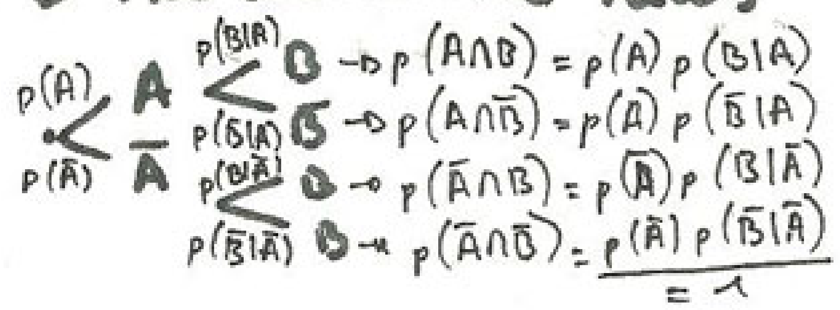
1. Probability Theory

→ Kolmogorov Axioms

Ω : sample space, $E \subseteq \Omega$: event

- $\forall E \subseteq \Omega$, $0 \leq P(E) \leq 1$
- $P(\Omega) = 1$
- $\forall E_1, E_2 \subseteq \Omega$, $E_1 \cap E_2 = \emptyset \Leftrightarrow P(E_1 \cup E_2) = P(E_1) + P(E_2)$

→ Probabilistic laws



$\Rightarrow$

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(B|A)$
- $P(A|B) = P(B|A)\, P(A) / P(B)$

- $A \wedge B$ are mutEx $\Rightarrow P(A|B) = P(B|A) = 0 \Rightarrow \begin{cases} P(A\cup B) = P(A) + P(B) \\ P(A \cap B) = 0 \end{cases}$

  $\Rightarrow$ Since wlog A depends on $\bar{B}$, $A \wedge B$ are dependent

- $A \wedge B$ are indep $\Rightarrow P(A|B) = P(A) \wedge P(B|A) = P(B) \Rightarrow \begin{cases} P(A\cup B) = P(A) + P(B) - P(A)P(B) \\ P(A\cap B) = P(A)P(B) \end{cases}$

  $\Rightarrow A \wedge B$ are not mutEx (indep. $\wedge$ mutEx are mutEx props.)

→ Distribution Functions

. Discrete

pmf: $\sum_{x \in \Omega} f(x) = 1$

$f(x) = P(x)$

cdf: $F(x) = \sum_{i=0}^{x} f(x_i)$

icdf: $F^{-1}(y) = x$ iff $F(x) = y$

. Continuous

pdf: $\int_a^b f(x) = 1 \quad (\Omega = (a,b))$

$\int_\alpha^\beta f(x)\,dx = P(\alpha \leq x \leq \beta)$

cdf: $F(x) = \int_{-\infty}^{x} f(x)\,dx$

icdf: $F^{-1}(y) = x$ iff $F(x) = y$

→ Moments

$$\underbrace{E}_{\text{expectation operator}}(X^m) = \begin{cases} \int_{-\infty}^{\infty} x^m f(x)\,dx & \text{if discrete} \\ \sum_{x \in \Omega} x^m f(x) & \text{if continuous} \end{cases}$$

Side-note: variance not constant within all R.V.s $\Rightarrow$ Heteroscedasticity

- moments are akin to differentiation
  - → 1st raw moment (= 1st central moment) $\Rightarrow$ mean $E(X)$
  - → 2nd raw moment → 2nd central moment $\Rightarrow$ variance $E(X^2) - E(X)^2$ (raw / central)
  - → 3rd $\Rightarrow$ skewness
  - → 4th $\Rightarrow$ kurtosis

- Jensen's inequality

$$X\,(:: \text{R.V.}) \wedge \varphi\,(:: \text{convex function}) \Rightarrow \varphi(E(X)) \leq E(\varphi(X))$$

note: lin. comb of R.V. is R.V.

- Moment generating function: $g(t) = E(e^{tX})$

→ Characteristic function: $\varphi(X) = E(e^{itX})$

1. integrate/sum w/ resp. to X.
2. differentiate number of times corresponding to moment
3. $t=0$ gives the raw moment e.g. $g'''(0)$ is 3rd raw m.

→ Variance & Co-variance

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

$$Cov(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y)$$

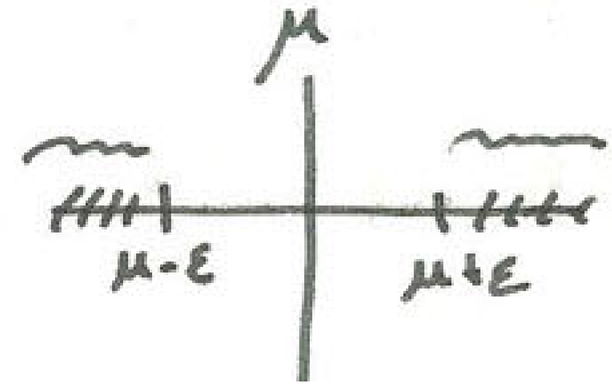
→ $Cov(X,X) = Var(X)$, $Cov(X,Y) = 0 \Leftrightarrow X, Y$ indep.

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

- Chebyshev's inequality

$$P(|X - \mu| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

$$\Rightarrow P(X \leq \mu - \varepsilon \vee X \geq \mu + \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$



→ Multivariate Probability

- Joint - marginal distributions

For x, y indep. ⇒

discrete:

	x_1	x_2	...	y-marginal
y_1		$p(x_1 \cap x_2)$	...	
y_2			...	
...	...	...	...	
x-marginal				1 ← $\underline{\underline{1}}$

continuous:

joint: $f(x,y) = f(x) f(y)$

x-marginal: $f(x) = \int_{\Omega_y} f(x,y)\, dy$

$f(x|y) = f(x,y) / f(x)$

⚠ ⇒ $E(XY) = \iint xy\, f(x,y)\, dx\, dy$

⇒ $E(X|Y) = \int x\, f(x|y)\, dx$

- Conditional distribution $P(Y=y | X=x) = \frac{P(Y=y \cap X=x)}{P(X=x)}$
- Chapman-Kolmogorov equation

$f(x_1, \ldots x_{n-1}) = \int_{-\infty}^{\infty} f(x_1, \ldots x_n)\, dx_n$ ← marginalisation of x_n

2. Information Theory

- Information value of message: $I(m) = -\log_2[p(m)]$
- Entropy of R.V. (measure of uncertainty): $H_{(b)}(X) = E(I(X))$

↳ discrete R.V. w/ pmf $p(x_i)$ $\quad H_b(X) = -\sum_{i=1}^{n} p(x_i) \log_b p(x_i)$

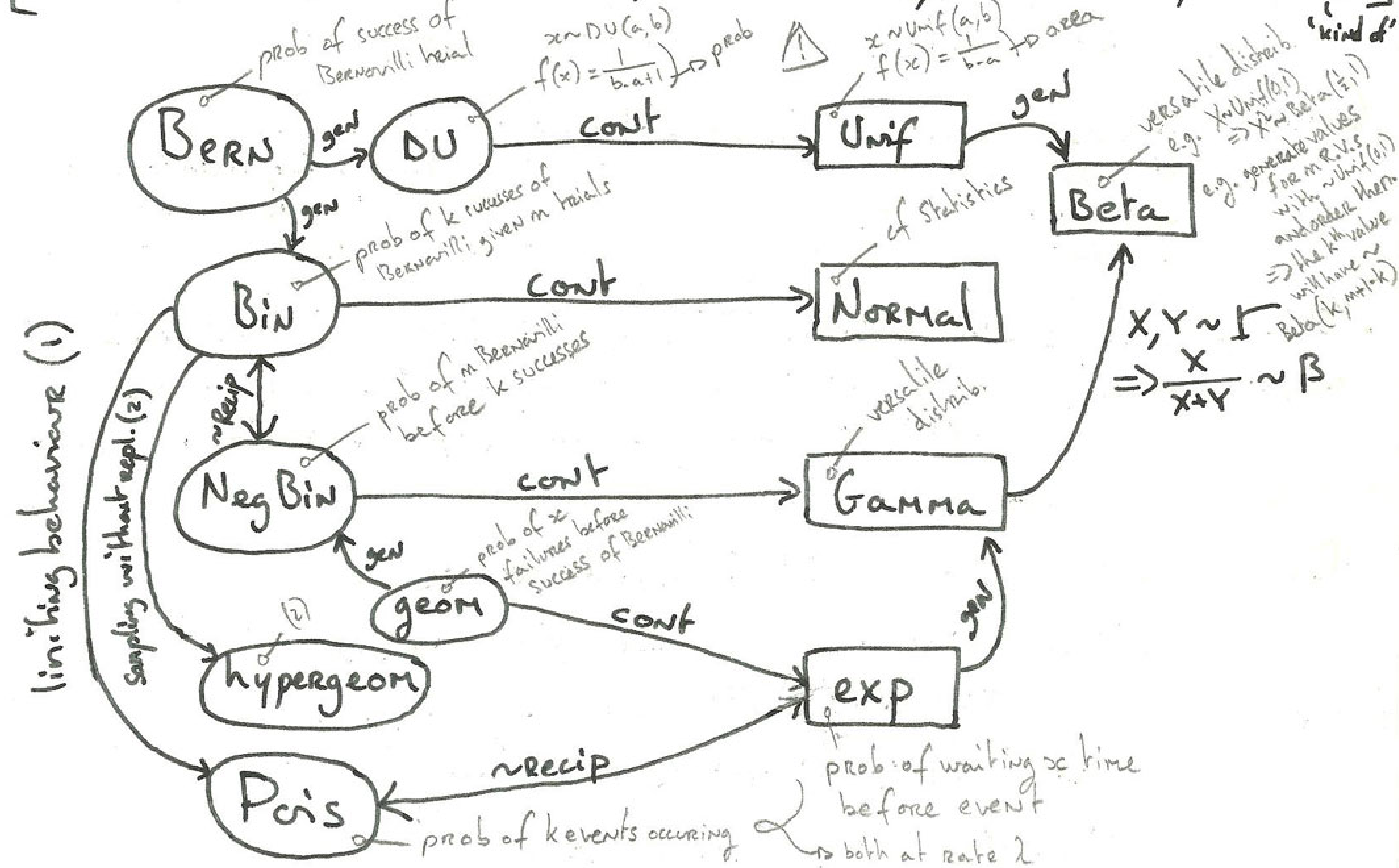
3. Combinatorics

→ Selection of k elements among n

	Repetition	∅ Repetition
Order → sequence, k-tuple	n^k (1,1,2)	(permutation: $P_{k=n!} = k!$) (1,2,3) k-permutation: $P^n_k = \frac{n!}{(n-k)!}$
∅ Order → subset / subbag	Multiset (bag) [1,1,2] $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$	combination C^n_k {1,2,3} $C^n_k = \binom{n}{k} = \binom{n}{n-k} = P^n_k / P_k = \frac{n!}{(n-k)!\,k!}$

4. Distribution Theory

[○ → Discrete ; □ → Continuous ; $\xrightarrow{\text{cont}}$ → Making continuous ; $\xrightarrow{\text{gen}}$ → generalize ; $\overset{\sim\text{recip}}{\longleftrightarrow}$ Reciprocal ~ 'kind of']



(1) $X \sim$ Count of successes in ∞ trials of Bin. $\Rightarrow E(X) = np = \lambda \; (\Rightarrow p = \frac{\lambda}{n})$

$$P(X) = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!\,k!} \cdot \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n(n-1)\ldots(n-k+1)}{n^k} \cdot \lim \frac{\lambda^k}{k!} \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= 1 \cdot \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \cdot \underbrace{\lim_{n\to\infty} \left(1-\lambda/n\right)^{-k}}_{=1}$$

$\lim_{\infty} \left(1-\frac{\lambda}{n}\right)^n$ let $n = -x\lambda$: $\underbrace{\lim_{\infty} \left(1-\frac{1}{x}\right)^{x}}_{e}{}^{-\lambda} = e^{-\lambda}$

$$P(X) = \frac{\lambda^k}{k!} e^{-\lambda}$$

(2) Hypergeom (k, m, n, N) → pmf: $\dfrac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$ "prob of k successes in n draws from popula° of size N containing m successes without replacement"

whereas Bin(n, p) can be understood as "prob of k successes from n draws with replacement"

Note: Monte-Carlo Method (approximation procedure based on randomness)

1. Define input domain
2. Generate inputs randomly from distrib
3. Check generat inp. deterministically
4. aggregate results

5. Stochastic Processes

Def: "bunch of numbers" ~ collection of R.V.s with underlying generating processes / distributions

<u>stochastic = random</u> (over time)

Classif:

index set	state space	
discrete	discrete	→ discrete markov
discrete	continuous	→ time series, markov chain monte carlo
continuous	discrete	→ continuous markov
continuous	continuous	→ brownian motion, etc.

→ variance increases with time

- IID property: R.V.s are indep. and identically distributed
- markov property: current state does not depend on previous states ~ akin to memorylessness (of exp distrib).

→ Poisson process

Def: counting process with indep. increments

→ interarrival times are IID with exp(λ)

(a.k.a. #events in an interval τ has ~ pois(λ))

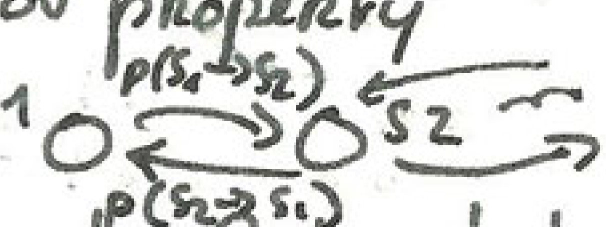
$$\Rightarrow P(N(s+\tau) - N(s) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

→ $N(0)=0 \wedge \forall t,\ N(t) \leq N(t+\varepsilon)$

→ Markov chains

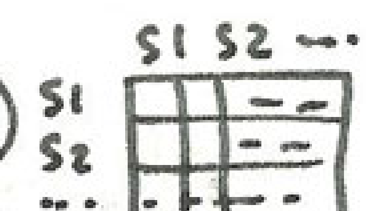
Def: stochastic / random process with markov property

→ usually represented with directed graph



- Absorbing markov chains: contains a subset of its state space such that once reached, other states (outside of subset) are unreachable

→ transition matrices (aka. stochastic / random matrices)



- For absorbing markov, it can be brought to canonical form (without TDMU in general) ← they don't match up!

$$P = \begin{pmatrix} Q & R \\ 0 & I_r \end{pmatrix}$$

Q: prob of transient state to tr. state; R: prob of transient state to absorbing state; 0: 0 matrix; I_r: identity → absorbing state

→ fundamental matrix: $N = \sum_{k=0}^{\infty} Q^k = (I_t - Q)^{-1}$

Note: relationship with statistics:

	popula°	sample	mean	theoretical
mean	μ	$\bar{x}$	$\hat{\mu}$	$E(X)$
variance	σ^2	s^2	$\hat{\sigma}^2$	$Var(X)$

1. Looking at data - distribu°

→ individuals ∈ popula°:
  ↳ characteristics → variables ↗ qualitative: pie chart, bargraph
  ↘ quantitative: stemplot, histogram, scatterplot...

$h = \frac{\text{count}}{\text{total}} \times w$

→ distribu° of a variable:
- → shape (sym^a, skewed)
- → center ($\bar{x}$, M)
- → spread (s^2, σ, IQR)
- → outliers ($\notin [Q_L - 1.5\,IQR, Q_U + 1.5\,IQR]$)

→ Numerical descrip°:
- $\bar{x} = \frac{1}{n}\sum x_i$
- $\sigma = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$
- $M = \begin{cases} n \text{ odd}: x_{\left(\frac{n+1}{2}\right)} \\ n \text{ even}: \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+1}{2}\right)}}{2} \end{cases}$

5 num. sum.

Min Q_L M Q_U max

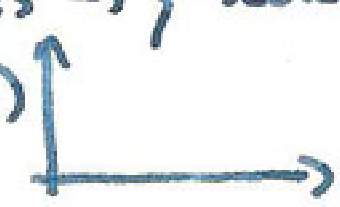
(box plot)

- Q_U / Q_L: media of values right/left of median, excluding the median.

$IQR = Q_U - Q_L$

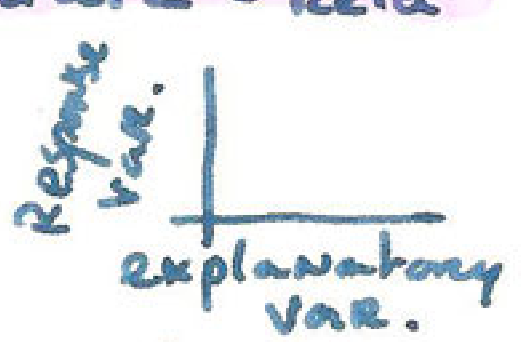
→ density curves:
- Normal dist:  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

  68 - 95 - 99.7 rule: $\mu \pm \sigma$, $\pm 2\sigma$, $\pm 3\sigma$
- standardiza°: z-scores: $z = \frac{x - \mu}{\sigma}$
- Normal quantile plot:
  - data are ranked → percentiles → z-score of percentile
  - Assess if dist is normal



2. Looking at data - rela°

→ scatterplot:



→ lurking var: not in study but explains response var.

→ confounding var: both var explain resp. but cannot distinguish effect

→ interpreta°:
- form: linear, curved, clustered
- dir°: positive, negative
- strength: weak, strong

→ correla°:
- Linear cor. coef: $R = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) = \frac{1}{n-1}\sum z(x_i)\, z(y_i)$
  → Not resistant to outliers ($\bar{x}, \bar{y}$) ⇒ influential pts
- Least-squares regression line: → predic°

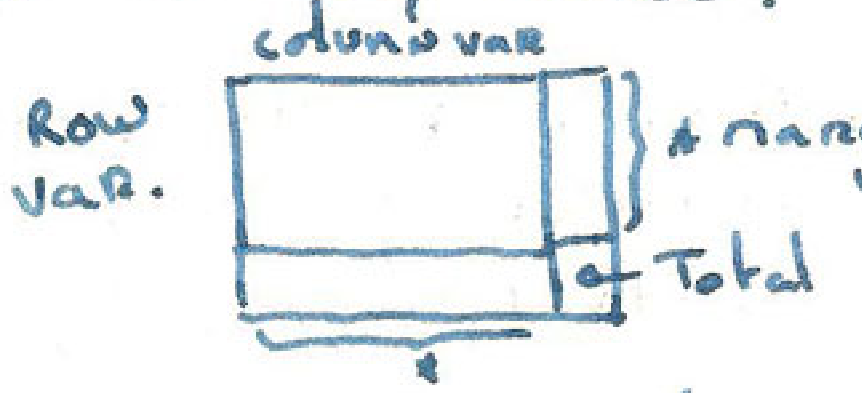
  $\hat{y} = b_1 x + b_0 \qquad \begin{cases} b_1 = R\frac{s_y}{s_x} \\ b_0 = \bar{y} - b_1\bar{x} \end{cases}$
- Coef of determina° R^2: percentage of variance in y that can be explained by changes in x.

→ Residuals : . Residual$_i$ = dist $(y_i - \hat{y})$ → $\hat{y}$ predicted by least-square Regression line

. Residual plot : → scattered ✓
→ pattern ✗
→ change width ?

→ caution before cor./reg. :
- always plot data
- influencial pts / outliers
- clumped data / wrong scaling
- lurking / confounding
- residual plot → associa° ~~⇒~~ causa°
- data from averages, extrapola°...

→ Two-way tables :



. joint distribu° : propor°(cell) = $\frac{\text{cell-value}}{\text{total}}$

. condi°l distribu° : $\frac{\text{cell}}{\text{Row-tot}}$ or $\frac{\text{cell}}{\text{col-tot}}$

→ simpson's paradox :

lurking var. in aggregated data

	A	B
	81/87.	234/270
	192/263	55/80
T	273/350	289/350

first two rows: A % > B %.
T: B % > A %.

3. Producing data

/ ~~Anecdotal data~~ / Available data / § Observa°l / Experimental { exp. units / treatment : factor / factor level

→ comparative experiments : → control (ø)
→ placebo (fake)

Bias : exp. favours certain outcomes (Lack of realism : design ≨ reality)

solu° : → Randomize
→ Replicate
→ Control

(double-blind : subj + exp don't know)

→ completely randomized designs : . individuals → random groups
. groups → random treatment

→ block / stratisfied designs :

. general : Subj ⟨ ♂ (e.g.) → CRD ⟨ G1 → A / G2 → B ⟩ compare
♀ (e.g.) → CRD ⟨ G1 → A / G2 → B ⟩ compare

CRD: Comp. Ran. Des.

. Matched pair :

Type I

subj → A → compare → B → conclude

A then B

Type II

subj 1 → A
subj 2 → B ⟩ compare

subj 1 & 2 closely matched

→ Sampling methods:
- convenience sampling ×
- voluntary response sampling ×
- Random sampling ✓
  - ↳ SRS (simple random sample)
  - → Stratisfied samples : pop → Criteria → SRS ; → ¬Criteria → SRS
  - → Multistage samples : pop → C_1 → $C_{1.2}$ → SRS ; → ¬$C_{1.2}$ → SRS ; → ¬C_1 → SRS
- Sampling surveys
  - ↳ nonresponse
  - ↳ response bias
  - ↳ wording effects
  - ↳ undercoverage

(- capture-recapture sampling:

	Y1	Y2
sample size:	200	120
nb tagged:		12

)

→ Sampling variability:

. Purpose of sampling → inference : use sample statistic to infer on popula° unknown parameter

. Statistic from SRS/ Comp. Rand. Experiments [CRE] has a sampling distribu° → how it varies in repeated data produc°

is subject to:
- → bias : center of sampling dist ≠ true parameter
- → variability : spread of samplin dist ⇒ margin of error

4. Probability

→ Randomness:

. Probability model for a random phenomenon
- → Sample space : set of all possible outcomes
  - → events : subsets of outcomes
  - → independent events : one does not influence the other
- → Assignment of probability :
  - → P (space) = 1
  - → P (event) ∈ [0, 1]
  - → $P(A^c) = 1 - P(A)$

. Probabilistic Rules:
- → $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - ⇒ A and B are disjoint : $P(A \cup B) = 0$
- → $P(A \cap B) = P(A)\,P(B \mid A)$
  - ⇒ A and B are independent : $P(B) = P(B \mid A)$
- → $P(A \mid B) = \dfrac{P(B \mid A)\,P(A)}{P(B)}$ (simplest case of Baye's rule)

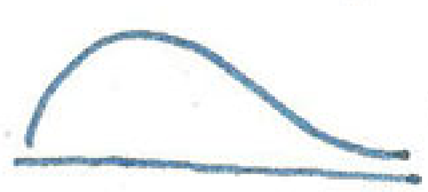
→ Random variables:

- discrete :

X	x_1 ... x_m
$P(X)$	p_1 ... p_m

$\Sigma\, p_i = 1$

- continuous :



Area = 1

→ Probability distribu°

- $\mu_X = \sum x_i p_i$ (for discrete)
  - → Law of large nb: sample size of random observations increases, sample means $\bar{x}$ get closer to μ_X
- $\sigma_X = \sqrt{\sigma_X^2} = \sqrt{\sum (x_i - \mu_X) p_i}$ (for discrete)

→ linear transf.

$\mu_{a+bX} = a + b\mu_X$

$\sigma^2_{a+bX} = b^2 \sigma_X^2$

→ Correlated events

$\mu_{X+Y} = \mu_X + \mu_Y$

$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$

$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$

if X and Y are independent: $\rho = 0 \Rightarrow \sigma^2_{X+Y} = \sigma^2_{X-Y}$

5. Sampling Distributions

→ Binomial distribu°

- Setting:
  - → categorical variable can take two values: Success or Failure
  - → total nb of outcomes of K is fixed
  - → outcomes / observa° are independent and have same probability
- Binomial dist of a count X of successes among n obs w/ probability p.

→ In statistical sampling:

- A popula° contains a propor° p of successes, the count X of successes in an SRS of size n has the binomial dist $B(n, p)$
- Binomial sampling dist for counts is used if pop size > 20 x SRS size

→ Binomial mean / variance of X

- $\mu_X = np$, $\sigma_X = \sqrt{np(1-p)}$
- $\hat{p} = \frac{X}{n} \Rightarrow \mu_{\hat{p}} = p$, $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ — sample propor° is unbiased estimator of popula° propor°

→ in an SRS of size n where X is the count of successes

→ $\hat{p}$ for a sample of size n, its dist is over several samples

→ Normal approxima° of binom dist

when $\begin{cases} np \geq 10 \\ n(1-p) \geq 10 \end{cases}$, $X \sim N\left(np, \sqrt{np(1-p)}\right)$

$\hat{p} = \frac{X}{n} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

→ Binomial probability

$p(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

→ Sampling dist for sample means

→ Concept of sampling dist of a statistic: distribu° of all possible $\underbrace{\text{combina° of}}_{\text{of size } n}$ possible values taken by the statistic

→ = probability dist of the statistic

→ $\left.\mu_{\bar{x}} = \mu \ , \ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\right\}$ averages are less variable than individual observa°

→ Central Limit Theorem

"When randomly sampling from any popula° with mean μ and std dev σ, when n is large enough, the sampling dist of $\bar{x}$ is approximately normal $N(\mu, \sigma/\sqrt{n})$"

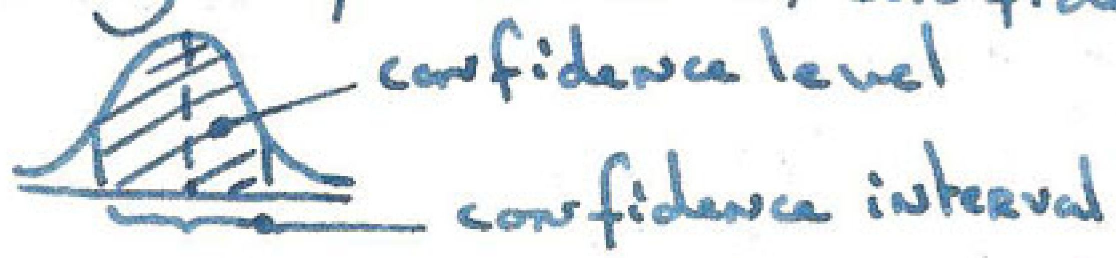
6. Introdution to inference

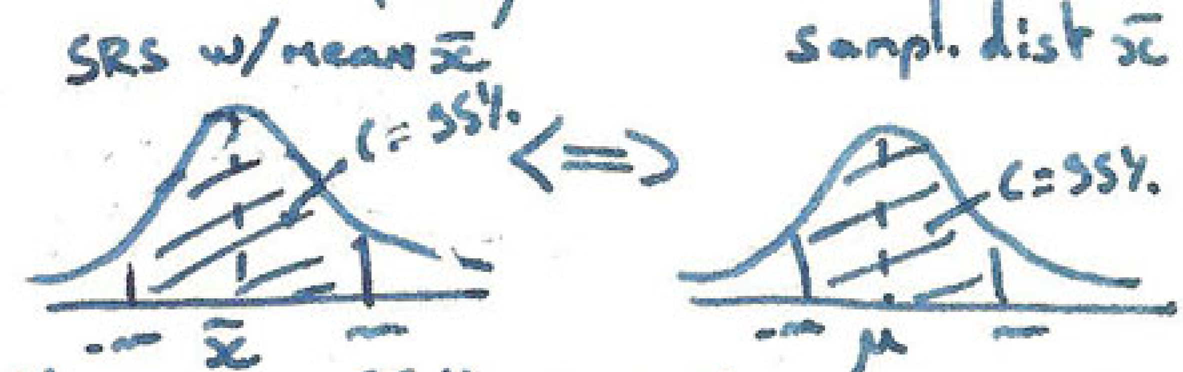
→ Defini°:

. Drawing conclus° about a popula° from sample data
. Methods: confindence intervals, tests of significance
. Appropriate when data are produced by SRS / CRE

→ Confidence intervals → statistical confidence

→ Range of values w/ confidence level C (%)





There is a 95% chance μ is within $\bar{x} \pm m$ ⟺ There is a 95% chance $\bar{x}$ is within $\mu \pm m$

→ to find C (probability [0, 1]) we need to standardize the sampling dist of $\bar{x}$: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

. practical use of z: z^* (Area under std normal curve between $-z^*$ and z^* is given by C: )

→ Thus confidence interval is: $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

$m = z^* \frac{\sigma}{\sqrt{n}}$

$\Rightarrow n = \left(\frac{z^* \sigma}{m}\right)^2$ to have a margin of error m in sample of size n.

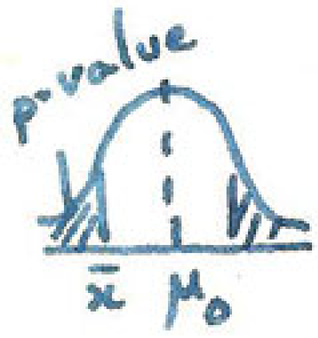
→ Test of significance → statistical significance

→ { Null hypothesis: specific statement about a parameter
Alternative hypothesis: $\neg H_0 = H_a$

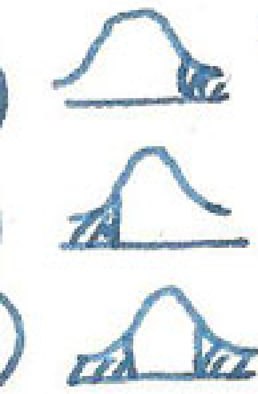
→ One-sided test: $\begin{cases} H_0 : \mu = c \\ H_a : \mu \neq c \end{cases}$

Two-sided test: $\begin{cases} H_0 : \mu = c \\ H_a : \mu < c \text{ or } H_a : \mu > c \end{cases}$

→ P-value :



- After finding the sampling distribu° of $\bar{x}$ assuming H_0, the p-value is found as the area under the curve for values at least as extreme, in the direc° of H_a as that of our SRS.
  - one-sided test : $\begin{cases} H_a : \mu > \mu_0 \to P(Z \geq z) \\ H_a : \mu < \mu \to P(Z \leq z) \end{cases}$ $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$
  - two-sided test : $H_a : \mu \neq \mu_0 \to p(Z \geq |z|)$



→ Significance level α

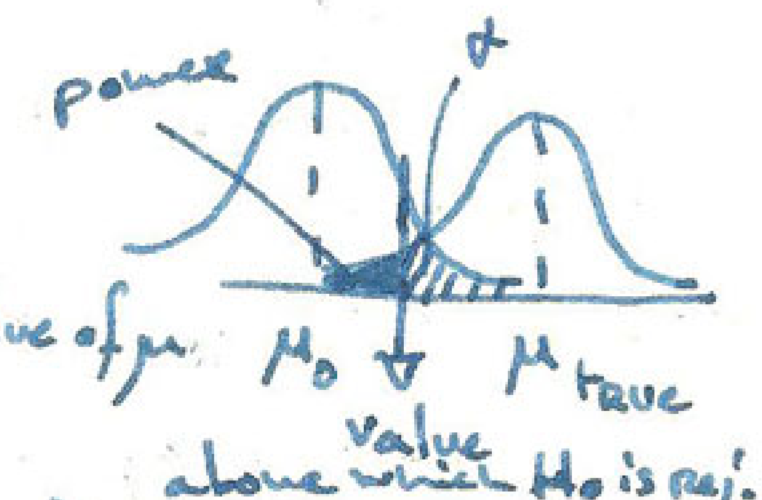
- Largest p-value tolerated : $\begin{cases} p \leq \alpha \Rightarrow \text{reject } H_0 \\ p > \alpha \Rightarrow \underline{\text{fail to reject } H_0} \end{cases}$

} C = 1 - α — confidence level
  └ confidence interval

p-value only supply evidence again

→ Power of a test of hypothesis

- With significance level α and $H_0 : \mu = \mu_0$, the power is the probability that the test will reject H_0 when {the alternative is true / compared with the true value of μ.



- Factors:
  - Size of the effect ($\nearrow \Rightarrow \nearrow$ power)
  - Significance level ($\searrow \alpha \Rightarrow \searrow$ power)
  - Sample size ($\nearrow n \Rightarrow \nearrow$ power)
  - variance / std dev ($\nearrow \sigma^2 \Rightarrow \searrow$ power)

→ Type I & II errors

- Type I : incorrectly reject H_0 → prob (type I) = α
- Type II : fail to reject correct H_0 → prob (type II) = β = 1 - power

7. Inference for Distributions

→ t-distribu°

- suppose σ is unknown, then we can approximate it with the std dev of the sample (s), thus when trying to infer (with the sampling distribu° of $\bar{x}$), we use the Standard Error: $SE_{\bar{x}} = s/\sqrt{n}$ (when $n > 30$, $SE_{\bar{x}} \nearrow$ so we use t-distribu°)
- Instead of infering with a <u>one-sample</u> z-statistic: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

  we use the <u>one-sample</u> t-statistic $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

  ⟹ it has the t distribu° w/ n-1 degrees of freedom

L4

→ t-statistic
. practical use of t : t^*

$\begin{cases} df = n-1 \\ \text{conf level} = C \end{cases}$

$m = t^* s/\sqrt{n}$

. in practice we can also use t-statistic for matched pairs } $H_0 : \mu_{x_1 - x_2} = 0$ ↳ we treat the dif. between samples as a popula°

. t is robust to slight devia°, but affected by outliers & skewness }
→ $n<15$: data must be normal, ∅ outliers
→ $15<n<40$: mild skewness, ∅ outliers
→ $n>40$: t-stat will always be valid

→ Inference for non-normal dist.
• small n + skewed : apply transforma° (e.g. log)
• sign test : → distribu°-free
→ e.g. matched pairs $\substack{\nearrow TA \\ \searrow TB}$ ↳ $\forall\, TA_i - TB_i \neq 0$, $TA_i - TB_i = n$
→ $B(n, \frac{1}{2})$ ↳ p-value

→ Comparing two means
→ Two-sample z-stat (σ is known)
. We have SRS_1 SRS_2 independent, we want to see if they belong to same popula°
- sampling dist of $\bar{x}_1 - \bar{x}_2$ is normal with std dev $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$\rightarrow z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

→ Two-sample t-stat (σ is unknown)
- $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, $df = \text{~~smallest~~}(n_1 - 1, n_2 - 1)$) — approx of $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$

• $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE}$, assuming $H_0 : \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$
→ $t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$

• confidence level $C \Rightarrow t^* \Rightarrow m = t^* SE$
(df = smallest$(n_1 - 1, n_2 - 1)$)) → p-value given for $t(df)$

Note : when 2 popula° have same variance, then we can use pooled two-sample t procedures : $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

8. Inference for proportions

→ single propor° (propor° of successes in a popula°)
• population size > 10 × size of SRS ⇒ std dev $\hat{p}$ has sampling dist : $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$
→ confidence interval: $\hat{p} \pm m$ with $m = z^* SE = z^* \sqrt{\hat{p}(1-\hat{p})/n}$

→ "plus four" adjustment for producing more confident interval for p. } $\tilde{p} = \frac{\text{count of successes} + 2}{\text{count of all obs} + 4}$ } $m = z^* \sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$

. significance test for p

$H_0 : p = p_0 \rightarrow z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$) valid when $\begin{cases} np_0 > 10 \\ n(1-p_0) > 10 \end{cases}$

↳ for desired margin of error m : $n = \left(\dfrac{z^*}{m}\right)^2 p^*(1-p^*)$

→ two proportions

. similar to 2-sample tests : $\begin{cases} \text{pop} > 20n \\ n_1 p_1 > 10,\ n_2 p_2 > 10 \\ n_1(1-p_1) > 10,\ n_2(1-p_2) > 10 \end{cases} \Rightarrow (\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

→ "plus four" CI : $\tilde{p}_1 = \dfrac{X_1 + 1}{n_1 + 2}$, $\tilde{p}_2 = \dfrac{X_2 + 1}{n_2 + 2} \Rightarrow (\tilde{p}_1 - \tilde{p}_2) \pm z^* \widetilde{SE}_{diff}$

$\widetilde{SE}_{diff} = \sqrt{\dfrac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \dfrac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}}$

→ pooled estimate $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2} \Rightarrow SE_{diff_{\underline{p}}} = \sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$

→ relative risk $RR = \dfrac{\hat{p}_1}{\hat{p}_2}$ → to summarize comparison of 2 proportions

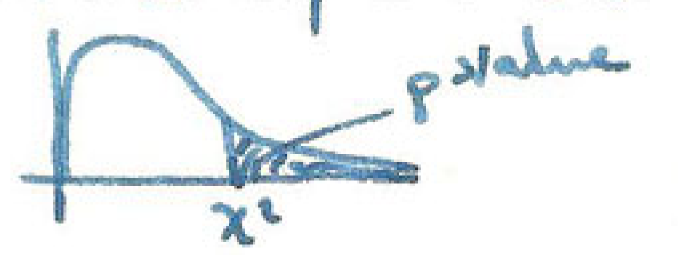
9. Inference for two-way tables

→ To compare row × column 2-way table, we assume no relaᵒ (null hypothesis) and find an expected count : $\dfrac{\text{Row tot} \times \text{column tot}}{n}$ (if H_0 is true)

→ We then apply the chi-square test :

$$X^2 = \Sigma \frac{(\text{obs count} - \text{expect count})^2}{\text{expect count}}$$

⇒ if H_0 is indeed true, the test has approximately a X^2 dist with $(r-1)(c-1)$ degrees of freedom



(↳ Conditional distribuᵒ: $\dfrac{\text{cell tot}}{\text{Row tot}}$ OR $\dfrac{\text{cell tot}}{\text{column tot}}$)

→ X^2 components : terms of the Σ → the largest ones point to the cells whose values differ the most to what would be expected by H_0

10. Inference for Regression

→ Simple linear regression

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ $\begin{cases} \beta_0, \beta_1 \text{ estimated from } \hat{y} \\ \sigma \text{ estimated by : } s = \sqrt{\dfrac{\Sigma e_i^2}{n-2}} \end{cases}$

- independent
- Residual : $\varepsilon_i = y_i - \hat{y}_i$
- normaly distributed $N(0, \sigma)$ → σ is the same for every x

$\boxed{SE_{b_0} = s\sqrt{\dfrac{1}{n} + \dfrac{\bar{x}^2}{\Sigma(x_i - \bar{x})^2}} \quad SE_{b_1} = \dfrac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}}}$

→ confidence interval (for β_1) → $b_1 \pm t^* SE_{b_1}$ → t for C & df: $n-2$

|| test of H_0 is based on t-statistic : $t = \dfrac{b_1}{SE_{b_1}}$

$\left[SE_{\hat{\mu}} = s\sqrt{\dfrac{1}{n} + \dfrac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} \right]$ (for μ_y) → $\hat{\mu}_y \pm t^* SE_{\hat{\mu}}$

→ estimated mean response for x^* : $\hat{\mu}_y = b_0 + b_1 x^*$

→ prediction interval :

$\hat{y} \pm t^* SE_{\hat{y}}$ $\left[SE_{\hat{y}} = s\sqrt{1 + \dfrac{1}{n} + \dfrac{(x^* - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}} \right]$

$\boxed{H_0 : \rho = 0 \text{ (cor. of } x \text{ and } y) \quad t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}}$