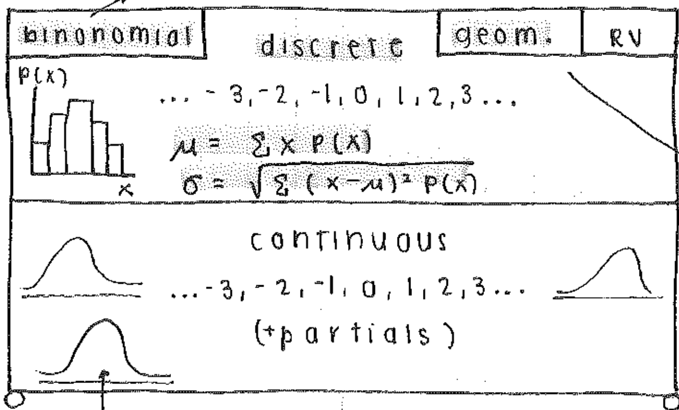
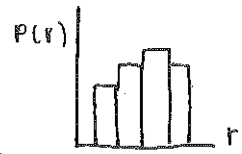
normal distribution

success, fail, set probability, independent, known # of trials

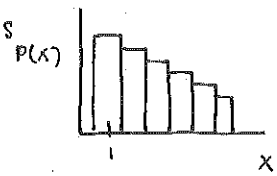




$\mu = np$
$\sigma = \sqrt{\mu q}$

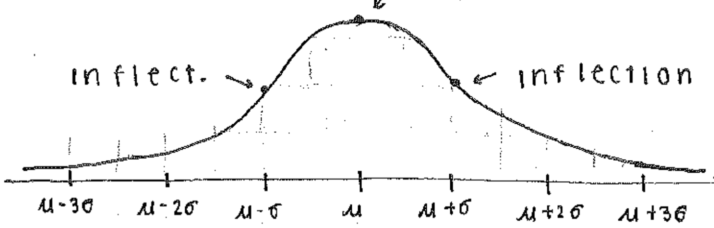
S/F, same probability, indep.
n unknown, go until first success

$\mu = 1/p$
$\sigma = \sqrt{q/p}$



bell curve

- normal curve: highest point centered over μ



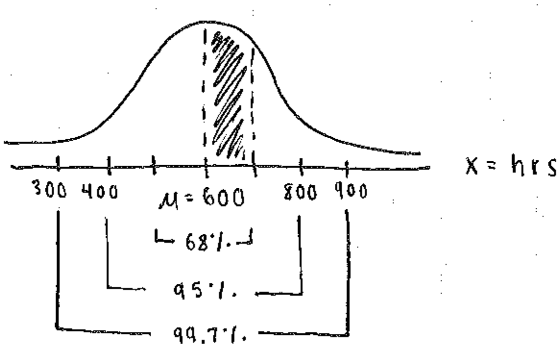
approaches but never crosses

mode = median
50th percentile

affected by mean and size of standard deviation
- mean affects location on # line
- stdev affects width

- empirical rule: area under normal curve
  ↳ probability
- 68% of area under curve falls between $\mu - 1\sigma$ and $\mu + 1\sigma$
- 95% between $\mu - 2\sigma$ and $\mu + 2\sigma$
- 99.7% between $\mu - 3\sigma$ and $\mu + 3\sigma$

ex. 1 sunshine radio



remember!
pic
p(x =)
. _ _ _ _
∃x a ...

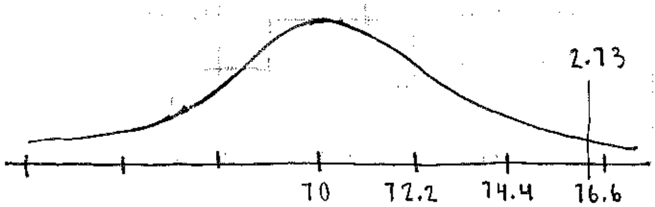
ex. micek's basketball scores

$z = \frac{x - \mu}{\sigma}$

chemistry: $x = 76,\ \mu = 70,\ \sigma = 2.2$

$z = \frac{76 - 70}{2.2} = 2.73$

$z = 2.73$ standard deviations

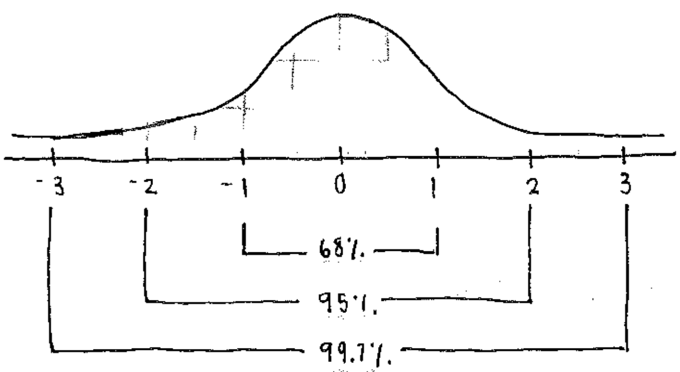


x = chem points

literature: $x = 89;\ \mu = 82,\ \sigma = 3.1$

$z = \frac{89 - 82}{3.1} = 2.26$

- with z-scores, you can compare different populations
- standard normal distribution: $\mu = 0,\ \sigma = 1$



z = SNC standard deviations

ex. pizza and cheese

- raw score $x = z\sigma + \mu$ ← mean; z score; stdev

ex. $z = 3.15 \quad \sigma = 10 \quad \mu = 75$

$3.15(10) + 75 = x$

$-1.7 = \frac{x - 75}{10} \quad x = 58$

7.2 #5 $\mu = 30,\ \sigma = 5$

a) z-score of $x = 25$

$z = \frac{25 - 30}{5} = -1$ stdev

b) $x = 42$

$z = \frac{42 - 30}{5} = \frac{12}{5} = 2.4$ stdev

c) $x = (-2)(5) + 30 = 20$

d) $x = 1.3(5) + 30 = 36.5$

7.5 (2-5, 8, 9, 11, 12, 16-28 E)

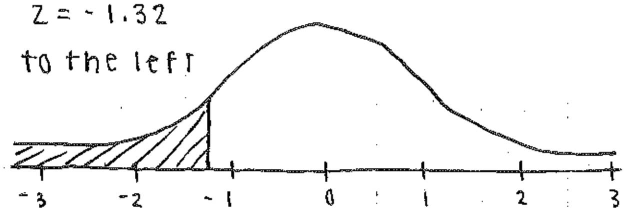
MLK edpuzzle

against all odds: normal calculations

- z-score is a standardized value for comparing standard deviations
  $\mu = 63.8$, $x = 70$, $\sigma = 4.2$
  $z = \frac{70 - 63.8}{4.2} = 1.48$ stdev above normal
- to find probability of being under a part of the curve, use z-table
- z = .98 for men z = 1.48 for women
  - .8635 of the male population don't make the cut
- can find percentages between 2 points
  - 93.06% of women don't make the cut
  - 56.36% of women are shorter than pardis sabeti (z=.16)
  - subtract percents, NOT z-scores
    ↳ 93.06 − 56.36 = 36.7% of the population
- pg 282 → waves
  - z-table tells you area to the left of z
  - for area to the right, do 1 − (value given by z-table)
    ↳ or use symmetry
  - for area between, do (area left of z_2) − (area left of z_1)

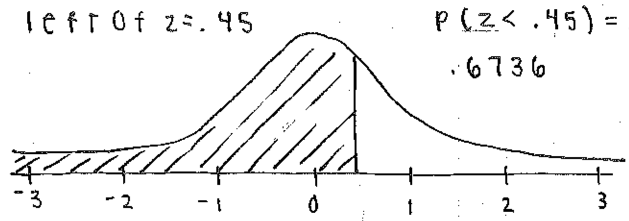
7.2 # 15 − 29 odd

15. Z = −1.32
to the left

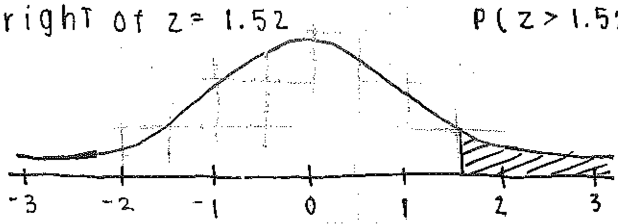


.0934 = P(Z < −1.32)

17. left of z = .45 P(z < .45) =
.6736



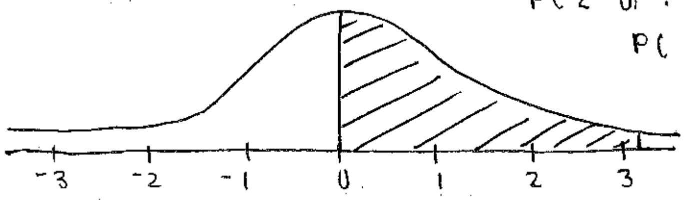
19. right of z = 1.52 P(z > 1.52) = .0643



23. between $z = 0$ and $z = 3.18$

$P(z=0) = .5000$ $P(z = 3.18) = .9993$

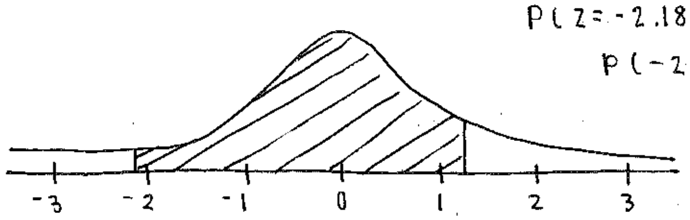
$P(0 < z < 3.18) = .4993$



25. between $z = -2.18$ and $z = 1.34$

$P(z = -2.18) = .0146$ $P(z = 1.34) = .9099$

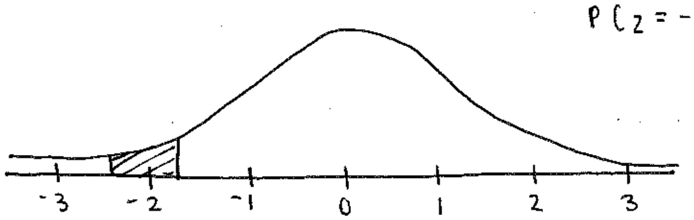
$P(-2.18 < z < 1.34) = .8953$



29. between $z = -2.42$ and $z = -1.77$

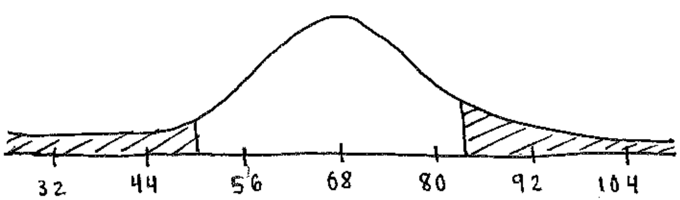
$P(z = -2.42) = .0078$ $P(z = -1.77) = .0384$

$P(-2.42 < z < -1.77) = .0306$



ex. tuna problem

$\mu = 68$, $\sigma = 12$



x = lbs

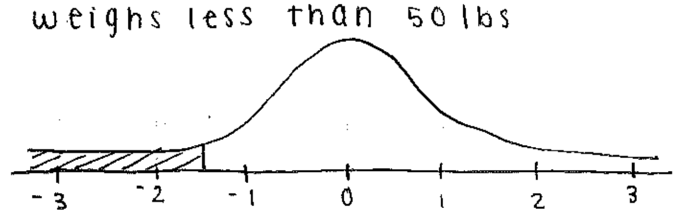
A. fish weighs less than 50 pounds

$x < 50$

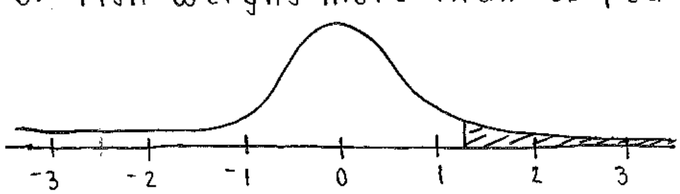
$z = \frac{x - \mu}{\sigma} = \frac{50 - 68}{12} = -1.5$

$P(z < -1.5) = .0668$

$\exists_x$ a 6.68% chance that a fish weighs less than 50 lbs



B. fish weighs more than 83 pounds



$z = \frac{83 - 68}{12} = 1.25$

$P(z = 1.25) = .8944$ $1 - .8944 = .1056$

$P(z > 1.25) = .1056$

$P(x > 83) = .1056 \rightarrow \exists_x$ a 10.56% chance that the fish weighs more than 83 lbs

C. between 50 and 80 lbs

NEVER use normalpdf

(next page)

- using calculator: (2nd) (vars) → (2)

lower = 50, upper = 83

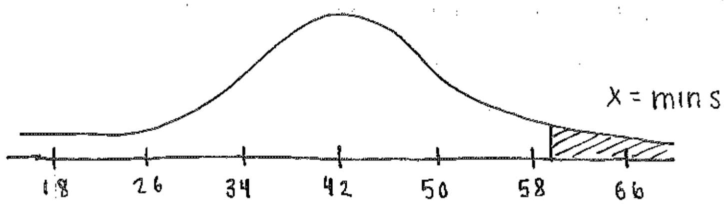
normalcdf (50, 83, 68, 12) = .8275 ← NOT notation

P(50 < x < 83) = .8275 → $\exists_x$ an 82.75% chance that a fish is

ex. Flight for Life helicopter

$\mu = 42$ $\sigma = 8$

c. prob. that response time > 60 min → choose something far out



lower: 60

upper: 200

$\mu = 42$

$\sigma = 8$

normalcdf = .0122 P(x > 60) = .0122

$\exists_x$ a 1.2224% chance that response time ≥ 60 min

7.3 (1-6, 15, 16, 22-25, 28)

pg 259 #9

50% of all inmates are serving time for drug dealing = p

n = 16

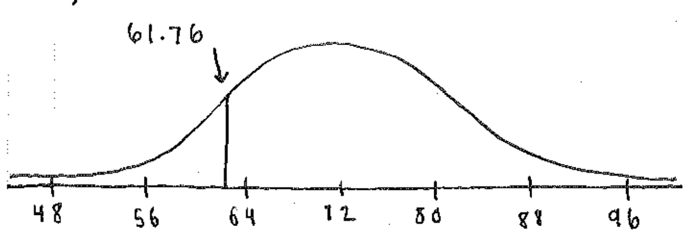
a) P(r ≥ 12)

binomcdf (16, .5, 11) 1 - .9616 = .0384

P(r ≥ 12) = .0384 → $\exists_x$ a 3.84% chance that 12+ inmates are in for drug dealing

ex. malaria prevention pill

$\mu = 72$ hours $\sigma = 8$ hours



a) fewer than 10% unprotected

find z-value where 10% to the left

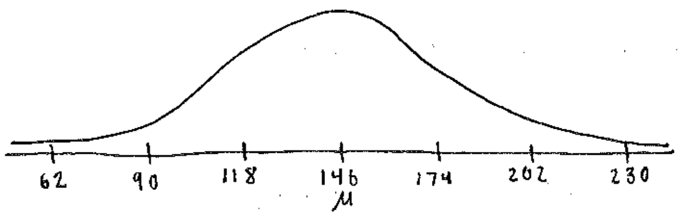
P(z < -1.28) = .1003

$x = (-1.28)(8) + 72 = 61.76$ hours

Jane Doe v. McDonald's

- μ = 146 pounds σ = 28lbs x = 397 lbs

$$z = \frac{397-146}{28} = 8.96 \text{ stdev}$$



Ms. Doe was literally 8.96 stdev from average weight; chair weight capacity does support most weights, but the probability of being this heavy is so rare

in an establishment like Mcdonalds

does data have a normal distribution?

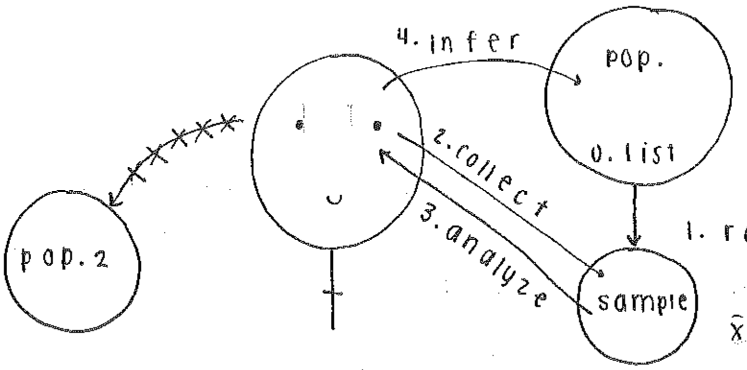
1. if histogram is relatively normal, distribution is normal
   - but if class width is different, it'll look different
2. outliers
3. pearson's index: $\frac{3(\bar{x} - \text{median})}{s}$
   - if index is >1 or <-1, it's skewed
4. Normal quantile plot

★ check for normalcy on AP exam! ★

against all odds: is it normal?

- normal quantile plot = linear, then data is normal
  - in the farm, the NQP isn't linear at the left, showing that more eggs are smaller than normally expected
  - concave down ≈ skewed right
  - concave up ≈ skewed left

7.4 sampling distributions



μ, σ, σ^2, π, $\pi_1 - \pi_2$

1. random representative sample

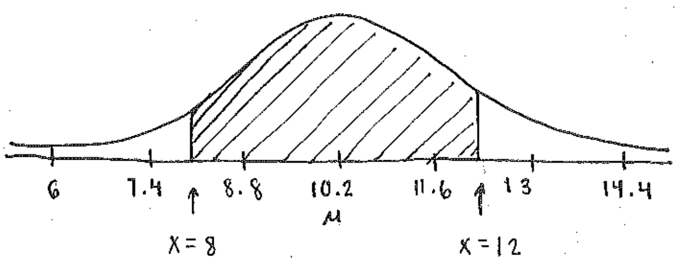
$\bar{x}$, s, s^2, $\hat{p}$, $\hat{p}_1 - \hat{p}_2$

- statistic is a descriptive measure of a sample
- parameter is a descriptive measure of a population
- we can use stats to infer about populations
  - confidence intervals (estimation)
  - hypothesis testing (decision)
- $\bar{x}$ distribution is average of a group → sampling distribution
  - normal curve just smaller stdev

fish in pond

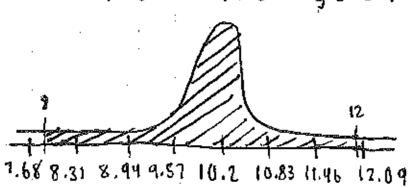
$\mu = 10.2"$ $\sigma = 1.4"$ $P(8 < x < 12)$ normalcdf(8, 12, 10.2, 1.4) = .8427

$P(8 < x < 12) = .8427$ → ∃x an 84.27% chance that a fish in the pond is between 8 and 12 inches

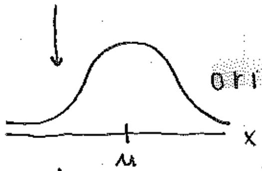


x = inches

P(8 < average of 5 < 12) = ?



original data distribution

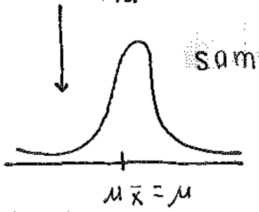


original parent

when $\downarrow \sigma_{\bar{x}}$, P(middle) ↑ and P(tails) ↓

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.4}{\sqrt{5}} = .6261$$

sampling distribution



normalcdf(8, 12, 10.2, $1.4/\sqrt{5}$) = .9978 → ∃x a 99.78% chance that the avg length of the 5 fish will be between 8-12"

$\bar{x}$ = average of a group

Theorem 7.1

★ have to start with a normal curve but n can be small or big ★

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

ex. heights of men $\mu = 68$ $\sigma = 3$ in

a) $P(67 < x < 69)$

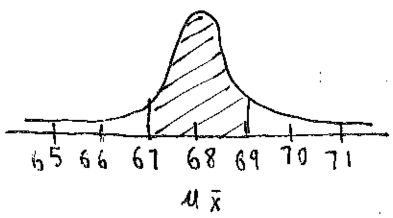
normalcdf(67, 69, 68, 3) = .2611

b) $P(67 < \bar{x}\ 69)$ 9 men selected

$\mu = 68$ in

$\sigma = \frac{\sigma}{\sqrt{9}} = \frac{3}{3} = 1$

$P(67 < \bar{x} < 69) = .6827 \rightarrow \exists_x$ a 68.27% chance that the average height of 9 18-year-old men is between 67 and 69 inches



$\bar{x}$ = average height of in inches of 9 men

what if the parent distribution wasn't normal?

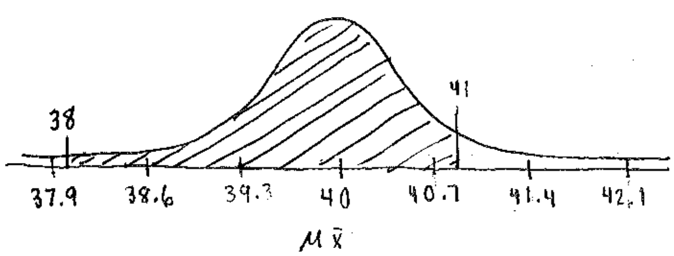
- theorem 7.2: the sampling distribution of non-normal curves will still be normal
  - $\mu_{\bar{x}} = \mu$
  - $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ "The Central Limit Theorem"
  - ★ any beginning or original distribution but n must be large★ ($n \geq 30$)

ex. professor + haircut

$\mu = 40$ min $\sigma = 4.2$ min $n = 36$ $P(38 < \bar{x} < 41)$

not normal, but $n \geq 30$

$n = 36 > 30$, ∴ central limit theorem invoked



$\bar{x}$ = average time of 36 visits

Pennies	Age	
1.	1	17
2.	3	1
3.	8	1
4.	8	1
5.	1	1
6.	1	5
7.	2	1
8.	5	5
9.	1	1
10.	1	1
11.	17	$\bar{x}$ = 3.4 years
12.	38	
13.	2	
14.	1	17
15.	34	1
16.	1	5
17.	1	1
18.	1	8
19.	5	17
20.	1	1
21.	1	1
22.	1	5
23.	1	8
24.	1	$\bar{x}$ = 6.4 years
25.	5	

distribution of pennies: skewed right
dimes: mound-shaped

against all odds: sample mean and control charts

- roulette → 36, 0, and 00; 38 numbers

  find mean of random variable (expected variable)

  \$1 bet

  $p(\text{win}) = 18/38 = .473$

  $p(\text{lose}) = 20/38 = .526$

  $p(\text{lose 13 times}) = p(\text{lose and lose and lose and} \ldots) = (.526)^{13} = .02367$.

  mean (expected value) = -5.3 cents

  - for every \$1, you can expect to lose 5.3 cents

  more bets = more likely to lose

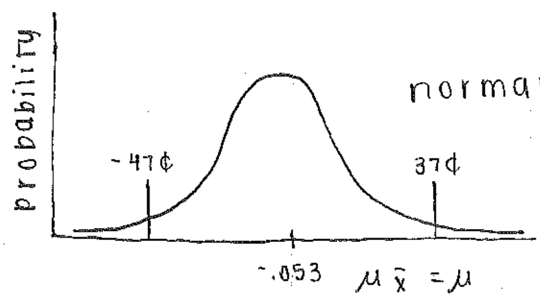
  50/50 chance when they first walk in

  stdev of a single bet $\sigma^2 = .997$, $\sigma = .998$ $\mu = -5.3$ per \$1 bet

  fifty \$1 bets

  mean winning = 10/50 = 20¢ = $\bar{x}$

  ↳ varies from sample to sample



mean of 50 \$1 bets ($\bar{x}$)

normal distribution

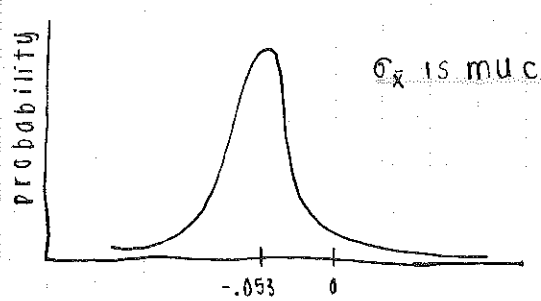
$n = 50$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.998}{\sqrt{50}} = .14$ (14 cents)

3 stdev = 42¢

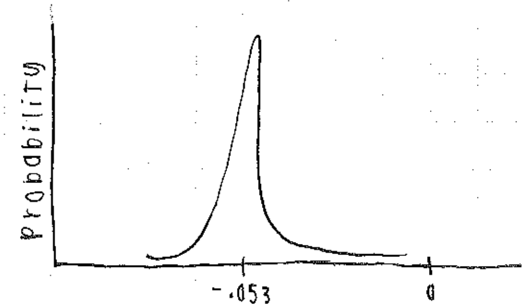
\$50 × -.47 =

\$50 × .37 =



mean of 1000 \$1 bets

$\sigma_{\bar{x}}$ is much lower



mean of 100,000 \$1 bets

- sampling distribution
  1. less spread than population
  2. normal distribution
  3. related to n, population
  4. larger sample size = smaller stdev.

dragon example: sample of 4 dragons

4.25

5.25

4.75

6

5

3.5

4.5

$\bar{x}$ distribution is normal and follows the above qualities

- increased sample size = more normal

- real life: only one sample
- CLT only really works when $n \geq 30$

review #20

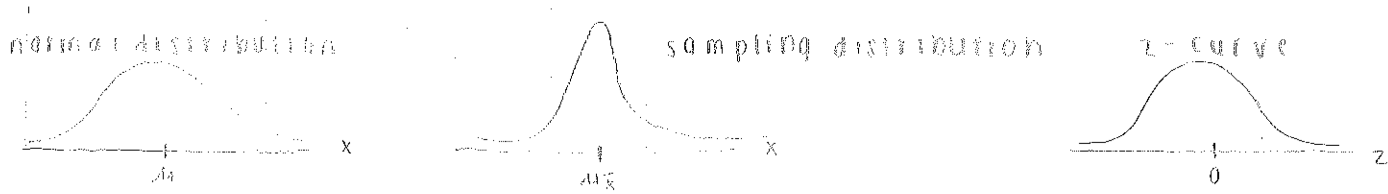
$\mu = 38$ $\sigma = 5$

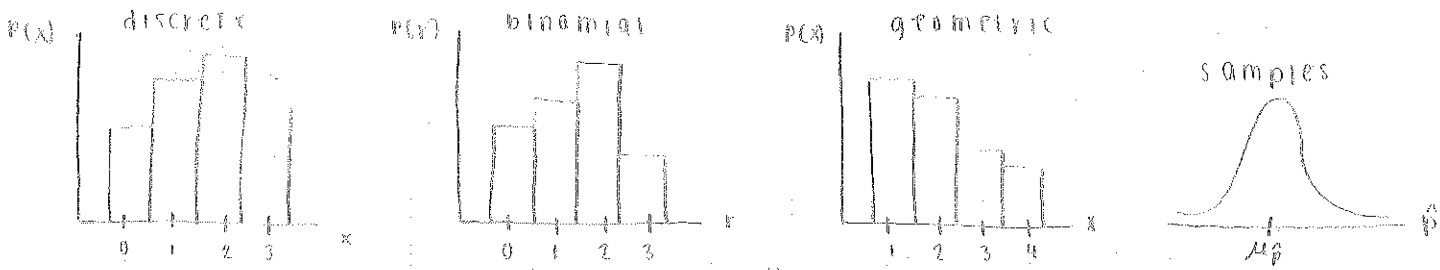
a) $P(X \leq 35)$ normalcdf (0, 35, 38, 5) = .2743

b) $n = 10$ $\mu_{\bar{x}} = 38$ $\sigma = 5/\sqrt{10}$

$P(\bar{x} \leq 35) = .0289$

estimation





estimating μ when σ is known

page 340 – confidence level table

c level z_c

checks:
- RRS
- indep.
- normal
- $n > 30$
- $n < .1N$

$n = \left(\frac{z_c \sigma}{E}\right)^2$

finding **confidence intervals**:

1. find **margin of error**

$mE = z_c\left(\frac{\sigma}{\sqrt{n}}\right)$

↑ margin of error   ↑ standard error

2. $\bar{x} - E < \mu < \bar{x} + E$

↑ population mean

3. sentence

ex. running $\sigma = 1.8$

$n = 90$ $\mu = 90$ $\bar{x} = 15.60$ min find .99 confidence interval for μ

$mE = z_c\left(\frac{\sigma}{\sqrt{90}}\right) = 2.58\left(\frac{1.8}{\sqrt{90}}\right) = .4895$ ↳ $z_c = 2.58$

$15.60 \pm .4895$ $15.11 < \mu < 16.09$

If we took 100 samples of $n = 90$, we expect to contain population mean μ of Julia's 2mile jogging times 99 times

homework: 8.1 (9 - 12, 15, 19 - 21)

estimating μ when σ is unknown

W.S. Gosset

chemist

Guiness Brewery

Ireland

1908

Student's t - distribution bc s varies

t- table

d.f. = degrees of freedom → n-1

if n = 4, d.f. = 3   95% confidence → 3.182

we can use sample standard deviation to estimate population σ

different t-distribution = different sample sizes

as n increases, t curve gets more like z-curve bc sample standard deviation gets more like population σ

t-score becomes z_c

we don't know population distribution so we can use t curve

$\bar{X} = 289.1$ $\bar{x} \pm t\left(\frac{s}{\sqrt{n}}\right) \rightarrow 289.1 \pm t\left(\frac{1.9120}{\sqrt{16}}\right)$ $t = 2.262$

$s = 1.9120$

ex. dental work

$n = 14$ d.f. = 13 $\bar{X} = 1.43$ $S = .38$ mm

90% confidence → $t = 1.771$

$1.43 \pm 1.771\left(\frac{.38}{\sqrt{14}}\right)$ $E = .1799$

$1.2501 < \mu < 1.6099$

if we took 100 samples of size n = 14, we expect to contain the population mean μ about 90 times

checks:

Random rep. sample

Independent

mound shaped → t-distribution

- if NQT is a line, it's mound-shaped
- box, histo, pearson

n cannot be bigger than 10% of population size $n \not> N/10$

- "it's reasonable to assume ______"

8.2 (11, 13-16, 20)

ex. shoplifting

$\bar{x} = \$327.67$ $s = \$29.31$ $n = 9$ → it is reasonable to assume open store for 90 weeks

90% confidence

↳ d.f. = 8, $t = 1.860$

$E = 1.86\left(\frac{29.31}{\sqrt{9}}\right) = 18.17$

327.67 $309.5 < \mu < 345.84$ dollars

if we took 100 samples with size $n = 9$, we expect to contain the population mean μ of the dollar amount attempted to shoplift 90 times

8.3 (11-16, 18, 19) EC project implicit, turn in to classroom

binomial: 2 outcomes S/F

indep.

n =

r = # of S / n

p and q

$\mu = np$ $\sigma = \sqrt{\mu q}$

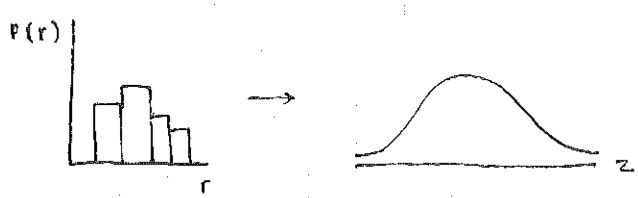
$\hat{p} = \frac{r}{n}$ $\hat{q} = 1 - \hat{p}$

$np > 5$ $nq > 5$ → normal approximation is justified

will be smaller

$n = \hat{p}\hat{q}\left(\frac{z_c}{E}\right)^2$ if you know $\hat{p}$

$n = \frac{1}{4}\left(\frac{z_c}{E}\right)^2$ if you don't know $\hat{p}$



can use c-lvl and z_c

1. $mE = z_c\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$
2. $\hat{p} - E < \overset{\pi}{p} < \hat{p} + E$
3. sentence

checks: RRs and independent

$n\hat{p} > 5$, $n\hat{q} > 5$

ex. smoking

$r = 250$

$n = 1{,}000$

$\hat{p} = \frac{250}{1{,}000} = .25$

$c = 95\%$

$z_c = 1.96$

$E = z_c\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96\sqrt{\frac{(.25)(.75)}{1{,}000}} = .0268$

$n\hat{p} = 250 > 5$

$n\hat{q} = 750 > 5$

∴ ANAIJ

$\hat{p} - E < \overset{\pi}{p} < \hat{p} + E \rightarrow .2232 < \overset{\pi}{p} < .2768$

if we took 100 samples of size $n = 1000$, we expect to contain the population percent π of large corp. employers who prefer nonsmokers 95 times confidence (arrow to 99%) E (arrow to .05)

25\. a) 99% sure that $\hat{p}$ will be .05 from p

$n = \frac{1}{4}\left(\frac{z_c}{E}\right)^2$ ← no preliminary estimate

$n = \frac{1}{4}\left(\frac{2.58}{.05}\right)^2 = 665.64 = 667-1$ ☺

b) $\hat{p} = .68 \quad \hat{q} = .32$

$n = (.68)(.32)\left(\frac{2.58}{.05}\right)^2 = 579.37 \rightarrow 580$

KEY TERMS

A **process** is a chain of steps that turns inputs into outputs. Every process has variation. **Common cause variation** is the variation due to day-to-day factors that influence the process. **Special cause variation** is the variation due to sudden, unexpected events that affect the process.

When a process is running smoothly, with its variables staying within an acceptable range, the process is **in control**. When the process becomes unstable or its variables are no longer within an acceptable range, the process is **out of control**.

A **run chart** is a scatterplot of the data values versus the order in which these values are collected. The chart displays process performance over time. Patterns and trends can be spotted and then investigated.

Control charts are used to monitor the output of a process. The charts are designed to signal when the process has been disturbed so that it is out of control. Control charts rely on samples taken over regular intervals. Sample statistics (for example, mean, standard deviation, range) are calculated for each sample. A control chart is a scatterplot of a sample statistic (the quality characteristic) versus the sample number. Figure 23.9 shows a generic control chart.

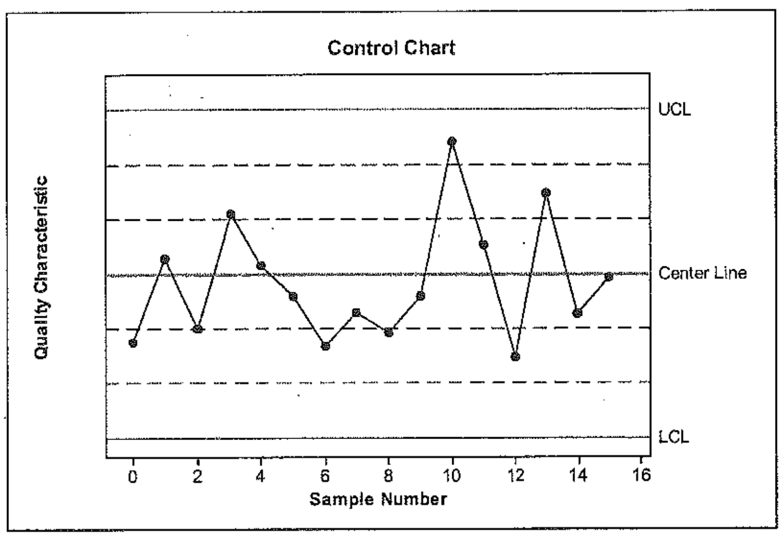


Figure 23.9. Generic control chart.

The **center line** on a control chart is generally the target value or the mean of the quality characteristic when the process is in control. The **upper control limit** (UCL) and **lower control limit** (LCL) on a control chart are generally set $\pm 3\ \sigma/\sqrt{n}$ from the center line.

An $\bar{x}$ **chart** is one example of a control chart. It is a scatterplot of the sample means versus the sample number. Scatterplots of sample standard deviations or sample ranges over time are two other types of control charts.

Decision rules consist of a set of rules used to identify when a process is becoming unstable or going out of control. Decision rules help quality control managers decide when to stop the process in order to fix problems or make adjustments.

THE VIDEO

Take out a piece of paper and be ready to write down answers to these questions as you watch the video.

1. What was W. Edwards Deming known for?

statistical process management

2. What is a process, statistically speaking? Give an example.

chain of steps that turn inputs into outputs

machine turning metal into tools

3. What does it mean for a process to be in control?

process is running smoothly with variables in an acceptable range

4. Why did Quest Diagnostics' lab need a statistical-quality-control intervention?

they weren't meeting deadlines and people were dissatisfied

5. In Quest's control chart, how did they determine where to set the upper and lower control limits?

3 standard deviations from the center line (target time)

6. How did Quest respond to what it learned from its control charts? What were the results of these changes?

remodeled the whole process with different groups for different tasks

variation of times got tighter around the target time