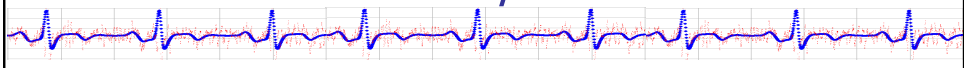
Empirical Research Methods in Information Science

IS 4800 / CS 6350

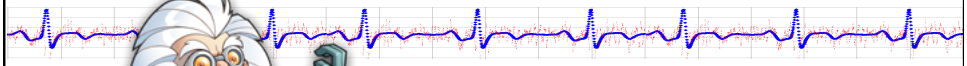


Lecture 21

Non-parametric Stats

1

Experiment



Data transforms?

2

When to Use Non-parametric Stats

- DV is nominal
- DV is ordinal
- DV is interval or ratio and
  - Skewed
  - Outliers
  - Multi-modal
  - Generally, non-normal
- aka "distribution-free tests"
- Note: Non-Parametric Stats still have assumptions, but much less restrictive

3

Non-parametric statistics

- Make few assumption about underlying population distribution.
- If your data does meet assumptions for parametric statistics (t-test etc), then non-parametric stats may be used.
- Provide less power than parametric equivalents
  - IF data is normal
- Generally operate on rank order-transformed or ranked data.
- A hodge-podge of techniques with no underlying theoretic framework.

4

R: tests for normality

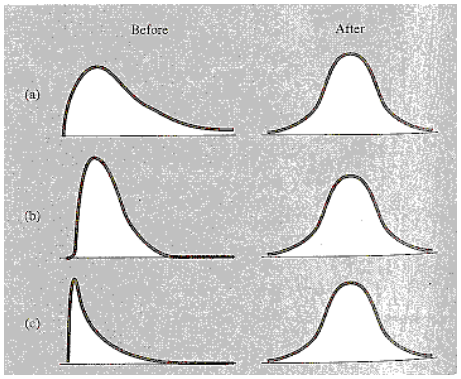
```
#Plot Q-Q line
>qqline(DV)

#Shapiro-Wilk Normality test
#Probability my data came from a
# normal distribution
>shapiro.test(DV)
```

5

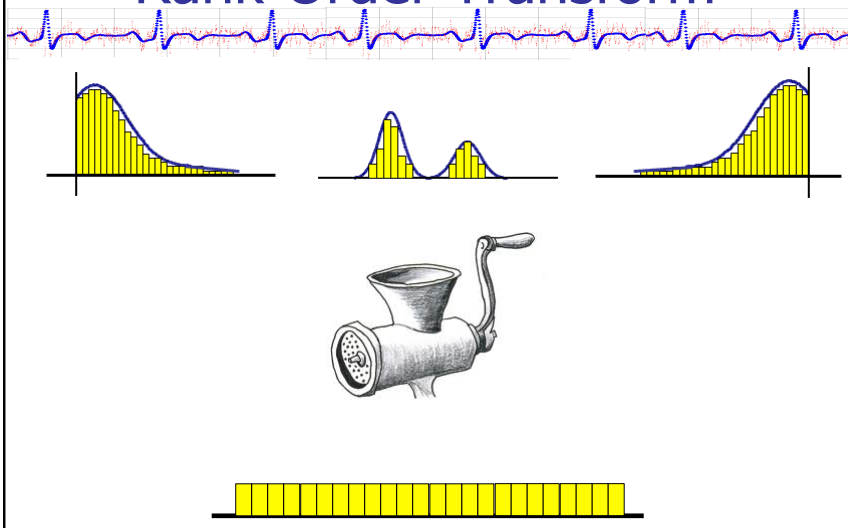
Other alternatives?

- Interval or Ratio DV, but skewed…



6

Rank-Order Transform



7

Rank-Order Transform

- Another kind of data transformation.
  - For non-uni-modal distributions or distributions with extreme outliers, or for ordinal measures
- Transform
  - Take all data from a given measure (from all groups), sort and assign ranking (ignore ties).
  - Then, replace the original values with the corresponding rank.

8

Rank-Order Transform Example

- Note For Ties: take average of ranks that would have been assigned
- DV Measure: Satisfaction
  - With sign in food court: 5.8, 7.1, 6.2, 6.7
  - Without sign: 2.3, 3.4, 1.7, 5.5
- Rankings:

1.7	2.3	3.4	5.5	5.8	6.2	6.7	7.1
1	2	3	4	5	6	7	8

- Resulting data for comparison:
  - With sign in food court: 5, 8, 6, 7
  - Without sign: 2, 3, 1, 4

9

Non-parametric Tests

Parametric Test	**Non-parametric Test**
Pearson correlation	Chi-square test for independence Spearman's rho *Just compute Pearson on the ranks*
t test for independent means	Wilcoxon rank-sum test aka Mann-Whitney U test
t test for dependent means	Wilcoxon signed rank test
One-way ANOVA	Kruskal-Wallis H-test
Full-factorial ANOVA	Aligned Rank Transform**

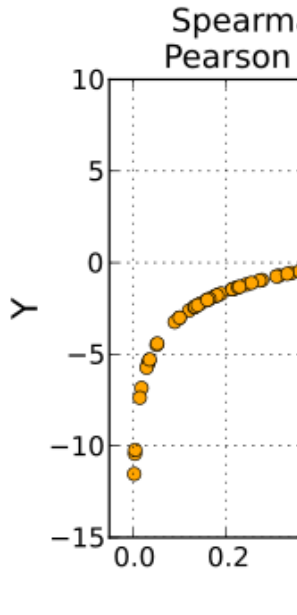
Note: Chi^sq on nominals, others on ordinal or better

X^2 Test for Independence in R

```
#use contingency table
chisq.test(table(x,y))

#e.g., data$favcolor, data$ownmac
chisq.test(table(data$favcolor,
                 data$ownmac))
```

11



Spearman measures degree of monotonicity.

12

Spearman's rho in R

```
> cor.test(x,y,method="spearman")
```

- Hyp test: likelihood we could have obtained a correlation this large, assuming H0 (true correlation is 0).

13

Wilcoxon Rank Sum

- Two groups of independent measurements
  - Cf: t-test for independent means
- For each group
  - Compute W =
    - Sum of ranks for group measures
    - Minus "mean rank" for group = $\sum_{i=1}^{N} i$ (corrects for size of group)
  - Pick smaller W and test

17

Wilcoxon Rank Sum (independent samples)

```
> wilcox.test(nonsmokers,smokers)
Wilcoxon rank sum test with
continuity correction
data: nonsmokers and smokers
W = 76.5, p-value = 0.04715
alternative hypothesis: true
location shift is not equal to 0
```

Report: W=76.5, p<.05

Wilcoxon in R more generally…

```
wilcox.test(dv-group1,
            dv-group2)
              -or-
#or
wilcox.test(dv ~ iv-factor,
paired=FALSE/TRUE, data=d)
```

19

Effect size – for both indep & paired tests

from "Discovering Statistics Using R"

$$r = \frac{z}{\sqrt{N}}$$

```
wilcoxModel <- wilcox.test(…)
z <- qnorm(wilcoxModel$p.value/2)
r <- z/sqrt(N)

Cohen: 0.1=Small, 0.3=Medium, 0.5=Large
```

20

Sample report

from "Discovering Statistics Using R"

- Depression levels in ecstasy users (Mdn = 17.50) did not differ significantly from alcohol users (Mdn = 16.00) the day after the drugs were taken, W = 35.5, p = 0.286, r = –. 25. However, by Wednesday, ecstasy users (Mdn = 33.50) were significantly more depressed than alcohol users (Mdn = 7.50), W = 4, p < .001, r = –. 78.

21

Wilcoxon signed rank

- Paired/dependent samples
  - Cf. paired sample t-test
- Compute difference for each pair of measures, discarding zeros
- Rank order transform the differences
- T_+=Sum of ranks from positive diffs
- T_-=Sum of ranks from negative diffs
- T = smaller of T_+ and T_-
  - Note: R calls this "V"

22

Wilcoxon signed rank in R

```
wilcox.test(dv-group1,
     dv-group2, paired=TRUE)
```

23

Wilcoxon Signed Rank Sample report

from "Discovering Statistics Using R"

- For ecstasy users, depression levels were significantly higher on Wednesday (Mdn = 33.50) than on Sunday (Mdn = 17.50), p = .047, r = –. 56. However, for alcohol users the opposite was true: depression levels were significantly lower on Wednesday (Mdn = 7.50) than on Sunday (Mdn = 16.0), p = .012, r = –. 45.

24

Differences between several independent groups

- Kruskall-Wallis (cf. 1-way ANOVA)
  - Based on ranked data
- H0: all samples came from same underlying population
- H1: there is at least one difference
- "H" statistic based on ranks within each group

25

Kruskal-Wallis in R

```
> kruskal.test(dv ~ ivfactor,
    data=d)

#posthocs – Wilcoxon Rank sum on
# all pairs of groups
> library(pgirmess)
> kruskalmc(dv ~ iv, data=d)
```


Sample output from kruskalmc

Significant Difference?

```
Multiple comparison test after Kruskal-Wallis, trea   t vs control
(two-tailed)
p.value: 0.05
Comparisons
                              obs.dif critical.dif difference
No Soya Meals-1 Soya Meal         2.2     15.63787      FALSE
No Soya Meals-4 Soya Meals        2.2     15.63787      FALSE
No Soya Meals-7 Soya Meals       19.0     15.63787       TRUE
```

Non-parametric equivalent of multi-factor ANOVA

- J. Wobbrock, et al. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11).

28

Homework

- Start T3
  - Experiment
    - Between or Within
    - At least 8 subjects.
  - Proposals due Friday.
    - 2 pages.
    - Include a power analysis. Should be prospective, based on a prior study.
- Read
  - Advanced study designs (B&A Ch 11).

29