060$.

Material 2 has significantly lower mean strength than each of materials 1, 3 and 4. No significant differences are found among materials 1, 3 and 4.

d) Perform a residual analysis to check the model assumptions:

   1. Normality
   2. Variance homogeneity

|||| Solution

We examine the residuals, which represent the variation remaining after accounting for joining method and material (see Section 8.3.4 in the textbook, pp. 336–337).

1. **Normality:** Use a normal Q-Q plot of the residuals. Approximately linear points support the normality assumption, whereas systematic curvature or large deviations from the line may indicate that the residuals are not normal.

   In this exercise, the points roughly follow a straight line, with some deviation at the upper end. Normality is plausible, although the small sample of 12 observations limits the assessment.

2. **Variance homogeneity:** Use boxplots of the residuals grouped by joining method and by material and compare their spreads. Notable differences may indicate that the error variance is not constant.

   In this case, the spreads are similar across joining methods. Materials 2 and 4 show less spread than materials 1 and 3. However, with only three observations per material, one for each joining method, this is insufficient to draw a firm conclusion about unequal variances.

Exam-style question

e) The two-way ANOVA gives a p-value of approximately 0.05041 for joining method. Which statement best describes this result at a significance level of $\alpha = 0.05$?

1) The result establishes that the joining methods have equal population mean strengths, since the null hypothesis is not rejected.
2) There is weak evidence of differences in mean strength between joining methods, but the result is not statistically significant at the 5% level.
3) There is approximately a 5.041% probability that the population mean strength is the same for all three joining methods.
4) A significant difference in mean strength can be concluded, since the p-value is sufficiently close to the 5% significance level.
5) The p-value must first be compared with $0.05/3$, since testing the joining-method factor involves comparing three joining methods.

|||| Solution

The p-value measures the evidence against the null hypothesis of equal population mean strengths. Here, $p \approx 0.05041$ corresponds to weak evidence against this hypothesis (see Table 3.1 in the textbook, p. 135).

Since $0.05041 > 0.05$, we do not reject the null hypothesis at the 5% significance level. The correct answer is therefore **2)**.

This does not establish that the population means are equal. It means that the evidence is insufficient to conclude that they differ at the chosen significance level.

8.7 Remoulade

|||| Exercise 8.7 Remoulade

A supermarket wants to find the best recipe for homemade "remoulade" sauce. A taste experiment is conducted with four dressing types (A, B, C, D) and three pickle types (I, II, III). One taste score is recorded for each combination of pickle type and dressing type, using a continuous scale from 0 to 5. The scores are shown below, together with the row and column averages:

	Dressing type				Row
Pickle type	A	B	C	D	average
I	4.0	3.0	3.8	2.4	3.30
II	4.3	3.1	3.3	1.9	3.15
III	3.9	2.3	3.0	2.4	2.90
Column average	4.07	2.80	3.37	2.23	

A two-way ANOVA model is fitted to the data. In the Python output below, six numerical values have been replaced by the symbols A-F:

```
            df     sum_sq   mean_sq          F    PR(>F)
Pickles      A   0.326667  0.163333          E  0.287133
Dressing     B   5.536667  1.845556          F  0.002273
Residual     C          D  0.105556        NaN       NaN
```

a) Calculate the missing values A, B and C in the ANOVA table.

|||| Solution

The entries A, B and C are the degrees of freedom for pickles, dressing and residuals, respectively (see the two-way ANOVA table in the formula compendium, Section 9, p. 39).

Each factor has one fewer degree of freedom than its number of levels. With $k = 3$ pickle types and $l = 4$ dressing types, we obtain

$$A = k - 1 = 3 - 1 = 2, \qquad B = l - 1 = 4 - 1 = 3.$$

With one observation per combination, the residual degrees of freedom are

$$C = (k-1)(l-1) = (3-1)(4-1) = 6.$$

b) Calculate the missing values D, E and F in the ANOVA table.

Solution

The entry D is the residual sum of squares, while E and F are the F-statistics for pickles and dressing, respectively. We use the relationships between sums of squares, mean squares and F-statistics in the two-way ANOVA table.

The residual mean square is the residual sum of squares divided by its degrees of freedom. Using $C = 6$, we therefore obtain

$$D = SSE = C \times MSE = 6 \times 0.105556 \approx 0.633.$$

For each factor, the F-statistic is its mean square divided by the residual mean square. Thus,

$$E = \frac{MS(\text{Pickles})}{MSE} = \frac{0.163333}{0.105556} \approx 1.547,$$

$$F = \frac{MS(\text{Dressing})}{MSE} = \frac{1.845556}{0.105556} \approx 17.48.$$

c) At a significance level of $\alpha = 0.05$, what can be concluded about the effects of dressing type and pickle type on mean taste score?

Solution

The two-way ANOVA provides a separate F-test for each factor. For each test, the null hypothesis is that the population mean taste score is the same across the levels of that factor (see "Group hypothesis testing" and "Block hypothesis testing" in the formula compendium, Section 9, p. 41).

We compare each p-value with $\alpha = 0.05$. From the ANOVA table,

$$p_{\text{dressing}} \approx 0.002273 < 0.05, \qquad p_{\text{pickles}} \approx 0.287133 > 0.05.$$

For dressing type, we reject the null hypothesis: there is evidence of a difference in mean taste score between at least two dressing types.

For pickle type, we do not reject the null hypothesis. No significant difference in mean taste score has been found between pickle types at the 5% significance level. This does not establish that their population means are equal.

Exam-style question

d) A similar taste experiment compares five pickle types and six dressing types, with one observation for each combination. An additive two-way ANOVA model is fitted to the data.
What are the degrees of freedom for pickle type, dressing type and residuals, respectively?

1) 4, 5, 19
2) 5, 6, 18
3) 4, 5, 9
4) 4, 5, 29
5) 4, 5, 20

|||| Solution

Each factor has one fewer degree of freedom than its number of levels. Therefore,

$$df_{\text{pickles}} = 5 - 1 = 4, \qquad df_{\text{dressing}} = 6 - 1 = 5.$$

The $5 \times 6 = 30$ observations provide $30 - 1 = 29$ total degrees of freedom. After subtracting those used by the two factors, the residual degrees of freedom are

$$df_{\text{residual}} = 29 - 4 - 5 = 20.$$

This is equivalent to the formula $(k-1)(l-1) = (5-1)(6-1) = 20$ (see the two-way ANOVA table in the formula compendium, Section 9, p. 39).

The correct answer is **5)**.

8.8 Transport times

|||| Exercise 8.8 Transport times

Delivery times are compared for three transport firms (A, B, C) and three item sizes (small, intermediate and large). One delivery time is recorded in days for each combination of firm and item size. The measurements are shown below, together with the row and column averages:

	Item size			Row
	Small	Intermediate	Large	average
Firm A	1.4	2.5	2.1	2.00
Firm B	0.8	1.8	1.9	1.50
Firm C	1.6	2.0	2.4	2.00
Column average	1.27	2.10	2.13	

A two-way ANOVA model is fitted to the data. In the Python output below, four numerical values have been replaced by the symbols A-D:

```
            df     sum_sq    mean_sq          F    PR(>F)
Firm       2.0          A          B   4.285714   0.10124
Itemsize   2.0   1.446667          C          D   0.01929
Residual   4.0   0.233333   0.058333        NaN       NaN
```

a) Calculate the missing values A, B, C and D in the ANOVA table.

|||| Solution

The entries A and B are the sum of squares and mean square for firm. The entries C and D are the mean square and F-statistic for item size (see the two-way ANOVA table in the formula compendium, Section 9, p. 39).

For each factor, the F-statistic is its mean square divided by the residual mean square. Rearranging this relationship, we first find the mean square for firm:

$$B = F_{\text{firm}} \times MSE = 4.285714 \times 0.058333 \approx 0.25.$$

The sum of squares is the mean square multiplied by its degrees of freedom. Since firm has 2 degrees of freedom,

$$A = 2B \approx 0.50.$$

For item size, we divide the sum of squares by its 2 degrees of freedom to obtain the mean square:

$$C = \frac{1.446667}{2} \approx 0.723333.$$

Finally, dividing this mean square by the residual mean square gives

$$D = \frac{C}{MSE} \approx \frac{0.723333}{0.058333} \approx 12.4.$$

b) At a significance level of $\alpha = 0.05$, what can be concluded about the effects of transport firm and item size on mean delivery time?

|||| Solution

The two-way ANOVA provides a separate F-test for each factor. For each test, the null hypothesis is that the population mean delivery time is the same across the levels of that factor (see "Group hypothesis testing" and "Block hypothesis testing" in the formula compendium, Section 9, p. 41).

We compare each p-value with $\alpha = 0.05$. From the ANOVA table,

$$p_{\text{item size}} \approx 0.01929 < 0.05, \qquad p_{\text{firm}} \approx 0.10124 > 0.05.$$

For item size, we reject the null hypothesis: there is evidence of a difference in mean delivery time between at least two item-size categories.

For transport firm, we do not reject the null hypothesis. No significant difference in mean delivery time has been found between firms at the 5% significance level. This does not establish that their population means are equal.

Exam-style question

c) A similar delivery-time study compares three transport firms and four item-size categories, with one observation for each combination. An additive two-way ANOVA gives $MSE = 0.08$ and an F-statistic of 5.0 for firm.

What is the mean square for firm, and how many residual degrees of freedom does the model have?

1) $MS(\text{Firm}) = 0.016$, residual $df = 6$
2) $MS(\text{Firm}) = 0.4$, residual $df = 12$
3) $MS(\text{Firm}) = 0.4$, residual $df = 6$
4) $MS(\text{Firm}) = 0.016$, residual $df = 5$
5) $MS(\text{Firm}) = 0.4$, residual $df = 5$

|||| Solution

The F-statistic is the factor mean square divided by the residual mean square (see the two-way ANOVA table in the formula compendium, Section 9, p. 39). Rearranging gives

$$MS(\text{Firm}) = F_{\text{firm}} \times MSE = 5.0 \times 0.08 = 0.4.$$

With $k = 3$ firms, $l = 4$ item-size categories and one observation per combination, the residual degrees of freedom are

$$df_{\text{residual}} = (k-1)(l-1) = (3-1)(4-1) = 6.$$

The correct answer is **3)**.