NOTESKARTS

noteskarts.com

B. Pharmacy — 8th Semester

BIOSTATISTICS AND RESEARCH METHODOLOGY

UNITS 4 — Blocking | Confounding | Regression Modeling | Statistical Software

Subject Code	Semester	Units	Platform
BP801T	8th Semester	Units 4 of 5	**Noteskarts.com**

UNITS 4 SYLLABUS AT A GLANCE

Sr.No.	Topic	Sub-topics Covered
1	**Blocking & Confounding in 2k Factorials**	*Confounding principles, 2^2 and 2^3 designs, blocks, alias structure, defining relations*
2	**Regression Modeling**	*Simple & Multiple Regression, ANOVA for regression, t-test for coefficients, R^2, Adjusted R^2*
3	**Statistical Software & Practical Components**	*Excel, SPSS, MINITAB, Design-Expert (DoE), R — Industrial & Clinical Trial Applications*

Blocking & Confounding in Two-Level Factorial Designs

In pharmaceutical experimentation, it is often impossible to complete all runs of a factorial experiment under identical homogeneous conditions. Blocking and confounding are techniques that manage this variability while preserving the ability to estimate important effects.

◆ Two-Level Factorial Design (2k Design)

A 2k factorial design studies k factors each at exactly 2 levels: Low (−1 or −) and High (+1 or +). Total runs = 2k.

Design	Factors (k)	Total Runs	Main Effects	Interactions
2^2	2	4	2 (A, B)	1 (AB)
2^3	3	8	3 (A, B, C)	4 (AB, AC, BC, ABC)
2^4	4	16	4 (A, B, C, D)	11 (2-fi, 3-fi, 4-fi)
2^5	5	32	5	26 interactions

◆ What is Blocking?

Blocking is the arrangement of experimental runs into groups (blocks) of homogeneous conditions to eliminate the effect of a nuisance variable. Within each block, conditions are as uniform as possible. Example: If a 2^3 design (8 runs) must be conducted over 2 days (days = blocks of 4 runs each), blocking assigns runs to days so that the DAY effect does not bias the factor estimates.

Why Blocking is Needed in Pharmaceutical Research:

- Different analytical batches (HPLC runs on different days)
- Different operators conducting experiments
- Different raw material lots used across experiment days
- Multiple pieces of equipment with slight calibration differences
- Temperature or humidity variation across experimental sessions

Key Principle:

Blocks should be arranged so that the block effects are orthogonal (independent) to the factor effects. This ensures that the block-to-block variability is separated from treatment effects and does not inflate experimental error.

◆ What is Confounding?

Confounding occurs when a factorial effect (main effect or interaction) is deliberately made indistinguishable from a block effect. The confounded effect cannot be estimated separately from the block effect — they are 'mixed together' or 'aliased'. This is an intentional trade-off: we sacrifice the ability to estimate a high-order interaction (which is assumed negligible) in order to run the experiment in blocks.

Confounding Terminology:

Term	Definition
Defining Relation (I)	The effect(s) confounded with the mean (identity I). Ex: I = ABC means ABC is confounded with blocks
Alias	Two effects that are confounded with each other and cannot be estimated separately
Alias Structure	Complete list of all aliased pairs in the design
Generator	The high-order effect chosen to confound with blocks. Usually the highest-order interaction
Resolution	Measure of the design's ability to separate effects. Higher resolution = better estimability

◆ Confounding in a 2^2 Design (2 Blocks of 2 Runs Each)

A 2^2 design has 4 runs with factors A, B, and interaction AB. If we must run the experiment in 2 blocks of 2, we confound the AB interaction with blocks.

Design Matrix for 2^2 (A = Binder%, B = Compression Force; Response = Dissolution %):

Run	A (Binder%)	B (Force kN)	AB (Interaction)	Block	Dissolution (%)
1	− (Low: 2%)	− (Low: 5 kN)	+ *(AB=−×−=+)*	Block 1	78
2	+ (High: 4%)	− (Low: 5 kN)	− *(AB=+×−=−)*	Block 2	84
3	− (Low: 2%)	+ (High: 10 kN)	− *(AB=−×+=−)*	Block 2	82
4	+ (High: 4%)	+ (High: 10 kN)	+ *(AB=+×+=+)*	Block 1	90

Block 1 = {Run 1, Run 4}: Runs where AB = + (conducted on Day 1) Block 2 = {Run 2, Run 3}: Runs where AB = − (conducted on Day 2) Defining Relation: I = AB → AB interaction is COMPLETELY CONFOUNDED with blocks.
Consequence: We CANNOT separately estimate the AB interaction and the Day effect.
Assumption: AB interaction is negligible, so this is an acceptable sacrifice.
Estimable: Main effects A and B are unconfounded and can be cleanly estimated.

Calculating Main Effects:

Effect A	$A = [(y_3+y_4)/2] - [(y_1+y_2)/2]$	High A average − Low A average
Effect B	$B = [(y_2+y_4)/2] - [(y_1+y_3)/2]$	High B average − Low B average

🔑 2^2 Blocked Design — Calculation

Runs: y_1=78, y_2=84, y_3=82, y_4=90 I = AB (AB confounded with blocks) Effect A = [(82+90)/2] − [(78+84)/2] = 86 − 81 = +5.0%
Interpretation: Increasing Binder% from 2% to 4% increases dissolution by 5.0% Effect B = [(84+90)/2] − [(78+82)/2] = 87 − 80 = +7.0%
Interpretation: Increasing Compression Force from 5 to 10 kN increases dissolution by 7.0% AB = Cannot be estimated — confounded with block (Day) effect.
Conclusion: Both A and B significantly improve dissolution. AB interaction cannot be evaluated.

◆ Confounding in a 2^3 Design (2 Blocks of 4 Runs Each)

A 2^3 design has 8 runs with factors A, B, C. If we must run in 2 blocks of 4, the highest-order interaction ABC is confounded with blocks.

Pharmaceutical Example: A = Binder%, B = Disintegrant%, C = Lubricant% → Response: Tablet Dissolution (%)

Run	A	B	C	ABC	Block	Dissolution %
1	−	−	−	−	Block 2	72
2	+	−	−	+	Block 1	80
3	−	+	−	+	Block 1	78
4	+	+	−	−	Block 2	86
5	−	−	+	+	Block 1	75
6	+	−	+	−	Block 2	83
7	−	+	+	−	Block 2	81
8	+	+	+	+	Block 1	90

Defining Relation: I = ABC Block 1 (ABC = +): Runs 2, 3, 5, 8 → Conducted on Day 1
Block 2 (ABC = −): Runs 1, 4, 6, 7 → Conducted on Day 2
Estimable effects: A, B, C, AB, AC, BC (all 6 effects clean) Confounded: ABC (3-factor interaction — assumed negligible in practice)
Rule: Always confound the HIGHEST ORDER interaction with blocks to preserve estimability of lower-order (more important) effects.

◆ Partial Confounding (Multiple Blocks, Replication)

When an experiment is replicated across multiple block sets, different effects can be confounded in each replicate. This allows recovery (estimation) of all effects across replicates — called Partial Confounding.

Replicate / Block Set	Effect Confounded with Blocks	Effects Estimable from This Rep
Replicate I (Blocks 1,2)	**ABC**	*A, B, C, AB, AC, BC — estimable with full info*
Replicate II (Blocks 3,4)	**AB**	*A, B, C, AC, BC, ABC — AB estimated from Rep I only*
Replicate III (Blocks 5,6)	**AC**	*A, B, C, AB, BC, ABC — AC estimated from Reps I & II*

◆ 2k Fractional Factorial Designs (Screening Designs)

When k is large (e.g., k=5, giving 32 runs), full factorial designs are too expensive. Fractional Factorial Designs run only a FRACTION (1/2, 1/4) of the full factorial, sacrificing estimation of high-order interactions to screen many factors efficiently.

Notation	Fraction	Runs	Resolution	Pharma Application
2^{5-1}	Half fraction	**16**	V — no 2fi aliased with 2fi	*Formulation optimization: 5 factors*
2^{5-2}	Quarter fraction	**8**	III — main effects aliased with 2fi	*Initial screening: 5 factors, 8 runs*
2^{6-2}	Quarter fraction	**16**	IV — 2fi aliased with 2fi	*Stability study: 6 variables*
2^{7-4}	1/16 fraction	**8**	III — Plackett-Burman type	*High-throughput screening: 7 factors*

Resolution Guide:

- Resolution III: No main effect is aliased with another main effect, but main effects ARE aliased with 2-factor interactions (2fi). Use only for screening.
- Resolution IV: No main effect aliased with 2fi, but 2fi aliased with 2fi. Compromise design.
- Resolution V: No main effect or 2fi aliased with each other. High-quality estimation. Preferred for optimization.

Regression Modeling — Hypothesis Testing

Hypothesis testing in regression determines whether the regression model as a whole, and each individual predictor, are statistically significant. This goes beyond just fitting a line — it validates whether the model has real predictive power.

◆ Simple Linear Regression — Model & Assumptions

Simple Linear Regression models the relationship between one dependent variable (Y) and one independent variable (X):

SLR Model	$Y = \beta_0 + \beta_1 X + \varepsilon$	β_0=intercept, β_1=slope, ε=random error ~$N(0,\sigma^2)$

Assumptions of Linear Regression (LINE):

- Linearity: True relationship between X and Y is linear
- Independence: Error terms are independent of each other
- Normality: Errors are normally distributed: $\varepsilon \sim N(0, \sigma^2)$
- Equal Variance (Homoscedasticity): Variance of errors is constant for all values of X

◆ Hypothesis Testing in Simple Linear Regression

Two levels of hypothesis testing exist in regression:

▸ A) Test for Significance of the Overall Model (F-test / ANOVA for Regression)

Tests whether the regression model explains a significant portion of the variation in Y. H_0: $\beta_1 = 0$ (X does NOT predict Y; model is useless) H_1: $\beta_1 \neq 0$ (X predicts Y; model is significant)

Source	SS	df	MS	F
Regression (Model)	SSR = $\Sigma(\hat{y}_i - \bar{y})^2$	p (= no. of predictors)	MSR = SSR/p	**MSR/MSE**
Error (Residual)	SSE = $\Sigma(y_i - \hat{y}_i)^2$	n−p−1	MSE = SSE/(n−p−1)	—
Total	SST = $\Sigma(y_i - \bar{y})^2$	n−1	—	—

Decision: If F_calc > F_critical(p, n−p−1, α) → Reject H_0 → Model is significant

▸ B) Test for Significance of Individual Regression Coefficients (t-test)

Tests whether each individual predictor (X) contributes significantly to explaining Y, after accounting for other predictors. H_0: $\beta_i = 0$ (X_i does not contribute) H_1: $\beta_i \neq 0$ (X_i contributes significantly)

t-test for $\hat{\beta}_i$	**$t = \hat{\beta}_i / SE(\hat{\beta}_i)$**	df = n−p−1, p = number of predictors
$SE(\hat{\beta}_1)$ for SLR	**$SE(\hat{\beta}_1) = \sqrt{[MSE / \Sigma(x_i - \bar{x})^2]}$**	MSE = Mean Square Error from ANOVA

Decision: If |t_calc| > t_critical(n−p−1, α/2) → Reject H_0 → Coefficient is significant

Note: In Simple Regression (p=1), the F-test for the overall model and the t-test for β_1 give IDENTICAL conclusions: $F = t^2$. They test the same hypothesis.

◆ Coefficient of Determination R^2 and Adjusted R^2

Statistic	Formula	Interpretation
R^2 (R-squared)	$R^2 = SSR/SST = 1 - SSE/SST$	Proportion of variance in Y explained by X. Range: 0 to 1.
Adjusted R^2	$R^2adj = 1 - (1-R^2)(n-1)/(n-p-1)$	Penalizes for adding non-significant predictors. Better for multiple regression.
RMSE	$RMSE = \sqrt{MSE} = \sqrt{[SSE/(n-p-1)]}$	Root Mean Square Error — SD of residuals
R^2 vs r	$R^2 = r^2$ in Simple Regression	r = Pearson correlation coefficient

Interpretation Guide for R^2:

- $R^2 > 0.90$ → Excellent fit (common in calibration curves, HPLC assays)
- $R^2 = 0.80–0.90$ → Good fit (typical in biological studies)
- $R^2 = 0.60–0.80$ → Moderate fit (clinical/epidemiological data — acceptable)
- $R^2 < 0.50$ → Poor fit — model needs revision or more predictors

◆ Multiple Linear Regression (MLR) — Hypothesis Testing

Multiple Linear Regression extends simple regression to k predictors:

MLR Model	$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_kX_k + \varepsilon$	k = number of predictors

Hypothesis Tests in MLR:

- Overall F-test: H_0: $\beta_1=\beta_2=...=\beta_k=0$ (no predictor is significant) — tests if any predictor matters
- Individual t-tests: H_0: $\beta_i=0$ for each i — tests if each predictor contributes, holding others constant
- Partial F-test: Tests significance of adding a subset of predictors to an existing model

Issue in MLR	Description	Pharma Consequence
Multicollinearity	High correlation between predictors (X variables)	*Inflated SE, unreliable coefficient estimates; use VIF (Variance Inflation Factor)*
Overfitting	Too many predictors relative to data points	*R^2 high but predictions poor on new data; use Adjusted R^2 or cross-validation*
Heteroscedasticity	Non-constant variance of residuals	*Violates assumption; use weighted regression or transform Y*
Outliers	Influential data points distorting regression	*Use Cook's distance, leverage plots to identify and investigate*

Statistical Software for Industrial & Clinical Trial Analysis

Modern pharmaceutical research and clinical trials rely heavily on statistical software for data analysis, visualization, and regulatory submission. Each software has specific strengths, and understanding their application is essential for a pharmaceutical professional.

◆ Microsoft Excel — Statistical Analysis

Microsoft Excel — Pharmaceutical Statistical Tool

Excel is the most widely used spreadsheet application with built-in statistical functions, Data Analysis ToolPak, and charting capabilities. It is the entry-level tool for basic pharmaceutical data analysis.

Key Statistical Functions in Excel:

Category	Excel Function	Pharmaceutical Use
Descriptive Statistics	AVERAGE(), STDEV(), MEDIAN(), MODE()	*Tablet weight, hardness, dissolution summary stats*
Confidence Intervals	CONFIDENCE.NORM(), CONFIDENCE.T()	*95% CI for drug concentration, QC limits*
t-tests (ToolPak)	t-Test: Paired / Two-Sample	*Compare drug effects before-after, Drug A vs Drug B*
ANOVA	ANOVA: Single Factor / Two Factor	*Compare batch dissolution, formulation optimization*
Regression	LINEST(), SLOPE(), INTERCEPT(), RSQ()	*Calibration curves, dose-response modeling*
Correlation	CORREL(), PEARSON()	*Dose-response correlation analysis*
Frequency	FREQUENCY(), COUNTIF()	*Tablet weight distribution, defect frequency*
Probability	NORM.DIST(), T.DIST(), CHISQ.DIST()	*Probability calculations for QC testing*

💻 Excel — Step-by-Step: t-test for Tablet Weight Comparison

Problem: Compare tablet weights of Batch A vs Batch B using t-test in Excel

1. 1. Enter data: Column A: Batch A weights (n=10): 498,502,497,503,499,501,500,500,498,502
2. Column B: Batch B weights (n=10): 495,505,492,508,497,503,496,504,498,502
3. Go to: Data Tab → Data Analysis → t-Test: Two-Sample Assuming Unequal Variances
4. 3. Input: Variable 1 Range: $A$1:$A$10 Variable 2 Range: $B$1:$B$10 Hypothesized Mean Difference: 0 Alpha: 0.05
5. 4. Output (Excel generates): t Stat: 0.285 t Critical two-tail: 2.101 P(T<=t) two-tail: 0.779
6. 5. Interpretation: t_calc (0.285) < t_critical (2.101) → Fail to Reject H_0 p-value (0.779) > 0.05 → No significant difference Conclusion: Both batches have equivalent tablet weights.

◆ SPSS (Statistical Package for the Social Sciences)

📈 IBM SPSS Statistics — Clinical Trial & Survey Analysis

SPSS (Statistical Package for the Social Sciences) is IBM's comprehensive statistical software widely used in clinical trials, pharmacovigilance studies, and epidemiological research. It provides a user-friendly GUI with powerful statistical procedures.

SPSS Modules for Pharmaceutical & Clinical Research:

SPSS Menu/Module	Statistical Procedure	Clinical Trial Application
Analyze → Descriptive Statistics	Mean, SD, Skewness, Kurtosis, Frequency tables	*Baseline demographics of clinical trial patients*
Analyze → Compare Means	One-sample t, Independent t, Paired t, One-way ANOVA	*Drug efficacy comparison between treatment arms*
Analyze → Correlate	Pearson, Spearman, Partial correlation	*Dose-response relationships, biomarker correlation*
Analyze → Regression	Linear, Logistic, Cox Regression	*Predicting outcomes, survival analysis, risk factors*
Analyze → Nonparametric Tests	Wilcoxon, Mann-Whitney, Kruskal-Wallis, Friedman	*Ordinal clinical outcomes (pain scales, ADR severity)*
Analyze → Survival	Kaplan-Meier, Log-Rank, Cox Proportional Hazards	*Time-to-event analysis in oncology/cardiology trials*
Analyze → Scale	Cronbach's alpha, Reliability analysis	*Questionnaire validation in patient-reported outcome studies*
Graphs → Chart Builder	Bar, Line, Scatter, Box plots, ROC curves	*Data visualization for clinical trial reports*

💻 SPSS — Step-by-Step: One-Way ANOVA for Drug Dissolution

Problem: Compare dissolution (%) of tablets from 3 formulations (F1, F2, F3) in SPSS 1. Data Setup in SPSS Data Editor: Variable 1 'Dissolution': Enter all dissolution values in one column Variable 2 'Formulation': Enter group codes (1=F1, 2=F2, 3=F3) in second column 2. Procedure: Analyze → Compare Means → One-Way ANOVA Dependent List: Dissolution Factor: Formulation Click 'Post Hoc': Select LSD, Tukey, or Bonferroni Click 'Options': Select Descriptive Statistics, Homogeneity of Variance Test 3. Key Output Interpretation: Levene's Test: $p>0.05$ → Equal variances assumed (ANOVA assumption met) ANOVA Table: $F(2,27)=18.43$, $p<0.001$ → Significant difference between formulations Post Hoc (LSD): F1 vs F2 ($p=0.003$*), F1 vs F3 ($p<0.001$*), F2 vs F3 ($p=0.042$*) 4. Conclusion: All three formulations differ significantly in dissolution ($p<0.05$). Use descriptive statistics output to identify which formulation is optimal.

◆ MINITAB® — Quality Control & Industrial Statistics

🏭 MINITAB® — Industrial Quality Control & Process Improvement

MINITAB is the industry-standard software for statistical quality control (SQC), Six Sigma projects, and process improvement in pharmaceutical manufacturing. It excels at DOE (Design of Experiments), SPC (Statistical Process Control), and Measurement System Analysis (MSA).

MINITAB Applications in Pharmaceutical Industry:

MINITAB Module	Pharmaceutical Application
Stat → Control Charts	Xbar-R, I-MR, P, NP, C charts for monitoring tablet weight, hardness, dissolution in continuous manufacturing
Stat → Quality Tools → Capability Analysis	Process Capability (Cpk, Ppk) for tablet weight uniformity — ICH Q1 stability, USP content uniformity
Stat → DOE (Design of Experiments)	Creating and analyzing 2k factorial, fractional factorial, RSM designs for formulation development
Stat → DOE → Response Surface	Box-Behnken/CCD analysis, response optimizer — identifies optimal formulation conditions (QbD)
Stat → Regression	Simple, multiple, binary logistic regression for stability data, dose-response analysis
Stat → ANOVA	Balanced ANOVA, General Linear Model for multi-factor pharmaceutical experiments
Stat → Reliability/Survival	Weibull analysis for shelf-life estimation, accelerated stability studies
Stat → Measurement System Analysis	Gage R&R studies for analytical instrument calibration and inter-analyst variability

◆ Design-Expert® (Design of Experiments Software)

Design-Expert® by Stat-Ease — Advanced DoE & RSM Software

Design-Expert is the leading specialized software for Design of Experiments and Response Surface Methodology in pharmaceutical R&D and formulation science. It is the preferred tool for QbD (Quality by Design) studies per ICH Q8.

Design-Expert Capabilities:

- Two-Level Factorial: Full 2k and fractional 2k-p designs with blocking
- Response Surface: CCD, Box-Behnken, Doehlert designs with 3D plots
- Mixture Designs: For formulation optimization (drug:excipient ratios must sum to 100%)
- Split-Plot Designs: For experiments where some factors are hard to change
- Optimal Designs: Custom designs for constrained experimental spaces
- Desirability Function: Simultaneously optimizes multiple responses

Workflow in Design-Expert for Pharmaceutical Formulation Optimization:

1. Create Design: Select design type (e.g., Box-Behnken, 3 factors)
2. Define Factors: Enter factor names, units, levels (e.g., Binder 2–6%, Force 5–15 kN)
3. Enter Responses: Define responses (Dissolution%, Hardness, Disintegration time)
4. Run Experiments: Software generates randomized run order
5. Enter Data: Input experimental results for each response
6. Analyze: Software fits models, ANOVA tables, diagnostics (R^2, lack-of-fit test)
7. Graphs: 3D Surface plots, Contour plots, Perturbation plots
8. Optimize: Desirability function — set goals for each response
9. Verify: Run confirmation experiments at the optimal point

Feature	Design-Expert	MINITAB
Primary focus	RSM, Mixture, Optimal designs	SPC, Gage R&R, ANOVA, general stats
Mixture designs	Yes — specialized for formulation ratios	Limited
3D response surface	Yes — interactive, color-coded	Yes, but less interactive
Desirability optimization	Yes — multi-response optimization	Via response optimizer
Pharma use	Formulation development, QbD	Manufacturing QC, Six Sigma
Regulatory acceptance	ICH Q8 referenced for DoE	FDA PAT guidance cited

◆ R — Open-Source Statistical Computing

R Programming Language — Statistical Computing & Graphics

R is a free, open-source programming language and environment for statistical computing and graphics. It is increasingly used in pharmaceutical research, clinical trials, and regulatory submissions due to its flexibility, reproducibility, and extensive package ecosystem.

Why R for Pharmaceutical & Clinical Research:

- Free and open-source — no licensing costs (unlike SAS, SPSS)
- FDA accepts R for statistical analysis in regulatory submissions (CDER 2021)
- 5000+ packages — comprehensive statistical procedures available
- Reproducible research — R scripts document the entire analysis
- Excellent graphics — ggplot2 for publication-quality figures
- Growing use in clinical trials for adaptive designs, Bayesian analysis, survival analysis

Essential R Packages for Pharmaceutical Statistics:

Package	Category	Functions
base R	Core Statistics	t.test(), aov(), lm(), cor(), chisq.test(), wilcox.test()
ggplot2	Visualization	Publication-quality graphs: scatter, box, bar, density, ROC plots
dplyr + tidyr	Data Manipulation	Data wrangling, reshaping, summarizing clinical trial datasets
survival	Survival Analysis	Kaplan-Meier curves, Cox PH model, log-rank test for oncology
nlme / lme4	Mixed Models	Random effects models for repeated measures clinical data

pwr	Power Analysis	Sample size calculation for t-test, ANOVA, correlation, proportion tests
BsMD	Bayesian Screening	Bayesian model discrimination for factorial screening designs
DoE.base	Design of Experiments	Factorial, fractional factorial design creation and analysis
BioequivalenceR	Bioequivalence	BE testing per FDA/EMA guidelines — TOST procedure, 90% CI
NONMEM / nlmixr	PK/PD Modeling	Population pharmacokinetic modeling for clinical trial data

◆ Online Statistical Software for Pharmaceutical Research

Software / Platform	Access & Cost	Best Use in Pharmacy
GraphPad Prism (Online)	**Subscription / Free trial**	*Dose-response curves, EC50 calculation, survival analysis — widely used in pharmacology*
Vassar Stats (vassar.net)	**Completely FREE web-based**	*t-test, ANOVA, chi-square, correlation — quick online analysis without software installation*
Social Science Statistics (socscistatistics.com)	**FREE web-based**	*t-test, ANOVA, Pearson/Spearman correlation, Mann-Whitney — ideal for students*
Statistics Kingdom (statskingdom.com)	**FREE**	*All basic tests with interpretation guidance — beginner-friendly*
JASP (jasp-stats.org)	**FREE, open-source desktop**	*Frequentist + Bayesian analysis; SEM; reproducible clinical research*
EasyCharts / PAST	**FREE**	*Ecological and pharmacological data visualization and multivariate analysis*
OpenEpi (openepi.com)	**FREE, WHO-supported**	*Epidemiology: sample size, odds ratio, relative risk for pharmacoepidemiology*
Biostat TGV (biostatgv.sentiweb.fr)	**FREE**	*French platform for biostatistics — survival analysis, test selection guide*

◆ Software Comparison for Industrial vs Clinical Trial Analysis

Feature	Excel	SPSS	MINITAB	Design-Expert	R
Cost	Low (Office)	High	Medium	Medium	FREE
Ease of Use	Very Easy	Easy (GUI)	Easy (GUI)	Easy (GUI)	Steep learning
Descriptive Stats	Yes	Yes	Yes	Limited	Yes
Clinical Trials	Basic	Best	Moderate	No	Excellent
Quality Control (SPC)	Basic	No	Best	No	Yes
DOE & RSM	No	No	Yes	Best	Yes (packages)
Survival Analysis	No	Yes	Yes	No	Best
Regulatory Acceptance	Common	Common	Common	Common	FDA accepted
Pharma Industry Use	Universal	Clinical	Manufacturing	Formulation R&D	Academia/CRO

◆ Expected Exam Questions — Units 4 & 5

Q	Question	Marks
1	What is confounding in 2k factorial designs? Why is it used? Explain with 2^2 example.	5–10
2	Explain blocking in two-level factorials. How are runs assigned to blocks using the ABC generator?	5–10
3	Construct a 2^3 design with 2 blocks, confounding ABC with blocks. Write the design matrix.	10
4	What is a defining relation? Explain alias structure with a 2^{3-1} fractional factorial.	5
5	What is resolution in fractional factorial designs? Differentiate Resolution III, IV, V.	5
6	Explain hypothesis testing in simple linear regression (F-test + t-test). Write the ANOVA table.	10
7	What is R^2 and adjusted R^2? How do they differ? What are good R^2 values for pharma data?	5
8	Perform hypothesis testing for given regression data: test significance of model and individual coefficient.	10

9	Explain Multiple Regression hypothesis testing. What are multicollinearity and its effects?	5
10	Describe the use of Excel (Data Analysis ToolPak) for: t-test, ANOVA, and regression in pharmacy.	5
11	What is SPSS? How is it used in clinical trial data analysis? Mention 5 key procedures.	5
12	What is MINITAB? Explain its application in pharmaceutical QC (SPC, Cpk, DOE).	5
13	Compare Design-Expert and MINITAB for formulation optimization. What is desirability function?	5
14	What is R? List 5 essential R packages for pharmaceutical and clinical trial statistics.	5
15	Compare Excel, SPSS, MINITAB, Design-Expert, and R for industrial and clinical trial use.	5–10
16	Write R code or MINITAB steps to perform one-way ANOVA on dissolution data from 3 batches.	5