جامعة كارنيجي ميلون في قطر
Carnegie Mellon Qatar

Probability and Statistics for Business Applications

Chapter 14: Sampling Variation and Quality

Fall 2012

Motivation

Example (HALT)

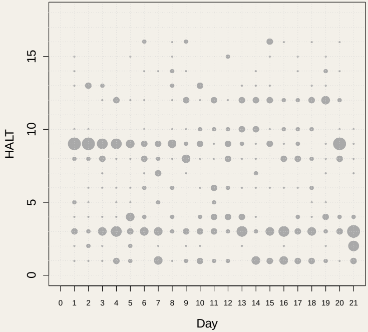
A manufacturer of GPS chips selects samples for highly accelerated life testing (HALT). How should managers monitor these tests to ensure proper functioning of the production process?

- Use control charts. Control charts determine whether a process is functioning as designed on the basis of properties of a sample of the production.
- Balance the two errors possible in all statistical decisions
  - Stopping a properly functioning process
  - Failing to detect a malfunctioning process

Motivation

Example (HALT - ctd)

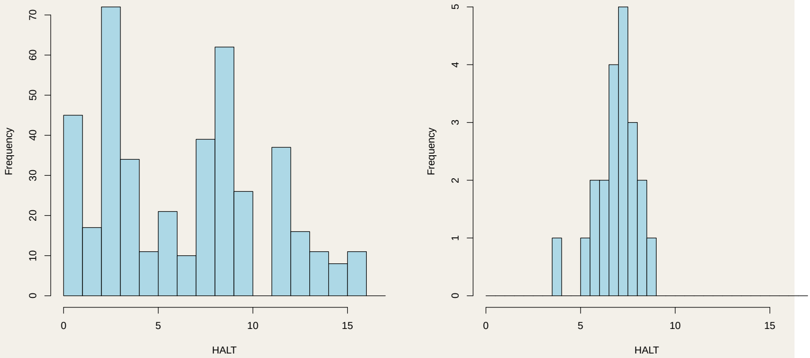
Random Variation or Change in Process?



Engineers process 20 chips through HALT each day and even when functioning properly there is variation among HALT scores (recorded as number of tests passed). Thus, we need to understand what to expect for HALT scores (e.g., on average chips should pass $\mu = 7$ tests with a standard deviation $\sigma = 4$).

Sampling Distribution of the Mean

Random Variation or Change in Process?



Definition (Sampling Distribution)

The probability distribution that describes how a statistic, such as the mean, varies from sample to sample.

Sampling Distribution of the Mean

Benefits of Averaging

- The sample-to-sample variance among mean HALT scores is smaller than the variance among individual HALT scores.
- The distribution of mean HALT scores appears more bell shaped than the distribution of individual HALT scores.

Normal Models

- Sample means are normally distributed if the individual values are normally distributed.
- Sample means are normally distributed because of the *Central Limit Theorem* (when sample size condition is satisfied).

The Sampling Distribution of the mean

When the population is normally distributed

Regardless of the sample size, the sampling distribution of the mean will be normally distributed, with

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

and

$$Var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

where μ is the population mean, σ^2 is the population variance and n is the sample size.

The Sampling Distribution of the mean

When the population is normally distributed

Example

For a random variable that is normally distributed, with $\mu = 80$ and $\sigma^2 = 100$, determine the probability that a simple random sample of 25 items will have a mean that is

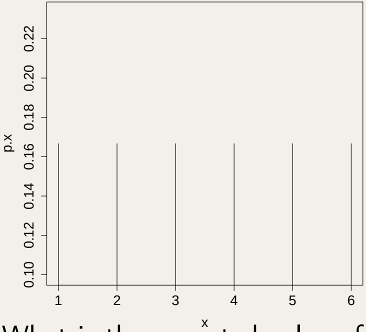
- greater than 78
- between 79 and 85
- less than 85

Sampling Distribution of the Mean

Central Limit Theorem

Example (Rolling one dice)

There are 6 possible outcomes when we roll one dice and record the up-face. The probability mass function (p.m.f) is sketched in the following figure.



x	$p(x)$
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$

What is the expected value of X, $\mu =$

Sampling Distribution of the Mean

Central Limit Theorem

Example (Rolling one dice -ctd)

Consider all possible samples of size 2. The probability of selecting each sample appears below

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Sampling Distribution of the Mean

Central Limit Theorem

Example (Rolling one dice -ctd)

The mean of each sample

	1	2	3	4	5	6
1	1.0	1.5	2.0	2.5	3.0	3.5
2	1.5	2.0	2.5	3.0	3.5	4.0
3	2.0	2.5	3.0	3.5	4.0	4.5
4	2.5	3.0	3.5	4.0	4.5	5.0
5	3.0	3.5	4.0	4.5	5.0	5.5
6	3.5	4.0	4.5	5.0	5.5	6.0

Let's fill out the following table

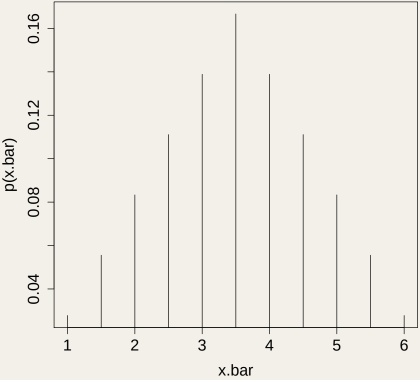
$\bar{x}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
Counts	1					6					1
$p(\bar{x})$	$\frac{1}{36}$					$\frac{6}{36}$					$\frac{1}{36}$

Sampling Distribution of the Mean

Central Limit Theorem

Example (Rolling one dice -ctd)

The probability mass function of the mean $\bar{X}$ is



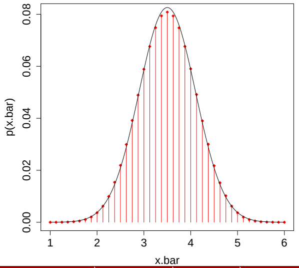
$\mu_{\bar{X}} = 3.5$ and $\sigma^2_{\bar{X}} = \frac{52.5}{36}$

Sampling Distribution of the Mean

Central Limit Theorem

If we were to take a great many samples of size n from the same population, we would end up with a great many different values for the sample mean. The resulting collection of sample means could then be viewed as a new random variable with its own mean and standard deaviation. The probability distribution of these sample means is called the distribution of sample means, or the sampling distribution of the mean.

For example, let $n = 8$. The probability mass function of the mean $\bar{X}$ is sketch below.



The Sampling Distribution of the mean

When the population is NOT normally distributed

The assumption of normality for a population is not always realistic, since in many cases the population is either not normally distributed or we have no knowledge about its actual distribution. However, provided that the sample size is large, the sample distribution of the **mean** can still be assumed to be normal.

Central Limit Theorem

For a large, simple random sample from a population that is not normally distributed, the sampling distribution of the mean will be approximately normally distributed, with

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \text{ and } \mathrm{Var}(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

where μ is the population mean, σ^2 is the population variance and n is the sample size.

The Sampling Distribution of the mean

Central Limit theorem

Sample Size Condition:

A normal model provides an accurate approximation to the sampling distribution of $\bar{X}$ if the sample size n is larger than 10 times the squared skewness and larger than 10 times the absolute value of the kurtosis.

$$n > 10 \cdot K_3^2 \text{ and } n > 10 \cdot |K_4|$$

Definition (Standard Error of the Mean)

A measure of the sample-to-sample variability in sample means.

The standard error is proportional to σ. As data become more variable, averages become more variable. The standard error is inversely proportional to the square root of n $(\sigma/\sqrt{n})$. The larger the sample size, the smaller the sampling variation of the averages.

The Sampling Distribution of the mean

When the population is NOT normally distributed

Example

An airline has found the luggage weight for individual air travelers on its transAtlantic route to have a mean of 80 pounds and a standard deviation of 20 pounds. The plane is consistenly fully booked and holds 100 passengers. The pilot insist on loading an extra 500 pounds of fuel whenever the total luggage weight exceeds 8300 pounds.

On what percentage of the flights will she end up having the extra fuel loaded?

The Sampling Distribution of the mean

Example (How many people in a car?)

A study of rush-hour traffic in Doha counts the number of people in each car crossing the "tilted roundabout". Suppose that this count has a mean of 1.5 and standard deviation 0.75 in the population of all cars that enter at this roundabout during rush hour.

- Could the exact distribution of the count be Normal?
- Traffic engineers estimate that the capacity of the roundabout is 700 cars per hour. What is the approximate distribution of the mean number of persons $\bar{X}$ in 700 randomly selected cars at this roundabout?
- What is the probability that 700 cars will carry more than 1075 people?

Control Limits

Definition

Boundaries that determine whether a process is out of control or should be allowed to continue.

Determining Control Limits

- Symmetric interval denoted as

$$\mu - L \leq \bar{X} \leq \mu + L$$

- Upper Control Limit (UCL) is $\mu + L$
- Lower Control Limit (LCL) is $\mu - L$

Control Limits

Type I and Type II Errors

Type I Error: the mistake of taking action when no action is needed.

Type II Error: the mistake of failing to take action when needed.

	Supervisor Chooses to Continue	Supervisor Chooses to Shut Down
Working as designed	✓	$\times_1$
Not working as designed	$\times_2$	✓

✓ represents a correct decision, $\times_1$ is the Type I error, and $\times_2$ is the Type II error.

Control Limits

Setting the Control Limits

- Specify the chance for a Type I error
- Based on parameters of the process

Example (HALT)

If production is shut down when the mean HALT score is less than 6 or more than 8, what is the chance of Type I error?

$$\begin{aligned}\Pr(\bar{X} < 6 \text{ or } \bar{X} > 8) &= 1 - \Pr(6 \leq \bar{X} \leq 8) \\ &= 0.27\end{aligned}$$

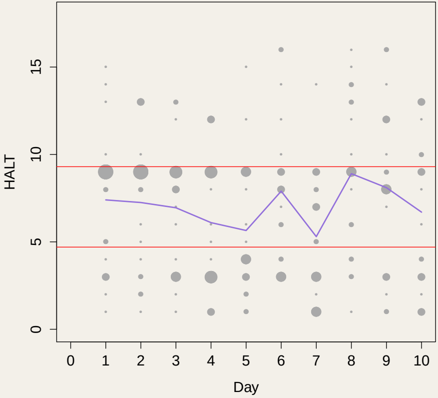
Control Limits

Balancing Type I and Type II Errors

- Wide control limits reduce the chance for a Type I error
- Narrow control limits reduce the chance for a Type II error
- Cannot simultaneously reduce the chances of both
- Control limits are determined by focusing on Type I errors (following convention in statistics).
- The chance of making a Type I error is typically set at 5% or 1%.

Using a Control Chart

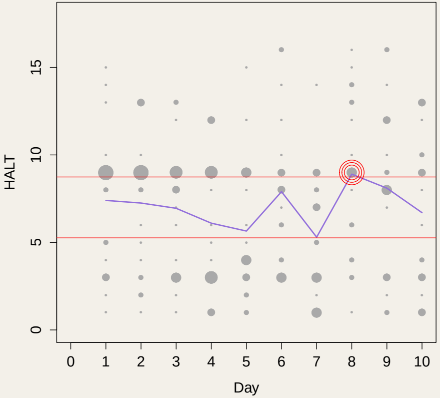
X Bar Chart: Tracks the Mean of Process



Shown are 99% control limits; process is in control

Using a Control Chart

X Bar Chart: Tracks the Mean of Process



Shown are 95% control limits; process incorrectly indicates that the process is out of control

Using a Control Chart

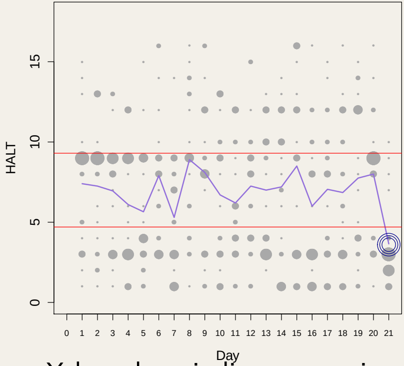
Repeated Testing

- The chance for a Type I error increases over consecutive points. (e.g., a 5% chance of a Type I error in any one day results in a 40% over 10 days)[1]
- Repeated testing eventually signals a problem.
- Typically the chance for Type I error is set to 0.0027 for any one point. This is the probability of a normal random variable falling more than three standard deviations from its mean. [2]

[1] $1 - 0.95^{10} \approx 0.401$
[2] $1 - 0.9973^{10} \approx 0.027$

Using a Control Chart

Recognizing a Problem



- The previous X-bar chart indicates a point outside the lower control limit.
- This can either be a Type I error or a real process problem. To verify the latter, management must be able to identify the problem.

Using a Control Chart

Control Limits For the X-Bar Chart

The $100(1-\alpha)\%$ control limits for monitoring averages of a sample of n measurements from a process with mean μ and standard deviation σ are

$$\mu \pm z_{\alpha/2}\sigma/\sqrt{n}$$

The multiplier $z_{\alpha/2}$ controls α, the chance of a Type I error. For example, $z_{0.025} = 1.96$ and $z_{0.005} = 2.58$.

Control Charts for Variation

Control charts can be used to monitor any sample statistic by comparing it to the corresponding parameter of the process. For a manufacturing process, it is also important to monitor the variability of a process.

Monitoring Process Variability

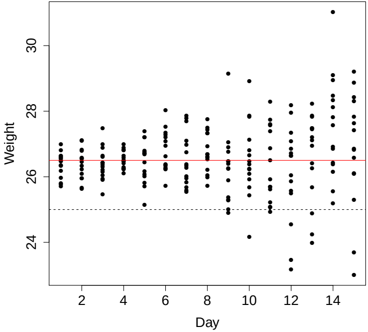
S-chart: tracks the standard deviation s from sample to sample.

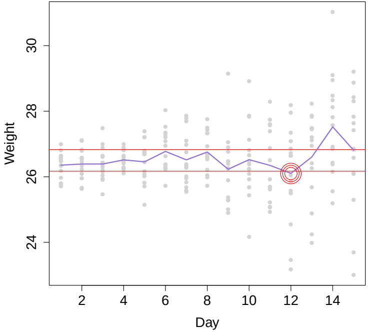
R-chart: tracks the range rather than the standard deviation from sample to sample.

Control Charts for Variation

Example (Package weights)

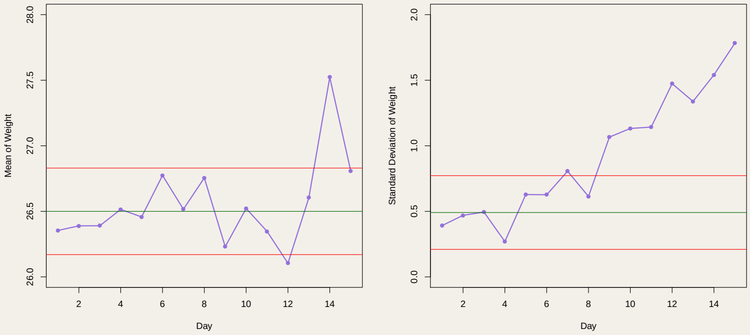
A food processor weighs samples of 15 packages of frozen food from each day's production. The packages are labeled to weigh 25 ounces. To accommodate for variation, the packaging system is designed to put 26.5 ounces on average into each package ($\sigma = 0.5$ ounces).





Control Charts for Variation

Example (Package weights-ctd)



The S-chart detects the problem more quickly than the $\bar{X}$ chart.

4M MONITORING A CALL CENTER

Motivation

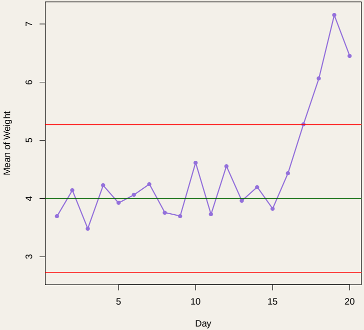
A bank wants a system for tracking calls related to its Internet bill-paying service. They are willing to monitor 50 calls per day.

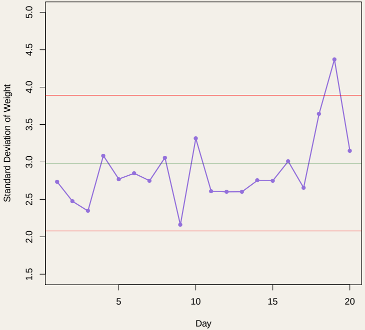
Method

Specify the parameters of the process based on past data. Check the sample size condition to verify appropriateness of the normal model. Calls average $\mu = 4$ min. with $s = 3$ min. Place limits three standard errors from the parameter

4M MONITORING A CALL CENTER

Mechanics





Message

The length of time required for the calls to this help line has changed. The average length has increased and the lengths have become more variable. Management should identify the reasons for this change.

Best Practices & Pitfalls

Best Practices

- Think hard about which attribute of the process to monitor.
- Use both X-bar charts and S-charts to monitor a process.
- Set the control limits from process characteristics, not data.
- Set the control limits before looking at the data.
- Carefully check before applying control limits to small samples.
- Recognize that control charts eventually signal a problem.

Pitfalls

- Do not concentrate on one error while ignoring the other.
- Do not assume that the process has failed if a value appears outside the control limits.