

Factorial designs

STAT 6160 • Experimental Design • Fall 2026

Siddharth Vishwanath

§1. Full factorials

Better popcorn

Should we increase the power, cook longer, or change the brand?

Factor	Low	High
<i>A</i> : power	Medium	High
<i>B</i> : time	Short	Long

One fresh batch per run. Response: **percentage of kernels that pop without burning.**

Keep the starting quantity, bowl, and preparation fixed.

An illustrative experiment. Table entries below are mean responses.

Changing one factor at a time

Start at medium power and a short cooking time.

Power / time	Short	Long
Medium	60	80
High	90	?

Longer cooking helps. Higher power helps.

What would you predict if we changed both?

All four combinations

Power / time	Short	Long
Medium	60	80
High	90	70

A **full factorial** includes every combination of factor settings.

Does longer cooking help? What comparison answers that question?

Coding a 2^2 experiment

Write $-$ for -1 (low) and $+$ for $+1$ (high), in the order A, B .

Setting	I	A : power	B : time	AB
$--$	1	-1	-1	$+1$
$+-$	1	$+1$	-1	-1
$-+$	1	-1	$+1$	-1
$++$	1	$+1$	$+1$	$+1$

$+-$ means **high power, short time**. The signs label settings, not outcomes.

A, B specify the design. I, A, B, AB form the **model matrix**. Multiply signs to get AB .

A third factor

Let C compare **store** (–) with **premium** (+) popcorn.

Power	Time	Store	Premium
Medium	Short	60	66
High	Short	90	96
Medium	Long	80	86
High	Long	70	76

Three two-level factors give $2^3 = 8$ **combinations**. In general: 2^k .

Now + – + means **high power, short time, premium brand**.

The 2^3 design matrix

The three signs in each setting are in the order A, B, C .

Setting	I	A	B	C	AB	AC	BC	ABC
---	+	-	-	-	+	+	+	-
+--	+	+	-	-	-	-	+	+
-+-	+	-	+	-	-	+	-	+
++-	+	+	+	-	+	-	-	-
--+	+	-	-	+	+	-	-	+
+ - +	+	+	-	+	-	+	-	-
- + +	+	-	+	+	-	-	+	-
+++	+	+	+	+	+	+	+	+

---: all low. +++: all high. I is the intercept column.

Standard order changes A fastest. **Randomize the actual run order.**

Eight means, eight coefficients

With coded settings $A, B, C \in \{-1, +1\}$,

$$\begin{aligned}\mathbb{E}(Y \mid A, B, C) = & \beta_0 + \beta_A A + \beta_B B + \beta_C C \\ & + \beta_{AB} AB + \beta_{AC} AC + \beta_{BC} BC \\ & + \beta_{ABC} ABC.\end{aligned}$$

The full model can represent **any eight combination means**.

Why do eight observations leave no residual degrees of freedom?

A pilot-plant experiment

Which operating conditions improve chemical yield?

Factor	–	+
T : temperature	160°C	180°C
C : concentration	20%	40%
K : catalyst type	A	B

Eight settings, **two independent runs per setting**, in randomized order.

Response: yield (%). An experimental unit is a complete reactor run.

Box, Hunter, and Hunter, §5.4. A real pilot-plant example simplified by the authors.

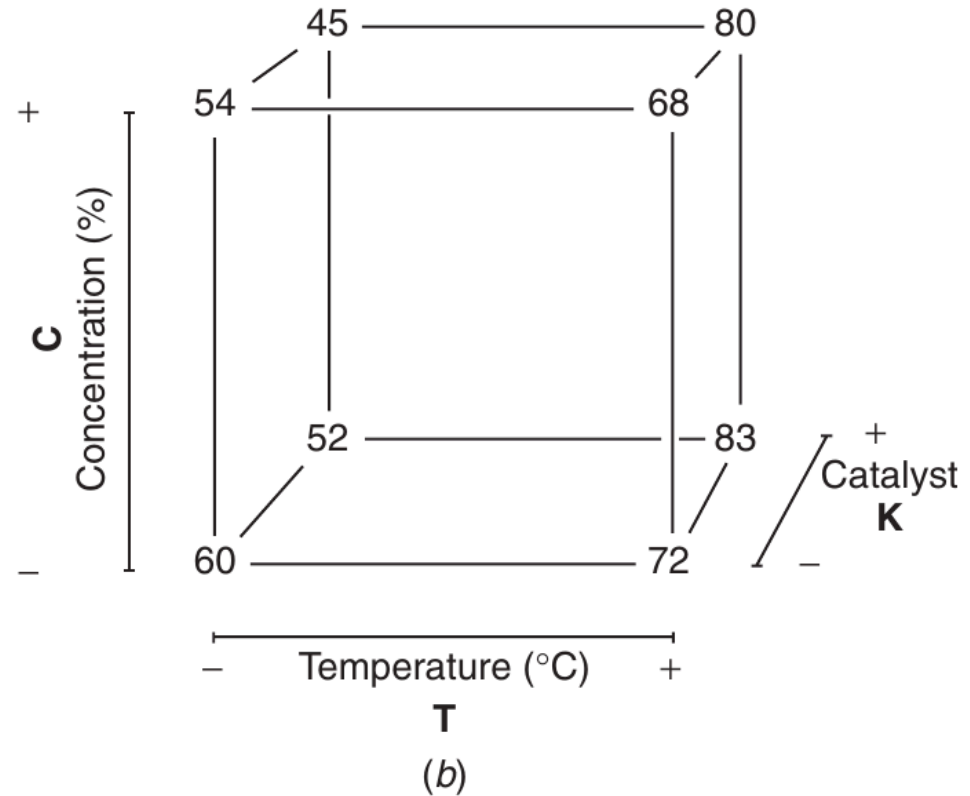
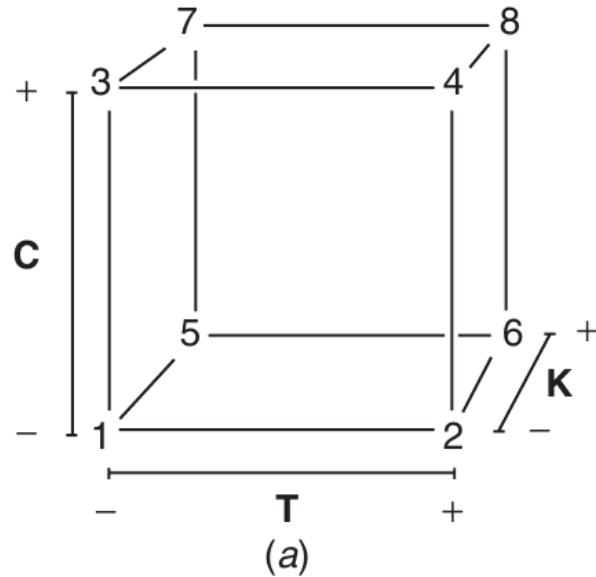
Pilot plant: the mean responses

Setting (T, C, K)	Temperature	Concentration	Catalyst	Mean yield (%)
---	160	20	A	60
+--	180	20	A	72
-+-	160	40	A	54
++-	180	40	A	68
--+	160	20	B	52
+ - +	180	20	B	83
- + +	160	40	B	45
+++	180	40	B	80

Each response is the mean of two runs. T, C, K take the roles of A, B, C in the coded matrix.

BHH, Table 5.2, p. 177. Temperature in °C; concentration in %.

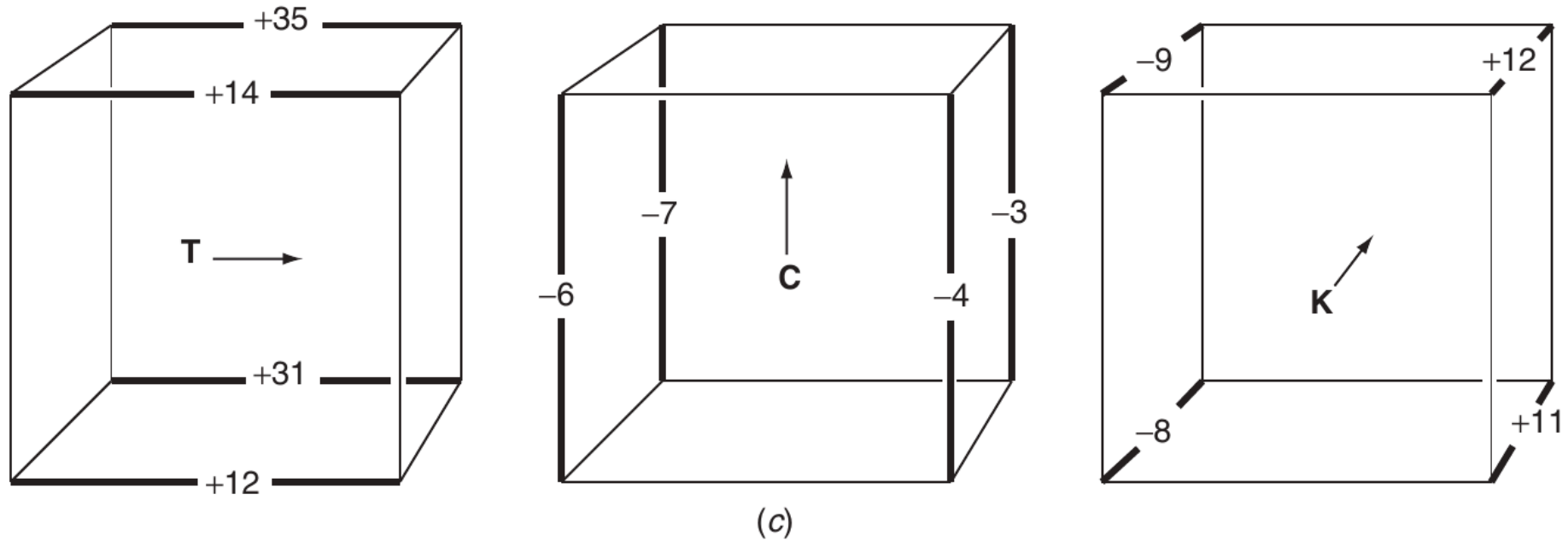
The 2^3 experiment as a cube



An edge comparison is a **simple effect**: change one factor, hold the others fixed.

BHH, Figure 5.3(a-b), p. 178.

Main effects on the cube



A **main effect** averages the four simple effects for that factor.

$$\hat{d}_T = \frac{12 + 14 + 31 + 35}{4}, \quad \hat{d}_C = \frac{-6 - 4 - 7 - 3}{4}, \quad \hat{d}_K = \frac{-8 + 11 - 9 + 12}{4}.$$

BHH, Figure 5.3(c), p. 178. Effects are in percentage points.

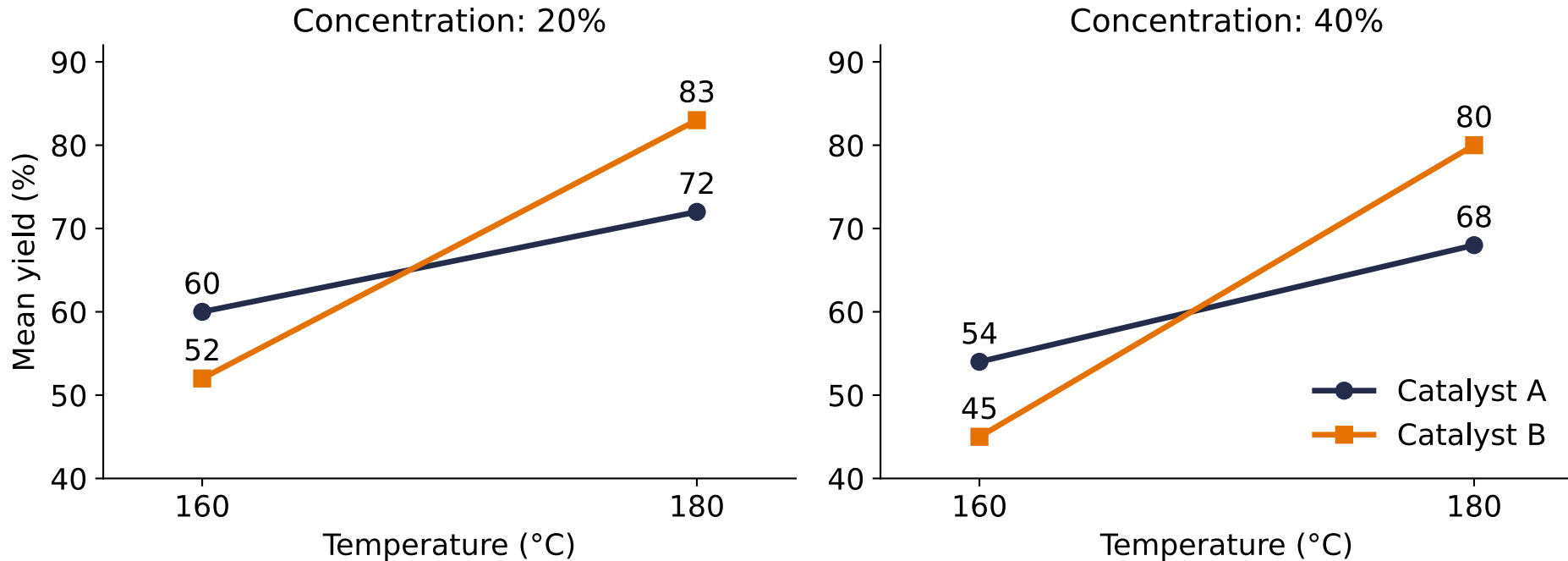
A difference of averages

$$\hat{d}_T = \bar{Y}_{T=+1} - \bar{Y}_{T=-1} = \frac{72 + 68 + 83 + 80}{4} - \frac{60 + 54 + 52 + 45}{4}.$$

Each average balances concentration and catalyst in exactly the same way.

Where are these two averages on the cube?

Are the temperature effects the same?



Points are combination means; lines connect means for the same catalyst.

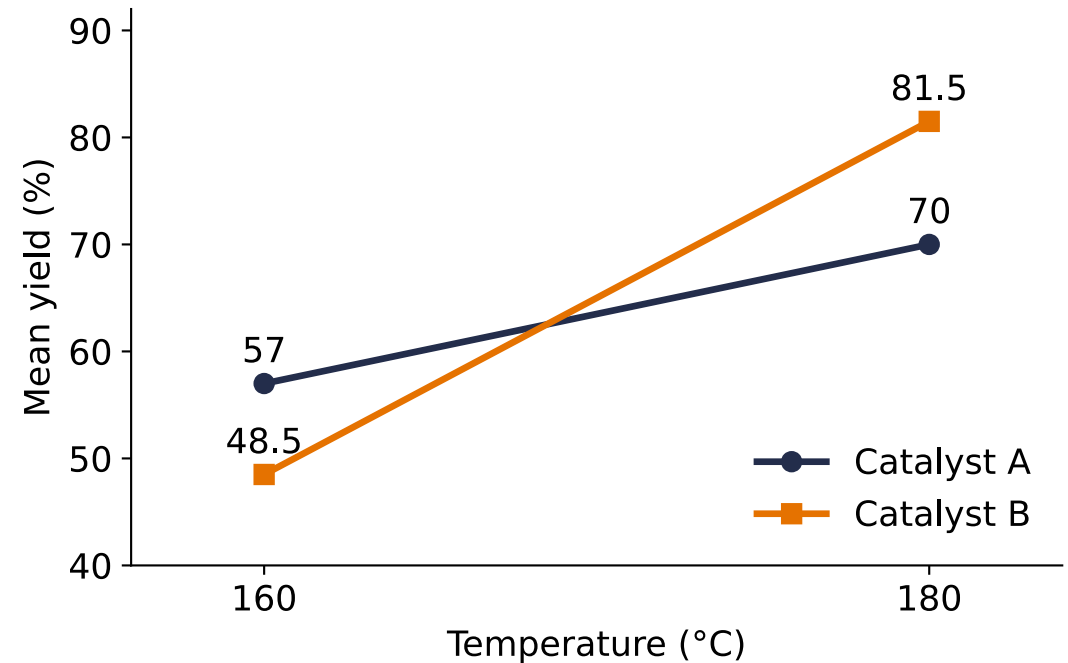
What does the average temperature effect of 23 leave out?

Averaging over concentration

Catalyst	160°C	180°C
A	57.0	70.0
B	48.5	81.5

Each entry averages the two concentrations.

For example, $57 = (60 + 54)/2$.



The same comparisons as BHH, Figure 5.5, p. 186.

First differences: change temperature

Let $\bar{y}_{t,k}$ average over concentration and replicates. Sign order: **temperature, catalyst**.

$$D_T(k) = \bar{y}_{+,k} - \bar{y}_{-,k}.$$

$$D_T(-) = 70 - 57, \quad D_T(+) = 81.5 - 48.5.$$

How does the effect of temperature change when we switch catalyst?

Second differences: change the effect

An interaction is a **difference of first differences**:

$$\begin{aligned}\Delta_K \Delta_T \bar{y} &= (\bar{y}_{++} - \bar{y}_{-+}) - (\bar{y}_{+-} - \bar{y}_{--}) . \\ &= (81.5 - 48.5) - (70 - 57) .\end{aligned}$$

Factorial convention: $\hat{d}_{TK} = \frac{1}{2} \Delta_K \Delta_T \bar{y}$.

The interaction term in a regression

For coded $T, K \in \{-1, +1\}$, averaging over concentration,

$$\mu(T, K) = \beta_0 + \beta_T T + \beta_K K + \beta_{TK} TK.$$

Taking a first difference in T gives

$$\mu(+1, K) - \mu(-1, K) = 2\beta_T + 2\beta_{TK} K.$$

Take another difference, now in K . Relate d_{TK} to β_{TK} .

Three-factor interactions

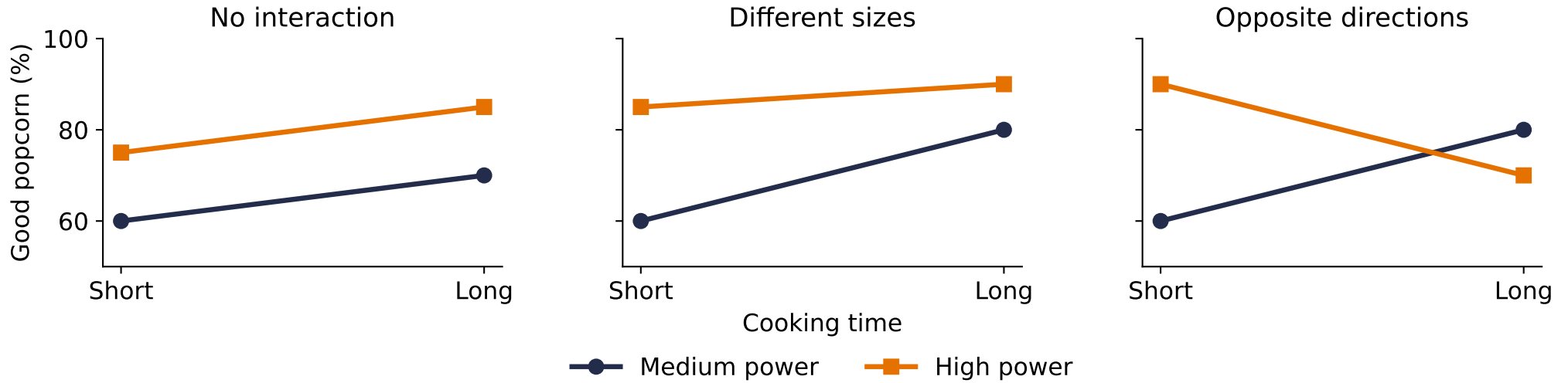
Does the temperature-by-catalyst interaction itself change with concentration?

$$\hat{d}_{TK|C=-} = \frac{(83 - 52) - (72 - 60)}{2},$$

$$\hat{d}_{TK|C=+} = \frac{(80 - 45) - (68 - 54)}{2}.$$

$$\hat{d}_{TCK} = \frac{\hat{d}_{TK|C=+} - \hat{d}_{TK|C=-}}{2}.$$

Interaction does not require crossing



An **interaction** means the effect of one factor depends on the setting of another.

The lines need not cross. These panels show population means.

Plot code in the appendix. See also Belzile, Fig. 5.1.

Interpreting interactions

- ▶ A main effect is an **average**. Catalyst's average effect is small even though the preferred catalyst changes with temperature.
- ▶ Interaction depends on the response scale and the settings studied.
- ▶ Noisy sample means can produce nonparallel lines. Estimate the **difference of effects**, with uncertainty.

“Significant for one catalyst, not significant for the other” does **not** establish an interaction.

One contrast for each effect

Multiply the factor columns to obtain each interaction column.

For any nonconstant sign column x_j and N total runs,

$$\hat{d}_j = \bar{Y}_{x_j=+1} - \bar{Y}_{x_j=-1} = \frac{2}{N} \sum_{i=1}^N x_{ij} Y_i.$$

In a balanced full factorial, $X^\top X = NI$. Therefore

$$\hat{\beta} = \frac{1}{N} X^\top \mathbf{Y}, \quad \hat{d}_j = 2\hat{\beta}_j.$$

Estimating the error variance

Let $q = 2^k$, with r independent runs at each setting. Then $N = qr$.

$$Y_{h\ell} = \mu_h + \epsilon_{h\ell}, \quad h = 1, \dots, q, \quad \ell = 1, \dots, r.$$

Assume independent errors with mean zero and common variance σ^2 .

$$s^2 = \frac{\sum_{h=1}^q \sum_{\ell=1}^r (Y_{h\ell} - \bar{Y}_h)^2}{q(r-1)}.$$

We lose one degree of freedom to each fitted cell mean: $\nu = N - q$.

The variance of an effect

Each contrast compares two **disjoint averages of $N/2$ runs**.

$$\text{Var} (\hat{d}_j) = \text{Var} (\bar{Y}_{x_j=+1}) + \text{Var} (\bar{Y}_{x_j=-1}).$$

Finish the variance calculation, then replace σ^2 by s^2 .

Standard errors and intervals

For a balanced full factorial,

$$\text{SE}(\hat{d}_j) = \frac{2s}{\sqrt{N}}, \quad \nu = N - 2^k.$$

With normal errors, an individual 95% interval is

$$\hat{d}_j \pm t_{0.975, \nu} \frac{2s}{\sqrt{N}}.$$

For the unscaled second difference $\hat{\Delta}_{AB} = 2\hat{d}_{AB}$, **multiply both the estimate and its SE by two.**

Factorial regression is ANOVA

A full 2^3 model and a one-way model with **eight treatment combinations** fit the same eight means.

The factorial model partitions the **7 treatment degrees of freedom**:

Source	df
A, B, C	$1 + 1 + 1$
AB, AC, BC	$1 + 1 + 1$
ABC	1
Error	$N - 8$

Both models have the same residuals and residual mean square.

One effect, one degree of freedom

For a balanced two-level factorial,

$$SS_j = \frac{N\hat{d}_j^2}{4}, \quad MS_j = SS_j.$$

$$F_j = \frac{SS_j}{s^2} = \left(\frac{\hat{d}_j}{2s/\sqrt{N}} \right)^2 = t_j^2.$$

The one-df ANOVA F -test and the two-sided contrast t -test give **the same p -value**.

Keep the signed effect and its interval when explaining the result.

Worked analysis: the pilot plant

The individual runs

Setting (T, C, K)	Yield 1	Yield 2	Mean	Within-setting variance
---	59	61	60	2
+--	74	70	72	8
-+-	50	58	54	32
++-	69	67	68	2
--+	50	54	52	8
+ - +	81	85	83	8
- + +	46	44	45	2
+++	79	81	80	2

Each replicate repeats the full reactor procedure. Repeated assays of one sample would estimate a different variance.

BHH, Table 5.3 and §5.7, pp. 183–184. Rows are in standard order, not run order.

Load the data

```
import numpy as np
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf

plan = pd.DataFrame({
    "T": [-1, 1] * 4,
    "C": [-1, -1, 1, 1] * 2,
    "K": [-1] * 4 + [1] * 4})
pilot = plan.loc[plan.index.repeat(2)].reset_index(drop=True)
pilot["y"] = [59, 61, 74, 70, 50, 58, 69, 67,
             50, 54, 81, 85, 46, 44, 79, 81]
pilot["order"] = [6, 13, 2, 4, 1, 16, 5, 10,
                 8, 12, 9, 14, 3, 11, 7, 15]
fit = smf.ols("y ~ T * C * K", data=pilot).fit()
```

$T * C * K$ includes all main effects, all two-factor interactions, and TCK .

Compute the error variance

For two runs at a setting,

$$s_h^2 = \frac{(Y_{h1} - Y_{h2})^2}{2}.$$

Pool the eight within-setting estimates:

$$s^2 = \frac{2 + 8 + 32 + 2 + 8 + 8 + 2 + 2}{8} = 8, \quad \nu = 8.$$

$$\text{SE}(\hat{d}_j) = \sqrt{\frac{8}{8} + \frac{8}{8}} = \sqrt{2} \approx 1.414.$$

Estimate effects and their uncertainty

```
effects = pd.DataFrame({  
    "Effect": 2*fit.params,  
    "SE": 2*fit.bse,  
    "Lower": 2*fit.conf_int()[0],  
    "Upper": 2*fit.conf_int()[1]  
}).drop("Intercept")  
print(effects.round(2).to_string())
```

	Effect	SE	Lower	Upper
T	23.0	1.41	19.74	26.26
C	-5.0	1.41	-8.26	-1.74
T:C	1.5	1.41	-1.76	4.76
K	1.5	1.41	-1.76	4.76
T:K	10.0	1.41	6.74	13.26
C:K	0.0	1.41	-3.26	3.26
T:C:K	0.5	1.41	-2.76	3.76

Effects and individual **95% intervals**, in percentage points. Each interval uses $t_{0.975,8} \approx 2.306$.

A small catalyst main effect does not make catalyst irrelevant.

Read the ANOVA table

```
av = sm.stats.anova_lm(fit)
print(av[["df", "sum_sq", "F"]]
      .round(3).to_string(na_rep=""))
```

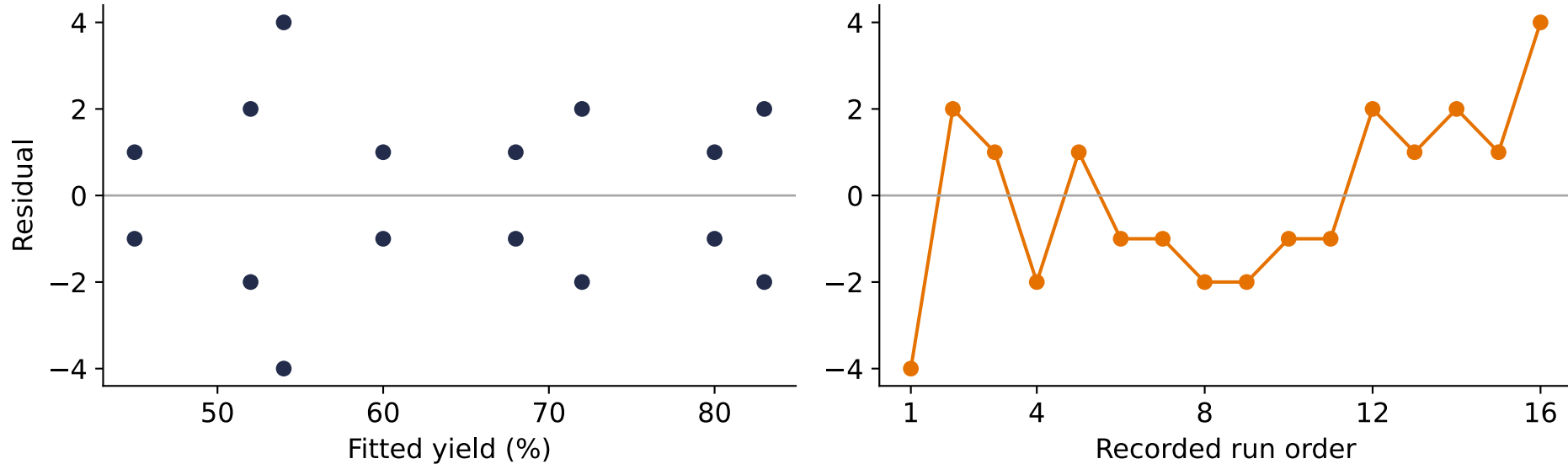
	df	sum_sq	F
T	1.0	2116.0	264.500
C	1.0	100.0	12.500
T:C	1.0	9.0	1.125
K	1.0	9.0	1.125
T:K	1.0	400.0	50.000
C:K	1.0	0.0	0.000
T:C:K	1.0	1.0	0.125
Residual	8.0	64.0	

For temperature by catalyst,

$$SS_{TK} = \frac{16(10)^2}{4} = 400, \quad F_{TK} = \frac{400}{8} = 50.$$

Check this against the squared t statistic for the same effect.

Check the residuals



Late-run residuals are positive. Investigate possible drift before relying on the independent-error model.

Why do the two residuals at each setting sum to zero?

Interpret and choose a follow-up

Comparison	Estimated change in yield
Raise temperature, catalyst A	+13 points
Raise temperature, catalyst B	+33 points
Switch A to B, at 160°C	−8.5 points
Switch A to B, at 180°C	+11.5 points

The best observed setting is 180° **C**, **20% concentration**, **catalyst B** (mean 83%).

Confirm with fresh runs. Two temperatures do not locate an optimum.

BHH §5.8: the catalyst difference prompted further investigation of the suppliers.

§2. Fractional factorials

A four-run budget

Return to popcorn: $A = \text{power}$, $B = \text{time}$, $C = \text{brand}$.

Eight combinations are possible. Suppose we can afford four.

A **fractional factorial** uses a carefully chosen subset. Choose every A, B combination and set $C = AB$.

A : power	B : time	C : brand
—	—	+
+	—	—
—	+	—
+	+	+

Which brand do we use at high power and a short time?

Does premium perform worse?

Power	Time	Brand	Mean (%)
Medium	Short	Premium	66
High	Short	Store	90
Medium	Long	Store	80
High	Long	Premium	76

$$\text{Premium minus store} = \frac{66 + 76}{2} - \frac{90 + 80}{2}.$$

Premium added 6 points at every power-and-time setting. What happened?

Two names for the same column

A	B	C	AB
–	–	+	+
+	–	–	–
–	+	–	–
+	+	+	+

Brand and power by time are **aliased**: C and AB are identical on these four combinations.

Could we fit separate coefficients for both columns?

What does the brand contrast estimate?

$$\beta_C C + \beta_{AB} AB = (\beta_C + \beta_{AB})C.$$

Write $\hat{b}_C$ for the fitted coefficient of this shared column. Then

$$\mathbb{E}(\hat{b}_C) = \beta_C + \beta_{AB}, \quad \mathbb{E}(2\hat{b}_C) = d_C + d_{AB}.$$

For our means, the contrast is $6 + (-20) = -14$ percentage points.

A large contrast identifies a combination of effects. Which effect caused it?

From a generator to a defining relation

The **generator** $C = AB$ tells us which brand to use for each power-and-time setting.

Multiply both sides by C :

$$C^2 = ABC \Rightarrow \boxed{I = ABC}.$$

Every coded factor squares to I , the all-ones column.

The **defining relation** lists the product columns equal to I on the chosen fraction.

Why is the intercept aliased with the three-factor interaction?

Find the other aliases

Multiply $I = ABC$ by the column of interest. For power,

$$A = A(ABC) = A^2BC = BC.$$

Find the aliases of time (B) and brand (C).

These are equalities of **design columns**, not claims that physical effects are equal.

When is a fraction useful?

A fraction saves runs by giving up the ability to separate some effects.

- ▶ **Sparsity:** only a few effects may be important.
- ▶ **Hierarchy:** lower-order effects are often larger than higher-order interactions.

Choose aliases that pair effects of interest with effects you can plausibly neglect.

Is neglecting power-by-time plausible for our popcorn?

How many combinations?

A **regular** 2^{k-p} **fraction** has k two-level factors and p independent generators.

$$q = 2^{k-p} = \frac{2^k}{2^p}.$$

One generator keeps half the combinations; two keep a quarter.

With r independent runs at each combination, $N = rq$.

Six factors, a quarter fraction: how many combinations? How many runs if $r = 2$?

Four factors in eight runs

Add D : **bowl shape** (shallow $-$, deep $+$).

Use all eight combinations of A, B, C , with $D = ABC$.

Combination	A	B	C	$D = ABC$
1	$-$	$-$	$-$	$-$
2	$+$	$-$	$-$	$+$
3	$-$	$+$	$-$	$+$
4	$+$	$+$	$-$	$-$
5	$-$	$-$	$+$	$+$
6	$+$	$-$	$+$	$-$
7	$-$	$+$	$+$	$-$
8	$+$	$+$	$+$	$+$

Check combination 6. Then randomize the order of the eight runs.

What does this half fraction protect?

With $D = ABC$,

$$I = ABCD, \quad A = BCD, \quad AB = CD.$$

Main effects are separate from two-factor interactions.

But **power by time** is aliased with **brand by bowl shape**.

Can we attribute a large AB/CD contrast to power by time alone?

Resolution

The **resolution** is the length of the shortest nonidentity word in the **complete defining relation**.

Defining relation	Resolution	An alias pair
$I = ABC$	III	$A = BC$
$I = ABCD$	IV	$AB = CD$
$I = ABCDE$	V	$AB = CDE$

Count letters after cancelling repeated factors. Do not count I .

Does resolution V separate two-factor from three-factor interactions?

What resolution guarantees

For regular two-level designs:

Resolution	Main effects vs. two-factor interactions	Two-factor interactions vs. each other
III	May be aliased	May be aliased
IV	Separate	May be aliased
V	Separate	Separate

Main effects are separate from one another in all three cases.

Higher-order aliases remain. **Check the aliases of the effects you care about.**

Protecting power by time

Add E : **oil amount** (none $-$, a fixed small amount $+$).

A 16-run half fraction uses A, B, C, D as basic factors and $E = ABCD$:

$$I = ABCDE, \quad A = BCDE, \quad AB = CDE.$$

This is 2_{V}^{5-1} . Power by time is separate from every other two-factor interaction.

What assumption lets us interpret the AB/CDE contrast as power by time?

Five factors on an eight-run budget

For the same five factors, try a **quarter fraction**. Choose A, B, C freely, then set $D = AB$ and $E = AC$.

Combination	A	B	C	$D = AB$	$E = AC$
1	—	—	—	+	+
2	+	—	—	—	—
3	—	+	—	—	+
4	+	+	—	+	—
5	—	—	+	+	—
6	+	—	+	—	+
7	—	+	+	—	—
8	+	+	+	+	+

$2^{5-2} = 8$ combinations. **What have these generators confounded with bowl shape and oil amount?**

Include every generator product

From $D = AB$ and $E = AC$, we get $I = ABD$ and $I = ACE$.

Their product also equals I :

$$(ABD)(ACE) = A^2BCDE = BCDE.$$

Thus the **complete defining relation** is

$$\boxed{I = ABD = ACE = BCDE}.$$

With p independent generators, there are 2^p **words**, including I .

What is its resolution?

A complete alias set

Multiply the target column by **every word**, including I .

For power, using $I = ABD = ACE = BCDE$:

$$A(I) = A, \quad A(ABD) = BD, \quad A(ACE) = CE, \quad A(BCDE) = ABCDE.$$

$$\boxed{A = BD = CE = ABCDE}.$$

Find the complete alias set for power by time (AB).

Check the complete relation

A six-factor quarter fraction uses A, B, C, E as basic factors and

$$D = ABC, \quad F = ABE.$$

Start with

$$I = ABCD = ABEF.$$

Find the missing defining word, the resolution, and the alias set for AB .

How many columns can we estimate?

In N runs,

$$\text{residual df} = N - \text{rank}(X).$$

An intercept and m linearly independent main-effect columns use $m + 1$ dimensions.

For a **16-run screening experiment**:

Main-effect columns	Total fitted columns	Residual df
13	$1 + 13$	
15	$1 + 15$	

What is the largest possible number of separate main-effect columns?

Saturated models

A fitted model is **saturated** when $\text{rank}(X) = N$.

- ▶ There are no residual degrees of freedom to estimate error variance.
- ▶ Adding an aliased interaction column does not increase the rank.
- ▶ Replicating combinations can estimate pure error, but the aliases remain.

Our 16-run, five-factor resolution V design has $1 + 5 + 10 = 16$ columns if we fit all main effects and two-factor interactions.

Where would an uncertainty estimate come from?

Can eight runs give resolution IV?

For five factors, six columns fit: I, A, B, C, D, E . Resolution IV requires more.

Consider these two groups:

$$\underbrace{I, A, B, C, D, E}_{6 \text{ columns}} \quad \underbrace{AB, AC, AD, AE}_{4 \text{ columns}}.$$

Resolution IV requires all ten to be distinct, mutually orthogonal columns.

Why can none coincide? Can ten orthogonal columns fit in eight rows?

Four new combinations

Return to power, time, and brand. Add the other half: $C = -AB$.

Power	Time	Brand	Mean (%)
Medium	Short	Store	60
High	Short	Premium	96
Medium	Long	Premium	86
High	Long	Store	70

$$\text{Premium minus store} = \frac{96 + 86}{2} - \frac{60 + 70}{2}.$$

What changed about the relation between C and AB ?

Separate brand from power by time

Fraction	Brand contrast	Estimates
Original: $C = AB$	-14	$d_C + d_{AB}$
Complement: $C = -AB$	26	$d_C - d_{AB}$

Together, the two halves give the full 2^3 design. Start with

$$\hat{d}_C + \hat{d}_{AB} = -14, \quad \hat{d}_C - \hat{d}_{AB} = 26.$$

Add and subtract the equations. Then recover $\hat{\beta}_C$ and $\hat{\beta}_{AB}$.

Choose the signs to reverse

For the four-factor half fraction, flip *D* **only**:

$$D = ABC \rightarrow D = -ABC.$$

The new eight combinations complete the full 2^4 design.

An **all-factor foldover** instead reverses every factor sign.

$$(-A)(-B)(-C) = -ABC, \quad (-A)(-B)(-C)(-D) = ABCD.$$

Which defining relation changes sign? Which stays the same?

What a foldover separates

Reversing all signs changes odd-order columns and preserves even-order columns.

Alias in the first fraction	Alias after reversing all signs
$A = BC$	$A = -BC$
$A = BCD$	$A = BCD$
$AB = CD$	$AB = CD$

Combining the fractions separates main effects from two-factor interactions.

Which aliases in the table remain?

Replicate or add combinations?

Spend four more runs on...	What do we gain?
The same four settings	Precision and pure-error information. Aliases remain.
The complementary four settings	Separate brand from power-by-time. No pure-error df with one run per setting.

Which addresses the misleading brand conclusion?

For sequential runs, record phase or day and randomize within each phase.

Before running the experiment

For the popcorn study:

1. Which effects and interactions could change our recommendation?
2. Does the proposed design separate them?
3. Where will the uncertainty estimate come from?
4. What follow-up would resolve an ambiguous result?

A smaller design needs stronger assumptions. Make those assumptions explicit.

Reading: Owen, Chapter 7, especially §§7.1 and 7.3–7.7.

Appendix

Plotting the patterns: setup

```
import matplotlib.pyplot as plt

patterns = [
    ("No interaction", [60, 70], [75, 85]),
    ("Different sizes", [60, 80], [85, 90]),
    ("Opposite directions", [60, 80], [90, 70]),
]
colors = ["#232D4B", "#E57200"]
```

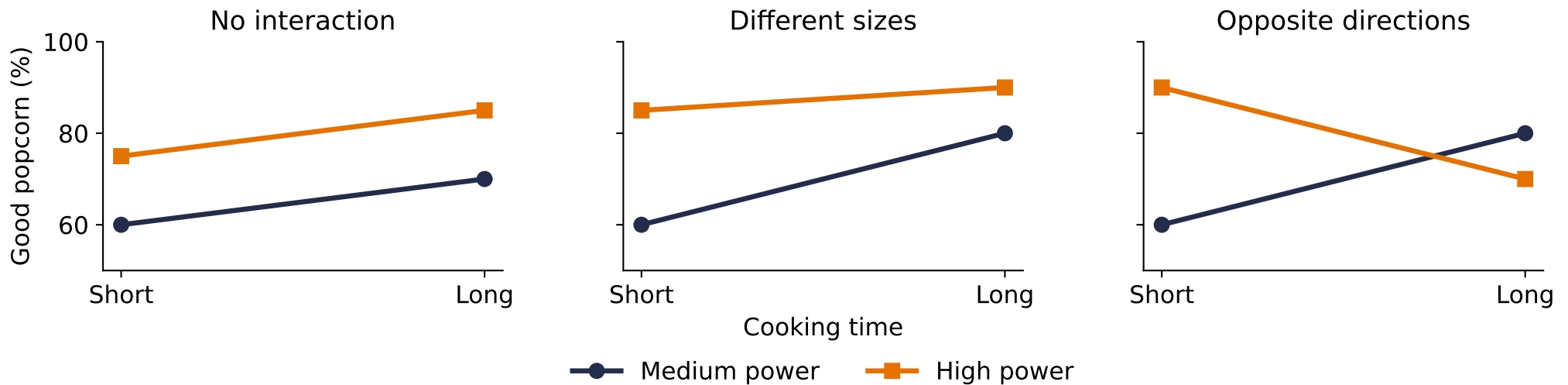
Each pair of lists gives the mean response at short and long times.

Plotting the patterns: helper

```
def interaction_panels(patterns):
    fig, axes = plt.subplots(1, 3, figsize=(11, 3.3),
                             sharey=True)
    for ax, (title, medium, high) in zip(axes, patterns):
        for y, name, color, marker in zip(
            [medium, high], ["Medium power", "High power"],
            colors, ["o", "s"]):
            ax.plot([0, 1], y, marker=marker, color=color,
                    label=name, lw=2.5, ms=7)
        ax.set(title=title, xticks=[0, 1],
               xticklabels=["Short", "Long"], ylim=(50, 100))
        ax.spines[["top", "right"]].set_visible(False)
    axes[0].set_ylabel("Good popcorn (%)")
    axes[1].set_xlabel("Cooking time")
    axes[1].legend(loc="upper center", bbox_to_anchor=(0.5, -0.3),
                  ncol=2, frameon=False)
    fig.tight_layout()
    return fig
```

Plotting the patterns: display

```
interaction_panels(patterns)
plt.show()
```



Compare the **slopes**, not just whether the lines cross.

Why interactions can be hard to detect

For a 2×2 design with n independent runs per cell and common error variance σ^2 ,

$$\text{Var} (\hat{d}_B) = \frac{\sigma^2}{n}, \quad \text{Var} (\hat{\Delta}_{AB}) = \frac{4\sigma^2}{n}.$$

But $d_{AB} = \Delta_{AB}/2$, so

$$\text{Var} (\hat{d}_{AB}) = \text{Var} (\hat{d}_B).$$

The difficulty depends on **signal relative to noise**, and on how many interactions we search for.

An interaction half as large needs four times as many runs to reach the same signal-to-noise ratio.

A contrast for every effect

In a balanced full factorial, let x_j be a nonconstant sign column.

$$\hat{d}_j = \bar{Y}_{x_j=+1} - \bar{Y}_{x_j=-1} = \frac{2}{N} \sum_{i=1}^N x_{ij} Y_i.$$

For an interaction, multiply the factor signs first.

The sign columns are **orthogonal**. With independent, equal-variance errors, the effect estimates are uncorrelated.

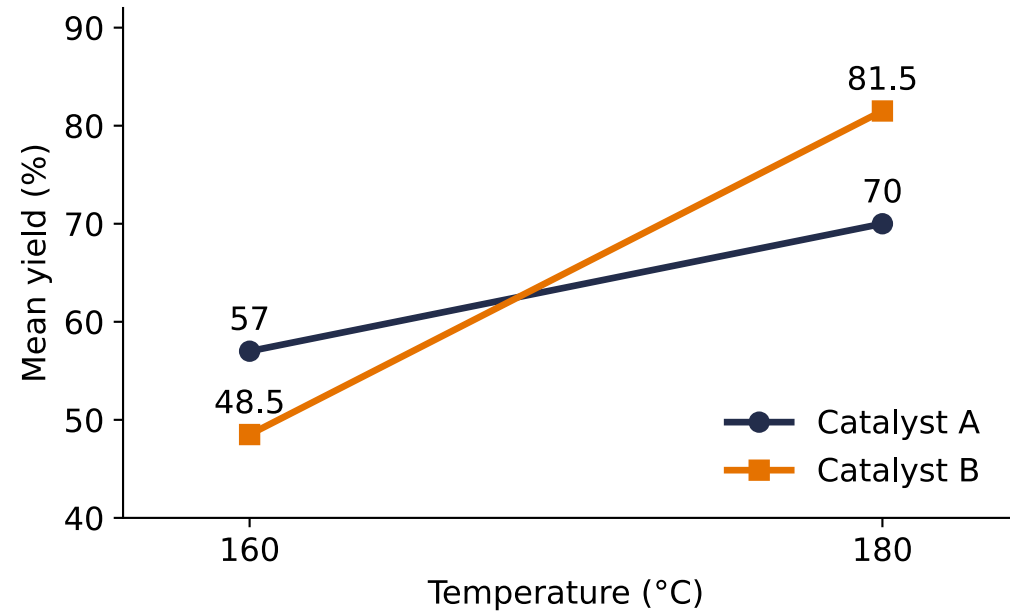
Build the AB column for the eight popcorn combinations.

Pilot-plant point plots: helper

```
def pilot_pointplot(data, ax):  
    for k, label, color, marker in [  
        (-1, "Catalyst A", "#232D4B", "o"),  
        (1, "Catalyst B", "#E57200", "s")]:  
        means = data.loc[data.K == k].groupby("T").y.mean()  
        ax.plot([160, 180], means, marker=marker,  
                color=color, lw=2.5, ms=7, label=label)  
        for x, y in zip([160, 180], means):  
            ax.annotate(f"{y:g}", (x, y), xytext=(0, 8),  
                       textcoords="offset points", ha="center")  
    ax.set(xticks=[160, 180], xlim=(156, 184), ylim=(40, 92),  
          xlabel="Temperature (°C)", ylabel="Mean yield (%)")  
    ax.spines[["top", "right"]].set_visible(False)
```

Pilot-plant point plots: display

```
fig, ax = plt.subplots(  
    figsize=(5.5, 3.5))  
pilot_pointplot(pilot, ax)  
ax.legend(frameon=False,  
          loc="lower right")  
fig.tight_layout()  
plt.show()
```



Use `pilot.loc[pilot.C == -1]` to plot concentration 20% separately.

Residual plots: code

```
fig, axes = plt.subplots(1, 2, figsize=(10, 3.2))
axes[0].scatter(fit.fittedvalues, fit.resid, color="#232D4B")
axes[0].set(xlabel="Fitted yield (%)", ylabel="Residual")
order = np.argsort(pilot.order.to_numpy())
axes[1].plot(pilot.order.iloc[order], fit.resid.iloc[order],
             "o-", color="#E57200")
axes[1].set(xlabel="Recorded run order", xticks=[1, 4, 8, 12, 16])
for ax in axes:
    ax.axhline(0, color="0.65", lw=1)
    ax.spines[["top", "right"]].set_visible(False)
fig.tight_layout()
plt.show()
```

Reading

- ▶ Owen, *A First Course in Experimental Design*: §§5.2–5.3, Chapter 6, and §§7.1, 7.4, 7.7.
- ▶ Box, Hunter, and Hunter, *Statistics for Experimenters*, 2nd ed.: §§5.4–5.10 and §§6.3–6.4, 6.8.
- ▶ Belzile, *Experimental Design and Statistical Methods*: §5.2, especially **Figure 5.1** (pp. 86–87).