## Basic Probability
### Definitions
**Sample Space:** Set of all possible outcomes
**Event:** Subset of sample space
**Probability:** Numerical measure of chance
### Set Operations
$A \cap B$ = Both A and B occur
$A \cup B$ = A or B (or both) occur
$A^c$ = A does not occur
### Probability Rules
**Intersection:** $\Pr(A \cap B) = \Pr(A \mid B)\Pr(B)$
**Union:** $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
**Complement:** $\Pr(A^c) = 1 - \Pr(A)$
**Conditional:** $\Pr(A \mid B) = \frac{\Pr(A\cap B)}{\Pr(B)}$
### Independence
$A$ and $B$ independent $\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$
If independent: $\Pr(A \mid B) = \Pr(A)$
## Discrete Random Variables
### PMF & CDF
**PMF:** $f(x) = \Pr(X = x)$
**CDF:** $F(a) = \Pr(X \le a) = \sum_{x\le a} f(x)$
**Expected Value:** $E[X] = \sum_{x\in S} x f(x)$
**Variance:** $\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
### Linear Transformations
$E[aX + b] = aE[X] + b$
$\mathrm{Var}[aX + b] = a^2 \,\mathrm{Var}[X]$
### Conditional Expectation
$E[X \mid Y = y] = \sum_x x \Pr(X = x \mid Y = y)$
**Law of Iterated Expectations:** $E[E[X \mid Y]] = E[X]$
## Moment-Generating Functions
### Definition
$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$ (discrete)
$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$ (continuous)
### Finding Moments
$r$-th moment: $\mu_r = E[X^r] = M_X^{(r)}(0)$
### Common MGFs (quick)
Bernoulli($p$): $M(t) = (1 - p) + pe^t$
Binomial($n, p$): $M(t) = (1 - p + pe^t)^n$
Poisson($\lambda$): $M(t) = \exp(\lambda(e^t - 1))$
Exponential($\lambda$): $M(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$
Gamma($\alpha, \theta$): $M(t) = (1 - \theta t)^{-\alpha}$ for $t < 1/\theta$
### Standardized Moments
$\hat{\mu}_k = \frac{E[(X-\mu)^k]}{E[(X-\mu)^2]^{k/2}}$
**Skewness:** $\hat{\mu}_3$
**Kurtosis:** $\hat{\mu}_4$
## Discrete Distributions
### Bernoulli
$X \sim \mathrm{Bernoulli}(p)$, $P(X = 1) = p$, $P(X = 0) = 1 - p$
$E[X] = p$, $\mathrm{Var}[X] = p(1 - p)$
### Binomial
$X \sim \mathrm{Binomial}(n, p)$
$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
$E[X] = np$, $\mathrm{Var}[X] = np(1 - p)$
### Geometric
$X \sim \mathrm{Geometric}(p)$ counts trials to first success
$P(X = x) = p(1 - p)^{x-1}$, $x = 1, 2, \ldots$
$E[X] = \frac{1}{p}$, $\mathrm{Var}[X] = \frac{1-p}{p^2}$
### Negative Binomial
$X \sim \mathrm{NegBin}(r, p)$ counts trials to $r$-th success
$P(X = x) = \binom{x-1}{r-1} p^r (1 - p)^{x-r}$
$E[X] = \frac{r}{p}$, $\mathrm{Var}[X] = \frac{r(1-p)}{p^2}$
### Poisson
$X \sim \mathrm{Poisson}(\lambda)$
$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
$E[X] = \lambda$, $\mathrm{Var}[X] = \lambda$

## Continuous Random Variables
### PDF & CDF
**PDF:** $f(x) \ge 0$, $\int_S f(x)dx = 1$
**CDF:** $F(a) = \Pr(X \le a) = \int_{-\infty}^{a} f(x)dx$
**Expected Value:** $E[X] = \int_{-\infty}^{\infty} x f(x)dx$
**Variance:** $\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx$
### Quantiles
$p$-th quantile $\pi_p$ satisfies $p = \int_{-\infty}^{\pi_p} f(x)dx$
## Continuous Distributions
### Uniform $[a, b]$
$f(x) = \frac{1}{b-a}$ for $a < x < b$, else 0
$F(x) = 0$ if $x < a$, $\frac{x-a}{b-a}$ if $a \le x < b$, 1 if $x \ge b$
$E[X] = \frac{a+b}{2}$, $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$
### Normal
$Z \sim N(0, 1)$ has $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
$X \sim N(\mu, \sigma^2)$ has $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Standardize: $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$
Empirical rule: 68/95/99.7% within $\mu \pm 1/2/3\sigma$
### Exponential
$W \sim \mathrm{Exp}(\lambda)$ with $f(w) = \lambda e^{-\lambda w}$ for $w \ge 0$
$F(w) = 1 - e^{-\lambda w}$, $E[W] = 1/\lambda$, $\mathrm{Var}[W] = 1/\lambda^2$
Memoryless: $\Pr(W > s + t \mid W > s) = \Pr(W > t)$
### Gamma
$X \sim \mathrm{Gamma}(\alpha, \theta)$
$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$ for $x \ge 0$
$E[X] = \alpha\theta$, $\mathrm{Var}[X] = \alpha\theta^2$
$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$, $\Gamma(t) = (t - 1)\Gamma(t - 1)$
## Special Cases & Relationships
**Exponential:** $\mathrm{Gamma}(1, \theta) = \mathrm{Exp}(\lambda)$, $\theta = 1/\lambda$
**Geometric:** $\mathrm{NegBin}(1, p) = \mathrm{Geom}(p)$
**Binomial $\to$ Poisson:** $n \to \infty$, $p \to 0$, $np \to \lambda$
## Useful Formulas
**Variance Identity:** $\mathrm{Var}[X] = E[X^2] - (E[X])^2$
**Combinatorics:** $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
**DeMorgan's Laws:** $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$
## Bivariate Distributions
### Expectations & Covariance
$E[X + Y] = E[X] + E[Y]$
$\mathrm{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$
Interpretation: $> 0$ positive association, $< 0$ negative, $= 0$ no linear link
$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$
$\mathrm{Cov}(X, Y + Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$
Independence $\Rightarrow \mathrm{Cov}(X, Y) = 0$, converse need not hold
### Variance of a Sum
$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$
If $X$ and $Y$ independent: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
### Correlation Coefficient
$\rho = \frac{\mathrm{Cov}(X,Y)}{\mathrm{SD}(X)\,\mathrm{SD}(Y)}$
Scale-free measure, $\rho \in [-1, 1]$, $\rho = 0$ means no linear association
### Conditional Distributions
- Discrete: $g(x \mid y) = \frac{f(x,y)}{f_Y(y)}$ for $f_Y(y) > 0$
- Independence $\Rightarrow g(x \mid y) = f_X(x)$
- Continuous joint pdf $f(x, y)$: $f(x, y) \ge 0$, $\iint f(x, y)dxdy = 1$
- Continuous marginals: $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$
- $X \perp Y$ iff $f(x, y) = f_X(x)f_Y(y)$
### Bivariate Normal
**PDF:** $f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{q(x,y)}{2}\right)$
$q(x, y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$

Marginals are normal with means $\mu_X, \mu_Y$ and variances $\sigma_X^2, \sigma_Y^2$
$X+Y$ is normal when $(X,Y)$ jointly normal
Conditional $Y|X=x$ normal with $E[Y|x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$
$\text{Var}(Y|x) = \sigma_Y^2(1-\rho^2)$
$\sigma_{XY} = \rho\sigma_X\sigma_Y$ so $E[Y|x] = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_X)$

## Functions of Random Variables

### Distribution Function Technique

If $Y = u(X)$, $G(y) = \Pr(Y \le y) = \Pr(u(X) \le y)$, density $g(y) = G'(y)$
Example: $X \sim N(0,1)$, $Y = e^X$ (log-normal) gives $G(y) = \Phi(\ln y)$, $g(y) = \phi(\ln y)\frac{1}{y}$ for $y > 0$

### Change-of-Variable (shortcut)

If $Y = u(X)$ is one-to-one on its support and $X = v(Y)$ is the inverse, then

$$g_Y(y) = f_X(v(y)) \cdot |v'(y)| \quad (y \in S_Y).$$

Equivalent form (when $u$ differentiable, monotone): $g_Y(y) = \frac{f_X(x)}{|u'(x)|}$ evaluated at $x = v(y)$.

### Change-of-Variable (1 Variable)

If $Y = u(X)$, define $v()$ such that $X = v(Y)$.
Suppose $c_1 < x < c_2$ and $d_1 = u(c_1) < y < d_2 = u(c_2)$ ($v()$ is increasing),

$$G(y) = \int_{c_1}^{v(y)} f(x)dx, \qquad d_1 < y < d_2$$

$$G'(y) = g(y) = f(v(y)) \cdot v'(y), \qquad d_1 < y < d_2$$

**Log-normal example:**
$u(x) = \exp(x)$, $v(y) = \ln(y)$, $v'(y) = 1/y$

$$g(y) = \phi(\ln y) \cdot \frac{1}{y} \qquad \text{when } y > 0$$

Suppose $c_1 < x < c_2$ and $d_1 = u(c_1) > y > d_2 = u(c_2)$ ($v()$ is decreasing),

$$G(y) = \int_{v(y)}^{c_2} f(x)dx, \qquad d_2 < y < d_1$$

$$G'(y) = g(y) = f(v(y)) \cdot (-v'(y)), \qquad d_2 < y < d_1$$

Note: The negative sign appears because $v'(y) < 0$ when $v()$ is decreasing.
In general, we have:

$$g(y) = f(v(y)) \cdot |v'(y)|, \qquad y \in S_Y$$

### Non-monotone / many-to-one case

If $y = u(x)$ has multiple inverse branches $\{x_j(y)\}$ over the support, then

$$g_Y(y) = \sum_j f_X(x_j(y)) \cdot \left|\frac{d}{dy}x_j(y)\right|.$$

Example: $Y = X^2$ with $X$ continuous,

$$g_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, \quad y > 0.$$

### Quick examples

Linear: if $Y = aX + b$ ($a \neq 0$), then $g_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.
Min/max CDF trick: if $M = \max(X_1, \ldots, X_n)$ iid with CDF $F$, then $F_M(m) = F(m)^n$; if $m = \min$, $F_m(m) = 1 - (1-F(m))^n$.

### Cauchy Distribution

$W \sim U(-\pi/2, \pi/2)$, $X = \tan W$ gives standard Cauchy
$f_X(x) = \frac{1}{\pi(1+x^2)}$, heavy tails, symmetric about 0
Median = mode = 0; mean and variance do not exist
Equivalent views: ratio of two independent standard normals; $t$ with 1 degree of freedom
Graph of pdf known as the witch of Agnesi

### Multivariate Transformations

For $(X_1, X_2)$ with joint pdf $f$, let $Y_1 = u_1(X_1, X_2)$, $Y_2 = u_2(X_1, X_2)$ invertible to $X_i = v_i(Y_1, Y_2)$

Joint pdf $g(y_1, y_2) = \begin{vmatrix} \frac{\partial v_1}{\partial y_1} & \frac{\partial v_1}{\partial y_2} \\ \frac{\partial v_2}{\partial y_1} & \frac{\partial v_2}{\partial y_2} \end{vmatrix} f(v_1(y_1,y_2), v_2(y_1,y_2))$

Jacobian determinant captures area/volume distortion from the transformation

### Common Jacobians

Sum/difference: $U = X+Y$, $V = X-Y \Rightarrow x = \frac{u+v}{2}$, $y = \frac{u-v}{2}$, $|J| = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \frac{1}{2}$.
Polar (from $(X,Y)$ to $(R,\Theta)$): $x = r\cos\theta$, $y = r\sin\theta$, $|J| = r$.

## Limit Theorems & Inequalities

### Sample Mean

$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ for iid $X_i$ with mean $\mu$, variance $\sigma^2$
$E[\overline{X}] = \mu$, $\text{Var}[\overline{X}] = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

### Law of Large Numbers

Sample mean converges in probability to $\mu$ as $n$ grows

### Central Limit Theorem

Standardized mean $W = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$ for large $n$
Justifies normal approximations even when data not normal, provided finite mean/variance

### Inequalities

**Markov:** For nonnegative $X$, $\Pr(X \ge a) \le \frac{E[X]}{a}$
**Chebyshev:** $\Pr(|X - E[X]| \ge r) \le \frac{\text{Var}[X]}{r^2}$

### MGF Method Insight

If mgfs $M_n(t)$ converge to $M(t)$ near 0, then $X_n \Rightarrow X$ with mgf $M(t)$; used to establish the CLT limit

## Quick Approximations (Pre Lec 14)

### Normal approximation

If $X \sim Binomial(n,p)$ with large $n$ and $np, n(1-p)$ not too small:

$$\frac{X - np}{\sqrt{np(1-p)}} \approx N(0,1) \qquad X \approx N(np,\ np(1-p))$$

### Poisson approximation

If $X \sim Binomial(n,p)$ with $n$ large and $p$ small, $\lambda = np$:

$$X \approx Poisson(\lambda) \quad \Rightarrow \quad \Pr(X = k) \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

### Sums of independent r.v.'s

For independent $X_i$: $E[\sum X_i] = \sum E[X_i]$, $Var(\sum X_i) = \sum Var(X_i)$ (no covariances).

## Key Concepts

**Monty Hall Problem:** Switching doors wins with probability $\frac{2}{3}$
**Bayes' Theorem:** $\Pr(A \mid B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$
**Total Probability:** $\Pr(A) = \sum_i \Pr(A \mid B_i)\Pr(B_i)$ for partition $\{B_i\}$

## Estimation (Lec 14)

### Population vs. Sample (Sample analog principle)

Population (unknown): $Y$, distribution, $\mu = E[Y]$, $\sigma^2 = Var(Y)$, $Cov(X,Y)$, $Corr(X,Y)$
Sample (data): $\{y_1, \ldots, y_n\}$, histogram, $\overline{y}$, $s^2$, $s_{xy}$, $r$

### Core estimators

$$\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \overline{y})^2$$

$$s^2 = \frac{\sum_{i=1}^n y_i^2 - n\overline{y}^2}{n-1}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) \qquad r = \frac{s_{xy}}{s_x s_y}$$

$$s_{xy} = \frac{\sum_{i=1}^n x_i y_i - n\overline{x}\,\overline{y}}{n-1}$$

**SE (standard error):** SD of an estimator, estimated by plugging in sample estimates.

$$SE(\overline{X}) = \frac{s}{\sqrt{n}}$$

$$\text{(true) } SD(\overline{X}) = \frac{\sigma}{\sqrt{n}}, \quad Var(\overline{X}) = \frac{\sigma^2}{n}.$$

95% "rule of thumb": margin of error $\approx 2 \cdot SE$.

**Portfolio / linear combo**

$$E[aX + bY] = aE[X] + bE[Y]$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X, Y)$$

$$Var(wX + (1-w)Y) = w^2 Var(X) + (1-w)^2 Var(Y) + 2w(1-w)Cov(X, Y)$$

**Interpretation:** diversification can reduce variance even if $Corr(X, Y) > 0$ (unless perfect correlation).

**Variance-minimizing mix (2 assets)**

For $P = wX + (1-w)Y$ with $Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$, $Cov(X, Y) = \sigma_{XY}$:

$$Var(P) = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_{XY}.$$

Minimum-variance weight on $X$:

$$w^* = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}} = \frac{\sigma_Y^2 - \rho\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y}.$$

**Sampling + bias (what can go wrong?)**

**Uncertainty:** sample $\neq$ population.

**Bias:** sample not representative (nonrandom selection, nonresponse, measurement error).

**Sampling designs:** simple random sample; stratified sample; cluster sample.

**Unbiasedness vs. consistency**

Unbiased: $E[\hat{\theta}] = \theta$. Consistent: $\hat{\theta} \xrightarrow{p} \theta$.

Examples for $E[X_i] = \mu$:

- $\overline{X}$ is unbiased and consistent.
- $X_1$ is unbiased but not consistent.
- $\frac{1}{n+1}\sum_{i=1}^n X_i = \frac{n}{n+1}\overline{X}$ is biased but consistent.

**Why $n-1$ in $s^2$ and $s_{xy}$?**

Using $\overline{y}$ costs one degree of freedom (Bessel correction); $s^2$ is unbiased for $\sigma^2$ under iid sampling.

**Consistency & Convergence (Lec 15)**

**Convergence in probability**

$S_n \xrightarrow{p} \mu$ iff $\lim_{n\to\infty} \Pr(|S_n - \mu| \geq \epsilon) = 0$ for all $\epsilon > 0$.
Notation: plim $S_n = \mu$; "$S_n$ consistent for $\mu$".

**LLN + Chebyshev**

If $Y_1, \ldots, Y_n$ iid, $E[Y_i] = \mu$, $Var(Y_i) = \sigma^2 < \infty$, then $\overline{Y} \xrightarrow{p} \mu$.
Chebyshev: $\Pr(|X - E[X]| \geq r) \leq \frac{Var(X)}{r^2}$.

$$Var(\overline{Y}) = \frac{\sigma^2}{n} \Rightarrow \Pr(|\overline{Y} - \mu| \geq r) \leq \frac{\sigma^2/n}{r^2} \to 0$$

**Generalized Chebyshev + MSE trick**

Generalized Chebyshev: for any $W$, $\Pr(|W| \geq r) \leq \frac{E[W^2]}{r^2}$.
To prove $\tilde{Y} \xrightarrow{p} \mu$: set $W = \tilde{Y} - \mu$ and show $E[W^2] \to 0$.
MSE: $E[(\tilde{Y} - \mu)^2] = Bias(\tilde{Y})^2 + Var(\tilde{Y})$.
Sufficient: $MSE \to 0 \Rightarrow$ consistent.
For a sequence $\theta_n$:

$$\lim_{n\to\infty} E[\theta_n] = \mu, \ \lim_{n\to\infty} Var(\theta_n) = 0 \Rightarrow \theta_n \xrightarrow{p} \mu$$

**SE for a sample proportion**

If $X_i \in \{0, 1\}$, $p = E[X_i]$, $\hat{p} = \overline{X}$:

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

$$\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right) \quad \text{(CLT)}.$$

**Maximum Likelihood (Lec 16)**

**Likelihood / log-likelihood**

Given iid sample $x_1, \ldots, x_n$ with pdf/pmf $f(x_i; \theta)$:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta), \qquad \mathcal{L}(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta)$$

MLE: $\hat{\theta} = \arg\max_\theta L(\theta)$ (equivalently maximize $\mathcal{L}$).

**Normal MLE ($X_i \sim N(\mu, \sigma^2)$)**

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2$$

Note: $\hat{\sigma}^2_{MLE}$ biased (uses $n$), but consistent.

**Bernoulli MLE ($X_i \sim Bern(p)$)**

$$L(p) = \prod p^{x_i}(1-p)^{1-x_i} \Rightarrow \hat{p}_{MLE} = \overline{X}$$

**Asymptotics (high level):** MLEs are consistent, asymptotically normal, and asymptotically efficient (under correct specification).

**Confidence Intervals (Lec 16)**

**Large sample CI for a mean**

By CLT, $\overline{X} \approx N(\mu, \sigma^2/n)$ and $Var(\overline{X}) = \sigma^2/n \approx s^2/n$.

$$CI_{1-\alpha}: \ \overline{X} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Common: 95% uses $z_{0.025} \approx 1.96$.

**One-sided confidence bounds (Practice 1)**

Upper $(1-\alpha)$ bound: $\mu \leq \overline{X} + z_{1-\alpha} \cdot SE(\overline{X})$.
Lower $(1-\alpha)$ bound: $\mu \geq \overline{X} - z_{1-\alpha} \cdot SE(\overline{X})$.
For a proportion: replace $SE(\overline{X})$ by $\sqrt{\hat{p}(1-\hat{p})/n}$ (CI) or $\sqrt{p_0(1-p_0)/n}$ (test).

**Small sample CI for a mean (Normal data)**

If $X_i \sim N(\mu, \sigma^2)$ and $\sigma$ unknown, for any $n > 1$:

$$\frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1} \Rightarrow \overline{X} \pm t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}$$

**CI for a difference in means (independent samples)**

Large $n$:

$$(\overline{X}_1 - \overline{X}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Small $n$, equal variances (Normal):

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2, n_1+n_2-2} \cdot s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

**CI for a proportion (large $n$)**

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

**Interpretation**

Method yields coverage in repeated samples: $\Pr(\mu \in CI) = 1 - \alpha$ (not: "$\Pr(\mu \in CI \mid data)$").

**Hypothesis Testing Basics (Lec 17)**

**General structure**

**Null** $H_0$ vs **alternative** $H_a$; choose significance level $\alpha$.
Test statistic $T$; p-value = $\Pr(\text{as/extreme as } T \mid H_0)$.
Decision: reject if p-value $< \alpha$.

**Type I / II and power**

Type I error: reject true $H_0$ (probability $\alpha$).
Type II error: fail to reject false $H_0$ (probability $\beta$).
Power $= 1 - \beta$ increases with larger $n$, larger effect size, larger $\alpha$.

**One-sample mean test (large $n$)**

$$H_0: \mu = \mu_0, \qquad z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \approx N(0, 1) \ (H_0)$$

Two-sided at 5%: reject if $|z| > 1.96$.
p-value (two-sided): $2(1 - \Phi(|z|))$.

**One-sided vs. two-sided**

Right-tail: $H_a: \mu > \mu_0$, reject for large $z$; p-value $= 1 - \Phi(z)$.
Left-tail: $H_a: \mu < \mu_0$, reject for small $z$; p-value $= \Phi(z)$.
Two-sided is the conservative default unless direction is justified *a priori*.

**CI $\leftrightarrow$ test link**

For many large-sample mean/proportion settings: reject $H_0$ at level $\alpha$ iff the $(1-\alpha)$ CI excludes the null value.

**Power + sample size (Practice 2: Catalogs)**

One-sided test $H_0 : \mu = \mu_0$ vs $H_a : \mu > \mu_0$, known $\sigma$: reject if $\overline{X} > \mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$.

If true mean is $\mu_1 = \mu_0 + \delta$, then power is

$$1 - \beta = \Pr(\text{reject} \mid \mu = \mu_1) = 1 - \Phi\left(z_{1-\alpha} - \frac{\delta\sqrt{n}}{\sigma}\right).$$

Solve for $n$:

$$n = \left(\frac{(z_{1-\alpha} + z_{1-\beta})\sigma}{\delta}\right)^2.$$

**Two-Sample Tests (Lec 17–19)**

**Independent samples (means)**

Equal-variance pooled SD:

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

Large-sample equal-variance z:

$$z = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \approx N(0,1)$$

Unequal-variance (large $n$) z:

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx N(0,1)$$

**Small sample t-tests**

One-sample: $t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$ (Normal data).

Two-sample equal-var: $t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$.

Welch (unequal var):

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx t_{d'} \quad d' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

**Useful $t$ critical values (5%)**

One-sided ($t_{0.95,df}$): df 4: 2.132; df 9: 1.833; df 14: 1.761; df 29: 1.699; df $\infty$: 1.645.

Two-sided ($t_{0.975,df}$): df 4: 2.776; df 9: 2.262; df 14: 2.145; df 29: 2.045; df $\infty$: 1.96.

**Paired samples**

Differences $d_i = x_{1i} - x_{2i}$, $\overline{d} = \frac{1}{n}\sum d_i$:

$$s_d = \sqrt{\frac{1}{n-1}\sum(d_i - \overline{d})^2}, \qquad t = \frac{\overline{d}}{s_d/\sqrt{n}} \sim t_{n-1}$$

**Equality of variances (F test)**

$$H_0 : \sigma_1^2 = \sigma_2^2, \qquad F = \frac{s_1^2}{s_2^2} \sim F_{n_1-1,n_2-1}\ (H_0)$$

Convention: put the larger sample variance in numerator so $F \geq 1$ (then use appropriate tail).

**Tests with proportions (large samples)**

One proportion:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1)$$

Interpretation: use $p_0$ in the SE under $H_0$.

95% CI for $p$ (plug-in):

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Two proportions (independent):

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \qquad z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \approx N(0,1)$$

Note: pooled $\hat{p}$ is used for the SE *under* $H_0 : p_1 = p_2$.

CI for $(p_1 - p_2)$ (plug-in, large $n$):

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

**Univariate Nonparametric Tests**

**Review of Distributions**

We have covered several distributions derived from the Normal distribution:

- **Chi-square Distribution** ($\chi_k^2$):
  - If $Z_1, \ldots, Z_k \sim N(0,1)$ are independent, then $X = \sum_{i=1}^k Z_i^2 \sim \chi_k^2$.
  - Notation: $V \sim \chi^2(n)$.
  - Mean: $k$, Variance: $2k$.
  - PDF:
  $$f_k(x) = \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}, \quad x > 0.$$
- **F Distribution** ($F_{m,n}$):
  - If $U \sim \chi_m^2$ and $V \sim \chi_n^2$ are independent, then $X = \frac{U/m}{V/n} \sim F_{m,n}$.
  - Used for testing equality of variances ($H_0 : \sigma_1^2 = \sigma_2^2$).
- **Student's t Distribution** ($t_n$):
  - If $Z \sim N(0,1)$ and $U \sim \chi_n^2$ are independent, then $X = \frac{Z}{\sqrt{U/n}} \sim t_n$.
  - Used for small sample tests of means.

**K-Multinomial Random Variables**

Let $Y \sim km(n, p_1, p_2, \ldots, p_k)$.

- $Y$ is a vector of $k$ counts $(Y_1, \ldots, Y_k)$.
- $\sum p_i = 1$.
- $n$ total trials. $Y_k$ is the number of times we observe the $k$-th outcome.

**The Univariate Chi-square Test**

We want to test if a sample comes from a specific multinomial distribution (Goodness of Fit).

- **Hypothesis**: The data is drawn from a multinomial distribution with probabilities $p_1, \ldots, p_k$.
- **Test Statistic (Pearson's Chi-square)**:

$$Q_{k-1} = \sum_{i=1}^k \frac{(Y_i - np_i)^2}{np_i} \sim \chi_{k-1}^2$$

- Here, $Y_i$ is the *Observed* count ($O_i$) and $np_i$ is the *Expected* count ($E_i$).

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

**Proof for k=2**

For $k = 2$, the Chi-square statistic is:

$$Q_1 = \frac{(Y_1 - np_1)^2}{np_1} + \frac{(Y_2 - np_2)^2}{np_2}$$

Since $Y_2 = n - Y_1$ and $p_2 = 1 - p_1$, this simplifies to:

$$Q_1 = \frac{(Y_1 - np_1)^2}{np_1} + \frac{((n - Y_1) - n(1-p_1))^2}{n(1-p_1)} = \frac{(Y_1 - np_1)^2}{np_1(1-p_1)}.$$

$$Q_1 = \frac{(Y_1 - np_1)^2}{np_1(1-p_1)}$$

This is exactly the square of the z-statistic for a (single) proportion:

$$z_k = \frac{Y_k - np_k}{\sqrt{np_k(1-p_k)}}.$$

In particular,

$$z_k^2 = \frac{(Y_k - np_k)^2}{np_k(1-p_k)}.$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \implies z^2 = \frac{(Y_1/n - p_1)^2}{p_1(1-p_1)/n} = \frac{(Y_1 - np_1)^2}{np_1(1-p_1)}$$

Thus, for $k = 2$, the Chi-square test is equivalent to the two-sided z-test for proportions, and

$$Q_1 = z^2 \approx \chi^2(1).$$

**Bivariate Nonparametric Tests**

**Bivariate Chi-square Test (Test of Independence / Homogeneity)**

Used to test if two (or more) groups have the same distribution of categorical outcomes.

- **Setup**: Data is arranged in a contingency table with $J$ groups (columns) and $K$ values (rows).
- **Hypothesis**: $H_0$: The distributions of the values are the same across groups.
- **Expected Counts ($E_{ij}$)**: Under $H_0$, the best estimate for the probability of being in category $i$ is the pooled proportion:

$$\hat{p}_i = \frac{\sum_{j=1}^{J} Y_{ij}}{N}$$

Then, $E_{ij} = \hat{p}_i \times$ (Total count for Group $j$).

$$E_{ij} = \hat{p}_i \sum_{k=1}^{K} Y_{kj}$$

$$\hat{p}_i = \frac{\sum_{j=1}^{J} Y_{ij}}{N}, \quad E_{ij} = \hat{p}_i \sum_{k=1}^{K} Y_{kj}.$$

$$E_{ij} = \frac{(\text{Row Total}_i) \times (\text{Column Total}_j)}{N}$$

- **Test Statistic**:

$$Q = \sum_{j=1}^{J} \sum_{i=1}^{K} \frac{(Y_{ij} - E_{ij})^2}{E_{ij}}$$

- **Distribution**: Under $H_0$, $Q \sim \chi^2_{(J-1)(K-1)}$.

$$Q = \sum_{j=1}^{J} \sum_{i=1}^{K} \frac{(Y_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{((J-1)\times(K-1))}.$$

- **Condition**: Works well if expected counts in each cell are at least 5.

**Fisher's Exact Test**

- Used when sample sizes are small (e.g., cell counts $< 5$) where the Chi-square approximation fails.
- Does not rely on the CLT or large samples.
- Computationally intensive for large tables.

**Median Tests**

Useful when data is skewed or has outliers (t-tests might be invalid).

**0.0.1 One Sample Median Test**

- $H_0 : \text{Median}(X) = m_0$.
- Under $H_0$, we expect 50% of observations to be below $m_0$.
- Let $C$ = count of observations below $m_0$.
- Under $H_0$, $C \sim \text{Binomial}(n, 0.5)$. We calculate the p-value using the Binomial distribution.

*Two Sample Median Test Tests if two populations have the same median.

1. Combine the two samples and compute the **pooled median**.
2. For each sample, count the number of observations above and below the pooled median.
3. Create a $2 \times 2$ contingency table with these counts.
4. Perform a Chi-square test (or Fisher's exact test) on this table.

**Which Test? (Lec 17–21)**

**Mean / proportion decision guide**

**One sample:** Normal data + small $n \Rightarrow$ one-sample $t$; large $n \Rightarrow z$ (CLT).

Binary outcome $\Rightarrow$ one-sample proportion $z$ (large $n$).

**Two samples:** Paired $\Rightarrow$ paired $t$ on differences.

Independent means: equal var? pooled $t$ (small $n$) / pooled $z$ (large $n$); unequal var $\Rightarrow$ Welch $t$ / unequal-var $z$.

Two proportions $\Rightarrow$ two-proportion $z$ (pooled SE under $H_0$).

**Categorical outcomes guide**

One multinomial sample vs known $p_i \Rightarrow \chi^2$ GOF.

Two-way table (independence/homogeneity) $\Rightarrow \chi^2$ with df $(J-1)(K-1)$.

Small expected cell counts $\Rightarrow$ Fisher's exact.

**P-values (quick)**

Two-sided: $p = 2(1-\Phi(|z|))$ for $z$; similarly $p = 2(1-F_{t,df}(|t|))$ for $t$.

Right-tail: $p = 1-\Phi(z)$; left-tail: $p = \Phi(z)$.

**Critical values (common)**

$z_{0.10} \approx 1.28$ (80% CI), $z_{0.05} \approx 1.645$ (90% CI), $z_{0.025} \approx 1.96$ (95% CI), $z_{0.005} \approx 2.576$ (99% CI).

**Inference Templates (Lec 14–21)**

**Test/CI workflow**

1) Write $H_0, H_a$ and choose $\alpha$. 2) Compute statistic. 3) Get p-value (or compare to critical value). 4) State decision + interpretation in context.

**Standard errors (common)**

**Mean:** $SE(\overline{X}) = s/\sqrt{n}$.

**Diff of means (indep):** $SE(\overline{X}_1 - \overline{X}_2) = \sqrt{s_1^2/n_1 + s_2^2/n_2}$.

**Proportion:** $SE(\hat{p}) \approx \sqrt{\hat{p}(1-\hat{p})/n}$.

**Diff of proportions:**

$$SE(\hat{p}_1 - \hat{p}_2) \approx \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

**Useful null distributions from Normal data**

If $X_i \sim N(\mu, \sigma^2)$ iid:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \qquad \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

If $X_i$ and $Y_j$ are Normal samples with variances $\sigma_1^2, \sigma_2^2$:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

**Two-sided $t$ p-value template**

If $t_{df}$ computed, then $p = 2(1 - F_{t,df}(|t|))$.

**$\chi^2$ df reminders**

GOF with $k$ categories: df $= k-1$.

Independence with $K \times J$ table: df $= (K-1)(J-1)$.

If you estimate parameters in GOF, subtract those estimated parameters from df.

**Useful $\chi^2$ critical values**

At 5%: df 1: 3.841; df 2: 5.991; df 3: 7.815; df 4: 9.488; df 5: 11.070; df 6: 12.592.

At 1%: df 1: 6.635; df 2: 9.210; df 3: 11.345; df 4: 13.277.

**Common interpretation pitfalls**

Fail to reject $\neq$ accept $H_0$.

Statistical significance $\neq$ practical significance.

Small p-value means data unlikely under $H_0$ (not: $H_0$ is unlikely).

CI width shrinks like $1/\sqrt{n}$.

**Computation Recipes (Lec 14–21)**

**1-sample mean ($H_0 : \mu = \mu_0$)**

Compute $\overline{x}$, $s$, $n$.

Large $n$: $z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$.

Small $n$ (Normal): $t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$ with df $n-1$.

Two-sided p-value: $2(1-\Phi(|z|))$ or $2(1-F_{t,df}(|t|))$.

**2 independent means ($H_0 : \mu_1 = \mu_2$)**

Compute $\overline{x}_1, \overline{x}_2, s_1, s_2, n_1, n_2$.

Large $n$: $z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$.

Small $n$ (Normal): pooled $t$ if equal variances; Welch $t$ otherwise.

**Paired means ($H_0 : \Delta = 0$)**

Compute differences $d_i = x_{1i} - x_{2i}$, then test $H_0 : E[d] = 0$ via $t = \frac{\overline{d}}{s_d/\sqrt{n}}$.

**1-sample proportion ($H_0 : p = p_0$)**

Compute $\hat{p} = x/n$. Use $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$ (SE under $H_0$).

**2 proportions** ($H_0 : p_1 = p_2$)
Compute $\hat{p}_1 = x_1/n_1$, $\hat{p}_2 = x_2/n_2$, pooled $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$, then

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}.$$

**$\chi^2$ GOF**
Compute expected counts $E_i = np_i$ from null model, then

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \quad df = k - 1.$$

**$\chi^2$ independence / homogeneity**
Compute row totals, column totals, $N$, then $E_{ij} = \frac{(\text{row } i)(\text{col } j)}{N}$ and

$$Q = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad df = (K-1)(J-1).$$

**Randomized Experiments (Lec 22)**

**Potential outcomes**
$Y_{0i}$: outcome if not treated; $Y_{1i}$: outcome if treated; $D_i \in \{0,1\}$ treatment indicator.

$$Y_i = (1 - D_i)Y_{0i} + D_i Y_{1i} = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

**Causal estimands**
ATE: $E[Y_{1i} - Y_{0i}]$.
TT: $E[Y_{1i} - Y_{0i} \mid D_i = 1]$.

**Selection bias decomposition**
Observed difference ("confounding comparison"):

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] = TT + SB$$

where $TT = E[Y_{1i} - Y_{0i} \mid D_i = 1]$ and $SB = E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$. Random assignment $\Rightarrow (Y_{0i}, Y_{1i}) \perp D_i \Rightarrow SB = 0$, so diff-in-means identifies ATE.

**ITT (intent-to-treat)**
If $Z_i$ is random assignment (offer/encouragement), ITT is:

$$ITT = E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]$$

Useful under non-compliance (keeps randomization intact).

**Diff-in-means estimator (simple RCT)**
With treated group $T$ and control group $C$:

$$\widehat{ATE} = \overline{Y}_T - \overline{Y}_C, \qquad SE(\widehat{ATE}) \approx \sqrt{\frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}}$$

**Compliance language**
$Z_i$ = assigned; $D_i$ = received. If perfect compliance ($D_i = Z_i$), then ITT = ATE. If not, ITT remains valid for the effect of *assignment*.

**Heterogeneous effects + balance checks**
Treatment effects may vary across subgroups; randomization implies groups should look similar on *baseline* covariates in expectation (can check with comparisons).

**Why randomization helps**
Random assignment makes treatment independent of unobserved determinants of $Y$ ("exogenous variation"): treated and control are comparable in expectation.

**Breza et al. (2021) design (example)**
Two-stage randomization (county $\rightarrow$ zipcode) enables causal measurement with large-scale online ad delivery; primary outcomes measured at county/zip levels.

**What observational studies compute (HBR effect)**
Often reported:

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0]$$

This is a **confounding comparison** unless treatment is as-good-as-random.

**Observed outcome identity**
Even though $(Y_{0i}, Y_{1i})$ are both defined for each $i$, we only observe one of them:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i.$$

**Full decomposition (TT + selection bias)**
Let $\Delta_i = Y_{1i} - Y_{0i}$. Define $TT = E[\Delta_i \mid D_i = 1]$ and $SB = E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]$. Then

$$E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] = TT + SB.$$

**Key equalities under random assignment**
If $D_i$ is randomly assigned (independent of $(Y_{0i}, Y_{1i})$), then

$$\begin{aligned} E[Y_{0i} \mid D_i = 1] &= E[Y_{0i} \mid D_i = 0] \\ \Rightarrow \quad E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] &= ATE. \end{aligned}$$

**Stylized example (why confounding can be huge)**
If the causal effect is modest (e.g., $ATE \approx 5$) but treated units have much higher baseline outcomes $Y_{0i}$, then $SB$ can be large and the observational difference can be huge even when the true causal effect is small.

**Practical checklist for randomized experiments**
- Define population + unit of randomization (individual / cluster).
- Define treatment $D_i$ and outcomes $Y_i$ (timing matters).
- Verify balance: compare baseline covariates across treatment/control (means, proportions, $\chi^2$ tables).
- Handle non-compliance: report ITT (effect of assignment).
- Consider spillovers: if treatment affects untreated units, simple comparisons may not recover the desired ATE.

**Two-stage randomization intuition (Breza)**
Stage 1 creates **different treatment intensities** (high vs low) at the county level; stage 2 randomizes treatment at the zipcode level within counties.
Comparisons can identify intent-to-treat effects of being in treated zipcodes and can be used to study spillovers when intensity varies.

**ATE vs TT (heterogeneous effects)**
If treatment effects vary across people, then $ATE = E[Y_{1i} - Y_{0i}]$ may differ from $TT = E[Y_{1i} - Y_{0i} \mid D_i = 1]$ because the treated subset can be different.

**How to check randomization (balance)**
Pick baseline variables measured *before* treatment. Compare treatment vs control: mean differences ($t/z$ tests), proportions (two-proportion $z$), and categorical distributions ($\chi^2$/Fisher).
Interpretation: you expect some imbalance by chance; systematic imbalance suggests randomization problems or attrition.

**Noncompliance**
If assignment $Z_i$ does not equal receipt $D_i$, report ITT (effect of $Z_i$). "As-treated" comparisons reintroduce selection bias. **If you need the effect of actually receiving treatment:** under additional assumptions, a common "treatment-on-treated" estimand is

$$\frac{ITT}{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]}.$$

**Online ads: why "targeted search" is tricky**
If ads are shown because a user searches for relevant keywords, then $D_i$ is correlated with purchase intent, so $E[Y_{0i} \mid D_i = 1] > E[Y_{0i} \mid D_i = 0]$ (positive selection bias). Randomized field experiments create exogenous variation in $D_i$.

**Breza et al. (2021) outcomes (as reported in lecture)**
- Travel: average distance traveled decreased by about 0.993 percentage points in high-intensity counties vs low-intensity counties (pre-holiday window).
- COVID: treated zipcodes saw about a 3.5% reduction in infections vs controls (measured post-holiday).