STAT 5000

Statistical Methods I

Week 4

Fall 2024

Dr. Danica Ommen

Unit 1

Model-Based Inference:

Diagnostics

Scenario

- Randomized Experiment
  - Two treatments
  - Is there a difference in the mean value of the response variable between the two treatments?
- Observational Study
  - Two populations
  - One sample from each population
  - Is there a difference in the mean value of the variable between the two populations?

Inference for $\mu_1 - \mu_2$

- Researchers may be interested in estimating the mean difference between the two treatments/populations under study.
- Assumptions:
  - $Y_{11}, Y_{12}, \ldots, Y_{1n_1}$ are i.i.d. $N(\mu_1, \sigma^2)$
  - $Y_{21}, Y_{22}, \ldots, Y_{2n_2}$ are i.i.d. $N(\mu_2, \sigma^2)$
  - So, population variances are equal
  - Y_{1i} and Y_{2j} are independent for all i and j
- Implication:

$$\Rightarrow T = \frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

All of these assumptions should be assessed ...

Consequences of Violating Assumptions:

- If model assumptions are violated,
  - T will not have a $t_{n_1+n_2-2}$ distribution
  - Means that
    - p-value for hypothesis test will be wrong
    - Confidence level for interval will be wrong

Robustness of Model-based Inference:

- How far off will true p-value and true confidence level be if model assumptions are violated?
  - Not far off - we can still use two-sample inference procedure.
  - Far off - we cannot use two-sample inference procedure.
- Research studies have established when violations of model assumptions will result in large differences between true p-values and confidence intervals compared to values obtained from two-sample inference procedure.

Assessing Independence

- Study should be designed to achieve independent responses
- Independence may not hold if some sample members are related
  - students in same class
  - genetic relationships
  - soil samples taken close together
- Check by plotting observations versus relevant variables like time or location
- Look for possible clusters in which sample members may not respond independently

Assessing Independence

- Two-Sample Inference Procedure is not robust to violating this assumption. Effects of correlated responses (random errors) include
  - Standard error formulas are incorrect $Var(\overline{Y}_1 - \overline{Y}_2) \neq \sigma^2(1/n_1 + 1/n_2)$
  - t procedures are in serious trouble
  - Confidence intervals will not have correct coverage probabilities
- Remedies - Use another statistical procedure
  - If clustering - reanalyze using appropriate methods
  - If time effects - use time-series models
  - If location effects - use spatial models

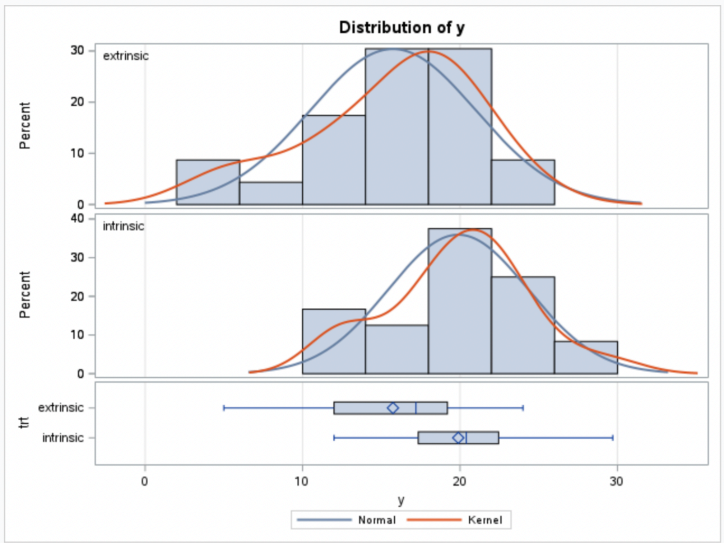
Assessing Equal Variances

- Graphical Methods
- Numerical Summaries
- F-test for equality of variances
- Brown-Forsythe test

Assessing Equal Variances: Graphical Method

- Construct residual plots, histograms, or boxplots of values for each group/population
- Look for
  - Outliers in each sample
  - Differences in IQR, Range
  - Differences in shape of sample distributions

Assessing Equal Variances: Graphical Method



Assessing Equal Variances: Summary Statistics

- Check the ratio of sample standard deviations

$$\frac{max\{S_1,\ S_2\}}{min\{S_1,\ S_2\}}$$

  - Between 1 and 2 - little impact
  - Between 2 and 3 - potential impact
  - Greater than 3 - likely impact

F-distribution

- Let U be a chi-square random variable with ν_1 df
- Let V be a chi-square random variable with ν_2 df
- Suppose U and V are independent, then

$$F = \frac{U/\nu_1}{V/\nu_2}$$

is an F-distributed random variable with ν_1 and ν_1 df.

F-distribution

- Let $Y_{1i} \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ for $i = 1, 2, \ldots, n_1$
- Let $Y_{2j} \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$ for $j = 1, 2, \ldots, n_2$
- Suppose Y_{1i} and Y_{2j} are independent for all i and j
- Then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

Assessing Equal Variances: F-test

- Reject $H_o : \sigma_1^2 = \sigma_2^2$ if

$$F_{max} = \frac{max\{S_1^2,\ S_2^2\}}{min\{S_1^2,\ S_2^2\}} \geq F_{(a,b),1-\alpha/2}$$

  where
  $a = n_1 - 1, b = n_2 - 1$ if $S_1^2 > S_2^2$
  $a = n_2 - 1, b = n_1 - 1$ if $S_2^2 > S_1^2$

- Very sensitive to normal distribution assumption
- Not recommended as the only check

Assessing Equal Variances: F-test

Output from the `ttest` procedure in SAS:

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	22	23	1.40	0.4304

Assessing Equal Variances: Brown-Forsythe Test

- Conduct a two sample t-test on the absolute deviations from the sample medians to assess homogeneous variability
- Available with the `hovtest=` option in the SAS `glm` procedure
- *The Statistical Sleuth*, Section 4.5.3 refers it to Levene's test
- Levene (1960) used absolute deviations from sample means

Assessing Equal Variances: Brown-Forsythe Test

- Compute $Z_{1j} = |Y_{1j} - \text{median}_1|$ for $j = 1, \ldots, n_1$ and $Z_{2j} = |Y_{2j} - \text{median}_2|$ for $j = 1, \ldots, n_2$
- Compute

$$\bar{Z}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} Z_{1j} \qquad \text{and} \qquad S^2_{Z_1} = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(Z_{1j} - \bar{Z}_1)^2$$

and

$$\bar{Z}_2 = \frac{1}{n_1}\sum_{j=1}^{n_2} Z_{2j} \qquad \text{and} \qquad S^2_{Z_2} = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(Z_{2j} - \bar{Z}_2)^2$$

Assessing Equal Variances: Brown-Forsythe Test

- Compute $S_{Z,p}^2 = \dfrac{(n_1 - 1)S_{Z_1}^2 + (n_2 - 1)S_{Z_2}^2}{n_1 + n_2 - 2}$
- Reject $H_0 : \sigma_1^2 = \sigma_2^2$ if

$$\left| \frac{\bar{Z}_1 - \bar{Z}_2}{S_{Z,p}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| > t_{(n_1+n_2-2),1-\alpha/2}$$

- Same as

$$\left| \frac{\bar{Z}_1 - \bar{Z}_2}{S_{Z,p}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|^2 > F_{(1,n_1+n_2-2),1-\alpha}$$

Assessing Equal Variances: Brown-Forsythe Test

```
/* Use the GLM procedure to perform the Brown-Forsythe
   test for homogeneous variances */

     proc glm data=set1 alpha=.05 ;
      class trt;
      model y = trt ;
      means trt / hovtest=bf;
      format trt trt.;
      run;
```

Brown and Forsythe's Test for Homogeneity of y Variance
ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
trt	1	3.8953	3.8953	0.36	0.5543
Error	45	493.7	10.9720		

Consequences of Unequal Variances

- Need to estimate quantity $Var(\overline{Y}_1 - \overline{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
- Unbiased estimator is $\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$
- Two-Sample Inference uses estimate $S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$
- If $n_1 = n_2 = n$, two estimates are equal (df are different).

$$\begin{aligned} S_p^2 &= \frac{(n-1)S_1^2 + (n-1)S_2^2}{2n-2} = \frac{S_1^2 + S_2^2}{2} \\ S_p^2 \left(\frac{1}{n} + \frac{1}{n} \right) &= \left(\frac{S_1^2 + S_2^2}{2} \right) \left(\frac{2}{n} \right) = \frac{S_1^2}{n} + \frac{S_2^2}{n} \end{aligned}$$

Consequences of Unequal Variances

- Minor if sample sizes are equal, especially when df is large.
- Minor if the ratio of variances is within a factor of 4 (or 10).
- The worst case is when $n1 \ll n2$ and $\sigma_1^2 > \sigma_2^2$, i.e., the group with the smaller sample size has the larger variance.
  - Empirical type I error rate is not controlled at the nominal rate.
  - While controlling $\alpha = 5\%$, the test may actually achieve 10% or 20% type I error rate.

Remedy for Unequal Variances

- Approximate Inference
  - Approximate t-test
- Variance Stabilizing Transformations
  - Logarithmic
  - Power

Remedy for Unequal Variances: Satterthwaite Approximation

Use separate sample variances for the two samples. Then

$$T^* = \frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

has an approximate t-distribution with

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$$

degrees of freedom. This is the Cochran-Satterthwaite approximation, and $\min(n_1 - 1, n_2 - 1) \leq \nu \leq n_1 + n_2 - 2$

Remedy for Unequal Variances: Satterthwaite Approximation

trt	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
extrinsic		15.7391	13.4677	18.0105	5.2526	4.0623	7.4343
intrinsic		19.8875	18.0119	21.7631	4.4418	3.4522	6.2308
Diff (1-2)	Pooled	-4.1484	-7.0018	-1.2950	4.8551	4.0270	6.1152
Diff (1-2)	Satterthwaite	-4.1484	-7.0156	-1.2812			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	45	-2.93	0.0053
Satterthwaite	Unequal	43.117	-2.92	0.0056

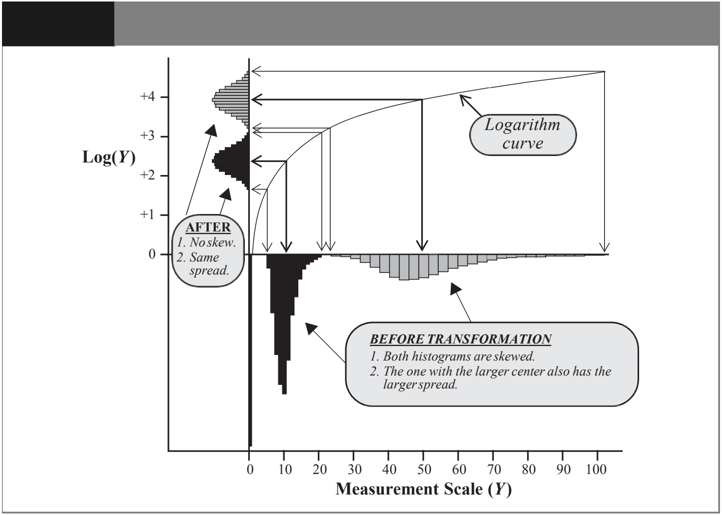
Remedy for Unequal Variances: Welch Approximation

- Welch test (*The Statistical Sleuth* Section 4.3.2)
- Very similar results to two-sample inference when samples sizes are nearly equal
- Better performance with unequal sample sizes AND unequal variances

Remedy for Unequal Variances: Transformation

- Replace Y_{ij} with $X_{ij} = h(Y_{ij})$
- Perform inference on X_{ij}'s $\Rightarrow$ e.g., compare $\bar{X}_1$ with $\bar{X}_2$
- Back transform estimates to get conclusions on the Y scale
  - only approximate conclusions about population means on the Y scale
- How can $X_{ij} = h(Y_{ij})$ affect heterogeneity?
  - Consider $h(Y) = log(Y)$.

Remedy for Unequal Variances: Log Transformation



Remedy for Unequal Variances: Log Transformation

- The logarithm function has a steep slope for small Y values, almost flat for large Y values.
- Small values are ‘stretched’ $\Rightarrow$ larger variance
- Large values are ‘shrunk’ $\Rightarrow$ smaller variance

Remedy for Unequal Variances: Transformation

- Choosing the transformation
  - Trial and error: transform and check histogram
  - Rules of thumb:
    - Data are all positive - use $\log(Y)$
    - Data are proportions - use $arcsin(\sqrt{Y})$
    - Data are counts - use $\sqrt{Y}$
  - Use transformation based on science
    (square root of area, cube root of volume)
  - Adjust for a variance-mean relationship
    (common for variance to increase with the mean)

Remedy for Unequal Variances: Transformation

- If $Var(Y) = g(E(Y))$, then a variance stabilizing transformation can be obtained from

$$h(y) \propto \int \frac{1}{\sqrt{g(x)}}\, dx$$

- Some Examples/Rules of Thumb:
  - If var $\propto$ mean, then $g(x) = x$ and $h(y) = \sqrt{y}$
  - If var $\propto$ mean², then $g(x) = x^2$ and $h(y) = log(y)$
  - If var $\propto$ mean(1-mean), then $g(x) = x(1-x)$ and $h(y) = \sin^{-1}(\sqrt{y})$

Remedy for Unequal Variances: Power Transformation

- Power transformation

$$X = \begin{cases} Y^{\lambda} & (\lambda \neq 0) \\ \ln Y & (\lambda = 0) \end{cases}$$

- estimate λ using maximum likelihood or use the variance-mean relationship
- Using variance-mean relationship
  - works for 2 groups or many groups (ANOVA)
  - Compute $\overline{Y}$ and S_Y for each group
  - Regress $\log(S_Y)$ on $\log \overline{Y}$ and estimate the slope (β)
  - Use transformation Y^{λ} with $\lambda = 1 - \beta$

Remedy for Unequal Variances: Power Transformation

- When does this work?
  - Population standard deviation is proportional to a power of the population mean:

$$\sigma = \sqrt{Var(Y)} = \kappa\mu^{\beta}$$

or

$$Var(Y) = \sigma^2 = \left[\kappa\mu^{\beta}\right]^2 = f(\mu)$$

  - Use the delta method to obtain the transformation:

$$X = g(Y) = Y^{\lambda}$$

Power Transformation Derivation

- Consider the Taylor series expansion

$$X = g(Y) \approx g(\mu) + (Y - \mu)g'(\mu)$$

- Then an approximation for $Var(g(Y))$ is

$$Var(g(Y)) \approx [g'(\mu)]^2 \, Var(Y)$$

- This is called the *delta method*

Power Transformation Derivation

- For $X = g(Y) = Y^{\lambda}$ we have

$$\frac{dX}{dY} = g'(Y) = \lambda Y^{\lambda-1}$$

- From the delta method

$$\begin{aligned} Var(X) &\approx \left(\lambda\mu^{(\lambda-1)}\right)^2 \times \left(\kappa\mu^{\beta}\right)^2 \\ &= \kappa^2\,\lambda^2\,\mu^{2(\lambda-1+\beta)} \end{aligned}$$

- when $\lambda = 1 - \beta$ then $\lambda - 1 + \beta = 0$ and $Var(X) \approx k^2\,\lambda^2$ is approximately constant
- Analyze the transformed data: e.g., $X_{11} = log(Y_{11}), X_{12} = log(Y_{12}), ..., X_{2,n_2} = log(Y_{2,n_2})$

Power Transformation Derivation

- What if β = 0.12?
  Usually round to a "reasonable" value,
  i.e., use $\beta \approx 0$ and $\lambda = 1$.
- Caution: Some researchers estimate the slope from the regression of log($Var(Y)$) on log($\overline{Y}$). Then use the transformation $Z = Y^{\lambda}$ with $\lambda = 1 - \beta/2$.
  - Issues / concerns: Interpretation of results are more difficult (e.g., the expectation of log(Y) is not the logarithm of $E(Y)$)

Assessing Normality

- Graphical Methods
- Numerical Summaries
- Tests for Normality
  - Presence of Outliers

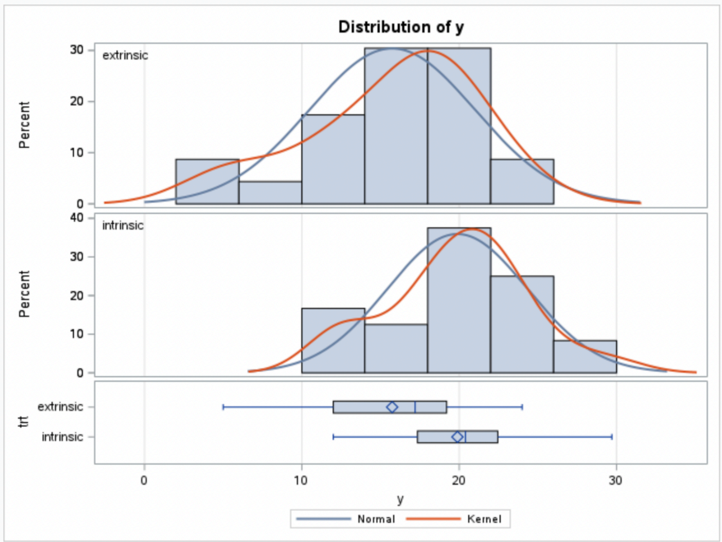
Assessing Normality: **Graphical Methods**

- Histogram of values within each group/population
  - Look for symmetric, bell shape
- Normal probability plot within each group/population
  - Compare empirical cumulative distribution function (CDF) to CDF for theoretical normal distribution
  - Most commonly done using quantiles (Q-Q plot): plot empirical quantiles against expected quantiles from normal distribution

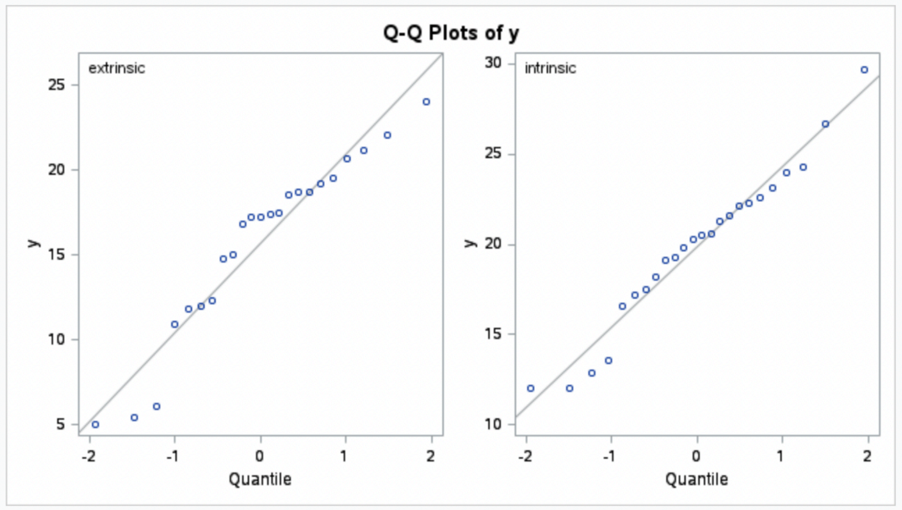
Assessing Normality: Normal Q-Q Plot

- Order *residuals* from smallest to largest (say $X_{(1)}, \ldots, X_{(n)}$)
- Compute expected quantiles $(q_{(1)}, \ldots, q_{(n)})$ from a standard normal distribution.
  - Expected quantiles can be calculated with tables.
  - General approximation: $q_i = \Phi^{-1}\left(\frac{i}{n+1}\right)$
  - Blom approximation $q_i = \Phi^{-1}\left(\frac{i-.375}{n+.25}\right)$
  - For $i = 5$, $n = 9$, $q_5 = \Phi^{-1}\left(\frac{5}{10}\right) = 0$
- Scatterplot of $X_{(i)}$ vs q_i should be close to a straight line with slope σ
- Curved patterns indicate non-normal distributions (or presence of outliers)

Assessing Normality: Graphical Methods



Assessing Normality: Graphical Methods



Assessing Normality: Numerical Summaries

- For any normal distribution
  - Mean and median should be equal
  - Skewness = $E(Y-\mu)^3/\sigma^3 = 0$
    (Skewness measures the asymmetry)
  - kurtosis = $E(Y-\mu)^4/\sigma^4 = 3$
  - excess kurtosis = kurtosis - 3
    (estimated by the `univariate` procedure in SAS)
  - The sample kurtosis measures the heaviness of the tails of the data distribution.
  - positive value: long-tail; negative value: short-tail

Assessing Normality: Numerical Summaries

The UNIVARIATE Procedure
Variable: y

trt=intrinsic

Moments			
N	24	Sum Weights	24
Mean	19.8875	Sum Observations	477.3
Std Deviation	4.44182811	Variance	19.729837
Skewness	-0.0769953	Kurtosis	0.07803298
Uncorrected SS	9946.09	Corrected SS	453.78625
Coeff Variation	22.3347737	Std Error Mean	0.90668437

Basic Statistical Measures			
Location		Variability	
Mean	19.88750	Std Deviation	4.44183
Median	20.40000	Variance	19.72984
Mode	12.00000	Range	17.70000
		Interquartile Range	5.10000

The UNIVARIATE Procedure
Variable: y

trt=extrinsic

Moments			
N	23	Sum Weights	23
Mean	15.7391304	Sum Observations	362
Std Deviation	5.25259582	Variance	27.5897628
Skewness	-0.76156	Kurtosis	-0.0935406
Uncorrected SS	6304.54	Corrected SS	606.974783
Coeff Variation	33.3728464	Std Error Mean	1.09524194

Basic Statistical Measures			
Location		Variability	
Mean	15.73913	Std Deviation	5.25260
Median	17.20000	Variance	27.58976
Mode	17.20000	Range	19.00000
		Interquartile Range	7.20000

Note: The mode displayed is the smallest of 2 modes with a count of 2.

Assessing Normality: Tests

- Many proposed tests for normality
- Tests based on empirical cdf's: Kolmogorov-Smirnov, Anderson-Darling, etc.
- Tests based on skewness or kurtosis
- Chi-square goodness-of-fit tests
- Tests based on normal probability plots: Shapiro-Wilk, correlation tests
- Normality is almost always rejected for large sample sizes.

Consecutive of Non-Normality

- Large samples $\rightarrow$ few consequences (Central Limit Theorem)
- Small samples
  - Sample distributions have same shape and
    - equal sample sizes $\rightarrow$ very little impact
    - different sample sizes $\rightarrow$ potential impact if distributions are skewed
  - Sample distributions have different shapes $\rightarrow$ impact

Remedy for Non-Normality

- Transformation (especially for skewness)
- Discussed earlier (under remedies for unequal variances)
- Detect and eliminate outliers
- Non-parametric tests

Testing Normality: Outliers

- Outlier: one (or a few) very unusual observation(s)
- Always an issue if outliers are from a non-target population
- Goal: make inferences for the target population

  Data: from a mix of the target population and an outlier population

- Detect and eliminate outliers
- Reduce the effects of outliers by using "robust" procedures

Testing Normality: Outliers

- Analyze data with and without suspected outliers to see if inferences change
- Remove data only if one can argue that observations are from a different population. Remove any other observations from that different population.
- Acknowledge deletion of outliers in final report

Unit 1

Non-Parametric Inference

Wilcoxon Rank-Sum Test

- Independence
- Null hypothesis: two populations have the same distribution
  - Distribution is not required to be normal.
  - Implies equal medians, percentiles, means and variances
- Can test against one- or two-sided alternative
- Can compute “exact" p-values based on the null distribution of the ranks

Wilcoxon Rank-Sum Test

- Order the combined $n_1 + n_2$ observations (small to large)
- Assign ranks
  - Smallest gets rank=1, second smallest gets rank=2, etc...
  - For tied observations, average the ranks
- Compute the sum of the ranks for one group (call it W)
- Assuming H_0 is true, compute:

$$E_0(W) = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \text{and} \quad V_0(W) = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

- Large sample Z-test: $z = \dfrac{|W - E_0(W)| - 0.5}{\sqrt{V_0(W)}}$
- Approximate p-value $= 2 * P(Z > |z|)$

Wilcoxon Rank-Sum Test: Example

- Randomized experiment with two treatments
  - Animals A, B, and C receive a standard drug: observations are 3,8,4
  - Animals D, E, and F receive the new drug: observations are 7,9,11
- Null Hypothesis: Distribution of response variable is the same for both groups.
- Alternative Hypothesis: Distribution of response variable is different between the two groups.

Wilcoxon Rank-Sum Test: Example

Order the combined sample:

Observation (Y)	Animal ID	Treatment Group	Rank
3	A	standard	1
4	C	standard	2
7	D	new drug	$\boxed{3}$
8	B	standard	4
9	E	new drug	$\boxed{5}$
11	F	new drug	$\boxed{6}$

- Test statistic: $W = 3 + 5 + 6 = 14$

Wilcoxon Rank-Sum Test: Example

- $E_0(W) = (3)(7)/2 = 10.5$ and $V_0(W) = (3)(3)(7)/12 = 5.25$
- z = $\dfrac{|14 - 10.5| - 0.5}{\sqrt{5.25}}$ = 1.31
- Approximate p-value is $2 * P(Z > 1.31) = 0.19$
- Fail to reject the null hypothesis
- We do not have significant evidence that the response distributions for the new drug and the standard drug are different.

Wilcoxon Rank-Sum Test: Example

Computing the Exact p-value

- There are $\binom{6}{3} = 20$ ways to assign to the treatment groups
- The observed ranks for the treatment group are 3, 5, 6
- There is only one other random assignment that could produce a larger value of W: subjects with ranks 4, 5, 6 are assigned to the treatment group
- Two other random assignments are equally extreme in the other direction
  - ranks 1, 2, 4 (observed)
  - ranks 1, 2, 3 (more extreme than observed)
- The exact two-sided p-value is

$$\begin{aligned} p &= \frac{\text{\# of sets as extreme or more extreme than observed}}{\text{\# of possible sets of ranks}} \\ &= 4/20 = 0.20 \end{aligned}$$

Wilcoxon Rank-Sum Test: SAS Example

```
/* Evaluate the Wilcoxon test.  The 'exact'
   option requests the exact randomization
   p-value for the test.  After the slash you
   can include the 'MC' option to get a Monte
   Carlo approximation to the exact p-value. */

proc npar1way data=set1 wilcoxon;
    class trt;
    var y;
    exact wilcoxon / MC N=20000 Seed=7892441
        alpha=.05 maxtime=20;
    format trt $trt.;
run;
```

Wilcoxon Rank-Sum Test: SAS Example

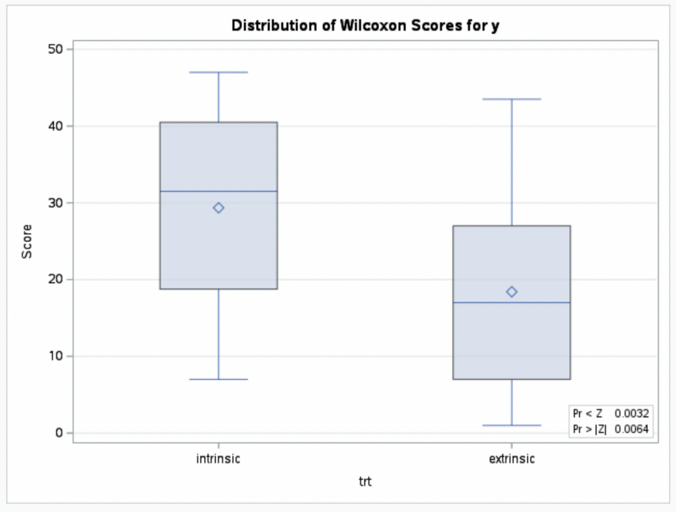
The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable y Classified by Variable trt					
trt	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
intrinsic	24	704.50	576.0	46.973057	29.354167
extrinsic	23	423.50	552.0	46.973057	18.413043
Average scores were used for ties.					

Wilcoxon Two-Sample Test					
				t Approximation	
Statistic (S)	Z	Pr < Z	Pr > |Z|	Pr < Z	Pr > |Z|
423.5000	-2.7250	0.0032	0.0064	0.0045	0.0091
Z includes a continuity correction of 0.5.					

Monte Carlo Estimates for the Exact Test					
Probability	Estimate	95% Confidence Limits		Samples	Seed
Pr <= S	0.0029	0.0021	0.0036	20000	7892441
Pr >= |S - Mean|	0.0051	0.0041	0.0060		

Wilcoxon Rank-Sum Test: SAS Example



Questions?

Contact me:

Email: dmommen@iastate.edu
Visit Student Office Hours: Thur @ 10-11 am