MADDY KUBIK
PERIOD 3
AP

MEASURES OF CENTRAL TENDENCY

- MEAN, MEDIAN, MODE, TRIMMED MEAN
- ARITHMETIC MEAN ($\overline{x}$) = sample (μ) = population
  - usually meant when one says average
  - sum / # of items
  - easily affected by outliers and pulled in that direction
- MEDIAN
  - middle value when ordered small → large or large → small
  - position is emphasized not numeric values, so it is resistant to outliers (advantage over mean)
- MODE
  - most frequently repeated value
  - often most appropriate (average shoe size)
  - can be bi/trimodal, may be no mode
  - outliers don't affect
- TRIMMED MEANS
  - mean that resists extremes
  - eliminates pull of extremely low or high values in data set
  - find 10% trimmed mean = don't count bottom and top 10% of data

$\sum$ = capital sigma = sum

POP STANDARD DEVIATION

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

SAMPLE STANDARD DEVIATION:

$$s = \sqrt{\frac{\sum (x-\overline{x})^2}{n-1}}$$

MEASURES OF VARIATION

- One number may not represent an entire set of numbers well so, a cross reference is the measure of dispersion/variation/fluctuation/spread
- 3 measures of variation = range, standard deviation, variance, coefficient of variation

- advantages/disadvantages of the range = tells difference between the largest and smallest value, but does not tell us how much other values vary from one another or the mean
  ↓ easily computed
- the measurement that helps us see how the data is different from the mean is standard deviation
- why divide by n-1 sometimes and n others?
  - n-1 ↓ sample standard deviation (s)
  - n ↑ population standard deviation (σ)
- mean and standard deviation go hand in hand, NOT median

MADDY KUBIK
PERIOD 3-AP

3.3 BOX PLOTS

MEDIAN = 50^{th} Percentile
↓ a value at which x percent of data values falls at it or below it

PAGE 112 #3 - Not necessarily because the scores all could have been extremely high which would make 82 a non-passing score.

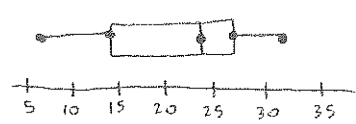
QUARTILES - special percentile, divide data into 4^{th}'s
- 1st = Q_1 / LQ (lower quartile) = 25^{th} percentile
- 2nd = Q_2 = 50^{th} percentile = MEDIAN
- UQ = upper quartile → 3rd = Q_3 = 75^{th} percentile

FIND QUARTILES:
1. Arrange numbers small → large
2. Median = Q_2
3. Look at data left of median, find median of subgroup = Q_1
4. Look at data right of median, find median of subgroup = Q_3

INTERQUARTILE RANGE = $Q_3 - Q_1$ → resistant to outliers
(measures spread of middle)

BOX AND WHISKER PLOTS:
1. # line
2. median = •
3. LQ UQ = • •
4. small, large = • •
5. draw box + lines

GENERAL BOX/WHISKER



CENTER = 23
SHAPE = SKEWED LEFT
SPREAD = IQR = 27-18 = 9
OUTLIERS = no low outliers
no high outliers

PAGE 114
#10 = HOW

Chapter 4

MADDY KUBIK
PERIOD 3

4.2 notes

LINE OF BEST FIT OR LEAST SQUARES LINE

$y_p = a + bx$

or $\hat{y}$ — y-int — slope

mean

$a = \bar{y} - b\bar{x}$

$b = \dfrac{SSxy}{SSx}$

super sum

$SSxy = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

$SSx = \sum x^2 - \dfrac{(\sum x)^2}{n}$

n — number of points in plo

LINE OF BEST FIT: sum of the squares of the vertical distances from the points to the line be made as small as possible

1. ORGANIZED TABLE
2. FIND SSx
3. FIND SSxy
4. FIND $\bar{x}$
5. FIND $\bar{y}$
6. FIND b (slope)
7. FIND a (y-int)
8. $\hat{y} = a + bx$
9. GRAPH

x	y
$\bar{x}$	$\bar{y}$
0	a
x	$\hat{y}$

→ use if can't use $(\bar{x}, y)$ or $(0, a)$

PREDICTIONS: interpolation = GOOD
extrapolation = BAD

GUIDED EXERCISE 3 = GOOD EXAMPLE
PAGE 138

PAGE 144 #6

x	y	xy	x^2
16.2	7.2	116.64	262.44
9.9	8.8	87.12	98.01
19.5	7.9	154.05	380.25
19.7	8.1	159.57	388.09
9.8	10.3	100.94	96.04

$\sum x = 75.10$ $\quad \sum y = 42.30$ $\quad \sum xy = 618.32$ $\quad \sum x^2 = 1{,}224.83$

$\dfrac{\sum x}{n} = \bar{x} = 15.02$ $\quad \dfrac{\sum y}{n} = \bar{y} = 8.46$

$(\sum x)^2 = 5640.01$

$b = \dfrac{SSxy}{SSx} = \dfrac{-17.03}{96.83} = -.1759$

$SSxy = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

$SSxy = 618.32 - \dfrac{(75.10)(4}{5}$

$SSxy = -17.03$

$SSx = \sum x^2 - \dfrac{(\sum x)^2}{n}$

$SSx = 1{,}224.83 - \dfrac{(75.10)^2}{5}$

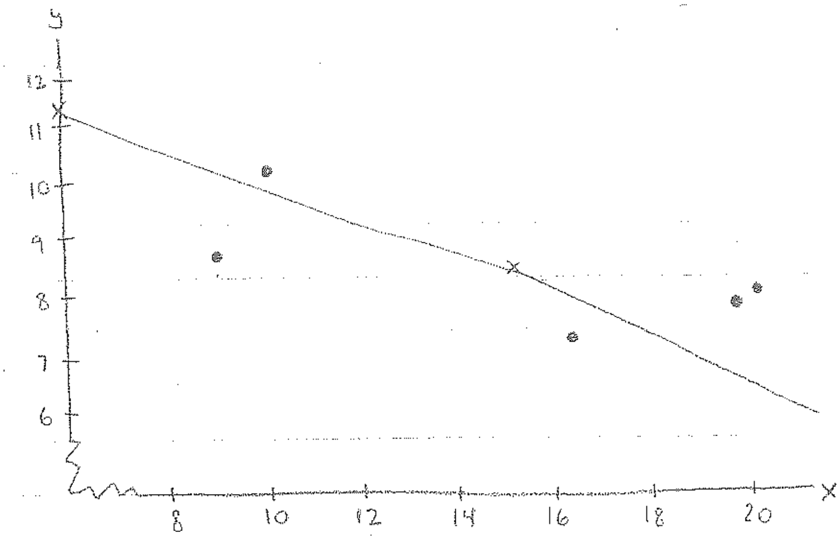
$SSx = 96.83$

$a = \bar{y} - b\bar{x}$

$a = 8.46 - (-.1759)(15.02)$

$a = 11.10$

x	y
15.02	8.46
0	11.10



plug in

$\hat{y} = 11.1 + (-.1759x)$

d) $x = 17$

$\hat{y} = 8.11 = \$8,110$

4.3 HW = (1-5h), 6 or 8

MADDY KUBIK
PERIOD 3

4.3 COEFFICIENT OF CORRELATION (r)

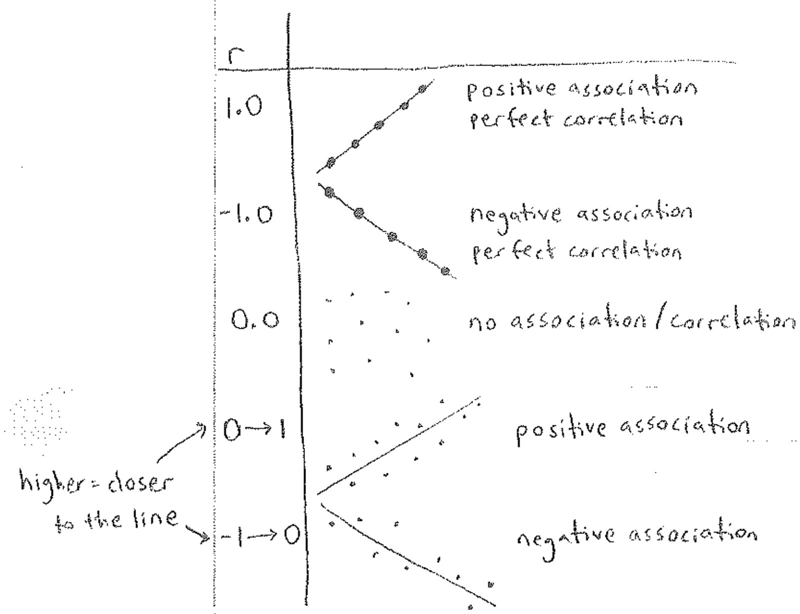
- CARL PEERSON - 1856

$$r = \frac{SSxy}{\sqrt{SSx\,SSy}}$$

$$-1 \le r \le 1$$

$$SSxy = \Sigma xy - \frac{\Sigma x \Sigma y}{n} \qquad SSx = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$SSy = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$



COEFFICIENT OF DETERMINATION: (r^2)
- another way of determining the strength of relationship between x + y

- must classify r → low, mod, high
- sentence for r^2 → 44.61% of the variation in the y (GPA) is explained by the Least Squares Regression Line and variation in the x (IQ).

↓

This means that 55% is unexplained by IQ.

↓

GPA can be affected by family life, motivation, extracurriculars, etc.

PRACTICE PROBLEM ⟶

PAGE 156 # 8

x = age	y = % fatal accidents	xy	x^2	y^2
37	5	185	1369	25
47	8	376	2209	64
57	10	570	3249	100
67	16	1,072	4489	256
77	30	2,310	5929	900
87	43	3,741	7569	1849
$\Sigma x = 372$	$\Sigma y = 112$	$\Sigma xy = 8,254$	$\Sigma x^2 = 24,814$	$\Sigma y^2 = 3,194$
$\bar{x} = 62$	$\bar{y} = 18.67$			
$(\Sigma x)^2 = 138,384$	$(\Sigma y)^2 = 12,544$			

$r = \frac{SS_{xy}}{\sqrt{SS_x \times SS_y}}$

SS_{xy}: $8254 - \frac{(372)(112)}{6} = 1,310$

SS_x: $24,814 - \frac{138,384}{6} = 1,750$

$SS_y = 3,194 - \frac{12,544}{6} = 1,103.3$

$\sqrt{(1,750)(1,103.3)} = 1389.52$

$r = .9428$ = high

$r^2 = 88.89\%$ of the variati[on] in the percent of fatal accid[ents] is explained by the LSRL a[nd] variation in the age of drive[rs]

Chapter 5

MADDY KUBIK
11-30-10
PERIOD 3

5.1 INTRO TO PROBABILITY

- PROBABILITY - measure of chance
- P(A) = Probability of event A occurring "P of A"
- $\frac{\text{\# of desired results}}{\text{total \# of results}}$ = Probability

Favorable : non-Favorable = ODDS

- $0 \le P \le 1 \rightarrow$ certain
  ↓ no chance - impossible

chances of yellow marble
P(y) = $0/6$ = impossible
chances of marble
P(m) = $6/6 = 1$ = certain
chances of black = $2/6 = 1/3 = .33$

P(U|F) = probability of you GIVEN THAT you are female

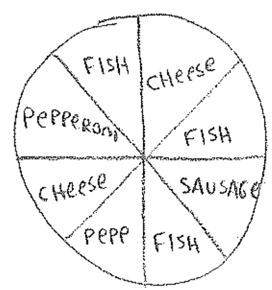
- ROLLING A DIE:
  - P(3) = $1/6$ = 16.7%
  - P(even) = $3/6 = 1/2$
  - P(5+) = $2/6 = 1/3$
  - P(1') = $5/6$
    not a 1

P(not A) = P(not event) = "Complement of an event"
P(A')
P(A) + P(A') = 1
So, P(A') = 1 - P(A)

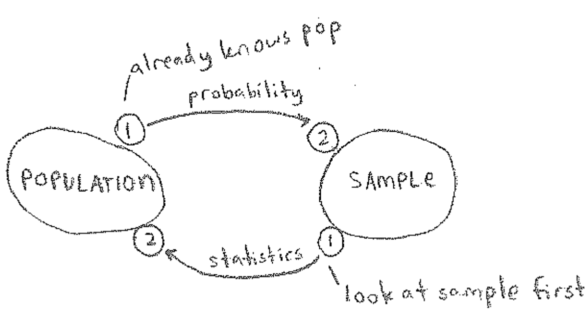
1 - P(A)
1 - P(1)
1 - $1/6$ = $5/6$

- SAMPLE SPACE OF DIE:
  list of outcomes
  S = {1, 2, 3, 4, 5, 6}

PIZZA:



P(cheese) = $2/8 = 1/4$
P(fish) = $3/8$
P(meat) = $6/8 = 3/4$
P(cheese') = $1 - 1/4 = 3/4$



- Fermont and Pascal = introduced probability
  ↓ French

STANDARD DEVIATION EXAMPLES

SAMPLE

① 2, 4, 6, 8, 10

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \quad n-1 \downarrow 5-1$$

$\bar{x} = \frac{30}{5} = 6$

$x-\bar{x}$	$(x-\bar{x})^2$
-4	16
-2	+ 4
0	+ 0
2	+ 4
4	+ 16

$40 \div n-1 = 20 \div 4 = 10 \rightarrow \sqrt{10} = \boxed{3.16}$

$\sqrt{s^2}$

POPULATION

② 19.8, 43.8, 36.1, 52.4, 63.1, 28.7, 46.3

$\mu = 41.46$

$$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{n}} \quad n \uparrow 7$$

$x-\mu$	$(x-\mu)^2$
-21.66	469.16
2.34	5.48
-5.36	28.73
10.94	119.68
21.64	468.29
-12.76	162.82
4.84	23.43

$\sum = \frac{1,276.72}{7}$ $\quad \sigma^2 = 182.5$

$\sqrt{182.5} = \boxed{13.5}$

- disadvantage of standard deviation = depends on unit of measurement

COEFFICIENT OF VARIATION:

xpresses standard eviation as a % of ample or pop mean

SAMPLE $= \frac{s}{\bar{x}} \cdot 100$

POP $= \frac{\sigma}{\mu} \cdot 100$

CV RATINGS

when all #'s are same → 0-13% = tight / cluster / banded / low

15-33% = moderately low

35-65% = moderate

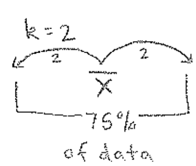
67-90% = moderately high

90-100% = scattered / wide / high

100+% = data gone WILD

CHEBYSHEV'S THEOREM

- For ANY set of data and for any constant k greater than 1, the proportion of the data that must lie within k standard deviations on either side of the mean is at least $1 - \frac{1}{k^2}$



$1 - \frac{1}{2^2} = \frac{3}{4} = 75\%$

Chapter 6

MADDY KUBIK
1-3-11
PERIOD 3

6.1 RANDOM VARIABLES

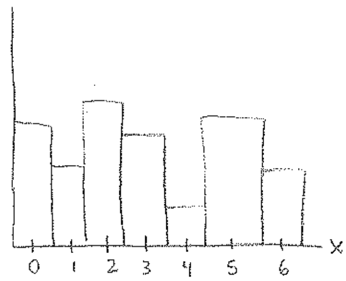
DISCRETE- quantatative observation that are countable (finite), no decimals/fractions (number students

CONTINUOUS- countless (infinite) observations, decimals + fractions + square roots (time, height, temp)

PROBABILITY DISTRIBUTION

DISCRETE: P(x)

- all probabilities add to 1 = entire sample space shown, mutually exclusive



- frequency distributions have means + standard deviations

x = value of random variable

$\mu = \sum x \cdot P(x)$ = expected value $\quad \sigma = \sqrt{\sum (x-\mu)^2 P(x)}$

example

number of times buyers view infomercial before buying

x (# viewings)	$P(x)$	$xP(x)$	$x-\mu$	$(x-\mu)^2$	$(x-\mu)^2 P(x)$

$\mu = \sum x P(x)$ $\quad \sigma = \sqrt{\sum (x-\mu)^2 P(x)}$

MADDY KUBIK
1-5-11
PERIOD 3

6.2 BINOMIAL EXPERIMENTS
(aka Bernoulli)

PAGE 225

1. Who was Jacob Bernoulli? - Swiss mathmatician who studied binomial experiments in late 1600's
2. Problems with exactly 2 outcomes: Binomial (Bernoulli) experiments
3. Describe the central problem of a binomial experiment: to find probability of r successes out of n trials
4. Binomial experiments work only with dependent situations, T/F? - FALSE (only Independent)
5. Each faculty member has been asked to recommend a car to micek.
   # of trials, n = number of people asked
   # of possible outcomes = number of types of cars
   Binomial experiment? - no

PRACTICE

1. 80% of the time a nurse can respond in 3 min
   73 room calls last night
   find P(nurses respond to 62 in 3 min)

trial = room call
s = nurse comes within 3 min
f = nurse comes after 3 min
n = 73
r = 62
p = .80   anti-p = q = 1-.80 = .2

BINOMIAL EXPERIMENT if independent and the same conditions each trial

2. travel agent has 4 different packages
   99% enjoy Hawaii package
   95% Europe
   96% Alaska
   97% New York
   sold 51 last month
   find P(enjoying 43)

trial = going on vacation
s = enjoying
f = not enjoying
n = 51
r = 43
p = no constant probability of success

NOT BINOMIAL no identical conditions, not independent

$P(r) = C_{n,r} p^r q^{n-r}$

binomial coefficient

EXAMPLE 4 PAGE 230

79% concerned about privacy on Internet

find p(6) out of 10 users

$p(6) = C_{10,6}(.79)^6(.21)^{10-6}$

$p(6) = (210)(.2431)(.0019)$

$p(6) = \boxed{.09928}$

EXAMPLE 5 PAGE 231

find probability that AT LEAST 8 seeds will germinate out of 10

* use Table 3 so you don't have to do a 4-in-1 problem

find n=10, on page A4

look over to probability of .70

add p(8)+p(9)+p(10)

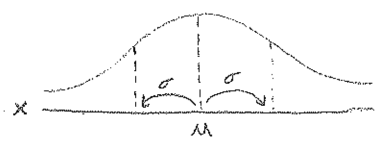
write sentence once answer is found - 38.20% chance that at least 8 seeds will germinate.

Chapter 7

MADDY KUBIK
1-13-11
PERIOD 3

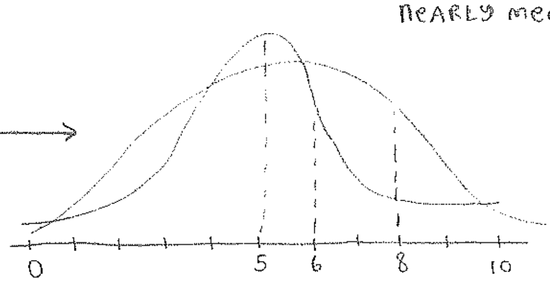
7.1 CONTINUOUS RANDOM VARIABLES AND THE NORMAL CURVE

- NORMAL CURVE AKA BELL CURVE OR GAUSSIAN CURVE (CARL GAUSS)



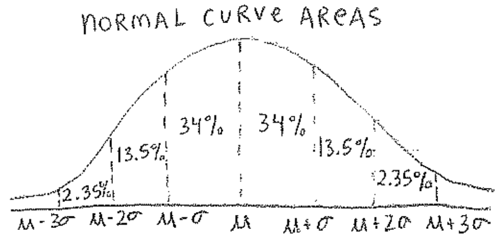
SYMMETRICAL
PEAK AT μ - mean/median/mode
OUTLIER = 3 σ above or below the μ
NEARLY MEETS X-AXIS BUT NEVER TOUCHES
AREA UNDER THE CURVE ADDS TO 1

DIFFERENT SHAPES



$\mu = 5$ $\quad$ $\mu = 5$
$\sigma = 1$ $\quad$ $\sigma = 3$
SHARP $\quad$ GRADUAL SLOPE

NORMAL CURVE AREAS

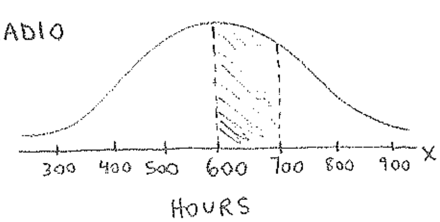


EMPIRICAL RULE:

For a symmetrical bell-shaped curve... 68% of data values will lie within one standard deviation above or below μ.
95% = 2 below or above μ.
99.7% = 3 below or above μ.

PAGE 260 EXAMPLE 1

SUNSHINE RADIO

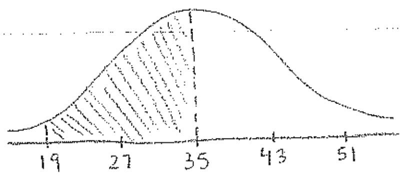


$P(600 \leq x \leq 700) = .3413 = 34.13\%$ chance that a radio lasts between 600 and 700 hours.

PAGE 261-GUIDED EXERCISE 4

$\mu = 35$ BUSHELS

$\sigma = 8$ BUSHELS



$P(19 \leq x \leq 35) = .475 = 47.5\%$ chance

MADDY KUBIK
1-18-11
PERIOD 3

7.2 NOTES

COMPARING NORMAL CURVES

example:

TEST SCORES

MATT = CHEMISTRY, $x = 76$, $\mu = 70$

JOHN = HISTORY, $x = 95$, $\mu = 86$

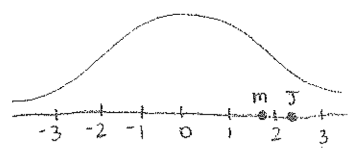
Z-SCORES: $z = \frac{x-\mu}{\sigma}$ or $\frac{x-\bar{x}}{s}$

$z_{MATT} = \frac{76-70}{3.1} = 1.935$

HOW MANY STEPS AWAY

$z_{JOHN} = \frac{95-86}{4.2} = 2.143$

Z-SCORE STANDARD CURVE



MICEK'S COMMUTE TO BIOLA:

If given Z-score, but not x (raw value)
$x = z\sigma + \mu$

$\mu = 44$ minutes

$\sigma = 9$ minutes

Tuesday A.M = 78 minutes

$z = \frac{78-44}{9} = 3.78$ = almost 4 standard deviations

Thursday A.M = 37 minutes

$z = \frac{37-44}{9} = -.78$ = almost 1 below mean

TRANSFORMATION OF NORMAL DISTRIBUTION TO STANDARD NORMAL DISTRIBUTION

σ is the unit of length on z axi

$\mu = 8$

$\sigma = 2$

Shift curve so μ is 0

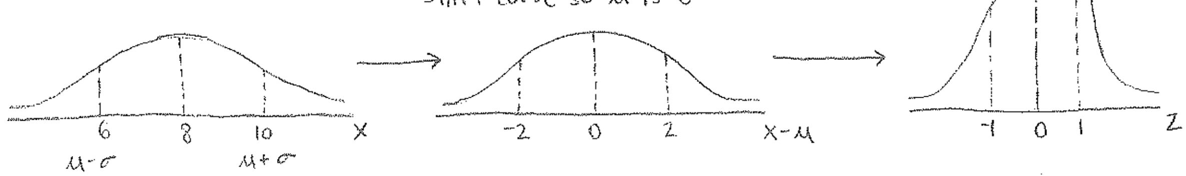


TABLE 4: Z-TABLE

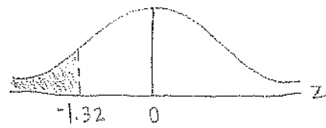
Z-TABLE PROVIDES AREA TO THE LEFT OF Z

PAGE 281: IMPORTANT

EXAMPLES:

PAGE 286 #11

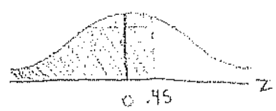
left of $z = -1.32$



$\boxed{.0934}$

#13

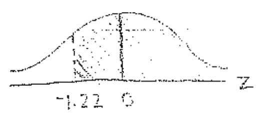
left of $z = 0.45$



$\boxed{.6736}$

#17

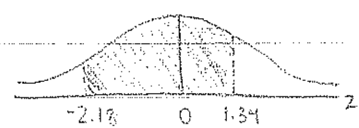
right of $z = -1.22$



$1 - .1112 = \boxed{.8888}$

#23

between -2.18 and 1.34



$-2.18 = .0146$ $\quad .9099 - .0146 = \boxed{.8953}$

$1.34 = .9099$

MADDY KUBIK
1-20-11
PERIOD 3

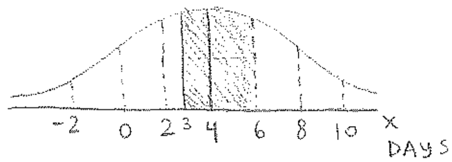
7.3 NOTES - MORE NORMAL CURVES

- DISCRETE can be graphed on normal curve too
- EXAMPLES: *ASSUMING normal distribution

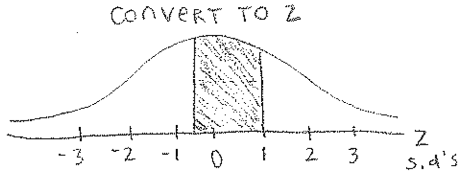
$\mu = 4$ days

$\sigma = 2$ days

$P(3 \leq x \leq 6)$



CONVERT TO Z



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

$Z = \frac{3-4}{2} = -.5 \rightarrow .3085$ $\quad \frac{6-4}{2} = 1 \rightarrow .8413$

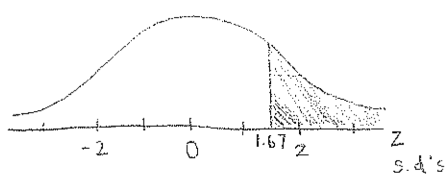
$.8413 - .3085 = \boxed{.5328}$ - 53.28% that a college student picked at random works out between 3 and 6 days a week.

PAGE 296 #23

ACT: $\mu = 18$

$\sigma = 6$

d) $P(x > 28)$

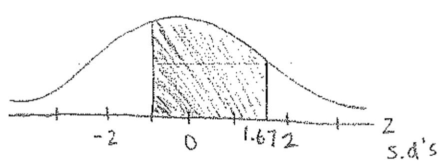


$Z = \frac{28-18}{6} = 1.67 \rightarrow .9525$

$1 - .9525 = \boxed{.0475}$ - 4.75% chance that a H.S Senior scores 28 or higher on ACT math assuming normal distribution.

f) $P(12 \leq x \leq 28)$

$Z = \frac{12-18}{6} = -1 \rightarrow .1587$



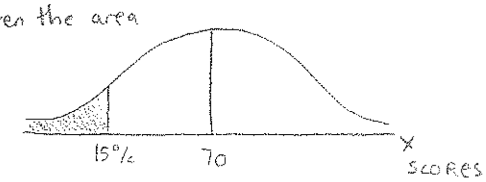
$Z = \frac{28-18}{6} = 1.67 \rightarrow .9525$

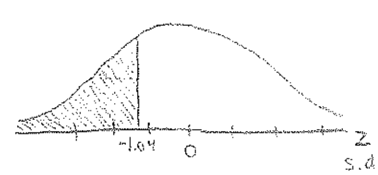
$.9525 - .1587 = \boxed{.7938}$ - 79.38% chance that a H.S Senior scores between 12 and 28 on ACT math assuming normal distribution.

INVERSE NORMAL DISTRIBUTION - given the area

$\mu = 70$

$\sigma = 9$





- look for .15 in z-table, find closest one .1492 → z-score = -1.04

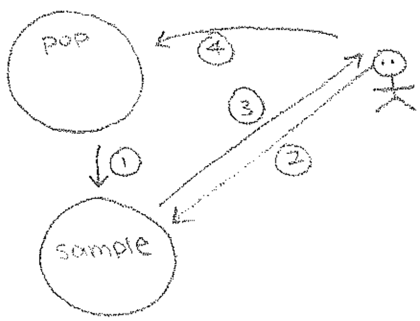
$-1.04 = \frac{x-70}{9}$ $\quad \boxed{x = 60.64}$

Chapter 8

8.1

Sampling Distributions

- Big ideas
- population - all measurements of intrest
- sample - a subset of the measurements from the population
- random sample - a representative sample, a sample that accurately reflects the population, it's what we are intereste in and worked in the projects
- draw the picture



1) obtain sample
2) look at sample
3) analyze sample
4) conclusion

- Sample Statistics and Population parameters
- statistic - numerical descriptive measure of a sample, a # that represents a sample ($\bar{x}$, s, s^2, $\hat{p}$)
- parameter - numerical descriptive measure of a population, a # that represents a population (μ, σ, σ^2, p or π, ρ, α, β)
- use sample statistics to infer about population parameters, don't have access
- make inference by estimation, decision
- Chapter 1-4, descriptive statistics, organize, summerize numbers
- Chapter 5-7, probability theory, distribution
- Chapter 8-10, inferential

collect	organize
analyze	interpret

MADDY KUBIK
2-7-11
PERIOD 3

8.1 THEOREM

USES SAMPLE:

- $\bar{x}$ distribution is a normal distribution
- mean of $\bar{x}$ distribution is μ
- standard deviation is $\frac{\sigma}{\sqrt{n}}$ = standard error

TO GET Z-SCORES:

$\mu_{\bar{x}} = \mu$

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

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

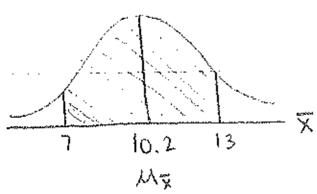
EXAMPLE:

$P(7 \leq \bar{x} \leq 13)$

$\mu = 10.2''$

$\sigma = 1.4''$

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



$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{7 - 10.2}{.6261} = -5.11$

$z = \frac{13 - 10.2}{.6261} = 4.47$

99.99% chance that the average of your catch will be between 7 and 13 inches.

probability goes up because σ decreases

8.2 CENTRAL LIMIT THEOREM

same formulas as 8.1 theorem

- if x distribution is not normal, $\bar{x}$ still NORMAL!
- works for big sample size ($n > 25$)

EXAMPLE PROBLEMS

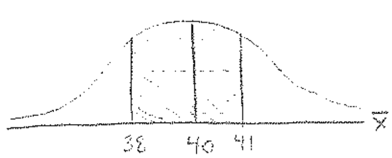
1. $n = 36$

Don't know if normal - use CLT

$P(38 \leq \bar{x} \leq 41)$

$\mu = 40$

$\sigma_{\bar{x}} = \frac{4.2}{\sqrt{36}} = .7$



$z = \frac{38 - 40}{.7} = -2.86 \rightarrow .0021$

$z = \frac{41 - 40}{.7} = 1.43 \rightarrow .9236$

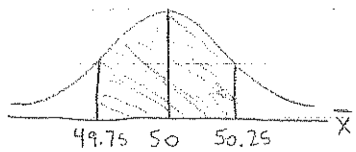
[illegible] that average time between 38 and 41 min

2. COFFEE

$n = 100$

$\mu = 50$ lbs

$\sigma_{\bar{x}} = .1$



$P(49.75 \leq \bar{x} \leq 50.25)$

98.76% chance that average coffee weight between 49.75 and 50.25 lbs.

$$Z = \frac{50.25 - 50}{.1} = 2.5$$

$$Z = \frac{49.75 - 50}{.1} = -2.5$$

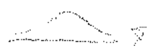
Central Limit Theorem

X distribution of data (original)

normal

$\downarrow$

$\bar{x}$ distribution (sample data)



$\mu_{\bar{x}} = \mu$

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

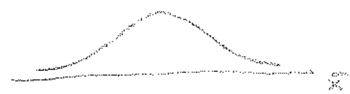
$z = \frac{\bar{x}-\mu}{\sigma}$

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

"Standard error"

non-normal or ??

$\bar{x}$ distribution (sample data)



$\mu_{\bar{x}} = \mu$

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

$z = \bar{x}-\mu_{\bar{x}} / (\sigma/\sqrt{n})$

8.2

central limit theorm

disappear

normal data skewed

+ $n = 30^{+}$ BIG

$\hookrightarrow 25$

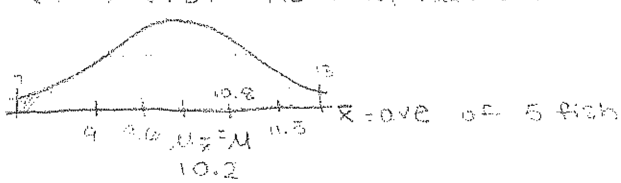
Theorm 8.1

works:

normal, n = ? big or small

P(average trout length 7 → 13") =

$P(7 < \bar{x} < 13) =$ normal cdf (7, 13, 10.2, .6261) = .99995 9638

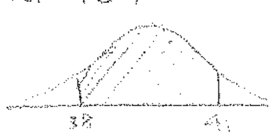


$\sigma = \sigma/\sqrt{n} = .6261$

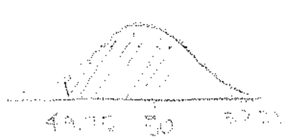
There exist a 99.99% chance that the kid at the fishing pond will take home an average of between 7 and 13"

$P(38 \leq \bar{x} \leq 41) =$ normalcdf (38, 41, 40, .7) = .9213

$\mu = 40$, $\sigma = 4.2/\sqrt{36}$



$P(49.75 \leq \bar{x} \leq 50.25) = .9876$



$\mu = 50$

$\sigma_{\bar{x}} = 1/\sqrt{100} = .1$

Chapter 9

MADDY KUBIK
2-15-11
PERIOD 3

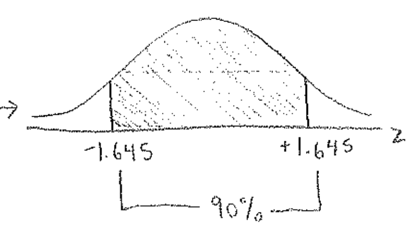
9.1 ESTIMATING μ WITH LARGE SAMPLES

POPULATION PARAMETERS: μ, σ, p
SAMPLE STATISTICS: $s, \overline{x}$ ← aka point estimates

$\sigma \approx s$ ← standard deviations
$\mu \overset{?}{\underset{?}{\approx}} \overline{x}$ ← error of estimate $= \overline{x} - \mu$

CONFIDENCE LEVELS - on z-table

LEVEL OF CONFIDENCE	Z_c
70%	1.04
75%	1.15
80%	1.28
85%	1.44
90%	1.645
95%	1.96
98%	2.33
99%	2.58



If $n \geq 30$, $E \approx Z_c \frac{s}{\sqrt{n}}$

standard error

max error tolerance for $\mu - \overline{x}$

- The confidence interval for μ = $\overline{x} - E < \mu < \overline{x} + E$
- Write sentence - We conclude with ___ confidence that the population mean μ is between ___ and ___.

EXAMPLE - 99% confidence

$s = 1.80$ min

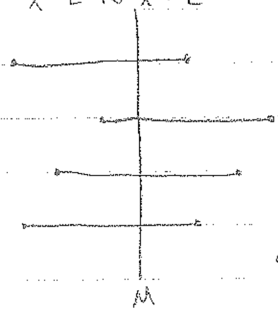
$\overline{x} = 15.60$ min

$E = 2.58\left(\frac{1.8}{\sqrt{90}}\right) = .4895$

$15.11 < \mu < 16.09$

We conclude with 99% confidence that the population mean μ is between 15.11 and 16.09 minutes.

$\bar{X} - E$ to $\bar{x} + E$



Read paragraph under Julia jogging problem page 346.

90% confidence means we will capture the real mean 9 times out of 10.

PAGE 352 #9

A) $s = 12.7$

$\bar{x} = 146.5$

B) 80% confidence - $Z_c = 1.28$

$E = (1.28)\left(\dfrac{12.7}{\sqrt{35}}\right) = 2.75$

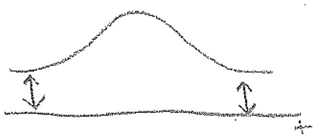
$143.75 < \mu < 149.25$

We conclude with 80% confidence that...

Estimating μ with Small Samples

- W.S. Gosset (worked at Guinness) - first to reconize the importance of developed (1908) statistical methods for small sample

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ $E = t_c(s/\sqrt{n})$ $\bar{x} - E < \mu < \bar{x} + E$ $d.f. = n-1$



thicker tails than z curve

- $n=14$, $df = 14-1 = 13$, $c = 90\%$, $t_c = 1.771$

$E = t_c(s/\sqrt{n}) = 1.771(.38/\sqrt{14}) = .1799$

$\bar{x} \pm E$ $1.43 - .1799 < \mu < 1.43 + .1799$ $1.2501 < \mu < 1.6099$

We conclude with 90% confidence that the population mean μ of depth perception is between 1.2501 and 1.6099 millimeters. We would capture the mean 9/10 times.

- $n=9$, $df=8$, $\mu = 327.67$, $s = 29.31$, $t_c = 1.86$

$E = t_c(s/\sqrt{n}) = 1.86(29.31/\sqrt{9}) = 18.1722$ $309.50 < \mu < 345.84$

We conclude with 90% confidence that the population mean μ of shoplifting attempts is between 309.5 and 345.84. We would capture the mean 9/10 times.

Estimating p in the Binomial Distribution

- $\mu \neq \mu$, $p = \pi$ = population parameters

- $\bar{x}$, $\hat{p}$, b, r = sample statistics
- $\hat{p} \to p$, binomial distribution, CIF, r/n = # of trials ⟵ # of success
- $E = z_c \sqrt{\hat{p}\hat{q}/n}$

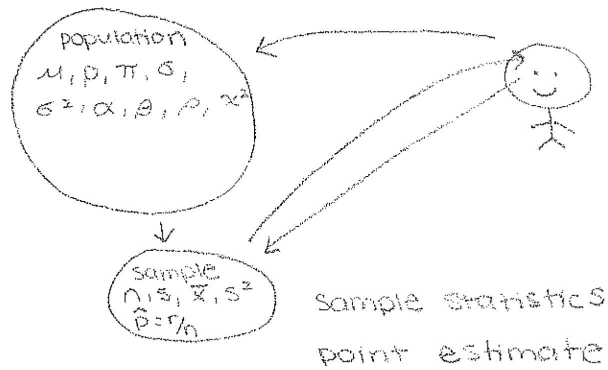
as sample size →, closer to normal

np>10 and nq>10 (✓ for largeness)

$\hat{p} \pm E$ ______ < p < ______

Confidence Intervals

$n \geq 30$



$\mu \approx \bar{x}$ (error in estimate)

$\sigma \approx s$

confidence level

$C = 0 \to 1$, .70, .75, .80, .85, .90, .95, .99

A if $n \geq 30$, $E = z_c \, s/\sqrt{n}$

max error tolerate — critical value of confidence

B $\bar{x} - E < \mu < \bar{x} + E$ $\quad \bar{x} \pm E$

the confidence interval for pop. mean(μ):

$E \approx z_c \, s/\sqrt{n}$; 99%

$= 2.58(1.8/\sqrt{90}) = .4895$

$15.60 - .4895 < \mu < 15.6 + .4895$

$15.11 < \mu < 16.09$

We conclude with 99% confidence that the population mean μ of jogging times for Julia is between 15.11 and 16.09 min.

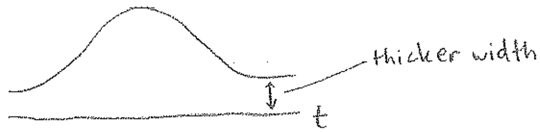
On a 99% confidence interval, if take 100 times, we would get μ 99 times or more.

MADDY KUBIK
2-17-11
PERIOD 3

9.2 NOTES- ESTIMATING μ WITH SMALL NUMBERS

1908 - W.S GOSSET - t distribution - GUINNESS BREWING CO

small samples - $n < 30$



$t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$ = t-score

$E = t_c \dfrac{s}{\sqrt{n}}$ = max error

$\overline{x} - E < \mu < \overline{x} + E$ = confidence interval

*use t-table that Mr. micek passed out

degrees of freedom: $n-1$

EXAMPLE:

(1) $\overline{x} = 1.43$ mm

$s = .38$ mm

$n = 14$

90% confidence

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

$1.25 < \mu < 1.61$

We conclude with 90% confidence that the population mean μ of errors in depth perception is between 1.25 and 1.61 mm.

(2) $\overline{x} = 327.6667$

$s = 29.3087$

$n = 9$

90% confidence

$E = (1.860)\left(\dfrac{29.31}{\sqrt{9}}\right) = 18.17$

$309.5 < \mu < 345.84$

We conclude with 90% confidence that the population mean μ of dollars of items shoplifted is between \$309.50 and \$345.84.

AGAINST ALL ODDS - VIDEO 19 - CONFIDENCE INTERVALS

- margin of error in voting polls - sampling error
- every sample slightly different from true population
- polls provide snapshot of what people are thinking

BLOOD PRESSURE EXAMPLE

$\bar{x} = 130$

assume - independent observations, normal distribution, and we know $s = 20$ points

margin of error: 15.2

95% confident - $\bar{x} - 15.2 < \mu < \bar{x} + 15.2$

↓

95% of time, confidence interval will encompass μ

* - higher confidence level, wider the interval
* - standard deviation increases, interval increases

MADDY KUBIK
2-22-11
PERIOD 3

9.3 NOTES - ESTIMATING P

LESLIE KISH = first one to use margin of error

POINT ESTIMATE FOR P:

$\hat{p} = \frac{r}{n}$ $E = z_c\sqrt{\frac{pq}{n}}$

- confidence interval for p
- large samples = $\underbrace{n\hat{p} > 5,\ nq > 5}_{\text{BOOK}}$ $\underbrace{n\hat{p} \geq 10,\ n\hat{q} \geq 10}_{\text{MICEK}}$

If these conditions are met, we can use normal z curve.

PAGE 366 - EXAMPLE 5

n = 800 r = 600

$\hat{p} = 600/800 = .75$

$\hat{q} = .25$

$np = (800)(.75) = 600 > 5$

$nq = (800)(.25) = 200 > 5$

99% confidence

$$E = (2.58)\sqrt{\frac{(.75)(.25)}{800}} = .0395$$

$0.71 < p < 0.79$

① We conclude with 99% confidence that the population proportion p...

② If we collected 100 samples we would capture population proportion p 99 times.

PAGE 372 #8

95% confidence

$np = (100)(.85) = 85 > 10$ $nq = (100)(.15) = 15 > 10$

*So, we can use z distribution

$$E = (1.96)\sqrt{\frac{(.85)(.15)}{100}} = .0699$$

$.78 < p < .92$

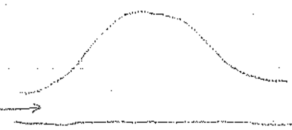
① We conclude with 95% confidence...

② If we collected 100 samples of 100 pharmacists, we would capture...

VIDEO 21 NOTES

REVIEW OF 9.2

- t-table = as it goes down, turns into z
- use t when σ is unknown and $n < 30$
- 1908 Guiness Brewery - Gosset
- t-curve, tails higher - larger probability of getting →



- degrees of freedom = $n-1$

Chapter 10

MADDY KUBIK
2-24-11
PERIOD 3

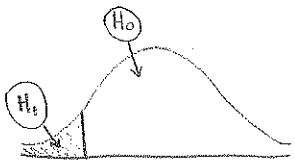
10.1 HYPOTHESIS TESTING

HYPOTHESIS - an assumption or belief about a parameter

HYPOTHESIS TESTING - procedure based on sample information, by which one "accepts" or "rejects" hypothesis

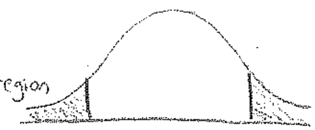
NULL HYPOTHESIS - H_0 - A hypothesis that is set up to test whether it can be rejected

ALTERNATE HYPOTHESIS - H_1 or H_a - the hypothesis to be accepted if H_0 is rejected



$H_1: \mu < k$

$H_1: \mu > k$



$H_1: \mu \neq k$

PAGE 394 EXERCISE 2

$H_0: \mu = 24$

$H_1: \mu > 24$

HYPOTHESIS ERRORS:

	ACCEPT H_0	REJECT H_0
H_0 TRUE	YAY ☺	TYPE 1 ERROR → putting innocent person in prison
H_0 FALSE	TYPE 2 ERROR → setting a guilty person free	YAY ☺

PAGE 399 #3

No we haven't PROVEN anythin. If it lands in shaded region, we fail to accept null.

LEVEL OF SIGNIFICANCE = α = P(reject H_0 when it is TRUE)

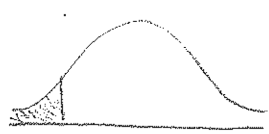
alpha → α = .01/.05/.10

$1-\alpha$ = .99/.95/.90 = confidence intervals

beta → B = P(reject H_0 when H_0 FALSE)

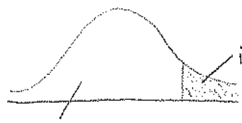
POWER OF A TEST = P(1-B)

GREAT HYPOTHESIS TEST:

1. H_0 - null hypothesis
2. H_1 - alternate hypothesis   picture 
3. critical value - Z_0 or t_0
4. convert test statistic to standard score (z, t)

MADDY KUBIK
3-4-11
PERIOD 3

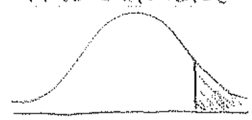
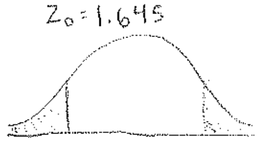
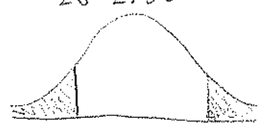
10.2 HYPOTHESIS TESTING WITH μ (LARGE SAMPLES)



if value lands in shaded, we accept H_1

if it lands in H_0 region, we fail to reject H_0

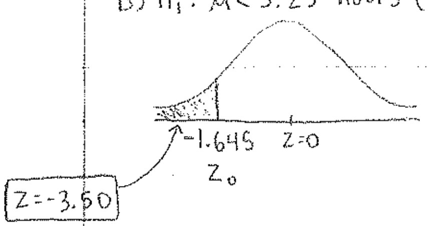
α = probability of rejecting H_0 when H_0 true

$\alpha = 0.05$	$\alpha = 0.01$
95% confidence	99% confidence
	
$Z_0 = 1.645$	$Z_0 = 2.33$
	
$Z_0 = \pm 1.96$	$Z_0 = \pm 2.58$

GUIDED EXERCISE 4:

A) $H_0: \mu = 5.25$ hours (sentence)

B) $H_1: \mu < 5.25$ hours (sentence)



$$Z = \frac{\bar{x} - \mu}{(\sigma/\sqrt{n})} = \frac{4.9 - 5.25}{(.6/\sqrt{36})} = -3.50$$

EXAMPLE:

OLD ROUTE - $\mu = 1.5$ hours $\sigma = 10$ minutes

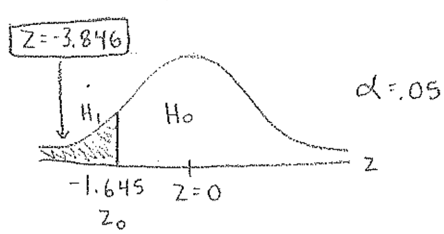
NEW ROUTE - $n = 41$ times $\bar{x} = 1.4$ hours

A) $H_0: \mu = 1.5$ hours - No difference in routes

B) $H_1: \mu < 1.5$ hours

C) $\bar{x} \to z$

$$\frac{1.4 - 1.5}{(.1\bar{6})/\sqrt{41}} = -3.846$$



We will accept H_1 at the significance level of .05. There exists sufficient statistical evidence to suggest that the new route is better than the old route.

EXAMPLE 4 - PAGE 408

$\mu = 8.5$ years

$n = 48$

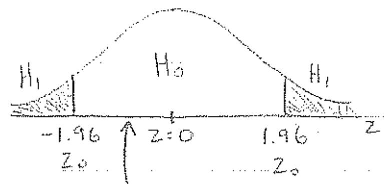
$\bar{x} = 7.91$ years

$s = 3.62$ years

$\alpha = .05$

$H_0 : \mu = 8.50$

$H_1 : \mu \neq 8.50$



$$z = \frac{7.91 - 8.50}{(3.62/\sqrt{48})} = \boxed{-1.13}$$ - We fail to reject H_0

EXAMPLE:

$\mu = 130°F$

$n = 81$

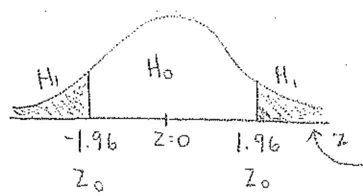
$\bar{x} = 131.08°F$

$\sigma = 1.5$

$\alpha = .05$

$H_0 : \mu = 130°F$ - no difference in the company's claim

$H_1 : \mu \neq 130°F$



$$z = \frac{131.08 - 130}{(1.5/\sqrt{81})} = \boxed{6.48}$$

We accept H_1 at the significance level of .05.

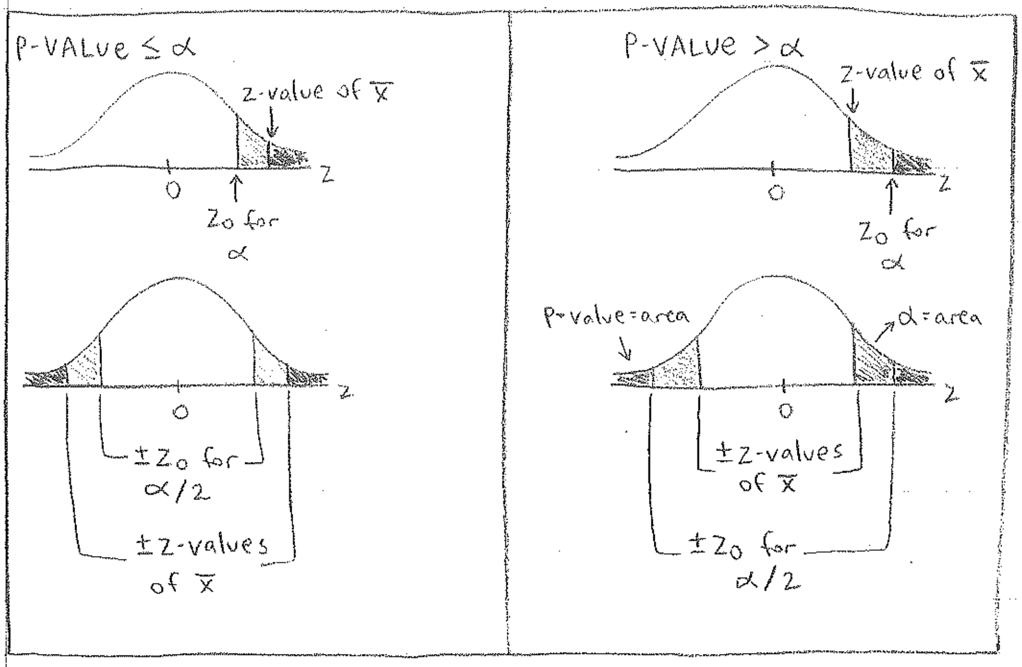
MADDY KUBIK
3-8-11
PERIOD 3

10.3 NOTES - HYPOTHESIS TESTING P-VALUES

P-VALUE - smallest level of significance for which observed sample statistic tells us to reject H_0.

P-VALUE $\leq \alpha$ = REJECT H_0

P-VALUE $> \alpha$ = DO NOT REJECT H_0



TWO-TAIL TEST = double value

FINDING P-VALUE FOR TESTS OF μ

very precise α → P-VALUE = P($\bar{x}$ computed from sample size n $\geq$ observed $\bar{x}$)

*areas in tail/tails beyond observed sample statistic

PAGE 418 EXAMPLE 6:

$H_0: \mu = 576$

$H_1: \mu > 576$

$n = 36$

$\bar{x} = \$615$

$s = \$120$

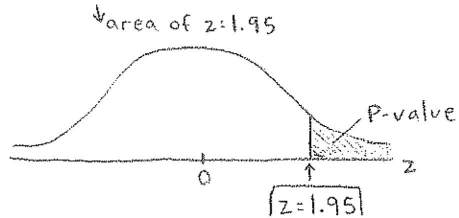
① convert $\bar{x} \rightarrow z$

$$z = \frac{615-576}{(120/\sqrt{36})} = 1.95$$

② P-value = $P(\bar{x} \geq 615) = P(z \geq 1.95)$

$= 1 - \underset{\text{area of } z=1.95}{0.9744} = 0.0256$

③ GRAPH



④ CONCLUDE:

old conclusion + → The P-value of 0.0256 is the smallest level of significance for which we reject Ho.

- STATISTICALLY SIGNIFICANT - if data lands far from H_0, we reject H_0
- NOT STATISTICALLY SIGNIFICANT - do not reject H_0

- P-VALUES tell us all levels of significance for which we reject H_0.
- P-VALUE $= 0.0358$

  * fail to reject H_0 for $\alpha = .01$ because P-value is greater than .01.

EXAMPLE:

$\mu = 41$ min

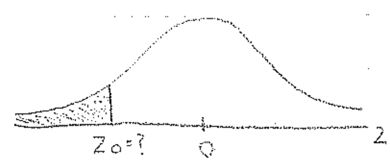
$n = 40$

$\bar{x} = 38$ min

$s = 11.2$ min

$H_0: \mu = 41$ - There is no change/no difference

$H_1: \mu < 41$



$$\frac{38-41}{11.2/\sqrt{40}} = -1.69 = z \qquad \text{p-value} = .0455$$

Reject H_0, accept H_1 at $\alpha = .05$

Accept H_0 at $\alpha = .01$

MADDY KUBIK
3-10-11
PERIOD 3
AP

10.4 NOTES - HYPOTHESIS TESTS OF μ (SMALL SAMPLES)

$n<30$

$$t=\frac{\bar{x}-\mu}{s/\sqrt{n}}$$

* If two-tail test, on t-table divide α by 2
  - if $\alpha=.05$, look at .025 on table

- Hypothesis test same as large samples but use t instead of z

PAGE 428 EXAMPLE 9

$H_o: \mu=5500$ pounds

$n=6$, $d.f=5$

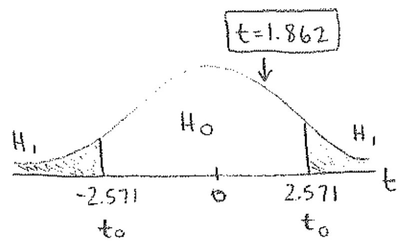
$\bar{x}=5690$ pounds

$s=250$

$H_1: \mu\neq5500$

$\alpha=.05$

$t_o=\pm2.571$



$$t=\frac{5690-5500}{250/\sqrt{6}}=1.862$$

* Fail to reject H_o at significance level of .05.
  PAGE 429 = good conclusion

PAGE 433 #7

$H_o: \mu=4.8$

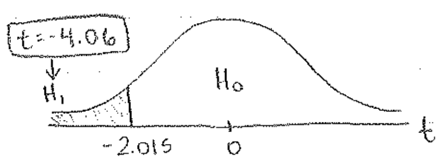
$n=6$, $d.f=5$

$\bar{x}=4.07$

$s=.44$

$H_1: \mu<4.8$

$\alpha=.05$



$$t=\frac{4.07-4.8}{.44/\sqrt{6}}=-4.06$$

P-value = between .0025 and .005

* Reject Ho and accept H_1 at the significance level of .05. There exists sufficient statistical evidence to suggest that the RBC count for this patient is lower than the average of 4.8 for healthy adult females.

PAGE 434 #10

$H_0: \mu = 8.8$ - there is no change in the average catch in Homser Lake

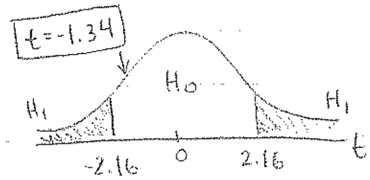
$n = 14$

$\overline{X} = 7.36$

$s = 4.03$

$H_1: \mu \neq 8.8$

$\alpha = .05$



$$t = \frac{7.36 - 8.8}{4.03/\sqrt{14}} = -1.34$$

p-value = double p because two-tail

$.2 < p < .3$

$.10 \times 2$ $\quad$ $.15 \times 2$

* We fail to reject Ho at significance level of .05.

MADDY KUBIK
3-14-11
PERIOD 3

10.5 NOTES- HYPOTHESIS TESTING WITH PROPORTIONS

ESTIMATE FOR p-sample test statistic

$\hat{p} = r/n$

$\mu = p$

$\sigma = \sqrt{\frac{pq}{n}}$

SAME AS BEFORE

LEVEL OF SIGNIFICANCE	$\alpha = .05$	$\alpha = .01$
z_0 left-tailed test	-1.645	-2.33
z_0 right-tailed test	1.645	2.33
$\pm z_0$ two-tailed tests	± 1.96	± 2.58

$np > 10$ ✓

$nq > 10$ ✓ If yes, we can use normal z-curve

$\hat{p} \to z$

$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$

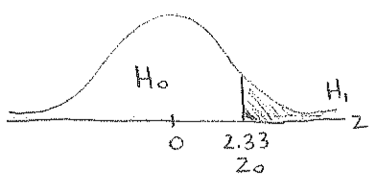
EXAMPLE 12 PAGE 437

$H_0: p = .30$

$H_1: p > .30$

$\alpha = .01$

$z_0 = 2.33$



$np = (225)(.30) = 67.5 > 10$ ✓

$nq = (225)(.70) = 157.5 > 10$ ✓

$\hat{p} = .39$

$z = \frac{.39 - .30}{\sqrt{\frac{(.3)(.7)}{225}}} = 2.95$

PAGE 441 #1

$p = .70$ - same as national average

$q = .30$ $np = (32)(.70) = 22.4 > 10$ ✓

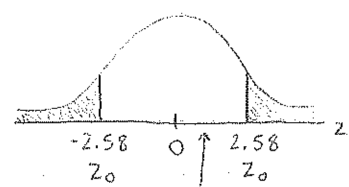
$n = 32$ $nq = (32)(.30) = 9.6 < 10$

$\hat{p} = .75$ *9.6 is almost 10, so we will procee[d]

$\alpha = .01$ with caution. Use z-curve.

$H_0: p = .70$

$H_1: p \neq .70$



$z = \frac{.75 - .70}{\sqrt{\frac{(.70)(.30)}{32}}} = \boxed{.62}$

p-value $= 1 - .7324 =$

one tail $\longrightarrow .2676\,(2) = \boxed{.5352}$

We fail to reject Ho at α of .01. There exists no sufficient statistical evidence to show that 70% of arrests in Rock Springs are young males. The p-value is .5352.

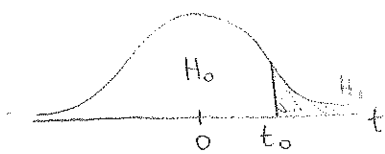
Chapter 11

MADDY KUBIK
3-23-11 (AP)
PERIOD 3

11.1 NOTES - HYPOTHESIS TESTS WITH DEPENDENT SAMPLES

- comparing before and after data
- differences between two dependent samples - "paired" or "matched"
- ADVANTAGES
  - reduces outside factors
  - lowers variance
- $\bar{d}$ = mean difference between paired data (is a sample mean = $\bar{x}$)
- $S_d = \sqrt{\dfrac{\Sigma(d-\bar{d})^2}{n-1}}$

- $H_0: \mu_d = 0$ - the population mean difference is 0.
- $H_1: \mu_d >< \neq 0$



- $\bar{d} \rightarrow$ t-score

  $\dfrac{\bar{d}-0}{s/\sqrt{n}}$

PAGE 462 #6
$\alpha = .01$

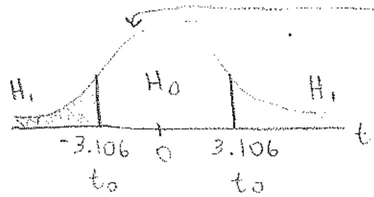
p-value = STAT, TESTS, 2-T-TEST, put in stats
p-value = .4315

d- in LIST
-6.9
-4.6
-.2
.2
1.8
2.9
3.1
2.9
2.1
-.6
-3.8
-5.2

$\bar{d} = -.8417$
$Sd = 3.57$
$H_0: \mu_d = 0$ - SENTENCE
$H_1: \mu_d \neq 0$

$\bar{d} \rightarrow t$

$t = \dfrac{-.8417-0}{3.57/\sqrt{12}} = \boxed{-.8157}$



Do not reject H_0 at α of .01. There exists sufficient statistical evidence to show that there is no difference.

MADDY KUBIK
4-11-11
PERIOD 3
AP

11.3 - INFERENCES ABOUT DIFFERENCE OF 2 MEANS

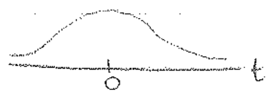
- SMALL, INDEPENDENT samples - t-distribution
- STEPS:

① $H_0: \mu_1 = \mu_2$ - No difference...

② $H_1: \mu_1 \neq \mu_2$

③ Find t_0

degrees freedom = $n_1 + n_2 - 2$



④ $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

POOLED STANDARD DEVIATION

$s = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$

EXAMPLE 7 = GOOD EXAMPLE

CONFIDENCE INTERVALS FOR $\mu_1 - \mu_2$

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

$E = t_0 s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

CONCLUDE:

BOTH + = $\mu_1 > \mu_2$

BOTH - = $\mu_1 < \mu_2$

+ AND - = CANNOT CONCLUDE

EXAMPLE 8 = CONFIDENCE INTERVAL

MADDY KUBIK
3-29-11
PERIOD 3
AP

Similar to 11.2

11.4 NOTES - HYPOTHESIS TESTING WITH PROPORTIONS

$\hat{p}$ = pooled estimate $= \frac{r_1 + r_2}{n_1 + n_2}$

$\hat{p}_1 = \frac{r_2}{n_1}$ $\quad \hat{p}_2 = \frac{r_2}{n_2}$

$\hat{p}_1 - \hat{p}_2 \rightarrow z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$ $\quad \hat{q} = 1 - \hat{p}$

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

Z_0 VALUES

	$\alpha = .05$	$\alpha = .01$
LEFT	-1.645	-2.33
RIGHT	1.645	2.33
TWO	±1.96	±2.58

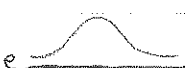
$H_0: p_1 - p_2 = 0$

CONFIDENCE INTERVALS FOR $p_1 - p_2$

$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$

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

CHECK:

① $n_1\hat{p}_1 \quad n_2\hat{p}_2 > 10$ so use  z
$n_1\hat{q}_1 \quad n_2\hat{q}_2$

② TOO BIG?
$n < \frac{1}{10}N$

③ Random Assignment

✱ CONCLUSION

c% interval for $p_1 - p_2$

- only NEGATIVE VALUES = $p_1 - p_2 < 0$ so $p_1 < p_2$
- only POSITIVE VALUES = $p_1 - p_2 > 0$ so $p_1 > p_2$
- both POSITIVE + NEGATIVE VALUES = can't conclude that either are larger. Reduce confidence interval to smaller value to conclude.

AP - HomewORK - 11.4 (1, 3-5, 8, 11)

Chapter 12

MADDY KUBIK
4-21-11
PERIOD 3
AP

12.2 NOTES - χ^2 GOODNESS OF FIT

- Ask about whether or not a population follows (fits) a given pattern

$\chi^2 = 0$ Ho: "pop fits the distribution"

$\chi^2 > 0$ H_1: "pop has a different distribution"

- $\chi^2 = \Sigma (O-E)^2/E$
- Compute E - % of n (# of items in distribution)
- d.f = E-1 (# of entries)

EXAMPLE - PINE PARK SCHOOL DISTRICT

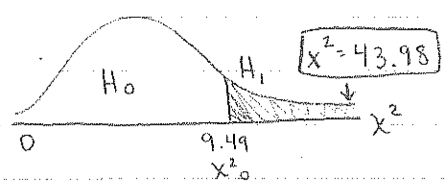
	1999	2007-O	E	$(O-E)^2$	$(O-E)^2/E$
VACATION	4%	3	(.04)(137) = 5.48	6.15	1.12
SALARY	65%	77	89.05	145.20	1.63
SAFETY	13%	9	17.81	77.62	4.36
HEALTH	12%	41	16.44	603	36.69
OVERTIME	6%	7	8.22	1.49	.1812
		137 = n			$\Sigma = 43.98 = \chi^2$

Ho: $\chi^2 = 0$ - 2007 + 1999 follow the same distribution

H_1: $\chi^2 > 0$

d.f: E-1 = 5-1 = 4

$\chi^2_o = 9.49$



Reject Ho at α of .05.

12.3 NOTES ON BACK →

12.3 NOTES - TESTING A SINGLE VARIANCE OR STANDARD DEVIATION

- Variance σ^2 of population
- n = number of measurements
- s^2 = variance of sample
- s = standard deviation of sample

$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ d.f = n-1

$H_0: \sigma^2 = k$ - No change in variance, variance remains the same

$H_1: \sigma^2 > k$ - if variance gets larger, more scattered - MOST COMMON

$\sigma^2 < k$

$\sigma^2 \neq k$

MADDY KUBIK
4-28-11
PERIOD 3
AP

12.4 NOTES - CONFIDENCE INTERVALS

LSRL = $y = a + bx$

$a = \bar{y} - b\bar{x}$

$b = \dfrac{SSxy}{SSx}$

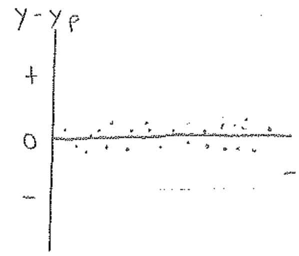
$SSxy = \Sigma xy - \dfrac{(\Sigma x)(\Sigma y)}{n}$

$SSx = \Sigma x^2 - \dfrac{(\Sigma x)^2}{n}$

y_o = observed y

y_p = predicted y

$y - y_p$ = residual



RESIDUAL PLOT

- if no pattern, LSRL good representation of data
- if pattern, use transformation

Se = standard error of estimate - measures differences between y and y_p (residuals), measures the spread

$$Se = \sqrt{\frac{SSy - bSSxy}{n-2}}$$

Se = 0 when perfect correlation (positive or negative)

Se = low when r is close to 1

Se = high when r is close to 0

Se on CALC
STAT, TEST, LINREGTTEST
LOOK FOR "s"

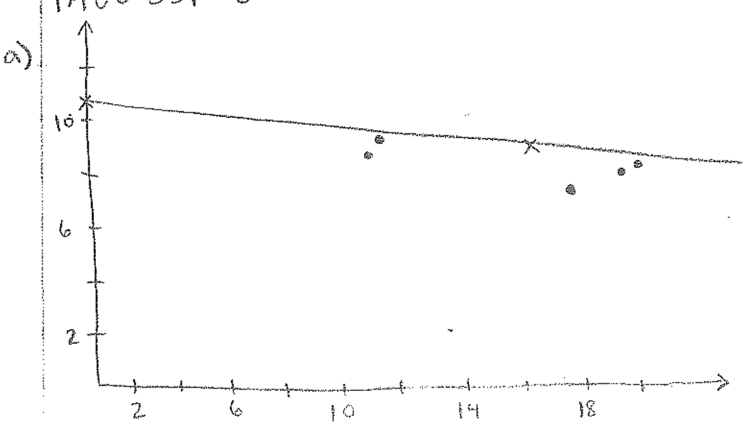
CONFIDENCE INTERVALS

*KNOW THIS → $E = t_c Se \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{SSx}}$ d.f = n - 2

- confidence interval for real y = $y_p - E < y < y_p + E$

PAGE 551 #6

a)



*neg association

*mod correlation

*none

*As x increases, y decreases

b) $\bar{x} = 15.02$

$\bar{y} = 8.46$ $n-2 = 5-2 = 3 = d.f$

$b = -.1758$

c)

0	15.02
11.1	8.46

$y = 11.1 + -.1758$

d) when $x = 17$, $y = 8.1118$

$$E = t_c S_e \sqrt{1 + \frac{1}{5} + \frac{(17-15.02)^2}{96.828}}$$ $$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 1224.83 - \frac{(75.1)^2}{5} = 96.828$$

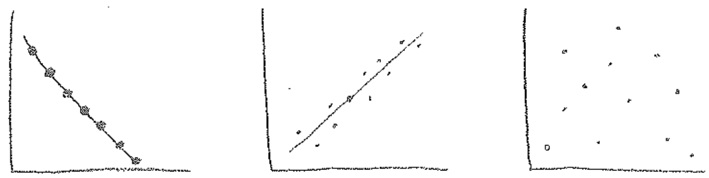
$E = 1.4578$

$y_p - E < y < y_p + E$

$6.65 < y < 9.57$ - We conclude with 75% confidence that in this city, the per capita is between 6.65 and 9.57 (thousands of dollars).

MADDY KUBIK
4-26-11
PERIOD 3
AP

12.5 NOTES - TESTING THE CORRELATION + SLOPE



PERFECT CORRELATION $r=-1$ | $r \approx .8$ | NO LINEAR CORRELATION $r=0$

LINE OF BEST FIT:

ρ = "rho" — $y=a+bx$ - sample statistic

β = "beta" — $y=\alpha+\beta x$ - population parameter

Get same t-score for testing ρ and β

ASSUMPTIONS: PAGE 554

① The set is a random sample

② The y values have a normal distribution

$r \rightarrow t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ AP TEST ONLY

*d.f = n-2

To test ρ:

$H_0: \rho = 0$ - x and y have no linear correlation.

$H_1: \rho > 0$

< 0

$\neq 0$

EXAMPLE PAGE 553

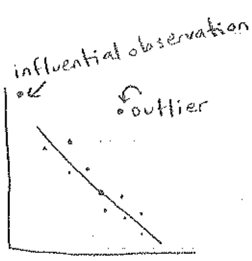
-find $y=a+bx$

STAT, CALC, 8-LINREG

$y=10.13+2.55x$

$r=.887$ - high correlation

$t=\dfrac{.887\sqrt{6-2}}{\sqrt{1-.887^2}}=3.84$



ANALYZE: +/- association

lo/mo/hi correlation

outliers

trend

To test β:

$H_0: \beta = 0$ - The population slope is 0.

$H_1: \beta > 0$

< 0

$\neq 0$

$b \rightarrow t = \dfrac{b}{\frac{S_e}{\sqrt{SS_x}}}$ $\quad S_e = \sqrt{\dfrac{SS_y - bSS_{xy}}{n-2}}$ AP TEST ONLY

*d.f = n-2

→

TI-84

CALCULATOR:

$L_1 = x$

$L_2 = y$

STAT-TESTS- F: LinRegTTest

- gives t-value
  - p-value
  - d.f
  - a = y-intercept
  - b = slope
  - s
  - r^2 = coefficient of determination
  - r = correlation coefficient

EXAMPLE PROBLEM

PAGE 561 #2

$r = .187$

$b = .172$

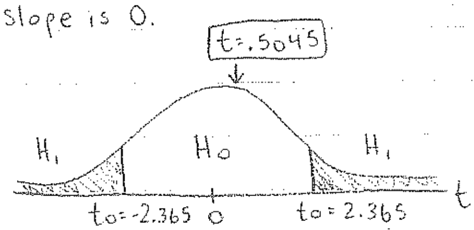
$\alpha = .05$

$H_0: \rho = 0$ - No linear correlation between student tuition and professor salaries at public universities.

$\beta = 0$ - Slope is 0.

$H_1: \rho \neq 0$

$\beta \neq 0$



d.f = 9-2 = 7

t = .5045

p-value = .6294

We fail to reject Ho at the α of .05. There exists sufficient statistical evidence to suggest that there is no linear correlation and the slope is 0. The p-value of .6294 provides weak evidence against Ho.