Chapter 1

- **Population:** entire collection of outcomes
- **Sample:** Subset of population that is obs.
- Conceptual vs. Tangible Population: Measuring 1 object 5 times vs. 5 diff objects
- Controlled Experiment vs Observational Study

- **Sample Mean**

  $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

- **Deviation:** Distance from mean

  → $(x_1 - \bar{x}), (x_2 - \bar{x}), \dots (x_n - \bar{x})$

- **Sample Variance**

  $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

- **Sample Stdev:**

  $s = \sqrt{s^2}$ — units in terms of orig.

- **Linear Transformations of Datasets:**

  → $\bar{y} = a + b\bar{x}$ $(y_i = ax_i + b)$

  → $s_y^2 = b^2 s_x^2$

  → $s_y = |b| s_x$

- **Outliers:**

  Can't just throw out, must examine

- **Sample Median**

  Middle number when sorted

  Resistant to outliers

  Special Case of Trimmed Mean

- **Quartiles**
  - Q_1: $(0.25)(n+1)$
  - Q_3: $(0.75)(n+1)$
  - Q_2: median
  - IQR: $Q_3 - Q_1$

Describing Data:

- **Shape**

  → mode(s): unimodal, bimodal, etc.

  → Symmetry: symm., skew, etc.

- **Outliers**

  → unusual observations

- **Center**

  → mean / median

- **Variability**

  → StDev / IQR

VSCO

Use mean + stdev when unimodal, ~symmetric

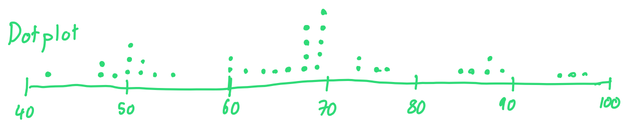
Graphical Summaries:

- **Stem & Leaf Plot**

Key: 5|0 = 5.0

Stem	Leaf
0	13678
1	012225667
2	0267
3	04
4	39

- **Dotplot**



- **Histogram**

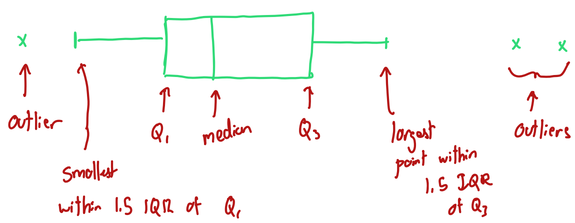
  → Like dotplot but w/ intervals - classes of equal width (partitions)

  → Frequency (count), relative frequency (%)

- **Boxplot**
  - 5 number summary: min, Q_1, median, Q_3, max
  - Generally:

    → Outlier if more than 1.5 IQR from Q_1/Q_3

    → Extreme outlier if more than 3 IQR from Q_1/Q_3



Chapter 2

- Experiment: Process that results in outcome that cannot be predicted w/ certainty.
- Sample Space (S): Set of all possible outcomes
- Event (E): Subset of S (collection of outcomes)
  - Simple $\Rightarrow |E| = 1$; Compound $\Rightarrow$ otherwise

Axioms of Probability:
- $P(S) = 1$
- $0 \le P(E) \le 1$
- If $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

  Generally, $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$ for mutually exclusive events

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(B \cap A^c) = P(B) - P(B \cap A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- De Morgan's: $P((A \cup B)^c) = P(A^c \cap B^c)$
  $P((A \cap B)^c) = P(A^c \cup B^c)$
- $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$
- Law of Total Probability:

  If $A_1, A_2, \dots, A_n$ mutually exclusive and exhaustive, then:

  $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \dots + P(A_n \cap B)$

  $P(B) = P(B \mid A_1) P(A_1) + P(B \mid A_2) P(A_2) + \dots$

- Bayes:

  $P(B) = P(A \cap B) + P(A^c \cap B)$

  $P(B) = P(B \mid A) P(A) + P(B \mid A^c) P(A^c)$

- Independence:

  Iff A, B independent:

  $P(A \mid B) = P(A)$

  $P(B \mid A) = P(B)$

  $P(A \cap B) = P(A) P(B)$

- Mutual Independence:

  $P(A_i)$ is same no matter what other A_j s occur.

  Can multiply probabilities.

Random Variables:

PMF (discrete): $P(X = x)$

CDF (both): $P(X \le x)$

- for discrete, "steps"

Discrete:

μ: $E[X] = \sum_x x \cdot P(X = x)$

$Var(X) = \sigma^2 = E[(X - \mu)^2]$

$= \sum_x (x - \mu)^2 P(X = x)$

$SD(X) = \sqrt{Var(X)}$

$Var(X) = E[X^2] - (E[X])^2$

Continuous:

PDF: $P(X = x)$, $f(x)$ – not actual probability

$P(a < x < b) = \int_a^b f(x)\,dx$

$\int_{-\infty}^{\infty} f(x) = 1$

$f(x) \ge 0 \quad \forall x$

CDF: Cumulative Dist. Func

$F(x) = \int_{-\infty}^{x} f(t)\,dt$

Mean:

$E[X] = \mu = \int_{-\infty}^{\infty} x f(x)\,dx$

$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,dx$

Linear Functions of R.V.s

1. $E[aX + b] = a E[X] + b$

   $Var(aX + b) = a^2 Var(X)$

   $SD(aX + b) = |a| SD(X)$

2. Independent RVs $X_1, X_2, \dots, X_n$

   $E[c_1 X_1 + c_2 X_2 + \dots] = c_1 E[X_1] + c_2 E[X_2] + \dots$

   $Var(c_1 X_1 + c_2 X_2 + \dots) = c_1^2 Var(X_1) + c_2^2 Var(X_2) + \dots$,

   Note: $E[X - Y] = E[X] - E[Y]$

   $Var(X - Y) = Var(X) + Var(Y)$

3. $X_1, X_2, \dots, X_n$ SRS from population μ, σ^2

   $T = X_1 + X_2 + \dots + X_n$

   $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$

   $E[T] = n E[X_i]$ $\qquad E[\bar{X}] = E[X]$

   $Var(T) = n Var(X_i)$ $\qquad Var(\bar{X}) = Var(X)/n$

   $SD(T) = \sqrt{n}\, SD(X)$ $\qquad SD(\bar{X}) = SD(X)/\sqrt{n}$

   Note: In SRS case, DO NOT just use $X_1 + X_2 + \dots + X_n = nX$. Diff SDs.

$SD(\bar{X}) = \dfrac{SD(X)}{\sqrt{n}}$

JOINT DISTRIBUTIONS OF R.Vs

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

DISCRETE:

$$\sum_x \sum_y p(x,y) = 1$$

$$P_X(x) = P(X=x) = \sum_y p(X=x, y)$$

$$P_Y(y) = P(Y=y) = \sum_x P(x, Y=y)$$

CONTINUOUS

$$P(a \le X \le b \cap c \le Y \le d) = \int_a^b \int_c^d f(x,y)\, dy\, dx$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

$$E[h(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) f(x,y)\, dy\, dx$$

$$\sum_x \sum_y h(x,y) f(x,y)\, dy\, dx$$

CONDITIONAL

$$P(Y|X) = \frac{p(x,y)}{P_X(x)}$$

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

INDEPENDENCE

X, Y independent iff:

$$P_{XY}(x,y) = P_X(x)\, P_Y(y)$$

$$f_{XY}(x,y) = f_X(x)\, f_Y(y)$$

Then,

$$P_{X|Y=y}(x) = P_X(x)$$

$$f_{X|Y=y}(x) = f_X(x)$$

CHAPTER 3

BIAS: Systematic error in measurement

RANDOM ERROR: differs between measurements, averages out to 0.

MEASURED $\approx$ TRUE + BIAS + RANDOM ERROR

$\mu =$ TRUE + BIAS

$\sigma = \sigma_{\text{MEASUREMENT ERROR}}$

ACCURACY: $\mu -$ TRUE $=$ BIAS

PRECISION: Uncertainty, σ

Chapter 4

4.1: Bernoulli Distribution

X : 0 for failure, 1 for success
$(1 - p = q)$ (p)

$E[x] = p$

$Var(x) = p(1-p)$

Conditions to be Bernoulli:

- Binary Success / Failure
- Independent Trials

4.2: Binomial Distribution

X : # of successes in n Bernoulli trials

$X \sim Bin(n, p)$

- 5% rule: Samples considered independent if $n \leq 5\%$ of population

5% rule

VSCO — athleen
drive shop cash

- $P(X=x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$

$\mu = np$
$s^2 = np(1-p)$

- $E[x] = np$
- $Var[x] = np(1-p)$
- Sample Proportions (estimate p):

$\hat{p} = \frac{\# \text{ success}}{\# \text{ trials}}$

$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = SE(\hat{p})$

$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = SE(\hat{p})$