HW = 5~8, 11
1.1

Joyce Li

Period 3

08-17-2023

Level of Measurement

Nominal | Categorical Data: names, labels, listings ; order is not Important
||

Ordinal | Catagorical Data ; can be arranged in order, but differences between values cannot be found or have no meaning: Ranking ; Ex = Low, Medium, High.

Interval | Quantitative Data (#) ; can be arranged by order, differences are meaningful; No natural zero or starting point.

Ratio | Highest Level of Quantitative Data ; can be arranged in order, differences & division make sense, a starting zero.

Quiz
① Ages (in years) of students = Ratio
② IQs of Students = Interval
③ Ranking of Students Efforts (Good, Better, Best) : Ordinal

Statistics

Statistics Data
- found almost anywhere
- no statistical bias

Definition = Collection, Organization, analysis, interpretation of numerical information.
- Collection = Samples, Experiments, Census, Simulations → (SECS)

Quiz
① observational studies
② surveys
③ experiments
④ anecdotes

stat
↓
Sort A
↓
L_1

randInt → can have same #

TI-84 Calculator
① MATH → Prob → $\boxed{8}$
② lower & upper
③ n → sample size
④ Stor → List x

08-21-2023 STATS is ...

1. Collecting • sample • experiment • census • simulation	2. Organize
3. Analysis	4. Interpret

Survey → sample
observation → experiment

Organize : graph, chart, Table.

Analysis : center / shape / spread / Outliers, Gaps
Trends
Measures of central tendency
of dispersion / spread
Confidence intervals
Hypothesis test

Interpretation : Assumption
Pvalue & Alpha level
confidence level
Z score or T score or chi sq score
written summary

Population : Measurements of Entire group of interest
{ Data from ALL
Quantitive variable from ALL
Categorical variable from ALL

$N = x$ ← population size
Population = kind of people + Data from all

Sample : part of the population
{ Smaller section of those being measured
Data from
the variable of

$n = x$ → sample size

1.1 #8 a) mpg = variable
b) Quantitative
c) Average mile per gallon for all new hybrid small car

Random # Sheet
① # each car.
② Drop an object.
③ go up / right / left / down
④ 3 number as one, within 001 ~ 550
⑤ Go inspect the car

$\boxed{\text{Random Sample}}$ overcover → cover, but too much
Undercover → Didn't cover.

↓
from complete list

Google sheet
① number each locker
② Open google sheet
③ = RandBetween (1, 1500)
④ right down corner

→ other side

HW 1.2 (1~4, 10~12)(16)(20)
*

Joyce LT

Period 3

08-23-2023

ch 8 Pdf (before Test, probably not now)

Simpling { Simple random ①, convenience (X) ②, systematic ③, cluster ④, Stratified ⑤ }

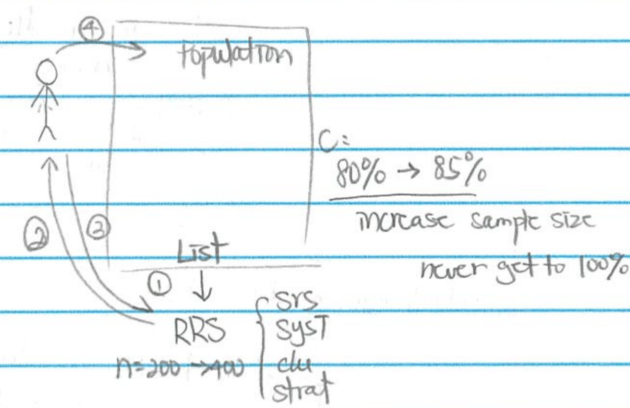
Overcover vs. Undercover

STATS is:
① Collecting
② Organize
③ Analysis
④ Interpret

① number & random choose
② People nearby
③ Start point, choose after O x x x O x x x ...
④ cluster (usually by location) (all of the sample in the cluster)
⑤ 分类 then random choose

Fight Bias !
- To begin well, get a LIST
- Simple Random Sample (SRS)
- Stratified Sample
- Cluster Sample
- Systematic Sample

Inferential Statistic Pic



EX 1.2 (20)
a) Stratified
b) Simple random (SRS)
c) Cluster
d) Systematic
e) Convenience

Bias = Read PPT
- Non-Res

$\boxed{n = 1,200}$

each 298 Patient with chest Pain → TI-84 = Random Int No Rep (1, 298, 149)
→ left over

→ pre-test → Treatment → 3 months
↓
post-test
↓
Compare

1.3 (7~10)

9 (b) UPO.

TI Clac RandInt NoRep (1, 18, 9).

\# each 18 OPHS 1~18 → R.A. → 9 HS UPO (Treatment group) → 10 weeks → Compare result

Left over → 9 HS No Placebo (control group) under... → 10 weeks → Compare result

Pretest
- Survey
- examine record

post-Test
same as pre-Test

no blind

9(c)

TI calc RandInt NoRep (1, 75, 40)

\# each 75 Smokers 1~75 → Rand. A. → 40 with drug (Treatment g.) → 2mon → Compare Results.

Left over → 35 with no drug placebo (control g.) → 2mon → Compare Results.

Pre-test
- survey → 1st, → packs
- X-ray

single blind

post-test
- survey.

Blocking (age, gender ...).

300 smoker
F → 160 Female
→ Teen ~ 20s (50) → (1, 50, 25) randIntNoRep with drug → placebo . . .
→ 30 - 40s (60)
→ 50+ (50)

M → 140 Male
→ Teen - 20s
→ 30 - 40s
→ 50+

Simulation (仿真/模拟) start with model, in some fashion, experiment, a procedure for imitating (simulation), repetition. [TI-84, Computer, RNT]

The imitation of chance behavior, based on a model that accurately reflects the experiment under consideration, is called a simulation

Simulation☆
- ① State the Prob.
- ② Assumptions
- ③☆ Assign digits. (similiar to real out come)
- ④ Many Reps. 25+ time
- ⑤ state Cond. Sim ≈ reality

Example: (follow ☆) flip coins
① Find Hot streak Coins (3+ same in roll)
② Head ≈ Tail equally likely. (50% - 50%) (Toss are Indep.)
③ 0, 1, 2 .... 9, RNT 1/10 (Indep.) (1 digit = 1 coin) (odd = H, Even = T)
$\underbrace{10\# = 10 \text{ Coin flap}}_{10\#s} = 1 \text{ Rep}$
④ Repeat many time (25+)
⑤ state Cond. ≈.

5.56 (23:59) ① 10 University student about Evening exam
② Assumptions: Y/N = 80%/20% (Student are alone).

5.57 ① simulation of free throws in real High School Basketball game rival/ close B.C.P
② Assumptions (Indep. shots) (15 feet & 10feet) (no fan's sound) (no wind) (no effect). (70%) (not injurded).
③ Assigned digit. 0,1,2,3,4,5,6 = $ 7,8,9 = Bluk
Rep 1 96746 (No) Rep 2 12149 (No). Rep 3 37823 (No) Rep 4 71868 (Yes) $x/25$
④ Rep. $x/25$. $5/25 = 20\%$
⑤ state cond.

6.16 ① simulate baseball hit major lead 20 at bats.
5.60 ② Indep. Pitcher = Reg. x colo Relcics. Late Season
③ 0,1,2 = Hit ⓧ (30% ≠ 32%) 00 ~ 31 = Hit, 32 ~ 99 = not hit

Joyce Li

Period 3

Statistic History 08-11-2023

① In the beginning, Statistics involved summarizing data by means of charts and table.

② what did the Chinese use statistics for? when?
- Chinese who used statistics for keeping state record. The Chinese under the Chou Dynasty, 2000 B.C, maintained extensive lists of revenue collection and government expenditures, availability of warriors.

③ who began the study of statistics?
Bills of Mortality - The study of statistics was begun by John Graunt. (English) (1620~74)

④ why wasn't statistical theory commonly use prior to 1930? what changed?
- Time consuming, Computer involved

⑤ what is inference?
- Generalizing on samples.

⑥ Where are the origins of the study of probability found?
- Pascal, Fermat.

⑦ who wrote the first book on the theory of randomness? what was the title?
- Cardan, Italian (1501~1570) The Book On Games of chance

how to lie with stats. 2 weeks

Graph = Information more quickly, efficiently, attractively

Title! name!

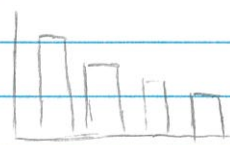
7 second, keep it simple

Analysis, trends, why?

Bar Graph = Show Differences squiggle

comparing several

Pareto charts = vertical bar graph arranged tall to small



ogive

Circle (Pie) graph.

- % of a whole.
- inside (#) or a key.

Line graph

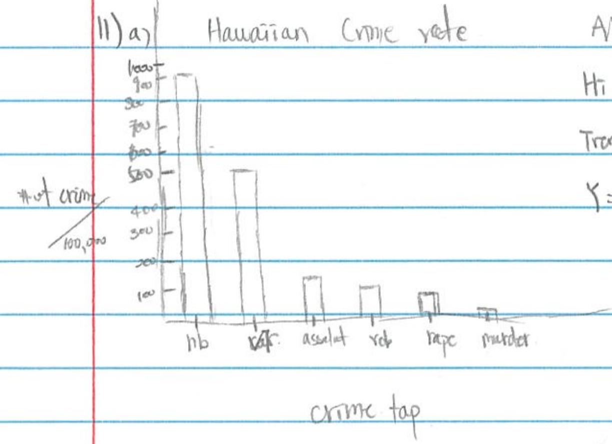
Picto graph = full symbol

Analysis

- Note the high & low
- Trend
- conjecture, why

1. Misleading.

11) a)



ANA:

Hi, Low house burlary, murder.

Trend: 难度, 程度 less violence, more frequen.

Y: easier target for HB & VT.

usually not super violent.

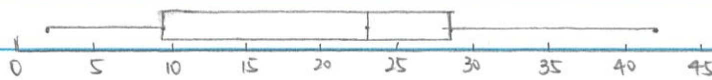
Joyce Li

Period 3

10-30-2023

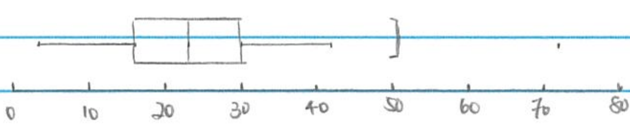
3.3 (7, 8, 11) + Billionaries

7. min = 2 Q_1 = 9.5 Med = 23 Q_3 = 28.5 max = 42



center = 23 (Med)
Spread = IQR, 19
shape = Uniform
Outlier = $Q_1 - 1.5 \times IQR = -26.5$
$Q_3 + 1.5 \times IQR = 57$
None

8. min = 3 Q_1 = 16 Med = 23 Q_3 = 30 max = 72



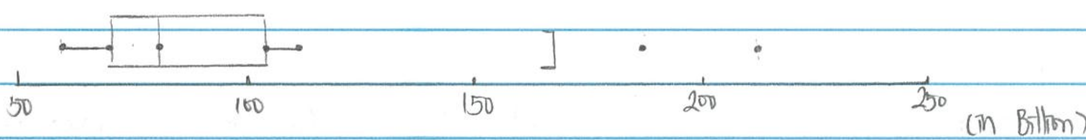
center = 23 (Med)
Spread = IQR, 14
shape = Uniform
Outlier = $Q_1 - 1.5 \times IQR = -18$
$Q_3 + 1.5 \times IQR = 51$
72

11) a) Lowest = CA, Highest = PA
b) PA
c) CA has smallest range, TX has the smallest interquartile range.
d) CA - (c)
PA - (b)
TX - (a)

Billionaries (in Billion)

1. 211
2. 180
3. 114
4. 107
5. 106
6. 104
7. 94.5
8. 93
9. 83.4
10. 80.7
11. 80.5
12. 79.2
13. 77.3
14. 76
15. 68
16. 64.4
17. 59
18. 59
19. 58.8
20. 57.6

Min = 57.6
Q_1 = 66.2
Med = 80.6
Q_3 = 105
Max = 211.



center = 80.6 (Med)

spred = IQR = 38.8 = "small"

shape = skewed right

outlier = $Q_1 - 1.5 \times IQR = 8$

$Q_3 + 1.5 \times IQR = 163.2$

180, 211

Unit 5 (ch 4) Probability 2

- Probabilty = A measure of chance, likehood of an event happening
- Probabilty Notation = P(A) (Pro of event "A")
- Prob Found = f / n (# of desired / # of total result)
- Fraction (reduced), Decimal, % s = all good
- $0 \le P \le 1$
- contemplate certain $P(C) = 1 \rightarrow$ certain
- Ponder Impossible $P(D) = 0 \rightarrow$ Impossible

P(E) = 0.8, 0.9, 0.95
highly likely

P(F) = 0.15, 0.1, 0.05
highly unlikely

depent on situation

P(str) = #F/n = 3/6 = 1/2 = 0.5 event chance, moderate likely

- Complement : P(not the event). P(not A), P(A') or $P(A^c)$ or $P(\bar{A})$
- P(A) + P(A') = 1
- P(A') = 1 − P(A)

Prob vs. Stat

- stats : the sample is known
- conclusion about population
- Prob = the population is known
- ask about sample, chance
- stats reason from the sample toward the population
- Prob reason from the population toward the sample

$$\text{Pop} \underset{\text{stats}}{\overset{\text{Prob}}{\rightleftarrows}} \text{sample}$$

Odd

- special type of Prob
- Fav to UnFav
- Fav = Fav' or Fav / Fav'
- Odd of tails on a coin = T : T' → 1 = 1
- Odd of a 6 on a die = F : F' → 1 = 5

4.1(11,14~17,19).

100 - tail   km - head

LOL #5 (Law of Large numbers)

- Fallacy of the short run
- Law of Large numbers

Sample Space

- $S = \{\text{the set of all possible outcomes}\}$
- Coin Toss = $S = \{h, t\}$
- 2 coin = $S = \left\{ \begin{array}{cc} S & G \\ t & t \\ h & t \\ t & h \\ h & h \end{array} \right\}$

round 1 Dr. K

round 2 coco cola

round 3 Pepsi

① 病	⑥ :	⑪ 自杀
② 病	⑦ :	⑫ 车
③ :	⑧ gunshot (自)	⑬ 自杀
④ :	⑨ drug	⑭ 车
⑤ :	⑩ gun (他)	⑮ 车

→ 非常不常见 & 非常常见 (大部分无危险).

$$S = \left\{ \begin{array}{ccc} 1 & 2 & 3 \\ C & P & K \\ C & K & P \\ P & C & K \\ P & K & C \\ K & C & P \\ K & P & C \end{array} \right\}$$

$P(\text{right}) = \frac{E}{n} = \frac{1}{6} = 16\frac{2}{3}\%$ = "unlikely".

Expected #: $E(x) = n \cdot P = 25 \times \frac{1}{6} = 4.1\bar{6}$

Joyce Li

09-05-2023

Lace of shoes

Hypothesis More Lace than none lace. G2 G3

Lace None Lace

M= ~~IF~~ 2 4 3 ~~4~~ 14 M=

F= ~~IF IF~~ 2 1 2 ~~6~~ 15 F= T 2

29

M 3 1 5 8 16 M 1 1

F 5 4 3 5 17 F 5

M ~~7~~ 1 2 2 1 4 ~~8~~ 10 M 1

F ~~1~~ 2 1 2 1 3 9 19 F = 1 2 J

64

Good: Observation = none changing

Saw People

Collect data

Poor: Limited in Location (G Building)

Limit in People (Senior & Junior)(Undercover)(Teacher)

None sure of population

No List

Limited in Time (rush)

Overcover (morning, Tue) Undercover(noon, Wed..., Oct).

Population Lace waring Partice of ALL sts at OPHS early Sept 2023

n. 50+ , gender (guess) , Grade Sr / Jrs

random selection.

not a survey.

Describe the struggle.

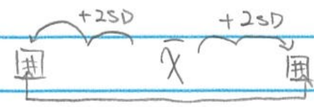
+ 3.3 (7, 8, 11) Video

Chebyshev's Theorem

$1 - 1/k^2$, $1 - \frac{1}{k^2}$

$k \neq 1$, $k > 1$

Standard Deviation 2sd : $1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$



75% of # in that area

3sd : $1 - \frac{1}{3^2} = \frac{8}{9} = 88.9\%$

4sd : $1 - \frac{1}{4^2} = \frac{15}{16} = 93.8\%$

Textbook = P111 bluebox

1 sentence conclusion.

EX : $\bar{x} - 4s$ to $\bar{x} + 4s$

$29.1 - 4(1.7)$ to $29.1 + 4(1.7)$

22.3 to 35.9

At least 93.8% of the students would fit into the group that volunteered from 22.3 to 35.9 hours each semester

+ 3.2 (4, 16, 17) 17 (c) (go with 88.9%)

P99 ① / chem ② / Q_1, Q_2, Q_3 ③

① P99 . bit 99% of the people

②

③ Q_2 = 50% , Median

Q_1 = 25%

Q_3 = 75%

18, 20, 21, 21, 21, 21, 22, 22, 22, 23, 23, 23, 23 | 23, 23, 24, 24, 25, 25, 25, 26, 27, 27, 30, 32, 35.

$Q_1 = 22$ $Q_2 = 23$ $Q_3 = 25$

Outlier :

$Q_1 - 1.5 \times IQR = 17.5$ P25 Med P75

$Q_3 + 1.5 \times IQR = 29.5$ P50

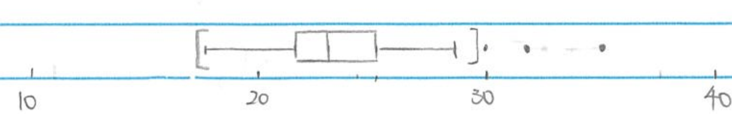
30, 32, 35

IQR

$Q_3 - Q_1 = 3$.

Ana =

center = 23 (Median)

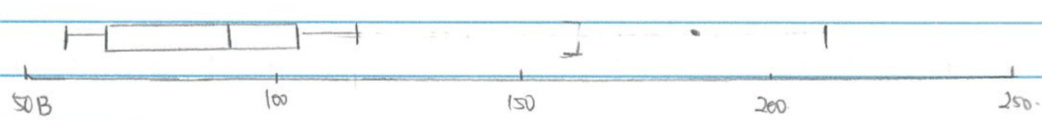


Shape : Skewed right

spread = IQR = 3 "tight"

① 211 B → Max.
② Elon M. 180B
③ Jeff B. 114B
④ Larry E 107B
⑤ Warren B 106B → $Q_3 = 105$
⑥ Bill G. 104B
⑦ Blomberg 94.5B
⑧ Helu 93B
⑨ Ambani 83.4B
⑩ Ballmer 80.7 → Q_2(Med) = 80.6.
⑪ 80.5
⑫ 79.2
⑬ 77.6
⑭ 76
⑮ 68 → $Q_1 = 66.2$
⑯ 64.4
⑰ 59
⑱ 59
⑲ 58.8
⑳ 57.6 → Min

3.3.11
a) low = CA high: Penn
b) Penn
c) CA, Tex
d) (a) Texas
(b) Penn
(c) CA



$ in Billion

Center = 80.6 (Med)
Spread = IQR, 38.8 = "Small"
Shape : Skewed right
J.W. Tukey. Outlier : $Q_1 - 1.5 \times IQR = 8$.
$Q_3 + 1.5 \times IQR = 163.2$
180, 211.

IQR
CV
S
S^2
~~Range~~ ×
chebychv

+ Don't assume people know a certain thing
+ No too long Question or long Survey
+ No leading (scandal)(corruption)
+ Double Negative
+ context effect
+ Pre-test

- more about struggle
- Limitation
- why (?)

09-27-2023
- face to face survey is bad (Response bias)

Survey - Do you believe there are currently Terrorists

- Do you and your parent regularly discuss the news?
  - "Regularly"
  - How big is the news
  - assuming
  - parents
  - Yes or No

Please log the daily amount of time (min) spent discussing news with a trusted adult.

Mon ☐ state
☐ national
☐ global

Tue ☐ ...
☐ ...
☐ ... ...

2.3(1,7)

2.3 #7

min = 1.0 round 1.0 → 1

max = 29.8 29.8 → 30

0	1 4
.	8 9 8 9 7 9 5
1	4 2 4 1 3 2
.	6 7 7 6 5 5 5 5 5
2	
.	
3	0
.	

Tar in your lung

2|7 = 27 mg of Tar

ctr: mode: 15~17

shape: skwed low.

sprd = ch3

out: 30.

Survey 09-19-2023

Ex: - people are effect by others
- Word (Leading Suggest)
- Peer influence (group setting)
- Verbal (face, voice, body language)
- Cities
- 2016 vs. 2021 (tend to give younger higher for recent)

+ Survey
+ Olympics

• Not really. I didn't put much attention on sport.

+ writen (not verbal)
+ more individual
+ body more

- still in group setting
- Word (pharse) (suggest)
- Cities name
- word
- weak & Strong Opinion.

• 1 2 3 4 Ordinal Data

+ writen
+ more individua
+ rank opinion
+ no city (no suggestion)
+ even # choices

- still in group setting
- no option for "not watch"
- actually happened in 2021
- not center
- equal spacing.

P4.
+ Right Now
+ number.

09-21-2023

good	bad

AOL
+ balanced / non-suggestive question
+ equal spacing
+ no slang

OP sur
+ even options.

AOL
- Picture, association
- influence (word & picture)

OP sur
- too long.
- Survey in group setting.
- response bias
- too many word
- "I believe" "should be made" leading
- unequal space

Subway Survey

Good

Bad
- Volutary Response Bias
- No enough option
- leading suggestive

Jamba juice

Good
+ Short, less word.
+ Not leading
+ even options

Bad
- Volutary Response Bias
- Confusion option.
- No Incentive
- Name, address, phone (X)

BBC
- Non-Random selective.
- Influence leading

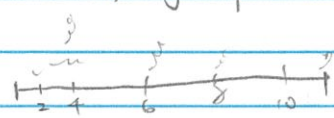
+ Professional Organization.

2.1 (6, 8, 15, 16, 22, 23).

Shape of Histograms {
special bar graph

- Rectangular / Uniform
- Bimodal
- Skewed Right — Look for the tail
- Skewed Left
- Symmetric
- Symmetric / Normal

Dot Pot = need # Line
similar to vertical graph
Quantitive data, mid size
not high spread.



(median).
shape
spread: chapter 3.

# of classes, usually 5~10,

$W = \frac{max - min}{\# \ classes}$ (round up) (3.1 ~ 3.9 = 4)

Begin with smallest value

Use classe limits

$W = \frac{44-14}{10} = \frac{30}{10} = 3.0$ (Don't round)

Classe ← Limite

14~ 16

lower limit → 17

20

↓

41~ 44 ← ~~cant~~ car include max #

2.1 41~47 # 6 4/40 = 5%

Stem & Leaf

1. Max & Min.
2. splitting data, 1|2 ← 12   7 → 0|7
3. let the data flow
4. 2nd graph, put it in order

* Title, key, Analysis.

Small &

1. max = 22

mm = 2

2. tens   ones

3.

0	7 3 5 6 2 2 7 3
1	0 0 4 0
2	2 0

P3 "off" from 1 min.

4.

0	2 2 3 3 5 6 7 7
1	0 0 0 4
2	0 2

2|0 = 20 secs away

center = 7 (median).

shape = skewed higher.

outlier = N/A.

"2 line / stem"

0-4	0	2 2 3 3
5-9	·	5 6 7 7
	1	0 0 0 4
	·	
	2	0 2
	·	

1|4 = 14 sec off

Ana - C

S

S

O

1. max = 25

mm = 3

2.

F		M
7 6 5 5 5 3 5 7	0	5 5 6 5 6 5 5
8 ~~7~~ 5	1	0 0 7
5 ~~4~~ 0 0 5	2	5 5

P3

3.

F		M
7 7 6 5 5 5 5 3	0	5 5 5 5 5 6 6
8 5	1	0 0 7
5 5 0 0	2	5 5

ctr = 7

Skew high

spread

outlier = 25?

N/A

0|2| Female

2|5 = 25 min Male

= 0|2|5 =

4.3 Study Guide

① Read P177 ~ 184

② When the outcomes are equally likely, how do we find P(A).

$$P(A) = \frac{\text{Number of outcomes favorable to the event A}}{\text{Number of outcomes in the sample space.}}$$

③

④ a tree diagram

⑥

⑦ eleven factorial = $11! = 11 \times 10 \times 9 \cdots \times 1$

$0! = 1 = 1!$

Permutation (order matter) $P_{n,r} = \frac{n!}{(n-r)!}$ ↑ order combinations.

Combination (order don't matter) $C_{n,r} = \frac{n!}{r!(n-r)!}$

n = # of trials r = # of successes

$C_{10,4}$ = the # of combination of 4 success out of 10 trials.

nothing break without enthusiasm.

$_{10}C_4 = C_{10,4} = \frac{10!}{4!(10-4)!} = \frac{10!}{4!\,6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4! \cdot 1} = \frac{10 \cdot \cancel{9}^{3} \cdot \cancel{8} \cdot 7}{\cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot 1} = 210$

$P_{10,4}$ = the # of Permutation of 4 success out of 10 trials.

$P_{10,4} = \frac{10!}{(10-4)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{1} = 5040$

$P_{5,2} = 20$ $P_{5,3} = 60$ $P_{5,4} = 120$ $P_{5,5} = 120$.

$C_{5,5} = 1$

4.3 (8, 9, 17 ~ 22)

5.1 Random Variables and Probability Distributions

Random Variables, x — 5.1 (1, 3, 10, 12, 14).

Quantitative measurement, variable (x) — ① continuous

Ex = amount of snow fall — ② discrete

Non-Ex = nominal or ordinal

Discrete Random Variables

① countable, finite, (-3, -2, 0, 3, 7)

② No fraction, decimals, partial

③ # of voters in a election

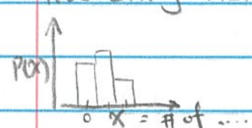
Continuous

① Quantitative, infinite countless.

② -2, -1, 0, 1, 2 and Fraction, decimals, Partial in between.

③ Ex: temp in room right now, or height

Probability Distribution

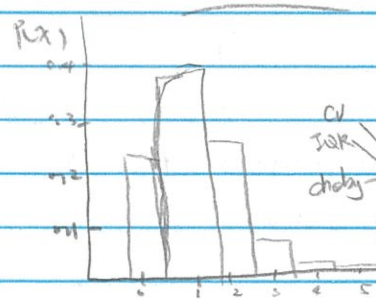


Expected value = $\mu = \Sigma x \cdot p(x)$

Valid = add up to 1

spread $\sigma^2 = Var(x) = \Sigma(x-\mu)^2 \cdot P(x)$

$\sigma = SD(x) = \sqrt{\Sigma(x-\mu)^2 \cdot P(x)}$



Ana:

center = μ = 1.253

shape = skewed high.

CV, IQR, cheby — spread: σ = 0.9698 ≈ high

σ^2 σ (3σ?) outlier: 4+

x = # of repeat offender

5.1, 3) a) 0.25 + 0.6 + 0.15 = 1 Valid

b) 0.25 + 0.6 + 0.20 = 1.05 Non-Valid

14) ~~a~~ d) $\mu = \bar{x} = \Sigma x \cdot P(x) \rightarrow x \cdot P(x)$

= 1.253

e) $\sigma = \sqrt{\Sigma_{L_3}(x-\mu)^2 \cdot P(x)}$ $\mu \approx 1.25$

= 0.9698

÷ 1.253 = 0.7740 high spread

$L_1 = x \rightarrow$ List (1-var)

$L_2 = P(x) \rightarrow$ FreqList

x	$P(x)$	$x \cdot P(x)$	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 \cdot P(x)$

a) $P(x \geq 1) = P(x=1) + P(x=2) \dots + P(x=5)$

$= 1 - P(x=0) = 1 - 0.237 = 0.763$

b) $P(x \geq 2) = P(x=2) + P(x=3) + \dots + P(x=5)$

$= 1 - (0.237 + 0.396) = 0.367$

5.2 (16, 18, 24)

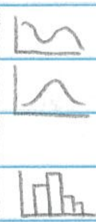
Binomial - Bernoulli (Swiss, 1600s)
↓
2 possible outcomes { Success
Failed

% → · Probability of r successes out of n trial r/n → independent situation
· p - probability of successes $p' = q$ $p + p' = 1$ $p + q = 1$

# of trial (asked people), n = 500 } non-binomial
# of outcome possible = 500

Discrete Random Variable

Continuous RV
..., -0.2, -0.1, 0, 0.1, 0.2, 1 ...
Discrete $\boxed{\text{Binomial}}$
..., -3, -2, -1, 0, 1, 2, 3, ...



↑
S/F, Ind., P=, n=, r/n

Ex: ppt P_3
Trial = room call   Q = 0.2
Ind = Yes   n = 72
S = 3⁻ min get   r = 63
F = 3⁺ min get
P = 0.8

Ex: Trial = yes of date   Ind = Yes   S = say yes   F = say no   P = 71% = 0.71   Q = 0.29   n = 20   r = 18+

$$P(r) = \frac{n!}{r!(n-r)!} p^r q^{n-r} = C_{n,r} p^r q^{n-1}$$

n=6   p=0.3   q=0.7   r=5

$\boxed{2nd} \rightarrow \boxed{\text{vars}}$ (distr) → A: binompdf
X value = r

$P(r) = C_{6,5} \times 0.3^5 \times 0.7^{6-5}$
$P(r) = 6 \times 0.00243 \times 0.7 = 0.0102$
$P(r=5) = 0.0102$

Expected Number [of zombie]
$n \cdot p$

n=6   p=0.7   q=0.3   r=4   Table A11
$P(r=4) = 0.324$
$P(r \geq 4) = P(4) + P(5) + P(6) = 0.324 + 0.303 + 0.118 = 0.745$

Compound Event (2 or more events)

pair of dices = paradice (?)

both, Also, intersection, overlap

P(A and B) $\qquad P(A \cap B)$

$\parallel$

$P(A) \times P(B)$

$P(5 \text{ and } 5) = P(5) \times P(5) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 0.02\bar{7}$ "highly unlikely"

① Replace & ② No Replace $\quad P(\text{Ace and Ace}) \overset{①}{=} P(\text{Ace}) \times P(\text{Ace}) = \frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169} = 0.0059$ "highly unlikely"

$\overset{②}{=} P(A) \times P(A|A) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221} = 0.0045$ "highly unlikely"

Dependency $\qquad$ given

$P(G \text{ and } B) = P(G) \times P(B|B) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6} = 16.\bar{6}\%$ "unlikely"

$P(\text{all work}) = 0.977^6 = 0.8697$

Yes: 87% Live, 13% Die

No: 87% Live, 13% Die

P(A or B) = as long as someone got it done.  $P(A \cup B)$ either, add

$= P(A) + P(B)$

$P(\text{King}) + P(\text{Ace}) = \frac{4}{51} + \frac{4}{51} = \frac{8}{51}$ ← (-1 "Queen" of Grand dad)

$= 0.1569$ "pretty unlikely"

$P(\text{King}) + P(\text{heart}) = \frac{4}{52} + \left(\frac{13}{52} - \frac{1}{52}\right) = \frac{16}{52} = 0.3077$

non-disjoint or non-mutually exclusive

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

~~4.1~~ 4.2(9, 13, 15, 18, 22)

8 black (blk), 7 brown (br), 6 blue (blu), 4 str (str), 4 green (gre), 3 gray (gra), 1 purple (pur) = 33

$P(\text{blk}, \text{blk}) = P(\text{blk}) \cdot P(\text{blk}|\text{blk}) = \frac{8}{33} \cdot \frac{7}{32} = \frac{56}{1056}$

or + $\quad P(\text{br}, \text{br}) = P(\text{br}) \cdot P(\text{br}|\text{br}) = \frac{7}{33} \cdot \frac{6}{32} = \frac{42}{1056}$

or + $\quad P(\text{blu}, \text{blu}) = P(\text{blu}) \cdot P(\text{blu}|\text{blu}) = \frac{6}{33} \cdot \frac{5}{32} = \frac{30}{1056}$ $\qquad = \frac{158}{1056} = 0.1496$ pretty Unlikely

or + $\quad P(\text{gre}) \cdot P(\text{gre}|\text{gre}) = \frac{4}{33} \cdot \frac{3}{32} = \frac{12}{1056}$

or + $\quad P(\text{str}) \cdot P(\text{str}|\text{str}) = \frac{4}{33} \cdot \frac{3}{32} = \frac{12}{1056}$

or + $\quad P(\text{gra}) \cdot P(\text{gra}|\text{gra}) = \frac{3}{33} \cdot \frac{2}{32} = \frac{6}{1056}$

or + $\quad P(\text{pur}) \cdot P(\text{pur}|\text{pur}) = \frac{1}{33} \cdot \frac{0}{32} = 0 \; \frac{0}{1056}$

odd

$F = F'$

$P(1,2,3,4,5,6,7) = 0.6 \times 0.2^6 = 3.84 \times 10^{-5} = 0.0000384 = 0.00384 \rightarrow 99.99616$

$\frac{11}{13} = 0.8462$ = highly likely" reality = "not enough likely" the sensitivity

$\frac{2}{13} = 0.1538$ = highly unlikely" false-negative

$\frac{50}{70} = 0.714285$ = = moderate"

$\frac{20}{70} = 0.285714$ = low likely"

P(present & +) $\frac{110}{200} = 55\%$

P(present & −) $\frac{20}{200} = 10\%$

$P(A) = P(A|B) \leftarrow$ independent

a) $\frac{470}{1000} = P(N) = 47\%$ $P(M) =$ P

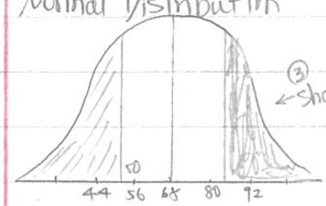
b) $P(N|W) = \frac{420}{500} = 84\%$ $P(N) \neq P(N|M)$ Dependent

~~c)~~ d) $P(N \& W) = \frac{420}{1000} = 42\%$.

$P(N \text{ or } M) = P(N) + P(M) = \frac{470}{1000} + \frac{390}{1000} = \frac{860}{1000} = 86\%$

4.2 (29, 30) Pig , online. Edpuzzle

Normal Distribution $\mu = 68$ $\sigma = 12$



③ ← shading

A) $x = 50$ $Z = \frac{50-68}{12} = -1.5$ $P(X<50) = 0.0668$ ← ② answer.

∃X a 6.68% chance that a Pacific Yellofin Tuna caught weight < 50 lbs.

B) $x = 85$ $Z = \frac{85-68}{12} = 1.42$ $P(X>85) = 1 - 0.9222 = 0.0778$.

X = lbs

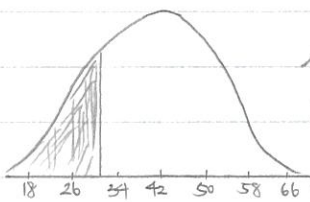
Z = STD

∃X a 7.78% chance that a Pacific Yellowfin Tuna caught weight > 85 lbs.

①notation $\boxed{2nd} \rightarrow \boxed{vars}$ (distr) $\rightarrow \boxed{2}$ normalcdf(④words

c) $P(50 < X < 85) = 0.8549$.

∃X a 85.49% chance that a Pacific Yellowfin Tuna caught weight between 50 & 85 lbs.



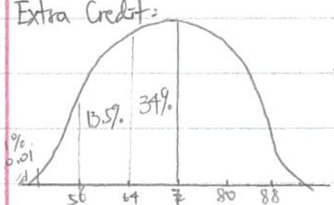
X = min

$\mu = 42$, $\sigma = 8$, $x = 30$

$Z = \frac{30-42}{8} = -1.5$ $P(X<30) = 0.0668$

∃X a 6.68% chance that response time of emergency will take less than 30 minutes.

Extra Credit:



$\mu = 72$ $\sigma = 8$. $\boxed{distr} \rightarrow \boxed{3}$ invNorm(

1% unprotected. 53.39 → round down

53 hours.

6.3 (26, 28, 36) Extra = movie video, 37.

chapter 5 (9)

a) $n = 16$ $P(r \geq 12) = 0.028 + 0.009 + 0.002 + 0 + 0 = 0.039$

$p = 0.5$

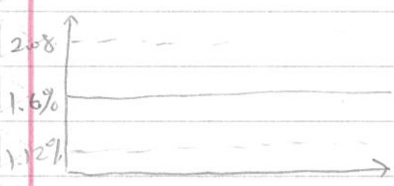
$q = 0.5$

$\frac{r}{n} = \frac{12^+}{16}$

∃X a 3.9% chance that there will be 12 or more federal inmates are serving time for drug dealing.

c) $\mu = np = 16(0.5)$

Statisical Process Control.



out of control.

① ~~1~~ one point 3 σ out side.

3 or 2 point 3 or 2 σ

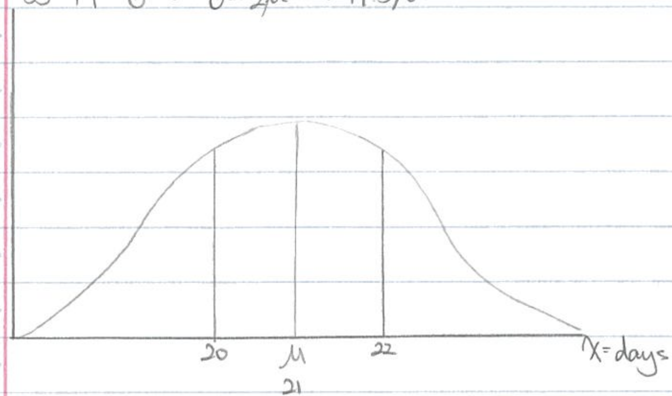
4 or 5 point ~~1~~ 1 and 2 σ same side

8 in Row on the same side of Mean.

6.1 (1, 7, 8, 10)

$\mu = 35 \quad \sigma = 8$

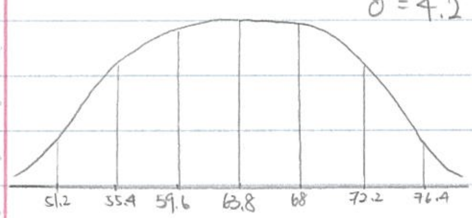
$35 \sim 19 \quad \sigma \rightarrow \sigma - 2\mu \rightarrow 47.5\%$



21 days or fewer = 50% → 500

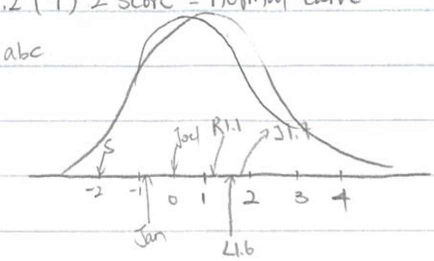
19 days to 23 day = 75% → 950

$\sigma = 4.2''$



6.2 (9) z score = normal curve

abc



d) Raw Score : $x = z\sigma + \mu$

Juan : $1.7 = \frac{x-150}{20}$ Joel : $0 = \frac{x-150}{20}$

$34 = x - 150$ $150 = x$

$x = 184$

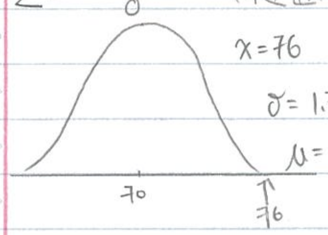
Jan : $-0.8 = \frac{x-150}{20}$

$-16 = x - 150$

$x = 134$

P294 figure 6-18

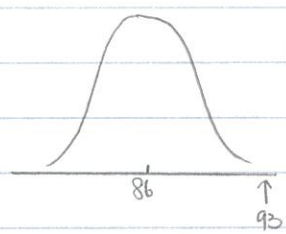
$Z = \frac{x - \mu}{\sigma}$ (不是越大越好，看情况)



$x = 76$

$\sigma = 1.75$

$\mu = 70$



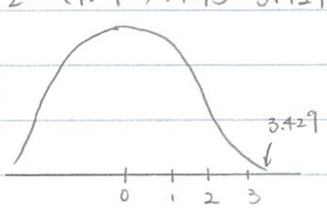
$x = 93$

$\sigma = 2.25$

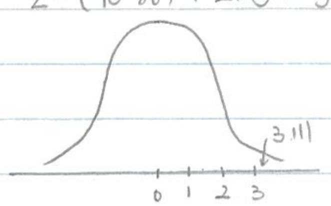
$\mu = 86$

$Z = (76-70) \div 1.75 = 3.429$

$Z = (93-86) \div 2.25 = 3.111$



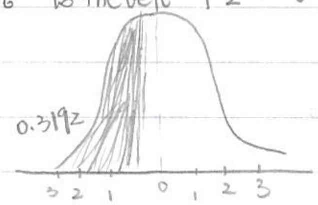
← Better.



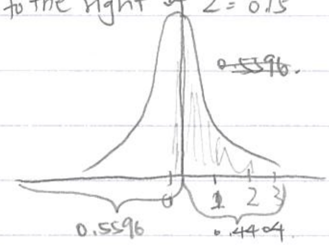
How many standard Deviation away from μ

percentiles - less or equal to a certain value (%)   Ex: 95th percentiles - 95th are below = Top 5%

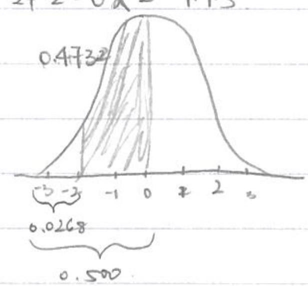
6.2 16 To the left of z = -0.47.



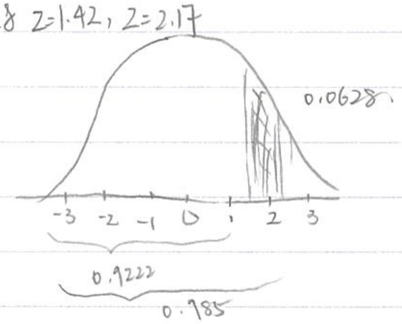
20 to the right of z = 0.15



24 z = 0 & z = -1.93.



28 z=1.42, z=2.17

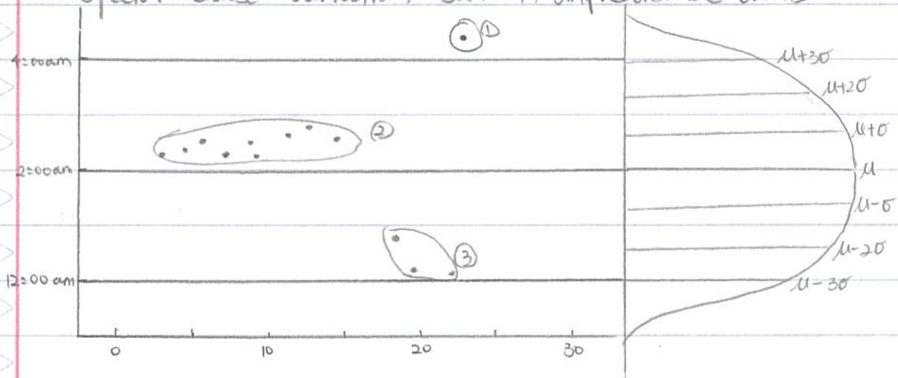


6.2 (4,8,9) (14~40) ← (16, 20, 24, 28, 32).

Control Charts

Common cause variation = day-to-day factors

Special Cause Variation = sudden, unpredictable events



Out of Control:

① 1 point fall beyond 3σ

② 9 consecutive points on one side of μ.

③ 2 of the 3 consecutive points lies beyond 2σ on the same side

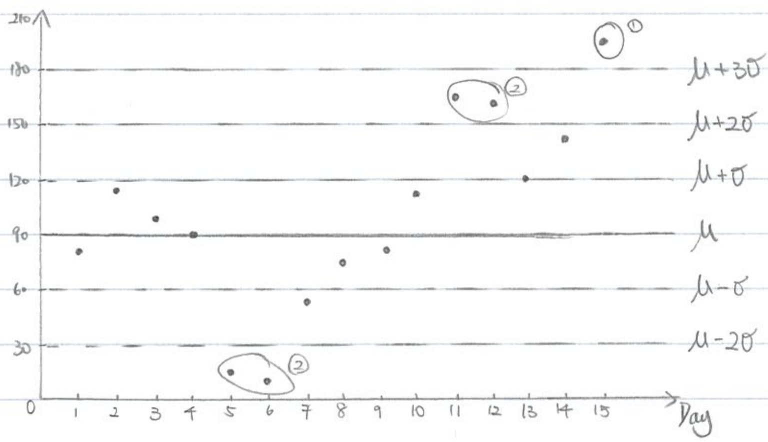
6.1 (12~15)

15) $\mu = 90$

$\sigma = 30$

Max = 195

Min = 10



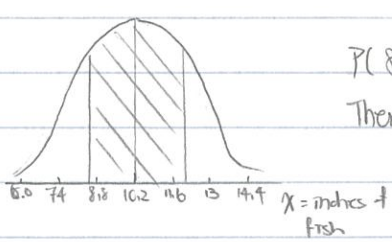
① out of control Sign I

② out of control Sign III

Review

$\mu = 10.2''$

$\sigma = 1.4''$



$P(8 < x < 12) = 0.8427$

There exist a 84.27% chance that the size of the fish in the fishing pond is between 8" and 12".

$\mu, \sigma, \sigma, \rho, \hat{p}$, $y = \alpha + \beta x$

Population → List

↓ ①

① Inferential → Random sample

collect ② Analysis ③

$\bar{x}, s, s^2, r, r/n, y = a + bx$

6.4 Sampling Distribution

Population = All measurement of interest

Sample = A subset of the measurements from the population

random sample:

Probability Theory - The Central Limit Theorem

Normal Probability Distribution

① The $\bar{x}$ distribution is a normal distribution.

② The mean of $\bar{x}$ distribution is μ.

③ The standard deviation

Orig x Distr

Normal

$\mu_{\bar{x}} = \mu$ $\bar{x}$ = Avg of group

$\sigma_{\bar{x}} = \sigma/\sqrt{n}$ ← "st error"

$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

*n can be any # (Big or small)

normalcdf

Lower: 8

upper: 12

$\mu = 10.2$

$\sigma = \sigma/\sqrt{n} = 1.4/\sqrt{5}$

σ goes small

$P(8 < \bar{x} < 12) = \underline{99.77\%}$

higher

The Central Limit

$\bar{x}$ = Avg of group

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

$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

* n is BIG ≥ 30

$\mu = 68$

$\sigma = 3$

$P(67 < \bar{x} < 69) = 0.6827$.

67 68 69 $\bar{x}$ = avg of 9 guys.

$\mu = 40$

$\sigma = 4.2$

$P(38 < \bar{x} < 41) = 92.13\%$.

p of avg = Orig x Distr

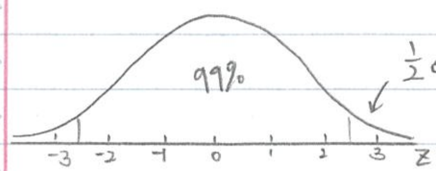
$n = 36 \boxed{\geq 30}$

0, 0, 0, 0, 6.13

0, 0, 0, 0, 0, 377

6.5 (14, 16, 18)

7.1 Estimation



$\frac{1}{2}$ of 1% = 0.005

-2.57 or -2.58

Level of Confidence C.	Z_c
0%	
80%	1.28
90%	1.645
95%	1.96
99%	2.58

A) $E = Z_c\left(\frac{\sigma}{\sqrt{n}}\right)$

b) $\bar{x} - E < \mu < \bar{x} + E$

$\sigma = 1.80$, $\bar{x} = 15.60$, $n = 90$, c-level = 0.99

$E = (2.58)\left(\frac{1.80}{\sqrt{90}}\right) = 0.4895$

$15.6 - 0.4895 < \mu < 15.6 + 0.4895$

$15.11 < \mu < 16.09$

If we took 100 samples we expect to catch the population mean (μ) of JJI 99 occasions.

15) c-level = 80%, $n = 15$, $\sigma = 0.33$

$E = (1.28)\left(\frac{0.33}{\sqrt{15}}\right) = 0.1090$

If we took 10 sample we expect to catch the Pop. mean(μ) of hummingbird 80% time

0.33 - 0.1890 =

$3.041 < \mu < 3.226$

$n = ?$ $\quad n\uparrow \quad c\uparrow \quad n = \left(\frac{Z_c \sigma}{E}\right)^2$

d) $n = \left(\frac{1.28 \times 0.33}{0.08}\right)^2 = 27.88 \approx 28$ (always around up!) (Even .1, .2 ...)

7.1 (15, 18, 19, 22)

7.2 (15, 16, 18)

$E = Z_c\left(\frac{\sigma}{\sqrt{n}}\right)$ ← population

C-level	Z_c
80%	1.28
90%	1.645
95%	1.96
99%	2.58

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

If we took ....

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

t-formula = confidence intervals

$\bar{x} \pm t\left(\frac{s}{\sqrt{n}}\right)$

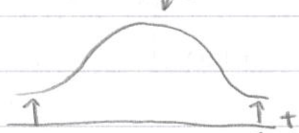
W.S. Gossett - t-curve

chemist

Guiness Brewery

Ireland 1908

shorter



σ is unknown

more space for error

$\bar{x} = 289$, $s = 1.8$, $n = 10$ d.f. $= n-1$.

$289 \pm t\left(\frac{1.8}{\sqrt{10}}\right) \rightarrow 289 \pm 1.3$

$t = 2.262$

$287.7 < \mu < 290.3$

20 years $p = 0.80$

P / F $n = 20$ students

Ind.

Geomatric Distribution

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

S/F Ind
$p =$
$n = ?$
$x =$ 1st success

continuse	RV
... -2, -1, 0, 1, 2 ... (& everything in between)	
Geo.P Disc Bino	

Bino → S/F, Ind
→ p r/n
$\mu = np$
$\sigma = \sqrt{npq}$

first success → n = # of trial on which we get our first success.
(geometric probability distribution) ↳ 第几把为首次成功，止于成功
n is not fixed number

$$P(n) = p(1-p)^{n-1}$$

$$\mu = \frac{1}{p} \qquad \sigma = \frac{\sqrt{1-p}}{p}$$

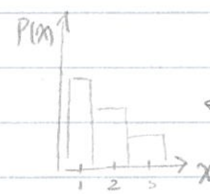
n = # of binomial trial → first success
p = probability of success on each trial
↳ should stay the same.

8.26

Disk Drives Defective = S
D D W = F
Ind
$p = 0.03$
$q = 0.97$

x	$P(x)$
1	0.03
2	F·S 1st Failed, 2nd Success $0.97 \cdot 0.03 = 0.0291$
3	F→F→S $(0.97)^2 \cdot 0.03 = 0.0282$
4	F·F·F·S $(0.97)^3 \cdot 0.03 = 0.0274$
5	F·F·F·F·S $(0.97)^4 \cdot 0.03 = 0.0266$

$P(X=n) = q^{n-1}p$



← always step down.

8.41

S = { HHH, TTT
HHT, TTH
HTT, THH
HTH, THT }

$p = 0.75$
$q = 0.25$

$P(X=3) = 0.25^{3-1} \cdot 0.75 = 0.0469$

5.4 (10, 11)
8.26, 8.41

5 trials = shot (basket) Binomial

Success = in $n=5$

Failed = miss $r=4$

$p = 0.7$

$q = 0.3$

Independent

4th successed Geomatric

trial = catch the ball

Success = catch Independent

Failed = miss $P(X=n) = q^{n-1} \cdot p$ $\mu = \frac{1}{p}$

$n = ?$ $r=4$ $\sigma = \frac{\sqrt{q}}{p}$

stop = caught $P(X=4) = 0.3^{4-1} \cdot 0.7 = 0.189$

$p=0.7$ $q=0.3$

Independent

$P = 0.25$ $P(X=n) = q^{n-1} \cdot p$

$q = 0.75$ $P(X=17) = 0.75^{17-1} \cdot 0.25$

Success = win $= 0.0025$

Failed = lose calculator

$n = 17$th 2nd → vars → distr → F: geometpdf

stop = 17th get (win). $P = p$ $X = $ stop

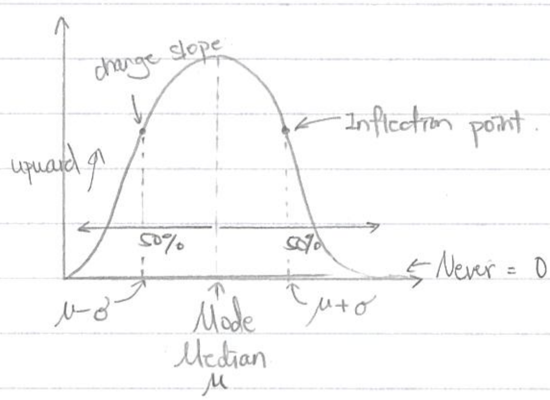
ch6 Normal Distribution

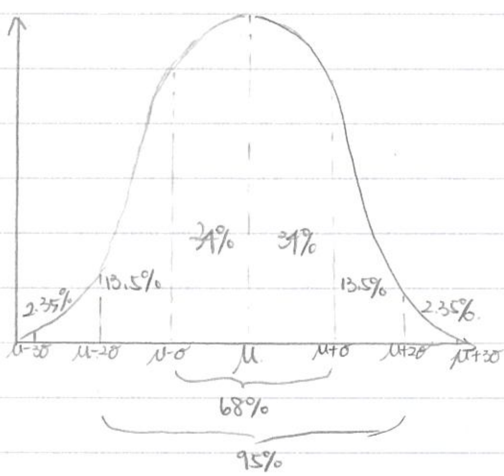
Continuous RV

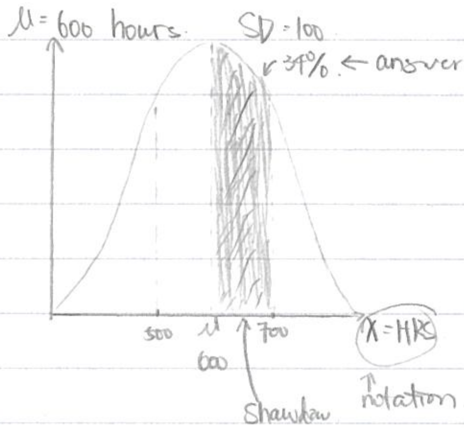
$-2, -1, 0, 1, 2$ (and everything)

← normal (bell)

Geo Discrete BT ← same







[illegible] Li

① High bias, ~~Low~~ variability. high variability

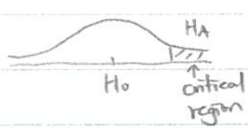
② Low bias, Low variability. :) Good job

③ Low bias, High variability

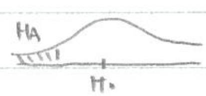
④ High bias, low variability.

Hypothesis = a assumption or belief about parameter. "accept" or "reject"
Null Hypothesis = H_0. The champ. "No change" "No Difference" Tradition.
Alternate Hypothesis = H_1 or H_a. the challenger, better way accepted if Null reject.

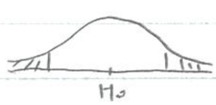
$H_0 : \mu = 3$ hrs
$H_A : \mu > 3$ hrs



$H_0 : \mu = 3$ hrs
$H_A : \mu < 3$ hrs



$H_0 = \mu = 3$ hrs
$H_A = \mu \neq 3$ hrs

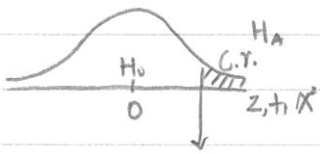


The true states of nature	Accept Null	Reject Null
Null hypothesis is true →	Correct Conclusion	Type I Error rejected a true null hypothesis α = Level of Significance 0.01, 0.05, 0.10 $1 - \alpha = c$
Null hypothesis is False	Type II Error accept a false null hypothesis β =	Correct conclusion Power = $1 - \beta$

4 Ingredients to Stats Test   8.1(16,17)

① H_0 → stay in population $\alpha = 0.05$

② H_1 or H_A $>, <, \neq$ C-level $= 0.95$



p-value $= 0.2805$ $\boxed{\geq}$ $\alpha = 0.05$

"wern"

③ C.V. critical value   weak evident vs. null

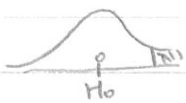
④ $\bar{x} = \ldots \rightarrow z$.   P-value $= 0.0280$ $\boxed{<}$ $\alpha = 0.05$

"Sern"

strong evident vs. null

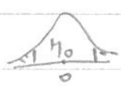
17) a) $H_0: \mu = 16.4'$

b) $H_1: \mu > 16.4'$



If the p-value is low, the H_0 should go.

d: $H_1: \mu \neq 16.4'$



$\boxed{8.1(21,24)\ 8.2(11,12)\ \ (\text{Pepp})}$

P457

S known, σ unknown, $df = n-1$, $\bar{x} \rightarrow t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ $\bar{x} \rightarrow z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

t-table → df → find t → two or one-tail → Range $0.02 < \text{p-value} < 0.05$

8.2(18, ~~19~~, 20)

18) $\mu = 14$, $\bar{x} = 15.1$, $S = 2.51$, $\alpha = 0.01$, $n = 10$, $df = 9$. higher (1-tail).

$t_c = \pm 2.821$. $t = (15.1 - 14) / [(2.51)/\sqrt{10}] = 1.39$.

$H_0 = \mu = 14$ Insufficient SSes

$H_A: \mu > 14$ $0.1 > \text{p-value} > 0.075$ $\boxed{>}$ $\alpha = 0.01$ "wern"

20) $\mu = 77$, $\bar{x} = 71.4$, $S = 20.65$, $n = 20$, $df = 19$, $\alpha = 0.05$, less (1-tail).

$t_c = \pm 1.729$ $t = (71.4 - 77) / (20.65/\sqrt{20}) = -1.21$

$H_0: \mu = 77$ Insufficient SSes

$H_A: \mu < 77$ $0.125 > p\text{-v} > 0.100$ $\boxed{\geq}$ $\alpha = 0.05$ "wern

$Z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \hat{p}$ Z_c = critical value

8.3 (9, 10, 12)

$p = 0.67$ $\hat{p} \rightarrow Z = (0.5526 - 0.67) / \sqrt{(0.67)(0.33)/38}$

$n = 38$ $= -1.5391$

$r = 21$ ≈ -1.54

$\alpha = 0.05$

"less"

$\hat{p} = 0.5526$

% Intervals Tests

- RRS
- Ind $n = \frac{N}{10}$
- $np \geq 5$ "ANAIJ"
  nq

8.3 (9)

purple / white

8 Review (5~7) with TI-84

Monday Test

Hyposis Test (Dependent)

- 2 groups
- Before / After
- Matching link (natural match)

Advantages:

reduce the danger of extra factor

reduce variation

$\bar{d}$ mean different between the paired data

$= \bar{x}$ (sample mean) (difference 差).

Sd sample standard deviation

T-Test

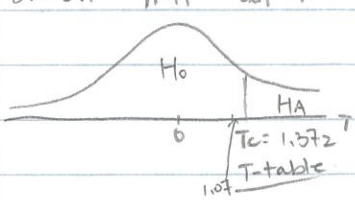
$t = 1.07$

$P = 0.1553$

$H_0: \mu_d = 0$ The average difference of L & R ~~mate's~~ feet of op mate ≈ 0

$H_A = \mu_d > 0$

$\alpha = 0.1$ $n = 11$ $df = 10$



$\bar{d} \rightarrow t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}}$

$= 1.07$

H0 HA fail to reject the null

∃x Insssess op mate feet different ~~than~~ is > 0

$\therefore \mu_d = 0$

p-value = 0.1553 $\boxed{\geq}$ $\alpha = 0.1$

"wcm"

Ex: 2 tests for each patients (dependent)

$H_0: \mu_d = 0$. $H_A: \mu_d > 0$.

$T = 6.0661$ $P = 9 \times 10^{-4}$

8.4 (9, 12, 20) vid 28

① Test $\mu_1 - \mu_2$ when σ_1 & σ_2 is know.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$Z_c \to \alpha$

8.5 (15, 17, 18). Extra: Perils. highlight

② Test $\mu_1 - \mu_2$ when σ_1 & σ_2 is unknow.

$\sigma \to s$.

$$\cancel{Z} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$t_c \to$ table

8.5 (19, 25, 26)

15. Test → ③ 17. Test → ③

Z-Test of Difference of 2 Independent Mean

H_0: $\underset{\text{child}}{\mu_1} = \underset{\text{adult}}{\mu_2}$ The average REM sleep time of children is same as adult.

$H_A: \mu_1 > \mu_2$ $\alpha = 0.01$ $Z_c = 2.58$.

$n_1 = 10$ $n_2 = 10$

$\bar{X}_1 = 2.8$ $\bar{X}_2 = 2.1$

$\sigma_1 = 0.5$ $\sigma_2 = 0.7$

26, L3, L4.

T-Test of D of 2 Ind Mean

$\alpha = 0.05$

H_0: $\overset{I}{\mu_1} = \overset{S}{\mu_2}$

$H_A: \mu_1 < \mu_2$

$t = -1.1263$ $P = 0.1402$

19) L5, L6.

T-Test of D of 2 Ind Mean

$\alpha = 0.01$

H_0: $\overset{NE}{\mu_1} = \overset{RM}{\mu_2}$

$H_A: \mu_1 < \mu_2$

$t = -0.9524$ $P = 0.1761$

Joyce Li

Period 3

Quiz

① Dependent
② Independent
③ Dependent
④ Dependent.
⑤ Independent.
⑥ ~~Dependent~~ Ind.

have	have not.	
$n_1 = 792$	$n_2 = 490$.	→ 20 months

death rate:

$\hat{p}_1 = 14.2\%$	$\hat{p}_2 = 19.8\%$	$\boxed{6 = \text{2-Prop Z Test}}$
$x_1 = 112$	$x_2 = 97$	

$p = 0.0074 \; \boxed{<} \; \alpha = 0.01$

"Sean"

chi - square.

#9. $\alpha = 0.05$ 3x2

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

O	E	O-E	$(O-E)^2$	$(O-E)^2/E$
62	49.02	12.98	168.48	3.44
68	74.22	-6.22	38.69	0.5212
56	62.76	-6.76	45.69	0.7280
45	57.98	-12.98	168.48	2.91
94	87.78	6.22	38.69	0.4408
81	74.24	6.76	45.69	0.6154

$\chi^2 = \Sigma = 8.64$

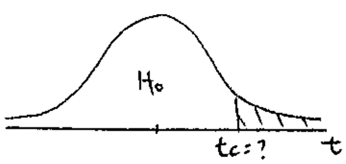
Inference Template

Problem Number: 9.3 Pssx Ex 8.

Name: ____________________

Name of Test/Interval:
Null and Alternative Hypotheses: (in words and symbols) H_0 : $\rho = 0$ $\beta = 0$ H_A = $\rho = 0$ $\beta = 0$
Check of Conditions:

Math Box: (standard deviation, test statistic, p-value, margin of error, sketch, etc…)



$r \rightarrow$, $b \nearrow$ $t = 7.56$

Decision and conclusion:

- $\cancel{H_0}$ H_A $\alpha = 0.01$
- ∃X sses x (age of wolves) and y (weight) has positive correlation and positive sloop.
- p-value = 0.0003 $\boxed{<}$ $\alpha = 0.01$

"Sean"

Inference Template

Name: ____________________

Problem Number: 9.3 Ex 6

Name of Test/Interval:

T-Test of population correlation ρ

Null and Alternative Hypotheses:
(in words and symbols) $H_0 = \rho = 0$ x (% of people go to college) and y (income)

$\beta = 0$ have no correlation

population sloop ≈ 0.

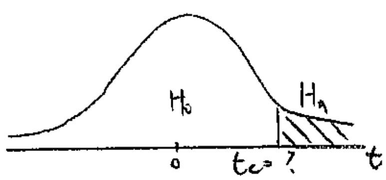
$H_A: \rho > 0$ $\beta > 0$.

Check of Conditions: X.

CALC → List (#) → stat → TESTS → F: LinReg TTest.

Math Box: (standard deviation, test statistic, p-value, margin of error, sketch, etc…)

$$r \to t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \qquad d.f = n-2$$



$$b \to t = \frac{b}{S_e}\sqrt{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}$$

t = t, but r ≠ b.

Decision and conclusion:

③ Test of Difference of Proportions. $p_1 - p_2$

$p \to$ 数据,统计出来的
can be unknown.

Large values $\hat{p}_1 - \hat{p}_2 {}^{(1)} = \frac{r_1}{n_1} - \frac{r_2}{n_2}$

$\mu = p_1 - p_2$ $\quad \sigma = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2} {}^{(2)}$.

$$Z = \frac{\hat{p}_1 - \hat{p}_2 {}^{(1)}}{\sqrt{\frac{\bar{p}\bar{q}^{(2)}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

stat → TESTS → 6: 2-Prop Z Test $\quad x_1 = r_1,\ x_2 = r_2$ $\qquad$ 8.5 (29, 33~35).

7.4 confidence intervals for $\mu_1 - \mu_2$ (σ_1 & σ_2 × or ✓) & $p_1 - p_2$. $\qquad$ 7.4 (18~19)

① $\mu_1 - \mu_2$, σ_1 & σ_2 known

$E = z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

② $\mu_1 - \mu_2$, σ_1 & σ_2 unknown.

$E \approx t_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$. $\xrightarrow{\text{same}}$

$\bar{x}_1$ & $\bar{x}_2$ = sample mean $\xrightarrow{\text{same}}$

n_1 & n_2 = sample size.

d.f. $= n_1 - 1$ → go with the smaller one
$= n_2 - 1$

7.4 stat → TESTS → 9 or 0 or B
① ② ③

③ $p_1 - p_2$

$E = z_c \hat{\sigma} = z_c \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$

$(\hat{p}_1 - \hat{p}_2) - E \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + E$ → $\boxed{\text{unit!}}$

notation → $\mu_1 - \mu_2$

for ① & ② $\qquad$ for ③

$+ \longrightarrow + \Rightarrow \mu_1 - \mu_2 > 0$

$\Rightarrow \mu_1 > \mu_2$ $\qquad \pi_1 > \pi_2$

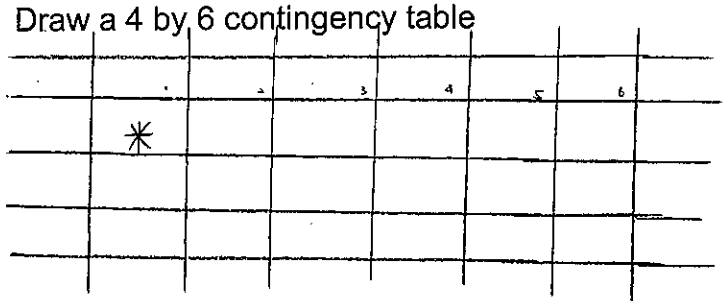
∴ Group 1 who ... brain wave is greater than Group 2.

$- \longrightarrow - \Rightarrow \mu_1 < \mu_2$ $\qquad \pi_1 < \pi_2$

$- \longrightarrow + \Rightarrow \mu_1 \approx \mu_2$ $\qquad \pi_1 \approx \pi_2$

10.1 Chi-Square Study Guide

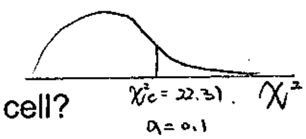
1. Read pp. 624 through 630
2. Draw a 4 by 6 contingency table



3. Put a * in cell #7
4. Chi-squared distributions test the <u>independence of 2 factors</u> always whole #
5. T/F The graph of χ^2 is symmetrical   Irony of factors   O - E observe - Expect

6. T/F As the d.f. increase, the χ^2 distribution becomes more skewed right
7. Where's the mode/high point on a χ^2 distribution? n-2
8. How does one calculate the critical value for χ^2?

   $\alpha = ?$ $df = ?$ → χ^2 table



9. Which is correct for finding the expected frequency of a cell?

$$\frac{\text{row total}}{\text{sample total}} * \frac{\text{column total}}{\text{sample total}} * \frac{\text{sample total}}{1} \quad \overset{\text{Same}}{\text{OR}} \quad \frac{\text{row total * column total}}{\text{sample total}}$$

10. Find E for cell #9 table 10-2

    $E = (100)(60)/(300) = 20$

11. Symbolize observed frequencies

12. χ^2 measures <u>O-E / Independent</u>
13. Use the χ^2 formula to justify the authors' claim in guided exercise 3 that χ^2 =13.31. Show your work.

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

14. Is that "it"? What else do we do?

15. How does one calculate the d.f.?

    $df = (\text{row}-1)(\text{column}-1)$

16. Find the d.f. for your contingency table in #2 above.

    $df = (4-1)(6-1) = 15.$

17. <u>Do</u> 10.1 (9)

10.2 Chi Square Goodness of Fit

$E = \% \times n$
sample size

$\chi^2 = \sum \frac{(O-E)^2}{E}$

O = observed

$H_0: \chi^2 = 0$

$H_A: \chi^2 > 0$

d.f. = k − 1 → # of categories in the distribution.

P640 Table 10-8

10.2 (6, 10, 13) + PPUSD

13) $E = 1215 \times 10\% = 121.5$ $\alpha = 0.01$

$1215 \times 0.03 = 36.45$ $\chi^2 = 13.6829$

$1215 \times 0.38 = 461.7$ pvalue = 0.0178

$1215 \times 0.41 = 498.15$ $\chi^2_c = 15.09$

$1215 \times 0.06 = 72.9$ d.f. = 6 − 1 = 5

$1215 \times 0.02 = 24.3$

Sample stat		population
s	s.d.	σ
s^2	varience	σ^2

Theorm 10.1 $\chi^2_{score} = \frac{(n-1)s^2}{\sigma^2}$ $H_0: \sigma^2 = k$ number

χ^2_c = → Table & d.f. $H_A: \sigma^2 \geq k$ → most, $\leq$ → some

10.3 (6, 8, 9) 4 : c.c.

scatterplot

9.2 15) $\Sigma x = 112.8$ $\Sigma y = 32.4$ $\Sigma x^2 = 2167.14$ $\Sigma y^2 = 90.14$ $\Sigma xy = 665.3$ $r \approx 0.764$

$x = 24$ $y = ?$

$y = a + bx$ $a = -17.204$ $b = 1.202$

stat → CALC → $\boxed{8}$ → ① $r^2 = 0.5836$ $r = 0.7640$

$-1 \le r \le 1$

ANA = -+/- Assocation

-As ...

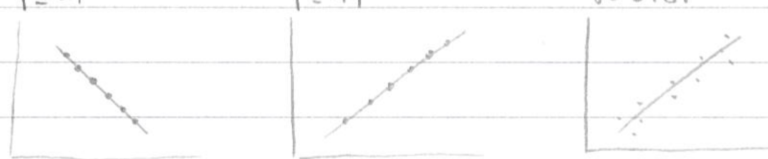
- Influential (24, 13)

outlier (

- r = correlation.

r^2 = coefficient of Determination

r = -1 r = +1 r = 0.81



r = 0 r = 0

r = 0.7640 Moderatly strong linear correlation

$r^2 = 0.5836$ 58.36% of variation in y (violent crime) is explained by variation in x (High no school) the Least Square Regression Line.

CALCULATOR → r & r^2 → mode 2 STAT DIAGNOSTICS

trend line =

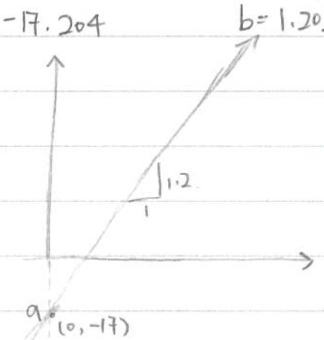
From ① → y= → vars → 5= → EQ → 1= → Zoom → 9=

point =

$\boxed{2nd}$ → trace → calc → 1= → x = ?

a = y-intercept b = sloop

a = -17.204 b = 1.202



$\hat{y} = a + bx$

LSRL

$\hat{y} = -17.20 + 1.20x$

$\hat{y} = -17.204 + 1.202(24)$

$\hat{y} = 11.64$.

ρ = "rho" population correlation, coefficient.

β = "Beta" population sloop

9.3 (7, 9, 10) Test (only b & f) (ex. g) Thurs (9 & 10)

9 in class.

confidant Interval of β → Data in List → stat → TEST → G=

1	1998	26
2	1986	38
3	2015	9
4	1997	27
5	2012	12
6	1996	28
7	2019	$\boxed{5}$
8	1989	35
9	2014	10
10	2000	24
11	2010	14 2
12	2017	7
13	2002	22
14	2016	8
15	1992	32
16	2006	18
17	2016	8
18	2019	$\boxed{5}$
19	2012	12
20	2006	18
21	2021	3
22	2016	~~12~~ 8
23	2000	24
24	2014	10
25	2012	12

median ≈ 15

mean: > 15 ≈ 18, 19, 20

spread: 12~14 High spread

12 $\quad 150/10 = \cancel{150}$. 15

14

12

18

12

28

24

8

10

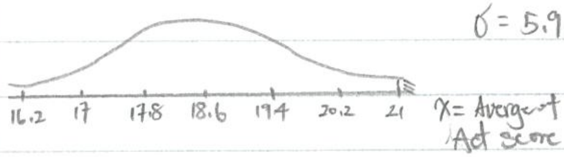
12

ACT

$\mu = 18.6 \quad \sigma = 5.9 \quad n = 50 \quad P(\bar{x} > 21) \approx 0.0020$

$\sigma = 5.9/\sqrt{50} = 0.8344$

There exit a 0.2% chance that 50 student score mean is higher than 21



$\mu = \Sigma x \cdot P(x)$

$1,000,000 / day × 365 → 365,000,000 → 3,650,000,000

11 by hand

9.2 (10~12, 15)

Nov. 9th Quiz 9.1~9.2

93% => 4SD.

$\bar{x} = 20.35$

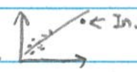
$20.35 + (4 \times 9.79) = 59.51$

$20.35 - (4 \times 9.79) = -18.81$

At least 93.8% of the data is captured between -18.81 to 59.51 cm/gm/sec.

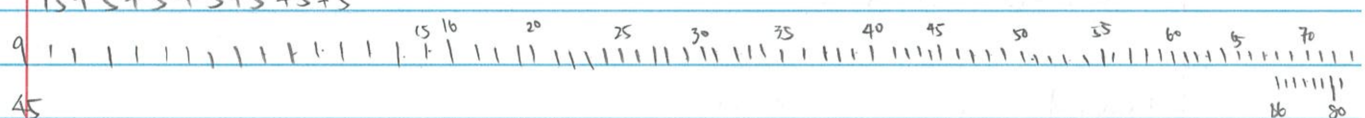
HT WS
62'' 59''

Analysis Ex
- + Association / + linear association
- Trend (with word) : the greater the wing spans, the higher the height.
- outlier (can point) / outlier : (60, 80)
  Influensial Score. = the point is follow the trend, but "out" of the data.



- Correlation : Not Now.

15 + 5 + 5 + 5 + 5 + 5 + 5 + 5



45

9.2 #11 $b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} = \frac{7(4015) - (329)(115)}{7(18,263) - (108,241)} = -0.4964$

$\Sigma x = 329$ $\bar{x} = 47$ $\Sigma x^2 = 18,263$

$\Sigma y = 115$ $\bar{y} = 16.43$ $(\Sigma x)^2 = 108,241$

$\Sigma xy = 4015$

$a = \bar{y} - b\bar{x} = 16.43 - (-0.4964)(47) = 39.76$

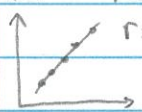
$\hat{y} = 39.76 - 0.4964x$

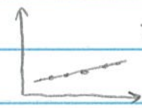
f) $\hat{y} = 39.76 - 0.4964(25) = 27.35\ \%$

Formula way*
↑
9.1 (13, 15, 18) ⇐ easier way

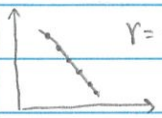
Correlation (Karl Pearson)

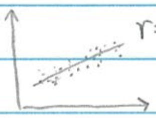
(r). $-1 \le r \le 1$ $r \ne b$.

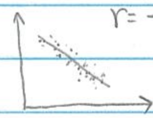

$r = 1$

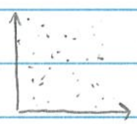

$r = 1$ $b \to$ smaller

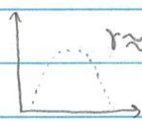
Perfect Linear Correlation
"PLC"


$r = -1$ Perfect Linear negative Correlation.
"PLnC"


$r = 0.8$ or 0.9 ....
"strong "+" LC"


$r = -0.8$ or -0.9
"Strong "-" LC"


$r \approx 0$


$r \approx 0$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

Ditermanation = r^2 $0 \le r^2 \le 1$

$(-0.969)^2 = 0.9490$

$R^2 = 0.9490 = 94.90\%$ of the variation in y (# of muggings) is explaind by the LSRL and the variation in x (# of UPOs).

13) $\Sigma x = 63$ $(\Sigma x)^2 = 3969$

$\Sigma y = 650$ $(\Sigma y)^2 = 422{,}500$

$\Sigma x^2 = 1089$

$\Sigma y^2 = 95{,}350$

$\Sigma xy = 9930$

$$r = \frac{5(9930) - (63)(650)}{\sqrt{5(1089) - 3969}\ \sqrt{5(95{,}350) - 422{,}500}} = \frac{8700}{(38.42)(232.92)} = 0.9722$$

Strong Positive Linear Correlation

$r^2 = 0.9453 = 94.53\%$ of the variation in y (weight of the pony) is explan by the least squre regresson line and the variation in x (age of the pony)

Joyce Li

Period 4.

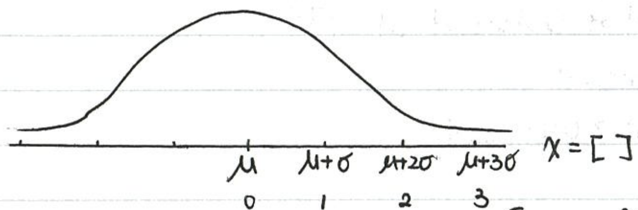
$\mu=?$ $\sigma=?$ ② have x find area% ③ have area% find x.

6.3 ② $P(?<x<?)=(\ )\%$ ① calc: 2nd → distr → $\boxed{2}$ normalcdf. ∃x ( )% chance that [ ] is [</>/?<?]

③ find x calc: 2nd → distr → $\boxed{3}$ invNorm.

6.5 A random sample

$\mu=?$ $\sigma_{\bar{x}}=\sigma/\sqrt{n}$ $P(?<\bar{x}<?)=(\ )\%$ ① ∃x ( )% chance that a sample of [ ] is [</>/?<?/]



68% - 34%

95% - 13.5%

99.7% - 2.35%

$Z=\frac{x-\mu}{\sigma}$ or $Z=\frac{x-\mu}{\sigma/\sqrt{n}}$

& 7.2

7.1 necessary sample size = $n=\left(\frac{Z_c\sigma}{E}\right)^2$ $\bar{x}-E<\mu<\bar{x}+E$ (unit!)

$E=z$ or $t \times \left(\frac{\sigma \text{ or } s}{\sqrt{n}}\right)$ calc: stat → TEST → $\boxed{7:}$ZInt / $\boxed{8:}$TInt.

7.3 necessary sample size: $n=pq\left(\frac{Z_c}{E}\right)^2$ $\hat{p}-E<p<\hat{p}+E$

$E=z \times \left(\sqrt{\frac{pq}{n}}\right)$ calc = stat → TEST → $\boxed{A:}$ 1-PropZInt.

If we took x sample we expect to catch population mean μ (or) population % p of [ ] y times. (% for p, n= ...)

7.4 Z-T Interval of Difference of 2 Mean / Proportion (Ind) 80% 1.28 95% 1.96

< p_1-p_2 < # calc: stat → TEST → $\boxed{9}$ $\boxed{0}$ $\boxed{B}$ 90% 1.645 99% 2.58.

$\mu_1-\mu_2$

$- \to + \Rightarrow \mu_1 \approx \mu_2$ $p_1 \approx p_2$ [ ] is equal to / less than / greater than [ ].

$- \to - \Rightarrow \mu_1 < \mu_2$ $p_1 < p_2$

$+ \to + \Rightarrow \mu_1 > \mu_2$ $p_1 > p_2$

8.1~8.3 Z/T-Test of 1-mean/p (Ind) 8.4. T-Test of Dep Mean

calc: TEST → $\boxed{1}$ $\boxed{2}$ $\boxed{5}$ $H_0: \mu/p=$ calc: Ln (differences) → TEST → $\boxed{2}$.

8.5 Z/T-TEST of Difference of 2 Ind. μ/p.

calc: TEST → $\boxed{3}$ $\boxed{4}$ $\boxed{6}$ $\hat{p}=r/n$

9.1 & 9.2 $\hat{y}=a+bx$ and r, r^2 calc: stat → CALC → $\boxed{8:}$ mode → STAT DIANOSTCS.

Ana: +/- ~~Associ~~ Association -r = x (strong...) +/- linear correlation

Influential / Outliers. $-r^2 = x$ % of variation y ( ) is explained by LSRL and variance in $\hat{y}$ ( ).

9.3 T-Test of Population correlation ρ and β.

$r \to t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ calc = # in List → stat → TEST → $\boxed{F:}$. Ho: $\rho = 0$ x[ ] & y[ ] have no Correlation.

b ↗ β Interval : ↘ $\boxed{G:}$. $\beta = 0$ Population slope of LSRL is flat.

10.1 Chi-Square Test of Ind. $\boxed{\text{H}}$

Calc: $\boxed{2nd}$ → matrix → Edit.

stat → TEST → $\boxed{C}$

$E = \frac{\text{rowtotal} \times \text{column}}{\text{sample total}}$

d.f. = (row−1)(column−1)

$\chi^2 = \Sigma \frac{(O-E)^2}{E}$

10.2 chi-Square Test of Goodness of fit. #

d.f. = # of categoriess−1

calc = stat → edit → O=L1, E=L2

TEST → $\boxed{D:}$

E = % × n

$\chi^2 = \Sigma \frac{(O-E)^2}{E}$

10.3 Chi-Square Test of Variance

Ho = σ^2 = #

$\chi^2_{scr} = \frac{(n-1)s^2}{\sigma^2}$

Calc = $\boxed{2nd}$ → distr → $\boxed{8:}$ (p-value).