תכנון ניסויים וניתוח שונות

סילבוס

פרופ' דוד שטיינברג

אוניברסיטת תל אביב, החוג לסטטיסטיקה וחקר ביצועים

סוכם ע"י עידו גרינברג, 2018

תוכן עניינים

סיכום

1. עקרונות עיקריים ל<u>**תכנון ניסוי**</u>:
   a. חזרות.
   b. **הקצאה אקראית**.
   c. חלוקה ל**בלוקים** לפי גורמי השפעה ידועים.
   d. **מבנה פקטוריאלי**/היררכי של השפעות (הפרדה בין השפעות מסדר ראשון לבין אינטראקציות).
   e. **חשאיות**.
2. התמודדות עם <u>**היעדר נורמליות**</u> של השארים במבחני השוואה:
   a. כפיית נורמליות דרך טרנספורמציה.
   b. מודל חלופי מפורש להתפלגות המתאימה.
   c. **מבחן פרמוטציות**: p-value הוא האחוזון מבין הפרמוטציות על הקצאת יחידות הניסוי.
   d. מבחן דרגות: החלפת התוצאות המספריות בדרגותיהן (**Wilcoxon Rank Sum Test**).
   e. Bootstrap: מבחן פרמוטציות עם resampling בהקצאות.
3. <u>**מדגם מזווג (paired)**</u> – מדגם לפני/אחרי:
   a. ניתן להשתמש ישירות ב-$\mathbf{x_i=x_{after}-x_{before}}$ (עד כדי טרנספורמציה כגון log) במקום ב-x_{befire},x_{after} בנפרד.
   b. **Wilcoxon Signed Rank Test**: $W=\sum_{i=1}^{N} sign(x_i)rank(|x_i|)$ – אין הנחת נורמליות על השארים אך יש התחשבות (חלקית) בגודל ההפרשים.
4. <u>**השוואת פרופורציות**</u> – האם תופעה מסוימת שכיחה יותר באוכלוסיה אחת מבאחרת?
   a. שיטה I: הפרש השכיחויות $\hat{p}_1-\hat{p}_2$ מתכנס להתפלגות נורמלית עם גודל המדגם (הפרש בין שני בינומים = סכומים של הרבה ברנולים) 🡸 p-value.
   b. שיטה II: **Pearson χ^2-test** – ניתן להכללכה ל-k אוכלוסיות: כמה כל פרופורציה רחוקה מהממוצע על כל המדגם?
   c. שיטה III: **Fischer exact test**: גישה דומה לפירסון, עם סטטיסטי אחר, שאמורה להיות מובהקת יותר במדגמים קטנים.
5. <u>**ניסויים פקטוריאליים**</u> – בחינת השפעת M משתנים נשלטים בינאריים (או k-נאריים) על משתנה תלוי:
   a. יש $\mathbf{2^M}$ פקטורים – המשתנים עצמם וכל צירופי האינטראקציות ביניהם – אם כי ניתן לבחון רק חלק מהם (לדוגמה רק אינטראקציות מסדר נמוך, או רק אינטראקציות בין חלק מהמשתנים).
   b. האם הפקטורים משפיעים? – ניתן לבנות מודל רגרסיה ליניארית ולהשוות את התפלגות המקדמים להתפלגות נורמלית (עם שונות המתאימה לשונות המדגמית – הנאמדת ע"י השונויות בתוך הקבוצות המוגדרות ע"י הפקטורים).
   c. ניסוי עם **תצפית יחידה לכל פקטור** (קומבינציית ערכי משתנים):
      i. יש לשלול את ההשפעה של חלק מהפקטורים (באופן גס לפי מודל הרגרסיה, אפילו שאין אמדן אפריורי לשונות המדגמית; או באמצעות הקירוב של **Lenth's PSE**).
      ii. מהמקדמים של הפקטורים חסרי ההשפעה במודל הרגרסיה ניתן לגזור את השונות המדגמית, ומכאן לבחון את יתר הפקטורים.
   d. **ניסוי פקטורים חלקי** – פחות תצפיות מצירופי משתנים:
      i. לא תיתכן הפרדה בין כל צירופי המשתנים – בהכרח ישנם פקטורים שתכנון הניסוי לא מאפשר להפריד בין השפעותיהם (***aliases***).
      ii. יש להניח אפריורית צירופים שאינם בעלי השפעה (כגון אינטראקציות מסדר גבוה), ולבחור את צירופי הערכים עבורם תהיינה תצפיות (באמצעות ***הקשר המגדיר***) כך שכל גורם משפיע יהיה alias רק עם גורמים שכהנחה אינם בעלי השפעה.
6. <u>**ANOVA**</u> (Analysis Of Variance) – בחינת ההשפעה של משתנה נשלט בעל k ערכים על משתנה תלוי:
   a. ויזואליזציה: **boxplot**.
   b. מודל: $\boldsymbol{Y_{ij}=\mu+\alpha_i+\epsilon_{ij}}$ (<u>שארים נורמלים</u> <u>ב"ת</u> עם <u>שונות שאינה תלויה בקבוצה</u>).

c. האם יש הבדל בין הקבוצות? $\boldsymbol{F = \frac{MSB}{MSW} \sim F_{k-1,N-k}}$ (עבור k=2 שקול למבחן t עם שונויות שוות).
d. האם יש הבדל בין זוג קבוצות ספציפי? עבור שתי קבוצות m, l תחת H_0, $\boldsymbol{\bar{Y}_m - \bar{Y}_l \sim t}$. עם זאת, כאשר משווים מספר גדול של זוגות קבוצות, הסבירות לשגיאה מסוג I בזוג אחד לפחות (***FWER***) עולה בהרבה על הסבירות שהוגדרה כנדרשת עבור כל זוג בפני עצמו (***IER***). ניתן לתקן זאת באחת השיטות:
   i. לבחון רק מספר מצומצם של זוגות (לדוג' $\{m - vs - all\}_m$ במקום $\{m - vs - l\}_{m,l}$).
   ii. **Bonferroni**: לקבוע את FWER הרצוי ולהקטין את IER בהתאם תחת הנחות שמרניות.
   iii. **Tukey**: חישוב p-value מפורש עבור $\binom{k}{2}$ השוואות בין k קבוצות בגדלים שווים. עבור גדלים שונים / פחות השוואות, מתקבל מבחן שמרני יותר, ולכן מומלץ לבחון מראש אם יש עדיפות על Bonferroni.
   iv. **Scheffe**: הכללה של Tukey עבור השוואות בין כל זוגות ה-contrasts האפשריים (כל ***contrast*** הוא צ"ל של קבוצות).
e. <u>**הימנעות מהנחות ה-ANOVA**</u>:
   i. **מבחן F** אמור להיות יחסית בלתי-רגיש להנחת הנורמליות.
   ii. לחלופין, הסתמכות על דירוג הנתונים במקום על הערכים מבטל את הצורך בנורמליות ושונות קבועה, ומצריך רק אי תלות בין הדגימות.
   iii. פיתרון I: **Wilcoxon Rank Sum עם תיקון Bonferroni**.
   iv. פיתרון II: **Kruskal-Wallis** (**KW**) – חישוב p-value ל-SSB מנורמל על הדירוגים ($\frac{12\sum_1^k n_i(\bar{R}_i - \bar{R})^2}{N(N+1)}$), לפי פרמוטציות או (עבור מדגם גדול מספיק) לפי קירוב ל-χ^2_{k-1}.
f. <u>**Two-way ANOVA**</u> – בחינת ההשפעה של שני משתנים נשלטים ("ניסוי כפול"):
   i. מודל: $\boldsymbol{Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}}$.
      1. (תחת האילוץ של נרמול סכומי ההשפעות ל-0 עבור כל i בנפרד וכל j בנפרד).
   ii. האם המשתנים או האינטראקציה ביניהם משפיעים? $\boldsymbol{F_{\{A/B/AB\}} := \frac{MS\{A/B/AB\}}{MSW} \sim F}$.
g. <u>**ANOVA עם אפקטים מקריים**</u> (בניגוד לאפקטים קבועים): מודל ANOVA שבו k הקבוצות המתוצפתות הן רק מדגם מקרי מתוך אוכלוסיית הקבוצות המלאה.

7. <u>**ניסוי בלוקים – RCBD**</u> (Randomized Complete **Block** Design) – הפחתת השפעת גורמים מרעישים:
a. הכללה של מדגם מזווג: בהינתן גורם חיצוני החשוד כמשפיע (***Nuisance Factor***), יחידות הניסוי יחולקו ל***בלוקים*** (קבוצות "דומות" לפי אותו גורם חיצוני), ובכל בלוק היחידות יוקצו אקראית לקבוצות הניסוי.
b. מודל (דומה ל-Two-way ANOVA כאשר B נתפס כגורם מרעיש): $\boldsymbol{Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}}$.
   i. הנחה: אין אינטראקציה בין הגורם המרעיש לגורם הנבדק.
c. בדומה ל-ANOVA, ניתן לבחון השפעות ע"י **מבחן F** או ע"י **השוואת זוגות של קבוצות/contrasts**.
d. מבחן **Friedman** – הכללה של KW לניסוי בלוקים: מבחן המבוסס על דירוג התצפיות בלבד.
e. <u>**ריבוע לטיני**</u>: הקצאה של יחידות הניסוי לפי שני גורמים משפיעים, כך שכל אחד מהגורמים מתחלק בצורה שווה בין קבוצות הניסוי.
   i. ניתן לממש הקצאה מבוססת ריבוע לטיני בתוך או במקום בלוקים של RCBD.

טבלה מסכמת

<table>
<tr><th rowspan="2"></th><th rowspan="2">Is there affecting factor?</th><th colspan="2">Which factors do affect?</th></tr>
<tr><th>Value-based
(assuming Normality)</th><th>Rank-based
(exploiting less info)</th></tr>
<tr><td>2 Groups</td><td colspan="2">t-test</td><td>Wilcoxon Rank Sum / Permutations</td></tr>
<tr><td>2 Groups – paired</td><td colspan="2">t-test for differences</td><td>Wilcoxon Signed Rank (values are partially used)</td></tr>
<tr><td>K Groups – compare proportion (binary Y)</td><td colspan="2">p-value of p2-p1 (only for K=2) /
Pearson χ^2-test / Fischer exact test</td><td>–</td></tr>
<tr><td>1D K Groups</td><td>F-test for MSB/MSW</td><td rowspan="3">Bonferroni / Tukey / Scheffe</td><td rowspan="2">Wilcoxon + Bonferroni / KW</td></tr>
<tr><td>2D $K_1 \times K_2$ Groups</td><td rowspan="2">F-test for MS{A/B/AB}/MSW</td></tr>
<tr><td>K Groups by Blocks (possibly 2D blocks as Latin square)</td><td>Friedman</td></tr>
<tr><td>2^M Groups</td><td>F-test on MSB/MSW?</td><td rowspan="2">Regression residuals analysis</td><td rowspan="2"></td></tr>
<tr><td>2^M Groups – single sample (or 2^{M-P})</td><td>Lenth’s PSE on regression coefficients</td></tr>
</table>

(1) מבוא

1. **ניסוי vs. תצפית** – בניסוי יש שליטה בתנאים שמגבירה את היכולת לנטרל גורמים מסבירים זרים ולהסיק מסקנות. **חיפוש ונטרול גורמים מסבירים** (**"מקורות שונות"**) הוא אחד הדברים החשובים ביותר בתכנון וניתוח תוצאות ניסוי.

(2) עקרונות לתכנון ניסוי

1. **הקצאה אקראית**: יחידות הניסוי מחולקות לקבוצות הניסוי באופן אקראי.
2. **חזרות**: כל קבוצת ניסוי כוללת הרבה פריטים, כך שניתן להבחין בין שונות שמוסברת ע"י המשתנה הנבדק לבין הפיזור הטבעי בניסוי.
3. **חלוקה לבלוקים**:
   a. כדי לצמצם השפעת גורמים זרים, מומלץ לחלק את יחידות הניסוי לבלוקים הומוגניים ככל האפשר ("הומוגניים" במונחים של אותם גורמים זרים, עד כמה שידוע), ולהקצות יחידות ניסוי לקבוצות שונות בתוך כל בלוק בנפרד.
   b. לדוגמה, בניסוי חקלאי האדמה תחולק לריבועים ובתוך כל ריבוע יחולקו שטחים שונים לקבוצות הניסוי השונות; ובניסוי על 1000 פריטים סדורים, לכל n פריטים בנפרד (לפי הסדר) יוגרל איזה פריט יוקצה לכל אחת מ-n הקבוצות.
   c. בפרט תמיד מומלץ לבחון את תוצאות הניסוי כפונקציה של המספר הסידורי של יחידת הניסוי ולחפש תופעות מגמתיות או מחזוריות – הסדר של הדאטא מכיל לעיתים מקורות שונות משמעותיים.
4. מבנה פקטוריאלי?
5. מבנה היררכי?
6. **סמיות (חשאיות)**: היחידה הנבדקת בניסוי ("מטופל"), ובעדיפות שניה גם הגורם המיישם את הניסוי ("מטפל"), ובעדיפות שלישית גם הגורם המנתח את תוצאות הניסוי – אינם יודעים איזו יחידת ניסוי הוקצתה לאיזו קבוצת ניסוי.

(3) הנחת הנורמליות במבחני השוואה

1. נורמליות של התפלגות נתונים שימושית לחישובים של גדלים סטנדרטיים כמו p-value, confidence interval וכו'. בהיעדר נורמליות של התפלגות תוצאות ניסוי השוואה, ניתן לנסות **לכפות נורמליות דרך טרנספורמציה** על התוצאות דוגמת log, sqrt, inv וכו', או להציע **מודל חלופי** למודל הנורמלי (שיתאים לתוצאות) ולנתח אותו.
2. כמו כן, קיימים מבחנים אלטרנטיביים שאינם מניחים משפחה פרמטרית מסוימת של התפלגויות – **מבחנים א-פרמטריים**:
   a. **מבחן פרמוטציות** מלא: בחינת תוצאות הניסוי עבור כלל ההקצאות אקראיות של יחידות הניסוי (תחת השערת האפס – שאין הבדל בין ההתפלגות בשתי האוכלוסיות – שינוי ההקצאה לא משפיע על התוצאה הנמדדת). אם עבור מרבית ההקצאות האקראיות מתקבלות תוצאות קיצוניות פחות מהתוצאות שהתקבלו בפועל, ניתן לדחות את השערת האפס.
      i. מבחן פרמוטציות חלקי (מבחן **מונטה קרלו**): הגרלה של חלק מההקצאות בלבד.
      ii. **Bootstrap**: מבחן פרמוטציות חלקי עם resampling בהקצאות.
   b. **מבחן דרגות** (**Wilcoxon Rank Sum Test**): מבחן פרמוטציות שבו הערכים של תוצאות הניסוי ממוינים ומוחלפים באינדקס שלהם (לדוג' 3,7,5 יוחלפו ב-1,3,2).
      i. התפלגות האינדקסים (בניגוד להתפלגות התוצאות) אינה תלויה בניסוי וניתנת למידול גלובלי. סטטיסטי המבחן הוא סכום הדרגות של קבוצה בגודל n מתוך N:

$$W := \sum_{1}^{n} s_i$$

$$E(W) = n\frac{N+1}{2} \qquad Var(W) = n(N-n)(N+1)/12$$

$$p - value \approx 1 - \phi\left(\frac{W - E(W)}{\sqrt{Var(W)}}\right) \quad (for\ large\ enough\ dataset)$$

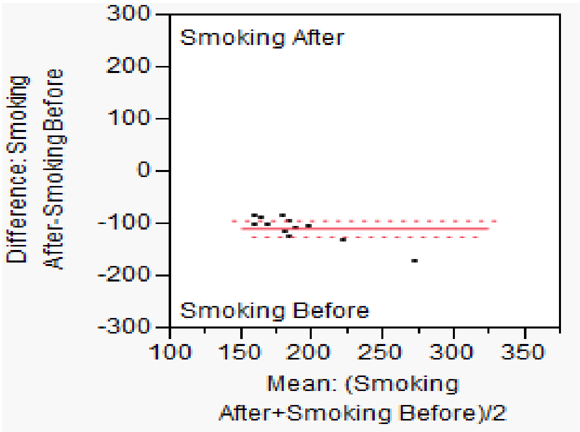
ii. **Mann–Whitney *U* test**: מבחן דרגות שקול המחושב באופן אחר.

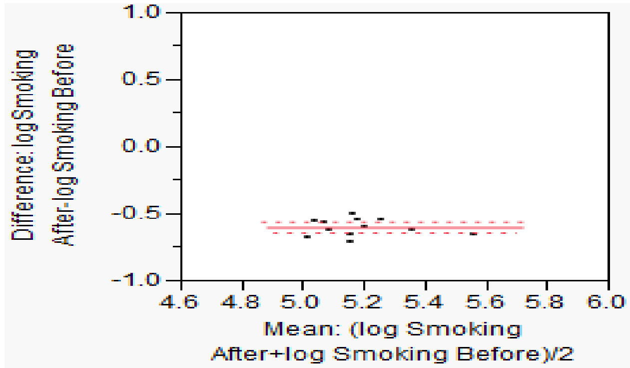
iii. המחיר של ויתור על אינפורמציה מהתוצאות (דרגות בלבד במקום ערכים) הוא עוצמה פחותה (לרוב לא במידה משמעותית) של המבחן ביחס ל-t-test (כלומר "קשה" יותר לדחות את השערת האפס בצורה מובהקת).

c. כאמור, מבחנים א-פרמטריים אינם מניחים שהתצפיות מגיעות ממשפחה מסוימת של התפלגויות, ולכן ההשערה האלטרנטיבית בניסוי H_A אינה יכולה לבטא טווח ערכים של פרמטר. במקום זאת, **ההשערה האלטרנטיבית מבטאת את האפשרות ששתי ההתפלגויות שונות באופן שמגדיל את ההסתברות P(a<b)** עבור תצפיות a,b מאוכלוסיות A,B. השערה אלטרנטיבית זו מזוהה לעיתים קרובות לצרכי הניסוי עם ההשערה של-B תוחלת גדולה מל-A, אף על פי שהדבר אינו בהכרח שקול.

(4) מדגם מזווג: ניסוי עם השוואה ל-Baseline

1. אידיאליזציה של עיקרון הבלוקים, שהינה אפשרית כאשר ניתן לייצר "גרסה" של אותה יחידת ניסוי עבור כל אחת מקבוצות הניסוי (לדוגמה מדידת מטופל לפני ואחרי טיפול; בחינת אותו נבדק בימים שונים בתנאים שונים; פיצול של יחידת ניסוי ושיוך כל תת-יחידה לקבוצה אחרת; סימולציה של קבוצות ניסוי שונות על קבוצות זהות של קלטים...).
2. קבוצת ביקורת בניסוי כזה תהיה יחידות ניסוי שחולקו לכמה חלקים אך כולם שויכו לקבוצת ברירת המחדל של הניסוי – כך ניתן לשערך את השונות הטבעית בתוך יחידת ניסוי.
3. **בחירת סטטיסטי ההשוואה** יכולה להיעשות בסיוע תרשים פיזור מתאים – ההפרש x_{after}-x_{before} כתלות בממוצע $(x_{before}+x_{after})/2$. לעיתים קרובות ניתן יהיה לראות הפרשים גדולים יותר היכן שהערכים גדולים יותר. במקרה כזה ניתן למדוד יחסים במקום הפרשים, או באופן שקול את log ההפרשים.





4. מבחן מועיל להשוואה על פני מדגם מזווג הוא **Wilcoxon Signed Rank**:
   a. בדומה לגרסה המקורית, התוצאות (x1…xn) מוחלפות בדירוג שלהן (1…n), אך הפעם עבור $x_i = x_{after} - x_{before}$ ולפי הערך המוחלט |xi|. בנוסף מוגדר zi:=(xi>0). הסטטיסטי הוא $\boldsymbol{W \coloneqq \sum_1^n i z_i}$.
   b. תחת השערת האפס, zi הם 0 או 1 בהסתברויות שוות וב"ת (משתני ברנולי סימטריים ב"ת), לכן $E(W) = \frac{1}{2}\left(\frac{n(n+1)}{2}\right)$ ו-$Var(W) = n(n+1)(2n+1)/24$, ועבור מדגם גדול מספיק ניתן להשתמש במשפט הגבול המרכזי.

(5) השוואת פרופורציות

1. מטרה: לבדוק אם יש הבדל בשכיחות p של תופעה מסוימת בין שתי אוכלוסיות. קיימים 3 מבחנים מקובלים.
2. **השוואת פרופורציות ישירה**: עבור מדגמים גדולים ניתן לבדוק מובהקות ל-0 $H_0: p_1 - p_2 = 0$ לפי:

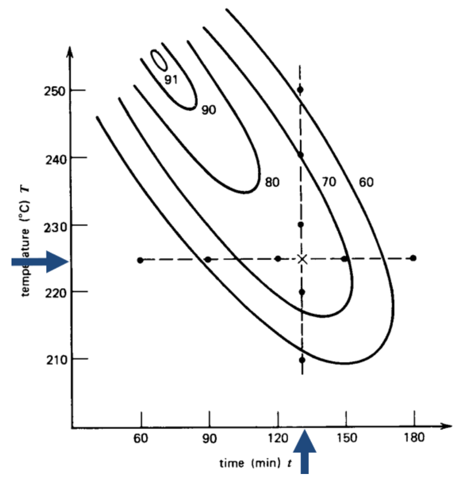
$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

   בין אם ישירות ל-$\hat{p}_1 - \hat{p}_2$ או ל-$Z \coloneqq (\hat{p}_1 - \hat{p}_2)/std$.
3. **Pearson χ^2-test**: לבדוק כמה $\hat{p}_1, \hat{p}_2$ רחוקים מ-$\hat{p} \coloneqq \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$, וכנ"ל עבור $1 - \hat{p}_1, 1 - \hat{p}_2$.
   a. הסטטיסטי הוא $\sum \frac{(\boldsymbol{O_i - E_i})^2}{\boldsymbol{E_i}}$, כאשר O_i הם הפרופורציות p1,p2,1-p1,1-p2 כפי שנצפו בניסוי, ו-E_i הם הפרופורציות הצפויות תחת H_0 ו-$\hat{p}$ – לדוגמה $E_i = \frac{n_1}{\sum n_j}\hat{p}$.
      i. ככל שהסטטיסטי הנ"ל גדול יותר, כך הוא מתאים פחות ל-H_0.
   b. **יתרון: ניתן להכללה** ל-N אוכלוסיות ולמשתנים עם M ערכים (ולא רק תופעות בינאריות) – הסטטיסטי מתפלג במקרה זה $\chi^2\big((N-1)(M-1)\big)$.
   c. ראה χ^2-test Fit tests / Approximated וכן Independence tests בסיכומים של Statistical Theory.
4. **Fisher exact test**:
   a. גישה דומה לפירסון, אבל הפעם הסטטיסטים הם p_i ו-$1 - p_i$ עצמם באופן ישיר, המתפלגים היפרגאומטרית.
   b. **יתרון: מובהק יותר במדגמים קטנים**.

(6) ניסויים פקטוריאליים

1. ניסוי הבוחן השפעה של משתנים נשלטים (כגון פרמטרים של אלגוריתם) על משתנה אחר.
2. **One factor at a time** – כאשר רוצים צירוף פקטורים שממקסם משתנה מסוים: עבור הפקטור הראשון – לבחור ערכים שרירותיים לפקטורים האחרים ולמצוא את המקסימום של המשתנה. עבור כל פקטור בהמשך – להשתמש בערכים האופטימליים של המשתנים שנבדקו ובערכים השרירותיים של אלה שלא נבדקו.

a. זהו למעשה חיפוש חמדני שסורק כל ציר (פקטור) פעם אחת – סה"כ $K \cdot M$ בדיקות עבור M פקטורים עם K ערכים.
b. זו שיטה יעילה תחת ההנחה שהמשתנה הממוקסם כפונקציה של הפקטורים מקיים $y(\{x_i\}) = \sum f_i(x_i)$ – כלומר **בהנחה שאין אינטראקציות בין ההשפעות של הפקטורים**.
c. עם זאת, בגלל ההיצמדות השרירותית לצירים של הפקטורים, קשר בין הפקטורים המשפיעים מוביל ל**תוצאות לא נכונות אפילו כשהבעיה קמורה**:



ניסוי פקטורים מלא

3. **Full factorial design – ניסוי פקטורים מלא**: עבור M משתנים משפיעים עם K ערכים כ"א – לבחון כל אחת מ-K^M תתי-האוכלוסיות בנפרד.
   a. מצריך הרבה דאטא וקשה לחילוץ תובנות כלליות.
   b. ניתן להשתמש בכמות סבירה של דאטא ולבחון חתכים (כלומר משתנים) שונים באופן גס כדי להכווין ניסויי המשך.
4. את ה**אינטראקציה** בין שני פקטורים משפיעים ניתן למדוד באופן הבא (נניח שהפקטורים בינאריים, ונסמן ב-y(A,B) את ממוצע התצפיות של y עבור ערכים מסוימים של A,B):

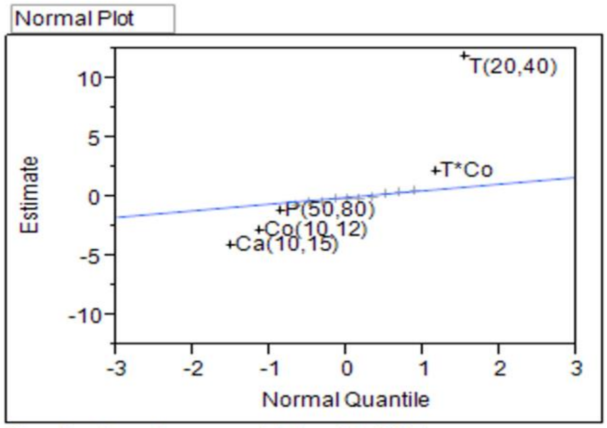
Interaction(A,B) = effect(B|A+)-effect(B|A-) =

= (y(A+,B+)-y(A+,B-)) - (y(A-,B+)-y(A-,B-)) = (y(A+,B+)+y(A-,B-)) – (y(A+,B-)+y(A-,B+))

   a. בפרט מתקבל שהאינטראקציה הינה סימטרית בין A ל-B.
5. ניתן **למדל תלות בשני פקטורים עם אינטראקציה** ע"י מודל לינארי באופן הבא (המודל מניח תצפיות iid ושגיאות iid): $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{12} X_{1i} X_{2i} + \epsilon_i$. למשתנה האינטראקציה ניתן לחשב מובהקות ורווח סמך כמו לפקטורים הרגילים.
   a. עבור M פקטורים, כדי למדל את כל האינטראקציות בין m פקטורים נדרשים $\binom{M}{m}$ משתנים חדשים. האינטראקציה בין 3 פקטורים, לדוגמה, מחושבת באופן הבא: Intr(ABC) = Intr(AB|C+)-Intr(AB|C-), כאשר Intr(AB) מחושב כמוגדר לעיל.
6. **אמדן שונות מדגמית**: את השונות המדגמית (שאינה מוסברת ע"י הפקטורים) ניתן לאמוד כממוצע משוקלל של K^M השונויות השונות בקבוצות הניסוי השונות (משוקלל לפי דרגות החופש = גודל הקבוצה פחות 1).
7. ניתן לפשט את מערך הניסוי ולצמצם K→2 ע"י חלוקת הסקאלה של כל פקטור באופן גס ל"גבוה" ו"נמוך". חלוקה זו מפחיתה את דיוק מדידת ההשפעה, אבל מאפשרת למדוד קיום השפעה של כל פקטור עם $\sim 2^M$ דאטא במקום $\sim K^M$. זה כמובן הכרחי במקרה הנפוץ ש-$K = \infty$ (טווח רציף של ערכים).
8. **האם הפקטורים משפיעים?**

a. אם ממדלים את כל האינטראקציות מקבלים בסה"כ 2^M משתנים. אם מניחים (H0) שהפקטורים אינם משפיעים, אזי המקדמים של המשתנים מתפלגים נורמלית סביב 0 (תחת ההנחה שהשגיאות iid נורמליות).

b. ניתן לבדוק נורמליות של הבטאות סביב 0, למשל, ע"י תרשים אחוזונים מנורמל באופן ייעודי להתפלגות נורמלית (qqnorm) – שם תחת H0 אמור להתקבל קירוב של קו ישר העובר ב-0.



Blue line has slope equal to Lenth's PSE.

c. מעבר לבחינת ההשערה שהפקטורים משפיעים, ניתן לקבל באופן זה הכוונה לפקטורים החשודים כמשמעותיים (אלו שחורגים באופן משמעותי מהקו (במקרה שבו יש משהו דומה לקו)).

Sorted Parameter Estimates

Term	Estimate	Relative Std Error	Pseudo t-Ratio	Pseudo p-Value
T(20,40)	12	0.25	21.33	<.0001*
Ca(10,15)	-4	0.25	-7.11	0.0009*
Co(10,12)	-2.75	0.25	-4.89	0.0045*
T*Co	2.25	0.25	4.00	0.0103*
P(50,80)	-1.125	0.25	-2.00	0.1019
T*P	-0.625	0.25	-1.11	0.3171
Ca*T	0.5	0.25	0.89	0.4148
Ca*P	0.375	0.25	0.67	0.5345
Ca*T*P	-0.375	0.25	-0.67	0.5345
T*P*Co	-0.375	0.25	-0.67	0.5345
Ca*T*Co	0.25	0.25	0.44	0.6753
P*Co	-0.125	0.25	-0.22	0.8329
Ca*P*Co	-0.125	0.25	-0.22	0.8329
Ca*T*P*Co	-0.125	0.25	-0.22	0.8329
Ca*Co	0	0.25	0.00	1.0000

9. ניסוי K^M עם **תצפית יחידה בכל קומבינציה**:

a. <u>תמצית</u>: בטאות של צירופי-פקטורים שאינם נראים בעלי השפעה, מבטאות את השונות שנותרת לאחר שמחלקים את התצפיות לשתי קבוצות שכל אחת מהן מושפעת באותה המידה מכלל צירופי-הפקטורים המשפיעים. לכן שונות זו היא השונות המדגמית (ה"רעש" ע"פ המודל).

b. בבדיקת ההשערה שהפקטורים משפיעים, ניתן לגלות שפקטורים או צירופי-פקטורים מסוימים אינם משפיעים באופן מובהק. בדוגמה למעלה נראה שכל הצירופים של 3 פקטורים ומעלה הם כאלה (קרי אין אינטראקציות מובהקות בין 3 פקטורים או יותר).

c. במצב כזה, היות שהבטאות המתאימות מתפלגות $\beta_{non-influential} \sim N(0, \sigma^2/N)$ (כאשר N מספר התצפיות הכולל ו-σ^2 השונות המדגמית – זו של השגיאה), ניתן להשתמש בהן כדי לאמוד את השונות המדגמית. בדוגמה למעלה יש 5 אינטראקציות בין 3 משתנים ומעלה שאנו מניחים

שאינן משפיעות, לכל אחת מהן בטא המחושבת לפי כל 16 התצפיות, ולכן **אומדן לשונות המדגמית** יהיה $MSW = \frac{16}{5}\sum\hat{\beta}^2_{non-influential}$ עבור המדגם כולו ו-$\widehat{SE} = \sqrt{MSW/16}$ פר תצפית.

d. כעת ניתן לבחון את יתר צירופי-הפקטורים (אלו שלא שללנו את השפעתם בשלב הראשוני) ביחס לשונות המדגמית.

10. **Lenth's PSE** (Pseudo-Standard-Error):

a. כפי שהוסבר למעלה, בניסוי פקטורים מלא עיקר הפיזור מוסבר לרוב ע"י מספר קטן של צירופי-פקטורים, וניתן לנצל את שאר צירופי-הפקטורים כדי לאמוד את שגיאת התקן.

b. לנץ הציע דרך עקומה אך אובייקטיבית יותר לעשות זאת (לא צריך לבחור את הצירופים הבלתי-משפיעים בעין לפי גרף אחוזונים):

i. $s_0 \coloneqq 1.5 \cdot med(|\hat{\beta}_i|)$

ii. $PSE \coloneqq 1.5 \cdot med\{|\hat{\beta}_i| \mid |\hat{\beta}_i| < 2.5 s_0\}$

iii. $t_i \coloneqq \hat{\beta}_i / PSE$

c. לנץ טען שניתן לקרב את t_i תחת H0 (לפיה צירוף-הפקטורים i הינו חסר השפעה) למשתנה $t(2^M/3)$ (התפלגות t עם שליש-מספר-צירופי-הפקטורים כדרגות חופש). בפועל ספי המובהקות של t_i נקבעו גלובלית על סמך סימולציות.

סיכום הגישה בניסויים 2^k

1. אמידת ההשפעות.
2. מהמודל הראשוני:
   – אם יש חזרות, האומד $MSW = \frac{\sum_{i=1}^{2^k} s_i^2}{2^k}$ משמש להסקה על ההשפעות.
   – אם אין חזרות, השיטה של Lenth יחד עם ה- normal probability plot משמשים להסקה על ההשפעות.
3. איפוס מקדמים במודל.
4. ניתוח שאריות.
5. ניתוח וסיכום תוצאות.

ניסוי פקטורים חלקי

11. **ניסוי 2^{M-P} – ניסוי פקטורים חלקי** (עבור פקטורים בינאריים שערכיהם ±1):

a. תחת הנחה אפריורית שאינטראקציות מסדר גבוה הינן בעלות השפעה זניחה, ניתן לבחון M פקטורים גם עם פחות מ-2^M תצפיות.

b. בדוגמה עם 5 פקטורים, השפעות עד סדר שני כוללות 5+10=15 צירופי-פקטורים מתוך 32, כך שניתן להסתפק בחצי מהדגימות.

c. **בחירת מחצית התצפיות כך שיכילו את כל הצירופים הרלוונטיים**: היות שמניחים שהאינטראקציה הגבוהה ביותר $\Pi_{i=1}^M X_i$ אינה משפיעה, אזי ניתן לקבע אותה לערך 1 (או 1-), ומכאן לקבל את צירופי הערכים הנדרשים לתצפיות ע"י $X_M = \Pi_{i=1}^{M-1} X_i$ – כלומר M-1 הפקטורים הראשונים מקבלים את כל הצירופים האפשריים, והפקטור ה-M נקבע לפיהם. כך מתקבל **ניסוי פקטורים חלקי 2^{M-1}**.

d. **הקשר המגדיר של הניסוי** הוא האילוצים המאפיינים את הנקודות הנבחרות לניסוי. בדוגמה: $\Pi_{i=1}^M X_i = 1$. באופן כללי עבור פקטורים בינאריים הקשר המגדיר הוא קבוצה של אילוצים מהצורה $\Pi_{certain-i's} X_i = 1$ (או באופן שקול 1-=).

e. השפעתו של צירוף הפקטורים $\Pi_{i=1}^M X_i$ אינה מתבטאת כלל בניסוי משום שנבדק רק ערך אחד של גורם זה. כל שאר הגורמים (צירופי-הפקטורים) כן מקבלים בניסוי ערכים שונים ולכן משפיעים, אולם השפעתם מתערבבת עם גורמים אחרים ולא ניתן לבודד אותה באופן מלא.

f. גורמים שהשפעתם מתערבבת בניסוי מכונים **aliases**, והמטרה של הקשר המגדיר היא שכל גורם משפיע יהיה alias עם גורם שאינו משפיע, כך שבפועל השפעתו תישאר מבודדת. בדוגמה לעיל כל זוג $X_i, \Pi_{j\neq i}X_j$ הם aliases, כמו גם כל זוג $X_iX_j, \Pi_{l\neq i,j}X_l$, ולכן בהנחה שאין אינטראקציות משפיעות מסדר>2, מתקבל שכל גורם משפיע הינו מבודד.

g. באופן כללי ב**ניסוי** $\mathbf{2^{M-p}}$ יש להשתמש בכל הקומבינציות של M-p הפקטורים הראשונים ולהשלים את p הפקטורים הנוספים בהתאם לקשר המגדיר של הניסוי.

i. לא מוסבר איך להכליל קשר זה עבור $p \geq 1$. נראה שזה תלוי בהנחות על הגורמים המשפיעים, אך תחת ההנחה שהגורמים המשפיעים הם האינטראקציות מסדרים נמוכים כנראה שיש טבלאות כלליות לקביעת הקשר המגדיר עבור M,p שונים.

h. ה**רזולוציה** של הניסוי (**design resolution**) מוגדרת להיות אורך המילה הקצרה ביותר בקשר המגדיר. ככלל רזולוציה קטנה מ-3 היא בעייתית. בידוד האינטראקציות מסדר נמוך הטוב ביותר שניתן להשיג ברזולוציות נמוכות הוא:

i. 3: השפעות ראשיות מסוימות מתערבבות עם אינטראקציות זוגיות.

ii. 4: אינטראקציות זוגיות מסוימות מתערבבות עם אחרות.

iii. 5: אין ערבוב בין אינטראקציות מסדרים 1-2.

(7) ANOVA: ניתוח שונות בין יותר משתי קבוצות

1. הבעיה: בחינת השונות בין $2 < k$ תתי קבוצות של דאטא.
2. **<u>הנחות</u>**: דגימות <u>**iid**</u> <u>**נורמליות**</u> מקבוצות עם תוחלות שונות ו<u>**שונויות שוות**</u> ($x \sim N(\mu_i, \sigma^2)$).

ויזואליזציה

1. עבור מעט נתונים ניתן לעשות ישירות plot(group_id,x) (ובמידת הצורך לעבור לסקאלה לוגריתמית).
2. **Boxplot**.

האם ישנם הבדלים בין הקבוצות?

ANOVA = AOV = **Analysis Of Variance**:

1. המודל מניח שכל קבוצה מתפלגת נורמלית עם שונות זהה, ובודק אם קיימים הבדלים בתוחלות:

$$Y_{i,j} \sim N(\mu_i, \sigma) = N(\mu + \alpha_i, \sigma)$$

או באופן שקול – The One-Way ANOVA Model:

$$\boldsymbol{Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}} \qquad \epsilon_{i,j} \sim N(0, \sigma^2),\ \ \sum \alpha_i = 0$$

2. תחת השערת האפס אין שונות בין הקבוצות ($\alpha_i \equiv 0$), ולכן $\bar{Y_i} - \bar{Y}$ מתפלג נורמלית סביב 0 עם שונות הנגזרת מ-σ ומ-n_i לפי משפט הגבול המרכזי. עם זאת, σ אינו ידוע, ואם נשווה את $\bar{Y_i} - \bar{Y}$ לסטיית התקן המדגמית של הדאטא בפועל, לא נדע אם מקור ההבדלים בין $\bar{Y_i} - \bar{Y}$ השונים הינו בשונות בתוך הקבוצות או בין הקבוצות.
3. במקום זאת **נשווה בצורה ישירה בין השונות בתוך הקבוצות לבין השונות בין הקבוצות:**

a. Sum Squares Total = Between + Within:

$$\boldsymbol{SST} = \sum\left(Y_{i,j} - \bar{Y}\right)^2 = \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{n_i(\bar{Y_i} - \bar{Y})^2} + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{i,j} - \bar{Y_i}\right)^2 = \boldsymbol{SSB + SSW}$$

i. נרמול: $\boldsymbol{MST} \coloneqq SST/(N-1)$, $\boldsymbol{MSB} \coloneqq SSB/(k-1)$, $\boldsymbol{MSW} \coloneqq SSW/(N-k)$ (Mean/Sum Squared Total/Between/Within).

ii. תחת H_0, כל אחד מהשלושה MST,MSB,MSW הוא אמדן ח"ה ל-σ^2.
iii. **תחת H_1, $E[MSB] > E[MSW] = \sigma^2$**.
iv. סטטיסטי המבחן $F := \frac{MSB}{MSW}$ מתפלג תחת H_0 כמו מנת משתני χ^2 (ריבועי משתנים נורמליים) – **התפלגות $F_{k-1,N-k}$**, וכן p-value=$P\left(F_{k-1,N-k} > F\right)$.

b. **עבור $k = 2$ מתקבל p-value זהה למבחן t** דו"צ תחת הנחת שונויות שוות.
c. כרגיל, דיוק המודל עלול להיות **רגיש להנחות – אי תלות** של הדגימות, **נורמליות** של הדגימות, **ושונות זהה** בתוך הקבוצות.

היכן נמצאים ההבדלים ומהם?

1. אם נמצאו הבדלים בין הקבוצות, ניתן לבצע השוואה בזוגות לזיהוי הבדלים פרטניים.
2. בהנחה $\epsilon_{ij} \sim N$ מתקבל $\frac{(\bar{Y}_m-\mu_m)-(\bar{Y}_l-\mu_l)}{\sqrt{MSW}\sqrt{\frac{1}{n_m}+\frac{1}{n_l}}} \sim t_{N-k}$ (האמנם? ההפרש בין שתי התפלגויות t לא אמור להיות התפלגות t...) ומכאן ניתן לגזור p-value להשערה $\mu_m = \mu_l$ (H_0).
3. רמת מובהקות במקרה של השוואות מרובות:
   a. **IER** (Individual Error Rate): הסבירות לטעות (מסוג I) בהשוואה פרטנית של זוג.
   b. **FWER** (Family-Wise Error Rate): הסבירות לטעות אחת לפחות בעת השוואת זוגות מרובים.
   c. במקרה שלנו כברירת מחדל ישנן $\binom{k}{2}$, ולכן כדי לקבל את FWER הרצוי, נדרש IER קטן בהתאם.
   d. כדי לאפשר IER גדול יותר (ולקבל עוצמה גבוהה יותר למבחן המשותף), ניתן לצמצם את מספר ההשוואות (למשל להשוות כל קבוצה רק לקבוצת הביקורת או לממוצע הכללי, ולא כל זוג קבוצות).
4. **תיקון Bonferroni**: עבור M השוואות, ניתן לבחון כל השוואה עם רמת מובהקות α/M.
   a. מתקבל מבחן שמרני מהרצוי: $P(there's\ an\ error) \le \sum_i P(error\ in\ i)$.
5. **שיטת Tukey**: רווח סמך "סימולטני" לכל הזוגות $\mu_m - \mu_l$ (המתאים לסבירות α ש<u>לפחות אחד</u> מהפרשי התוחלות יחרוג מהרווח): רווח סמך לסטטיסטי $\max_{m,l} \frac{\bar{Y}_m-\bar{Y}_l}{\sqrt{MSW(1/n_m+1/n_l)}}$.
   a. עבור גדלי מדגמים שווים ($n_i \equiv n$), ניתן לקבל בדיוק את רמת המובהקות הרצויה α, ולכן המבחן עדיף על פני תיקון Bonferroni כאשר מבצעים השוואה של כל $\binom{k}{2}$ הזוגות.
   b. עבור גדלי מדגמים שונים מתקבל מבחן שמרני יותר מהרצוי.
   c. במקרה של $m < \binom{k}{2}$ השוואות, יש לחשב את גודל רווח הסמך (שתלוי בגודל הקבוצות בדאטא אך לא בדאטא עצמו) עבור כל שיטה, ולבחור מראש את המבחן עם רווח הסמך הקטן יותר.

ר"ס Tukey :

- $$(\bar{y}_i-\bar{y}_j) \pm \frac{1}{\sqrt{2}} q_\alpha(k, N-k) \sqrt{MSW\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

ר"ס Bonferroni :

$$(\bar{y}_i-\bar{y}_j) \pm t_{1-\frac{\alpha}{2m}, N-k} \sqrt{MSW\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

6. **שיטת Scheffe**:
   a. **Linear Contrast** של תוחלות $\{\mu_i\}_i$: $\psi = \sum_i c_i \mu_i$ כאשר $\sum c_i = 0$.
      i. קונטרסט לינארי מבטא השוואה בין אוסף קבוצות אחד ($c_i > 0$) לבין אוסף קבוצות אחר ($c_i < 0$).
      ii. דוגמה (מקרה פרטי) – השוואה פרטנית בין שתי קבוצות: $\mu_i = 1, \mu_j = -1, \mu_{l\neq i,j} = 0$.
      iii. דוגמה – השוואה של קבוצת הביקורת לממוצע הקבוצות האחרות: $\psi = \mu_1 - \sum_{i=2}^{k} \mu_i/(k-1)$.

iv. **השוואה בין אוספי קבוצות בקונטרסט יחיד**: עבור דגימות נורמליות ב"ת $\frac{\hat{\psi}-\psi}{\sqrt{MSW\sum_{i=1}^{k}c_i^2/n_i}}\sim t_{N-k}$ (לא ברור לי למה), כאשר $\psi=0$ תחת H_0.

b. **Scheffe**: תחת הנחות ה-ANOVA ו-H_0, מתקיים $P\left(\max_{con}\frac{\widehat{\psi_{con}}^2}{(k-1)MSW\sum\frac{c_i^2}{n_i}}\right)\leq F_{1-\alpha,k-1,N-k}$.

i. הסטטיסטי של Scheffe הוא הכללה של Tukey עבור קונטרסטים.

ii. הסטטיסטי הוא **המקסימום על כל הקונטרסטים האפשריים**, לרבות כל ההשוואות הזוגיות, ולכן שמרני יותר מ-Tukey.

7. סיכום:

a. השוואה בין כל הזוגות: Tukey.

b. השוואה בין זוגות מסוימים של קבוצות: Bonferroni או Tukey (לפי רווח הסמך הקטן יותר; ככל שמספר הזוגות גדל יש עדיפות ל-Tukey).

c. השוואה בין אוספים מסוימים של קבוצות: Bonferroni או Scheffe (לפי רווח הסמך הקטן יותר; ככל שמספר האוספים גדל יש עדיפות ל-Scheffe).

השוואות מרובות ללא הנחות ה-ANOVA

1. הסתמכות על **דירוג הנתונים** בלבד ($x_1 \dots x_N \to 1 \dots N$) **מצריך הנחת אי תלות** בין הדגימות אך **לא מצריך את הנחות הנורמליות או השונות הקבועה**.
2. פיתרון פשוט: **Wilcoxon Rank Sum Test** (ראה פרק 3) עם תיקון Bonferroni.
3. מבחן **Kruskal-Wallis** – הסטטיסטי הוא פשוט SSB מנורמל על הדירוגים:

$$kw=\frac{12}{N(N+1)}\sum_{i=1}^{k}n_i(\bar{R_i}-\bar{R})^2 \qquad \left(\bar{R}=\frac{N+1}{2}\right)$$

a. עבור מדגמים גדולים ($n_i\geq 5$), kw מתפלג בקירוב χ_{k-1}^2.

b. עבור מדגמים קטנים, ניתן לחשב p-value ל-kw ישירות לפי בדיקת כל הפרמוטציות האפשריות על הדרגות.

ניתוח שונות עם אפקטים מקריים

1. מודל ANOVA עם **אפקטים מקריים**: הקטגוריות הנמדדות הן רק מדגם מאוכלוסיית הקטגוריות.

a. דוגמה: מדידה של משכורות כדורגלנים ב-5 ליגות שונות, כמדגם למשכורות בכלל הליגות בעולם.

b. ההשפעות $\alpha_1 \dots \alpha_k$ הן כעת משתנים מקריים המתפלגים $N(0,\sigma_a^2)$.

c. σ_a^2 זו **שונות האוכלוסיה**, השונה מ**שונות הדגימה** σ^2. כל דגימה מקיימת $Y_{ij}\sim N(\mu,\sigma_a^2+\sigma^2)$, כאשר **קיימת שונות משותפת** σ_a^2 (קורלציה $\sigma_a^2/(\sigma_a^2+\sigma^2)$) לדגימות מאותה הקטגוריה.

d. בזהה למקרה של אפקטים קבועים, סטטיסטי המבחן הוא $F=\frac{MSB}{MSW}\sim F_{k-1,N-k}$.

2. אמידת התוחלת הכללית עבור אפקטים מקריים עם מדגם מאוזן: $\frac{\bar{Y}-\mu}{\sqrt{\frac{SSB}{N(k-1)}}}\sim t_{k-1}$.

a. יש k-1 דרגות חופש בלבד, משום שהמידע הנאסף אודות μ מוגבל ל-k הקטגוריות הנדגמות גם כאשר יש דגימות רבות (אבל יש גם מספר סופי של דגימות, אז למה לא פחות מ-k-1?).

3. אומד חסר הטיה ל-σ^2 הוא MSW.
4. אומד טבעי ל-σ_a^2 הוא $\frac{MSB-MSW}{N/k}$ (עלול לצאת שלילי, ובמקרה זה נהוג להניח כי $\sigma_a^2=0$).

(8) ניסוי עם הקצאה אקראית בבלוקים

1. **Randomized Complete Block Design** (**RCBD**):

a. מקור הטרמינולוגיה בניסויים חקלאיים (השפעת דשן / שיטת גידול על היבול).
b. הבעיה: השוואה בין I קבוצות (דשנים) כאשר יש גורם משפיע (גורם הטרדה, **Nuisance Factor**) עם J ערכים שונים (חלקות אדמה = בלוקים), כאשר הגורם המשפיע (בלוקים) ידוע מראש והקצאת יחידות הניסוי (שתילים) לקבוצות (דשנים) ניתנת לשליטה.
c. הפיתרון: עבור כל ערך של J (בלוק), הדאטא יחולק באקראי ל-I קבוצות שוות שיוקצו באקראי לקבוצות הניסוי.
d. זוהי הכללה של מדגם מזווג (לפני-אחרי עבור אותו נבדק) למספר קבוצות ניסוי (קבוצות שונות עבור אותו בלוק).

2. המודל:

$$\boldsymbol{Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}} \qquad \sum\alpha_i = \sum\beta_j = 0,\ \epsilon_{ij} \sim N(0,\sigma^2)$$

3. הסקה סטטיסטית:

$$\boldsymbol{SSA} \coloneqq J\sum_{i=1}^{I}(\bar{y}_{i\cdot} - \bar{y})^2\,, \boldsymbol{SSB} \coloneqq I\sum_{j=1}^{J}\left(\bar{y}_{\cdot j} - \bar{y}\right)^2, \boldsymbol{SSE} \coloneqq \sum_{i,j}\left(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}\right)^2, \boldsymbol{SST} \coloneqq \sum_{i,j}\left(y_{ij} - \bar{y}\right)^2$$

a. $SST = SSA + SSB + SSE$
b. $MSA \coloneqq \frac{SSA}{I-1},\ \ MSB \coloneqq \frac{SSB}{J-1},\ \ MSE \coloneqq \frac{SSE}{(I-1)(J-1)},\ \ MST \coloneqq \frac{SST}{IJ-1}$
c. $E(MSA) = \sigma^2 + \frac{J}{I-1}\sum\alpha_i^2,\ \ E(MSB) = \sigma^2 + \frac{I}{J-1}\sum\beta_j^2,\ \ E(MSE) = \sigma^2$
d. $H_0\colon \alpha_i \equiv 0$ (for blocks: $\beta_j \equiv 0$)
e. **Test statistic for groups:** $\boldsymbol{F_A \coloneqq \frac{MSA}{MSE} \sim F_{I-1,(I-1)(J-1)}}$ (under H_0)
   i. $\bar{Y}_{i\cdot}$ is unbiased estimator of $\mu + \alpha_i$ with variance σ^2/J.
   ii. MSE is unbiased estimator of σ^2.
   iii. Comparing pairs or contrasts of groups can be done similarly to ANOVA.
   iv. **It is claimed that F-test is relatively insensitive to slight a-Normality of $\boldsymbol{\epsilon_{ij}}$.**
f. Test statistic for blocks: $F_B \coloneqq \frac{MSB}{MSE} \sim F_{J-1,(I-1)(J-1)}$ (under H_0)
   i. Relevant for future experiments design.

4. **מבחן פרידמן** (**Friedman**):
a. מבחן על דירוג התצפיות במקום ערכיהן – כל תצפית מקבלת את דירוגה בתוך הבלוק.
b. סטטיסטי המבחן (מניח בגרסה זו תצפית יחידה לכל בלוק-קבוצה; כמו כן, התפלגות χ^2 היא קירוב בלבד עבור מספר גדול של תצפיות בכל קבוצה – אחרת יש לגזור p-value ממבחן פרמוטציות):

$$Q_i \coloneqq \left(\boldsymbol{\bar{R}_{i\cdot}} - \boldsymbol{\frac{I+1}{2}}\right)^2, \quad Q \coloneqq \frac{12J}{I(I+1)}\sum_{i=1}^{I} Q_i \sim \boldsymbol{\chi^2(I-1)}$$

c. ב-R: library(coin); help(friedman_test).

5. הערות כלליות: מודל RCBD מנטרל השפעה של גורמים ידועים, אולם מאבד דרגת חופש עבור כל בלוק, וכן מניח שהשפעת קבוצות הניסוי α_i קבועה בבלוקים השונים (ללא אינטראקציה עם הגורם המשפיע החיצוני). כמו כן המודל מתקשה יותר לעבוד עם תצפיות חסרות.

ריבוע לטיני

1. כאשר יחידות הניסוי מאופיינות ע"י שני משתנים (לדוגמה קו גובה וקו רוחב בתוך בלוק חקלאי; יום ופריט בניסוי על פריטים שונים בימים שונים), ניתן לסדר את יחידות הניסוי בטבלה ריבועית לפי שני המשתנים ולהקצותן לקבוצות הניסוי כך שלכל קבוצה ייצוג שווה בכל שורה ובכל עמודה.

a. סידור בריבוע לטיני מניח שכל אחד מהמשתנים החיצוניים משפיע, אך ללא אינטארקציה ביניהם.

(9) Fixed Two-way ANOVA: ניתוח שונות דו-כיווני עם אפקטים קבועים

1. ניתוח שונות דו-כיווני = בחינה של שני גורמים משפיעים בו"ז ("ניסוי כפול").
   a. דוגמה: בחינת התועלת של שני טיפולים רפואיים (קבוצות הניסוי: בלי/עם טיפול א X בלי/עם ב).
2. המודל:

$$\boldsymbol{Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}}$$

   a. ההשפעות הראשיות והאינטראקציות בכל שורה/עמודה נסכמות ל-0.
   b. אינטראקציות מוסיפות הרבה דרגות חופש. ניתן לבחון גרפית קיום אינטראקציות ע"י plot(x=i,y=mean y over k, by=j) – בהיעדר אינטראקציות (כלומר אם ההשפעות אדיטיביות יתקבלו גרפים מקבילים (בעלי מרחקים אחידים זה מזה לאורך ציר x) בקירוב.
3. ניתן לפרק את השונויות המדגמיות בניסוי ל-$\boldsymbol{SST = SSA + SSB + SSAB + SSW}$.
   a. חישוב השונויות המדגמיות עבור מספר תצפיות קבוע m לכל תא:

$$SSA = mJ\sum_{i=1}^{I}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$$

$$SSB = mI\sum_{j=1}^{J}(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2$$

$$SSAB = m\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2$$

$$SSW = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{m}(y_{ijk} - \bar{y}_{ij\bullet})^2$$

   b. סטטיסטי המבחן:

$$F_A \coloneqq \frac{MSA}{MSW} \sim F_{I-1,(m-1)IJ}, \quad F_B \coloneqq \frac{MSB}{MSW} \sim F_{J-1,(m-1)IJ}, \quad F_{AB} \coloneqq \frac{MSAB}{MSW} \sim F_{(I-1)(J-1),(m-1)IJ}$$

4. הנחת הנורמליות ניתנת לבחינה ע"י qnorm (Normal Quantile plot) על השאריות לאחר החלת המודל.

Random & Mixed Two-way ANOVA

1. המודל:

$$\boldsymbol{Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \epsilon_{ijk}} \qquad a_i \sim N(0, \sigma_a^2), b_j \sim N(0, \sigma_b^2), (ab)_{ij} \sim N(0, \sigma_{ab}^2), \epsilon_{ijk} \sim N(0, \sigma^2)$$

2. סטטיסטי המבחן לבחינת השערות ה-0:

$$E(MSA) = \sigma^2 + m\sigma_{ab}^2 + Jm\sigma_a^2 \qquad F_A = \frac{MSA}{MSAB} \sim F_{I-1,(I-1)(J-1)}$$

$$E(MSB) = \sigma^2 + m\sigma_{ab}^2 + Im\sigma_b^2 \qquad F_B = \frac{MSB}{MSAB} \sim F_{J-1,(I-1)(J-1)}$$

$$E(MSAB) = \sigma^2 + m\sigma_{ab}^2 \qquad F_{AB} = \frac{MSAB}{MSW} \sim F_{(I-1)(J-1),(m-1)IJ}$$

3. ניתן לעבוד עם מודל דומה במקרה המעורב (Mixed Two-way ANOVA): גורם משפיע אחד קבוע ואחד מקרי.

(10) אופטימיזציה בייסיאנית

1. ***Bayesian optimization***: approach in decision theory that treats the goal function/cost as random function (usually through dependence of that function on unknown parameters, i.e. $L(\theta, action)$), and takes actions according to that stochastic cost.
2. Bayesian optimization can be useful in experimental design when trying to estimate parameter or find optimal action. That's particularly true if the experimental data is costly to gather – either because the experiment is expensive or because it is conducted as part of the process which is tried to be optimized (***exploration-optimization tradeoff***).
3. Bayesian optimization is based on a prior distribution assigned to the parameter from which the action is derived, and not from convexity of $L(\theta)$ as in gradient-based optimization. In addition to avoiding the convexity assumption, it naturally handles stochasticity in the results, which is simply reflected in the posterior distribution.
4. Examples of implementations of Bayesian optimization in experimental design:
    a. ***Bayesian experimental design***: quantify both experiment cost and information gain, and choose the next sample to maximize the (expected) utility:

    $$U(a) = \int \left(\int (info(a,x) - cost(x)) dp_{\theta,a}(x) \right) d\pi(\theta)$$

        i. Information gain $info(a,x)$ can be quantified, for example, through Shannon's entropy (after - before), or as KL-divergence between prior & posterior $\pi(\theta)$.

    b. ***Thompson sampling***: for each sample, draw θ randomly from its current distribution $\pi(\theta)$, then choose optimal action assuming this θ, and update the distribution $\pi(\theta)$ according to the new observations.