Hypothesis Testing

Null Hypothesis $H_0$

Alternative Hypothesis $H_1$

$Z$-statistic

Significance level

$P$-value

Making a test of significance

Significance tests for classifying and counting

2-sample $Z$-test

2 sample test for RC DB experiment

tests with more than 2 classes ($\chi^2$-test)
If the null hypothesis is true, what's the chance of getting a result as extreme as this or more extreme?

It is very unlikely that chance alone could result in such an extreme value.

*test statistic* measures the difference between data and what's expected under $H_0$

$$Z = \frac{\text{observed} - \text{expected}}{\text{SE}}$$

assuming $H_0$ is true

- how many SEs away an observed value is from its expected value when expected value is computed using $H_0$
Observed **significance level** is the chance of getting a test statistic as extreme or more than the observed one.

This is usually denoted $P$, and is called the **p-value**.
$H_0$ - null hypothesis

$H_1$ - alternative hypothesis.

$z$ - statistic.

$P$ - value. $\neq$ NOT the chance of $H_0$ being true.

$P$ is: if the null is true, it's the chance of observing data as extreme or more extreme, just due to chance variation.
Argument by contradiction

1. Assume that null hypothesis is true.
2. Compute the chance of the data, assuming $H_0$ — $P$-value.
3. Reject null hypothesis if the chance is too small.

Significance levels.

- $P < 5\%$ — statistically significant
- $P < 1\%$ — highly statistically significant

For arbitrary but conventional thresholds
An administrator claims that mean score in reading test is above 260.

A random sample of 8th graders has reading scores with mean 265, SD 55.

Does the data support the administrator's claim?

$H_0$ - mean test score is $\leq 260$

$H_1$ - mean test score $> 260$

test statistic $Z = \frac{\text{observed} - \text{expected}}{\text{SE}}$

$= \frac{265 - 260}{\frac{55}{\sqrt{85}}} = \frac{5}{\frac{55}{\sqrt{85}}} \approx 0.838$

$\frac{1}{2} (100 - A) = 21\%$

$P$-value is 21\%.

Do not have enough evidence to reject $H_0$. 
Light bulbs. - mean lifetime 750 hours.

Sample of 36 light bulbs. mean lifetime 725
SD 60

Is there evidence to reject the claim?

\[ H_0: \text{mean lifetime is } 750 \]

\[ H_1: \text{mean lifetime } < 750 \]

\[ Z \text{-statistic} = \frac{\text{observed - expected value under } H_0}{SE} \]

\[ = \frac{725 - 750}{60/\sqrt{36}} = -2.5 \]

\[ p\text{-value } < 0.6\% \]

Evidence against \( H_0 \).

Reject \( H_0 \) at 10\% level

Result is highly statistically significant
Sample

\[ SE_{\text{sum}} = \sqrt{\text{#draws}} \times SD_{\text{box}} \]

\[
SE_{\text{average of draws}} = \frac{SE_{\text{sum}}}{\text{#draws}} = \frac{\sqrt{\text{#draws}}}{\text{#draws}} \times SD_{\text{box}}
\]

\[
= \frac{SD_{\text{box}}}{\sqrt{\text{#draws}}}
\]
Box Plot

- lower quartile
- median
- upper quartile

IQR

Smallest data item (no small outliers in this case)

Largest non-outlier

Outlier

1.5 IQR above the upper quartile
Classification & counting.

Vaccine - effective
not effective.

$H_0$ - new vaccine is the same as the old one.

$H_0$ - 25% effective

Sample of size 2000, new vaccine was effective for 534 people.

$Z = \frac{\text{observed} - \text{expected under } H_0}{\text{SE}}$

$= \frac{534 - 500}{19.2} = 1.77$

$\frac{1}{2}(100 - A) = 4\%$

Reject $H_0$ at the 5% significance level.
Compared a value computed from a sample with a fixed expected value.

Is the parameter value for population 1 different from that for population 2?
Population 1

chips ahoy!
chocolate chip cookies	parameter - mean # chips in a cookie

Sample 1

Size 84
Mean 25.4
SD 3.56

generic brand C.C.

Sample 2

Size 72
Mean 29.6
SD 5.64

Does the data in the two samples allow us to say whether the mean # chips is different between the two brands?
Null hypothesis, $H_0$

the means of the two populations
are the same.

Alternative hypothesis $H_A$

the means differ

two-sample $z$-statistic $= \frac{\text{observed difference} - \text{expected difference}}{\text{SE for difference}}$

$= \frac{(29.6 - 25.4) - 0}{??}$

$\text{SE for difference of independent samples}$

$= \sqrt{(\text{first SE})^2 + (\text{second SE})^2}$

$= \sqrt{2.56^2 + 4.54^2}$

$= 6.07$ always larger than the
$\text{SE of either of the}$
$2$ $\text{samples.}$
SE for sample 1 = \frac{SD}{\sqrt{\frac{\text{Sample Size}}{2}}} = \frac{3.56}{\sqrt{84}} = 0.39.

SE for sample 2 = \frac{SD}{\sqrt{\frac{\text{Sample Size}}{2}}} = \frac{5.64}{\sqrt{72}} = 0.665.

z = \frac{4.2}{\sqrt{0.39^2 + 0.665^2}} = \frac{4.2}{0.77} = 5.5

\text{P-value} < 1\% - \text{highly significant}

Reject H_0 that the mean number of chips is the same.