MATH 3339
Statistics for the Sciences
Sec 7.3;7.4;7.6

Wendy Wang
wwang60@central.uh.edu

Lecture 14 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

1. More Examples of Confidence Intervals for population means
2. Sample Size
3. Checking for Normality
4. Estimating Proportions
5. Estimating Variance and Standard Deviation

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Interval for μ Recap

- Z-confidence interval, given the population standard deviation, σ is **known**

$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

- T-confidence interval, given that the population standard deviation, σ is **unknown**

$$\bar{x} \pm t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example

Suppose your class is investigating the weights of Snickers 1−ounce fun−size candy bars to see if the customers are getting full value for their money. Assume the weights are Normally distributed. Several candy bars are randomly selected and weighed with sensitive balances borrowed from the physics lab. The weights are:

$$0.95, \quad 1.02, \quad 0.98, \quad 0.97, \quad 1.05, \quad 1.01, \quad 0.98, \quad 1.00$$

We want to determine a 90% confidence interval for the true mean weight of these candy bars.

σ is NOT given,

t- interval

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

You try

From each case determine if we want to use **a)** Z-confidence interval or **b)** T-confidence interval

1. We assume the monthly rent for an apartment is approximately Normally distributed with a standard deviation of \$90. We find the mean monthly rent for a random sample of 10 apartments is \$640. Find a 99% confidence interval for the mean monthly rent for apartments.

   σ is given, $\sigma = 90$

   (a)

2. From a sample of 16 apartments the mean is \$508 with a standard deviation of \$78. Find a 99% confidence interval for the mean monthly rent for apartments.

   $s = 78$, σ is <u>NOT</u> given

   (b)

UNIVERSITY of HOUSTON DEPARTMENT OF MATHEMATICS

What will happen if changing Confidence Levels C

Suppose we have a 99% confidence interval for the population mean, μ of (2.272, 17.728). If we change the confidence level to 92% what is the confidence interval?

$99\% \longrightarrow 92\%$

narrower.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

You try

3. If the 90% confidence limits for the population mean are 34 and 46, which of the following **could** be the 99% confidence limits
   a) (36, 41)
   b) (39, 41)
   c) (30, 50)
   d) (39, 43)
   e) (38, 45)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Changing Sample Size

The mean of a random sample of n measurements is equal to $\bar{x} = 33.9$. Assume $\sigma = 3.3$. Determine the margin of error for a 95% confidence interval for the population mean when the sample size is

1. n = 100
2. n = 400

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Behavior of Confidence Intervals with different n

Notice that as the sample size increases the width of the interval decreases. Or the confidence interval becomes thinner.

- First: Mathematically, notice that we are dividing by a larger number, so that will decrease the quotient.

- Second: Intuitively, as the sample size increases the accuracy of the estimation becomes better, thus the point estimate is getting closer to the population mean and the interval does not need to be as wide.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Choosing Sample Size

You can have both a high confidence while at the same time a small margin of error by taking enough observations.

- Sample size for confidence intervals of means.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Starting Salary

- We want to estimate annual starting salaries for college graduates with degrees in business administration. To determine this we need a sample.
- Assume that a 95% confidence interval estimate of the population mean annual starting salary is desired.
- Assume the standard deviation is $\sigma = \$7,500$.
- How large a sample should be taken if the desired margin of error is $m = \$500$?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Choosing Sample Size

You can have both a high confidence while at the same time a small margin of error by taking enough observations.

- Sample size for confidence intervals of means.

$$n > \left(\frac{z_{\alpha/2}\sigma}{m} \right)^2$$

Starting Salary

- We want to estimate annual starting salaries for college graduates with degrees in business administration. To determine this we need a sample.
- Assume that a 95% confidence interval estimate of the population mean annual starting salary is desired.
- Assume the standard deviation is $\sigma = \$7,500$.
- How large a sample should be taken if the desired margin of error is $m = \$500$?
- How large a sample should be taken if the desired margin of error is $m = \$100$?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

You try

4. Given $\sigma = 7500$ and a 95% confidence level. What should be the sample size if we want the margin of error to be $m = \$100$?

   a) 7125
   b) 147
   c) 21,609
   d) 21,610

$n > (\quad)^2$

$= \underline{21609}$

$\underline{21610}$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Assumptions for Estimating the Population Mean

1. The sample has to be as a result of a simple random sample (SRS).
2. The distribution of the population has to be Normal. By the Central Limit Theorem if our sample size is larger than 30 then the sample means have a Normal distribution.

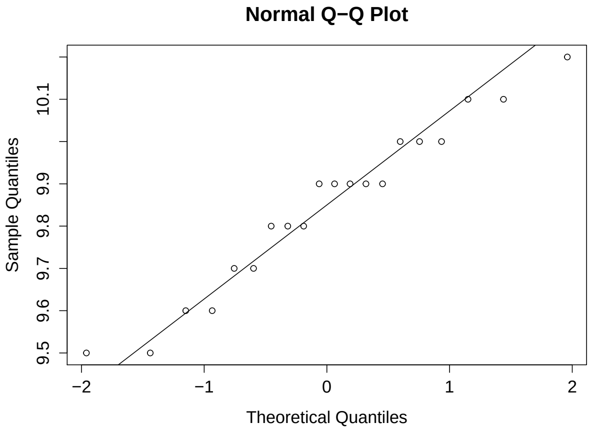
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Assessing the Normality of data

- In order to use the t-interval and the confidence interval for variances we need to judge whether data are approximately Normal.
- Most useful tool for assessing Normality is a graph called **Normal quantile plot**.
- This plots the value against what the z-score should be if we were using a Normal distribution. For example from the coffee machine Q1 = 9.7 so this is the 25th percentile, thus if we have a Normal distribution, z = -0.67. We would put a point at (-.67, 9.7).
- If the plots follow a straight line then we can say that we have an approximate Normal distribution.
- Rcode: qqnorm(x) then qqline(x).

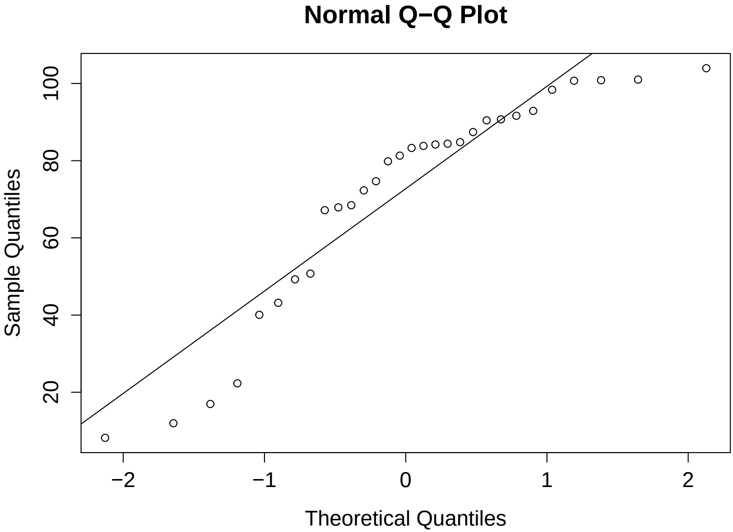
UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example of a Normal distribution



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Normal Quantile Plot of Course Score



Brown M&Ms

We took a sample of 100 m&ms candies and found 25 of them to be brown. What can we say about the proportion (percent) of m&ms that are brown?

$$\hat{p} = 0.25 = \frac{25}{100}$$

is a point estimate

population p,

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Intervals for Proportions

Before any inferences can be made about a proportions, certain conditions must be satisfied:

1. The sample must be an SRS from the population of interest.
2. The population must be at least 10 times the size of the sample.
3. The number of successes and number of failures must each be at least 10 (both $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$).

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Can we use the Confidence Interval?

We want to determine the proportion of M&Ms that are brown. In which of the following situations can we create a confidence interval.

1. We take a sample of 25 and found 8 to be brown.

   $8 < 10$.

2. We look at one bag approximately to have 50 M&Ms and notice that 14 are brown.

   SRS 10%

3. We want to determine the proportion of brown M&Ms in one small bag (approximately 50 M&Ms), we take a simple random sample of 25 from the bag a found 8 to be brown.

   10

4. We randomly select 100 M&Ms from the entire produced M&Ms and found 25 to be brown.

   Yes.

Determining the 99% Confidence Interval

Suppose we want a 99% confidence interval to estimate the proportion of M&Ms that are brown. For a random sample of 100 we found 25 to be brown.

1. Point estimate = $\hat{p} = 0.25$. $= \frac{25}{100}$
2. Confidence Level = $1 - \alpha = 0.99$, $\alpha/2 = 0.01/2 = 0.005$
3. Critical value = $z_{0.005} = 2.576$ In R: qnorm(1.99/2)
4. Standard error = $\sqrt{\frac{0.25(1-0.25)}{100}} = 0.0433$. $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
5. Margin of error = $2.576 \times 0.0433 = 0.1115$
6. Confidence Interval: $0.25 \pm 0.1115 = (0.1385, 0.3615)$
7. Interpretation: We are 99% confident that the percent of all M&Ms that are brown are between 13.85% and 36.15%.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Interval for Proportions, p

The $1 - \alpha$ confidence interval for proportions is:

$$\underset{\text{point estimate}}{\hat{p}} \pm \underbrace{z_{\alpha/2}}_{\text{critical value}} \underbrace{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}_{\text{st. error}}$$

Rcode

$\hat{p}$ $\pm$ C.V ~~qnorm~~

```
> 0.25+c(-1,1)*qnorm(1.99/2)*sqrt(.25*(1-.25)/100)
[1] 0.1384633 0.3615367
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Candy

pop. mean: μ $\quad \bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$ $\qquad \bar{x} \pm t_{(\frac{\alpha}{2}, df)} \cdot \frac{s}{\sqrt{n}}$

A bag of M&Ms was randomly selected from the grocery store shelf, and the color counts were:

Brown	14	Red	14	Yellow	5
Orange	7	Green	6	Blue	10

Determine a 96.5% confidence interval for the proportion of blue M&Ms. p

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$C = 0.965$

$$\hat{p}_{blue} = \frac{10}{14+14+5+7+6+10} = \frac{10}{56}$$

$n = 56$, $\quad z_{\frac{\alpha}{2}} = \text{qnorm}\left(\frac{1+0.965}{2}\right)$

UNIVERSITY of HOUSTON DEPARTMENT OF MATHEMATICS

Choosing Sample Size

- We use the margin of error to determine the sample size and solve for n.
- For the confidence interval for the mean it is:

$$n > \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2 \quad \text{— ①}$$

- For the confidence interval for proportion it is:

$$n > p^*(1 - p^*)\left(\frac{z_{\alpha/2}}{m}\right)^2 \quad \text{— ②}$$

Where $p^* = 0.5$ if not given a prior proportion.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example of Determining Sample Size

Mars Inc. claims that they produce M&Ms with the following distributions:

Brown	30%	Red	20%	Yellow	20%
Orange	10%	Green	10%	Blue	10%

How many M&Ms must be sampled to construct the 98% confidence interval for the proportion of red M&Ms in that bag if we want a margin of error of $\pm$ .15?

$$p^* = 0.20$$

$$n > p^*(1-p^*)\left(\frac{z_{\frac{\alpha}{2}}}{m}\right)^2$$

$$= 0.2\,(1-0.2) \times \left(\frac{\text{qnorm}\left(\frac{1.98}{2}\right)}{0.15}\right)^2$$

$$= 38.48\ldots \approx \boxed{39}$$

need at least 39

Another Example

Suppose that prior to conducting a coin-flipping experiment, we suspect that the coin is fair. How many times would we have to flip the coin in order to obtain a 96.5% confidence interval of width of at most .16 for the probability of flipping a head?

$$n > p^*(1-p^*) \cdot \left(\frac{z_{\frac{\alpha}{2}}}{m}\right)^2$$

$p^* = 0.5$ $\qquad z_{\frac{\alpha}{2}} = qnorm\left(\frac{1.965}{2}\right)$

width $= 2 \times ME$ $\qquad ME = \frac{\text{width}}{2} = 0.08$

$n > 173.64$

$n = 174$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Estimating the Variance

- A sample size of n is drawn from a Normal population with variance σ^2.
- The sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. We can obtain a value of s^2 from the sample.
- This computed sample variance is used as a point estimate of σ^2. Hence the statistic s^2 is an estimator of σ^2.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

The Distribution for Variance

- $\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ See section 6.7 or the extra notes on CASA.
- Where X^2 has the chi-square distribution with degrees of freedom $\nu = n - 1$.
- Find $P(X^2 \leq 15.033)$ with $\nu = 6$ using chi-square table and R.

```
> pchisq(15.033,6)
[1] 0.9799984
```

- Find c such that $P(X^2 \geq c) = 0.02$ with $\nu = 10$.

```
> qchisq(1-0.02,10)
[1] 21.16077
```

Setting up the Confidence Interval

- Thus: $P(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}) = 1 - \alpha$
- Similarly to the confidence intervals of μ and p we want to solve for σ^2.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Interval for σ^2

A $100(1-\alpha)\%$ confidence interval for the variance σ^2 of a normal population has lower limit

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}$$

2 3

and upper limit

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$

$\frac{20}{3}$ $\frac{20}{2}$

A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in the interval for σ^2, where $\alpha/2$ is the area in the upper tail of the chi-squre distribution.

Example

A coffee machine dispenses coffee into paper cups. Here are the amounts measured in a random sample of 20 cups.

$$9.9, 9.7, 10.0, 10.1, 9.9, 9.6, 9.8, 9.8, 10.0, 9.5,$$
$$9.7, 10.1, 9.9, 9.6, 10.2, 9.8, 10.0, 9.9, 9.5, 9.9$$

σ is NOT t-int

- Determine a 95% confidence interval for the mean amount of coffee dispensed from this machine.
- Determine a 95% confidence interval for the standard deviation of the amount of coffee dispensed from this machine.

R Code

$(n-1)$ s^2

```
> lcl=(length(coffee)-1)*var(coffee)/qchisq(0.025,(length(coffee)-1),
lower.tail = F)
> ucl=(length(coffee)-1)*var(coffee)/qchisq(0.025,(length(coffee)-1))
> c(lcl,ucl)
[1] 0.02281421 0.08415187
> sqrt(c(lcl,ucl))
[1] 0.1510437 0.2900894
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example College Placement Tests

A random sample of 20 students obtained a mean of $\bar{x} = 72$ and a variance of $s^2 = 16$ on a college placement test in mathematics. Assuming the scores to be normally distributed, construct a 98% confidence interval for the variance.


```
> C=0.98
> alpha=1-C
> alpha
[1] 0.02
> alpha/2
[1] 0.01
> lcl=(20-1)*16/qchisq(0.01,20-1,lower.tail = F)
> lcl
[1] 8.399909
> ucl=(20-1)*16/qchisq(0.01,20-1)
> ucl
[1] 39.82848
```

qchisq(1-0.01,20-1)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS