

Final Review Sheet

Yeongseo Lim

Probability & Statistics for Biological & Physical Sciences & Engineering

1 Useful Laws/Theorem/Properties

1.1 DeMorgan's Laws

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

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

Note here that whenever you are distributing the complement (c), switch from union to intersection (or intersection to union) needs to be made.

1.2 Set Properties

1. Distributive properties

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

2. Commutative properties

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

3. Associative properties

$$(A \cup B) \cup C = A \cup (B \cup C)$$

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

1.3 Axioms of Probability

1. Nonnegativity

For any event A

$$P(A) \geq 0$$

2. Normalization

$$P(\Omega) = 1$$

Note from axiom 2, we obtain the complimentary rule $P(A) = 1 - P(A^c)$

3. Finite Additivity

When A_1 and A_2 are mutually exclusive events $\rightarrow (A_1 \cap A_2) = \emptyset$, then

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

3.* Countable Additivity

When $A_1, A_2, A_3 \dots$ are sequence of mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

More explicitly,

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

From Axioms of Probability, we derive

1. $P(\emptyset) = 0$
2. $0 \leq P(A) \leq 1$
3. If $A \subseteq B$, then $P(A) \leq P(B)$ (monotonicity)

1.4 Inclusion-exclusion Principle

The way that I like to remember this is that the **odd** number of intersections **add** (+) and the **even** number of intersections **subtract** (-). For example

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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

Also, note that you alternate between (-) and (+) for every even/odd intersection of events

2 Conditional Probability

Probability of event A happening given event B already occurred

$$P(A|B)$$

Given $P(B) > 0$,

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

Rearranging, we can get that

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

Note that $P(A|B) \neq P(B|A)$

3 Law of Total Probability

When B_1, B_2, \dots are mutually exclusive and exhaustive (i.e union of all $B_i = \Omega$)

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

*Note: when there are only A and B, law of total probability says

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \end{aligned}$$

4 Bayes Rule

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_k P(A|B_k)P(B_k)}$$

5 Independence

When knowing outcome of event B does not affect the outcome of event A.

$$P(A|B) = P(A)$$

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

***NOTE:**

- Mutually exclusive \neq independent
If we know one occurred, we automatically know the other one did not
- If A and B are independent
 - A and B^c
 - B and A^c
 - A^c and B^c

are also independent

6 Counting & Combinatorics

6.1 Multiplication rule

If event A can be done m ways, and for each of these outcomes event B can be done n ways, then the sequence (A, B) can be done $\mathbf{m} \times \mathbf{n}$ ways

6.2 Combinatorics

Combinatorics is a useful method for counting the cardinality of the sample size or event when it is difficult to calculate the cardinality. We will go through three cases.

6.2.1 Sample with replacement

Let's say we have a set amount of " n " possible outcomes and we want to draw a sample " k " times. For example, let us consider the case of flipping a coin ($n = 2$). In this scenario, each trial has the same probability distribution of outcome because an outcome is not removed from future trials once it is drawn. The general formula then is:

$$n^k \text{ possible outcomes}$$

6.2.2 Sample without replacement (order matters)

Let us consider a case where we remove the outcome once it has been selected. Assume there are n distinct objects, select k without replacing after each draw, and keep track of the order. Hence, **order matters**.

$${}_nP_k = \frac{n!}{(n-k)!}$$

This is called **permutation** (n permute k).

***NOTE:** When $k = n$, ${}_nP_k = n!$

More Examples:

- Rearranging objects with repeats (several types of objects, more than one object in a type)
 - Anagrams with repeat letters
- Rearranging objects with restrictions
 - Anagrams where first letter must be vowel
 - Anagrams where all vowels come before all consonants, etc

6.2.3 Sample without replacement (order does not matter)

Let us say that we are conducting an experiment where **order does not matter** i.e. $\{a, b\} = \{b, a\} \rightarrow$ they are considered the same combination

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = {}_nC_k$$

This is called **combinations** (n choose k).

7 Random Variables

A function that maps elements of the sample space of an experiment to real number values. We can call random variable X discrete if the set of possible values (measurement) we can observe from X is a discrete set, i.e. either finite set or countably infinite set.

7.1 Discrete RV: PMF (probability mass function)

Defines the probability distribution for a discrete random variable

$$P_X(x) = P(X = x)$$

Properties

1. $P(X_i) > 0$
2. $\sum_{i=1}^n P(X_i) = 1$

7.1.1 Geometric Distribution

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

For when the question asks, “what is the probability that (something) happens on the ‘ k^{th} ’ trial”?

Geometric Series:

$$\frac{ar^k}{1 - r} = \frac{\text{first term}}{1 - \text{geometric ratio}} \quad \text{when } |r| < 1$$

7.1.2 Binomial Distribution

1. **n** identical trials
2. Two possible outcomes (**success/failure**)
3. Result of each trials are **independent**
4. Drawing with replacement
5. Sample size is small (< 0.05 of the population)
6. The probability of success (**p**) is **constant** from trial to trial

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

7.1.3 Hypergeometric Distribution

1. Finite population
2. Two types in population (success/failure)
3. m are success (known), $N - m$ are failures
4. Randomly draw a sample of size n **without replacement**
5. Drawn from the population ignoring the order (\rightarrow combination)

$$\text{hypergeometric}(n, m, N)$$

$$P(X = x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

Note that for $n < 0.05N$, then $\text{Hypergeometric}(n, m, N) \approx \text{Bin}(n, \frac{m}{N})$

7.2 Continuous RV: PDF (probability density function)

1. $f(x)$ is defined for all $x \in \mathbb{R}$
2. $f(x) \geq 0$ (non-negative)
3. $\int_{-\infty}^{\infty} f(x) dx = 1$

Note that when graphed, y-axis is not the probability, but probability density (area under the curve). Hence, we can only find the probability of ranges of value. This also means that continuous RV has NO probability at discrete point (but does NOT mean the event won't happen at that point). If $P(X = c)$, then

$$\int_c^c f(x) dx = 0$$

Another useful property is that

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = P(a \leq X < b)$$

Note for **PMF**, equal signs in the inequality matters.

Additionally, if a PDF is not defined on some interval, do not integrate over that region to find probability!

7.3 Cumulative Distribution Function (CDF)

CDF can be for PDF or PMF. Broadly, CDF is defined as

$$F_X(x) = P(X \leq x)$$

To get probability from CDF, we can simply do: $P(a \leq X \leq b) = F(b) - F(a)$

PMF \rightarrow CDF:

$$F_X(x) = \sum_i^x P(X = x)$$

PDF \rightarrow CDF:

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

8 CDF Method

If Y is a transformed variable of X, we can find the PDF of Y by first finding the CDF of Y, then taking the derivative. The CDF method can be summarized into 3 steps:

1. Find possible range of transformed variable (in this case Y)
2. Find the CDF of Y
3. Take the derivative of CDF of Y to find the PDF of Y

Useful **calculus equations** for CDF method

$$\frac{d}{dx} \int_0^{h(x)} g(u) du = g(h(x))h'(x)$$

$$\frac{d}{dx} \int_{g(x)}^{h(x)} f(u) du = f(h(x))h'(x) - f(g(x))g'(x)$$

9 Expected Values

If X is **discrete**

$$E(X) = \sum_x xP(X = x) = \mu$$

If X is **continuous**

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx = \mu$$

Several things to note:

- The expected value of a RV that takes on all possible values with equal probability is the average/mean
- $E(X)$ does not have to be a value the RV can actually take
- Can determine $E(X)$ from PMF, PDF, or CDF (need to convert to PMF/PDF first)
- $E(X^2)$ is NOT equal to $[E(X)]^2$

Properties

$$E(aX^n + b) = aE(X^n) + b$$

$$E(c) = c$$

where c is the constant

9.1 Law of the Unconscious Statistician (LOTUS)

If $g = g(x)$ is any function, then

If X is **discrete**

$$E(g(X)) = \sum_x g(x)P(X = x)$$

If X is **continuous**

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx$$

9.2 Linearity of Expectation

For any $X_1, X_2, X_3, \dots, X_n$

$$E(X_1 + X_2 + X_3 + \dots + X_n) = E(X_1) + E(X_2) + E(X_3) + \dots + E(X_n)$$

10 Variance

Variance measures the spread of a distribution. Higher variance means higher spread (i.e further away from the mean)

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

Properties

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

$$Var(c) = 0$$

11 Standard Deviation

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

Properties

$$\sigma_{ax+b} = |a|\sigma_x$$

12 Joint Distribution

When two (or more) random variables are defined on the same sample space Ω

12.1 Discrete Joint Distribution

The **joint PMF** of X,Y

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

Similar to other PMFs, the following **properties** hold

1. $P_{X,Y}(x_i, y_i) \geq 0$ for all i
2. $\sum_x \sum_y P_{X,Y}(x,y) = 1$

The **marginal PMF** of X,Y

$$P_X(x) = \sum_{\text{all } y} P_{X,Y}(x,y)$$

$$P_Y(y) = \sum_{\text{all } x} P_{X,Y}(x,y)$$

Independence

$$\text{Joint} = (\text{marginal})(\text{marginal})$$

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

*Note: Independence means that every element of partition of x is independent of every element of partition of y. The above equality must be true for all possible values of x,y for X,Y to be independent.

12.2 Continuous Joint Distribution

The **joint PDF** of X,Y

$$f_{XY}(x, y)$$

Similar to other PDFs, the following **properties** hold

1. $f_{X,Y}(x_i, y_i) \geq 0$ for all $x, y \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x_i, y_i) dx dy = 1$

The **marginal PDF** of X,Y

$$\begin{aligned} f_X(x) &= \int_{\text{all } y} f_{XY}(x, y) dy \\ f_Y(y) &= \int_{\text{all } x} f_{XY}(x, y) dx \end{aligned}$$

Independence

$$\begin{aligned} \text{Joint} &= (\text{marginal})(\text{marginal}) \\ f_{XY}(x, y) &= f_X(x)f_Y(y) \end{aligned}$$

for all possible values of x,y

13 Covariance

Measure of linear association between random variables

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

*Note:

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

Hence,

$$\text{Var}(X) = \text{Cov}(X, X)$$

Interpretations

1. Positive correlation

$$\text{Cov}(X, Y) > 0$$

- (a) If X(or Y) **increases**, on average Y(or X) also **increases**
- (b) If X(or Y) **decreases**, on average Y(or X) also **decreases**

2. Negative correlation

$$\text{Cov}(X, Y) < 0$$

- (a) If X(or Y) **increases**, on average Y(or X) **decreases**

(b) If X (or Y) **decreases**, on average Y (or X) **increases**

3. Uncorrelated (no linear association)

$$\text{Cov}(X, Y) = 0$$

(a) $\text{Cov}(X, Y) = 0$ **DOES NOT MEAN** X and Y are independent

(b) There are examples of random variables X and Y such that $\text{Cov}(X, Y) = 0$ where X and Y are NOT independent. One example is a standard normal Z and its square Z^2 are uncorrelated:

$$\text{Cov}(Z, Z^2) = E(Z^3) = 0$$

But Z and Z^2 are clearly dependent because once you know the value of Z , the value of Z^2 is completely determined.

Properties

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$
3. $\text{Cov}(X, bY) = b\text{Cov}(X, Y)$
4. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
5. $\text{Cov}(c(X, Y)) = \text{Cov}(cX, cY) = c^2\text{Cov}(X, Y)$
6. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$

Variance

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

Independence

If X and Y are independent

1. $E(XY) = E(X)E(Y)$
2. $\text{Cov}(X, Y) = 0$

*NOTE for (2), the inverse is NOT TRUE!

14 Correlation

Covariance gives direction (+, -, 0), but does not tell us about the strength of the relationship
Correlation gives us both direction and magnitude (correlation coefficient p)

$$P_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Correlation is always between -1 and 1

15 Moment Generating Function(MGF)

Useful for

1. Finding **moments** of X

$$M_X(t) = E(e^{tx})$$

2. **Identifying** the **distribution** of X

If $M_X(t) = M_Y(t)$, X and Y have identical distribution

To compute $E(e^{tx})$, recall LOTUS

For **discrete** random variable

$$E(e^{tx}) = \sum_x e^{tx} P(X = x)$$

For **continuous** random variable

$$E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

Finding nth moments

k^{th} derivative of the MGF evaluated at 0 corresponds to k^{th} moment of X

$$M_X^{(k)}(0) = E(X^k)$$

*NOTE: You can still use LOTUS to find k^{th} moment

1. If k is really large (or it is not MGF of common distribution), use LOTUS
2. If k is small or is MGF of common distribution (or MGF is simple to compute), use MGF

Independence

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

16 Useful Functions

16.1 Euler-Gamma Function

For $\alpha > 0$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

Things to remember:

For $\alpha > 1$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

For integer $n \geq 1$

$$\Gamma(n) = (n - 1)!$$

Important Values:

$$\Gamma(1) = \Gamma(2) = 1$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

16.2 Gamma Function

For $\alpha, \beta > 0, X \sim \text{Gamma}(\alpha, \beta)$ can be expressed as

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, & x > 0 \\ 0, & elsewhere \end{cases}$$

Since

$$\int_0^\infty f(x) dx = 1$$

We can rearrange the above equation to get

$$\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \beta^\alpha \Gamma(\alpha)$$

*NOTE: For 6.1 and 6.2, to use these properties to solve integrals quickly, your bounds HAVE TO MATCH (i.e. the integral bound has to be your essential domain)!

*NOTE: You can do this manipulation for basically any PMF or PDF to compute seemingly difficult integral/summation (refer to week 8 problem 2)

Additional Properties

$$\text{Gamma}(1, \frac{1}{\lambda}) = \text{Exp}(\lambda)$$

As $\alpha \rightarrow \infty$, $\text{Gamma}(\alpha, \beta) \rightarrow \text{Normal}$

17 Normal Distribution

Refer to the distribution table for the PDF, mean, variance, and MGF

$$X \sim \text{Normal}(\mu, \sigma^2)$$

17.1 Sum of Normal Distribution

Given two independent normal random variables X and Y, we can use the independent property of MGF to get

$$aX + bY \sim \text{Normal}(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2)$$

Note that this is doing the same thing as finding transformed function's expected value and variance.

To generalize,

$$X + Y \sim \text{Normal}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

$$X - Y \sim \text{Normal}(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$$

Note that you always ADD the VARIANCE!

By similar logic, sum of n *i.i.d* (independent and identically distributed) normal random variables $x_1, x_2 \dots, x_n$ gives

$$X_1 + X_2 + \dots + X_n \sim \text{Normal}(n\mu, n\sigma^2)$$

17.2 Standard Normal Distribution

$$X \sim \text{Normal}(0, 1)$$

$$\text{PDF : } f(z) = \frac{e^{-\frac{z^2}{2}}}{\sigma\sqrt{2\pi}} \text{ for } -\infty < z < \infty$$

$$\text{MGF : } M_Z(t) = e^{\frac{t^2}{2}}$$

17.3 Z-score Transformation

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

17.4 Standard Normal Table

Note that the table gives the probability that the random variable takes a value less than or equal to value of interest (CDF)

$$F_Z(z) = \phi(z) = P(Z < z) = P(Z \leq z)$$

Hence,

$$P(Z \leq b) = \phi(b)$$

$$P(Z \geq a) = 1 - \phi(a) = \phi(-a)$$

$$P(a \leq Z \leq b) = \phi(b) - \phi(a)$$

If Z is not on the table,

- If Z is too small,

$$\phi(z) \approx 0$$

- If Z is too large,

$$\phi(z) \approx 1$$

18 Central Limit Theorem (CLT)

Can approximate sum of ANY *i.i.d* random variables as Normal given that one of the following conditions is met

Conditions:

1. $n \geq 30$
2. $np > 5, n(1-p) > 5, np(1-p) > 5$ (for Bernoulli/Binomial)
3. When the population is symmetric about its mean (i.e. uniform), CLT works even for $n \geq 10$

When the question involves **sum of *i.i.d* random variable**

$$Z = \frac{\sum_{i=1}^n x_i - n\mu}{\sigma\sqrt{n}}$$

Note that this is again following the form for standard normal distribution (remember Z-score transformation). Hence, μ and σ here are that of x_i

When the question involves sample **mean**

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

Similarly, μ here is just the expected value of the sample mean, and $\frac{\sigma}{\sqrt{n}}$ is the standard deviation of the sample

18.1 Continuity Correction

If x_i is **DISCRETE** and **INTEGER-VALUED**, we should apply continuity correction. This is because we are approximating a discrete random variable to a continuous distribution.

$$\begin{aligned} P(X \geq A) &= P(X \geq A - 0.5) \\ P(X \leq A) &= P(X \leq A + 0.5) \\ P(X = A) &= P(A - 0.5 \leq X \leq A + 0.5) \end{aligned}$$

19 Point Estimation

19.1 Method of Moments (MoM)

1. Determine the theoretical moment
 - (a) We need as many non-zero theoretical moments as unknown parameters
2. Equate each k^{th} theoretical moment to its corresponding sample moment, and solve for the unknown parameter
3. Plug in the sampled values to find values of the sample moments and solve for parameters

The k^{th} **sample** moment

$$m_k = \frac{\sum_{i=1}^n x_i^k}{n}$$

The k^{th} **theoretical** moment

$$E(X^k)$$

*NOTE:

- Don't leave your point estimation in m_k (expand them out)
- Put $\hat{}$ to indicate that you are estimating the parameter (i.e. $\hat{\alpha}$)

19.2 Maximum Likelihood Estimate (MLE)

Finding the value(s) of the parameter(s) that maximizes the likelihood of the random sample we observe
If **PMF**

$$L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n P_{x_i}(x_i|\theta)$$

If **PDF**

$$L(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f_{x_i}(x_i|\theta)$$

1. Find likelihood function $L(\theta)$
2. Find θ where $L(\theta)$ is maximized
3. If given x_1, \dots, x_n , plug them into your function for the estimated parameter

For finding where $L(\theta)$ is maximized

1. If the essential domain does not include the unknown parameter
 - Use $l(\theta) = \ln(L(\theta))$
 - Set $l'(\theta) = 0$
2. If the essential domain includes the unknown parameter
 - Refer to my week 11 slide (p23-)

Helpful log properties:

$$\begin{aligned} \ln(a^b) &= b(\log(a)) \\ \ln(ab) &= \log(a) + \log(b) \\ \ln\left(\frac{a}{b}\right) &= \log(a) - \log(b) \\ \ln(e^a) &= a \end{aligned}$$

*NOTE:

- Put $\hat{\cdot}$ to indicate that you are estimating the parameter (i.e. $\hat{\alpha}$)

20 Confidence Interval

Interpretation: we are $100(1-\alpha)\%$ confident that the true population parameter will fall within the confidence interval

20.1 Estimating Population Mean (1 sample)

Known σ

$$P(\bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Unknown σ

$$P(\bar{X}_n - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}) = 1 - \alpha$$

20.2 Estimating Population Mean (2 samples)

Known σ_1^2 and σ_2^2

$$P((\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}) = 1 - \alpha$$

Unknown σ_1^2 and σ_2^2 but are equal

$$P((\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2}, n_1+n_2-2} \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2}, n_1+n_2-2} \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}) = 1 - \alpha$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Unknown σ_1^2 and σ_2^2 and are unequal

$$P((\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2}, k} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2}, k} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}) = 1 - \alpha$$

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

20.3 Estimating Population Variance (1 sample)

Known μ

$$P\left(\frac{(n)\tilde{S}^2}{\chi_{\frac{\alpha}{2}, n}^2} \leq \sigma^2 \leq \frac{(n)\tilde{S}^2}{\chi_{1-\frac{\alpha}{2}, n}^2}\right) = 1 - \alpha$$

Unknown μ

$$P\left(\frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}\right) = 1 - \alpha$$

20.4 Population Variance (2 samples)

$$P\left(\frac{S_1^2}{S_2^2} F_{1-\frac{\alpha}{2}, n_2-1, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}, n_2-1, n_1-1}\right) = 1 - \alpha$$

20.5 Estimating Population Proportion

$$P(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = 1 - \alpha$$

*NOTE: When \hat{p} is not given, we can take a conservative approach to set $\hat{p} = 0.5$

21 Hypothesis Testing

21.1 Errors

21.1.1 Type 1 Error

Errors in rejecting H_0 when H_0 is true

$$P(\text{reject } H_0 | H_0 \text{ is true})$$

21.1.2 Type 2 Error

Errors in not rejecting H_0 when H_a is true

$$P(\text{fail to reject } H_0 | H_a \text{ is true})$$

21.2 Terminology

- α = type 1 error = significance level
- β = type 2 error = $1 - \text{power}$
- Power = the probability that we would correctly reject the null hypothesis
- P-value = probability of observing a value at least as extreme as your "observed value" assuming that the null hypothesis is correct

21.3 Hypothesis Testing Summary

1. Set up hypothesis
2. Find the critical region
3. Find the test statistic
4. Evaluate the result
 - Reject H_0 if P-value is $< \alpha$. Otherwise, fail to reject H_0
 - Reject H_0 if within the critical region. Otherwise, fail to reject H_0

What test to use:

Sample	Desired	Known	Test
1	μ	σ^2	Z
1	μ	s^2	T
1	σ^2	μ	χ^2_n
1	σ^2		χ^2_{n-1}
2	$\mu_1 = \mu_2$	$\sigma_1^2 = \sigma_2^2$	Pooled T-test
2	$\mu_1 = \mu_2$	$\sigma_1^2 \neq \sigma_2^2$	2 sample T-test
2	$\mu_1 = \mu_2$	paired sample	Paired T-test
2	$\sigma_1^2 = \sigma_2^2$	s_1^2, s_2^2	F-test

21.4 Notation/Convention

- For **Chi** and **F** values, the cut-off represents the area to the **left** (percentile)
- For **Z** and **T** values, the cut-off represents the area to the **right**

21.5 One Sample Tests

21.5.1 1 Sample Z-Test

Testing $H_0 : \mu = \mu_0$ vs. $H_a :$

H_a	P-value	Critical region of size α
$\mu < \mu_0$	$P(Z \leq z_{obs})$	$z_{obs} \leq -Z_\alpha$
$\mu > \mu_0$	$P(Z \geq z_{obs})$	$z_{obs} \geq Z_\alpha$
$\mu \neq \mu_0$	$2P(Z \geq z_{obs})$	$ z_{obs} \geq Z_{\frac{\alpha}{2}}$

$$z_{obs} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

21.5.2 1 Sample T-Test

Testing $H_0 : \mu = \mu_0$ vs. $H_a :$

H_a	P-value	Critical region of size α
$\mu < \mu_0$	$P(T \leq t_{obs})$	$t_{obs} \leq -t_{\alpha, n-1}$
$\mu > \mu_0$	$P(T \geq t_{obs})$	$t_{obs} \geq t_{\alpha, n-1}$
$\mu \neq \mu_0$	$2P(T \geq t_{obs})$	$ t_{obs} \geq t_{\frac{\alpha}{2}, n-1}$

$$t_{obs} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

21.5.3 Chi-square Test

Testing $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_a :$

Known μ

H_a	P-value	Critical region of size α
$\sigma^2 < \sigma_0^2$	$P(\chi^2 \leq \chi^2_{obs})$	$\chi^2_{obs} \leq \chi^2_{1-\alpha, n}$
$\sigma^2 > \sigma_0^2$	$P(\chi^2 \geq \chi^2_{obs})$	$\chi^2_{obs} \geq \chi^2_{\alpha, n}$
$\sigma^2 \neq \sigma_0^2$	$2P(\chi^2 \geq \chi^2_{obs})$	$\chi^2_{obs} \geq \chi^2_{\frac{\alpha}{2}, n}$ or $\chi^2_{obs} \leq \chi^2_{1-\frac{\alpha}{2}, n}$

$$\chi^2_{obs} = \frac{n \tilde{S}^2}{\sigma^2}$$

Unknown μ

H_a	P-value	Critical region of size α
$\sigma^2 < \sigma_0^2$	$P(\chi^2 \leq \chi^2_{obs})$	$\chi^2_{obs} \leq \chi^2_{1-\alpha, n-1}$
$\sigma^2 > \sigma_0^2$	$P(\chi^2 \geq \chi^2_{obs})$	$\chi^2_{obs} \geq \chi^2_{\alpha, n-1}$
$\sigma^2 \neq \sigma_0^2$	$2P(\chi^2 \geq \chi^2_{obs})$	$\chi^2_{obs} \geq \chi^2_{\frac{\alpha}{2}, n-1}$ or $\chi^2_{obs} \leq \chi^2_{1-\frac{\alpha}{2}, n-1}$

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

21.6 Two Samples Tests

21.6.1 Pooled T-Test

Testing $H_0 : \mu_1 = \mu_2$ vs. $H_a :$

H_a	P-value	Critical region of size α
$\mu_1 - \mu_2 < 0$	$P(T \leq t_{obs})$	$t_{obs} \leq -t_{\alpha, n_1 + n_2 - 2}$
$\mu_1 - \mu_2 > 0$	$P(T \geq t_{obs})$	$t_{obs} \geq t_{\alpha, n_1 + n_2 - 2}$
$\mu_1 - \mu_2 \neq 0$	$2P(T \geq t_{obs})$	$ t_{obs} \geq t_{\frac{\alpha}{2}, n_1 + n_2 - 2}$

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

21.6.2 2 Samples T-Test (Satterthwaite-Welch)

Testing $H_0 : \mu_1 = \mu_2$ vs. $H_a :$

H_a	P-value	Critical region of size α
$\mu_1 - \mu_2 < 0$	$P(T \leq t_{obs})$	$t_{obs} \leq -t_{\alpha, k}$
$\mu_1 - \mu_2 > 0$	$P(T \geq t_{obs})$	$t_{obs} \geq t_{\alpha, k}$
$\mu_1 - \mu_2 \neq 0$	$2P(T \geq t_{obs})$	$ t_{obs} \geq t_{\frac{\alpha}{2}, k}$

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2)^2}{n_1 - 1} + \frac{(s_2^2)^2}{n_2 - 1}}$$

*NOTE: round down k to the nearest integer

21.6.3 Paired T-Test

- Take the paired difference
- Calculate sample mean and variance from the paired difference
- Ends up being a 1 sample T-test
- n = number of pairs

21.6.4 F-Test

Testing $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_a :$

H_a	P-value	Critical region of size α
$\sigma_1^2 < \sigma_2^2$	$P(F \leq F_{obs})$	$F_{obs} \leq F_{1-\alpha, n_2 - 1, n_1 - 1}$
$\sigma_1^2 > \sigma_2^2$	$P(F \geq F_{obs})$	$F_{obs} \geq F_{\alpha, n_2 - 1, n_1 - 1}$
$\sigma^2 \neq \sigma_0^2$	$2P(F \geq F_{obs})$	$F_{obs} \leq F_{1-\frac{\alpha}{2}, n_2 - 1, n_1 - 1}$ or $F_{obs} \geq F_{\frac{\alpha}{2}, n_2 - 1, n_1 - 1}$

$$F = \frac{\sigma_1^2}{\sigma_2^2} \frac{S_2^2}{S_1^2} \sim F_{n_2 - 1, n_1 - 1}$$

22 Chi-square Tests

22.1 The Goodness of Fit Tests (GOF)

A test for determining if observed values in a random sample follow a specified probability distribution

$$H_0 : P_i = P_{i,0} \text{ for } i = 1, 2, \dots, k$$

$$H_a : \text{some } P_i \neq P_{i,0}$$

Pearson's Statistic

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Condition

$$\text{All } E_i \geq 5$$

*NOTE: If this condition is not met, then we have to consolidate

Critical Region

$$\chi^2 \geq \chi^2_{\alpha, k-1}$$

22.2 Test of Homogeneity

Testing whether several multinomial populations follow the same multinomial proportions

22.2.1 CASE 1: Proportions are specified

$$H_0 : P_{1,j} = P_{2,j} = P_0 \text{ for } i = 1, 2, \dots, k$$

$$H_a : \text{some } P_{1,j} \neq P_{2,j}$$

Statistic

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{m(k-1)}$$

$$E_{ij} = n_i p_0$$

Condition

$$\text{All } E_{ij} \geq 5$$

Critical Region

$$\chi^2 \geq \chi^2_{\alpha, m(k-1)}$$

22.2.2 CASE 2: Proportions are not specified (but told they are equal)

$$H_0 : P_{11} = P_{21}, P_{12} = P_{22}, P_{13} = P_{23}$$

$$H_a : \text{some } P_{1,i} \neq P_{2,i}$$

Statistic

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \sim \chi^2_{(\textcolor{red}{m-1})(k-1)}$$

$$\hat{E}_{ij} = \frac{(\text{row sum } i)(\text{column sum } j)}{\text{total population}}$$

Condition

$$\text{All } \hat{E}_{ij} \geq 5$$

Critical Region

$$\chi^2 \geq \chi^2_{\alpha, (\textcolor{red}{m-1})(k-1)}$$

22.2.3 Partial information given about the category distribution

Reduce your degrees of freedom when estimating the unknown(s) by MLE

22.3 Test of independence

Testing whether two classifications are statistically independent in a single multinomial population where each member of the population is categorized by two criteria

- Mathematically identical to the test of homogeneity (case 2)

$$H_0 : \text{factors A and B are independent}$$

$$H_a : \text{factors A and B are not independent}$$

Statistic

$$\chi^2 = \sum_{i=1}^a \sum_{j=1}^b \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \sim \chi^2_{(\textcolor{red}{a-1})(b-1)}$$

$$\hat{E}_{ij} = \frac{\sum \text{row } i \sum \text{col } j}{N}$$

Condition

$$\text{All } \hat{E}_{ij} \geq 5$$

Critical Region

$$\chi^2 \geq \chi^2_{1-\alpha, (\textcolor{red}{a-1})(b-1)}$$

23 One-way ANOVA

- Extension of 2-sample t-test
- Test for a difference among the means of two (or more) normal population means
- NOTE that rejecting the null hypothesis tells us that there is a difference between treatment, BUT it does not tell us which treatment means are statistically significantly different

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

$$H_a : \text{some } \mu_i \neq \mu_j$$

Conditions

1. All $X_{i,j}$ are independent random variables
2. $X_{i,j} \sim \text{Normal}(\mu_i, \sigma^2)$
3. All $X_{i,j}$ have the same variance

General idea: Comparing the average variability within the population to the average variability between the population

For **m** samples of size **n**

MSTr (mean square for treatments)

$$\frac{SS_b}{m-1}$$

$$= \frac{n}{m-1} \sum_{i=1}^m (\bar{X}_{i\cdot} - \bar{X}_{..})^2$$

where

$$SS_b = n \sum_{i=1}^m (\bar{X}_{i\cdot} - \bar{X}_{..})^2$$

MSE (mean square for error)

$$\begin{aligned} & \frac{SS_W}{m(n-1)} \\ &= \frac{1}{m} \sum_{i=1}^m S_i^2 \end{aligned}$$

where

$$SS_W = (n-1) \sum_{i=1}^m S_i^2 = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$$

Test Statistic

$$F = \frac{MSTr}{MSE} \sim F_{(m-1), m(n-1)}$$

If $F > F_{\alpha, (m-1), m(n-1)}$ → reject H_0

NOTE:

$$SS_T = SS_b + SS_W$$

 $\bar{X}_{i\cdot}$ = sample mean of i $\bar{X}_{..}$ = grand mean