Table of Contents

2.1. Outlier Tests

There are 2 types of outlier tests.

1. Single outlier detection
2. Multiple outlier detection

Doornbos test

- For single outlier detection
- Uses externally studentized values (CHECK NORMALITY)
- Tests if each observation is an outlier, conducts n tests in total
- Similar to outlier detection in regression analysis (ne demekse artik)

Grubbs test

- Uses standardized values
- Approximates criterion through the t-distribution (CHECK NORMALITY)
- **Based on underlying normal distribution** and uses exact criterion
- Similar to outlier detection in regression analysis (ne demekse artik)

Dixon test

- Uses difference between ordered values instead of standardized or studentized values (CHECK NORMALITY)

Hampel's rule

- Considered as a non-parametric test
- Multiple detection
- An observation is called an outlier absolute normalized value > 3,5 (for exact formulas see slide 31)

Tukey's rule

- Non-parametric
- Multiple detection

- Embedded in boxplot, outliers are visible outside the whiskers

```
PROC BOXPLOT DATA=COLOR;
PLOT COUNT*REGION/BOXSTYLE=SCHEMATIC;
RUN;QUIT;
```

Comments

- Multiple testler (hampel ve tukey) k tane extreme valuenun outlier olup olmadigini test eder
- Single outlier detect eden testlerde birden fazla outlier olup olmadigini test ederken 2 yontem mumkun
    1. Distan ice (en extremden daha az extreme)
    2. Icten disa **(tercih edilen yontem bu)**
        - N tane gozlemden k tanesinin outlier oldugunu test ederken. Gozlemleri siralayip en kucuk n-k gozlemi n-k+1.ye karsi test ediyorsun. n+k-1 outliersa ondan daha ekstrem olan k tanesi zaten outlier olur ve isimiz biter. Degilse onu da en kucuk n-k ya katip bi buyugu test ediyoruz. Boyle boyle n.yi test edene kadar devam edebiliriz.

2.2. Two independent samples

t-test

- Tests if two sample means are equal
- Assumes normality but robust against deviations from normality
- Essentially assumes two variances are equal. But it can be extended to handle unequal variances. SAS automatically produces results for both cases.
- When you test if the variances are equal before applying the t-test, because of type 1 error you may incorrectly think that variances are equal

```
PROC TTEST DATA=CUS;
CLASS BATCH;
VAR CU; RUN;
```

Data must be in long format. (long format ne ola ki?)

Homogenity Tests

Testing if two variances are equal
Heteroskedastic = unequal variance.

F-test

- Assumes normality
- Automatically conducted during the t-test in SAS

Bartlett's test

- Also assumes normality, it is highly sensitive for it. (do not use if not sure that the data is normally distributed). Even tough there is a modification for handling non-normal data, it is not implemented in SAS.
- It can be extended to more than 2 groups

```
PROC GLM DATA=CUS; CLASS BATCH;
MODEL CU = BATCH;
MEANS BATCH / HOVTEST=BARTLETT; RUN;QUIT;
```

Levene's and Brown&Forsythe test

- Nonparametric
- Brown & Forsythe adjustment uses median instead of mean in calculations. (Mean ve median arasinda ciddi bir sapma oldugunu dusunuyorsak, ya da sample size kucukse bunu kullanalim.) BF Levene gore daha robust bir test. Bunu kullanmak overallda daha iyi gibi.

  ```
  PROC GLM DATA=CUS; CLASS BATCH;
  MODEL CU = BATCH;
  MEANS BATCH / HOVTEST=Options; RUN;QUIT;
  ```

  Levene's test: HOVTEST=LEVENE;
  Brown & Forsythe test: HOVTEST=BF;

Wilcoxon rank sum

- Nonparametric
- Tests if two medians are equal
- Assumes that observed variable is continuous
- Assumes that two groups come from distributions of identical shape
    - Last bullet implies equality of variances, but it is not that strict.
- Might be sensitive to ties, when calculating the statistic ranks the observations (2 numarali continuous assumptioni bu tielari engellemek icin onemli. Tie oldugunda average rank alip devam ediyor.)

```
PROC NPAR1WAY DATA=CUS; CLASS BATCH;
VAR CU;
EXACT WILCOXON/MC; RUN;
CORRECT=NO
```

- Exact hesaplama suresi gozlem sayisi arttikca exponential artiyor. Monte carlo simulasyon versiyonu asagidaki gibi.
- *Bu kodun sonuna bi yere CORRECT=NO ekleyince continuity correctioni yapmiyormus. Koyalim ozellikle belirtmisler week3.1de*

Mann-whitney test

- Although having subtle differences, it is the mathematical equivalent of Wilcoxon rank sum test

3.1. Independent samples (continuation of 2.2)

Kolmogorov-smirnoff test

- Only two assumptions
  - Two samples are independent
  - Outcome variable is ordinal or numerical

```
PROC NPAR1WAY DATA=CUS; CLASS BATCH;
VAR CU;
EXACT KS /MC; RUN;
```

Binary outcome tests

- Binary durumunda degiskenin Bernoulli triallar ile dagildigi assume edilir. Null hypothesis her iki samplein Bernoulli success rateinin, yani p lerin, esit oldugudur.

Chi-square test

- Summarizes the observations in 2x2 matrix in binary rows and columns
- It is an asymptotic test, so not fit for small sample size, when n <= 20 go for Fisher's exact test
- Assumes chi-square distribution underneath which implies normality
- 2x2 matrixteki hucrelerden herbirinin 5in uzerinde olmasi lazim. Yine large sample sizein onemini goruyoruz.
- Can be used for more than 2 groups
- Can handle nominal or categorical variables , yani 4x5 bi matrixle calisabilir. Binary olayinin disina cikabiliyor.

```
PROC FREQ DATA=CUS;
TABLES BATCH*[binary variable]/ CHISQ;
RUN;
```

Fisher's exact test

- Exact test, so does not require large sample size
- It is considered as conservative, not recommended by all statisticians
- Automatically calculated when chi-square test is conducted in SAS

3.2. Correlation

Pearson's correlation

- Measure of linear dependency, non-linear iliskileri yakalayamiyor
- Implies normality, when the assumption may be violated z-transformation of it can be utilized (z transformation zaten dagilimi normalize etmek icin kullaniliyor)
- Bunlarin hepsi mutlak degerli (phi coefficient icin de gecerli)
  - $|p| <= 0.3$ little correlation
  - $|p| <= 0.5$ weak correlation

    - |p|<=0.7 medium correlation
    - otherwise correlation

Normal pearson:

```
PROC CORR DATA=GRADES PEARSON;
VAR PRE EXAM; RUN;
```

Z-transformationla:

```
PROC CORR DATA=GRADES
PEARSON FISHER(BIASADJ=NO);
VAR PRE EXAM; RUN;
```

Intraclass correlation coefficient

- Special case of pearson, see slide 11 of week 3.2

Phi correlation coefficient (binary variable)

- binary version of pearson correlation

```
PROC FREQ DATA=GRADES;
TABLES [column1]*[column2]/CHISQ; RUN;
```

Statistic	DF	Value	Prob
Chi-Square	1	0.4167	0.5186
Likelihood Ratio Chi-Square	1	0.4153	0.5193
Continuity Adj. Chi-Square	1	0.0116	0.9143
Mantel-Haenszel Chi-Square	1	0.3889	0.5329
Phi Coefficient		-0.1667	
Contingency Coefficient		0.1644	
Cramer's V		-0.1667	
WARNING: 75% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Spearman's rank correlation

- nonparametric, adi uzerinde rank uzerinden ilerliyor
- called spearman's rho correlation in SAS

```
PROC CORR DATA=GRADES SPEARMAN;
VAR PRE EXAM; RUN;
```

Kendall's tau correlation

- original definition is based on concordance
    - pairs x1,y1 and x2,y2 are concordant if, (x1-x2)*(y1-y2) > 0, discordant otherwise (demek ki sadece x ve y degerlerinin birlikte azalip artmasina bakiyor)
- $\tau=2P(x_r-x_s\ y_r-y_s>0)-1,\ r\neq s$. Tau=0 ise independent oluyor
- formulu yukardaki gibi, demek ki ne kadar artip azaldigini degil sadece artma ve azalmanin birlikte olup olmadigini control ediyor.
- Uses large sample theory (sanki sample sayisi yuksek olsa daha iyi olur gibi, zaten slideta yazdigi uzere buldugu istatistigi normal quantilei ile karsilastiriyor)
- Adjusts the math in case of ties

```
PROC CORR DATA=GRADES KENDALL;
VAR [variable(sample)1] [variable2]; RUN;
```

(screenshottaki 0.27 tau, 0.17 p-value)

Kendall Tau b Correlation Coefficients, N = 15
Prob > |tau| under H0: Tau=0

	PRE	EXAM
PRE	1.00000	0.27412 0.1736
EXAM	0.27412 0.1736	1.00000

COPULALARLA EVAM SLIDE 40 WEEK3.2

4.1. Paired samples

Paired t-test (continuous variables)

- assumes normality
- null hypothesis is mu1 - mu2 = 0, or mu1/mu2 = 1, or log difference between means is zero (sonuncusu slidelarda medianlarin esitligi olarak da belirtilmis. Tam olarak baglayamadim konuyu. Slide 5)

Sign test (continuous variables)

- nonparametric, no distributional assumptions
- null hypothesis is medians of two samples are equal (bu durumda small sample size icin kullanilabilir)
- teker teker her pair digerinden yuksek mi diye bakiyor. Tie oldugunda o pairi yok sayarak istatistigi hesapliyor. Sonra da Bernoulli triallardan (p=0.5) olusan binomial distributionla karsilastirarak p-value hesaplaniyor.
- Farklarin ne kadar buyuk olduguna bakmiyor, sadece adi uzerinde sign'a bakiyor.

Wilcoxon signed rank test (continuous variables)

- Assumes a symmetric distribution (check the histogram or kernel density graph before running the test)
- Null hypothesis is medians are equal
- In case of ties, sample size decreases just like sign test
- Takes differences of pairs and ranks them for calculating the test statistic, for ties in differences takes average rank
- Can handle the ties (Tielar icin variance'i adjust ediyor)
- Asympotic statistic. Compares the statistic with quantiles of normal distribution (sanirsam mean ve variance'i duzelterek) (bu sebepten kaynaklaniyor symmetric distribution assumptioni)
- For Wilcoxon rank sign test the p-value is exact up to n=20 in SAS

SAS code for continuous paired test

- Bu testleri run etmeden once manuel olarak iki variable birbirinden cikartman gerekiyor. #amelasyon
- bu farklari asagida zd zl olarak gordugun yerlere yaziyorsun. Zd=difference, zl= log difference, zr = ratio
- Median ve mean farklari (null hypothesis) otomatik sifir olarak ataniyor
- For Wilcoxon rank sign test the p-value is exact up to n=20 in SAS

```
PROC UNIVARIATE DATA=GRADES NORMAL;
VAR ZD ZL ZR (cesitli variable);
HISTOGRAM ZD ZL ZR (cesitli variable);
```

```
RUN;
```

Mcnemar test (binary variables)

- The null hypothesis is marginal homogeneity. For details see slide 26. (aslinda iki variablein da 1 olma ihtimalinin ayni oldugunu null hypo yapiyorsun)
- Compares the statistic with chi square distribution to calculate p-values

```
PROC FREQ DATA=GRADES;
TABLES B1*B2/AGREE;
RUN;
```

- For small sample size use the exact version

```
PROC FREQ DATA=GRADES;
TABLES B1*B2;
EXACT MCNEM;ne
RUN;
```

Kappa agreement

- Binary variablelarin correlationi gibi bir sey
- Ters correlation durumunda negative de olabilir
- Degerleri asagidaki gibi yorumlamak lazim
- Sensitive to row and column totals, should not rare or non-event

• High:
• Substantial:
• Moderate:
• Fair:
• Poor:
$0.80 < \kappa \leq 1$ $0.60 < \kappa \leq 0.80$ $0.40 < \kappa \leq 0.60$ $0.20 < \kappa \leq 0.40$ $0 < \kappa \leq 0.20$

- SAS provides kappa statistic with McNemar test. ASE value in the table is the standard deviation for Kappa.

```
PROC FREQ DATA=GRADES;
TABLES B1*B2/AGREE;
RUN;
```

4.2. One-way anova

- fixed or random effect
- random: a random collection of a population is included in test, the interest is in the full group
- fixed: all groups are included and interest is on these groups only

mathematical model: $yij = \mu + \alpha i + eij$

assumption fixed: alpha_i is an unknown fixed parameter. All alphas sum up to 0.
Assumption random: alpha_i is a normally distributed random parameter with mean 0

Statistical assumptions of ANOVA:

- Normality of the residuals
- Homogeneity of residual variance across groups
- Normality of random effects
- Random effects and residuals are independent from each other

Confidence intervalsiz hali icin CL'leri sil.
Fixed effect:
PROC MIXED DATA=ANOVA METHOD=TYPE3 CL;
CLASS TRT;
MODEL RESP = TRT /SOLUTION OUTP=PRED(predi istedign isimle degistirebilirsin) CL;
RUN;

Random effect:
PROC MIXED DATA=ANOVA METHOD=TYPE3 CL;
CLASS ID;
MODEL RESP = /SOLUTION OUTP=PRED(predi istedign isimle degistirebilirsin) CL;
RANDOM ID;
RUN;

- kirmiziyla yazidigim yerler residuallari falan almak icin. Diger tum outputlar da orda oluyordu yanlis hatirlamiyorsam. Arkasindan anova assumptionlarini test ediyoruz hemen:

normality:
PROC UNIVARIATE DATA=PRED NORMAL;
VAR RESID;
RUN;

Homogeneity of variance: burda bf kullanilmis uygununu yukariya bakip dusunup bulmak lazim
PROC GLM DATA=PRED;
CLASS TRT;
MODEL RESID = TRT;
MEANS TRT/ HOVTEST=BF;
RUN;

Baktik assumptionlar tutmuyor anovayi kullanamiyoruz. One way fixed anovanin non-parametric alternatifi var kruskal-wallis diye onu asagida (5.1 icinde) gorebiliriz.

5.1. Normality tests

Graphical techniques

- Gozunle normal mi degil mi diye bakiyorsun. Baya subjectif sacma bi olay ama tekniktir nihayetinde.

```
PROC UNIVARIATE DATA=[datanin adu];
VAR [incelenek variablein adi];
PROBPLOT [variable adi]/NORMAL(MU=est SIGMA=est);
RUN;
```

Kolmogorov-smirnoff test

- Used to test against a prefined normal distribution (mu ve sigmasini belirtmen lazim. Aslinda normal mi diye degil, spesifik olarak verilen bu normal dagilima benziyor mu diye test ediyor)

```
PROC UNIVARIATE DATA=[data adi];
VAR [incelenen variable adi];
HISTOGRAM [variable]/NORMAL(MU=100 SIGMA=3);
RUN;
```

Cramer-von mises

- Cok bi olayi yok bu testin hangi durumda kullanilacak onu da pek cozemedim.

Anderson-darling

- Almost the same thing with cramer – von mises, but gives more weight to the tails of the distribution
- Has often the same power as the Shapiro-Wilk, but on overall it is somewhat less powerful.
- Is less sensitive to ties
- Recommended as omnibus test

Shapiro-wilk

- Tries to fit a regression line to come up with a normality test. (see slide 19)
- A very sensitive and powerful test against non- symmetric departures from normality
- Probably the most powerful test of all others and thus recommended as omnibus test
- Sensitive to ties in the data: when ties are a result of rounding, than normality can incorrectly be rejected

```
PROC UNIVARIATE DATA=CU NORMAL; VAR CU;
RUN;
```

Skewness and kurtosis tests

- Boyle testler de var ama tam ne ise yaradiklarini cozemedim. Bi tik manuel bi is. Yukardaki kodun sonucunda cikan tablodan degerleri alip slidelardaki critical valuelarla karsilastirmak gerekiyor. (slide 37-38de hesaplari var daha onceki slidelarda da degisik sample sizelar icin crtitical valuelar var)

Kruskal-wallis (non-parametric alternative of one way fixed anova)

- Extension of Wilcoxon rank sum test for more than two medians

$yij = \theta + \alpha i + eij$

• With θ an overall median
• With αi the fixed effect of group i on the median
• With eij i.i.d. F having a median 0

```
PROC NPAR1WAY DATA = [DATANIN ADI] WILCOXON;
CLASS [TREATMENT];
VAR [RESPONSE VARIABLE];
RUN;
```

5.2. Randomness

bunun notlari defterde ve slidelarda sali eklersem eklerim.

6.1. multiple testing

for multiple testing we will continue to assumptions of ANOVA such as normality of residuals, homogeneity of variance across variance groups. see ANOVA part.

Least square difference

LSD approach: first run ANOVA to see if there is a difference, if ANOVA says there is a difference, continue with pairwise comparisons. (richardin dedigini hatirla burda: aslinda ilk olarak anova testini yapip multiple ile devam edince bastaki type 1 errorun uzerine yeni type 1 error ihtimalleriyle devam ediyorsun. direkt olarak multiple teste gecmek bu acidan mantikli olsa da kendim dusununce zaten multiple testingle anovanin ne baglantisi var, sadece go veriyor diye dusunuyorum.)

```
PROC MIXED DATA=ANOVA METHOD=TYPE3;
        CLASS TRT;
        MODEL RESP = TRT/SOLUTION CL;
        LSMEANS TRT/DIFF;
RUN;
```

Bonferoni and Sidak correction

FWER ucmasin diye pvalue ayarlamaca. Not: experimentwise correction nonparametric testlerde de uygulanabilir (see slide 37)

- Both corrections are considered conservative since they ignore dependence
- For independent tests, sidak is exact, bonferoni is an approximation

Ustteki koda ekle bunlari

```
LSMEANS TRT/DIFF ADJUST=BON;
LSMEANS TRT/DIFF ADJUST=SIDAK;
```

To add a control group:

```
LSMEANS TRT/DIFF=CONTROL('0') ADJUST=BON;
```

- Both sidak and bonferoni work best on small number of groups

Tukey (studentized range test)

- Focusses on all pairwise comparisons
- Tukey's test is more powerful than Bonferroni correction in case of all pair-wise comparisons
- It is more difficult for unbalanced data

```
LSMEANS TRT/DIFF ADJUST=TUKEY;
```

Scheffe

- The most conservative one
- Tam ne ise yariyor anlamis degilim

```
LSMEANS TRT/DIFF ADJUST=SCHEFFE;
```

Duncan's multiple range

- Tries to cluster means
- Sorts the treatment means and compares means pair-wise

```
PROC GLM DATA=ANOVA;
CLASS TRT;
MODEL RESP = TRT;
MEANS TRT/DUNCAN;
RUN;
```

Dunnett's one to many test

- For the cases of comparison with a control group
- In case of all pairwise comparisons similar to Tukey
- P-value is calculated from multivariate t-dist (i think this implies normality)
- Because all treatments are compared against a control group an imbalanced design is beneficial

```
PROC MIXED DATA=ANOVA METHOD=TYPE3;
CLASS TRT;
MODEL RESP = TRT/SOLUTION;
LSMEANS TRT / DIFF ADJUST=DUNNETT;
RUN;
```

Comparison of techniques from slides

The LSD is frequently applied:
• It is the only test that uses the p-value of the F-test for the joint hypothesis
• Is very powerful, but not most powerful when just a few groups from many groups are truly different
• It does not protect against false differences when just a few groups from many are truly different
• Is easy to misuse as approach
• It is very popular, but the disadvantages are often overlooked

Bonferroni/Sidak corrections
• Can be used for any type of comparisons
(contrasts or characteristics)
• Are almost exact for small number of tests
• Can be improved (Holm-Bonferroni)

Tukey's test compares all pairs
• Protects against false differences
• Is not always the most powerful, but better than Bonferroni/Sidak for larger groups
• Is more complicated for imbalances

Scheffé for all contrasts
• Is a very generic procedure since it can handle any type of group comparisons
• Is less powerful than many other more specialized tests (e.g. one-to-many)

Duncan's multiple range test
• The only test that tries to cluster the groups
• Does not protect FWER, which is considered a serious drawback

Dunnett's one-to-many
• Is the Tukey's test for all comparisons against a control group
• It is more complicated for imbalances (although a multivariate t-distribution could be used)
• It protects the FWER, where Scheffé and Bonferroni/Sidak does not do this for this task
• It is more powerful than Bonferroni/Sidak for larger number of groups
• Optimal sample sizes are frequently not applied (unfortunately)

6.2. Two way ANOVA

- fixed ve random effect olayinin nasil oldugunu zaten one way anovada cozmustuk.

To visualize the results first average volumes (if the measurements are done)

```
PROC SORT DATA=INPUT;
BY SUBJECT ONCOLOGIST REPEAT;
RUN;
```

```
PROC MEANS DATA=INPUT NOPRINT;
VAR VOLUME;
BY SUBJECT ONCOLOGIST;
OUTPUT OUT=MEANS MEAN=VOLUME;
RUN;
```

Then run actual anova

```
PROC MIXED DATA=MEANS METHOD=TYPE3;
CLASS SUBJECT ONCOLOGIST;
MODEL VOLUME=ONCOLOGIST/SOLUTION DDFM=SAT;(kirmizi yer onemli bir correction yapiyor)
RANDOM SUBJECT;
LSMEANS ONCOLOGIST;
RUN;
```

Bu ornekte subject random effect, oncologist fixed effect.

To add confidence intervals use this one:

```
PROC MIXED DATA=MEANS METHOD=TYPE3 COVTEST CL;
```

7.1 nested two way anova

Friedman test (non parametric alternative for repeated two way anova)

- Null hypo: beta_i for every treatment equals zero, alternative = there exists a beta_i different than zero
- Fixed effect ve random effect sayisi ne kadar buyukse o kadar iyi approximate ediyor

```
PROC FREQ DATA=MEANS;
TABLES SUBJECT*ONCOLOGIST*VOLUME / CMH2
SCORES=RANK NOPRINT;
RUN;
```

Summary Statistics for Oncologist by VOLUME
Controlling for Subject

Cochran-Mantel-Haenszel Statistics (Based on Rank Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	0.2400	0.6242
2	Row Mean Scores Differ	4	12.9333	0.0116

Total Sample Size = 30

$J-1$

Q

$P(\chi^2_{J-1} \geq Q)$

Nested effects ANOVA

- Doctor hasta organlari olcumu orneginde her doctor farkli bir hastaninkini olctugunde hasta farklililarindan ortaya cikan randomnessi dikkate alan bir anova

- Fixed icinde fixed, random icinde random, random icinde fixed ve fixed icinde random olmak uzere kurulabilir. See slide 14-17

Two-way (mixed) Nested Model:

```
PROC SORT DATA=INPUT;
BY SUBJECT ONCOLOGIST REPEAT;
RUN;


PROC MIXED DATA=INPUT METHOD=TYPE3 COVTEST CL;
CLASS SUBJECT ONCOLOGIST;
MODEL VOLUME=ONCOLOGIST/SOLUTION DDFM=SAT;
RANDOM SUBJECT(ONCOLOGIST);
RUN;
```

Interaction effect

Factor A dependent effect of factor B

• An interaction effect between two factors

• Is fixed when both factors are fixed, otherwise the interaction effect is random

• Means that the effect of one factor depends on the level of the other factor

• The difference in volumes between two oncologists changes with subject

• Visually the connecting lines between volumes of subjects over oncologists should be parallel

```
PROC SORT DATA=INPUT;
BY SUBJECT ONCOLOGIST REPEAT;
RUN;

PROC MIXED DATA=INPUT METHOD=TYPE3 COVTEST CL;
CLASS SUBJECT ONCOLOGIST;
MODEL VOLUME=ONCOLOGIST/SOLUTION DDFM=SAT;
RANDOM SUBJECT SUBJECT*ONCOLOGIST;
RUN;
```