Role of statistics in research

- Statistical methods help us to collect, organize, summarize, analyze, interpret, & present data
- “The role of statistics is not to discover truth. The role of statistics is to resolve disagreements between people.” - Milton Friedman
- “...the purpose of statistics is to organize a useful argument from quantitative evidence, using a form of principled rhetoric.” - Robert P. Abelson

Ways of Collecting Data

- Designed Experiments -
  - effects of independent variables on dependent variables
  - random assignment of “subjects” to conditions
- Correlational Studies -
  - associations among predictor & criterion variables
  - “subjects” come with their own set of variables
- Both types can be combined into a single study/analysis: ANCOVA

Correlational Studies

- Correlational studies measure the association between variables
- Usually, the variables are not manipulated by investigator
  - each event/subject comes with own set of variables
  - but values on variables differ across events/subjects
- Regression measures association between predictor & criterion variables
  - e.g., measure the association between annual income (criterion variable) and parent’s income, years of education, race, gender (predictor variables)
Independent Variables

- Most Psychology studies collect data in designed experiments
- Experiments usually involve collection of data in various experimental conditions
  - conditions differ in terms of 1 or more independent variables
  - e.g., conditions in memory experiment defined by:
    ‣ type of items (faces vs words) being studied
    ‣ time interval between study & test phases

Dependent Variable

- The variable(s) that is(are) measured and constitute the data, or results, that will be analyzed.
- Designed experiments measure the effects of independent variables on dependent variables

Examples of dependent variables

Examples of dependent variables include reaction time, response accuracy, number of items recalled in memory test, event-related brain potentials, heart rate, number of offspring, number of aggressive or affiliative behaviours, etc.

Hypothetical Recognition Memory Experiment

2 independent variables

Retention Interval

- short (1 minute)
- long (1 hour)

Study Items

- faces
  - # of items correctly recognized during test
- words
  - # of items correctly recognized during test

1 dependent variable measured in 4 experimental conditions defined by combinations of independent variables
Random Assignment

- In Psychology, designed experiments use subjects that also come with their own set of intrinsic characteristics
- These characteristics (personality, motivation, intelligence, etc.) probably affect dependent variable
- HOWEVER, in most experiments, subjects are randomly assigned to experimental conditions
- So, effects of subject differences should be UNRELATED TO EFFECTS OF INDEPENDENT VARIABLES
  - big advantage of designed experiments over correlational studies

Exploratory vs Confirmatory Analyses

- Exploratory Data Analysis
  - first major proponent was John Tukey
  - goal: discover & summarize interesting aspects of data
  - discover interesting hypotheses to test
- Confirmatory Data Analysis
  - data are gathered & analyzed to evaluate specific a priori hypotheses
  - example: clinical drug trials
- Important not to confuse two types of analyses
  - replication crisis in Psychology related to confusion about two types of research

Modes of Statistical Analyses

- Exploratory vs. Confirmatory
- Descriptive vs. Inferential

Descriptive Statistics

- Descriptive statistics:
  - describes important characteristics of the sample
  - uses graphs & statistics e.g., mean or standard deviation
Describing Shapes of Distributions

- Unimodal vs Multi-modal
- Symmetric vs Skewed
- Light vs. Heavy tails (kurtosis)

Inferential Statistics

- Inferential statistics:
  - uses sample to make claims about a population
  - e.g., estimate population parameters from sample statistics
  - e.g., investigate differences among population by examining differences among samples [effect size & association strength]

Inference: Samples to Populations

- Population: all events (subjects, scores, etc)) of interest
- Sample: subset of population
  - random sample: each member of population has equal chance of being selected
  - convenience sample (e.g., psychology undergraduates)
- Inference depends on quality of relation between sample & population. Is sample representative of population?

Descriptive & Inferential Statistics

- Population parameters $\mu$, $\sigma^2$
- Sample statistics mean, variance $\bar{Y}$, $s^2$
- Statistical theory e.g., sampling distributions
- (Inferential Statistics)
- (Descriptive Statistics)
A woman in the US has just given birth to a full-term baby weighing 291 kg. Is this weight unusually low?

- In US, population of Caucasian (non-Hispanic) full-term infants has a mean weight of 3480 g and a standard deviation of 462 g.
- The weights are distributed approximately normally.
- A weight of 2910 g is 1.23 standard deviations below the mean:
  \[ z = \frac{(2910 - 3480)}{462} = -1.23 \]
- What is the probability of observing a weight that is at least this low?
z test

- In US, population of Caucasian (non-Hispanic) full-term infants has a mean weight of 3480 g and a standard deviation of 462 g.
- The weights are distributed approximately normally.
- A weight of 2910 g is 1.23 standard deviations below the mean:
  - $z = (2910 - 3480) / 462 = -1.23$
- What is the probability of observing a weight that is at least this low?
  - $p(z < -1.23) = 0.109$

z test for means

- Consider situation when we want to evaluate a group mean
  - e.g., measure birth weight of 100 Native-American full-term babies
  - mean = 3350 g; standard deviation = 425 g
- Is group mean of 3350 g unusually low?
  - $z = \frac{(\bar{Y} - \mu)\sqrt{n}}{\sigma}$
  - if sample was drawn from population of Caucasian birth weights, then group mean is unusually low
z test for means

- Sample mean = 93 (n=20)
- Population standard deviation = 10
- Is sample drawn from a population with a mean = 100?
- Calculate z and compare to “critical values” of ±1.96 or ±2.56

Assume Means are Normally Distributed
Parameters of Normal Sampling Distribution

\[ \mu_Y = 100 \]
\[ \sigma_Y = \frac{10}{\sqrt{20}} = 2.236 \]

Calculate z:

\[ z = \frac{\bar{Y} - \mu_Y}{\sigma_Y} = \frac{93 - 100}{2.236} = -3.13 \]

Our sample mean is 3.13 standard deviations below the expected mean.

Central Limit Theorem

\[ \mu_Y = \mu \]
\[ \sigma_Y = \sigma / \sqrt{n} \]

Figure 1: The theoretical sampling distribution of the mean.

z test for means

- z test assumes that group means are distributed normally
- If scores are distributed normally, then means are, too
- Suppose scores are NOT distributed normally?

CENTRAL LIMIT THEOREM:

- Irrespective of how the scores are distributed, the sample means will be distributed normally, provided that the sample size (n) is sufficiently large

Uniform Distribution of Scores

Sampling Distribution of Means (n=2)

Sampling Distribution of Means (n=8)

Sampling Distribution of Means (n=4)
t test for means
- sample: n = 100; mean = 3350 g; standard deviation = 425 g
- assume: sample drawn from population with μ for Caucasian births
  - distribution of SCORES: u = 3480; sd = σ = ?; distribution = ?
  - distribution of MEANS: u = 3480
    - sd = σ/sqrt(n); distribution = NORMAL (central limit theorem)
- what is probability of getting sample mean (n=100) < 3350 g?
  - will use sample SD to estimate population SD:
    \[ z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim \frac{\bar{Y} - \mu}{\hat{\sigma}/\sqrt{n}} \]

Although t looks similar to z, t is NOT distributed normally because the sampling distribution of the variance (σ² std dev) is skewed, especially for small sample sizes.

**t test**
- estimate of population standard deviation
  \[ \hat{\sigma} = s = \sqrt{\frac{\sum(Y_i - \bar{Y})^2}{n-1}} \]
- t statistic
  \[ t = \frac{\bar{Y} - \mu}{s/\sqrt{n}} \]

**t distribution(s)**
- t distribution is a family of distributions
- Low degrees-of-freedom: t has heavier tails than normal distribution
- t & normal distributions increasingly similar as df increases

**t test for means**
- sample: n = 100; mean = 3350 g; standard deviation = 425 g
- assume: sample drawn from population with μ for Caucasian births
  - distribution of SCORES: u = 3480; s = 425; distribution = ?
  - distribution of MEANS: u = 3480
    - SEM = s/sqrt(n) = 42.5; distribution = NORMAL (central limit theorem)
- what is probability of getting sample mean (n=100) < 3350 g?
  - t = (3350-3480)/42.5 = -3.059, df = n-1 = 99
  - p(t(99) < -3.059) = pt(-3.059,df=99) = 0.0014

**t statistic**
\[ t = \frac{\bar{Y} - \mu}{s/\sqrt{n}} \]
one-sample t-test example
- sample mean = 93 (n=20)
- population standard deviation = ? (estimate from sample)
- is sample drawn from a population with a mean = 100?
- calculate t and compare to “critical values”

```math
\[
t = \frac{(Y - \mu)}{(\sigma / \sqrt{n})}
\]
```

\[
t = \frac{(93 - 100)}{\sqrt{20}} = -2.236
\]

the probability of obtaining t at least this extreme when H0 is true
\[
\text{pt}(2.236, df=(20-1)) + \text{pt}(2.236, df=(20-1), lower, tail=FALSE)
\]

p = 0.0375

What does a significant p-value mean?
- A significant p-value indicates that the result is unusual
  - assuming null hypothesis is true and assumptions are correct
- That is ALL it means
  - (1-p) is not equal to the probability of replicating the result...

Power (Type I & Type II Errors)

<table>
<thead>
<tr>
<th>decision</th>
<th>H0 is True</th>
<th>H0 is False</th>
</tr>
</thead>
<tbody>
<tr>
<td>reject H0</td>
<td>Type I (p = ( \alpha ))</td>
<td>Correct (p = ( 1 - \beta ) = power)</td>
</tr>
<tr>
<td>do not reject H0</td>
<td>Correct (p = ( 1 - \alpha ))</td>
<td>Type II error (p = ( \beta ))</td>
</tr>
</tbody>
</table>

Type I Error: reject H0 when it is true (alpha)
Type II Error: fail to reject H0 when it is false (beta)
Power = Probability of rejecting false H0 (1-beta)
What does a significant p-value mean?

- A significant p-value indicates that the result is unusual
  - assuming null hypothesis is true and assumptions are correct
- That is **ALL** it means
  - (1-p) is **not** equal to the probability of replicating the result
  - p is not equal to the probability that H0 is TRUE...

---

t test for single mean

R code

```r
set.seed(555410)
mu <- 0 # population mean
n <- 20 # sample size
stddev <- 1 # population sd
R <- 10000
t.val <- rep(0,R)
p.val <- rep(0,R)
for(kk in 1:R){
  the.sample <- rnorm(n,mu,randdev)
  t.results <- t.test(the.sample)
  t.val[kk] <- t.results$statistic
  p.val[kk] <- t.results$p.value
}
```

Notice that the range of acceptable scores — which do not cause us to reject the null hypothesis — is smaller

```
[1] 0.04999945
[1] 0.02499999 + 0.02499946 # the JOINT probability of getting a score <=95.6 OR >=104.38
```

Correctly Fail to Reject H0

```
80
20
45
```

Correctly Reject H0

```
1000
100
900
```

---

Result of Significance Test

<table>
<thead>
<tr>
<th>True state of H0</th>
<th>Type I Error (false alarm)</th>
<th>Type II Error (false miss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>Correctly Fail to Reject H0</td>
<td>45 out of 125 &quot;discoveries&quot; are false alarms</td>
</tr>
<tr>
<td>False</td>
<td>Correctly Reject H0</td>
<td></td>
</tr>
</tbody>
</table>

---

- **P** can be used to calculate
  - Any value, \( Y \) can be converted to a standard score using the formula
    \[
    z = \frac{Y - \mu}{\sigma}
    \]
  - \( z \) is distributed as a normal variable with
    - \( \mu = 0 \)
    - \( \sigma = 1 \)
  - \( z \) scores. This section shows how to use such a table to conduct
    - a test.
  - \( z \) from .025 to .05. If we set \( \alpha = 0.05 \) or \( (1-\alpha) = 0.95 \), our observed
    - \( p \) value
    - \( t \) value
    - \( z \) score
    - \( t \) statistic
      \[
      t = \frac{Y - \mu}{s/\sqrt{n}}
      \]
  - \( s \) is the standard deviation of the sample
    - \( s \) is the standard deviation of the population mean
    - \( \mu \) and \( \sigma \) are the population mean and standard deviation
    - \( \hat{\mu} \) is the sample mean

---

\( \hat{\mu} \) is from \( \mu \) by an amount \( \pm 2 \) standard errors.

---

\( \bar{X} \) is from \( \mu \) by an amount \( \pm 2 \) standard errors.

---

\( \hat{\mu} \) is not significantly different from \( \mu \) (\( \mu \) or \( \hat{\mu} \)).

---

\( \bar{X} \) is not significantly different from \( \mu \) (\( \mu \) or \( \bar{X} \)).
What does a significant p-value mean?

- A significant p-value indicates that the result is unusual
  - assuming null hypothesis is true and assumptions are correct
  - That is ALL it means
  - (1-p) is not equal to the probability of replicating the result
  - p is not equal to the probability that H0 is TRUE
  - p is not a measure of the evidence in favour of H0
    ‣ when H0 is true, all p values are equally likely!

What do p-values mean?  

- A p value is the probability of obtaining a result that is at least as extreme as observed result when the null hypothesis is true
  - p value = p(result GIVEN H0 is TRUE)
  - measures compatibility of our data with a specified model
- Used properly, p values can control our Type I error rate re: H0
  - in the long run, our Type I error rate will equal alpha

Lykken DT, Psychol Bulletin, 1968

“Statistical significance is perhaps the least important attribute of a good experiment; it is never a sufficient condition for claiming that a theory has been usefully corroborated, that a meaningful empirical fact has been established, or that an experimental report ought to be published.”

What do p-values mean?  

<table>
<thead>
<tr>
<th>P-VALUE</th>
<th>INTERPRETATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>HIGHLY SIGNIFICANT</td>
</tr>
<tr>
<td>0.02</td>
<td>SIGNIFICANT</td>
</tr>
<tr>
<td>0.03</td>
<td>CALCULATED</td>
</tr>
<tr>
<td>0.04</td>
<td>ON THE EDGE OF SIGNIFICANCE</td>
</tr>
<tr>
<td>0.05</td>
<td>HIGHLY SUGGESTIVE</td>
</tr>
<tr>
<td>0.06</td>
<td>SIGNIFICANT AT THE 5% LEVEL</td>
</tr>
<tr>
<td>0.07</td>
<td>HIGHLY SUGGESTIVE</td>
</tr>
<tr>
<td>0.08</td>
<td>HIGHLY SUGGESTIVE</td>
</tr>
<tr>
<td>0.09</td>
<td>HIGHLY SUGGESTIVE</td>
</tr>
<tr>
<td>0.1</td>
<td>HIGHER LOOK AT FURTHER INTERESTING SUBGROUP ANALYSIS</td>
</tr>
</tbody>
</table>

Please note that the p-value table is an approximation and should be used for illustrative purposes only.