STATISTICAL TESTING

1. **Box Model:** Every legitimate test of significance involves a box model. The test gets at the question of whether an observed difference is real, or just a chance variation. A real difference is one that says something about the box, and doesn't just reflect a fluke of sampling.

2. **Null Hypothesis:** The null hypothesis says that an observed difference just reflects chance variation. The alternate hypothesis says that the observed difference is real.

3. **Test Statistic:** A test statistic measures the difference between the data and what is expected if the null hypothesis is true. You must be able to compute the probability values for the test statistic; that is, the test statistics must follow the normal curve (z-test), or a t-curve (t-test), or a chi-squared curve (χ^2 - test), or some other known probability variable.

4. **P-Value:** The probability value or the observed significance level of a test is the chance or probability of getting the test statistic as extreme as or more extreme than the observed one. The probability is computed on the basis that the null hypothesis is correct. The smaller this probability is, the stronger the evidence against the null hypothesis.

Chapter 27: More Tests for Averages and Percentages

Two–Sample z-Tests

Example: A psychologist develops a test that measures social insight. She gives the test to a random sample of female students and a random sample of male students. She then compares the average scores for the two groups.

Example: Does regular physical therapy help lower back pain? An experiment is conducted and randomly assigns patients with lower back pain to two groups, a treatment group and a control group. Over a five-week period both groups receive regular examination and advice. The treatment group also gets bi-weekly physical therapy. At the conclusion of the experiment, the level of improvement is compared for both groups.

Example: A bank wants to know which of two incentive plans will most increase the use of its credit cards. It randomly divides 200 of its credit card customers into two groups, one will be offered the first incentive plan, and the other group will get the second incentive plan. After 10 weeks, the average amounts of credit-card purchases for each group are compared.

The standard error for the difference of two independent quantities is $\sqrt{a^2 + b^2}$ where

a **is the SE for the first quantity, and**

b **is the SE for the second quantity.**

Example 1.

Box A has an average of 65 and an SD of 10.
Box B has an average of 69 and an SD of 12.

We take a simple random sample of 300 from box A and an independent simple random sample of 500 from box B. What do you expect the difference (sample AV from Box A minus sample average from Box B) between the sample averages to be? Find the SE (standard error) for the difference between the sample averages.

Example 1.

Box A has an average of 65 and an SD of 10.
Box B has an average of 69 and an SD of 12.

A → Draw 300 & consider the AV
B → Draw 500 & consider the AV.

We take a simple random sample of 300 from box A and an independent simple random sample of 500 from box B. What do you expect the difference (sample AV from Box A minus sample average from Box B) between the sample averages to be? Find the SE (standard error) for the difference between the sample averages.

$65 - 69 = -4$ SE = ?

SE for AV of draws from A $= \frac{10\sqrt{300}}{300} = .58$

SE for AV of draws from B $= \frac{12\sqrt{500}}{500} = .54$

SE for the difference

$= \sqrt{(.58)^2 + (.54)^2} = \boxed{.79}$

Example 2.

County A has 47% Republicans.
County B has 45% Republicans.

We take a simple random sample of 300 from County A and an independent simple random sample of 500 from County B. What do you expect the difference (sample % from Box A minus sample % from Box B) between the sample percentages to be? Find the SE (standard error) for the difference between the sample percentages.

Example 2.

County A has 47% Republicans.
County B has 45% Republicans.

We take a simple random sample of 300 from County A and an independent simple random sample of 500 from County B. What do you expect the difference (sample % from Box A minus sample % from Box B) between the sample percentages to be? Find the SE (standard error) for the difference between the sample percentages.

$\underset{A}{\boxed{47\%\ 1s,\ 53\%\ 0s}} \rightarrow 300$ + consider % 1s drawn

$\underset{B}{\boxed{45\%\ 1s,\ 55\%\ 0s}} \rightarrow 500$ "

SE for % 1s drawn from A $= \frac{\sqrt{(.47)(.53)}\sqrt{300}}{300} \times 100\%$

$= 2.88\%$

SE for % 1 drawn from B $= \frac{\sqrt{(.45)(.55)}\sqrt{500}}{500} \times 100\%$

$= 2.22\%$

SE for difference %s

$= \sqrt{(2.88)^2 + (2.22)^2} = 3.63\%$

Example 3. A billboard company takes a simple random sample of 500 Salt Lake residents and finds that only 26 know that Calvin Coolidge was the 30th President. They then put up several yellow billboards that say "Calvin Coolidge was the 30th President". After a few weeks, they take an independent simple random sample of 500 residents, and 158 answer correctly. Is this evidence that people see the billboards?

Example 3. A billboard company takes a simple random sample of 500 Salt Lake residents and finds that only 26 know that Calvin Coolidge was the 30th President. They then put up several yellow billboards that say "Calvin Coolidge was the 30th President". After a few weeks, they take an independent simple random sample of 500 residents, and 158 answer correctly. Is this evidence that people see the billboards?

[?0s, ?1s] → 500 + consider % 1s drawn

Box A (Before)

[?0s, ?1s] → 500 + consider % 1s drawn

Box B: (after)

Null: no difference in % people who know (The billboards had no effect.)

Alt: more after

SE for % 1s from A = $\frac{\text{Box SD}\sqrt{500}}{500} \times 100\%$

$\approx \frac{\sqrt{\frac{26}{500} \cdot \frac{474}{500}} \times \sqrt{500}}{500} \times 100\% = 1\%$

SE for % 1s from B =

$\approx \frac{\sqrt{\frac{158}{500} \cdot \frac{342}{500}} \times \sqrt{500}}{500} \times 100\% = 2\%$

SE for diff $\approx \sqrt{1+4} = \sqrt{5} = 2.23$

$\frac{\text{observed diff} - \text{expected diff}}{\text{SE for diff}} = \frac{31.6\% - 5.2\%}{2.23}$

25.2

Example 4.

Both box A and Box B contain only 0s and 1s. You draw 100 times with replacement from box A and observe 54 1s. You draw 400 times with replacement from box B and observe 176 1s. Can you infer that the percentage of 1s in box A is larger than the percentage of 1s in Box B?

Example 4.

$\frac{176}{400} \times 100\%$

54% for A
44% for B

Both box A and Box B contain only 0s and 1s. You draw 100 times with replacement from box A and observe 54 1s. You draw 400 times with replacement from box B and observe 176 1s. Can you infer that the percentage of 1s in box A is larger than the percentage of 1s in Box B?

Box A : [? 0s, ? 1s] → Draw 100 + consider % 1s

Box B : [? 0s, ? 1s] → Draw 400 + consider % 1s.

Null: % 1s in A = % 1s in B

Alternative: % 1s in A > % in B

Test statistic: $\frac{\text{\% 1s drawn from A} - \text{\% 1s drawn from B}}{\text{SE for difference}}$

z-test

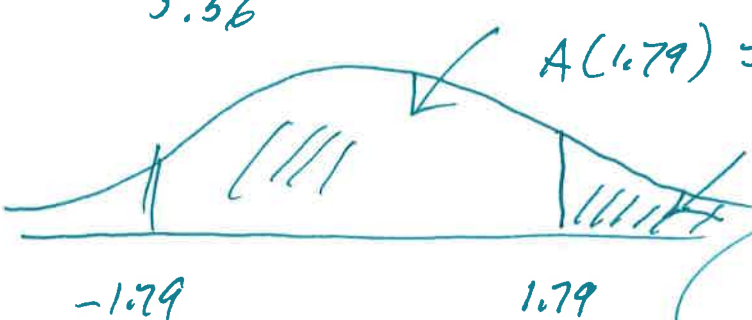
SE for % 1s from A ≈ $\frac{\sqrt{(.54)(.46)} \times \sqrt{100}}{100} \times 100\% = 4.98\%$

SE for % 1s from B ≈ $\frac{\sqrt{(.44)(.56)} \times \sqrt{400}}{400} \times 100\% = 2.48\%$

SE for difference = $\sqrt{(4.98)^2 + (2.48)^2} = 5.56\%$

$\frac{54-44}{5.56} = 1.79$

A(1.79) = 92%



p-value ≈ 4%

It is probably greater!

Example 5.

A simple random sample of 400 families is taken in Washington D.C. and the average of the 400 incomes is $61,000 with an SD of $9000.

Another simple random sample of 900 families is taken in Philadelphia: the average of the 900 incomes is $57,000 with an SD of $15,000.

Are they richer in Washington than in Philadelphia or is the sample difference just to chance error?

Example: 5

A simple random sample of 400 families is taken in Washington D.C. and the average of the 400 incomes is $61,000 with an SD of $9000.

Another simple random sample of 900 families is taken in Philadelphia: the average of the 900 incomes is $57,000 with an SD of $15,000.

Are they richer in Washington than in Philadelphia or is the sample difference just to chance error?

A: [Incomes for Wash. D.C. families] → Draw 400 & consider the AV of draws.

B: [Incomes for Phil. families] → Draw 900 & consider the AV of draws.

Null: ~~A Box AV = Box B AV~~
AV of Box A = AV of Box B

SE for AV of draws from A $= \frac{(9000)\sqrt{400}}{400} = \boxed{450}$

SE for AV of draws from B
$= \frac{(15,000)\sqrt{900}}{900} = \boxed{500}$

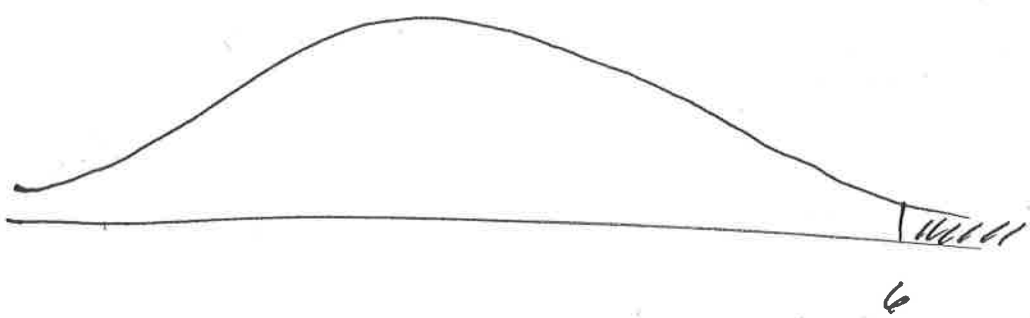
SE for the difference $\sqrt{450^2 + 500^2} = 672.68$

Test Statistic: Difference in AV of draws.

p-value:

$$\frac{61{,}000 - 57{,}000}{672.68}$$

$$= 5.94$$



p-value is nearly 0.

Reject!!

The two-sample z-test assumes that the two samples are independent. The samples should be large enough that the difference will follow the normal curve.

The two-sample z-tests can also be used for randomized, controlled, double-blind experiments!

Suppose there is a box of tickets. Each ticket has two numbers; one shows what the response would be to treatment A; the other, to treatment B. For each ticket, only one of the two numbers can be observed. Some tickets are drawn at random without replacement from the box, and the responses to treatment A are observed. Then a second sample is drawn at random without replacement from the remaining tickets. In the second sample, the responses to treatment B are observed. The SE for the difference between the two sample averages can be conservatively estimated as follows:

(i) compute the SEs for the averages as if drawing with replacement; Error: inflates the SE

(ii) combine the SEs as if the two samples were independent.

Error: deflates the SE

2 wrongs make a right!

Example 7.

There are 200 subjects in a small clinical trial on vitamin C. Half the subjects are assigned at random to treatment (2000 mg of vitamin C Daily) and half to control (2000 mg of placebo). Over the period of the experiment, the treatment group averaged 2.3 colds, and the SD was 3.1. The controls did a little worse: they averaged 2.6 colds, and the SD was 2.9. Is the difference in averages statistically significant?

Example

There are 200 subjects in a small clinical trial on vitamin C. Half the subjects are assigned at random to treatment (2000 mg of vitamin C Daily) and half to control (2000 mg of placebo). Over the period of the experiment, the treatment group averaged 2.3 colds, and the SD was 3.1. The controls did a little worse: they averaged 2.6 colds, and the SD was 2.9. Is the difference in averages statistically significant?

200 numbers
number of colds for each indiv.

Draw 100 & apply treatment & consider the AV of draws.

randomly assigned.

Draw 100 & consider AV of draws (Placebo)

Null Hypothesis : No difference

Test statistic : $\dfrac{\text{observed difference}}{\text{SE for difference}}$

SE for T : $\dfrac{(3.1)\sqrt{100}}{100} = .31$

SE for C : $\dfrac{(2.9)\sqrt{100}}{100} = .29$

Mistake Inflated: We assumed drawing with replacement.

p-value $\dfrac{(2.3-2.6)}{\sqrt{(.31)^2+(.29)^2}} = -.707$

$A(.707) = 52\%$

52%

p-value ≈ 24%

Mistake cuts it down

The SEs are not independent.

Example 8:

Many observational studies conclude that low-fat diets protect against cancer and cardiovascular "events" (heart attacks, stroke, and so forth). In 2006, the Women's Health Initiative (WHI) published its results. This was a large-scale randomized trial on women who had reached menopause. As one part of the study, 48,835 women were randomized: 19,541 were assigned to the treatment group and put on a low-fat diet. The other 29,294 women were assigned to the control group and ate as they normally would. Subjects were followed for 8 years. The investigators found that 1,357 women on the low-fat diet experienced at least one cardiovascular event, compared to 2,088 in the control group. Can the difference between the two groups be explained by chance? What do you conclude about the effect of the low-fat diet?

Ex 8:

$$\frac{1357}{19{,}541} \times 100\%$$

| ?0s, ?1s | → 19,541 |

T

1,357

6.94%

| ?0s, ?1s | → 29,294

C

2,088

$$\frac{2088}{29{,}294} \times 100\% = 7.13\%$$

Null: no difference, just chance

Alt:

SEs:

SE for % of 1s drawn in the treatment box

$$\frac{\sqrt{(.0694)(.9306)} \times \sqrt{19541}}{19541} \times 100\%$$

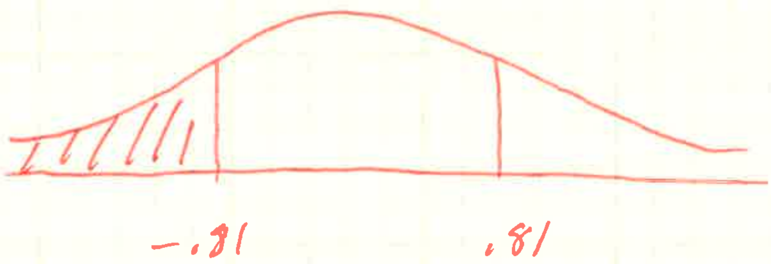
$$= .18\%$$

SE for % 1s drawn from control box

$$\frac{\sqrt{(.0713)(.9287)} \times \sqrt{29{,}294}}{29{,}294} \times 100\%$$

$$= .15\%$$

$$\frac{6.94 - 7.13}{\sqrt{(.18)^2 + (.15)^2}} = \frac{-.19}{.234} = -.81$$



$A(.81) \approx 58\%$

p-value $\approx 21\%$

or $.21$

The difference could be just do to chance variation, we do not reject the null hypothesis.

Example 9.

In a randomized, controlled, double blind experiment to test the effectiveness of "cholestyramine" there were 3,906 subjects. 1,906 subjects were randomly assigned to the treatment group with the remaining 1,900 subjects assigned to the control group. The results:

Treatment group: 155 heart attacks

Control group: 187 heart attacks

What do you conclude?

$\left(\frac{155}{1906} \times 100\%\right) = 8.13\%$, $\frac{187}{190} \times 100\% = 9.84\%$

[?0s, ?1s] → 1906 [?0s, ?1s] → 1900

Treatment, 8.13% Control, 9.84%

Null hypothesis: no difference, just chance variation

SE for T: $\dfrac{\sqrt{\frac{155}{1906} \cdot \frac{1751}{1906}} \times \sqrt{1906}}{1906} \times 100\% = \boxed{.626\%}$

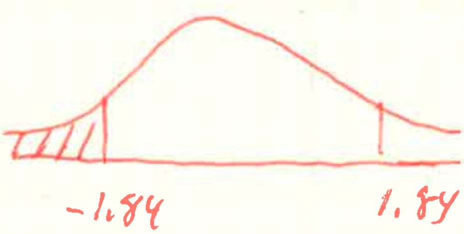
SE for C: $\dfrac{\sqrt{\frac{187}{1900} \cdot \frac{1713}{1900}} \times \sqrt{1900}}{1900} \times 100\% = \boxed{.683\%}$

SE for difference =

$$\sqrt{(.626)^2 + (.683)^2} = .926\%$$

Test statistic z

$$\frac{8.13\% - 9.84\%}{.926\%} = -1.84$$



$A(1.84) \approx 94\%$

p-value = 3%

Example 10.

Rosuvastatin is a cholesterol-lowering medication that was recently tested using a randomized, controlled, double-blind experiment. The researchers wanted to know whether or not Rosuvastatin protected against cardiovascular death. Of the 8901 subjects in the Rosuvastatin group, 83 died from cardiovascular causes; of the 8901 subjects in the placebo group, 157 died from cardiovascular causes. (**Source:** The *New England Journal of Medicine*, November 2008.)

$$T: \frac{83}{8901} \times 100\% = .93\%$$

Example 10.

$$C: \frac{157}{8901} \times 100\% = 1.76\%$$

Rosuvastatin is a cholesterol-lowering medication that was recently tested using a randomized, controlled, double-blind experiment. The researchers wanted to know whether or not Rosuvastatin protected against cardiovascular death. Of the 8901 subjects in the Rosuvastatin group, 83 died from cardiovascular causes; of the 8901 subjects in the placebo group, 157 died from cardiovascular causes. (**Source:** The *New England Journal of Medicine*, November 2008.)

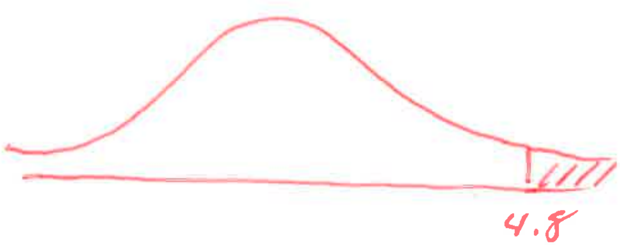
Null hypothesis: no difference

$$\text{SE for T: } \frac{\sqrt{\frac{83}{8901} \cdot \frac{8818}{8901}} \times \sqrt{8901}}{8901} \times 100\% = .102\%$$

$$\text{SE for C: } \frac{\sqrt{\frac{157}{8901} \cdot \frac{8744}{8901}} \times \sqrt{8901}}{8901} \times 100\% = .139\%$$

$$\text{SE for difference} = \sqrt{(.102)^2 + (.139)^2} = .172$$

$$\text{Test statistic: } \frac{1.76\% - .93\%}{.172\%} = 4.8$$



p-value ≈ 0

Reject the null, Rosuvastin is effective!

But what about the side effects?