***Learning Statistics with R - University of Adelaide***

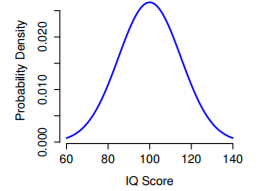
***Part IV – Statistical Theory***

**Chapter 10 – Estimating Unknown Quantities From A Sample**

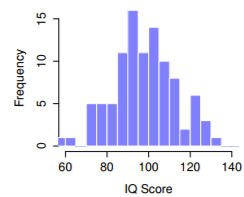
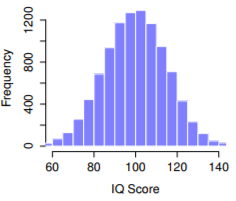
* Role of descriptive statistics = to concisely summarize what we DO know
* Role of inferential statistics = to learn what we do NOT know *from what we do*
* 2 big ideas: **estimation** and **hypothesis testing**.
* Estimation theory doesn’t make sense until you understand sampling

**10.1 - Samples, populations and sampling**

* ALL learning requires you to make assumptions
* 1st task: To come up w/ some fairly general assumptions about data that make sense 🡪 where **sampling theory** comes in
* **Probability Theory** = foundation upon which all statistical theory builds + **Sampling Theory** = the frame around which you can build the rest of the house.
* **Sampling theory** plays a huge role in *specifying the assumptions upon which statistical inferences rely.*
* Drawing inferences *from the sample* + *about the population.*
* Typical data set available to us is finite + incomplete.
* A **sample** is a *concrete*, while a **population** is a more abstract + refers to the set of ALL POSSIBLE observations you want to draw conclusions about (is generally much bigger than the sample)
* Ideal world = researcher begins a study w/ clear idea of what the **population of interest** b/c the process of designing a study + testing hypotheses about data the study produces depends on the population about which you want to make statements.
* usually researchers have vague ideas of what the population is + designs the study as best they can on that basis.
* Irrespective of how population is defined, the critical point is the **sample** = a subset of a population
* Goal = to use knowledge of the sample to draw inferences about the properties of the population.
* Relationship between the 2 depends on the *procedure by which the sample was selected* = **sampling method**
* The fact that the same sampling procedure can lead to different results each time we take a sample (i.e. sample of 5 cards w/out replacement repeated) means we refer to it as a **random process**
* Simplified: *a process has an element of randomness to it whenever it is possible to repeat the process and get different answers each time.*
* A procedure in which every member of the population has the same chance of being selected is a **simple random sample**.
* If we know that a sampling scheme is biased, that a sample doesn’t tell you very much about a population
* A simple random sample makes the data analysis much easier.
* Can be done **with or without replacement**
* w/ replacement = can observe the same population member multiple times in a sample
* Most statistical theory is based on the assumption the data arise from a SRS *with* replacement
* Rarely matters in real life if population of interest is large (> 10 entities), the difference between w/ + w/out replacement is negligible
* Almost impossible to obtain SRS from most populations
* **Stratified sampling** =population is divided into several different subpopulations, or **strata**.
* collect a SRS *from each of the strata*.
* sometimes easier to do than SRS, especially when population is already divided into distinct strata + can also be more efficient, especially when some subpopulations are rare.
* Ex: studying schizophrenia 🡺 better to divide population into 2 strata (schizophrenic + not-schizophrenic) + sample an equal number of people from each group.
* If you selected people randomly, you’d get so few schizophrenic people in the sample the study would be useless.
* This is **oversampling** = makes a deliberate attempt to over-represent rare groups.
* **Snowball sampling** = especially useful when sampling from a hidden/hard to access population (especially common in social sciences)
* Ex: opinion poll among transgender people + team might only have contact details for a few trans folks
* Stage 1: Survey starts by asking them to participate + at the end of the survey, participants are asked to provide contact details for other people who might want to participate.
* Stage 2 = new contacts are surveyed + process continues until we have sufficient data
* Big advantage to snowball sampling = gets you data in situations that might otherwise be impossible to get any.
* Main disadvantage = sample is *highly* NON-random + is so in ways that’re difficult to address
* procedure can also be unethical if not handled well b/c hidden populations are often hidden for a reason.
* might end up outing people who don’t want to be outed + can be intrusive to use people’s social networks to study them.
* Very hard to get people’s informed consent before contacting them, yet in many cases the simple act of contacting them + saying “we want to study you” can be hurtful.
* Social networks are complex things + just b/c you can use them to get data doesn’t always mean you should.
* **Convenience sampling** = samples are chosen in a way convenient to the researcher + not selected at random from population of interest.
* Snowball sampling = 1 type of convenience
* Studies on undergraduate psychology students automatically means data are restricted to a single subpopulation.
* Students usually get to pick studies they participate in, so a sample is a **self-selected subset** of psychology students, NOT a *randomly*-selected subset.
* In real life, most studies are convenience samples of 1 form or another
* Sometimes a severe limitation, but not always
* It *can* matter if data are not a SRS but it’s not quite as bad as it sounds.
* Some types of biased samples are entirely unproblematic
* **Stratified Sampling** = actually know what the bias is b/c YOU created it *deliberately*, often to *increase the effectiveness* of the study
* There’re statistical techniques you can use to adjust for biases you’ve introduced
* More generally though, remember that random sampling = a *means to an end, not the end in itself.*
* A bias in sampling method is only a problem if it causes you to draw the wrong conclusions.
* Don’t need sample to be randomly generated in *every* respect: only to the phenomenon of interest
* Ex: For a memory study, can sample randomly from all human beings currently alive w/ 1 exception
* Can only sample people born on a Monday OR Can only sample randomly from Australia.
* To generalize results to the population of ALL living humans, study 1 is better b/c we have no reason to think being born on a Monday has any interesting relationship to memory capacity
* Being Australian might matter b/c Australia is a wealthy, industrialized country w/ a very well-developed education system.
* People will have had life experiences much more similar to experiences of those who designed the tests + this shared experience might easily translate into similar beliefs the test
* When designing studies, it’s important to think about *what population you care about +* to try hard to sample in a way *appropriate to THAT population*.
* Usually forced to put up w/ a “sample of convenience”, but if so at least spend some time thinking about what the dangers of this practice might be.
* Secondly, if criticizing a study b/c of a sample of convenience rather than laboriously sampling randomly from the entire human population, have the courtesy to offer a specific theory as to how this might have distorted the results.
* Everyone in science is aware of this issue + does what they can to alleviate it.
* In most cases, populations scientists care about are *concrete* things that *actually exist* in the real world
* *Statisticians*, however, are interested in real world data + real science in the same way scientists are, but they also operate in the realm of *pure abstraction* in the way mathematicians do.
* As a consequence, **statistical theory** tends to be a bit abstract in how a population is defined.
* In much the same way psychological researchers operationalize abstract theoretical ideas in terms of concrete measurements, statisticians operationalize the concept of a “population” in terms of *mathematical objects* they know how to work w/ = **probability distributions**.
* Ex: IQ scores
* To a psychologist, population of interest = a group of actual humans who have IQ scores.
* A statistician “simplifies” this by **operationally defining** the population as: *the probability distribution of IQ scores*



* IQ tests are designed so the average IQ = 100 + the SD = 15 + the distribution of scores is normal
* These values = **population parameters** (characteristics of the entire population)
* Select 100 + 1K people at random from this normal population + administer an IQ test for an SRS

* Histogram is roughly right shape, but a very crude approximation to the true population distribution, + sample mean = fairly close to population mean but not identical.
* **Sample statistics** are properties of a sampled data set, + although they can be fairly similar to the *true* population values, they are NOT the same.
* **Sample statistics** = things you can calculate from a data set + **population parameters** = things you want to learn about.

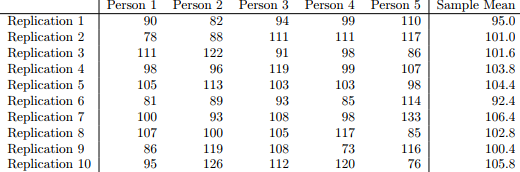
10.1 - Samples, populations and sampling

**10.2 The Law Of Large Numbers**

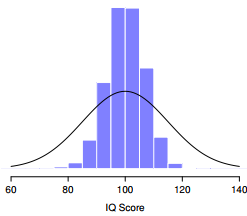
* If we want sample statistics to be much closer to population parameters = *collect more data.*
* A larger sample = much better approximation to true population distribution than a smaller one
* Also reflected in the sample statistics: mean +SD for a larger sample turns are closer to the true population’s
* Large samples generally give you better information (so obvious it shouldn’t need to be said)
* Jacob Bernoulli – one of the founders of probability theory – formalized this idea back in 1713: “For even the most stupid of men, by some instinct of nature, by himself and without any instruction (which is a remarkable thing), is convinced that the more observations have been made, the less danger there is of wandering from one’s goal”
* More data will give better answers. But why is this so?
* **The Law Of Large Numbers** = mathematical law that applies to many different sample statistics, but the simplest way to think about it is as a *law about averages*.
* Sample mean = most obvious example of a statistic that relies on averaging (obviously)
* When applied to the sample mean, the law of large numbers states “As the sample gets larger, sample mean tends to get closer to the true population mean”
* Or, to say it a little bit more precisely “as the sample size “approaches” infinity ( N 🡪 ∞) the sample mean approaches the population mean (X¯ 🡪 µ).
* Technically, the law of large numbers pertains to any sample statistic that can be described as an *average of independent quantities*.
* It’s also possible to write many other sample statistics as averages of 1 form or another.
* Variance of a sample can be rewritten as a kind of average + so is subject to the law of large numbers
* The minimum value of a sample, however, CANNOT be written as an average of *anything* + is therefore not governed by the law of large numbers
* This is 1 of the most important tools for statistical theory.
* The law of large numbers is the thing we can use to justify our belief that collecting more + more data eventually leads us to the truth.
* For any particular data set, the sample statistics calculated from it will be wrong, but the law of large numbers says if we keep collecting more data, they will tend to get closer + closer to the true population parameters.

**10.3 Sampling Distributions And The Central Limit Theorem**

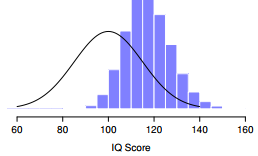
* The law of large numbers is a very powerful tool, but it’s good enough to answer all our questions.
* Gives us a *long run guarantee* =In the long run, if we were somehow able to collect an infinite amount of data, law of large numbers guarantees our sample statistics will be *correct*.
* John Maynard Keynes famously argued in economics that a long run guarantee is of little use in real life 🡪 “a misleading guide to current affairs. In the long run we are all dead. Economists set themselves too easy, too useless a task, if in tempestuous seasons they can only tell us, that when the storm is long past, the ocean is flat again”
* Not enough to know we will eventually arrive at a right answer when calculating a sample mean.
* Knowing an infinitely large data tells us the exact value of a population mean is cold comfort when actual data set has a sample size n = 100.
* In real life, we must know something about the *behavior* of the sample mean when calculated from a more modest data set
* Consider a very modest experiment w/ n = 5 people + measure IQ scores.
* Now replicate the experiment/repeat the procedure as closely as possible = randomly sample 5 new people + measure their IQ + do this 10 times

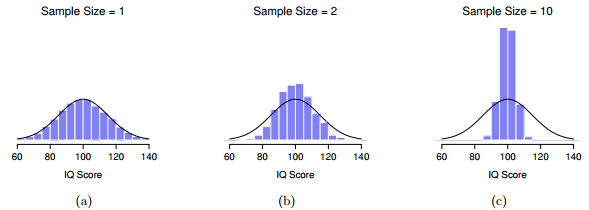


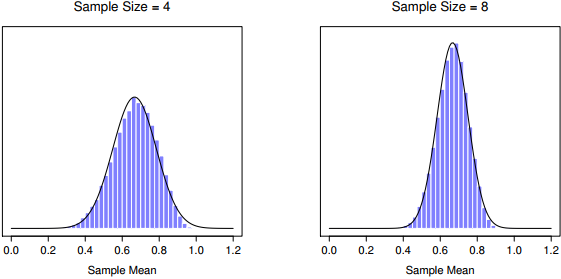
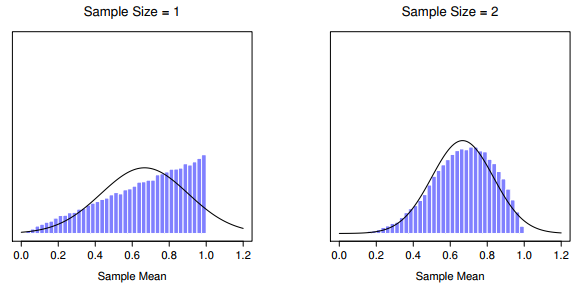
* Every time I replicate the experiment, over time, I’d be amassing a new data set, in which *every experiment generates a single data point*.



* The average of 5 IQ scores is usually between 90-110, but more importantly, if we replicate an experiment over + over again, we end up w/ a distribution of sample means **= the sampling distribution** of the mean.
* **Sampling distributions** are another important theoretical idea in statistics + are crucial for understanding behavior of small samples
* If I repeat the experiment, the sampling distribution tells me that I can expect to see a sample mean anywhere between 80-120
* 1 thing to keep in mind when thinking about sampling distributions is that *any sample statistic you might care to calculate has a sampling distribution*.
* Suppose I wrote down the largest IQ score in the experiment instead of the mean



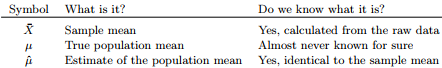
* Doing this over + over again gives a very different sampling distribution = **sampling distribution of the maximum**.
* Not surprisingly, if you pick 5 random people + then find the person w/ the highest IQ score, they’re going to have an above average IQ, typically between 100-140
* Most of the time you’ll end up with someone whose IQ is measured in the 100 to 140 range
* The sampling distribution of the mean changes as a function of sample size.
* Intuitively, if you only have a few observations, the sample mean is likely to be quite inaccurate: if you replicate a small experiment + recalculate the mean you’ll get a very different answer.
* In other words, *the sampling distribution is quite wide*.
* If you replicate a large experiment + recalculate the sample mean you’ll probably get the same answer you got last time, so the *sampling distribution will be very narrow*.
* 
* The bigger the sample size of 10,000 samples, the narrower the sampling distribution gets + more closely clustered around the true population mean
* We can *quantify* this effect by calculating the SD of the sampling distribution = the **standard error**.
* Since we’re usually interested in the standard error of the sample mean, we often use **SEM**.
* *What if the population distribution isn’t normal?*
* **No matter what shape a population distribution is, as N increases, the sampling distribution of the mean starts to look more like a normal distribution.**
* Started w/ a “ramped” distribution + comparing the triangular-shaped histogram to the bell curve plotted (black line), the population distribution doesn’t look very normal at all.



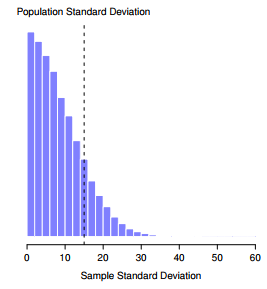
* BY the time we reach a sample size of n = 8, the sampling distribution is almost perfectly normal.
* In other words, as long as your sample size isn’t tiny, the sampling distribution of the mean will be approximately normal no matter what your population distribution looks like!
* It seems like we have evidence for all the following claims about sampling distribution of the mean:
* Mean of the sampling distribution is the same as the mean of the population
* SD of the sampling distribution (i.e., the standard error) gets smaller as the sample size increases
* Shape of a sampling distribution becomes normal as the sample size increases
* Not only are all these true, the **central limit theorem (CLT)** proves all 3 of them
* It tells us if the population distribution has mean = µ + SD = σ, the sampling distribution of the mean also has mean = µ, + SEM = σ/n
* B/c we divide the population SD σ by the square root of the sample size n, SEM gets smaller as sample size increases.
* It also tells us the shape of the sampling distribution becomes normal as n increases
* The CLT is a bit more general than this section implies.
* This is 1situation where CLT holds: when taking an average across lots of independent events drawn from the same distribution.
* CLT is much broader than this 🡪 a whole class of things called **U-statistics**, all of which satisfy CLT + therefore become normally distributed for large sample sizes.
* The mean is 1 such statistic, but is not the only one.
* This is useful for all sorts of things: tells us why large experiments are more reliable than small ones, + b/c it gives an explicit formula for SE, it tells us how much more *reliable* a large experiment is, tells us why the normal distribution is normal.
* In real experiments, many things we want to measure are actually averages of lots of different quantities (arguably, *general* intelligence as measured by IQ = average of a large number of *specific* skills + abilities), + when that happens, the averaged quantity should follow a normal distribution
* B/c of this mathematical law, the normal distribution pops up over + over again in real data.

**10.4 Estimating Population Parameters**

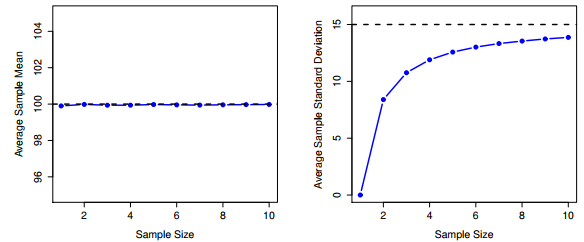
* How do we know that IQ scores have a true population mean = 100? (b/c people who designed the tests administered them to very large samples + rigged the scoring rules so their sample has mean = 100, an important part of designing a psychological measurement)
* Important to keep in mind this theoretical mean = 100 *only attaches to the population that the test designers used to design the tests*.
* Good test designers go to some lengths to provide **test norms** that can apply to *lots of different populations* (different age groups, nationalities, etc.).
* This is very handy, but of course almost every research project of interest involves looking at a *different* population of people to those used in the test norms.
* Suppose you wanted to measure the effect of low-level lead poisoning on cognitive functioning in Port Pirie, an industrial town w/ a lead smelter + want to compare IQ scores among people in Port Pirie to a comparable sample in Whyalla, an industrial town w/ a steel refinery.
* We’re going to have to estimate the population parameters from a sample of data since there’s no sensible norming data that can automatically be applied to South Australian industrial towns
* Suppose we go to Port Pirie + 100 of the locals are kind enough to sit through an IQ test w/ X¯ = 98.5
* What is the true mean IQ for the entire population of Port Pirie? Obviously, we don’t know that
* Our sampling isn’t exhaustive so we cannot give a definitive answer.
* 98.5 could be “giving a best guess” = the essence of **statistical estimation**
* **\*\*\*** need to be a lot more careful if really interested in this 🡪 can’t just compare IQ scores in Whyalla to Port Pirie + assume any differences are due to lead poisoning.
* Even if it were true, need to account for the fact that people already *believe lead pollution causes cognitive deficits* = different **demand effects** for the PP sample than the Whyalla sample
* Might end up w/ an **illusory group difference** in your data, caused by the fact that people *think* there is a real difference.
* Estimating the unknown population parameter is straightforward: Calculate the sample mean + use it as an estimate of the population mean
* Make sure you recognize the sample statistic + estimate of the population parameter are conceptually different things.
* A sample statistic = a *description* of *your data*, the estimate = a *guess* about the *population*.
* If true population mean is denoted µ, use µ^ to refer to population mean estimate
* Sample mean is denoted X¯ or sometimes m.
* However, in SRS, the estimate of the population mean is *identical* to the sample mean: if I observe a sample mean of X¯ = 98.5, my *estimate* of the population mean is also µ^ = 98.5.



* So far, estimation seems pretty simple
* In the case of the mean, estimate of the population parameter µ^ turned out to identical to the corresponding sample statistic X¯.
* However, that’s not always true.
* Think about how to construct an estimate of the population SD, σ^.
* Suppose I have a sample that contains a *single observation* = 20where you have no intuitions at all about what the true population values might be, and so it has a sample mean = 20
* B/c every observation in this sample is obviously = the sample mean, it has a sample SD = 0
* sample contains a single observation + therefore there is no variation observed w/in the sample.
* A sample SD of s = 0 is right, but as an estimate of the population SD, it feels completely insane
* The only reason we don’t see any variability in the sample is the sample is too small to display any
* If forced to make a best guess about population *mean*, it doesn’t feel completely insane to guess the population mean = 20 (wouldn’t feel very confident b/c we have 1 observation to work w/, but it’s still the best guess you can make)
* Suppose I now make a 2nd observation, now n = 2 + the complete sample = {20, 22}
* This time around, our sample is *just* large enough for us to be able to observe some variability: 2 observations is the bare minimum number needed for any variability to be observed
* For our new data set, sample mean X¯ = 21, + the sample SD s = 1.
* Again, as far as the population mean goes, best guess we can make is the sample mean = 21.
* SD is a little more complicated as the sample SD is only based on 2 observations, +, w/ only 2 observations, we haven’t given the population enough of a chance to reveal its true variability.
* Not just that we suspect the estimate is wrong (w/ only 2 observations we expect it to be wrong to some degree), the worry is that the error is **systematic**.
* Specifically, we suspect the sample SD is likely to be smaller than the population SD.
* This intuition feels right, but it would be nice to demonstrate this somehow.
* There are in proofs that confirm this intuition



* Even though the true population SD is 15, the average of the sample SDs is only 8.5, very different the sampling distribution of the mean
* Repeat for sample sizes from 1-10 + plot the average sample mean + SD as a function of sample size



* On average, average sample mean = population mean, an **unbiased estimator** = essentially the reason why your best estimate for the population mean is the sample mean
* Sample SD s is *smaller than the population SD σ* = a **biased** **estimator**.
* In other words, if we want to make a best guess σ^ about the value of the population SD σ, we should make sure our guess is a *little bit larger than the sample SD s*.
* **Unbiasedness** is a desirable characteristic for an estimator
* But there are other things that matter besides bias
* The fix to this systematic:
* 1st look at the variance (average of the squared deviations from the sample mean)



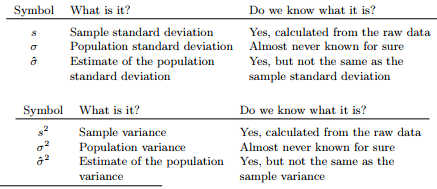
* Sample variance s^2 is a *biased estimator* of the population variance σ^2 .
* Only need to make a tiny tweak to transform this into an *unbiased* estimator 🡪 divide by N - 1 rather than by N.



* This is an unbiased estimator of the population variance σ
* Similar story applies for SD 🡪 divide by N - 1 rather than N = estimate of population SD becomes



* In a bizarre + counterintuitive twist, since ˆσ^2 is an unbiased estimator of σ^2, you’d assume taking the square root would be fine, + σ^ would be an unbiased estimator of σ.
* Weirdly, it’s not b/c there’s actually a subtle, tiny bias in σ^
* ˆσ^2 is + an unbiased estimate of the population variance σ^2, but when you take the square root, it turns out that σ^ IS a biased estimator of the population SD σ.
* The technical reasoning is b/c *non-linear transformations (e.g., square root) don’t commute w/ expectation*
* Fortunately, the bias is small, + in real life everyone uses σ^ + it works just fine.
* In practice, a lot of people tend to refer to ˆσ^ as the “sample SD” + technically, this is incorrect
* Sample SD should be equal to s (divide by N).
* These aren’t the same thing, either conceptually or numerically.
* 1 is a *property* of the *sample*, the other is an *estimated* *characteristic* of the *population*.
* However, in almost every real-life application, *we actually care about the estimate of the population parameter*, + so people always report σ^ rather than s (which IS the right number to report)
* People tend to get a little bit imprecise about terminology b/c “sample SD” is shorter than “estimated population SD”.
* It’s important to keep the 2 concepts separate: it’s never a good idea to confuse *known properties* of your sample w/ *guesses* about the population from which it came
* The moment you start thinking that s + ˆσ are the same thing, you start doing exactly that.



**10.5 Estimating A Confidence Interval**

* Statistics means never having to say you’re certain
* Every data set leaves us w/ some of uncertainty = estimates are never going to be perfectly accurate
* Want to *quantify* the amount of uncertainty attached to estimates/express the **degree of certainty** in a guess = **Confidence Interval** for a mean.
* We know from the CLT that a sampling distribution of the mean is approximately normal + from the normal distribution, there is a 95% chance a normally-distributed quantity, X¯, will fall w/in 2 SDs of the true mean.
* The 25th + 97.5th percentiles of the normal distribution are **{-1.959964, 1.959964}** = SD’s that X¯ falls between
* More correct answer = 95% chance a normally-distributed quantity falls w/in 1.96 SDs of true mean
* Recall that SD of a sampling distribution = the **standard error**, + **standard error of the mean = SEM.**
* When we put all these pieces together, there is a 95% probability a sample mean X¯ observed lies w/in 1.96 SE’s of the population mean.

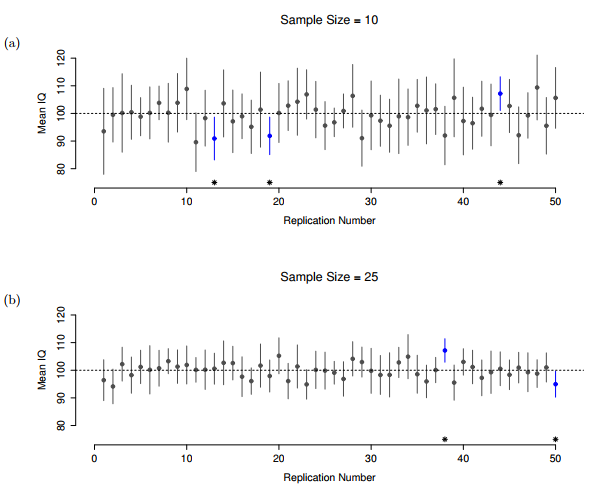
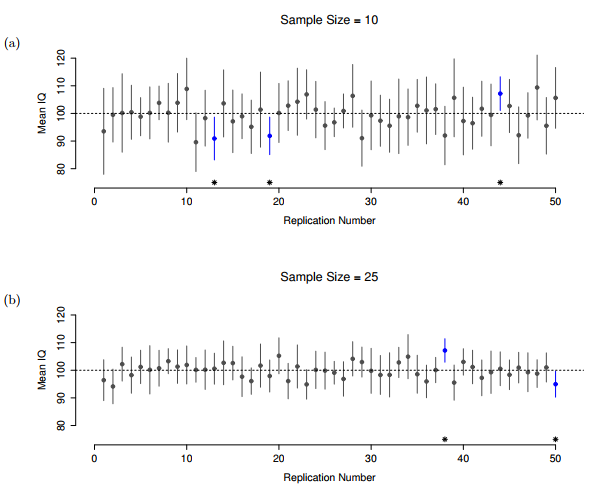


* where SEM = σ/n
* We can be 95% confident this is true, however, that’s not answering the question we’re actually interested in.
* The equation above tells us *what we should expect about the sample mean, given we know what the population parameters are*.
* What we want is to have this work the *other way around* 🡺 we want to know what we should believe about population parameters, given we have observed a particular sample.



* This says that this range of values has a 95% probability of containing the population mean µ = a **95% confidence interval**, denoted **CI(95)**.
* In short, as long as n is large enough for us to believe the sampling distribution of the mean is normal, we can write this as our formula for the 95% CI



* Nothing special about the number 1.96 = just happens to be the multiplier you use if you want a 95% CI.
* If I’d wanted a 70% CI, use **qnorm()**to calculate the 15th + 85th quantiles = {-1.036433, 1.036433}
* CI(70) would be the same as CI(95) except we’d use 1.04 as our magic number rather than 1.96.
* The formula given above for the 95% CI is approximately correct, notice it requires you to use the SEM, which in turn requires you to use the *true population SD σ*.
* Yet, we don’t actually know the true population parameters
* *B/c we don’t know the true value of σ, we have to use an estimate of the population SD ˆσ instead.*
* This is pretty straightforward to do, but has the consequence that we need to use the quantiles of the t-distribution rather than the normal distribution to calculate the multiplier, + the answer depends on sample size.
* When n is very large, we get pretty much the same value from both distributions, but when n is small, we get a much bigger number when we use the t distribution
* Bigger values mean the CI is wider, indicating we’re more uncertain about the true value of µ
* When we use the t-distribution instead of the normal distribution, we get bigger numbers, indicating we have more uncertainty
* We have that extra uncertainty b/c *our estimate of the population SD ˆσ might be wrong*
* If it’s wrong, it implies we’re a bit less sure about what our sampling distribution of the mean actually looks like + this uncertainty ends up getting reflected in a wider CI
* The hardest thing about CIs is understanding what they mean.
* Whenever people first encounter CIs, 1st instinct is almost always to say “there is a 95% probability the true mean lies inside the CI”.
* This intuitive definition relies very heavily on personal beliefs about the value of the population mean
* I say “I am 95% confident” b/c those are my *beliefs,* + talking about personal belief + confidence is a Bayesian idea.
* However, *CIs are NOT Bayesian tools*, but are frequentist tools,
* If using frequentist methods, it’s not appropriate to attach a Bayesian interpretation to them.
* Remember w/ frequentist probability, the only way we’re allowed to make “probability statements” is to talk about a *sequence of events*, + to *count up the frequencies of different kinds of events.*
* From that perspective, the interpretation of a 95% CI must have something to do w/ replication.
* Specifically: if we replicated the experiment over + over again + computed a 95% CI for each replication, then 95% of those intervals would contain the true mean.
* More generally, 95% of all CI’s constructed using this procedure should contain the true population mean.
* 
* See 50 CIs constructed for a 10 measured IQ scores experiments in the top + another 50 CIs for 25 measured 25 IQ scores experiments (bottom)
* Across the 100 replications, it turned out that exactly 95 of them contained the true mean.
* The critical difference here is that the *Bayesian* *claim* makes a probability statement about the population mean (refers to our uncertainty about the population mean), which is NOT allowed under the frequentist interpretation of probability b/c *you can’t “replicate” a population!*
* In the frequentist claim, the population mean is *fixed* + NO probabilistic claims can be made about it.
* CIs, however, ARE repeatable, so we CAN replicate experiments.
* Therefore, a frequentist is allowed to talk about *the probability that a CI (a random variable) contains the true mean, but is NOT allowed to talk about the probability the true population mean (NOT a repeatable event) falls w/in the CI.*
* **“We have obtained the CI using a procedure which, for 95% of all SRS’s, of the given size, produces an interval containing the parameter we are estimating”**
* Seems a little pedantic, but it does matter b/c the difference in interpretation leads to a difference in the mathematics.
* There is a Bayesian alternative to CIs, known as **credible intervals** + which in most situations are quite similar to CIs, but in other cases are drastically different.

