General:

- $E(b)$ means expected value of b
- **Mean:** $\mu = E(x)$ (mu)
  sample mean: $\bar{x}$
- **Variance:** σ^2 (sigma squared)
  $\sigma^2 = E((x-\mu)^2) = E(x^2) - E(x)^2$
  sample variance: $s^2 = \frac{\sum((x-\bar{x})^2)}{n-1}$
- **Standard Deviation:** σ (sigma)
  sample standard deviation: s
- **Random Process** can't be predicted
- **Complement** of A is A^c or A'
- $P(A \cap B)$ = probability of A **and** B
- $P(A \cup B)$ = probability of A **or** B
  $P(A \cup B) = (P(A \cap B))/P(B)$
- $P(A|B)$ = probability of A given B is true
- A, B independent if $P(A|B) = P(A)P(B)$
  therefore $P(A|B) = P(A)$
- **Reliability:** probability that it works
- **Discrete:** finite number of possible values
  Continuous: any value between a and b (e.g. any real number)
- Probability Mass Function (PMF) for discrete, Probability Density Function (PDF) for continuous
- Bernoulli random variable: has only 2 states: success or failure

5 Number Summary:

- 5 numbers:
  min
  Q_1 25%
  Q_2 50%
  Q_3 75%
  Max
- Q_x is a number (called a quartile) such that (25%, 50%, or 75%) of the data falls below that number
- Q_2 is also the median
- IQR: Inner Quartile Range = $Q_3 - Q_1$

Bayes Rule:

- When you've got a grid of the possible outcomes of two different events
- These edges are called marginals; they sum to 1
- if the two events are independent, each cell is the product of the corresponding marginals
  $P(A \cup B) = P(A)P(B)$
  example: fair dice roll is independent
- Two events in a grid are only independent if the property holds for every cell in the grid
- **Conditional Probability:** $P(A|B)$ = probability of A given that B is true $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- if A and B are independent, then $P(A|B) = P(A)$ because B doesn't affect A

Mutual Exclusion:

- A and B are mutually exclusive if $P(A \cap B) = 0$
- can be one or the other, or none, but can't be both
- mutually exclusive events can't be independent, because once you know one is true, you know the other is false

Combinations Formula:

- $\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{(n)(n-1)...(n-x+1)}{(x)(x-1)...(1)}$

Binomial Distribution:

- Discrete. Driven by p and n
  p: probability of success
  n: number of trials
- $\mu = np$
  $\sigma^2 = np(1-p)$
- PMF: $\binom{n}{x} p^x (1-p)^{n-x}$

Continuous Distributions:

- PDF: Probability Density Function
  $f(x) = \frac{d}{dx}(F(x))$
  $\int_{-\infty}^{\infty} f(x)dx = 1$
  $f(x) \geq 0$ for all x
- CDF: Cumulative Density Function
  $F(x) = \int_{-\infty}^{x} f(t)dt$
  $F(x) = P(X \leq x)$
- $P(a \leq X \leq b) = \int_a^b f(x)dx$
  $P(a \leq X \leq b) = F(b) - F(a)$
  $P(X = a) = 0$
- $\mu = E(x) = \int_a^b x f(x)dx$ ((a, b) = domain of $f(x)$)
  $\sigma^2 = E((X-\mu)^2) = E(x^2) - E(x)^2$
  $E(h(x)) = \int_a^b h(x)f(x)dx$ expected value of $h(x)$

Uniform Continuous Distribution:

- All outcomes have the same probability
- $\mu = \frac{B+A}{2}$
- $E(X^2) = (B^2 + AB + A^2)/3$
- $\sigma^2 = (B-A)^2/12$
- PDF: $f(x) = \frac{1}{B-A}$ for $A \leq x \leq B$
- CDF:
  $F(x) = 0$ for $x < A$
  $F(x) = \int_A^x \frac{1}{B-A}dx = \frac{x-A}{B-A}$ for $A \leq x \leq B$
  $F(x) = 1$ for $x > B$

Poisson distribution:

- Discrete. Driven by λ (lambda)
  Interval size is fixed, number of occurrences is varied
  λ: number of events occurring per interval
- $\mu = \sigma^2 = \lambda$
  $\sigma = \sqrt{\lambda}$
- PDF: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Gamma Distribution:

- Continuous equivalent of Poisson
- Driven by α and β
  Number of occurrences is fixed, interval length is varied
  α: number of events we're interested in
  β: rate at which events happen: "β time until the next event"
- $\lambda = 1/\beta$: shape parameter
  $\mu = \alpha\beta$
  $\sigma^2 = \alpha(\beta^2)$

Normal Distribution:

- Continuous. Defined in terms of μ and σ
- Empirical Rule: (probably don't use)
  $P(x = \mu \pm 1\sigma) : 68\%$
  $P(x = \mu \pm 2\sigma) : 95\%$
  $P(x = \mu \pm 3\sigma) : 99\%$
- **Important Z Values:**

Z	area to the right
1.6450	5.0%
1.9600	2.5%
2.3260	1.0%
2.5758	0.5%

Standard Normal Transformation:

- Z is like x in terms of μ and σ.
- For use with lookup tables
- $Z = \frac{x-\mu}{\sigma}$
- $\Phi(Z) =$NormCDF$(\sigma = 1, \mu = 0, x = Z)$
- Prof says this method is prone to error

Joint Distributions:

- $f(x, y)$
- **Independent** if you can split up $f(x, y) = g(x)h(y)$
- Mean and Variance are additive
- Standard Deviation is not additive

Distribution of Sample Totals:

- original distribution: μ_0, σ_0
- sample size: n
- mean: $\mu = n\mu_0$
- variance: $\sigma^2 = n\sigma_0^2$
- (not as easy for standard deviatoin (σ))

Distribution of Sample Mean:

- original distribution: μ_0, σ_0
- mean: $\mu = \mu_0$
- variance: $\sigma^2 = \frac{\sigma_0^2}{n}$

Confidence Interval - Mean:

- Only works for $n \geq 30$
- Mean: $\bar{x} \pm Z_{\frac{\alpha}{2}} \left(\frac{s}{\sqrt{n}}\right)$
  - $s =$ sample standard deviation
  - $\frac{s}{\sqrt{n}} =$ standard error=standard deviation of distribution of sample means
  - Only works for $n \geq 30$
- sample size in terms of confidence and interval width: $n = \left(2Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{L}\right)\right)^2$
  - $L =$ width of interval
    $\sigma =$ standard deviation of **population**

Confidence Interval - Mean - $n < 30$:

- Only works when parent distribution is normal
- use the t-distribution
- $n - 1 =$ degrees of freedom
- mean: $\bar{x} \pm \left(t_{\alpha/2,n-1}\right)\left(\frac{s}{\sqrt{n}}\right)$

Confidence Interval - Variance:

- You can only do this for $n < 30$ if the data is from a normal distribution
- distribution is a (right-skewed) χ^2 (chi-squared) distribution
- $n - 1 =$ degrees of freedom
- confidence interval for $\sigma^2 : \left(\frac{s^2(n-1)}{\chi^2_{\alpha/2,n-1}}, \frac{s^2(n-1)}{\chi^2_{1-(\alpha/2),n-1}}\right)$
  - $\chi^2_{x,y}$ is the chi-squared distribution with $P(X) \leq x$ and y degrees of freedom

Confidence Interval - Proportion:

- proportion of successes in sample of trials
- based on Bernoulli trials
  - each x is either 0 (fail) or 1 (success)
- $\hat{p} =$ sample proportion
  $p =$ population proportion
  $\hat{q}, q$ same, but for failure
  $q = 1 - p$, $\hat{q} = 1 - \hat{p}$
- you can only assume it's normal when $\hat{p} > 5$ and $1 - \hat{p} > 5$
- $\hat{p} = \frac{\sum x}{n}$
- Central Limit Theorem only applies for $\hat{p}n \geq 5$ and $\hat{q}n \geq 5$ and n much smaller than the population
- Confidence Interval: $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$
- Sample size: $n = 4Z^2_{\alpha/2}\hat{p}\hat{q}\frac{1}{L^2}$
  - If you don't have any idea about $\hat{p}$, use 0.5 to be maximally conservative

Prediction Interval:

- predict the next value that will occur
- $\bar{x} \pm Z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}$

- use s if you don't have the population standard deviation
- if $n < 30$ you can only use this if the parent population is normal
- the standard error is different (wider) than a confidence interval because we're comparing two different distributions:
  - variance of x: σ^2
  - variance of the sampling distribution of means: $\frac{\sigma^2}{n}$
- this follows from the fact that we had to use our sample mean to predict the next value, not the actual mean

Hypothesis Testing:

- Objective is always to reject the null hypothesis

- null hypothesis: H_0
  - usually $H_0 : \mu = \mu_0$
- alternative hypothesis: H_A

  1 tail right: $H_A : \mu > \mu_0$
  1 tail left: $H_A : \mu < \mu_0$
  2 tail: $H_A : \mu \neq \mu_0$
- `test statistic: $Z_t = \frac{\bar{x} - \mu}{s_\bar{x}}$ where $s_\bar{x} = \frac{s}{\sqrt{`

- critical value: Z_C (determined by α level)
  - for a 1 tailed test, it's Z_α
  - for a 2 tailed test, it's $Z_{\alpha/2}$
- p-value: area more extreme than text statistic
  - remember it's two-sided if it's a 2-tail test

Proportion Hypothesis:

- $Z_t = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$
- `equivalent of $S_\bar{x}$ is $\sqrt{p_0(1-p_0)/n}$`

Type 1 Error:

- false positive
- rejecting H_0 when it is true
- probability is equal to our alpha level (α)

Type 2 Error:

- false negative
- probability that we fail to reject when H_0 is false (we should have rejected)
- represents the probability of drawing a sample that just happens to support H_0 (be inside our critical range)
- depends on the actual **population** values μ and σ
- $P(\text{type 2 error}) = \frac{\bar{x}_c - \mu}{\sigma_{\bar{x}}}$
  - $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ standard error of sample from population
  - μ =real mean of population (not the one from H_0)
  - $\bar{x}_c$ critical value for sample mean (determined by H_0 and α)
- Example: if the real mean is exactly $\bar{x}_c$, then $P(T2) = 0.5$
- Technically, the minimum and maximum are 0 and 1 respectively, since it depends on the population values, and that could be anything.

Final stuff:

You calculate $Var(\bar{X} - \bar{y})$ differently depending on if the variances of the two distributions are equal or not

Difference of Means:

- actual variance of difference: $\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}$
  - does not matter if the two variances are (roughly) equal or not
  - don't usually get to use this because we don't have the population values
  - therefore must use s^2 instead of σ^2
  - (this is not the same thing as pooled variance)

Two Distributions, equal variance:

- Δ_0 is the expected difference between the means

- H_0 is usually that $\bar{X} - \bar{Y} = \Delta_0$
- variance considered equal if $\frac{1}{3} \leq \frac{s_X^2}{s_Y^2} \leq 3$
- pooled variance $s_p^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{n+m-2}$
  - weighted average of the sample variances
- test stat: $T = \frac{\bar{X} - \bar{Y} - \Delta_0}{s_p\sqrt{1/m + 1/n}}$

Two Distributions, unequal variance:

- always use t distribution
- test stat: $T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{s_X^2/m + s_Y^2/n}}$
- $DF = m + n - 2$

Paired Difference Test:

- `$D_i = X_i - Y_i$ mean difference is \bar{D}`

- test stat: $T = \frac{\bar{D} - \Delta_0}{s_D/\sqrt{n}}$
  - n is the number of paired differences, not the number of total observations

Proportion comparison:

- uses Z distribution
- common $\hat{p}$ is total successes divided by total observations
- test stat: $Z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}(1-\hat{p})(1/m + 1/n)}}$
  - $\hat{p}$ is success proportion of the overall study

- C.I: $(\hat{p}_X - \hat{p}_Y) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{m} + \frac{\hat{p}_y(1-\hat{p}_y)}{n}}$
- We use the common $\hat{p}$ for hypothesis tests but not C.I. because hypothesis tests assume that $\hat{p}_x = \hat{p}_y$

ANOVA:

- ANalysis Of VAriance
- $H_0 : \mu_1 = \mu_2 = \ldots = \mu_i$
  H_A: at least two means (μ's) are different
- trt = treatment
- err = error
- tot = total
- Assume normal population distribution with equal variance
- $X_{i,j}$ = jth sample from the ith treatment group

  - I = number of treatment groups, n_i = number of samples in treatment group i
- $\bar{X}_{GM}$ or just $\bar{X}$ = Grand Mean
- SS = Sum of Squares
  - $SS_{tot} = SS_{trt} + SS_{err}$
  - $SS_{tot} = \sum_{i=1}^{I} \sum_{j=0}^{n_i} \left(X_{i,j} - \bar{X} \right)^2$
  - $SS_{err} = \sum_{i=1}^{I} \sum_{j=0}^{n_i} \left(X_{i,j} - \bar{X}_i \right)^2$
  - $SS_{trt} = \sum_{i=1}^{I} \left(\bar{X}_i - \bar{X} \right)^2$
  -
- MS = Means Squared
  - $MS_{thing} = \frac{SS_{thing}}{DF_{thing}}$
- test statistic: $F_{test} = \frac{MS_{trt}}{MS_{err}}$
  - Describes how much error is due to treatment as opposed to (normally distributed) random errors
- critical value: $F_{\alpha, DF_{trt}, DF_{err}}$
- You can't use pair-wise ANOVA tests to determine which one of the distributions is actually different, because the uncertainty (introduced by the α level) propagates.
  - instead you have to use Tuke's method (not this class)

Regression:

- we assume different measurements are independent of each other
- model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_0$
  - β_0: intercept
  - β_1: slope
  - ϵ_i: (random) error term
    - assumed to have a normal distribution with $\mu = 0$
  - X_i: independent variable; predictor
  - Y_i: dependent variable; response
- $Q = \sum_{i=1}^{I} (Y_i - \beta_0 - X_i \beta_1)$
  - sum of squared vertical deviations
  - sum-of-squares minimizes this
- $S_{xx} = \sum (x_i - \bar{x})^2$
- $S_{xy} = \sum (y_i - \bar{y})(x_i - \bar{x})$
- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ ratio of joint variability of X and Y to variability of just X
  - hat ($\hat{}$) means estimate
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_0 \bar{x}$
- $\epsilon_i = y_i - \hat{y}_i$, error term, residual
  - actual - estimated
- total sum of squares $SST = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{1}{n} (\sum y^i)^2$
- error sum of squares$SSE = \sum (e_i^2) = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$
- regression sum of squares $SSR = \sum (\hat{y}_i - \bar{y})^2 = SST - SSR$
- **measure of fit:** $r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$
  - r^2 = proportion of variation in y explained by the linear relationship model with x
  - $0 \leq r^2 \leq 1$
  - $r^2 = 1$: all data perfectly on straight line
  - r^2 near zero: no linear relationship (may be other type of relationship)
- sample correlation: r
- you can do an ANOVA test for regression
  - $DF_{total} = n - 2$(because two estimates β_0, β_1)
  - $DF_{regression} = 1$
- `hypothesis test: $T = \frac{\hat{\beta}_1 - \`

  - `$s_\hat{\beta}_1 = \frac{s_\epsilon}{\sqrt`

  - H_0: $\beta_1 = \Delta_0$
  - s_ϵ = estimate of standard deviation of error term
  - Δ_0: expected slope (usually 0)
  - `confidence interval: $\hat{\beta}_1 \pm t_`

- confidence interval for mean response
  - mean y value at given value x^*
  - means that you're $(1-\alpha)100\%$ sure that the mean response will be inside this interval
  - variance $= \sigma_\epsilon^2 \left(\frac{1}{n} + \frac{(x^* - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right)$
  - C.I. $\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2, n-2} s_\epsilon \sqrt{\frac{1}{n} + \frac{(x^* - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$
- prediction interval
  - means you're $(1-\alpha)100\%$ sure that a new value will lie inside this interval
  - C.I. $\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2, n-2} s_\epsilon \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$

Law of Total Probability:

- If a probability is made up of sub-probabilities, the total probability is equal to the weighted average
- Example:
  - Lightbulb makers X and Y, probabilities of failure are $\Pr B_X$ and $\Pr B_Y$ respectively

- our lightbulb population is $\frac{6}{10}X$ and $\frac{4}{10}Y$
- A is total probability that any bulb will fail

- $\Pr(A) = \Pr(A|B_X)\Pr(B_X) + \Pr(A|B_Y)\Pr(B_Y)$
  $= \frac{99}{100} \cdot \frac{6}{10} + \frac{95}{100} \cdot \frac{4}{10} = \frac{594+380}{1000} = \frac{974}{1000}$

Long Range Frequency:

- as sample size increases, the relative frequencies of outcomes will approach their theoretical values.
- Example: A fair dice will give each number approximately $\frac{1}{2}$ of the time