SWARTHMORE

STAT011 Statistical Methods I

Lecture 14 Statistical Inference

Lu Chen
Swarthmore College
3/19/2019

Outline

- Review
  - Central limit theorem (CLT)
  - Level C confidence interval for a population mean
  - Level α z test for a population mean
- Motivation example - Emoji 😱😱😱
- Confidence interval - some cautions
- Hypothesis testing
  - The P-value and a statement of significance
  - A level C confidence interval and a level α two-sided test
  - Some cautions

Review - Central limit theorem (CLT)

Draw an SRS of size n from **any population** with mean μ and finite standard deviation σ. When ***n* is large**, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal:

$$\bar{x} \text{ is approximately } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{x} \overset{approx.}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- Central limit theorem holds for any distribution.
  - Population distribution is Normal.
  - Population distribution is not Normal.
  - Population distribution is Bernoulli.

Review - Confidence interval

Choose an SRS of size n from a population having unknown mean μ and known standard deviation σ. The margin of error for a level C confidence interval for μ is

$$m = z^* \frac{\sigma}{\sqrt{n}}.$$

Here, z^* is the value on the standard Normal curve with area C between the critical points $-z^*$ and z^*. The **level C confidence interval for μ** is

$$\bar{x} \pm m = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}.$$

The confidence level of this interval is exactly C when the population distribution is Normal and is approximately C when n is large in other cases.

Review - Confidence interval

Confidence level C	0.9	0.95	0.99
Critical point z^*	1.645	1.960	2.576

To find z^* , use `qnorm(1-(1-C)/2)`.

In the midterm exam,

- For 80% C.I., $z^* =$ `qnorm(0.9)`
- For 95% C.I., $z^* =$ `qnorm(0.975)`

Review - Hypothesis testing

To test the hypothesis $H_0 : \mu = \mu_0$ based on an SRS of size n from a population with unknown mean μ and known standard deviation σ, compute the **test statistic**

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of a standard Normal random variable Z, the P-value for a test of H_0 vs.

$$\begin{aligned} H_a &: \mu > \mu_0 \quad \text{is } P(Z \geq z) \\ H_a &: \mu < \mu_0 \quad \text{is } P(Z \leq z) \\ H_a &: \mu \neq \mu_0 \quad \text{is } 2P(Z \geq |z|) \end{aligned}$$

These P-values are exact if the population distribution is Normal and are approximately correct for large n in other cases.

Review - Hypothesis testing

z **Test**			
H_0	$\mu = \mu_0$		
H_a	$\mu > \mu_0$	$\mu < \mu_0$	$\mu \neq \mu_0$
Test statistic	$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \overset{approx.}{\sim} N(0, 1)$		
P**-value**	$P(Z \geq z)$ `1-pnorm(z)`	$P(Z \leq z)$ `pnorm(z)`	$2P(Z \geq \lvert z \rvert)$ `2*(1-pnorm(abs(z)))`
Conclusion	$P > \alpha$: statistically insignificant at level α; cannot reject H_0 $P \leq \alpha$: statistically significant at level α; reject H_0		

Emoji

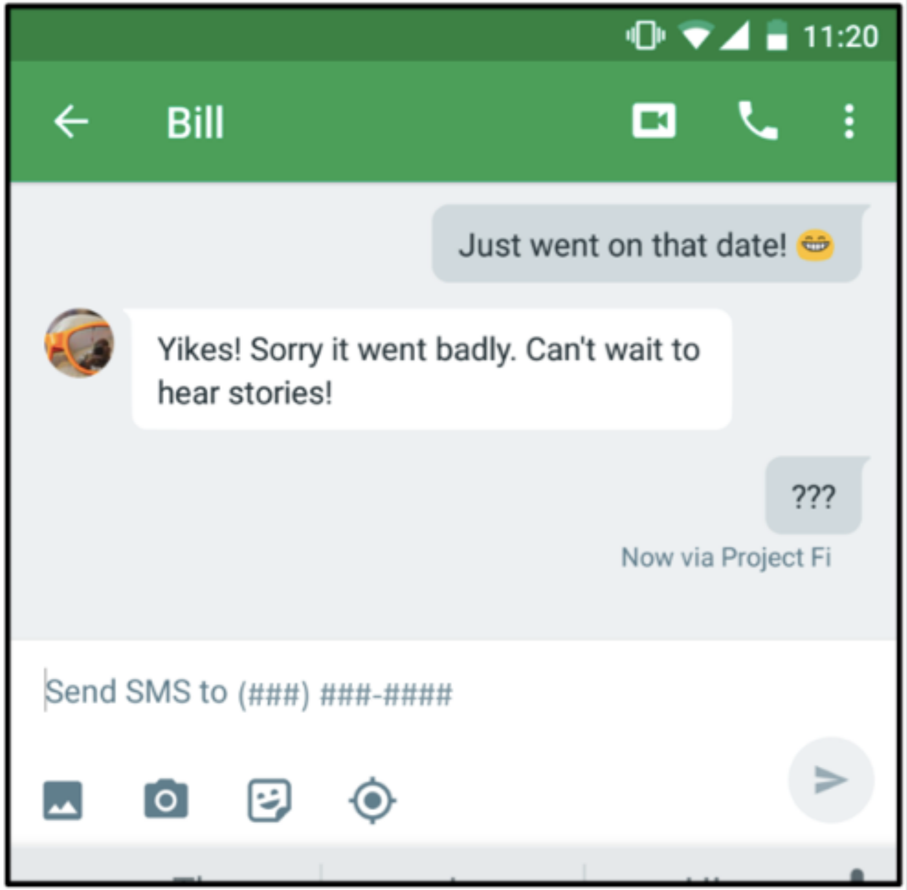
Investigating the Potential for Miscommunication Using Emoji

By Hannah Miller on April 5, 2016

Hey emoji users: Did you know that when you send your friend 😁 on your Nexus, they might see 😬 on their iPhone? And it's not just 😁; this type of thing can happen for all emoji (yes, even 💩). In a paper (download) that will be officially published at AAAI ICWSM in May, we show that this problem can cause people to misinterpret the emotion and the meaning of emoji-based communication, in some cases quite significantly. 😱, we know.

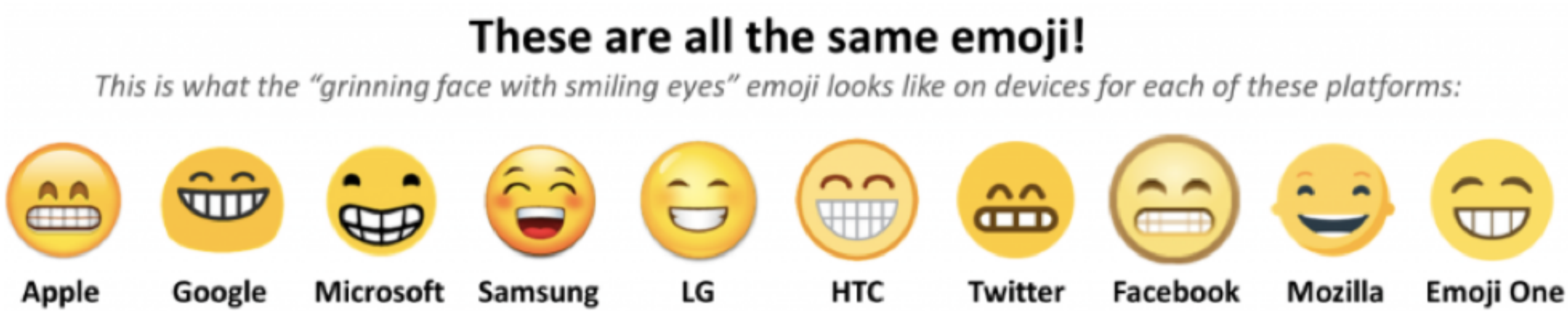
Emoji

Abby using a Google Nexus, texting Bill:



Bill using an iPhone, texting Abby:

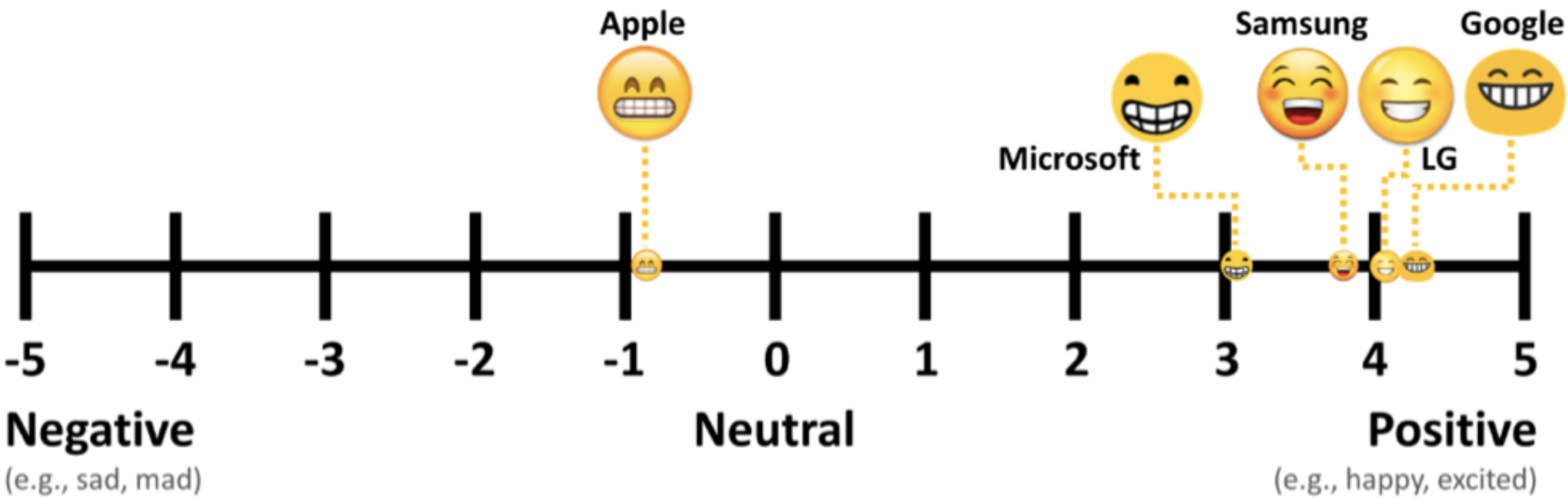
Emoji



- The investigators' questions of interest:
  - Across platform: Do people interpret one platforms rendering of an emoji character the same way that they interpret a different platform's rendering?
  - Within platform: Do people look at the exact same rendering of a given emoji and interpret it the same way?

Emoji

Across platform: the investigators assessed the emotional meaning of each rendering on a scale from -5 (strongly negative) to 5 (strongly positive).



Emoji

Within-platform score of mis-communication (25 emoji for each platform)

	Apple	Google	Microsoft	Samsung	LG
Top 3	3.64	3.26	4.40	3.69	2.59
	3.50	2.66	2.94	2.36	2.53
	2.72	2.61	2.35	2.29	2.51
•••	•••				
Bottom 3	1.25	1.13	1.12	1.23	1.30
	0.65	1.06	1.08	1.09	1.26
	0.45	0.62	0.66	1.08	0.63

Google, MS, Samsung and LG together

- Average score of mis-communication: 1.84
- Number of emoji's: 100
- Population standard deviation: 0.50

- Find the 95% confidence interval for the population mean score of mis-communication for the four platforms.
- Is the average score of mis-communication of the four platforms different from 2, which is the average score of Apple emoji?

Emoji

$\bar{x} = 1.84, \sigma = 0.5, n = 100, C = 0.95$

95% confidence interval

- $C = 0.95, z^* = 1.96$
- Margin of error

$$m = z^* \frac{\sigma}{\sqrt{n}} = 1.96 \frac{0.5}{\sqrt{100}} = 0.10$$

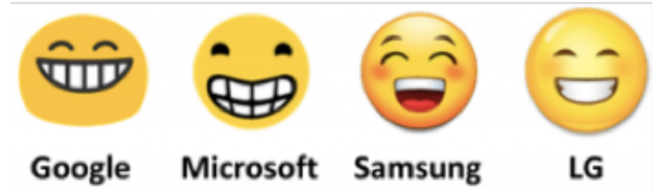
- 95% confidence interval $\bar{x} \pm m = 1.84 \pm 0.10$
- We are 95% confident (about the method) that the population mean score of mis-communication for the four platforms will be within [1.74, 1.94]

Confidence interval - some cautions

- The data should be an SRS from the population.
  - If not *simple* random sampling, other methods for calculating confidence intervals (C.I.s) are available.
  - If not simple *random* sampling, estimates may be biased and there is no statistical method for correction of bias.
- Outliers are very influential on C.I. because it is constructed based on mean $\bar{x}$.
- If the population distribution is Normal, the interval is exact; otherwise it is approximate.
  - Usually we require $n \geq 15$ for the approximation works well.
- Population standard deviation σ is assumed to be known. We will learn methods for calculating C.I. when σ is unknown.

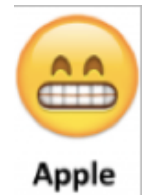
Emoji

Is the average score of mis-communication of the four from 2, which is the average score of Apple emoji?



$\bar{x} = 1.84, \sigma = 0.5, n = 100, \alpha = 0.05$

VS.

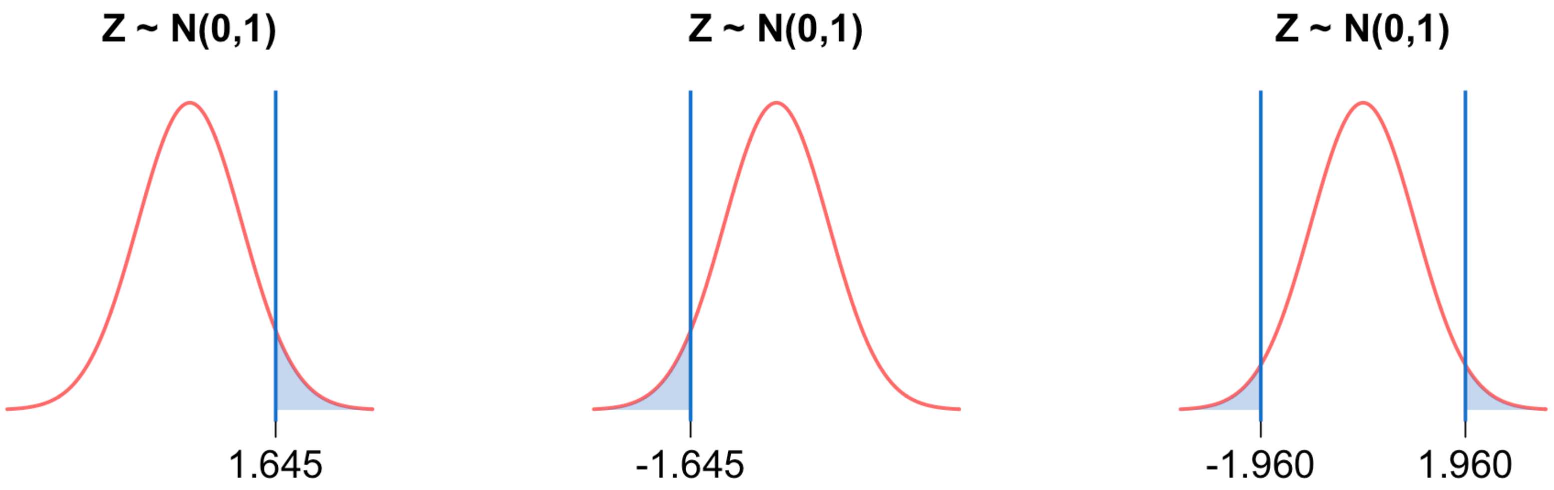


- $H_0 : \mu = 2; H_a : \mu \neq 2$
- $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.84 - 2}{0.5/\sqrt{100}} = \frac{-0.12}{0.05} = -3.2$$
- $2P(Z \geq |z|) = 2P(Z \geq 3.2) = 0.0014 < 0.05$ `2*(1-pnorm(3.2))`
- The test is significant at level 0.05 and we reject H_0. The mean score of mis-communication of the four platforms is significantly different from 2.
- **Note**: here we assume the mean score of Apple Emoji $= 2$ is fixed, which is not true since it is a sample mean. In Lecture 16, we will learn the method to compare two sample means.

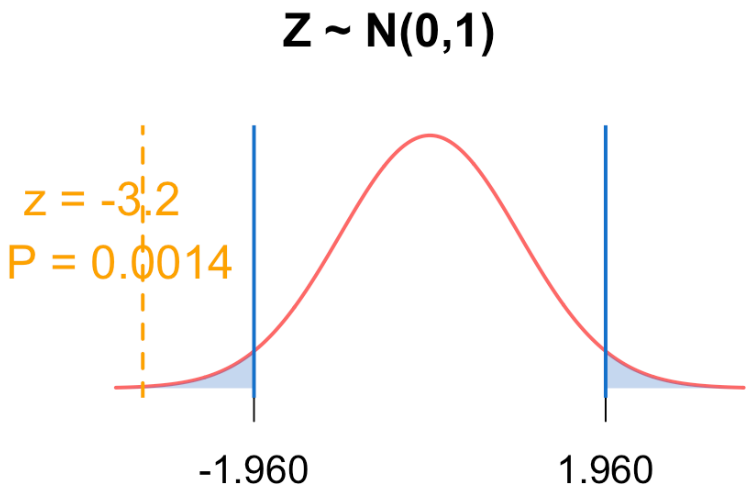
The P-value and a statement of significance

Level α test: we reject H_0 when $P \leq \alpha$. If $\alpha = 0.05$,

- For $H_a : \mu > \mu_0$, when $z \geq 1.645$, $P \leq$ `1-pnorm(1.645)` $= 0.05$
- For $H_a : \mu < \mu_0$, when $z \leq -1.645$, $P \leq$ `pnorm(-1.645)` $= 0.05$
- For $H_a : \mu \neq \mu_0$, when $z \geq 1.96$ or $z \leq -1.96$, $P \leq$ `2*(1-pnorm(1.96))` $= 0.05$



The P-value and a statement of significance



For this level $\alpha = 0.05$ two-sided test, there are two different ways to make a statement of significance (reject-or-not):

1. $P = 0.0014 < 0.05$, reject H_0
2. $z = -3.2 < -1.960$, reject H_0

Here 1.960 and -1.960 are called **critical values/points**.

- This is equivalent to a 95% confidence interval!

The confidence interval and a two-sided test

For the emoji example,

- 95% confidence interval for the population mean score of mis-communication is $1.84 \pm 1.960 \frac{0.5}{\sqrt{100}} = 1.84 \pm 0.10$
  - Apple emoji mean score 2 does not fall into the interval [1.74, 1.94], the population mean should be significantly different from 2.
  - 1.960 and -1.960 are the z^* values we used to calculate C.I..
- Level 0.05 hypothesis testing: $z = \frac{1.84-2}{0.5/\sqrt{100}} = -3.2$. Reject $H_0 : \mu = 2$ for
  - $P = 0.0014 < 0.05$ or
  - $z = -3.2 < -1.960$
- The 95% confidence interval does not contain 2 $\Leftrightarrow$ the level 0.05 two-sided test for $H_0 : \mu = 2$ is significant.

The confidence interval and a two-sided test

> A level α two-sided significance test rejects a hypothesis $H_0 : \mu = \mu_0$ exactly when the value μ_0 falls outside a level $C = 1 - \alpha$ confidence interval.

- This only holds for a two-sided test.
- A level α two-sided test is equivalent to a level $C = 1 - \alpha$ confidence interval in the statement of significance.
  - Level 0.01 two-sided test $\Leftrightarrow$ level 99% confidence interval
  - Level 0.05 two-sided test $\Leftrightarrow$ level 95% confidence interval
  - Level 0.10 two-sided test $\Leftrightarrow$ level 90% confidence interval

The confidence interval and a two-sided test

If the test is **one-sided**, its statement of significance is **not equivalent to** that from the confidence interval, for confidence interval is always "two-sided".

For a one-sided test $H_a : \mu > \mu_0$,

- Level 0.01 test $z^* = 2.326$; 99% confidence interval $z^* = 2.576$ and -2.576.
- Level 0.05 test $z^* = 1.645$; 95% confidence interval $z^* = 1.960$ and -1.960.
- Level 0.10 test $z^* = 1.282$; 90% confidence interval $z^* = 1.645$ and -1.645.

For a one-sided test $H_a : \mu < \mu_0$,

- Level 0.01 test $z^* = -2.326$; 99% confidence interval $z^* = 2.576$ and -2.576.
- Level 0.05 test $z^* = -1.645$; 95% confidence interval $z^* = 1.960$ and -1.960.
- Level 0.10 test $z^* = -1.282$; 90% confidence interval $z^* = 1.645$ and -1.645.

Confidence interval, test and significance

z^* values

α	$\mu > \mu_0$	$\mu < \mu_0$	$\mu \neq \mu_0$	C	**C.I.**
0.01	2.326 `qnorm(0.99)`	-2.326 `qnorm(0.01)`	± 2.576 `qnorm(1-0.01/2)`	**0.99**	± 2.576 `qnorm(1-0.01/2)`
0.05	1.645 `qnorm(0.95)`	-1.645 `qnorm(0.05)`	± 1.960 `qnorm(1-0.05/2)`	**0.95**	± 1.960 `qnorm(1-0.05/2)`
0.10	1.282 `qnorm(0.90)`	-1.282 `qnorm(0.10)`	± 1.645 `qnorm(1-0.10/2)`	**0.90**	± 1.645 `qnorm(1-0.10/2)`
a	`qnorm(1-a)`	`qnorm(a)`	`qnorm(1-a/2)`	$1 - a$	`qnorm(1-a/2)`

Hypothesis testing - some cautions

- A level α test rejects H_0 at
  - $P \leq \alpha$ or
  - $z \geq z^*, z \leq z^*$ (one-sided test) or
  - $z \geq z^*$ or $z \leq -z^*$ (two-sided test)
  - z^* only tells us to reject or not, but P-value shows the strength of rejection.
- A level α two-sided significance test is equivalent to a level C confidence interval, where $\alpha = 1 - C$ or $C = 1 - \alpha$.
- There is no sharp border between "significant" and "not significant", only increasingly strong evidence as the P-value decreases.
- Beware of searching for significance.
  - Choose H_a and level α before the test.

Hypothesis testing - some cautions

- Statistical significance does not suggest practical significance.
  - $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$. When n is large, even $\bar{x} - \mu_0$ is tiny, z could be large and the test could be statistically significant.
- Don't ingore lack of significance.
  - Sometimes, a practically meaningful result can hardly achieve statistical significance due to lack of large sample.
- Be aware of outliers.
  - $\bar{x}$ is not resistant. Always do exploratory data analysis before conducting any test.
- Statistical inference is not valid for all sets of data.
  - The z test assumes Normal distribution of $\bar{x}$.