Overview

- Course Information
- What is statistics?
- Ways of collecting data
- Modes of statistical analysis
- Sampling distributions & parameter estimation
- z tests & t tests

Course Management

- All lecture materials will be available on Avenue 2 Learn
  - reading assignments
  - lecture slides
  - labs & homework assignments
- Students are expected to install and use R
  - labs & exams will be in Psychology computer cluster (PC-154)
  - you may use your own laptop
- Labs, homework, & exams are open-book
  - you may use any/all aids to complete exams
  - you may collaborate on labs and homework assignments
Review of Statistical Inference

What is statistics?
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

Role of statistics in research (PPDAC)

- Understanding and defining the question
- How do we go about answering this question?
- What is already known, what are the data?
- Decisions: setting up a plan
- Collecting

Role of statistics in research

“...the purpose of statistics is to organize a useful argument from quantitative evidence, using a form of principled rhetoric.” - Robert P. Abelson

*rhetoric: the art of effective/persuasive speaking or writing
Ways of collecting data

**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 (e.g., ANCOVA)

**Correlational Studies**

- Measure the **association** between predictor & criterion variables
- Predictor variables are **not** manipulated by investigator
  - each event/subject comes with own set of variables
  - but values on variables differ across events/subjects
- Difficult to establish causal relation between variables

**Designed Experiments**

- Measure **causal effects** of independent variables on dependent variables
- Independent variables usually manipulated by experimenter
  - not always (e.g., “age” in developmental studies)
- Whenever possible, participants should be assigned randomly to conditions
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, subjects in most designed experiments 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.

Modes of Statistical Analysis

Descriptive vs. Inferential

Exploratory vs. Confirmatory

Descriptive vs. Inferential Statistics

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

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
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.
  - e.g., Is sample representative of population?

Sample Statistics vs. Population Parameters

- We can use sample statistics to estimate population parameters
  - The sample mean, $\bar{Y}$, is an unbiased estimate of population mean, $\mu$
  - The sample variance, $s^2$, is an unbiased estimate of population variance, $\sigma^2$
    - [N.B. True when using (n-1) in the denominator]
    - sample standard deviation, s, is a biased estimate of population variance, $\sigma$ [slightly too small]
  - The sample correlation, r, is a biased estimate of the population correlation, though the bias may be small when $n$ is large
- What is an “unbiased” estimate?
  - If the average value of many sample statistics (e.g., $\bar{Y}$) equals the population parameter (e.g., $\mu$), the statistic is an unbiased estimate of the parameter
Sampling Distributions of Mean and SD

- Distributions of statistics for 5000 samples (n=20)
  - Population mean = 100
  - Population variance = 100
  - Population sd = 10

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Mean</td>
<td>100</td>
<td>5.07</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>1066 (1053)</td>
<td>2.65</td>
</tr>
</tbody>
</table>

Interval Estimators & Confidence Intervals

- **Confidence Interval**
  - 95% Confidence Interval
    - An interval estimate of the value of a population parameter (e.g., population correlation)
    - E.g., the population correlation lies between “r-low” and “r-high”
    - CI_{95%} is calculated from data in your sample
    - The interval varies across samples
      - We wouldn’t expect it to be exactly the same for each random sample of (X,Y) values
    - In the long run, the interval contains the true population value 95% of the time
  - How can we calculate CI_{95%} for our correlation, r?
    - There are several methods... we first demonstrate the **percentile bootstrap method**
      - N.B. the method is not as important as understanding the meaning of the CI_{95%}

Calculating 95% Confidence Interval for ρ

- Our sample \( r = 0.52 \)
- What is CI_{95%} for ρ?
- **Percentile bootstrap method:**
  - Uses (X,Y) sample as estimate of (X,Y) population
  - Calculate r* on bootstrapped samples:
    - Randomly select 20 (X,Y) pairs from the data
    - Calculate r for each bootstrapped sample (r*)
    - Repeat MANY times
    - Display histogram of r*
  - Identify range of values containing 95% of r*
    - Range is PERCENTILE BOOTSTRAP estimate of CI_{95%} for ρ
distribution of bootstrapped r* values

- Calculate r* for many bootstrapped samples of data (n=20)
- (X,Y) data is an estimate of (X,Y) population
- Create bootstrapped sample:
  - Randomly select 20 (X,Y) pairs from data
  - Calculate r for each bootstrapped sample (r*)
  - Repeat MANY times
  - Display histogram of r*
- Identify range that contains 95% of r*
  - Range is PERCENTILE BOOTSTRAP estimate of 95% confidence interval for r
  - CI_{95%} = [0.17, 0.77]
  - CI_{95%} is our interval estimate of population parameter \( \rho \)

Confidence Interval of r

- Fisher's z transformation of r
- Transformed r's are approximately normal

Confidence Interval of r

- Fisher's z transformation of r
- Transformed r's are approximately normal
- Calculate 95% CI by calculating critical values that capture middle 95%
Confidence Interval of r

1. Convert r to z
2. Calculate standard deviation of sampling distribution 1/\sqrt{n-3}
3. Calculate values of z that cutoff lower/upper 2.5% of distribution
4. Calculate Confidence Interval defined by Fisher Z values
5. Transform Z values to r values

1 ( sampZ <- 0.5 * log((1+ourSampR)/(1-ourSampR)) )
[1] 0.5763398
2 ( zSE <- 1/sqrt(n-3) )
[1] 0.243
3 ( zCrit <- qnorm(0.975,0,1) ) # ± zCrit
[1] 1.96
4 ( zCI <- c(sampZ-zCrit*zSE, sampZ+zCrit*zSE) )
[1] 1.96
5 (rCI <- (exp(2*zCI)-1)/(exp(2*zCI)+1) )
[1] 0.1006368 0.7824667

A woman in the US has just given birth to a full-term baby weighing 291 kg. Is this weight unusually low?
Density Functions & Probability

Figure 2: The probability of randomly selecting a value of $x$ that is $\leq C$ -- i.e., $P(x \leq C)$ -- corresponds to the area under the probability density function that is to the left of $C$. $P(x \geq C)$ equals the area under the curve that is to the right of $C$.

**z test**

- $z$ is a standardized score: # standard deviations from mean
- used to evaluate individual scores and group mean when population variance is known
- when scores or means are distributed normally
  - $z$ is distributed normally with mean=0 and std dev=1
  - 95% of values fall between ±1.96
  - 99% of values fall between ±2.56

In the previous example, we knew that $\hat{Y} - \mu = 3480$ and $s = 462$. Now we want to know if our observed $\hat{Y}$ value is unusual. To determine this, we can use the $z$ test.

- The $z$ score – $z = \frac{\hat{Y} - \mu}{\sigma}$

- 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?

In the previous example, I used a computer to calculate the decision criteria for $z$ test.

<table>
<thead>
<tr>
<th>$z$ value</th>
<th>Decision Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt;1.96$</td>
<td>Reject null hypothesis</td>
</tr>
<tr>
<td>$&lt;-1.96$</td>
<td>Fail to reject null hypothesis</td>
</tr>
</tbody>
</table>

In the previous example, we considered any score falling below 95 or above 105 to be unusual. It is important to note, however, that getting an unusual score does not necessarily mean that the null hypothesis is false. For example, if the mean score is 105 and the standard deviation is 2.236, the probability of getting a score less than 95 or greater than 105 is only 0.025. Given their relatively low probability, it is reasonable to assert that scores that fall below 95 or above 105 are unusual. By this criterion, our observed mean score of 93 is unusual, and we therefore reject the null hypothesis that $\mu = 100$ in favor of the alternative hypothesis $\mu \neq 100$. This type of error — rejecting the null hypothesis when it is true — is referred to as a Type I error. The probability of making a Type I error is referred to as the significance level, which is set at 0.05 for this statistical test.

Nowadays, scientists are encouraged to publish the exact $z$ score rather than simply reporting whether a result is significant. This practice allows others to replicate the study and determine if similar results are obtained.

**Example:**

- 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?

Using R, we can calculate the probability:

```r
> pnorm(2910, mean=3480, sd=462)
[1] 0.109
```

In the previous example, we knew that $z = 2.56$ and $\alpha = 0.05$. Given these criteria, we conclude that the difference is statistically significant. This conclusion is based on the fact that the probability of getting a score of 75 or less, given that the null hypothesis is true, is less than $0.05$. Therefore, we conclude that the null hypothesis is false and that the observed differences are unlikely to be due to chance.
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 at least this low?
  - $p(z < -1.23) = 0.109$

z test for a group mean

z test for means

$$z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}}$$

- 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 test for means

- population mean & variance are known
  - mean = 3350 g; standard deviation = 425 g
- use z test
- convert observed mean to a z score: $$z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}}$$
Density Functions & Probability

**Example:** Normal Distribution

- For continuous distributions, the probability of selecting a value of $x$ that is less than or greater than some criterion, $C$, equals the area beneath the curve and to the left or right of $C$, respectively.

**Figure 2:** The probability of randomly selecting a value of $x$ that is $\leq C$—i.e., $P(x \leq C)$—corresponds to the area under the probability density function that is to the left of $C$. $P(x \geq C)$ equals the area under the curve that is to the right of $C$.

\[ [1] \ 0.01267143 \ # \text{prob of getting a score } \geq 105 \]

\[ >0.01267143 + 0.01267143 \]

\[ [1] \ 0.02534286 \ # \text{prob of getting a score } \leq 95 \text{ OR } \geq 105 \]

In other words, the probability of selecting a score that is less than 95 or greater than 105 is only 0.025.

Because the total probability is 1, the probability of selecting a score that is between 95 and 105 is $1 - 0.025 = 0.975$. Given their relatively low probability, it is reasonable to assert that scores that fall below 95 or above 105 are unusual. By this criterion, our observed mean score of 93 is unusual, and we therefore reject the null hypothesis that $\mu = 100$ in favor of the alternative hypothesis $\mu \neq 100$.

In our example, we considered any score falling below 95 or above 105 to be unusual. It is important to note, however, that getting an unusual score does not necessarily mean that the null hypothesis is false. After all, unusual scores are possible even when the null hypothesis is true. In fact, we expect to obtain an unusual score with a probability of 0.025 when the null hypothesis is true. Hence, it is possible that our decision to reject the null hypothesis is incorrect. This type of error—rejecting the null hypothesis when it is true—is called a Type I error. The probability of making this error is determined by the criteria we use to define a score as unusual. In this case, we used criteria (i.e., below 95 or above 105) which would lead to a Type I error 2.5% of the time. The probability of making a Type I error is referred to as $\alpha$, and so we would say that the Type I error rate, or $\alpha$, is 0.025 for this statistical test.

It is standard practice in Psychology to set $\alpha$ to either 0.05 or 0.01. If we set $\alpha = 0.05$, then our decision criteria would be 95.6 and 104.4:

\[ >\text{qnorm}(0.025,\text{mean} = 100,\text{sd} = 2.236,\text{lower.tail} = \text{TRUE}) \]

\[ \[1\] 95.61752 \ # \text{the probability of getting a score } \leq 95.6 \text{ is } 0.025... \]

\[ >\text{pnorm}(95.61752,\text{mean} = 100,\text{sd} = 2.236,\text{lower.tail} = \text{TRUE}) \]

\[ \[1\] 0.02499999 \]

\[ >\text{qnorm}(0.025,\text{mean} = 100,\text{sd} = 2.236,\text{lower.tail} = \text{FALSE}) \]

\[ \[1\] 104.3825 \ # \text{the probability of getting a score } \geq 104.38 \text{ is } 0.025... \]

\[ >\text{pnorm}(104.3825,\text{mean} = 100,\text{sd} = 2.236,\text{lower.tail} = \text{FALSE}) \]

\[ \[1\] 0.02500001 \]

\[ 3 \]

**t test for a group mean**

\[ z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \]

- Is group mean of 3350 g unusually low?
- $n = 100$; mean = 3350 g; standard deviation = 425 g
- Our null hypothesis (H0) is:
  - sample is drawn from population with parameters that are identical for Caucasian birth weights
  - distribution of BIRTHWEIGHTS: mean = $\mu = 3480$; sd = $\sigma = 462$; distribution = NORMAL
  - distribution of MEANS: mean = $\mu = 3480$; $\text{sd} = \sigma/\sqrt{n}$; distribution = NORMAL
- When H0 is true, what is probability of getting sample mean (n=100) < 3350 g?
  - Standard deviation of mean = Standard Error of Mean (SEM) = $462/\sqrt{100} = 46.2$
  - $z = (3350-3480) / 46.2 = -2.81$ [our mean is 2.81 standard deviations below $\mu$]
  - $p(z < -2.81) = \text{pnorm}(-2.81,0,1) = 0.0025$
- If sample was drawn from population of Caucasian birth weights, then group mean is unusually low

**t test for means**

- $t = \frac{\bar{Y} - \mu}{\text{SEM}}$

why use t instead of z?
**Effect of using estimate of σ**

- $z$ is defined with KNOWN population $\mu$ and $\sigma$
- only source of variation in $z$ is sampling error of mean
- using estimate of $\sigma$ introduces another source of variation in $z$
  - estimated $z$ depends on sample mean AND standard deviation
- does this affect our $z$ test?

$$z = \frac{\bar{X} - \mu}{\sigma \sqrt{n}}$$

**Effect of using estimate of population variation**

- William Gossett applied statistics to his work in the Guinness brewery
- Under the pseudonym, Student, he investigated effects of estimating $\sigma$ on $z$ test
  - sample variance is unbiased estimate of population variance
  - but sample standard deviation is a biased estimate of population standard deviation
  - sample SD underestimates population SD particularly for small sample sizes
- Discovered that using estimates of $\sigma$ lead to more extreme values of $z$ more frequently than predicted by statistical theory

**Effect of inflating $z$ score**

- calculating $z$ with estimated $\sigma$ inflates $z$ scores
- extreme $z$ values occur more frequently than expected when $H_0$ is True
- what effect does this have on our evaluation of $H_0$?
  - produces more Type I errors than expected

$$\hat{z} = \frac{\bar{X} - \mu}{\hat{\sigma} \sqrt{n}}$$

**Effect of using estimate of population variation**

- William Gossett applied statistics to his work in the Guinness brewery
- Under the pseudonym, Student, he investigated effects of estimating $\sigma$ on $z$ test
  - sample variance is unbiased estimate of population variance
  - but sample standard deviation is a biased estimate of population standard deviation
  - sample SD underestimates population SD particularly for small sample sizes
- Discovered that using estimates of $\sigma$ lead to more “extreme” values of $z$ than predicted by statistical theory
  - Caused an increase in Type I errors
    - especially for small samples
  - Devised a new test that corrected these errors
    - Student’s $t$ test
t distribution

- unimodal
- symmetrical around zero
- has 1 parameter:
  - degrees of freedom (df)
- df alters kurtosis
  - low df associated with narrower middle portion & heavier tails
- \( t \) approximately normal for \( df \approx 35 \)

Back to hypothesis testing

\[
t = \frac{X - \mu}{\sigma/\sqrt{n}}
\]

- When \( \sigma \) is NOT known
- estimated "\( z \)" is inflated
- our standardized score does not follow \( z \) distribution
- using "\( z \)" increases Type I error rate
- However, standardized score DOES follow a \( t \) distribution
- Therefore, our estimated "\( z \)" actually is a \( t \) statistic
- so we use critical values of \( t \), not \( z \), to evaluate null hypothesis

\[t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}\]

\[t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}}\]

- consider situation when we want to evaluate a group mean
  - e.g., measure birth weight of 100 Native-American full-term babies
  - our sample:
    - mean = 3350 g; standard deviation = 425 g
  - assuming our sample is drawn from typical population with UNKNOWN sigma
  - \( \mu = 3480; \sigma = \text{???} \); distribution’s shape = ??? [we will assume it is NORMAL]
  - is our sample mean of 3350 g unusually low?
When Ho is true, t will follow t distribution with 100-1=99 degrees of freedom. This t distribution is very similar to a standard normal distribution. We expect to get a sample t below the t=-1.66 approximately 5% of the time.

\[
t = \frac{X - \mu}{\sigma_X}
\]

- population mean is known (3480 g) but variance is unknown
- Sample: mean = 3350 g; standard deviation = 425 g; n = 100
- Ho: our sample was drawn from typical population
- assuming Ho is true, is our sample mean unusually low?
- convert observed mean to a t score:
- compare to critical value of t (t_{critical} = -1.66)
- observed t is more extreme than t_{critical}
- assuming Ho is true, our mean is unusually low
- reject null hypothesis (p < .05)

\[
t = \frac{(\bar{Y} - \mu)}{\sigma_{\bar{Y}}} = \frac{3380 - 3480}{425/\sqrt{100}} = -2.353
\]

\[
> qt(p=.05, df=100-1)
\]

> -1.660391

\[
t = \frac{(\bar{Y} - \mu)}{\sigma_{\bar{Y}}} = \frac{3380 - 3480}{425/\sqrt{100}} = -2.353
\]

\[
> qt(p=.05, df=100-1)
\]

[1] -1.660391

> boxplot(birthweight,ylab="birthweight (g)"
> t.test(x=birthweight, mu=3450, alternative="less")

One Sample t-test

data: birthweight

t = -2.3529, df = 99, p-value = 0.0103
alternative hypothesis: true mean is less than 3450
95 percent confidence interval:
-Inf 3420.567
sample estimates:
mean of x
3350