|  |  |  |
| --- | --- | --- |
| Slide 1 |  |  |
| Slide 2 |  |  |
| Slide 3 |  | Normally would see these as rows and columns. |
| Slide 4 |  | Each observed value:   * … is drawn from it’s own peculiar distribution * … is drawn from a dist with a common mean, but’s it’s own sd. * … is drawn fron a dist with it’s own mean, but a common sd. * … is drawn from the same dist – iid!!   Think of it as each observed value is pulled from the same underlying distribution.  doesn’t have to be normal.  doesn’t have to be known.  Ex Problems: Linguistic data or siblings/SAT scores |
| Slide 5 |  | Probability –  If we had a known population, we could use probability to describe an individual observation  Statistics -  Samples are assumed to represent populations, and we want to use our sample to infer “things” about the larger population |
| Slide 6 |  | Individual Observations  Every time I pull an observations, it’s coming from the same population  What is the RV? # tweets per day |
| Slide 7 |  | Independence assumption |
| Slide 8 |  | Large X is an rv  Small x is the observed values (essentially a placeholder for the values) |
| Slide 9 |  | no problem!! |
| Slide 10 |  | Each Xi might have it’s own pdf=f(xi)  pdf is a placeholder, not defined |
| Slide 11 |  | Recall from Probability… |
| Slide 12 |  | Probability density function’s  Makes the math make sense, not dealing with individual pdfs |
| Slide 13 |  | Because they are iid …  We don’t have to treat them as individuals, can create one pdf. Makes the math easier…    Focusing on notation – don’t necessarily need to get the math  Warning!! Need to meet the iid assumptions!! |
| Slide 14 |  | Don’t care about individuals – want summaries!!  NOT stalking Xi |
| Slide 15 |  | T is a place holder for any function (mean, var, etc)  linear on non-linear combination of the observed values    Y is an rv, because X is an rv  has it’s own distribution – different from X  Before: probability distribution (one value, or an interval)  Now: summaries (multiple values)  today - known dists (unrealistic)  next class – potentially unknown dist  Draw some pictures of probability dists vs sampling distribution of the mean |
| Slide 16 |  | Not a function of a parameter, must be random  Summarize data into a summary that makes a point or conclusion about the population  Just because they’re not a function of a parameter, doesn’t mean they can’t tell us something about a population  Estimator – like in linear regression  Test statistics – like a t-test statistic or wilcoxon |
| Slide 17 |  | Why is it random?  Because it’s from a random sample. Lots of possible samples!! |
| Slide 18 |  | Small x => sample data  Notation nightmares!! |
| Slide 19 |  | Big S – definition of the statistic |
| Slide 20 |  | small s – sample variance  Come back to n-1! |
| Slide 21 |  |  |
| Slide 22 |  |  |
| Slide 23 |  |  |
| Slide 24 |  |  |
| Slide 25 |  | Important!!  NOT saying the sample mean is the population mean  AM saying that the mean of all possible sample means is the population mean |
| Slide 26 |  |  |
| Slide 27 |  |  |
| Slide 28 |  |  |
| Slide 29 |  | Sample variance <> variance of the sample mean |
| Slide 30 |  | N-1!! Less important as n increases – but we just go ahead and do it anyway.  Aside; Using n does NOT lead to a biased sd. |
| Slide 31 |  |  |
| Slide 32 |  | Sampling distributions = Distribution of ALL the possible sample means (with a given n)  Now - enumerate and calculate  Code-along –  enumerate and calculate,  random sample with replacement  Next class –  sample only  and classical… |
| Slide 33 |  |  |
| Slide 34 |  | In the real world …  Which one can you create? Sample dist  use it to make assumptions about the other two  it’s an estimate  Population is real (to frequentists), but unseeable  Too large  changing?  Sampling distribution  theoretical  let’s you make inferences about your statistic! |
| Slide 35 |  | N – number of throws  Entire probability space – equally probably |
| Slide 36 |  |  |
| Slide 37 |  | Mean = E  3.5 is not observable – outside the probability space |
| Slide 38 |  |  |
| Slide 39 |  | Mean of a single number is that number… |
| Slide 40 |  | Not very interesting – equiprobable |
| Slide 41 |  | Mean – can’t get 3.5 |
| Slide 42 |  | Same X as the last … |
| Slide 43 |  |  |
| Slide 44 |  | Expand.grid = fully cross factored  All possible combintions |
| Slide 45 |  |  |
| Slide 46 |  | Repeats?? |
| Slide 47 |  | Look at that mean (E) |
| Slide 48 |  | Point out symmetry |
| Slide 49 |  | Possible sample means – NOT outcomes  Still talking about a known random variable and it’s distribution – fully enumerated |
| Slide 50 |  | Same mean!! |
| Slide 51 |  | Expectation of the sampling distribution of the sample means |
| Slide 52 |  |  |
| Slide 53 |  | Left: value x probability  Right:  add another die,  mean sample mean x probability |
| Slide 54 |  |  |
| Slide 55 |  |  |
| Slide 56 |  |  |
| Slide 57 |  | Here we did die, but it could be any observation – just added observations |
| Slide 58 |  | Why??  We can make conclusions about a populations from our sample – about a theoretical distribution  We can go from a population to a sampling distributions of sample means  No matter how many I add, same mean |
| Slide 59 |  |  |
| Slide 60 |  |  |
| Slide 61 |  |  |
| Slide 62 |  |  |
| Slide 63 |  |  |
| Slide 64 |  | Population – unknown/unseen  Sampling – theoretical (sorta)  What happens to the expectation of certain statistics as n increases? |
| Slide 65 |  | Spread – what measures that? |
| Slide 66 |  |  |
| Slide 67 |  |  |
| Slide 68 |  | Changes based on sqrt of n  What does this mean?  As the sample size grows, the var get smaller. Sample means get closer to the true population mean.  Known to be true. |
| Slide 69 |  | What happens at n=1? It IS the population.  Standard error formulas for other statistics are online – also ugly. Use R. |
| Slide 70 |  |  |
| Slide 71 |  |  |
| Slide 72 |  |  |
| Slide 73 |  |  |
| Slide 74 |  |  |
| Slide 75 |  |  |
| Slide 76 |  |  |
| Slide 77 |  |  |
| Slide 78 |  |  |
| Slide 79 |  |  |
| Slide 80 |  |  |
| Slide 81 |  |  |
| Slide 82 |  | Skip? Reading… |
| Slide 83 |  |  |