NOTESKARTS

noteskarts.com

B. Pharmacy — 8th Semester

BIOSTATISTICS AND RESEARCH METHODOLOGY

UNIT 5 — Factorial Design (2^2, 2^3) | Response Surface Methodology | CCD | Optimization

Subject	Semester	Unit	Platform
BP801T	8th Semester	Unit 5 of 5	**Noteskarts.com**

UNIT 5 SYLLABUS AT A GLANCE

S.N.	Topic	Sub-topics Covered
1	**Factorial Design: Definition & Concepts**	*Definition, factors, levels, treatment combinations, effects, interactions*
2	**2^2 Factorial Design**	*2^2 design matrix, main effects, interaction, ANOVA, pharma problem with analysis*
3	**2^3 Factorial Design**	*2^3 design matrix, 7 effects, Yates algorithm, interpretation, pharma problem*
4	**Advantages of Factorial Design**	*advantages vs one-factor-at-a-time (OFAT)*
5	**Response Surface Methodology (RSM)**	*Definition, when to use, concepts, RSM polynomial model*
6	**Central Composite Design (CCD)**	*Star points, face-centered, rotatable CCD, design matrix, pharma application*
7	**Historical Design (Definitive Screening Design)**	*History, HDA (Historical Data Analysis), archival data approach*
8	**Optimization Techniques**	*Graphical, numerical, desirability function, Lagrange multipliers, Pareto front*

Factorial Design — Definition

A Factorial Design is an experimental strategy in which all possible combinations of levels of two or more factors are investigated simultaneously in a single experiment. It is the most efficient method for studying the individual effects (main effects) and combined effects (interactions) of multiple variables on a response.

◆ Fundamental Terminology

Term	Definition with Pharmaceutical Example
Factor (Variable)	An independent variable studied in the experiment. Ex: Binder concentration, Compression force, Disintegrant level
Level	A specific value or setting of a factor. Ex: Binder at Low (2%) or High (4%) = 2 levels
Treatment Combination	A specific set of factor levels for a single experimental run. Ex: (Binder=2%, Force=5 kN) = one treatment combination
Response (Y)	The measured outcome variable. Ex: Dissolution %, Tablet hardness, Disintegration time
Main Effect	The average effect of changing ONE factor from its low to high level, averaged across all levels of other factors
Interaction Effect	The effect of ONE factor that DEPENDS on the level of another factor. If interaction is significant, main effects alone cannot tell the full story
Replicate	Repeating the entire experiment or individual runs to estimate experimental error
Randomization	Running experimental trials in random order to prevent systematic bias
Run	A single experimental trial at a specific treatment combination
Effect Estimate	Quantitative value showing how much a factor or interaction changes the response

◆ Notation System for Factorial Designs

A 2k design notation means: 2 levels per factor, k = number of factors, Total runs = 2k (without replication)

Design	Factors (k)	Levels/Factor	Total Runs	Pharma Application
2^2	2	2 (Low, High)	4	*Optimizing 2 formulation variables*

2^3	3	2 (Low, High)	8	*Screening 3 formulation factors*
2^4	4	2 (Low, High)	16	*Early-phase development: 4 variables*
2^5	5	2 (Low, High)	32	*Preformulation: 5 critical variables*
3^2	2	3 (Low,Mid,High)	9	*Curvature detection — needs center points*
3^3	3	3	27	*Full quadratic model — large design*

2^2 Factorial Design

The 2^2 (two-squared) factorial design studies 2 factors (A and B), each at 2 levels (Low = −1 and High = +1). It requires 4 treatment combinations (runs) and can estimate 3 effects: Main Effect A, Main Effect B, and Interaction AB.

◆ The 2^2 Design Matrix

Pharmaceutical Study: Effect of Factor A (Binder Concentration, %) and Factor B (Compression Force, kN) on Tablet Dissolution (%)

Factor Levels:

- Factor A (Binder%): Low (−) = 2%, High (+) = 6%
- Factor B (Force kN): Low (−) = 5 kN, High (+) = 15 kN

Run	A (Binder%)	B (Force kN)	AB (Interaction)	y (Dissolution%)	Treatment Label
1	− (2%)	− (5 kN)	+	72	*(1) — both low*
2	+ (6%)	− (5 kN)	−	80	*a — A high, B low*
3	− (2%)	+ (15 kN)	−	81	*b — A low, B high*
4	+ (6%)	+ (15 kN)	+	91	*ab — both high*

◆ Calculating Main Effects and Interaction in 2^2

Using effect estimates (contrast coefficients method):

Main Effect A	$A = [(a + ab)/2] - [(1) + b)/2]$	High A avg minus Low A avg
Main Effect B	$B = [(b + ab)/2] - [(1) + a)/2]$	High B avg minus Low B avg
Interaction AB	$AB = [(1) + ab)/2] - [(a + b)/2]$	Both-same avg minus cross avg

◆ ANOVA for 2^2 Design

Sum of Squares for each effect in a 2^2 design with n replicates per cell:

SS for any Effect	$SS = n \times (\text{Effect Estimate})^2 / 4$	n = number of replicates per treatment

With n=1 replicate: $SS_A = 1 \times (9.0)^2 / 4 = 81/4 = 20.25$; $SS_B = 100/4 = 25.0$; $SS_{AB} = 1/4 = 0.25$

Source	SS	df	MS	F	Decision (α=0.05)
A (Binder%)	20.25	1	20.25	—	*Significant if F>F_crit*
B (Force kN)	25.00	1	25.00	—	*Significant if F>F_crit*
AB (Interaction)	0.25	1	0.25	—	*Not significant (small SS)*
Error	—	n-4	MSE	—	—
Total (corrected)	—	n-1	—	—	—

Insight: With n=1 replicate and no error term in unreplicated 2^2 design, significance is assessed using Normal Probability Plot of Effects (Daniel's Plot) or Lenth's Method — effects falling off the straight line are significant.

2^3 Factorial Design

The 2^3 (two-cubed) factorial design studies 3 factors (A, B, C) each at 2 levels (Low = −1, High = +1). It requires 8 treatment combinations (runs) and can estimate 7 effects: 3 main effects (A, B, C) and 4 interactions (AB, AC, BC, ABC).

◆ The 2^3 Design Matrix (Standard Order)

Pharmaceutical Study: Factors = A (Binder%), B (Disintegrant%), C (Lubricant%); Response = Tablet Dissolution (%)

Factor Levels:

- A (Binder%): Low (−) = 2%, High (+) = 5%
- B (Disintegrant%): Low (−) = 2%, High (+) = 6%
- C (Lubricant%): Low (−) = 0.5%, High (+) = 2%

Run	A	B	C	AB	AC	BC	ABC	y (Dissolution%)
1	−	−	−	+	+	+	−	68
2	+	−	−	−	−	+	+	76
3	−	+	−	−	+	−	+	74
4	+	+	−	+	−	−	−	83
5	−	−	+	+	−	−	+	70
6	+	−	+	−	+	−	−	79
7	−	+	+	−	−	+	−	78
8	+	+	+	+	+	+	+	88

◆ Effect Estimation Formulas for 2^3 Design

For a 2^3 design with n replicates, each effect is calculated as (Contrast / 4n):

General Effect Formula	Effect = Contrast / (n × 2^(k-1))	k=3, so divide contrast by 4n

Contrast Formulas (using +/− signs from design matrix):

Contrast A	**A-contrast =** $-y_1+y_2-y_3+y_4-y_5+y_6-y_7+y_8$	Sum: (+) runs minus (−) runs for A
Contrast AB	**AB-contrast =** $+y_1-y_2-y_3+y_4+y_5-y_6-y_7+y_8$	Use AB column signs
Contrast ABC	**ABC-contrast =** $-y_1+y_2+y_3-y_4+y_5-y_6-y_7+y_8$	Three-factor interaction

◆ Yates Algorithm for Effect Calculation

Yates algorithm is a systematic method for computing all effects and contrasts in a 2k design without individually applying contrast formulas to each effect.

Steps of Yates Algorithm:

1. Write response values in STANDARD ORDER: (1), a, b, ab, c, ac, bc, abc
2. In Column 1: Add adjacent pairs for top half; subtract (upper minus lower) for bottom half
3. Repeat Column 2 using Column 1 values, same procedure
4. Repeat Column 3 (= k columns total for 2k design)
5. Final column = Contrasts; Divide by 2^(k-1) for effects; by 2k for grand mean

Run (Std Order)	y	Column 1	Column 2	Column 3 (Contrast)	Effect	Identity
(1)	68	68+76=144	144+157=301	301+316=617	617/8=77.1	*Grand Mean*
a	76	74+83=157	301+?=?	A-contrast=36	36/4=9.0	*Effect A*
b	74	68+70=138	AB=2	B-contrast=30	30/4=7.5	*Effect B*
ab	83	79+88=167	?	AB-contrast=2	2/4=0.5	*Effect AB*
c	70	76-68=8	C-contrast=14	C-contrast=14	14/4=3.5	*Effect C*
ac	79	83-74=9	AC=2	AC-contrast=2	2/4=0.5	*Effect AC*
bc	78	79-70=9	BC=4	BC-contrast=4	4/4=1.0	*Effect BC*
abc	88	88-79=9	ABC=0	ABC-contrast=0	0/4=0.0	*Effect ABC*

Advantages of Factorial Design

Factorial designs are vastly superior to the traditional One-Factor-At-A-Time (OFAT) approach. The advantages are both statistical (efficiency, power) and practical (detecting interactions, broader conclusions).

◆ Factorial Design vs One-Factor-At-A-Time (OFAT)

Feature	Factorial Design	OFAT (Traditional)
Number of experiments	Fewer runs for same information	More runs needed for same info
Interaction detection	YES — can detect and estimate interactions	NO — completely misses interactions
Effect estimation	Each effect estimated using ALL data	Each effect uses only subset of data
Statistical efficiency	High — each effect estimated with n×2^(k-1) observations	Low — each effect estimated with fewer observations
Validity of conclusions	Broader — conclusions valid across all factor levels	Narrow — valid only at fixed 'one factor' baseline
Confounding of effects	Controlled — orthogonal design	Possible confounding — not controlled
Optimal region discovery	Systematic approach to optimization	May miss optimum if interactions exist

◆ Advantages of Factorial Design

No.	Advantage	Explanation & Pharmaceutical Significance
1	**Interaction Detection**	The most critical advantage. Factorial design can detect whether the effect of Factor A depends on the level of Factor B (interaction). OFAT completely fails to detect interactions. Ex: Binder effect on dissolution may differ at different compression forces.
2	**Statistical Efficiency**	Each effect is estimated from ALL experimental runs, not just a subset. For a 2^3 design (8 runs), each main effect uses all 8 observations — equivalent to running 8 single-factor experiments simultaneously.

3	**Simultaneous Optimization**	Allows optimization of multiple response variables at the same time. Ex: Maximize dissolution AND hardness simultaneously using the same 8-run experiment.
4	**Reduced Experimentation**	Achieves more information with fewer total experiments compared to OFAT. A 2^3 factorial (8 runs) gives 7 independent estimates; OFAT needs 7+ experiments for the same, each studying only one factor.
5	**Broad Validity**	Conclusions apply across all combinations of factor levels, not just a single baseline. The optimum found applies under real manufacturing conditions where multiple factors vary simultaneously.
6	**Estimation of Curvature**	Adding center points to factorial designs allows testing for curvature (non-linearity) — signals need for RSM follow-up. Pure OFAT cannot test curvature in multiple dimensions.
7	**Randomization & Error Control**	Runs are randomized, preventing systematic bias. Blocking removes nuisance variable effects. OFAT experiments often lack randomization and may be confounded by time trends.
8	**Foundation for RSM**	Factorial screening identifies significant factors, which become inputs for Response Surface designs (CCD, BBD). This two-stage approach (screen → optimize) is the cornerstone of ICH Q8 QbD.
9	**Effect Hierarchy Principle**	Factorial analysis ranks all effects by magnitude — instantly reveals which factors and interactions matter most. Ex: B (10%) > A (9%) > C (3.5%) → Focus resources on B and A.
10	**Regulatory Acceptance**	ICH Q8 (Pharmaceutical Development), FDA PAT guidance, and ICH Q11 explicitly endorse factorial and RSM designs for formulation development and Design Space establishment.

Response Surface Methodology (RSM)

Response Surface Methodology (RSM) is a collection of mathematical and statistical techniques used to model and analyze problems in which a response of interest is influenced by several quantitative variables (factors), with the ultimate goal of OPTIMIZING the response. RSM moves BEYOND screening (factorial) to OPTIMIZATION — fitting a curved (second-order polynomial) surface to find the precise optimum conditions.

◆ Use RSM — The Two-Stage Strategy

Stage	Design Type	Purpose
Stage 1: Screening	**Full/Fractional Factorial (2k, 2k-p)**	Identify which factors are statistically significant

		(eliminate non-significant factors)
Stage 2: Optimization	**RSM Design (CCD, BBD, HDA)**	Fit second-order model; find optimal region; establish Design Space
Stage 3: Confirmation	**Verification experiments at optimum**	Confirm predicted optimum is achievable in practice (validation)

◆ RSM Second-Order Polynomial Model

RSM fits a full second-order (quadratic) polynomial model that includes linear, interaction, and squared terms:

RSM Model (2 factors)	$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_{12}X_1X_2 + \beta_{11}X_1^2 + \beta_{22}X_2^2$	β_0=intercept, β_1,β_2=linear, β_{12}=interaction, β_{11},β_{22}=quadratic
RSM Model (3 factors)	$Y = \beta_0 + \Sigma\beta_iX_i + \Sigma\beta_{ii}X_i^2 + \Sigma\Sigma\beta_{ij}X_iX_j$	Full second-order — 10 coefficients for 3 factors

Model Term	Description & Pharmaceutical Meaning
β_0 (Intercept)	Overall mean response at center point. Ex: Dissolution at mid-level of all factors = 82%
β_1, β_2 (Linear terms)	Direction and magnitude of effect of each factor. Ex: β_1=+6.5 means each unit increase in X1 adds 6.5% dissolution
β_{12} (Interaction term)	Combined effect of X1 and X2 together. Tilts the response surface. Significant interaction means saddle or ridge surface
β_{11}, β_{22} (Quadratic terms)	Curvature in the response. Negative β values indicate a maximum (optimum) exists. Positive = minimum exists

◆ Types of Response Surfaces

Surface Type	Characteristics	When Encountered
Mound (Maximum)	Convex surface with clear peak; β_{11} and β_{22} both negative; stationary point is maximum	*Drug release maximum — optimal formulation clearly identified*
Bowl (Minimum)	Concave surface with clear valley; β_{11} and β_{22} both positive; stationary point is minimum	*Minimizing particle size, residual moisture, impurity content*

Saddle Point	Neither max nor min; curvatures of opposite sign; ridge in one direction, valley in another	*Complex interactions; optimization requires canonical analysis*
Ridge	Maximum exists along a ridge line (multiple optima); one eigenvalue near zero	*Formulations with strong positive interaction between factors*
Stationary Ridge	Response plateaus across a range of factor values; robust zone identified	*Design Space establishment — confirms operating range*

Central Composite Design (CCD)

The Central Composite Design (CCD) is the most widely used RSM design.
It consists of three types of experimental points combined to fit a full second-order polynomial model efficiently:

1. Factorial Points (±1) — from the 2k or 2k-p factorial base design
2. Axial / Star Points (±α) — located on the axes at distance α from center
1. 3. Center Points (0,0) — replicated runs at the center for error estimation and curvature testing

◆ CCD Structure

Total CCD Runs	$N = 2k + 2k + n_0$	2k=factorial, 2k=axial, n_0=center points

Point Type	Coded Levels	Count (2-factor)	Purpose
Factorial Points	±1	4 (2^2)	Estimate main effects and 2-factor interactions; same as 2k design
Axial (Star) Points	±α	4 (2×k=4)	Estimate quadratic (curvature) terms β_{11} and β_{22}
Center Points	0, 0	3–6 (n_0)	Estimate pure error; check lack of fit; detect curvature

◆ Types of CCD Based on α Value

CCD Type	α Value	Axial Point Location	Pharmaceutical Use
Face-Centered CCD (FCCCD)	α = 1.0	On faces of the cube (within experimental region)	*When cannot exceed factor limits (e.g., drug% bounded by solubility)*

Circumscribed CCD (CCC)	**α = 2^(k/4) ≈ 1.414 (k=2)**	Outside the factorial cube (requires extrapolation)	*When wider factor ranges are feasible; most common for optimization*
Inscribed CCD (CCI)	**α < 1.0**	Inside the cube	*When extreme factor combinations are physically impossible*
Rotatable CCD	**α = 2^(k/4)**	Outside cube	*Prediction variance equal at all points equidistant from center — gold standard for RSM*

◆ 2-Factor CCD Design Matrix (Pharmaceutical Example)

Study: Optimize Tablet Dissolution (Y) by varying X1 = Binder% and X2 = Compression Force. α = 1.414 (Rotatable CCD)

Run	X1 (Coded)	X2 (Coded)	Point Type	Actual Levels	y (Dissolution %)
1	−1	−1	Factorial	*Binder 2%, Force 5 kN*	72
2	+1	−1	Factorial	*Binder 6%, Force 5 kN*	80
3	−1	+1	Factorial	*Binder 2%, Force 15 kN*	81
4	+1	+1	Factorial	*Binder 6%, Force 15 kN*	91
5	−1.414	0	Axial (Star)	*Binder 1.17%, Force 10 kN*	74
6	+1.414	0	Axial (Star)	*Binder 6.83%, Force 10 kN*	89
7	0	−1.414	Axial (Star)	*Binder 4%, Force 2.93 kN*	69
8	0	+1.414	Axial (Star)	*Binder 4%, Force 17.07 kN*	87
9	0	0	Center	*Binder 4%, Force 10 kN*	84
10	0	0	Center	*Binder 4%, Force 10 kN*	83
11	0	0	Center	*Binder 4%, Force 10 kN*	85

Color Code	Point Type	Role in Model Fitting
Teal background (Runs 1–4)	Factorial Points	*Estimate linear & interaction effects (β_1, β_2, β_{12})*
Purple background (Runs 5–8)	Axial / Star Points	*Estimate quadratic curvature (β_{11}, β_{22})*
Orange background (Runs 9–11)	Center Points (n_0=3)	*Pure error estimation; lack-of-fit test; curvature detection*

◆ Box-Behnken Design (BBD) — Alternative to CCD

The Box-Behnken Design is a rotatable or near-rotatable RSM design that does NOT include factorial corner points. All design points lie on a sphere (or cube edges/midpoints). It requires fewer runs than CCD for 3+ factors and avoids extreme treatment combinations.

Feature	CCD	BBD	Choice Guidance
Design Points	Factorial + Axial + Center	Edge midpoints + Center	*CCD: when factor extremes are feasible; BBD: when corners are problematic*
Factor levels	5 levels ($-\alpha, -1, 0, +1, +\alpha$)	3 levels (−1, 0, +1)	*BBD better for 3-level designs*
Runs (3 factors)	20 (with n_0=6)	15 (n_0=3)	*BBD more economical for 3 factors*
Runs (4 factors)	30	27	*CCD becomes more efficient for 4+ factors*
Extreme combinations	Includes corner points (all high or all low)	Avoids corner points	*BBD preferred if extreme combinations are physically unsafe*
Pharma application	Tablet optimization (2 factors)	Nanoparticle size optimization (3 factors)	*BBD widely used for 3-variable optimization in pharma*

Historical Design (Historical Data Analysis)

A Historical Design (also called Historical Data Analysis, HDA, or Definitive Screening Design in modern literature) refers to the use of existing, previously collected data to build and refine statistical models. Rather than designing new experiments, historical design extracts maximum information from data already available in databases, batch records, or archival studies.

◆ Definition and Concept

Historical Design involves the systematic analysis of existing experimental data (from past development batches, stability studies, manufacturing records) using regression modeling and factorial analysis techniques to:

- Identify critical process parameters (CPPs) and critical quality attributes (CQAs)
- Establish relationships between process variables and product quality
- Build predictive models without conducting new experiments
- Support retrospective DoE analysis for regulatory submissions
- Validate process understanding for process validation (PV Stage 1)

Aspect	Description
Data Source	Manufacturing batch records, clinical databases, stability study databases, literature data, archival lab notebooks
Design Type	Observational (retrospective) — not prospectively designed; no control over factor levels
Statistical Method	Multiple regression, ANOVA, principal component analysis (PCA), partial least squares (PLS)
Pharma Application	Process validation Stage 1 (process design), continued process verification (CPV), site transfer validation
Advantages	No new experiments needed — saves cost and time; uses real manufacturing data; compliant with real-world variability
Limitations	No control over factor ranges; potential confounding; multicollinearity common; observational not causal
Regulatory Context	FDA Guidance for Process Validation (2011); ICH Q10 Pharmaceutical Quality System; ICH Q8(R2)

◆ Definitive Screening Design (DSD) — Modern Historical Approach

The Definitive Screening Design (DSD) was introduced by Jones and Nachtsheim (2011) as an alternative to classical factorial screening. DSD can estimate main effects, two-factor interactions, AND quadratic effects with very few runs — often replacing both factorial screening AND RSM in a single economical design.

Properties of DSD:

- Three levels per factor (−1, 0, +1) — unlike two-level factorials
- Minimum runs: 2k+1 (where k = number of factors) — extremely economical
- Main effects are fully independent (orthogonal) of each other
- Main effects are orthogonal to two-factor interactions and quadratic effects
- No complete confounding of two-factor interactions with each other
- Can detect non-linearity (curvature) even in the screening phase

Feature	DSD	2k Factorial	CCD
Levels per factor	3 (−1, 0, +1)	2 (−1, +1)	5 (−α to +α)
Min runs (k=6)	13	64	54
Detects interactions	Partially	Fully	Fully
Detects curvature	Yes	No (needs center pts)	Yes
Pharma application	Early-phase screening + optimization	Screening only	Optimization only

Optimization Techniques in Pharmaceutical Formulation

Optimization is the process of finding the best possible set of factor (input variable) levels that achieves the desired response(s). In pharmaceutical development, optimization is the final step after RSM model fitting, using the fitted model to identify optimal formulation conditions.

◆ Graphical Optimization

Graphical optimization uses visual overlaid contour plots to identify the feasible region where all response specifications are simultaneously met.

Steps in Graphical Optimization:

1. Fit RSM model for each response (e.g., Dissolution ≥ 80%, Hardness 4–8 kg/cm², DT ≤ 15 min)
2. Generate contour plot for each response — shade the ACCEPTABLE region for each
3. OVERLAY all contour plots on the same graph
4. The OVERLAPPING white/unshaded region = feasible region where ALL specifications are met
5. Any point in the feasible region is an acceptable formulation — pick the center for robustness

◆ Numerical Optimization — Desirability Function

The Desirability Function (Derringer & Suich, 1980) converts multiple response optimization into a single numerical optimization problem. Each response is converted to a 'desirability' value d (0 to 1), and the overall desirability D = geometric mean of all d_i is maximized.

Overall Desirability D	$D = (d_1 \times d_2 \times d_3 \times ... \times d_n)^{(1/n)}$	D=0 means at least one response unacceptable; D=1 is ideal

◆ Other Optimization Techniques

Technique	Description & Pharmaceutical Application
Canonical Analysis	Mathematical transformation of second-order RSM model to identify stationary point (maximum, minimum, saddle). Uses eigenvalues of the matrix of quadratic coefficients to characterize the nature of the optimum.
Ridge Analysis	When stationary point is outside experimental region (extrapolation), ridge analysis finds the optimum on a sequence of spheres of increasing radius from center. Used when CCD analysis doesn't find clear optimum within design space.
Lagrange Multipliers	Mathematical optimization technique — find extremum of objective function subject to equality constraints. Ex: Maximize dissolution (Y1) subject to fixed total excipient weight (X1+X2+X3 = constant). Used in mixture designs.
Grid Search (Brute Force)	Systematically evaluate the fitted model at thousands of grid points across the factor space. Identify the grid point with highest predicted desirability. Simple but computationally intensive. Used in Design-Expert software.
Pareto Front Optimization	Multi-objective optimization finding the set of solutions where improving one response necessarily worsens another. Generates a 'Pareto front' of non-dominated solutions for decision-making. Used in bioequivalence formulation optimization.
Simplex Optimization	Sequential optimization technique moving through factor space using simplex (triangle/tetrahedron) of experimental points. Reflects worst point across best face. Useful for continuous optimization during manufacturing.
Artificial Neural Networks (ANN)	Non-parametric machine learning approach to model complex non-linear drug-formulation relationships. No model assumptions needed. Used when RSM models have poor fit or relationships are highly non-linear.
Particle Swarm Optimization (PSO)	Bio-inspired heuristic algorithm mimicking bird flocking. Multiple 'particles' explore factor space; best global solution emerges. Used for complex multi-response optimization in nanoparticle formulation.

◆ Optimization Workflow in Pharmaceutical QbD (ICH Q8)

Step	Activity	Tools & Outputs
1	**Define Target Product Profile (TPP)**	Desired quality attributes: dissolution ≥ 85%, hardness 4–8 kg/cm², etc.

2	**Identify Critical Quality Attributes (CQAs)**	Risk assessment (FMEA, Ishikawa); QbD target attributes identified
3	**Screening Experiment**	2k or DSD factorial; rank factors by significance — identify CPPs
4	**RSM Design**	CCD or BBD with significant factors only; fit second-order model
5	**Model Fitting & Diagnostics**	ANOVA, R^2, Lack-of-Fit, residual plots — confirm model adequacy
6	**Optimization**	Desirability function or graphical overlay — find optimal factor settings
7	**Design Space Definition**	Region where all CQAs meet specifications — regulatory acceptable range
8	**Control Strategy**	Process controls, specifications, PAR (Proven Acceptable Ranges) defined
9	**Confirmation Runs**	3–5 validation batches at optimal conditions — verify predicted values
10	**Regulatory Submission**	Include DoE, RSM, Design Space in CTD Section 3.2.P.2 (Pharmaceutical Development)

◆ Expected Exam Questions — Unit 5

Q	Question	Marks
1	Define Factorial Design. What is a 2k design? Give the notation, number of runs and effects for 2^2 and 2^3.	5
2	Construct a 2^2 factorial design for a pharmaceutical experiment. Calculate all main effects and interaction with given data.	10
3	Explain the 2^3 factorial design with design matrix in standard order. Calculate all 7 effects using given dissolution data.	10
4	What is the Yates algorithm? Apply it to a 2^3 design data to calculate all effects.	10
5	List and explain 5 advantages of factorial design over the one-factor-at-a-time (OFAT) approach.	5
6	What is Response Surface Methodology (RSM)? When is it used? Write the second-order polynomial model and explain each term.	5–10
7	What is a Central Composite Design (CCD)? Explain its three types of points with diagram. Calculate total runs for k=2 and k=3.	5–10
8	Differentiate between Face-Centered CCD, Circumscribed CCD, and Rotatable CCD with pharma examples.	5

9	Compare CCD and Box-Behnken Design (BBD). When would you prefer BBD over CCD?	5
10	What is a Historical Design? How is historical data used in pharmaceutical process optimization and process validation?	5
11	What is a Definitive Screening Design (DSD)? How does it differ from classical 2k factorial designs?	5
12	Explain graphical optimization (overlay contour plot) for pharmaceutical formulation optimization.	5
13	What is the Desirability Function? How is overall desirability D calculated? Solve a problem with 3 responses.	5–10
14	Write a short note on: (a) Ridge Analysis (b) Canonical Analysis (c) Pareto Front Optimization.	5
15	Describe the complete QbD optimization workflow (ICH Q8) from target product profile to regulatory submission.	10
16	A 2^2 design gave: (1)=72, a=80, b=81, ab=91. Calculate effects A, B, AB. Identify the most important factor.	5

BIOSTATISTICS & RESEARCH METHODOLOGY — ALL 5 UNITS COMPLETE!

Unit 1: Introduction, Central Tendency, Dispersion, Correlation
Unit 2: Regression, Probability, Hypothesis Testing, Parametric Tests (t-test, ANOVA, LSD)
Unit 3: Non-Parametric Tests, Research Design, Graphs, Clinical Trials
Unit 4: Blocking & Confounding, Regression Modeling, Statistical Software (Excel, SPSS, MINITAB, R)
Unit 5: Factorial Design (2^2, 2^3), RSM, CCD, Historical Design, Optimization

NOTESKARTS.COM | YouTube @Noteskarts | tests.noteskarts.com