Lesson 1

- Descriptive Stats
- Inferential Stats

Data Collection

Terms:

- Population - collection of all things we are interested in
  " " " " we can collect data from
- Sample - subset of the population → ☆ How do we know our sample is representative of the population?
- Data - information
- Data Types - Quantitative (numeric) & Qualitative (categorical)
- Random Sample = a sample in which all members of the population have the same chance of being selected.

Lesson 2

- What type of Data?

Quantitative

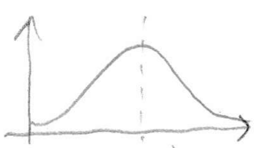
CUSS - what you should think of when you get a new dataset.

- C = center
- U = unusual points
- S = spread
- S = shape

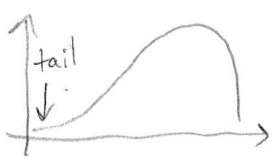
Visualizations:

- * Histogram - bars touch
- Boxplot
- Stem & Leaf Plot

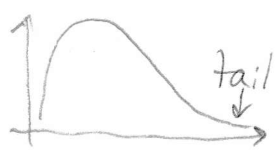
Shapes:



Symmetric (bell-shaped)



Left Skewed



Right Skewed

Skew chases the tail

Qualitative

Visualizations:

- Bar Graphs - bars do not touch
- Pie Chart

Numerics:

- Frequency (counts)
- Relative Freq. (percentages)

Lesson 2 - continued & Lesson 3

☆ Focus on Quantitative Data ☆

- Center (Location)
  - Median - Middle value; the value that has 50% of data <u>less</u> than it.
  - Mean (Arithmetic Average)

$-\cancel{1}\ \cancel{2}\ 3\ \cancel{4}\ \cancel{5000}$

median

Median doesn't look at the actual values, it just counts them.

$\overline{X} = \frac{1}{n}\sum X_i$

(mean of the X-values)

takes values into account

☆ mean is influenced by unusual points ☆

  - Centered Data: $X_i - \overline{X}$

WWTS (we want to show) $\sum (X_i - \overline{X}) = 0$; We want value c s.t. (such that) $\sum(X_i - c) = 0$

$\sum X_i - \sum \overline{X}$

$\sum X_i - n\overline{X}$

$\sum X_i - \sum X_i = 0$

$\Downarrow$

$c = \frac{1}{n}\sum X_i$

- Spread
  - $s^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2 \Rightarrow s = \sqrt{s^2}$

sample variance — sample standard deviation

- Five Number Summary

  Minimum of Data Set

  First Quartile (Q_1) = 25% mark
  (25% less, 75% greater)

  Median = 50% mark

  Third Quartile (Q_3) = 75% mark

  Maximum of Data Set

□ Range = length of data set = Max − Min

□ Inter-Quartile Range (IQR) = $Q_3 - Q_1$

☆ Our Outlier Rule: 1.5IQR Rule

↳ value that is too far away from the rest

① Calculate bounds

Lower = $Q_1 - 1.5 \cdot IQR$

Upper = $Q_3 + 1.5 \cdot IQR$

② Determine Outliers

- Data value is an outlier if $X_i <$ Lower
- Data value is an outlier if $X_i >$ upper

Lesson 4

- Sample Space - set of all possible outcomes in an experiment
- Event - some outcome of an experiment
  can be very complicated/abstract

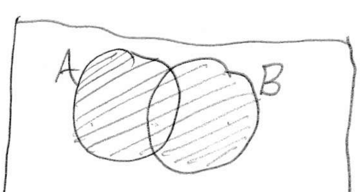
- Event Operations
  Let A, B be two events (any events)
  - A^c = "A complement" = Event that A does not occur
  - $A \cup B$ = "A union B" = At least one of A or B occurs
  - $A \cap B$ = "A intersect B" = Both A & B occur
  - $\emptyset$ = Empty Set = Event that no outcome occurs

- Probability
  - It is a function.
  - Basic Probability Computation: For an event A

$$\text{Prob}(A) = \mathbb{P}(A) = \frac{\#\text{ of ways A can occur}}{\text{Total }\#\text{ of Outcomes}}$$

Ex: $\mathbb{P}(\text{Rolling an odd }\#) = \frac{3}{6} = 1/2$

- Probability Relationships
  - $0 \leq \mathbb{P}(A) \leq 1$
  - Complement Rule: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \Leftrightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$
  - Addition Rule: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$



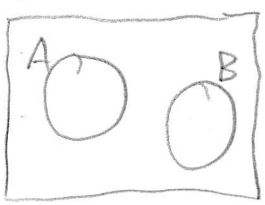
- $\mathbb{P}(\emptyset) = 0$; $\mathbb{P}(\text{Sample Space}) = 1$
- De Morgan's Laws:

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

- A & B are mutually exclusive (a.k.a. disjoint) if they cannot occur at the same time.

$\Rightarrow A \cap B = \emptyset$



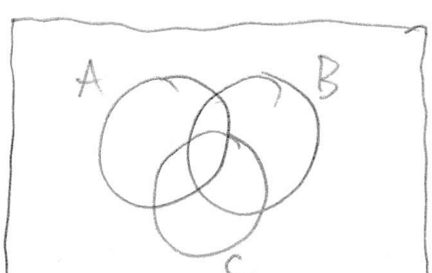
☆ If A & B are disjoint,

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(\emptyset)\\ &= \mathbb{P}(A) + \mathbb{P}(B)\end{aligned}$$

- Inclusion/Exclusion Property

We have three events; A, B, & C

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B \cap C)$$



Lesson 5

- Conditional Probability
  - we finding probabilities using some given information
  - Notation: $\mathbb{P}(A|B)$

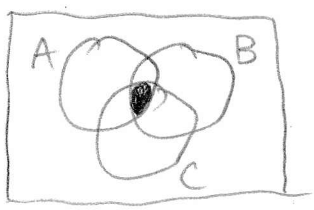
    read as "probability that A occurs given B has occurred."
  - Calculation: $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

    ☆ If A is a subset of B, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
  - Relationships: $\mathbb{P}(\cdot|B)$ satisfies all usual probability rules

    ☆ $\mathbb{P}(A|B) = 1 - \mathbb{P}(A^c|B)$

    ☆ $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(A \cap B|C)$

    $\hookrightarrow \dfrac{\mathbb{P}(A \cap C \cap B)}{\mathbb{P}(C)}$



- Multiplication Rule

  $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

  $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \cdot \mathbb{P}(B)$;

  $\Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A)$
- Law of Total Probability (Two events, A & B)

$$\mathbb{P}(A) = \mathbb{P}((A \cap B^c) \cup (A \cap B))$$
$$= \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) - \cancel{\mathbb{P}(\emptyset)}^{0}$$
$$= \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c) + \mathbb{P}(A|B)\mathbb{P}(B)$$

☆ $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)$

- Extended Law of Total Probability

Let $B_1, B_2, \ldots, B_n$ be events that partition the sample space

$$B_1 \cup B_2 \cup \cdots \cup B_n = \text{Sample Space}$$

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Lesson 6

☆ Book uses A' to denote A^c

☆ 2.17 - assume that each request only involves one type of software

$\Rightarrow A \cap B = \emptyset$

☆ Next week we start to use R.

- Independence
  - <u>Definition</u>: Two events are independent if knowing that one event occurred does not change the probability of the other occurring.
  - $\mathbb{P}(A|B) = \mathbb{P}(A)$; $\mathbb{P}(B|A) = \mathbb{P}(B)$

Suppose A & B are independent

☆ $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ ☆

- To check for independence
  - Calculate $\mathbb{P}(A \cap B)$, $\mathbb{P}(A)$, $\mathbb{P}(B)$

    IF $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, then A & B are independent

    IF $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$, then A & B are dependent

Independence

Hint: $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$

WWTS that B & A^c are independent

WWTS that $\mathbb{P}(B \cap A^c) = \mathbb{P}(B) \cdot \mathbb{P}(A^c)$

Proof: $\mathbb{P}(A^c) \cdot \mathbb{P}(B) = (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$

$= \mathbb{P}(B) - \mathbb{P}(A \cap B)$ (by independence of A&B)

$= \mathbb{P}(B \cap A^c)$ $\square$

Events A & B are events with $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$.

i) Let A & B be two independent events.

$\Rightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Since $\mathbb{P}(A), \mathbb{P}(B) \neq 0$, $\mathbb{P}(A) \cdot \mathbb{P}(B) \neq 0$

$\Rightarrow \mathbb{P}(A \cap B) \neq 0 \Rightarrow A \cap B \neq \emptyset$

$\Rightarrow$ Not disjoint

ii) Let A & B be two disjoint events

$\Rightarrow A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cap B) = 0$

Since $\mathbb{P}(A), \mathbb{P}(B) \neq 0$, $\mathbb{P}(A) \cdot \mathbb{P}(B) \neq 0$

$\Rightarrow \mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \Rightarrow$ A & B are not independent

Lesson 7 - Discrete Random Variable

- Characterisation of a discrete RV
  - $\Rightarrow$ Probability Mass Function (pmf)
    - It is always non-negative; $p(x) \geq 0$
    - Sum of its values must be 1; $\underbrace{\sum}_{\text{sum over all } x \text{ in the support}} p(x) = 1$

- Cumulative Distribution Function (CDF)

$$\underbrace{F}_{\text{CDF}}{}_{\underbrace{X}_{\text{RV}}}(\underbrace{k}_{\text{input}}) = \mathbb{P}(X \leq k)$$

- Describing RV's
  - ▫ Expected Value $= \mathbb{E}[\underbrace{X}_{\text{RV}}] = \underbrace{\sum}_{\text{sum over all values}} \overbrace{x}^{\text{values of RV}} \cdot \underbrace{p(x)}_{\text{probabilities}}$

Ex. #1 on L7: $\mathbb{E}[X] = 2(.05) + 3(.25) + 4(.6) + 5(.1) = 3.75$

  - ▫ Variance $= \text{Var}[X] = \mathbb{E}\left[(X - \underbrace{\mathbb{E}[X]}_{\text{just a number}})^2\right]$

☆ Computing expected values

$$\mathbb{E}[\underbrace{g}_{\text{some function}}(X)] = \sum g(x) \cdot p(x)$$

Ex. $\mathbb{E}[X^2] = 2^2 \cdot (.05) + 3^2(.25) + 4^2(.6) + 5^2(.1) = 14.55$

- Let $Y = a + bX$; a,b are constants

  - $\mathbb{E}[Y] = \mathbb{E}[a+bX] = \sum (a+bx)p(x)$

    $= \sum (ap(x) + bxp(x))$

    $= \sum ap(x) + \sum bxp(x)$

    $= a \cdot \sum p(x) + b \cdot \sum xp(x)$

    $= a \cdot 1 + b \cdot \mathbb{E}[X]$

    $= a + b\mathbb{E}[X]$

  ☆ Expectation is a linear operator ☆

  - $Var[Y] = Var[a+bX] = b^2 \cdot Var[X]$

  - $Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \geq 0$

☆Shortcut Formula

- Standard Deviation $= SD[X] = \sqrt{Var[X]}$

Lesson 7 Recap

- PMF : $p(x) = \mathbb{P}(\underset{\text{RV}}{X} = x)$ ← More important for discrete RV's
- CDF : $F_X(k) = \mathbb{P}(X \leq k)$
- $\mathbb{E}[X] = \sum x \cdot p(x)$; $\mathbb{E}[g(X)] = \sum g(x) \cdot p(x)$
  ↳ Long-run expectation of the sample average
- $Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- Properties :
  - $\mathbb{E}[a + bX] = a + b \cdot \mathbb{E}[X]$
  - $Var[a+bX] = b^2 \cdot Var[X] \Rightarrow SD[a+bX] = b \cdot SD[X]$

Paperwork

X = # of signatures required

Support = $\{2, 3, 4, 5\}$

$\mathbb{P}(X = k) = c \cdot k$ for some constant c

$\mathbb{P}(X=2) = 2c$

☆ Inversely proportional:
If a is inversely prop. to b
$\Rightarrow a = \frac{c}{b}$ for some constant c

To have a valid pmf, its values have to sum to 1

$\Rightarrow 2c + 3c + 4c + 5c = 1 \Rightarrow c = 1/14$

X	2	3	4	5
p(x)	2/14	3/14	4/14	5/14

; $\mathbb{P}(X = k) = \frac{k}{14}$

Lesson 8 - Bernoulli RV's, & Binomial RV's

- Bernoulli RV
  - Any success or failure experiment
    - Any experiment that has only two possible outcomes, "success" and "failure."
  - PMF

    Support: $\{0,1\}$

    $\mathbb{P}(X=1) = p$ ← different in every context (what we call a parameter)

    $\mathbb{P}(X=0) = 1-p$
  - $\mathbb{E}[X] = 0(1-p) + 1(p) = p$
  - $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p)$

    $\mathbb{E}[X^2] = 0^2(1-p) + 1^2(p) = p$

- Binomial RV ; $X \sim Binom(n,p)$
  - X is binomial if it counts the # of successes in n independent Bernoulli experiments.
  - PMF: Support $= \{0, 1, 2, \dots, n\}$

    $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$; $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ ← Counts # of ways to get k things out of n (order doesn't matter)
  - $\mathbb{E}[X] = np$
  - $\mathrm{Var}[X] = np(1-p)$

Lesson 8 - Continued

Binomial Distribution

- Parameters are n & p

  n ↑ size of group

  p ↑ probability of "success"

- When can we use it?
  - Need n <u>independent</u> <u>Bernoulli</u> Trials
  - Each <u>Bernoulli</u> trial must have the same prob. of success.
  - Each trial has the same two outcomes

- PMF: $\mathbb{P}(X=k) = \underbrace{\binom{n}{k}}_{\substack{\text{binomial}\\\text{coefficient}}} p^k(1-p)^{n-k}$; $\overbrace{k \in \{0,1,2,\dots,n\}}^{\text{Support}}$

$$\binom{n}{k} = \text{"n choose k"} = \frac{n!}{k!\,(n-k)!} \; ; \; n! = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(2)(1)$$

☆ $X \sim \text{Binom}(n,p)$, X = Sum of n independent Bernoulli(p) ☆

- $\mathbb{E}[X] = \mathbb{E}[\sum Y_i] = \sum \mathbb{E}[Y_i] = \sum p = np$

  $Y_i \sim \text{Bernoulli}(p)$

$$Var[X] = Var[\sum Y_i] \overset{\text{B/c the } Y_i\text{'s are independent}}{=} \sum Var[Y_i] = \sum p(1-p) = np(1-p)$$

Casella & Berger 3.5

Givens:

- Standard drug is effective 80% of the time
- Sample 100 patients w/ condition & give them the new drug
- 85 cases of the 100 result in a success of the new drug

Question:

- Is the new drug more effective than the standard?

Solution:

First, assume that standard treatment & new are equally effective. That is, the success rate for the new drug is 80%.

$$\underbrace{P(X \geq 85)}_{\text{probability of an event as or more extreme than we witnessed}} = \sum_{k=85}^{100} \binom{100}{k} (.8)^k (.2)^{100-k} \approx \underline{.1285}$$

$$\underbrace{n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n}_{\text{Stirling's Formula}} \; ; \qquad \lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1$$

Apply this to the Binomial pmf

$$\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \approx \left(\frac{n}{2\pi k(n-k)}\right)^{1/2} \left(\frac{n}{k} p\right)^k \left(\frac{n}{n-k}(1-p)\right)^{n-k}$$

Lesson 9 - Discrete RV's continued

• Geometric Random Variable

- Counts the number independent Bernoulli trials until the first success.

  Have the same outcomes with the same prob. of success

- Notation: $X \sim \text{Geom}(p)$

  prob. of success

- PMF: $\mathbb{P}(X=k) = (1-p)^{k-1} \cdot p = \underline{p(1-p)^{k-1}}$

  $k \in \{1, 2, 3, \dots\} = \mathbb{N}$

  support

- $\mathbb{E}[X] = \frac{1}{p}$; $\text{Var}[X] = \frac{1-p}{p^2}$

• Poisson Random Variable

- Counts the number of "arrivals" in a fixed window of time

  a.k.a. "events"

  "occurrences"

- Notation: $X \sim \text{Poisson}(\lambda)$

  "rate of arrivals"

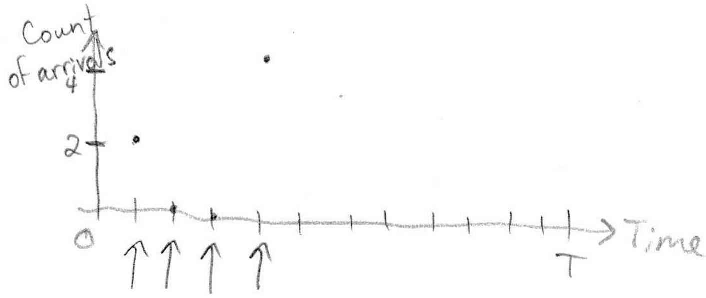
  or "average arrivals in that window"

- $\mathbb{E}[X] = \lambda = \text{Var}[X]$

- PMF: $\mathbb{P}(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$; $0! = 1$

  $k \in \{0, 1, 2, 3, \dots\}$

□ Poisson Process



Each time point is thought of as a Poisson RV

☆ Poisson processes scale with time! ☆

See last question on lesson 9

Lesson 10 - Exam 1 Review

- Exam Details
  - In the usual classroom, usual time, on Friday 9/15
  - Full 50 minutes for the exam
    I'll aim to show up 10-15 minutes early
  - Calculator allowed
  - Formula Sheet provided

- Topics
  - Week 1 - Lessons 1, 2, & 3
    - Basics are good to keep in mind, but no explicit questions on them.
    - ☆ - Shapes of Distributions
  - Week 2 - L 4, 5, & 6 - Probability
    - Relationships
      - ★ Event relationships - unions, intersections, & complements
      - ☆ Probability relationships - addition rule, complement rule, inclusion-exclusion principle, multiplication rule, law of total probability, independence
    - Conditional Probability
      - ☆ $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$
      - ☆ $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C) - \mathbb{P}(A \cap B|C)$
    - Independence

      How to check if two events are independent:

$$\mathbb{P}(A \cap B) \stackrel{?}{=} \mathbb{P}(A) \cdot \mathbb{P}(B)$$

- Week 3 - L7, L8 - Discrete RV's, Binomial
  - □ PMF, $\mathbb{E}[X]$, $Var[X]$, $\mathbb{E}[g(X)]$
    - all values ($p(x)$) ≥ 0
    - $\sum p(x) = 1$
  - □ Bernoulli & Binomial RV's

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B \mid C)\,\mathbb{P}(C)$$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B \cap C)\,\mathbb{P}(B \cap C)$$

Lesson 11 - Continuous RV's

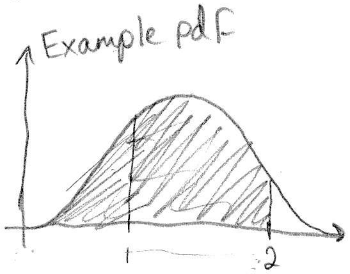
- Probability Density Function (pdf)

  $F(x)$ is a pdf if:

  - $F(X) \geq 0$ for all x
  - $\int_{\text{support}} F(x)\,dx = 1$

- Finding probabilities:

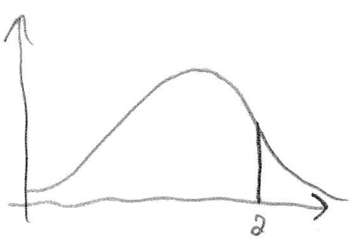
  *Probabilities as being areas under the pdf



$\mathbb{P}(1 \leq X \leq 2) = \int_1^2 F(x)\,dx$ ← pdf

$\mathbb{P}(X \leq 2) = \int_{\text{start of the support}}^{2} F(x)\,dx \; ; \; \mathbb{P}(X \geq 2) = \int_2^{\text{end of support}} F(x)\,dx$

CDF (for $\mathbb{P}(X \leq 2)$)



$\mathbb{P}(X = 2) = 0 \; ; \; \mathbb{P}(X = k) = 0$ ← For any continuous RV

2. $\underset{\text{pdf}}{F(x) = k(9 - x^2)}; \quad \underset{\text{support}}{0 < x < 3}$

$$\int_0^3 k(9 - x^2)\,dx = 1 \Rightarrow k \cdot \underbrace{\left(9x - \tfrac{1}{3}x^3\right)\Big|_0^3} = 1$$

$$\rightarrow k(27 - 9) = 1 \Rightarrow k \cdot 18 = 1 \Rightarrow k = \frac{1}{18}$$

Lessons 11 & 12 - Continuous RV's continued

- Characterization: pdf, $f(x)$

  Use pdf to find probabilities

- CDF (cumulative dist. function)

$$F(k) = \mathbb{P}(X \leq k) = \int_{\text{start of support}}^{k} f(x)\,dx$$; can ~~always~~ often compute a formula for CDF

- Expected Value

$$\mathbb{E}[g(X)] = \int_{\text{support}} g(x) \cdot f(x)\,dx$$

Particularly:

- $\mathbb{E}[X] = \int_{\text{support}} x \cdot f(x)\,dx$
- $\mathbb{E}[X^2] = \int_{\text{support}} x^2 \cdot f(x)\,dx$

- Variance

$$Var[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- Percentiles

Let k_p be the $(100 \cdot p)^{th}$-percentile

k_p is the value that has $(100 \cdot p)\%$ of the distribution less than it. (or equal to)

$k_{.5} = 50^{th}$ percentile = median

$$\int_{\text{start of support}}^{k_p} f(x)\,dx = \mathbb{P}(X \leq k_p) = p$$

$$\mathbb{P}(X \leq k_p) = \mathbb{P}(X < k_p)$$

Example:

$$\int_{\text{start of support}}^{k_{.5}} f(x)\,dx = .5$$

L11 #3 $f(x) = \frac{1}{18}(9-x^2)\ ;\ 0<x<3$

c)

$$F(c) = P(X \le c) = \int_0^c \frac{1}{18}(9-x^2)\,dx = \frac{1}{18}\left(9x - \frac{1}{3}x^3\right)\Big|_0^c$$

is in the support of X

$$= \frac{1}{18}\left(9c - \frac{1}{3}c^3\right)$$

$$F(c) = \begin{cases} 0, & c<0 \\ \frac{1}{18}\left(9c - \frac{1}{3}c^3\right), & c \in (0,3) \\ 1, & c>3 \end{cases}$$

R code:

- sapply(range of values, function to apply)

in our example: grid   integrate(f, 0, grid element)

Lesson 13 - Continuous RVs

- Genes on L9 asked for an "exact distribution."

  To identify a distribution, $\underbrace{\text{always give a pmf or pdf}}$

  In the case of common dist.'s, we can simply state the name & parameters.

- Devore, 4.19

Given CDF of X: $F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{4}\left[1+\ln(4/x)\right], & x \in (0,4] \leftarrow \text{support} \\ 1 & , x > 4 \end{cases}$

c) Find the pdf of X

$$\underset{\text{CDF} \uparrow}{F(x)} = \int_{\text{start of support}}^{x} f(z)\,dz$$

☆ PDF is the derivative of CDF ☆

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(\frac{x}{4}\left[1+\ln(4/x)\right]\right)$$

$$= \left(\tfrac{1}{4}\right)\left(1+\ln(4/x)\right) + \left(\tfrac{x}{4}\right)\left(\frac{1}{4/x}\cdot\left(\frac{-4}{x^2}\right)\right)$$

$$= \frac{1}{4} + \frac{1}{4}\ln(4/x) - \frac{\cancel{4}}{\cancel{x^2}}\cdot\frac{x}{\cancel{4}}\cdot\frac{\cancel{x}}{4}$$

$$= \frac{1}{4}\ln(4/x) = \frac{1}{4}\left(\ln(4)-\ln(x)\right)$$

$x \in (0,4]$

- PDF, Lesson 11: $F(3) = \mathbb{P}(X \le 3) = 1$

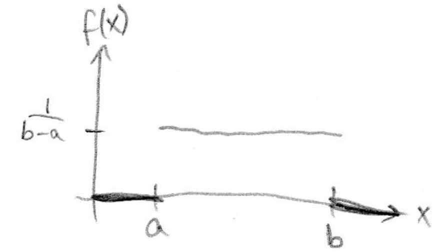
  Support $0 < x < 2$

• Uniform Distribution

- Support is an interval ($\to$ Finite), and the endpoints of the interval are the parameters of this distribution.

$X \sim \text{Uniform}(a, b)$; Support is $\underset{\text{interval}}{(a, b)}$

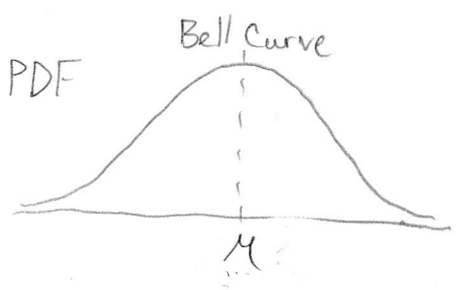
- PDF: just a flat line, $f(x) = \frac{1}{b-a}$; $x \in (a,b)$



- $\mathbb{E}[X] = \frac{a+b}{2}$

- $Var[X] = \frac{(b-a)^2}{12}$

• Normal Distribution
(Gaussian)



$X \sim \text{Normal}(\mu, \sigma)$

PDF: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

$x \in \mathbb{R} = (-\infty, \infty)$

μ = center (location parameter)
$= \mathbb{E}[X]$

σ = spread (shape parameter)
$= SD[X] \Rightarrow Var[X] = \sigma^2$

CDF: $\boxed{\Phi(z)} = F(z) = \mathbb{P}(X \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \leftarrow$ No closed form

☆ When $\mu = 0, \sigma = 1$, we have the standard normal dist.

$Z \sim \text{Normal}(0,1)$; $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$; $z \in (-\infty, \infty)$

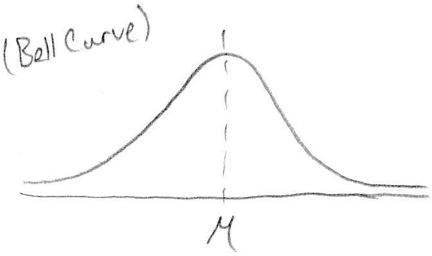
If $X \sim \text{Normal}(\mu, \sigma)$, then $\frac{X - \mu}{\sigma} \sim \text{Normal}(0,1)$

Lesson 13 - Continued

• Normal Random Variable

$X \sim \text{Normal}(\underset{\mathbb{E}[X]}{\mu}, \underset{SD[X]}{\sigma})$

PDF: (Bell Curve)



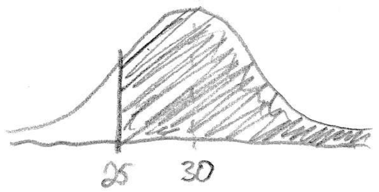
Reactor $X \sim \text{Normal}(\overset{\mu}{30}, \overset{\sigma}{6})$

a) $\mathbb{P}(X > 25) = 1 - \mathbb{P}(X \le 25)$

parameters

$= 1 - \underline{\text{pnorm}}(25, 30, 6) \approx .7977$

standard dev.

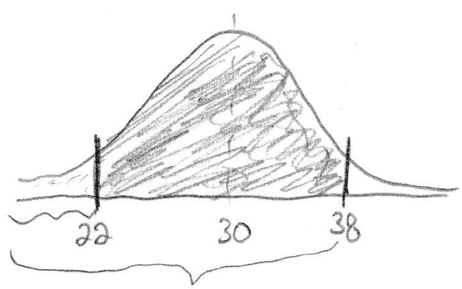


b) $\mathbb{P}(|X - \underset{\mu}{30}| < 8) = \mathbb{P}(-8 < X - 30 < 8)$

☆ For Continuous RV's $\mathbb{P}(X<k) = \mathbb{P}(X \le k)$ ☆

$= \mathbb{P}(22 < X < 38)$

$= \mathbb{P}(X < 38) - \mathbb{P}(X < 22)$



$= \text{pnorm}(38, 30, 6) - \text{pnorm}(22, 30, 6)$

$\approx .82$

c) For percentiles:

☆ p^{th} percentile is the support value that has $p\%$ of the dist. less than or equal to it ☆

$\text{qnorm}(.85, 30, 6) \approx 36.2186$

percentile we are looking for

☆ $\text{qnorm}(.5, \mu, \sigma) = \mu$ (median)

- Exponential Random Variable
  - Exponential RV's measure "waiting times"
    ↳ how long until the next event/arrival
  - $X \sim Exp(\lambda)$
    ↑ "rate of occurrence over time"
  - PDF: $F(x) = \lambda \cdot e^{-\lambda x}$; $x > 0$
  - CDF: $F(x) = \mathbb{P}(X \le x) = \int_0^x \lambda e^{-\lambda z} dz = \left(-e^{-\lambda z}\right)\Big|_0^x = 1 - e^{-\lambda x}$; $x > 0$
  - $\mathbb{E}[X] = \int_0^\infty \underbrace{x}_{u} \cdot \underbrace{\lambda e^{-\lambda x} dx}_{dv}$ $\begin{pmatrix} u = x & dv = \lambda e^{-\lambda x} dx \\ du = dx & v = -e^{-\lambda x} \end{pmatrix}$

$$= \left(\cancel{-xe^{-\lambda x}}^{\,0}\right)\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx$$

$$= (0 + 0) + \left(-\frac{1}{\lambda} e^{-\lambda x}\right)\Big|_0^\infty = 0 - \left(-\frac{1}{\lambda}\right) = 1/\lambda$$

  - $Var[X] = 1/\lambda^2$

Lesson 14 - Joint Distributions

Y $\qquad$ y

- Definition: Joint Distribution
  - Discrete: We consider two RV's, X & Y, but now we use a joint pmf to describe their probabilities

$$p(x,y) = \mathbb{P}(X=x \cap Y=y)$$

  - Continuous: Consider two RV's, X & Y; here we use a joint pdf

$$f(x,y)$$

what we integrate to get probabilities.

- Independence between two RV's.
  - In the discrete case, X & Y are independent if

$$\underset{\text{joint pmf}}{p(x,y)} = \underset{\text{pmf of X}}{p_X(x)} \cdot \underset{\text{pmf of Y}}{p_Y(y)} \quad \text{For all } x,y \text{ in the support}$$

  - In the continuous case, X & Y are independent if

$$\underset{\text{joint pdf}}{f(x,y)} = \underset{\text{pdf of X}}{f_X(x)} \cdot \underset{\text{pdf of Y}}{f_Y(y)} \quad \text{For all } x,y \text{ in the support}$$

- Marginal Distributions (individual dist.'s)

  - Discrete case: we know $p(x,y)$.

$$p_X(x) = \sum_{\substack{y \text{ support} \\ \text{values}}} p(x,y) \quad ; \quad p_Y(y) = \sum_{\substack{x \text{ support} \\ \text{values}}} p(x,y)$$

  - Continuous case; we know $f(x,y)$

$$f_X(x) = \int_{\substack{\text{support} \\ \text{of } Y}} f(x,y)\,dy \quad ; \quad f_Y(y) = \int_{\substack{\text{support} \\ \text{of } X}} f(x,y)\,dx$$

Lesson 14, #3

X	$\mathbb{P}(X=x)$
0	.7
2	.3

Y	$\mathbb{P}(Y=y)$
-1	.4
1	.6

$p_X(0)\cdot p_Y(-1) = (.7)(.4) = .28$

$\neq .3 = p(0,-1)$

$\Rightarrow$ X & Y are not independent.

#4

a)

X \ Y	0	1	2
0	.16	.24	.4
1	.04	.06	.1

☆ By independence, we know

$p(x,y) = p_X(x)\cdot p_Y(y)$

$p(0,0) = p_X(0)\cdot p_Y(0) = (.8)(.2) = .16$

b) $\mathbb{P}(X+Y \leq 1) = p(0,0) + p(0,1) + p(1,0) = .44$

#5 $f(x,y) = k(x+y)\,;\quad \begin{matrix} 0<x<2 \\ 0<y<2 \end{matrix}$

a) $\int_0^2\int_0^2 f(x,y)\,dx\,dy = 1 \Rightarrow \int_0^2\int_0^2 k(x+y)\,dx\,dy = 1$

$\Rightarrow k\int_0^2\left(\int_0^2 x+y\,dx\right)dy = 1$

$\rightarrow k\int_0^2 \left(\frac{1}{2}x^2 + xy\right)\Big|_0^2\,dy = 1$

$\Rightarrow k\int_0^2 2+2y\,dy = 1$

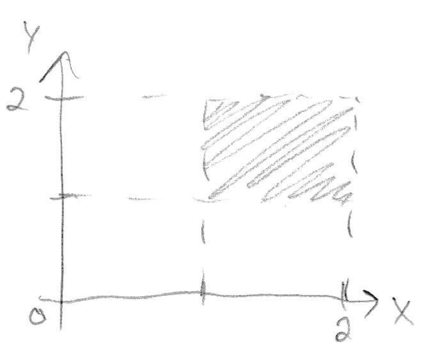
$\Rightarrow k\left(2y+y^2\right)\Big|_0^2 = 1 \Rightarrow k(4+4) = 1 \Rightarrow k = 1/8$

b) $f_X(x) = \int_0^2 \frac{1}{8}(x+y)\,dy = \frac{1}{8}\left(xy + \frac{1}{2}y^2\right)\Big|_0^2 = \frac{1}{8}(2x+2)\,;\quad 0<x<2$

$f_Y(y) = \frac{1}{8}(2y+2)\,;\quad 0<y<2$

d) $\mathbb{P}(X>1 \cap Y>1) = \int_1^2\int_1^2 \frac{1}{8}(x+y)\,dx\,dy = 3/8$

$\mathbb{P}(X\&Y \text{ in some region}) = \iint_{\text{region}} f(x,y)\,dx\,dy$



Lesson 15 - Joint Dist.'s Continued

- $\mathbb{E}[g(X,Y)] \longrightarrow$ ⭑Discrete case $= \sum_{x\text{ support}} \sum_{y\text{ support}} g(x,y)\cdot p(x,y)$

  ⭑Continuous Case $= \int_{x\text{ support}} \int_{y\text{ support}} g(x,y)\cdot f(x,y)\,dy\,dx$

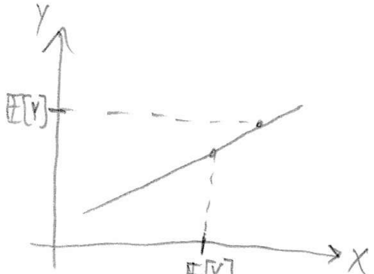
⭑ Commonly used formula:

$\mathbb{E}[XY] \xrightarrow{\text{discrete}} \sum_{x\text{ supp.}} \sum_{y\text{ supp}} xy\cdot p(x,y)$

$\mathbb{E}[XY] \xrightarrow{\text{continuous}} \int_{x\text{ supp.}} \int_{y\text{ supp.}} xy\cdot f(x,y)\,dy\,dx$

- Covariance - measures the <u>linear</u> relationship between two RV's.

$$\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \underbrace{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}_{\text{computation Formula}}$$



⭑ Can be positive or negative

$$\mathrm{Cov}[X,X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathrm{Var}[X]$$

- Correlation - true, unitless measure of two RVs <u>linear</u> relationship
  - Correlation is always between -1 and 1
  - Sign of the correlation tells us the slope of the linear relationship

  Magnitude tells us how close to a line our relationship is.

  Example: • Correlation $= 0 \Rightarrow$ No <u>linear</u> relation
  • Correlation $= 1 \Rightarrow$ perfect positive linear relation

$$\text{Corr}[X,Y] = \frac{\text{Cov}[X,Y]}{SD[X] \cdot SD[Y]} = \frac{\text{Cov}[X,Y]}{\sqrt{\text{Var}[X] \cdot \text{Var}[Y]}}$$

- Independence

☆ If X & Y are independent, then $\text{Cov}[X,Y] = 0$ ☆

$\text{Corr}[X,Y] = 0$

Beware: $\text{Cov}[X,Y] = 0 \not\Rightarrow X$ & Y are independent

- Random Sample

  Let $\{X_i\}_{i=1}^{n} = \{X_1, X_2, X_3, \ldots, X_n\}$

  each X_i is a random variable

  each X_i is a data value/observation

  $\{X_i\}_{i=1}^{n}$ is a random sample if:
  - each X_i comes from the same dist. (population)
  - all of the X_i's are independent

A random sample is "independent and identically distributed."

Lesson 16 – Linear Combinations of RVs

- Let $X_1, \dots, X_n$ be RVs Ex: $W = X_1 + 2X_2$

  Let $a_1, \dots, a_n$ be constants

  Linear combination of the X_i's is $\sum_{i=1}^{n} a_i X_i$

$$= a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

$\mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i]$; Ex. $\mathbb{E}[W] = \mathbb{E}[X_1 + 2X_2] = \mathbb{E}[X_1] + 2\mathbb{E}[X_2]$

$\mathrm{Var}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 \mathrm{Var}[X_i] + 2a_i a_j \underbrace{\sum_{i<j} \mathrm{Cov}[X_i, X_j]}$ → Covariances of all pairs of RVs

Ex. $\mathrm{Var}[X+Y+Z] = \mathrm{Var}[X] + \mathrm{Var}[Y] + \mathrm{Var}[Z]$
$$+ 2\mathrm{Cov}[X,Y] + 2\mathrm{Cov}[X,Z] + 2\mathrm{Cov}[Y,Z]$$

$\mathrm{Cov}[X,Y] + \mathrm{Cov}[Y,X]$

☆ When the X_i's are independent

$$\mathrm{Var}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 \mathrm{Var}[X_i]$$

- Linear Combinations of Normal RVs

  Suppose $\{X_i\}_{i=1}^{n} \overset{\perp\!\!\!\perp}{\sim} \mathrm{Normal}(\mu_i, \sigma_i)$ ← independent ⟶ $X_1 \sim \mathrm{Normal}(\mu_1, \sigma_1)$

  Then $\sum_{i=1}^{n} a_i X_i \sim \mathrm{Normal}\left(\underbrace{\sum_{i=1}^{n} a_i \mu_i}_{\text{mean}}, \underbrace{\sqrt{\sum_{i=1}^{n} a_i^2 \sigma_i^2}}_{\text{standard dev.}}\right)$

Shipping

Let T = total weight shipped in a week

$$= 5 \cdot \underbrace{S}_{\text{\# of small items shipped in total}} + 7M + 10L$$

$$\mathbb{E}[T] = 5\mathbb{E}[S] + 7\mathbb{E}[M] + 10\mathbb{E}[L] = 3400$$

$$\mathrm{Var}[T] = (25)(36) + 49(64) + 100(16) = 5636$$

$$SD[T] = \sqrt{5636}$$

Ethanol

$X_1, X_2, X_3 \overset{\text{II}}{\sim} N(24,4)$; $X_4 \sim N(22,3)$

b) $\mathbb{P}(X_1 > X_4) = \mathbb{P}(\underbrace{X_1 - X_4}_{\text{what is this distribution?}} > 0)$

$$X_1 - X_4 \sim \mathrm{Normal}(24-22, \sqrt{16+9})$$

$$\sim \mathrm{Normal}(2,5)$$

$$\Rightarrow \mathbb{P}(X_1 - X_4 > 0) = 1 - \mathbb{P}(X_1 - X_4 \le 0)$$

$$= 1 - \mathrm{pnorm}(0, \underset{\text{mean}}{2}, \underset{\text{sd}}{5}) \approx .6554$$

Lesson 17 - Sampling Distributions

$\underbrace{\text{Sampling}}$ → probability distribution of a statistic

- Two statistics of interest
  - Sum $= \sum_{i=1}^{n} X_i$ (adding data values)
  - Mean $= \frac{1}{n}\sum_{i=1}^{n} X_i$

☆ Suppose that the $\{X_i\}_{i=1}^{n} \overset{iid}{\sim} \text{Normal}(\mu, \sigma)$ ☆

Then: $\sum_{i=1}^{n} X_i \sim \text{Normal}(n\mu, \sqrt{n\sigma^2})$

$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \text{Normal}(\mu, \frac{\sigma}{\sqrt{n}})$

exact

- Central Limit Theorem (CLT)

If $\{X_i\}_{i=1}^{n}$ are iid (random sample) with mean μ and standard deviation σ, then:

$$\underbrace{\overline{X}_n}_{\text{Sample avg of } n \text{ data values}} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx \text{Normal}(\mu, \frac{\sigma}{\sqrt{n}})$$

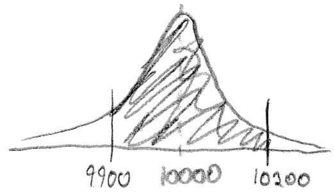
approximately

☆ great approximation when $n > 30$

Rivets $\{X_i\}_{i=1}^{40} \overset{iid}{\sim} (\overset{\mu}{10000}, \overset{\sigma}{500})$

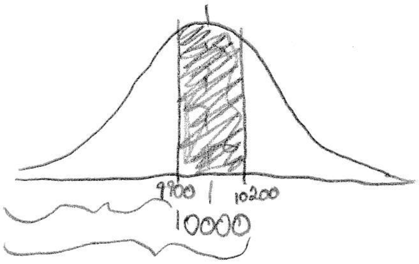
st.dev of $\overline{X}$ is less than the st.dev of X

a) $\overline{X}_{40} \approx \text{Normal}\left(10000, \frac{500}{\sqrt{40}}\right)$



b) $SE[\overline{X}_{40}] = \frac{500}{\sqrt{40}}$

c) $\mathbb{P}(9900 < \overline{X}_{40} < 10200) = \mathbb{P}(\overline{X}_{40} < 10200) - \mathbb{P}(\overline{X}_{40} < 9900)$



$= \text{pnorm}\left(10200, \overset{\mu}{10000}, \overset{\sigma}{\frac{500}{\sqrt{40}}}\right)$

$- \text{pnorm}\left(9900, 10000, \frac{500}{\sqrt{40}}\right)$

$\approx .8913$

$\mu = 0$
$sd = 1$

pnorm(1)

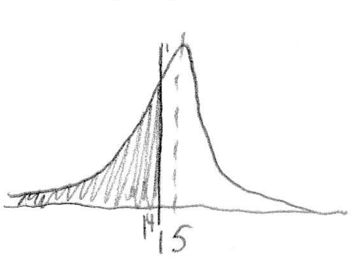
Uniform $\{X_i\}_{i=1}^{50} \overset{iid}{\sim} \underline{\text{Unif}}(10,20)$

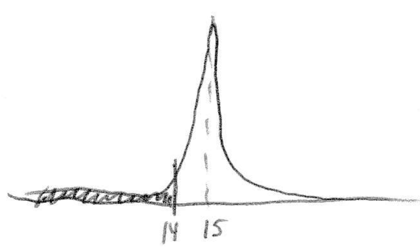
a) $\mathbb{E}[X] = \frac{10+20}{2} = 15$

$\text{Var}[X] = \frac{(20-10)^2}{12} = 25/3 \Rightarrow SD[X] = 5/\sqrt{3}$

b) $\overline{X}_{50} \approx \text{Normal}\left(15, \frac{5/\sqrt{3}}{\sqrt{50}}\right) \rightarrow \overline{X}_{50} \approx \text{Normal}\left(15, \frac{1}{\sqrt{6}}\right)$

c) $\mathbb{P}(\overline{X}_{50} < 14) = \text{pnorm}(14, \overset{\mu}{15}, \overset{\sigma}{1/\sqrt{6}}) \approx .0072$





Lesson 18 - Point Estimators & Interval Estimators

• Point Estimators

Definition: Suppose we want to estimate a parameter Θ.

Feature of the population.

sample statistic

Then a point estimator of Θ, denoted by $\hat{\Theta}$, is just a value that we think of as the "best guess" for Θ.

Example: L18 #1 $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$; p unknown

Goal: Estimate p

☆ $\mathbb{E}[\bar{X}] = \mathbb{E}[X]$

$\mathbb{E}[X_i] = p$, $Var[X_i] = p(1-p)$

Use $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ to estimate p.

↳ sample proportion of success.

☆ $\mathbb{E}[\hat{p}] = p$ Expected value of the estimator is the quantity we are trying to estimate.

$\Rightarrow$ we say $\hat{p}$ is unbiased for p.

☆ CLT: If $X_i \overset{iid}{\sim} (\mu, \sigma)$, then $\bar{X} \sim \text{Normal}(\mu, \frac{\sigma}{\sqrt{n}})$ for $n > 30$

$\hat{p} \sim \text{Normal}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ when $\left(\begin{matrix} np \geq 10 \\ n(1-p) \geq 10 \end{matrix} \text{ or } n > 30\right)$

how variable is our estimator

As $n \to \infty$, $\sqrt{\frac{p(1-p)}{n}} \to 0 \Rightarrow$ As the sample size gets larger, the spread of our estimator gets smaller, & the estimator gets more accurate.

- Interval Estimators

Definition: We want to estimate a parameter Θ. An interval estimator of Θ is a range of numbers that we are <u>confident</u> contains the true parameter value.

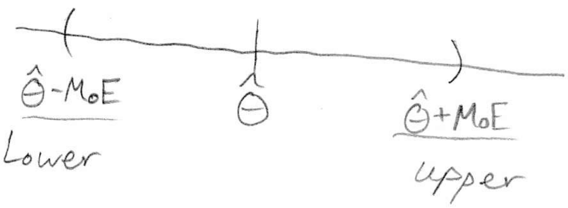
General formula: A $\underline{cl}$ Confidence Interval for Θ is given by:

conf. level

standard error of $\hat{\Theta}$

$$\hat{\Theta} \pm \underbrace{(\text{critical value})(SE[\hat{\Theta}])}_{MoE\ =\ \text{Margin of Error}}$$

pt. estimate



Interval Notation: (Lower, upper)

Interpreting a Confidence Interval (CI):

After a CI is constructed we may want to ask ourselves "what's the probability that Θ is in the CI?"

Example: We get a 95% confidence interval of <u>(.02, .16)</u>

☆ Nothing is random after the interval is constructed ☆

How to correctly interpret:

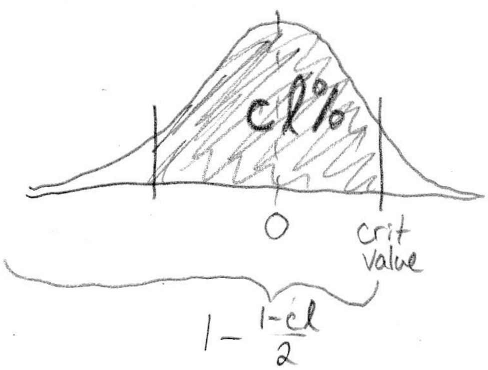
In the long run, if we repeatedly collect data and keep constructing 95% CI's, we expect 95% of those intervals to capture the parameter.

Say we have a confidence interval result to interpret:

"We are $\underset{\text{confidence level}}{\underline{\qquad\qquad}}$ % <u>confident</u> that the true $\underset{\text{parameter}}{\underline{\qquad\qquad}}$ is between $\underset{\text{lower}}{\underline{\qquad}}$ and $\underset{\text{upper}}{\underline{\qquad}}$."

• Critical Values for a Normal population mean ($\theta = \mu$)

For a given conf. level (cl), the critical value is the $\left(1-\frac{1-cl}{2}\right)$-percentile of the <u>standard normal dist.</u>



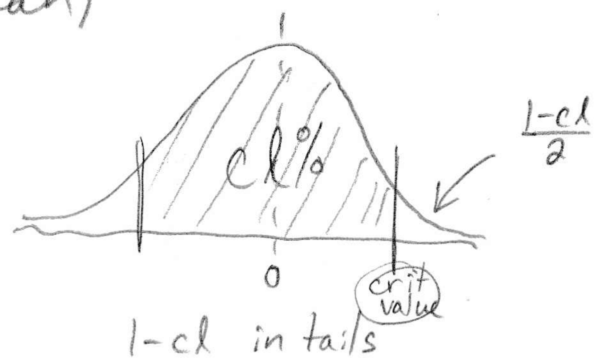
$$\text{Critical Value} = \text{qnorm}\left(1-\frac{1-cl}{2}\right)$$

Lesson 18 - Continued

In every confidence interval (CI), we have a confidence level (cl) which directly relates to critical value.

- CI's for μ (population mean)

$$Z_{1-\frac{1-cl}{2}} = \text{qnorm}\left(1-\frac{1-cl}{2}\right)$$



CI for μ: $\bar{X} \pm Z_{1-\frac{1-cl}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

Given cl

☆ Notice this involves σ

Example: Blades

a) $\sigma = 20$
$cl = .95$ → want to keep this as a decimal!
$n = 25$
$\bar{X} = 402$

$$Z_{.975} = \text{qnorm}(.975) = 1.96$$

$$402 \pm (1.96)\left(\frac{20}{\sqrt{25}}\right)$$

$$= (394.16, 409.84) = \text{all \#'s between } 394.16 \text{ \& } 409.84$$

We are 95% confident that the true mean power is between 394.16 kW and 409.84 kW.

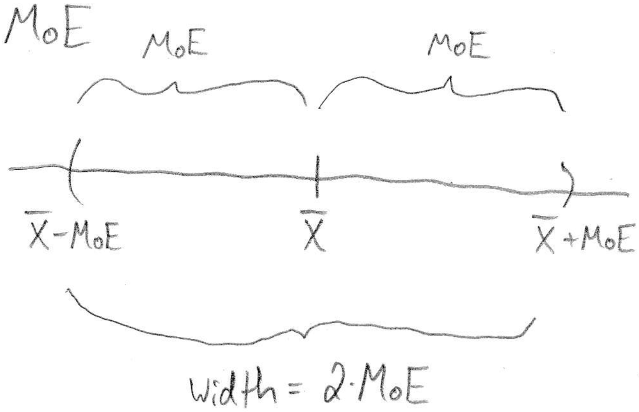
- Required Sample Sizes in CI's

How can we achieve a desired MoE?

$$MoE = z_{1-\frac{1-cl}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{\overset{\text{crit value}}{z} \cdot \sigma}{MoE} \right)^2$$

If we fix a cl, what sample size is needed to get a certain MoE?

$\bar{X} \pm MoE$



Example: #4

$\sigma = \sqrt{76}$

$cl = .95 \Rightarrow z_{.975} = 1.96$

$MoE = 3$

$$n = \left(\frac{1.96 \cdot \sqrt{76}}{3} \right)^2 = 32.44$$

$$\Downarrow$$

$n = 33$ ☆ Always Round Up! ☆

Lesson 19 - CI's for μ

- What happens if σ (population st. dev) is unknown?

Previously:

$\bar{X} \pm z \cdot \frac{\sigma}{\sqrt{n}}$

Now:

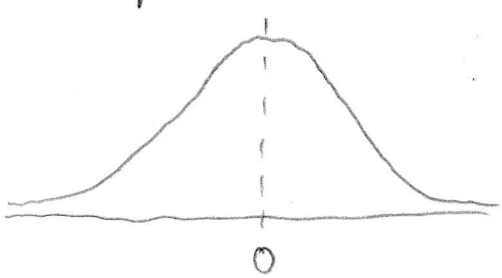
$\bar{X} \pm (\text{crit-value})\left(\frac{S}{\sqrt{n}}\right)$

come from the t-dist.

- Student's t-distribution (t)
  - ☆ Continuous dist. with support $(-\infty, \infty)$
  - ☆ Parameter - known as "degrees of freedom" (df)
    - calculated as $df = n-1$

sample size

  - ☆ Shape/Look - picture of pdf



$t \neq$ Normal!

b/c t-dist has more probability in the tails

i.e, more chance for extreme events!

As $n \to \infty$, $t \to$ Normal

  - ☆ Percentiles/ Critical Values

$$t_{1-\frac{1-cl}{2}} = qt\left(1-\frac{1-cl}{2}, df\right)$$

$n-1$

t-crit value

R code for t-crit values

Lesson 19 - Continued

Confidence Intervals for a population mean μ.

☆ When population (σ) st. dev is known

$$\bar{X} \pm z_{1-\frac{1-cl}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad ; \quad z_{1-\frac{1-cl}{2}} = qnorm\left(1-\frac{1-cl}{2}\right)$$

☆ When population (σ) st. dev is unknown, use S = sample st. dev

$$\bar{X} \pm t_{1-\frac{1-cl}{2}, df} \cdot \frac{S}{\sqrt{n}} \quad ; \quad t_{1-\frac{1-cl}{2}, df} = qt\left(1-\frac{1-cl}{2}, df\right)$$

$$df = n-1$$

Irisin | $\bar{X} = 4.3$, $S = .4$, $n = 6$, $cl = .90$

a) $t_{.95,5} = qt(.95, 5) = 2.015$

$$4.3 \pm (2.015)\left(\frac{.4}{\sqrt{6}}\right) = (3.971, 4.629)$$

We are 90% confident that the true mean Irisin level is between 3.971 mg/ml and 4.629 mg/ml.

☆ $\mathbb{E}[\bar{X}] = \mathbb{E}[X]$

- Prediction Interval

What if we want come up with an interval estimator for an observation rather than a parameter.

What has to change from the previous CI's?

☆ We have to change the MoE, specifically the variability is increased.

Prediction Interval Formula:

$$\underbrace{\bar{X} \pm t_{1-\frac{1-cl}{2},df} \cdot S \cdot \sqrt{1+\frac{1}{n}}} = \bar{X} \pm t_{1-\frac{1-cl}{2},df} \cdot S \cdot \sqrt{\frac{n+1}{n}}$$

CI for μ: $\bar{X} \pm \underbrace{t_{1-\frac{1-cl}{2},df} \cdot \frac{S}{\sqrt{n}}}_{MoE}$ $\rightarrow S \cdot \sqrt{\frac{1}{n}}$

Irisin $\bar{X} = 4.3,\ S = .4,\ n = 6,\ cl = .90$

b) $t_{.95,5} = 2.015$

$4.3 \pm (2.015)(.4)\sqrt{\frac{7}{6}} = (3.4299, 5.1701)$

We are 90% confident that the irisin level for a randomly selected subject will be between 3.4299 mg/ml and 5.1701 mg/ml.

Compare to CI for μ: $(3.971, 4.629)$

Lesson 22 - Hypothesis Testing Introduction

- What is a hypothesis test? How do we use these for estimation? (or inference)

→ ☆ First, make an assumption about parameters

☆ Second, collect data & quantify the probability of observing that sample under our initial assumption.

☆ If the probability we calculate is <u>too small</u>, we think that the initial assumption <u>is not supported</u>.

- Setting up Hypotheses

Always two hypotheses:

Null H_0: Initial Assumption ← will be able to use this for calculating probabilities.

Alternative H_a: Research Hypothesis

↳ "Thing we think we have evidence for"
"Thing we expect to see evidence for"

Rules:

□ Hypotheses should be in terms of a parameter

$H_0: \mu = 3$

□ Null and Alternative hypotheses are complements of each other

$H_0: \mu = 3$ $\qquad$ $H_0: \mu \leq 3$

$H_a: \mu \neq 3$ $\qquad$ $H_a: \mu > 3$

2. a) $H_0: \mu \leq 260$ → μ less than or <u>equal</u> to 260

$H_a: \mu > 260$

true mean reaction time

b) $H_0: \mu \geq 707$

$H_a: \mu < 707$

c) $H_0: \mu = 345$

$H_a: \mu \neq 345$

d) $\cancel{H_0: \mu \neq 35}$ $H_0: \mu = 35$

$\cancel{H_a: \mu = 35}$ $H_a: \mu \neq 35$

3.

		Reality	
		H_0 true	H_0 False
our conclusion	Reject H_0	Type 1 Error	
	Fail to Reject H_0		Type 2 Error

Type 1 Error - Throwing away a good assumption.

Type 2 Error - Keeping a bad assumption

- $\mathbb{P}$(Type 1 Error) = significance level = α
  - ☆ In STAT315, sig. level will be given
  - ☆ If not stated, use $\alpha = .05$

Lesson 23 - Hypothesis Testing Continued

- **Test Statistic** (Step T in our procedure)

General Test-stat Formula $= \dfrac{\hat{\theta} - \theta_0}{SE[\hat{\theta}]}$; $\hat{\theta}$ = point estimate of θ

θ_0 = value from the null hypothesis (initial assumption about θ)

Test-statistic for μ $= \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ or $\dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

population st. dev (σ) sample st. dev (s)

- Distribution of the test statistic

$$Z_{test} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

★ when original data is normal or when CLT applies.

$$t_{test} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{\underbrace{n-1}_{df}}$$

← t-dist with n-1 df.

- p-value

Definition: probability of witnessing a test statistic "as or more extreme" than the one we observed, under our null hypothesis.

Intuitively, the p-value is the probability of seeing our data, assuming H_0 is true.

- Result: compare p-value & significance level (α)

  If p-value $< \alpha$, then we reject H_0

  If p-value $> \alpha$, then we fail to reject H_0

- Conclusion Step

  ☆ Stating results in the context of the problem ☆

3. a) We reject H_0 at the .05 significance level, and conclude that we have evidence for the true mean being different from 7.

   c) We fail to reject H_0 at the .05 significance level, and conclude that we do not have evidence for the true mean being less than 4.1.

In the book, you will see

→ $H_0: \mu = 5$

$H_a: \mu > 5$

Reaction

treat it as $\mu = 260$

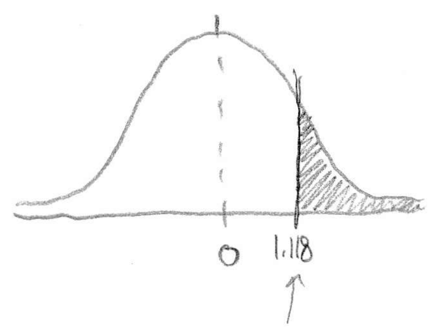
N $H_0: \mu \le 260$

A $H_a: \mu > 260$ → μ_0

T $Z_{test} = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{270-260}{40/\sqrt{20}} \approx 1.118 \sim N(0,1)$ D

R

p-value $= \mathbb{P}(Z > 1.118) \approx .1318 > .05$



C Therefore, we FTR H_0, and conclude that we do not have evidence that the true mean reaction time is greater than 260 ms.

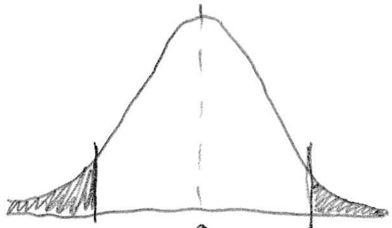
Lesson 24 - Hypothesis Tests & Confidence Intervals

Lesson 23, Inflation

Ⓝ $H_0: \mu = 35$

Ⓐ $H_a: \underbrace{\mu \neq 35}$

"Two-tailed test"



symmetric, so the probabilities shaded above are equal.

Ⓣ $t_{test} = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} = \dfrac{36.2 - 35}{2.3/\sqrt{20}} \approx 2.3333 \sim t_{19}$ Ⓓ

t_{19} ← n-1

$$\text{p-value} = 2 \cdot \mathbb{P}(t_{19} < {}^-2.3333)$$

$$= 2 \cdot pt({}^-2.3333, 19) \approx .0308 < .05$$ Ⓡ

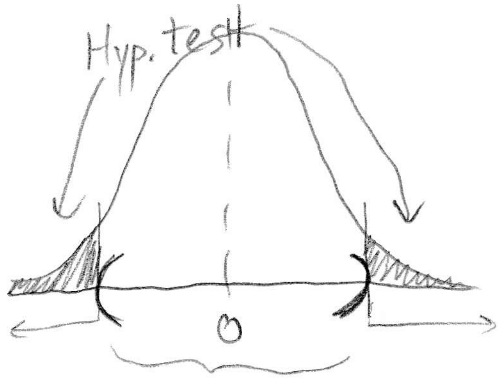
Ⓒ We reject H_0 at the .05 sig. level and conclude that we have evidence for the mean inflation pressure being different from 35 psi.

☆ t-distribution is a normal dist. divided by a Chi-squared dist. ☆

Chi-squared dist. → Squared Standard Normal.

Sample st. dev ⟹ t-dist always!

- Connection between CI's & Hypothesis Tests
  - Every CI corresponds to a two-tailed hyp. test.



A CI is a range of good estimates.

$$cl + \alpha = 1$$

⇒ A 95% CI corresponds to a two-tailed test at the .05 sig. level.

1. a) 95% CI = (18.7207, 21.2793)

b) Ⓝ $H_0: \mu = 22$

Ⓐ $H_a: \mu \neq 22$

Ⓒ Since 22, the hypothesized value, is not in our 95% CI for the mean, we reject H_0 at the .05 significance level and conclude that we have evidence for μ being different from 22.

- Inference for Proportions

p = population proportion

$\hat{p}$ = sample proportion

p_0 = hypothesized value

☆ For proportion inference, always use z.

CI: $\hat{p} \pm z_{1-\frac{1-cl}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Hypothesis Tests: $H_0: p \leq .5$ $\quad$ Ⓣ $z_{test} = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim$ Ⓓ $N(0,1)$

$H_a: p > .5$

$\boxed{\text{ADHD}}$ $H_0: p = .03$ $\quad \hat{p} = \frac{42}{1494}$

$H_a: p \neq .03$

Lesson 25 - Comparing Two Means

1. We want to estimate $\mu_1 - \mu_2$.

Point Estimate: $\bar{X}_1 - \bar{X}_2$; $\bar{X}_i$ = sample mean from population i.

Standard Error: $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (samples from the two populations are ind.)

$$SE[\bar{X}_1 - \bar{X}_2] = \sqrt{Var[\bar{X}_1 - \bar{X}_2]} = \sqrt{Var[\bar{X}_1] + Var[\bar{X}_2]} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

→ Assuming independence of $\bar{X}_1$ & $\bar{X}_2$.

2. When σ_1 & σ_2 are known:

CI: $(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{1-cL}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Test-statistic: $$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

→ hypothesized value

When σ_1 & σ_2 are unknown:
(only have s_1 & s_2)

CI: $(\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{1-cL}{2}, dF} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

test-statistic: $$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{dF}$$

→ use Welch-Satterthwaite approx.

Degrees of Freedom

- Maximum df (smallest CI width) = $n_1 + n_2 - 1$

★ Middle df (Best df estimate) = Welch-Satterthwaite df (approximation)

$$df = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2} \right\rfloor \rightarrow \text{round down!}$$

★ Use when $n_1 \neq n_2$ ★

- Minimum df (largest CI width) = $\min(n_1 - 1, n_2 - 1)$ ← minimum
  (use if unsure to use the others)

$\boxed{\text{Stopping}}$

→ $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

qt(.975, 14)

$$95\%\ CI: \quad (126.8 - 119.3) \pm (2.1448)\sqrt{\frac{5.4^2}{8} + \frac{5^2}{10}} = (2.1832, 12.8186)$$

Zero is not in the CI ⇒ Reject H_0!

Lesson 27 - ANOVA

- Here, we are testing for differences between many (more than 2) group/population means

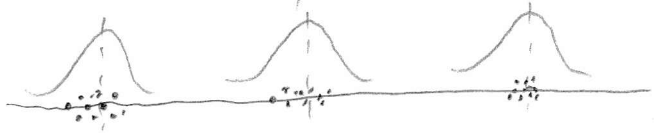
Ⓝ $H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ (k groups)

Ⓐ H_a: At least one mean is different

Notation:

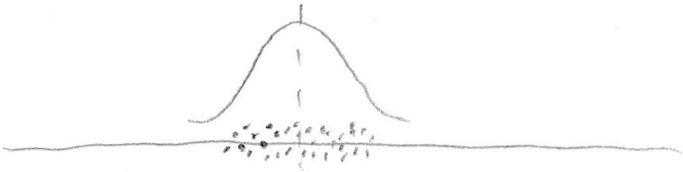
- k = # of groups
- N = overall sample size
- n_i = sample size of group i
- $\bar{X}_{..}$ = mean of the overall data set
- $\bar{X}_{i.}$ = mean of the sample from group i
- X_{ij} = j^{th} data value from the i^{th} group
- S_i^2 = sample variance from group i.

ANOVA = Analysis of Variance

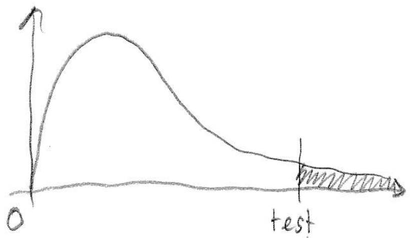


☆ Within each sample, st. dev is small

☆ Between groups, the st. dev is large



- F - distribution
  - ☆ Continuous → $pf(\text{test-stat}, \ldots)$ ← In R
  - ☆ Parameters: Numerator df & Denominator df
  - ☆ Notation: F_{df_1, df_2}
  - ☆ F-distribution is strictly positive



- Formulas

SSTR (sum of squares treatment) $= \sum_{i=1}^{k} n_i (\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$ (within group variation)

SSE (sum of squares error) $= \sum_{i=1}^{k} (n_i - 1) s_i^2$

$df_{SSTR} = k - 1$

$df_{SSE} = N - k$

$MSTR = \frac{SSTR}{k-1}$; $MSE = \frac{SSE}{N-k}$; $F_{test} = \frac{MSTR}{MSE} \sim F_{k-1, N-k}$

Source	df	SS	MS	F_{test}
Treatment	$k-1$	SSTR	$\frac{SSTR}{k-1}$	$\frac{MSTR}{MSE}$
Error	$N-k$	SSE	$\frac{SSE}{N-k}$	×
Total	$N-1$	SSTO $= SSTR + SSE$	×	×

Lesson 28 - ANOVA Part 2 (Multiple Comparisons)

Hypotheses from ANOVA:

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$

H_a: At least one mean is different

☆ Whenever we reject H_0 in ANOVA, we need to use multiple comparisons that is, we want to know which means are different!

For us, multiple comparisons will be CI's for $\mu_i - \mu_j$

↳ simultaneous comparisons

Imagine if we constructed independent CI's:

$\mathbb{P}$(All m CI's capture the true values)

$= \mathbb{P}(\text{1 CI captures the true value})^m = (cl)^m$

we have 6 CI's at the 95% cl.

$\Rightarrow$ the cl for all of our intervals is $(.95)^6 < .95$

For STAT315, we use the Tukey-Kramer method for mult. comp.'s

Here, we use a Q-distribution to account for the simultaneous CI's.

- always positive
- Parameters: k and $\sum_{i=1}^{k}(n_i - 1)$

# of groups

- Critical Value: $q_{cl, k, \Sigma(n_i - 1)} = \text{qtukey}(cl, k, \sum_{i=1}^{k}(n_i-1))$

get crit value in R

☆ We construct one CI for each group comparison:

$$(\bar{X}_{i\cdot} - \bar{X}_{j\cdot}) \pm q_{cl,k,\Sigma(n_i-1)} \sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

sample mean from group i

$\boxed{\text{Flicker}}$ $MSE = 2.394$

$q_{.95,3,16} = \text{qtukey}(.95,3,16) = \underline{3.6491}$

CI for $\mu_1 - \mu_2$: $(25.6 - 26.9) \pm 3.6491\sqrt{\frac{2.394}{2}\left(\frac{1}{8}+\frac{1}{5}\right)} = (-3.576, .976)$

" " $\mu_1 - \mu_3$: $= (-4.7561, -.4439)$

" $\mu_2 - \mu_3$ $= (-3.7175, 1.1175)$

Since the 95% CI for $\mu_1 - \mu_3$ does not contain zero, we have evidence that $\underline{\mu_1 \neq \mu_3}$ // Problem is completed

$\mu_1 = \mu_2$

$\mu_1 \neq \mu_3$

$\mu_2 = \mu_3$

Lesson 30 - Intro to Regression

- **Regression Analysis** – attempting to model a relationship between variables (random)
  - we try to fit a function to our data

  Example: Linear $y = 3 + 5x$

- **Deterministic Model**: relate variables in a fixed way

- **Probabilistic Model**: relate variables with a function, but there is error apparent.

  Consider: X & Y and we fit the function

  $$\boxed{y = f(x) + \varepsilon} \quad ; \quad \varepsilon \sim \text{Normal}(0, \sigma)$$

  epsilon = error

- **Simple Linear Regression (SLR)**
  - 1 response variable, y
  - 1 explanatory variable, x (predictor)

  ⇒ Want to use x to predict y

Model: $y = \beta_0 + \beta_1 x + \varepsilon \quad ; \quad \varepsilon \sim N(0, \sigma)$

intercept slope (parameters)

Fitted Line (Regression Line): $\hat{y} = b_0 + b_1 x$ (Error is removed)

Comes From Data

predicted y-value, estimated intercept, estimated slope

Lesson 31 - SLR Continued

☆ SLR Model: $y = \beta_0 + \beta_1 X + \varepsilon$; $\varepsilon \sim N(0, \sigma)$

β_0, β_1: intercept & slope parameters

☆ Fitted Line (Regression Line): $\hat{y} = b_0 + b_1 X$

b_0, b_1: estimates/statistics

Lesson 30 #7

$$y = \overset{\beta_0}{3} + \overset{\beta_1}{5}x + \varepsilon \; ; \; \varepsilon \sim N(0, 4)$$

Linear models always have $\mathbb{E}[\varepsilon] = 0$.

$$\Rightarrow y \sim N(\underline{3+5x}, 4)$$

c) $\mathbb{E}[Y \mid X=6] = 3 + 5(6) = 33$

d) $\mathbb{P}(Y < 30 \mid X=6) = \text{pnorm}(30, 33, 4)$

b) Interpret $\beta_1 = \overset{\text{true}}{\text{slope}}$

$y = \beta_0 + \beta_1(x)$

$y = \beta_0 + \beta_1(x+1)$ $\Rightarrow \beta_1$

β_1 is the expected <u>change</u> in y when x increases by one unit.

We expect y to increase by 5 when x is increased by 1.

- How do we calculate b_0 & b_1?

– Minimizing a "Loss" function

$$l(b_0, b_1) = \sum_{i=1}^{n} (\underset{\text{observed}}{y_i} - \underset{\text{predicted}}{\hat{y}_i})^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2$$

$$\Rightarrow \quad b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{S_{xy}}{S_{xx}} \qquad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \sum (x_i - \bar{x})^2$$

- **Using the Fitted Line**

At this step, we have $\hat{y} = b_0 + b_1 x$

☆ Prediction: $\hat{y}$ @ a certain x-value

☆ Residual: $\underset{\text{observed y-value}}{y_i} - \underset{\text{predicted y-value}}{\hat{y}_i}$

☆ Model Utility Test

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

$y = \beta_0 + \beta_1 X + \varepsilon$

$\beta_1 = 0 \Rightarrow y = \beta_0 + \varepsilon$

NOT USEFUL

Lesson 35 - Regression Diagnostics

- Assumptions of the SLR Model
  - Linearity
  - Normality (error term)
  - Homoskedasticity (constant variance)

- How do we check that these are met?
  Answer: with visualizations.
  - To check linearity: ☆ look at a scatterplot of data (how linear is it?)
    ☆ look at residual plot (we want no pattern)
  - To check Normality: ☆ look at QQ-plot
    ↳ quartile vs. quartile
    QQ-plot: graphs the quantiles (percentiles) of the residuals against quantiles of the normal distribution.
    ☆ We want QQ-plot to be <u>linear!</u> ☆
    ↳ closer this is to a line, the more "normal" it is.
  - To check homoskedasticity: ☆ look at residual plots
    plot residuals against x-values
    We want the spread of residuals to be the same across the x-values

5. Polynomial Regression

(specifically, quadratic regression)

Model: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$; $\varepsilon \sim Normal(0, \sigma)$

Fitted Curve: $\hat{y} = b_0 + b_1 x + b_2 x^2$

Coefficient of Multiple Determination $(R^2) = 1 - \frac{SSE}{SST} = \frac{SSR}{SST}$ — This can only increase as more predictors are added.

☆ <u>Adjusted</u> Coeff. of Mult. Determination $(R_a^2) = 1 - \left(\frac{n-1}{n-k-1}\right)\left(\frac{SSE}{SST}\right)$

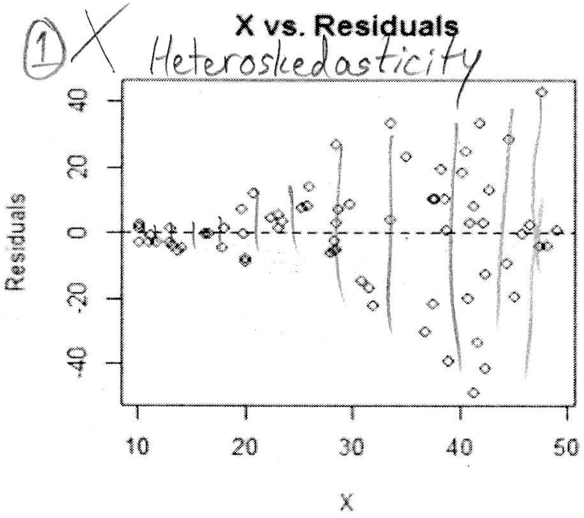
k = # of predictor variables

CI's for β_i: $b_i \pm t_{1-\frac{1-cl}{2}, df} \cdot SE[b_i]$ $(df = n - p)$

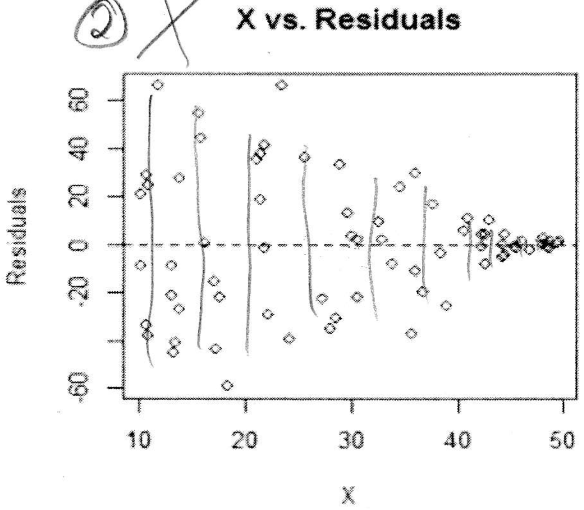
p = # of coefficients being estimated in the model.

Examples of residual plots

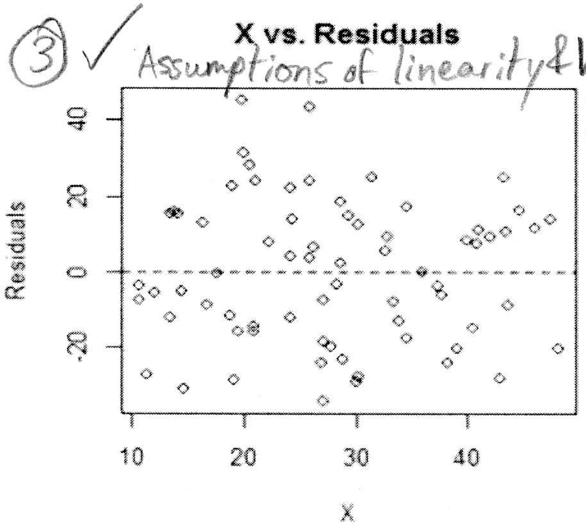
① ✗ Heteroskedasticity



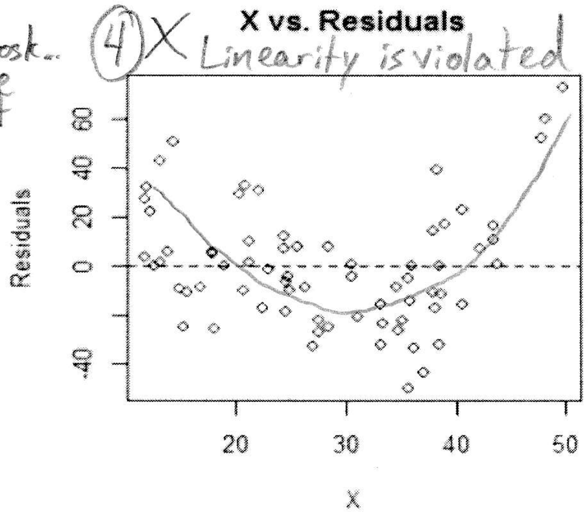
② ✗



③ ✓ Assumptions of linearity & homosk... are met



④ ✗ Linearity is violated



Lesson 36 & 37 - Multiple Regression (Linear)

(mult. linear regression)

Goal: In MLR, we want to explain the response variable Y using $X_1, \dots, X_p$ (p predictor variables) and a (linear) relationship such as:

(Model:) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$; $\varepsilon \sim Normal(0, \sigma)$

- Fitted plane: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$
- Interpretations:

  β_i : Expected change in Y when X_i is increased by 1,
  ($i = 1, 2, \dots, p$) while holding all other predictor variables constant.
  (while controlling for other predictor variables)

- Model Utility Test (Regression ANOVA)

  Ⓝ $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$

  Ⓐ H_a: At least one is nonzero

Source	df	SS	MS	Ⓣ F
Regression	p	SSR	$\frac{SSR}{p} = MSR$	$\frac{MSR}{MSE}$ Ⓓ $\sim F_{p, n-p-1}$
Error	$n-p-1$	SSE	$\frac{SSE}{n-p-1} = MSE$	×
Total	$n-1$	SST	×	×

p-value = $1 - pf\left(\frac{MSR}{MSE}, p, n-p-1\right)$

- Adjusted R^2

  $$R_a^2 = 1 - \left(\frac{n-1}{n-p-1}\right)\left(\frac{SSE}{SST}\right)$$

  ☆ Use to compare models! ☆

  always between 0 & 1

• Indicator Variables: just a variable (predictor) that is 0 or 1.
0 if a specified event doesn't occur
1 if " " " does occur

• Interaction Effects: a variable (a predictor) that is the product of an indicator variable and a "usual" predictor.

Suppose $X_1 = 0$ for region A
$X_1 = 1$ for region B
$\Longrightarrow$ Then $X_1 \cdot X_2$ (population size) $\to$ 0 (for region A)
$\to$ X_2 (for region B)

$\boxed{\text{Sonic}}$ a) $\hat{y} = 9.9 + 5x_1 + 9.6x_2$

b) $\hat{y}_{screwdriver} = 9.9 + 5x_1 + 0 = 9.9 + 5x_1$

$\hat{y}_{blaster} = 9.9 + 5x_1 + 9.6 = 19.5 + 5x_1$

Two different intercepts!

Example: Suppose $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1 X_2 + \varepsilon$

We fit: $\hat{y} = 9.9 + 5x_1 + 9.6x_1x_2$

$\hat{y}_{screwdriver} = 9.9 + 5x_1$

$\hat{y}_{blaster} = 9.9 + 5x_1 + 9.6x_1 = 9.9 + 14.6x_1$