Midterm Review

Davood Tofighi, Ph.D.
Department of Mathematics and Statistics, University of New Mexico
dtofighi@unm.edu

Review Questions for Exam 1- Part 1

Chapter 1: Introduction to Design and Analysis of Experiments

Conceptual Questions

1. What is the primary difference between an **observational study** and a **designed experiment**?
   - A) Observational studies actively manipulate variables, while designed experiments do not.
   - B) Observational studies establish causation, while designed experiments establish correlation.
   - C) Observational studies do not impose treatments, while designed experiments actively manipulate variables.
   - D) Observational studies always involve randomization, while designed experiments do not.
2. Explain why **randomization** is a key principle in the design of experiments.
3. What are the three basic principles of experimental design? Provide a brief explanation of each.

Computational Questions

1. Suppose a study is conducted on the effect of a new fertilizer on crop yield. There are **two fertilizer types (A, B)** and **three different soil conditions (1, 2, 3)**. If each combination of fertilizer type and soil condition is tested on **four plots**, how many total experimental units are there?
2. A completely randomized design is conducted with **4 treatments** and **5 replicates per treatment**. Construct an appropriate ANOVA table, identifying sources of variation and degrees of freedom.

Chapter 2: Basic Statistical Concepts in Experimental Design

Conceptual Questions

1. Define the following terms:
   - Population vs. Sample
   - Parameter vs. Statistic
   - Sampling Distribution
2. Why is it important for a sample to be **random** when conducting statistical inference?
3. Which of the following is NOT a property of the normal distribution?
   - A) It is symmetric about its mean.

- B) It has a mean of 0 and a standard deviation of 1 in all cases.
- C) The total area under the curve is 1.
- D) The mean, median, and mode are equal.

Computational Questions

1. The lifespan of a type of light bulb follows a normal distribution with a **mean of 1200 hours** and a **standard deviation of 100 hours**. If a bulb is randomly selected, what is the probability that it lasts **more than 1300 hours**?
2. A random sample of 40 observations is drawn from a normal population with $\mu = 50$ and $\sigma = 10$. Compute the probability that the sample mean **is greater than 52**.

Chapter 3: Completely Randomized Designs

Conceptual Questions

1. What is the purpose of **blocking** in an experiment? When should it be used?
2. Which of the following statements about **randomized complete block designs (RCBD)** is true?
   - A) It is used to remove known sources of variability.
   - B) It requires an equal number of observations for each treatment.
   - C) It assumes treatments and blocks are independent.
   - D) All of the above.
3. Describe the difference between **fixed effects** and **random effects** models.

Computational Questions

1. An experiment tests the effect of **three different diets** on the weight gain of pigs. A total of **15 pigs** are used, with **five pigs randomly assigned to each diet**. The following sum of squares are obtained:

   - **Total Sum of Squares (SST) = 400**
   - **Sum of Squares for Treatment (SSTr) = 300**
   - **Sum of Squares for Error (SSE) = 100**

   Construct the ANOVA table and test whether diet has a significant effect at $\alpha = 0.05$.

2. Suppose an experiment consists of **four treatment levels** with **six observations per treatment**. Calculate the **degrees of freedom** for the following:

   - Treatment
   - Error
   - Total

Chapter 4: Factorial Designs

Conceptual Questions

1. What is the main advantage of **factorial designs** compared to **one-factor-at-a-time (OFAT) experiments**?
2. A 2^2 **factorial design** involves two factors, each at two levels. How many treatment combinations are there?
3. How do you interpret the **main effect** and **interaction effect** in a factorial design?

Computational Questions

1. Consider a 2^2 **factorial experiment** where Factor A has levels **A1 and A2**, and Factor B has levels **B1 and B2**. The following treatment means were obtained:

Treatment	Mean Response
A1B1	20
A1B2	30
A2B1	25
A2B2	35

   Compute:

   - The **main effect** of A
   - The **main effect** of B
   - The **interaction effect** AB

2. A researcher conducts a 2^3 **factorial experiment** with **factors A, B, and C**. The experiment is replicated **twice**. How many total observations are required?

Chapter 5: Analysis of Variance (ANOVA)

Conceptual Questions

1. What assumptions must be met for a **one-way ANOVA** to be valid?
2. In a **two-way ANOVA**, how many hypotheses are tested, and what do they represent?
3. A researcher wants to compare the effectiveness of **three fertilizers** on crop yield. What type of ANOVA should be used?

Computational Questions

1. A researcher conducts a **one-way ANOVA** with **three groups** ($n_1 = 8, n_2 = 8, n_3 = 8$) and calculates the following sum of squares:

   - **SST = 90**
   - **SSTr = 60**
   - **SSE = 30**

   Construct the ANOVA table and test the significance at $\alpha = 0.05$.

2. In a **two-way ANOVA**, the sum of squares values are given:

- **SST = 180**
- **SSA = 60**
- **SSB = 40**
- **SSAB = 30**
- **SSE = 50**

Compute the **F-ratios** for Factor A, Factor B, and the interaction term, assuming the total sample size is **30**.

Solutions And Explanations

Exam 1: Review Questions - Part 1

Question 9: Probability of a light Bulb Lasting More than 1300 hours

- Given: $X \sim N(1200, 100^2)$
- Standardizing:

$$Z = \frac{1300 - 1200}{100} = 1$$

- Using the standard normal table:

$$P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

- **Answer: 0.1587** (approximately **15.87%**)

Question 10: Probability that Sample Mean > 52

- Given: $X \sim N(50, 10^2)$, sample size $n = 40$
- Standard deviation of sample mean:

$$\sigma_{X^-} = \frac{10}{\sqrt{40}} = 1.58$$

- Standardizing:

$$Z = \frac{52 - 50}{1.58} = 1.27$$

- From the normal table:

$$P(Z > 1.27) = 1 - 0.8980 = 0.1029$$

- **Answer: 0.1029** (approximately **10.29%**)

Question 14: ANOVA Table for Three Diets

Source	SS	df	MS	F	p-value
Treatment	300	2	150.000	18.0	0.000244
Error	100	12	8.333	-	-
Total	400	14	-	-	-

- The F-statistic **F = 18.0**, and the p-value **0.000244** is much less than 0.05.
- **Conclusion: Reject** H_0; at least one diet significantly affects weight gain.

Question 15: Degrees of Freedom

- **Treatment df** = $4 - 1 = 3$
- **Error df** = $(4 \times 6) - 4 = 20$
- **Total df** = $3 + 20 = 23$

Answer: (3, 20, 23)

Question 19: Main and Interaction Effects

- **Main Effect of A**:

$$\frac{A2B1 + A2B2}{2} - \frac{A1B1 + A1B2}{2} = \frac{25 + 35}{2} - \frac{20 + 30}{2} = 5$$

- **Main Effect of B**:

$$\frac{A1B2 + A2B2}{2} - \frac{A1B1 + A2B1}{2} = \frac{30 + 35}{2} - \frac{20 + 25}{2} = 10$$

- **Interaction Effect** AB:

$$\frac{A1B1 + A2B2}{2} - \frac{A1B2 + A2B1}{2} = \frac{20 + 35}{2} - \frac{30 + 25}{2} = 0$$

Answer: (5, 10, 0)

Question 20: Total Observations in a 2^3 Factorial Design

$$\text{Total Observations} = (2^3) \times 2 = 16$$

Answer: 16 observations

Question 24: One-Way ANOVA Table

Source	SS	df	MS	F	p-value
Treatment	60	2	30.000	21.0	0.00001
Error	30	21	1.428	-	-

Source	SS	df	MS	F	p-value
Total	90	23	-	-	-

- The F-statistic **F = 21.0**, and the p-value **0.00001** is much smaller than 0.05.
- **Conclusion: Reject** H_0; there is a significant effect of treatment on response.

Question 25: F-ratios for Two-Way ANOVA

- F_A **(Factor A)**

$$\frac{MS_A}{MSE} = \frac{60}{2.083} = 28.8$$

- F_B **(Factor B)**

$$\frac{MS_B}{MSE} = \frac{40}{4.167} = 9.6$$

- F_{AB} **(Interaction)**

$$\frac{MS_{AB}}{MSE} = \frac{30}{4.167} = 7.2$$

Answer: (28.8, 9.6, 7.2)

Exam 1: Review Questions - Part 2

Chapter 1: Introduction to Design and Analysis of Experiments

Conceptual Questions

1. **(Short Answer)** Explain the three fundamental principles of experimental design and how they contribute to the validity of an experiment.
2. **(Multiple Choice)** Which of the following is **not** a key advantage of a designed experiment over an observational study?
   - A) Ability to establish causality
   - B) Ability to control variability
   - C) Requires fewer resources in all cases
   - D) Allows for factor interactions to be studied
3. **(Explanation Required)** Why is randomization crucial in experimental design? Provide an example where lack of randomization leads to bias.

Numerical & Computational Problems

1. **(Completely Randomized Design)** An experimenter is testing the effect of **three different fertilizers** on plant growth. **Four plants** are assigned to each fertilizer treatment. Construct a **completely randomized design (CRD)** layout and describe the randomization procedure.

2. **(ANOVA Table Construction)** A researcher tests the effects of **four different training programs** on running performance. The following sum of squares values are obtained:
   - **Total SS = 500**, **Treatment SS = 300**, **Error SS = 200**
   - There are **5 subjects per group**

   Construct the **ANOVA table** and test whether training programs have a significant effect at $\alpha = 0.05$.

Chapter 2: Basic Statistical Concepts in Experimental Design

Conceptual & Proof-Based Questions

1. **(Conceptual)** Define the following and provide an example of each:
   - Population vs. Sample
   - Parameter vs. Statistic
   - Sampling Distribution
2. **(Proof)** Prove that the **sample mean is an unbiased estimator** of the population mean.
3. **(Short Answer)** Explain why **independent and identically distributed (i.i.d.) samples** are a key assumption in many experimental designs.

Numerical & R-Based Interpretation

1. **(Probability Computation)** A new drug extends survival time, which follows a **normal distribution with a mean of 18 months and a standard deviation of 4 months**. What is the probability that a randomly selected patient survives more than **22 months**?

2. **(R Output Interpretation)** Below is an R output of a simple **one-way ANOVA**:

```
Analysis of Variance Table

Response: Yield
Df Sum Sq Mean Sq F value Pr(>F)
Treatment  3  24.6   8.2     5.4    0.009
Residuals  16 24.3   1.5
```

   (a) How many total observations are in the dataset?

   (b) What is the decision at $\alpha = 0.05$?

   (c) What is the total sum of squares?

Chapter 3: Completely Randomized Designs (CRD) & Blocking Designs

Conceptual & Proof-Based Questions

1. **(Conceptual)** Describe when a **randomized complete block design (RCBD)** is preferred over a **completely randomized design (CRD)**.
2. **(Proof-Based)** Show that, under an ANOVA model, the total sum of squares can be decomposed as:

$$SS_{Total} = SS_{Treatments} + SS_{Error}$$

3. **(Short Answer)** What are the advantages of using blocking in an experiment?

Numerical & Computational Problems

1. **(Factorial ANOVA Computation)** A study examines the effects of **two different fertilizers (A, B)** and **two irrigation methods (X, Y)** on plant growth. The average growth (in cm) per treatment is given below:

Treatment	Mean Growth
A-X	15
A-Y	18
B-X	22
B-Y	25

Compute:

- The **main effect** of Fertilizer
- The **main effect** of Irrigation
- The **interaction effect** between Fertilizer and Irrigation

2. **(R Output Interpretation: Two-Way ANOVA)**

```
Analysis of Variance Table

Response: Growth
Df Sum Sq Mean Sq F value Pr(>F)
Fertilizer  1  45.6   45.6    12.3    0.004
Irrigation  1  35.2   35.2    9.4     0.009
Interaction 1  10.1   10.1    2.7     0.105
Residuals   16 59.8   3.74
```

(a) Which factors have a significant effect at $\alpha = 0.05$?

(b) Interpret the **interaction term** in the context of the study.

(c) Compute the total sum of squares.

Chapter 4: Factorial Designs

Conceptual & Proof-Based Questions

1. **(Conceptual)** Why are **factorial designs** superior to **one-factor-at-a-time (OFAT) designs**?
2. **(Proof-Based)** Show that the **expected mean square (EMS)** for a factorial ANOVA follows:

$$E(MS_Treatment) = \sigma^2 + \frac{n}{k}\sum(\tau_i^2)$$

3. **(Short Answer)** What does it mean if a **factor interaction is significant** in a factorial design?

Numerical & R-Based Interpretation

1. **(Factorial Design Computation)** A 2^3 **factorial design** is conducted with **three factors (A, B, C)** at two levels each. If each treatment combination is replicated **twice**, how many total observations are needed?
2. **(R Output for Factorial Design)** Given the following R output:

```
Analysis of Variance Table

Response: Time
Df Sum Sq Mean Sq F value Pr(>F)
A  1  30.2   30.2    8.5    0.007
B  1  10.1   10.1    2.8    0.103
A:B  1  25.3   25.3    7.1    0.012
Residuals  20  71.2   3.56
```

   (a) What does the **interaction term (A:B)** suggest?

   (b) Is Factor **B** significant at $\alpha = 0.05$?

   (c) Compute the total sum of squares.

Chapter 5: Analysis of Variance (ANOVA)

Conceptual & Proof-Based Questions

1. **(Conceptual)** What are the key **assumptions** of ANOVA?
2. **(Proof-Based)** Derive the **expected mean square (EMS)** for the error term in an ANOVA model.
3. **(Short Answer)** What is the **difference between a fixed-effects and a random-effects model** in ANOVA?

Numerical & R-Based Interpretation

1. **(ANOVA Computation)** A **one-way ANOVA** is conducted with **three groups** ($n_1 = 8, n_2 = 8, n_3 = 8$) and the following sum of squares:

   - **SST = 90**, **SSTr = 60**, **SSE = 30**

   Construct the **ANOVA table** and test the significance at $\alpha = 0.05$.
2. **(R Output for One-Way ANOVA)** Given the R output:

```
Analysis of Variance Table
```

```
Response: Score
Df Sum Sq Mean Sq F value Pr(>F)
Treatment  2  50.1   25.05   7.2    0.002
Residuals  27  93.8   3.47
```

(a) What is the total sample size?
(b) Is there a significant effect of treatment?
(c) Compute the total sum of squares.

Solution Key for Review Questions

Chapter 2: Basic Statistical Concepts in Experimental Design

Q9: Probability of Survival beyond 22 months

- Given: $X \sim N(18, 4^2)$
- Standardizing:

$$Z = \frac{22 - 18}{4} = 1$$

- Using the standard normal table:

$$P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587$$

- **Answer: 0.1587** (approximately **15.87%**)

Q10: R Output Interpretation (One-Way ANOVA)

Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	3	24.6	8.2	5.4
Residuals	16	24.3	1.5	

(a) **Total observations:**

$$df_{Total} = df_{Treatment} + df_{Residual} + 1 = 3 + 16 + 1 = 20$$

(b) **Decision at** $\alpha = 0.05$

Since $p = 0.009 < 0.05$, **reject** H_0 → The treatment effect is significant.

(c) **Total SS**

$$SS_{Total} = SS_{Treatment} + SS_{Residual} = 24.6 + 24.3 = 48.9$$

Chapter 3: Factorial Design

Q14: Main and Interaction Effects

Treatment	Mean Growth
A-X	15
A-Y	18
B-X	22
B-Y	25

- **Main Effect of A:**

$$\frac{B-X+B-Y}{2}-\frac{A-X+A-Y}{2}=\frac{22+25}{2}-\frac{15+18}{2}=7.0$$

- **Main Effect of B:**

$$\frac{A-Y+B-Y}{2}-\frac{A-X+B-X}{2}=\frac{18+25}{2}-\frac{15+22}{2}=3.0$$

- **Interaction Effect AB:**

$$\frac{A-X+B-Y}{2}-\frac{A-Y+B-X}{2}=\frac{15+25}{2}-\frac{18+22}{2}=0.0$$

Q15: Two-Way ANOVA (R Output Interpretation)

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fertilizer	1	45.6	45.6	12.3	0.004
Irrigation	1	35.2	35.2	9.4	0.009
Interaction	1	10.1	10.1	2.7	0.105
Residuals	16	59.8	3.74		

(a) **Which factors are significant at $\alpha = 0.05$?**

- Fertilizer ($p = 0.004$) $\longrightarrow$ **Significant**
- Irrigation ($p = 0.009$) $\longrightarrow$ **Significant**
- Interaction ($p = 0.105$) $\longrightarrow$ **Not significant**

(b) **Interpret the interaction term**

- Since $p = 0.105 > 0.05$, the interaction effect is **not significant**, meaning the combined effect of **Fertilizer × Irrigation** does not significantly influence plant growth.

(c) **Total sum of squares**

$$SS_{Total} = SS_{Fertilizer} + SS_{Irrigation} + SS_{Interaction} + SS_{Residual} = 45.6 + 35.2 + 10.1 + 59.8 = 150.7$$

Chapter 4: Factorial Designs

Q19: Total Observations in a 2^3 Factorial Design

- 2^3 factorial with **two replications per treatment**:

$$\text{Total observations} = (2^3) \times 2 = 16$$

Q20: Factorial Design (R Output Interpretation)

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	30.2	30.2	8.5	0.007
B	1	10.1	10.1	2.8	0.103
A:B	1	25.3	25.3	7.1	0.012
Residuals	20	71.2	3.56		

(a) **What does the interaction term (A:B) suggest?**

- Since $p = 0.012 < 0.05$, the interaction effect is **significant**, indicating that the effect of factor A on response time depends on the level of factor B.

(b) **Is Factor B significant at $\alpha = 0.05$?**

- $p = 0.103 > 0.05 \rightarrow$ **Not significant**

(c) **Total SS**

$$SS_{Total} = SS_A + SS_B + SS_{A:B} + SS_{Residual} = 30.2 + 10.1 + 25.3 + 71.2 = 136.8$$

Chapter 5: Analysis of Variance (ANOVA)

Q24: One-Way ANOVA

Source	SS	df	MS	F	p-value
Treatment	60	2	30.0	21.0	0.00001
Error	30	21	1.428	-	-
Total	90	23	-	-	-

Decision: Since $p = 0.00001 < 0.05$, **reject** H_0 $\rightarrow$ Treatment has a significant effect.

Q25: One-Way ANOVA (R Output Interpretation)

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	2	50.1	25.05	7.2	0.002

Factor	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	27	93.8	3.47		

(a) **Total sample size**

$$df_{Total} = df_{Treatment} + df_{Residual} + 1 = 2 + 27 + 1 = 30$$

(b) **Decision at** $\alpha = 0.05$

Since $p = 0.002 < 0.05$, **reject** H_0 $\rightarrow$ The treatment effect is significant.

(c) **Total SS**

$$SS_{Total} = SS_{Treatment} + SS_{Residual} = 50.1 + 93.8 = 143.9$$

Bibliography