**BASIC STATISTICS**
Brian Gleeson

**INTRODUCTION TO POPULATION STATISTICS**

**POPULATIONS STATISTICS**
- Why?
  - Because you can’t test everything/everyone

- What do population statistics tell you?
  - How confident can you be about the results of your sample

**VARIANCE: THE CORE OF ALL POPULATION STATISTICS**
- (Draw curves on board)
**DISCLAIMER**

- These methods are not the best methods
- They are the most common methods
  - What you’ll read in papers
  - What reviewers expect to see
  - What you’ll be able to do with the common tools

**NOTATION**

- Statistics: measures that arise from your sample
  - $y$ – some score in your sample
  - $s^2$ – variance of sample
  - $s$ – standard deviation of sample
- Parameters: measures that arise from the population
  - $\sigma^2$ – variance of population
  - $\sigma$ – standard deviation of population
  - est. $\sigma$ – estimate of $\sigma$
  - est. $\sigma^2$ – estimate of $\sigma^2$
- Other:
  - $SS$ = Sums of Squares
  - $\Sigma (y_i \text{ }- \text{ } \text{mean}(y))^2$

**MEASURES OF VARIANCE**

$$\frac{\Sigma(y - \bar{y})^2}{N}$$

$$\frac{\Sigma(y - \bar{y})^2}{N - 1}$$
CORRELATIONS

- Pearson's r
  - Linear correlation
- Spearman's r
  - Rank correlation

CONFIDENCE INTERVAL

- Can use instead of standard error
  - More informative
  - Really tells you how well you know the true mean
- Confidence Interval
  - "95% probability that the interval contains the true value of the population mean."
  - Not: "95% probability that the true value of the population mean is within the interval."
  - Do not use Microsoft Excel confidence interval.

HYPOTHESIS TESTING

- H₀ = Null Hypothesis
  - Independent variable has no effect
- Hₐ = Alternative Hypothesis
  - Independent variable has some effect
Hypothesis Testing

H_0 True

Reject H_0

Type 1 Error
p = α

OK
p = ‘power’

Do not Reject H_0

H_0 False

Type 2 Error
p = β

Testing a Single Group Mean (t-Test)

(Board)

Signal = Distance of Sample Mean from 0
Noise = Variance of Sample Distribution of the Mean.

1-Tailed vs. 2-Tailed Test

Only use 1-tailed if you have an a priori reason to do so

- e.g. Two point discrimination between finger and back

- e.g. Two point discrimination between index and middle finger

Compare 2 Independent Means

Requirements

- Samples are independent
- Sample sizes are equal
- Normal distributions
- Equal variance (depends on method)

Test: Student’s t-test for Independent Means

H_0: μ_1 = μ_2
H_A: μ_1 ≠ μ_2

(Board)

Typical Use

- Between-subjects design
COMPARE 2 DEPENDENT MEANS

- Requirements
  - Samples are dependent
  - ...same as previous
- Test:
  - Calculate difference between means
    - \( \mu_{\text{difference}} = \mu_1 - \mu_2 \)
  - (Board)
  - Student’s t-test for Single Mean
    - \( H_0: \mu_{\text{difference}} = 0 \)
    - \( H_A: \mu_{\text{difference}} \neq 0 \)
- Typical Use
  - Within-subjects design

SIGNIFICANCE LEVEL IN T-TESTS

An \( \alpha \)% significance means that there is a risk that \( \alpha \)% of the Group 1 distribution may belong to the Group 2 distribution

T-TEST: INTERPRETING RESULTS

<table>
<thead>
<tr>
<th>t-Test: Two-Sample Assuming Equal Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>Pooled Variance</td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
</tr>
<tr>
<td>df</td>
</tr>
<tr>
<td>t Stat</td>
</tr>
<tr>
<td>t Critical one-tail</td>
</tr>
<tr>
<td>t Critical two-tail</td>
</tr>
</tbody>
</table>

If \( p < 0.05 \) (\( \alpha \)), reject the null hypothesis

How to report these results:

- \( t(df) = t_{\text{obtained}} \), \( p = \ldots \)
- e.g. \( t(18) = -3.37 \), \( p = 0.003 \)

"Reject the null hypothesis" or "Fail to reject the null hypothesis"
Never "Accept the null/alternative hypothesis"

EXAMPLE
ANOVA

- ANOVA = ANalysis Of VAriance, a.k.a. F-test
- Use to analyze several groups means for differences.
- When only 2 groups, same as t-Test
- Does not tell you which group mean is different.
  - $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots$
  - $H_1$: At least one of the group means is different from all the rest
  - Post-hoc analysis necessary to find which groups are different

ANOVA ASSUMPTIONS

- Normal distributions and homogeneity of variance. Therefore, in a one-factor ANOVA, it is assumed that each of the populations is normally distributed with the same variance.
- In between-subjects analyses, it is assumed that each score is sampled randomly and independently.
- Research has shown that ANOVA is "robust" to violations of its assumptions.
- ANOVA tends to be conservative when its assumptions are violated.

ANOVA: TERMINOLOGY

- Response variable (Dependent variable): The primary variable of interest measured in the experiment
- Factor (Independent variable, Predictor variable): Variable that has an effect on the measurement of the response variable
- Factor levels (Treatment level): The particular values that a factor can have
- Two types of factor:
  - Fixed-Effect: Factor which its levels included in the study are the only levels of interest or maybe the only possible levels. (ex. gender, marital status)
  - Random-Effect: Factor which its levels included in the study are not the only ones we are interested in making inferences about (ex. samples of merchandise, users)
CLASSIFICATIONS OF ANOVA

- One-way vs. Multi-way (One-factor vs. Multi-factor)
- Fixed-effect vs. Random-effect model (Mixed-effect)
- Independent vs. Repeated measures
- Balanced vs. Unbalanced design

ONE-WAY VS. MULTI-WAY ANOVA

Depends on the number of factors in the study

Examples:
- Evaluate the effect of force feedback modalities on task execution time (One-way)
- Evaluate the effect of the force feedback modalities and users with haptic devices on execution time (Two-way)
- Evaluate the effect of age, gender, and drug dosage on the treatment of a disease (Three-way)

Fixed- vs. Random-Effect Model

Depends on the type of the factors in the study, which can vary by design

Let’s look at the factors of our last examples…
- Force feedback modalities
- Users
- Age, gender, and drug dosage

Independent vs. Repeated Measure

Independent measures:
Different subjects provides the scores for different factor levels

Repeated measures:
Same subjects provides the scores for different factor levels. Effect of subjects is treated as a random factor.
Balanced vs. Unbalanced design

Balanced design

Factor 1  Level 1  Level 2  Level 2
Factor 1  Level 1  Y1,1,1, Y1,1,2, Y1,1,3
Factor 1  Level 2  Y1,2,1, Y1,2,2, Y1,2,3

Unbalanced design

Factor 1  Level 1  Level 2  Level 2
Factor 1  Level 1  Y1,1,1, Y1,1,2
Factor 1  Level 2  Y1,2,1, Y1,2,2

The data analysis of an unbalanced design is more difficult. A good rule of thumb is to keep your design balanced!

ANOVA: INTERPRETING RESULTS

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>F value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>62.30891</td>
<td>3</td>
<td>20.76964</td>
<td>20.77442</td>
<td>5.521E-08</td>
<td>2.866266</td>
</tr>
<tr>
<td>Within Groups</td>
<td>35.9917025</td>
<td>36</td>
<td>0.99977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>98.3006124</td>
<td>39</td>
<td>2.57657</td>
<td>2.580228</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How to report these results:

- If \( p < 0.001 \), just say "\( p < 0.001 \)", don't give the exact value.
- "Reject the null hypothesis" or "Fail to reject the null hypothesis"
- Never "Accept the null/alternative hypothesis"

POST-HOC ANALYSIS

- After ANOVA, which means are different?
- Why not just t-test all pairs?
  - Group-wise error rate > \( \alpha \)
  - Each has 5% of being wrong
  - Error accumulates with multiple, dependent tests

POST-HOC ANALYSIS: A PRIORI TESTS

- Uses these tests for independent comparisons predicted by you hypothesis
- Orthogonal t-tests
  - Learn how to set up orthogonal tests, or just never test the same mean twice
  - Perform t-tests as normal
- For non-orthogonal tests: Bonferroni's Correction
- For \( N \) non-orthogonal tests: \( \alpha \rightarrow \alpha/N \)
- Some non-orthogonal tests
  - Dunn's
  - Denet's test
**POST-HOC ANALYSIS: REAL POST-HOC TESTS**

- Use these to check for things not specifically predicted by your hypothesis.
- Tukey’s Honest Significant Difference
  - All possible pair-wise comparisons
  - Conservative
- Scheffe’s Test
  - All imaginable comparisons
  - Very conservative

**EXAMPLE: 1-WAY ANOVA**

**EXAMPLE: POST HOC**

**EXAMPLE: 2-WAY ANOVA WITH INTERACTION**
EXAMPLE: POST-HOC ANALYSIS

Click on the group you want to test

Compare everything:
No Significant Effect

Look at effect of factor X1 only:
Significant Effect

REMEmBER THE PRECONDITION AND
ASSUMPTIONS

• All samples are independent
  • Essential
• Data must be normally distributed
  • How to tell
  o Look at histogram of data
  o Shapiro-Wilk test for Non-Normality
  • If not normally distributed
    o Transform data to make normal
• Variance the same for all groups
  • How to tell
    o Test: F-ratio, Levene’s F
  • If variances not equal
    o t-Test: use t-test for unequal variance (Welch’s t)
    o ANOVA: nothing you can do

Note: t-test and F-test are robust to violations of normality and variance, especially if the sample size is large (>30) and approximately the same for all groups.

DESIGN YOUR EXPERIMENT WELL

• Equal sample sizes
• Large Samples
• If possible, use metrics that give normally distributed data
• Avoid confounding factors
  • Independent variable should be the only thing that changes between tests
    o Constant Task and task difficulty
    o Beware of sound, sight, etc
    o Order
      o Randomize
      o Full factorial
      o Balanced/Unbalanced Latin Squares

© B. Gleeson, 2009
STATISTICS SOFTWARE YOU CAN USE

- Excel
- Matlab w/ Stats Toolbox
- Mathematica
- SAS/SPS/SPSS
- R/R Commander (open source)

HOW TO USE STATS TOOLS IN EXCEL

1. Click the Microsoft Office Button, and then click Excel Options.
2. Click Add-ins, and then in the Manage box, select Excel Add-ins.
3. Click Go.
4. In the Add-Ins available box, select the Analysis ToolPak check box, and then click OK.
5. Tip If Analysis ToolPak is not listed in the Add-Ins available box, click Browse to locate it.
6. If you are prompted that the Analysis ToolPak is not currently installed on your computer, click Yes to install it.

DATA CONDITIONING

Before you do stats, do this

What's wrong with these data?

10 12
N
2 4 6 8
0 20 40 60 80 100 120 140 160 180 200
-2 0
**Outlier Rejection**

- Only if justifiable
- Reject data a priori
- Reject data > nσ from mean
- 3σ, 6σ common
- Based on pooled-mean or subject-mean: your call

**Data Conditioning**

- Use a histogram to look at data distribution
- Or use statistical tests of normality, e.g., Shapiro-Wilk test for Non-Normality
- So now what?
  - Deal with it
  - Transform Data

**Histogram of Data**

**Histogram of Transformed Data: 1/y**
MEASURES OF EFFECT SIZE

You get meaningless units when you:

- Transform data
- Use some established non-SI measure (e.g. JND)
- Use some arbitrary performance metric (e.g. accuracy/time)
- Or, you don’t have the original data (comparing results in the literature)

Use a normalized measure of effect size

- Cohen’s d
  - effect size in sample. Most common.
- Hedges’s g
  - estimate of effect size in population
- Glass’s Δ
  - compare to control group

EXAMPLE

- Task Completion Experiment
  - Visual Feedback
  - Haptic Feedback
  - Independent Samples (between-subjects)

How big is the effect?

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Cohen’s d</th>
<th>Hedges’s g</th>
<th>Glass’s Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>small</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>med</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>large</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>

EXAMPLE

Task Completion Experiment
- Visual Feedback
- Haptic Feedback
- Independent Samples (between-subjects)

1. LOOK AT DATA
   - Plot data and visually inspect

MEASURES OF EFFECT SIZE

\[ d = \frac{\bar{Y}_1 - \bar{Y}_2}{\sigma_{pooled}} \]

where \( \sigma_{pooled} = \sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2}} \)

\[ g = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{est.} \sigma_{pooled}} \]

where \( \text{est.} \sigma_{pooled} = \sqrt{\frac{(N_1 - 1) \text{est.} \sigma_1^2 + (N_2 - 1) \text{est.} \sigma_2^2}{N_1 + N_2 - 2}} \)

\[ \Delta = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{est.} \sigma_{Control}} \]

where \( \text{est.} \sigma_{Control} = \sqrt{\text{est.} \sigma_{Control}^2} \)
2. REJECT OUTLIERS
Throw out bad data, if there is a reason to do so.
Here, reject (y-mean) > 3σ

3. LOOK AT DATA AGAIN
See what valid data looks like

4. LOOK AT HISTOGRAM
Check distribution shape

5. TRANSFORM DATA
Transform data to achieve normality
6. Look at Normalized Data

- Look for obvious effects or problems

![Graph showing completion time vs sample](image)

7. Calculate Statistics

- Results:
  - Mean Visual = 3.189 1/s
  - Mean Haptic = 3.059 1/s
  - Effect Size = 0.130 1/s
    - Calculate normalized effect size
      - $d = 0.111$ (small effect)
  - Test
    - $t(1969) = -2.824$, $p = 0.0048$
    - Statistically significant
- So what?
  - Use your judgment
  - Consider application
  - Refer to literature

![Graph showing inverse completion time](image)

8. Display Data for Publication/Report

- Make it readable
- Consider choice of error bars

![Graph showing inverse completion time](image)