**Hypothesis Testing:**

**p-value**

**SECTION 4.2**

- Randomization distribution
- p-value

---

**Two Plausible Explanations**

- If the sample data support the alternative, there are two plausible explanations:
  1. The alternative hypothesis ($H_a$) is true
  2. The null hypothesis ($H_0$) is true, and the sample results were just due to random chance

- Do the data provide enough evidence to rule out #2?

---

**Key Question**

*How unusual is it to see a sample statistic as extreme as that observed, if $H_0$ is true?*

- If it is very unusual, we have *statistically significant* evidence against the null hypothesis

*Today’s Question: How do we measure how unusual a sample statistic is, if $H_0$ is true?*

**SIMULATE** what would happen if $H_0$ were true!

---

**Cocaine Addiction**

Is Desipramine better than Lithium at treating cocaine addiction (preventing relapses)?

What parameter(s) are we interested in?

- a) Proportion
- b) Mean
- c) Difference in proportions
- d) Difference in means
- e) Correlation

---

**Cocaine Addiction**

Is Desipramine better than Lithium at treating cocaine addiction (preventing relapses)?

$p_D$: proportion of cocaine addicts who relapse after Desipramine

$p_L$: proportion of cocaine addicts who relapse after Lithium

What are the relevant hypotheses?

- a) $H_0: p_D = p_L$, $H_a: p_D \neq p_L$
- b) $H_0: p_D = p_L$, $H_a: p_D < p_L$
- c) $H_0: p_D = p_L$, $H_a: p_D > p_L$
- d) $H_0: p_D < p_L$, $H_a: p_D = p_L$
- e) $H_0: p_D > p_L$, $H_a: p_D = p_L$
Measuring Evidence against $H_0$

To see if a statistic provides evidence against $H_0$, we need to see what kind of sample statistics we would observe, just **by random chance, if $H_0$ were true**.

Cocaine Addiction

- "by random chance" means by the random assignment to the two treatment groups
- "if $H_0$ were true" means if the two drugs were equally effective at preventing relapses (equivalently: whether a person relapses or not does not depend on which drug is taken)
- Simulate what would happen just by random chance, if $H_0$ were true...
Randomization Distribution

A *randomization distribution* is a collection of statistics from samples simulated assuming the null hypothesis is true.

- The randomization distribution shows what types of statistics would be observed, just by random chance, if the null hypothesis were true.

Key Question

*How unusual would it be to see a sample statistic as extreme as that observed, if $H_0$ were true?*

- A randomization distribution allows us to assess this!

Randomization Distribution

In a hypothesis test for $H_0: \mu = 12$ vs $H_a: \mu < 12$, we have a sample with $n = 45$ and $\bar{x} = 10.2$.

What do we require about the method to produce randomization samples?

- a) $\mu = 12$
- b) $\mu < 12$
- c) $\bar{x} = 10.2$

Randomization Distribution

In a hypothesis test for $H_0: \mu = 12$ vs $H_a: \mu < 12$, we have a sample with $n = 45$ and $\bar{x} = 10.2$.

Where will the randomization distribution be centered?

- a) 10.2
- b) 12
- c) 45
- d) 1.8
Randomization Distribution Center

- A randomization distribution simulates samples assuming the null hypothesis is true, so

A randomization distribution is centered at the value of the parameter given in the null hypothesis.

Randomization Distribution

In a hypothesis test for $H_0: \mu = 12$ vs $H_a: \mu < 12$, we have a sample with $n = 45$ and $\bar{x} = 10.2$.

What will we look for on the randomization distribution?

a) How extreme 10.2 is
b) How extreme 12 is
c) How extreme 45 is
d) What the standard error is
e) How many randomization samples we collected

Randomization Distribution

In a hypothesis test for $H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 > \mu_2$, we have a sample with $\bar{x}_1 = 26$ and $\bar{x}_2 = 21$.

What do we require about the method to produce randomization samples?

a) $\mu_1 = \mu_2$
b) $\mu_1 > \mu_2$
c) $\bar{x}_1 = 26, \bar{x}_2 = 21$
d) $\bar{x}_1 - \bar{x}_2 = 5$

Randomization Distribution

In a hypothesis test for $H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 > \mu_2$, we have a sample with $\bar{x}_1 = 26$ and $\bar{x}_2 = 21$.

Where will the randomization distribution be centered?

a) 0
b) 1
c) 21
d) 26
e) 5

Randomization Distribution

In a hypothesis test for $H_0: \mu_1 = \mu_2$ vs $H_a: \mu_1 > \mu_2$, we have a sample with $\bar{x}_1 = 26$ and $\bar{x}_2 = 21$.

What do we look for on the randomization distribution?

a) The standard error
b) The center point
c) How extreme 26 is
d) How extreme 21 is
e) How extreme 5 is

Remember: Statistical Significance

When results as extreme as the observed sample statistic are unlikely to occur by random chance alone (assuming the null hypothesis is true), we say the sample results are statistically significant.
Do you think the results from the cocaine addiction experiment are significant?

\[ \hat{p}_D \ - \hat{p}_L = 0.333 \]

a) Yes
b) No

- The **p-value** is the chance of obtaining a sample statistic as extreme (or more extreme) than the observed sample statistic, if the null hypothesis is true.

- The p-value can be calculated as the proportion of statistics in a randomization distribution that are as extreme (or more extreme) than the observed sample statistic.

1. What kinds of statistics would we get, just by random chance, if the null hypothesis were true? (randomization distribution)

2. What proportion of these statistics are as extreme as our original sample statistic? (p-value)

If the two drugs are equal regarding cocaine relapse rates, we have a 2% chance of seeing a difference in proportions as extreme as -0.333.

- **p-values** can be calculated by randomization distributions:
  - simulate samples, assuming \( H_0 \) is true
  - calculate the statistic of interest for each sample
  - find the p-value as the proportion of simulated statistics as extreme as the observed statistic.
**Cocaine Addiction**

- In the cocaine addiction experiment, people were actually randomized to one of three groups: Desipramine, Lithium, or Placebo.
- Does Desipramine do better than just a placebo at preventing relapses?
- Does Lithium do better than just a placebo at preventing relapses?

**Is Desipramine better than a placebo at preventing relapses?**

- $p_D$: proportion of cocaine addicts who relapse after Desipramine
- $p_P$: proportion of cocaine addicts who relapse after placebo

What are the relevant hypotheses?

1. $H_0: p_D = p_P$, $H_a: p_D \neq p_P$
2. $H_0: p_D = p_P$, $H_a: p_D < p_P$
3. $H_0: p_D = p_P$, $H_a: p_D > p_P$
4. $H_0: p_D < p_P$, $H_a: p_D = p_P$
5. $H_0: p_D > p_P$, $H_a: p_D = p_P$

**Desipramine vs Placebo**

Distribution of statistic if $H_0$ true

If there were no difference between Desipramine and placebo regarding cocaine relapses, we would only see a difference as extreme as that observed 1 out of 1000 times.

**Significant?**

Do you think the Desipramine vs placebo results from the cocaine addiction experiment are significant?

- a) Yes
- b) No

**Lithium vs Placebo**

Distribution of statistic if $H_0$ true

If there were no difference between Lithium and placebo regarding cocaine relapses, we would see a difference as extreme as the one observed about 34% of the time.

**Significant?**

Do you think the Lithium vs placebo results from the cocaine addiction experiment are significant?

- a) Yes
- b) No
Alternative Hypothesis

- A **one-sided** alternative contains either > or <
- A **two-sided** alternative contains ≠

- The p-value is the proportion in the tail in the direction specified by $H_a$
- For a two-sided alternative, the p-value is twice the proportion in the smallest tail

---

Sleep or Caffeine for Memory?

<table>
<thead>
<tr>
<th>Words</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>sleep</td>
</tr>
<tr>
<td>11</td>
<td>sleep</td>
</tr>
<tr>
<td>13</td>
<td>sleep</td>
</tr>
<tr>
<td>14</td>
<td>sleep</td>
</tr>
<tr>
<td>15</td>
<td>sleep</td>
</tr>
<tr>
<td>16</td>
<td>sleep</td>
</tr>
<tr>
<td>17</td>
<td>sleep</td>
</tr>
<tr>
<td>18</td>
<td>sleep</td>
</tr>
<tr>
<td>21</td>
<td>sleep</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Words</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>caffeine</td>
</tr>
<tr>
<td>7</td>
<td>caffeine</td>
</tr>
<tr>
<td>10</td>
<td>caffeine</td>
</tr>
<tr>
<td>12</td>
<td>caffeine</td>
</tr>
<tr>
<td>13</td>
<td>caffeine</td>
</tr>
<tr>
<td>16</td>
<td>caffeine</td>
</tr>
<tr>
<td>14</td>
<td>caffeine</td>
</tr>
<tr>
<td>15</td>
<td>caffeine</td>
</tr>
<tr>
<td>18</td>
<td>caffeine</td>
</tr>
</tbody>
</table>

mean = 15.25  mean = 12.25

---

Sleep or Caffeine for Memory?

<table>
<thead>
<tr>
<th>Words</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>sleep</td>
</tr>
<tr>
<td>11</td>
<td>caffeine</td>
</tr>
<tr>
<td>13</td>
<td>caffeine</td>
</tr>
<tr>
<td>14</td>
<td>sleep</td>
</tr>
<tr>
<td>15</td>
<td>caffeine</td>
</tr>
<tr>
<td>16</td>
<td>sleep</td>
</tr>
<tr>
<td>17</td>
<td>caffeine</td>
</tr>
<tr>
<td>18</td>
<td>sleep</td>
</tr>
<tr>
<td>21</td>
<td>sleep</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Words</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>caffeine</td>
</tr>
<tr>
<td>7</td>
<td>sleep</td>
</tr>
<tr>
<td>10</td>
<td>caffeine</td>
</tr>
<tr>
<td>12</td>
<td>caffeine</td>
</tr>
<tr>
<td>13</td>
<td>caffeine</td>
</tr>
<tr>
<td>16</td>
<td>sleep</td>
</tr>
<tr>
<td>14</td>
<td>caffeine</td>
</tr>
<tr>
<td>15</td>
<td>caffeine</td>
</tr>
<tr>
<td>18</td>
<td>caffeine</td>
</tr>
</tbody>
</table>

sleep mean = 14.25  caffeine mean = 13.25

sleep mean – caffeine mean = 1
**Sleep or Caffeine for Memory?**

- **p-value**
  
  \[
  p-value = 2 \times 0.022 = 0.044
  \]

\[
\bar{x}_S - \bar{x}_C = 3
\]

\[
\bar{x}_S \hat{=} \bar{x}_C \text{ when } H_0 \text{ true}
\]

**p-value and $H_0$**

- If the p-value is small, then a statistic as extreme as that observed would be unlikely if the null hypothesis were true, providing significant evidence against $H_0$.
- The smaller the p-value, the stronger the evidence against the null hypothesis and in favor of the alternative.

**Summary**

- The randomization distribution shows what types of statistics would be observed, just by random chance, if the null hypothesis were true.
- A p-value is the chance of getting a statistic as extreme as that observed, if $H_0$ is true.
- A p-value can be calculated as the proportion of statistics in the randomization distribution as extreme as the observed sample statistic.
- The smaller the p-value, the greater the evidence against $H_0$.

**To Do**

- Read Section 4.2
- HW 4.1, 4.2 due Friday 3/6