Hypothesis Testing

SHENANDOAH UNIVERSITY

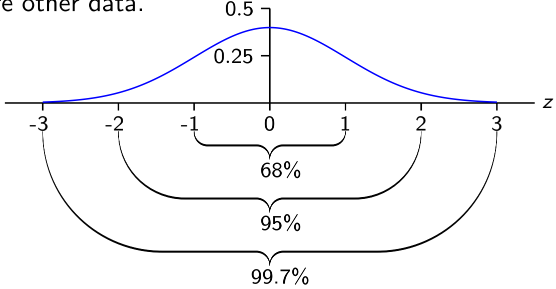
Dr. Ralph L. Wojtowicz

rwojtowi@su.edu

4 October 2019

The Normal Curve and Standard Units

- The **standard normal** (or Gaussian) curve is an ideal histogram to which we can compare other data.



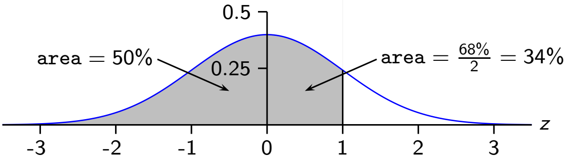
- If $x_1, \ldots, x_n$ is a list of numbers, we convert the values in the list to **standard units** using the following formula:

$$z_i = \frac{x_i - \text{mean}}{\text{SD}}$$

- z_i measures how far (in units of SD) x_i is from the mean (average) of the list

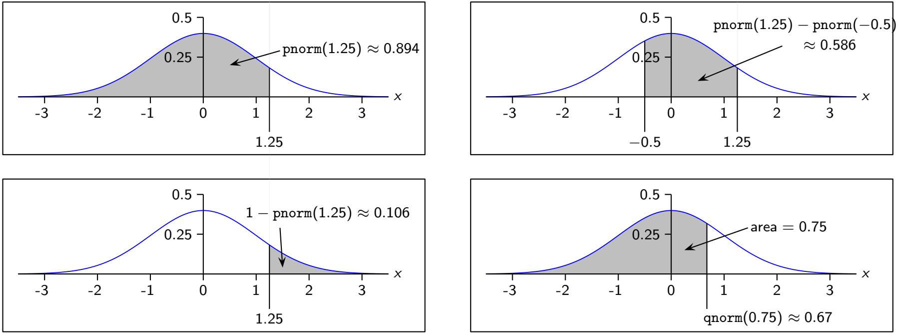
Finding Areas Under the Normal Curve

- Use one or more of the following to find areas under the normal curve:
  - Total area under the curve is 1 (that is, 100%)
  - The area is symmetric about vertical the line $x = 0$
  - (area to the left of x) $=$ (1 $-$ area to the right of x)
  - The 68%, 95%, 99.7% rules
  - The `pnorm` or `qnorm` functions in `R` (www.r-project.org)
  - Normal tables of values in many statistics textbooks.
- Example:



Total area $= 50\% + 34\% = 84\%$

Finding Areas Under the Normal Curve with R



Drs. Nullsheimer and Altshuler Debate the Contents of a Box of Numbered Tickets

- Drs. Nullsheimer and Altshuler are arguing about a large box of numbered tickets.
- Dr. Null says that the average is 50.
- Dr. Alt says the average is different from 50.
- They get tired of arguing and decide to look at some data.
- They draw a random sample of 500 tickets (with replacement).
- The average of the draws turns out to be 48 and the SD is 15.3.

Dr. Alt	The average is really below 50.
Dr. Null	Oh, come on, the difference is only 2, and the SD is 15.3. The difference is tiny relative to the SD. It's just chance.
Dr. Alt	Hmmm. Dr. Nullsheimer, I think we need to look at the SE, not the SD.
Dr. Null	Why is that?
Dr. Alt	Because the SE tells us how far the average of the sample is likely to be from its expected value — the average of the box.
Dr. Null	So, what's the SE?
Dr. Alt	Can we agree to estimate the SD of the box as 15.3, the SD of the data?
Dr. Null	I'll go along with you there.

Debate (continued)

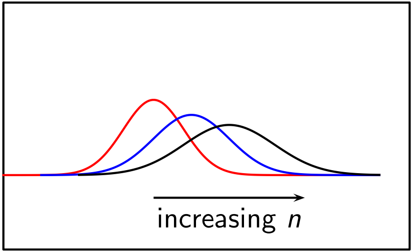
Dr. Alt	OK, then the SE for the average of the draws is about $15.3/\sqrt{500} \approx 0.7$.
Dr. Null	So?
Dr. Alt	The average of the draws is 48. You say it ought to be 50. If your theory is right, the average is about 3 SEs below its expected value.
Dr. Null	Where did you get the 3?
Dr. Alt	Well, $z = \dfrac{\text{observed value} - \text{expected value}}{\text{standard error}} = \dfrac{48-50}{0.7} \approx -3$
Dr. Null	You're going to tell me that 3 SEs is too many SEs to explain by chance?
Dr. Alt	That's my point. You can't explain the difference by chance. The difference is real. In other words, the average of the tickets in the box isn't 50, it's some other number.
Dr. Null	I thought the SE was about the difference between the sample average and its expected value.
Dr. Alt	Yes, yes. But the expected value of the sample average *is* the average of the tickets in the box.

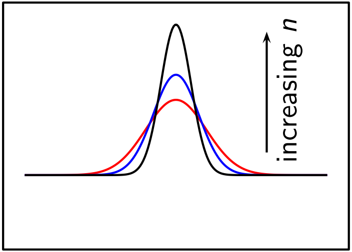
The Central Limit Theorem

When drawing at random with replacement from a box, the probability histogram for the sum (and the average) will follow the normal curve, even if the contents of the box do not. The histogram must be put into standard units, and the number of draws must be reasonably large.

$$\mathrm{EV}_{\mathrm{sum}} = n \cdot \mathrm{AV}_{\mathrm{box}} \qquad \mathrm{EV}_{\mathrm{av}} = \mathrm{AV}_{\mathrm{box}}$$

$$\mathrm{SE}_{\mathrm{sum}} = \sqrt{n} \cdot \mathrm{SD}_{\mathrm{box}} \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}_{\mathrm{box}}}{\sqrt{n}}$$





Variations on the Dialog Example

1. In the dialog, suppose the 500 tickets in the sample average 48 but the SD is 33.6. Who wins?

$$\mathrm{SE}_{\mathrm{av}} = \frac{33.6}{\sqrt{500}} \approx 1.5 \qquad \text{and} \qquad z = \frac{48-50}{1.5} = -1.3$$

2. Suppose 100 tickets are drawn, not 500, the sample average is 48 and the SD is 33.6. Who wins?

$$\mathrm{SE}_{\mathrm{av}} = \frac{33.6}{\sqrt{100}} \approx 3.36 \qquad \text{and} \qquad z = \frac{48-50}{3.36} = -0.59$$

3. Suppose 10,000 tickets are drawn, the sample average is 50.3 and the SD is 15. Who wins?

$$\mathrm{SE}_{\mathrm{av}} = \frac{15.0}{\sqrt{10,000}} = 0.15 \qquad \text{and} \qquad z = \frac{50.3-50}{0.15} = 2$$

Hypothesis Testing

- **Null Hypothesis**: The difference between the expected and observed values is due to chance.
- **Alternative Hypothesis**: No it's not!
- ***p*-Value**: If we assume that the null hypothesis is true, what is the chance of getting data as extreme or more extreme than the observed values?
- **Examples**:

$P(z < -3.00) \approx 0.1\%$	highly significant!	Accept Alternative
$P(z < -1.30) \approx 9.7\%$		Accept Null
$P(z < -0.59) \approx 27.8\%$		Accept Null
$P(z > -2.00) \approx 2.3\%$	significant	Accept Alternative

- **Small p-values support the alternative hypothesis.**

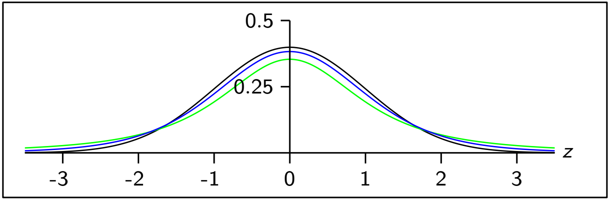
Coin-Tossing Example

A coin is tossed 10,000 times, and it lands heads 5,167 times. Is the chance of heads equal to 50%?

- **Null Hypothesis**: The coin is fair. The observed percentage 51.67% differs from the expected value 50.0% due to chance.
- **Alternative Hypothesis**: No, it's not fair!
- **Box model**: $\boxed{0 \quad 1}$
- $\text{AV}_{\text{box}} = 0.5$ $\qquad$ $\text{SD}_{\text{box}} = 0.5$
- $\text{EV}_{\text{av}} = 50\%$ $\qquad$ $\text{SE}_{\text{av}} = \frac{0.5}{\sqrt{10{,}000}} = 0.5\%$
- Test statistic $z = \dfrac{51.67 - 50.0}{0.5} \approx 3.3$
- ***p*-value** = area to the right of 3.3 under the normal curve $\approx 0.05\%$
- **Conclusion**: the coin is (very likely) not fair!

The t-Distribution and t-Test

- The Central Limit Theorem asserts that the sum (or average) of a random sample follows the normal curve if the sample size is reasonably large.
- If the sample size is small, then the appropriate probability distribution (ideal histogram) is a t-distribution.
- A t-distribution is similar to the normal curve but it is flatter in the middle and allows a bit more probability for extreme values.
- A t-distribution has a parameter called degrees of freedom (DOF) $= n - 1$ which influences its shape.
- As DOF increases, the t-distribution looks more like the normal curve.



Spectrophotometer Example I

A spectrophotometer is a device that can be used to detect concentrations of pollutants (such as CO) in air samples. A spectrophotometer is being calibrated with a manufactured sample. The technician makes five readings and gets

$$78 \quad 83 \quad 68 \quad 72 \quad 88$$

measured in ppm. If the device is working correctly, the expected value is 70.

- **Box model**: measurement = true value + bias + chance error
- Sample average = 77.8
- Sample SD = 8.1
- $\mathrm{SE}_{\mathrm{av}} = \frac{8.1}{\sqrt{4}} = 4.05$
- test statistic $z = \dfrac{77.8 - 70}{4.05} \approx 1.92$
- ***p*-Value**: area to the right of 1.92 under the t-distribution with $5 - 1 = 4$ degrees of freedom = 6.3%
- **Conclusion**: (Hesitantly) accept the null hypothesis.

Spectrophotometer Example II

Suppose the technician makes only three readings which are

$$71 \qquad 68 \qquad 79$$

measured in ppm. Is the device working correctly?

- Sample average $= 72.7$
- Sample SD $= 5.7$
- $\mathrm{SE}_{\mathrm{av}} = \frac{5.7}{\sqrt{3}} = 3.3$
- test statistic $z = \dfrac{72.7 - 70}{3.3} \approx 0.82$
- ***p*-Value**: area to the right of 0.82 under the t-distribution with $3 - 1 = 2$ degrees of freedom $= 25\%$
- **Conclusion**: Accept the null hypothesis.

Two-Sample z-Test (I)

400 draws are made at random with replacement from box A. The sample average and SD are 110 and 60. Independently, 100 draws are made at random with replacement from box B. The sample average and SD are 90 and 40. Are the averages of the tickets in the two boxes equal?

- **Null Hypothesis**: The boxes have the same average. The observed difference is simply due to chance.
- **Alternative Hypothesis**: No, it's not!
- $\mathrm{SE}_A = \frac{60}{\sqrt{400}} = 3$ $\qquad$ $\mathrm{SE}_B = \frac{40}{\sqrt{100}} = 4$
- $\mathrm{SE}_{\text{diff}} = \sqrt{3^2 + 4^2} = 5.$
- $z = \dfrac{\text{observed difference} - \text{expected difference}}{\text{standard error}} = \dfrac{(110 - 90) \ - \ 0}{5} = 4$
- ***p*-value** $=$ area to the right of 4 under the normal curve $\approx 0.0032\%$
- **Conclusion**: The averages are very likely **not** the same.

Two-Sample z-Test (II)

Repeat the last example but assume that the sample sizes were both 50.

- **Null Hypothesis**: The boxes have the same average. The observed difference is simply due to chance.
- **Alternative Hypothesis**: No, it's not!
- $\mathrm{SE}_A = \frac{60}{\sqrt{50}} = 8.5$ $\qquad$ $\mathrm{SE}_B = \frac{40}{\sqrt{50}} \approx 5.7$
- $\mathrm{SE}_{\mathrm{diff}} = \sqrt{8.5^2 + 5.7^2} \approx 10.2.$
- $z = \dfrac{(110 - 90) \ - \ 0}{10.2} = 1.96$
- ***p*-value** $=$ area to the right of 1.96 under the normal curve $\approx 2.5\%$
- **Conclusion**: The averages are likely **not** the same.