EWMBA 296 (Fall 2015)

Section 2

GSI: Fenella Carpena

October 29, 2015

Housekeeping: Announcements

- Take-Home Team Assignment # 1 due 11/2 at 11:59 PM
- Chapter 15, problem 52 (p. 382) and problem 54 (p. 383)
- Submit through bCourses – but if you are having technical issues with submission please feel free to email your submission to me (same deadline)
- **Please check bCourses > "People" tab > "Team Assignments" tab** to check you are assigned to the right group – important for your grades!

Agenda for Today

- Sampling Distribution of the Mean
- Control Limits
  - Practice Problems: Ch. 14, Q30(a) and Q26(a)
- Confidence Intervals
  - Practice Problem: Ch. 15, Q47
- Review 2014 Quiz 3 (to the extent possible)
- Margin of error (not covering today due to time constraints, but see textbook, lecture and section notes)

Sampling Distribution of the Mean

- The **sample mean**, denoted $\overline{X}$, is an average of n individual measurements from a population, where n is the sample size.
- $\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}$
- Each X_i in the right hand side above is an individual measurement drawn from the same population, so all X_i's have the same mean μ and the same sd σ.
- **The sample mean $\overline{X}$ has a sampling distribution.**
- Even though we may not know what the distribution of $\overline{X}$ is, the **Central Limit Theorem** tells us that for sufficiently large n, $\overline{X}$ is approximately normally distributed with mean μ and variance σ^2/n. That is,

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- Why do we care? If we know that $\overline{X}$ is approx. normally distributed, we can use the Z-table.

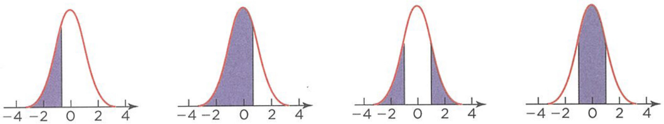
Control Limits

- A **control limit** is a range $\mu - L$ and $\mu + L$ such that if $\overline{X}$ lies within this range, then the process should continue. Otherwise, we should stop the process.
- "Recipe" for constructing control limits:
  - Step 0: Check that the sample size conditions are met: (1) $n > 10(K_3)^2$ and (2) $n > 10 \cdot |K_4|$
  - Step 1: Specify the parameters of the process when it is working normally, μ and σ. Also note sample size n.
  - Step 2: Select α, your tolerance for a Type I error rate.
  - Step 3: Find $z_{\alpha/2}$ in the Z-Table. You can find it by looking for the value of $z_{\alpha/2}$ which satisfies:
    - $P(Z \leq -z_{\alpha/2}) = \alpha/2$, or
    - $P(Z \geq z_{\alpha/2}) = \alpha/2$, or
    - $P(|Z| > z_{\alpha/2}) = \alpha$, or
    - $P(|Z| \leq z_{\alpha/2}) = 1 - \alpha$
    - All of the above methods should give you the same answer.
  - Step 4: Construct the control limits using the formula

$$\underbrace{\mu}_{\text{Step 1}} \pm \underbrace{z_{\alpha/2}}_{\text{Step 3}} * \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{Step 1}}$$

Practice Problem: Stine & Foster Ch 14, Q30(a)

Find control limits if the design of the production process sets $\alpha = 0.01$ with the following parameters $\mu = 1000$, $\sigma = 20$, $n = 25$ cases per batch. Assume that the sample size conditions have been verified.



z	$P(Z \le -z)$	$P(Z \le z)$	$P(\lvert Z\rvert > z)$	$P(\lvert Z\rvert \le z)$
1.2816	0.10	0.90	0.20	0.80
1.3408	0.09	0.91	0.18	0.82
1.4051	0.08	0.92	0.16	0.84
1.4758	0.07	0.93	0.14	0.86
1.5548	0.06	0.94	0.12	0.88
1.6449	0.05	0.95	0.10	0.90
1.7507	0.04	0.96	0.08	0.92
1.8808	0.03	0.97	0.06	0.94
1.9600	0.025	0.975	0.05	0.95
2.0537	0.02	0.98	0.04	0.96
2.3263	0.01	0.99	0.02	0.98
2.5758	0.005	0.995	0.01	0.99
2.8070	0.0025	0.9975	0.005	0.995
3.0902	0.001	0.999	0.002	0.998
3.2905	0.0005	0.9995	0.001	0.999
3.7190	0.0001	0.9999	0.0002	0.9998
3.8906	0.00005	0.99995	0.0001	0.9999
4.2649	0.00001	0.99999	0.00002	0.99998
4.4172	0.000005	0.999995	0.00001	0.99999

Practice Problem: Stine & Foster Ch 14, Q26(a)

A bottler carefully weights bottles coming off its production line to check that the system is filling the bottles with the correct amount of beverage. By design, the system is set to slightly overfill the bottles to allow for random variation in the packaging. The content weight is designed to be 1020 g with standard deviation 8 g. Assume that the weights are normally distributed.

(a) A proposed system shuts down the facility if the contents of a bottle weighs less than 1000 g. What is the chance of a Type I error in this system?

Confidence Intervals

- A **confidence interval** (CI) is a range of plausible values of a parameter based on a sample from the population.
- **Constructing CIs relies on the sampling distribution.**
- **Why do we care about CIs?** Typically, the population parameters (e.g., p and μ) are *unknown*. We want to estimate these population parameters using data from the sample, and the CI provides us with a range of plausible values.
- **What is the difference between confidence intervals and control limits?** In control limits, we know the value of μ, and the control limits $(\mu + L, \mu - L)$ tells us the interval where $\overline{X}$ should be found. In CIs, the set-up is reversed: we do *not* know μ, but we use $\overline{X}$ and $se(\overline{X})$ to determine a range of plausible values for μ.
- **How do we interpret CIs?** "We are 95% confident that the true population parameter is between ... and ..."

Confidence Intervals

	Proportion	**Mean**
Population parameter	p	μ
Sample statistic	$\widehat{p}$	$\overline{X}$
Conditions required	(1) Sample is from a SRS (2) Sample size conditions (SSC): $n\widehat{p}$ and $n(1-\widehat{p}) > 10$	(1) Sample is from a SRS (2) Sample size conditions (SSC): $n > 10(K_3)^2$ (squared skewness) and $n > 10 \cdot \lvert K_4 \rvert$ (absolute value of kurtosis)
Sampling distribution (conditions are *not* satisifed)	Based on binomial distribution	Student's t distribution (use T-Table)
Sampling distribution (conditions are satisfied)	$\widehat{p} \sim \mathcal{N}(p, \frac{p(1-p)}{n})$ (use Z-Table)	$\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ (use Z-Table)
Estimated standard error (se)	$\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$	$s/\sqrt{n}$

Confidence Intervals

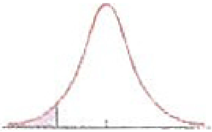
A “recipe” for constructing CIs (for both means and proportions)

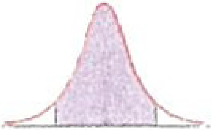
- **Step 0:** Check that the conditions
- **Step 1:** Find the sample statistic ($\widehat{p}$ or $\overline{X}$).
- **Step 2:** Find the “critical value” $z_{\alpha/2}$ or $t_{\alpha/2,n-1}$ depending on whether conditions from step 0 hold. Note: $\alpha = 1-$(confidence level).
- **Step 3:** Calculate the estimated standard error (se).
- **Step 4:** Construct the confidence interval using the formula:

$$\underbrace{\textit{sample statistic}}_{\text{Step 1}} \pm \underbrace{\textit{critical value}}_{\text{Step 2}} * \underbrace{\textit{se(sample statistic)}}_{\text{Step 3}}$$

$df = 150$

t	$P(T_{150} \le -t)$	$P(-t \le T_{150} \le t)$
1.287	0.1	0.8
1.655	0.05	0.9
1.976	0.025	0.95
2.351	0.01	0.98
2.609	0.005	0.99
3.145	0.001	0.998
3.357	0.0005	0.999
3.998	0.00005	0.9999





Practice Problem: Stine&Foster Ch 15, Q47

Direct mail advertisers send solicitations (junk mail) to thousands of potential customers hoping that some will buy the product. The response rate is usually quite low. Suppose a company wants to test the response to a new flyer and sends it to 1,000 randomly selected people. The company gets orders from 123 of the recipients and decides to do a mass mailing to everyone on its mailing list of over 200,000. Create a 95% confidence interval for the percentage of those people who will order something.

2014 Quiz 3 “Scottish Independence Referendum”

(1) Consider a recent poll for the 2014 Scottish Independence Referendum vote (“Yes” is for Scotland being an independent country, “No” otherwise.). Is the following statement true or false? “The individual poll responses from potential voters are normally distributed.”

2014 Quiz 3 "Scottish Independence Referendum"

(2) Suppose a junior analyst tells you that because the sample respondents is a representative sample of the population, then the poll will exactly inform you of the voting outcome. Provide 2 reasons why the polling data may not exactly match the voting outcome.

2014 Quiz 3 "Scottish Independence Referendum"

(3) Suppose that the proportion who voted "Yes" in the most recent poll is 40%. Ignore the "Don't Know" votes, so the proportion "No" is 60%. What is the standard error of the sampling distribution? Assume that the size of the sample was 625.

2014 Quiz 3 "Scottish Independence Referendum"

(4) Suppose you are extremely concerned about the possibility of a "Yes" vote. Your data analytics team suggests that you estimate a 99% CI for the proportion of the population voting "Yes." Please calculate this interval.

2014 Quiz 3 "Scottish Independence Referendum"

(5) You are the CEO of a company which heavily trades with partners based in the UK. Your CFO mentions that she was approached by an insurance company offering full coverage insurance against the potential loss from British pound volatility in the event of a Scottish "Yes" vote. You are considering whether or not to buy the insurance. What are the Type I and Type II errors in this context?

Final Points

- All solutions to Stine & Foster problems covered today are in `Section2_Notes.pdf`
- Quiz 2 Logistics
  - Covers Ch. 14.1, 14.2 + Ch. 15 (Lectures 3 and 4, from last Saturday)
  - You can bring a single-sided, 8.5 x 11 sheet of notes and use a calculator.
- Tip: I would recommend memorizing (or writing on your sheet) the z values from the Z-table of the 90%, 95%, and 99% CI, since these are the most common confidence levels.
  - 90% CI: $z_{0.10/2} = 1.6449$
  - 95% CI: $z_{0.05/2} = 1.9600$
  - 99% CI: $z_{0.01/2} = 2.5758$
- Lecture this Saturday (10/24) will cover Chapter 16 (Statistical Tests) and Chapter 17 Section 1, 2, and 4 (Comparison)
- Please read Hawthorne "mini-case" for this Saturday's class (see Reed's email)