MIT 14.32/14.320 **Spring 2023**
J. Angrist

Lecture Note 2

Sampling Distributions and Statistical Inference

1 Populations and Samples

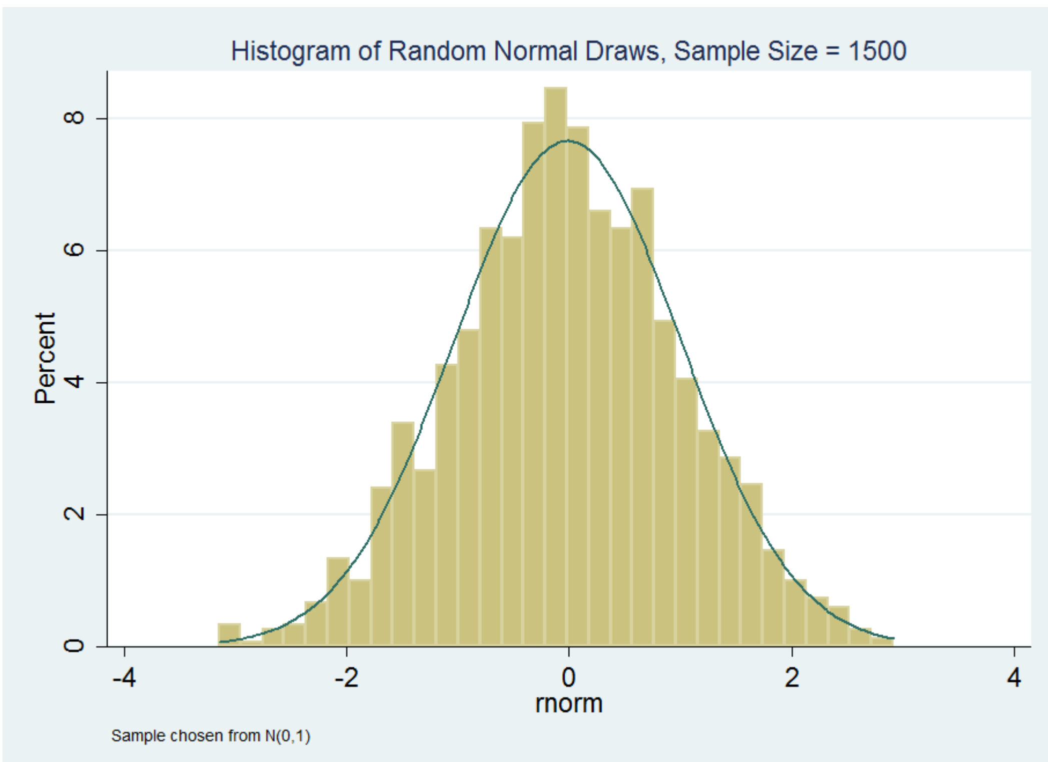
1.1 Big Ideas

We're interested in features of a population, like the population of economics majors (Are there really fewer women than men studying this awesome subject? We use statistical methods to evaluate the evidence)

- Moments and functions of moments – mean, variance, covariance, regression coefficients of various kinds – characterize populations of interest

We learn about population parameters by drawing samples. The most interesting samples are drawn from real data. But sometimes we sample from computer-generated theoretical distributions, just to see what the resulting samples look like.

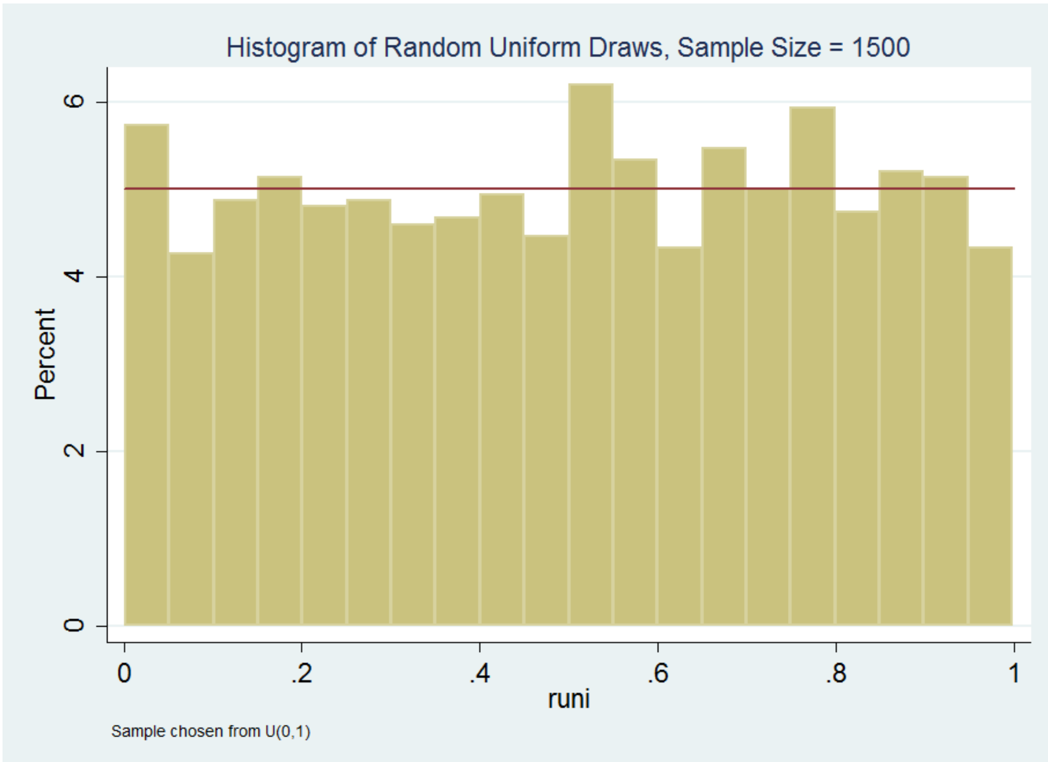
- Sampling from a Normal distribution (density and histogram below)



Q. What's this distribution a good model for?

- The Normal distribution is a *thing*: it's got a formula, it's *parametric*.
  - We write: $X_i \sim N(\mu, \sigma^2)$. What are the parameters of a Normal distribution?

- Sampling from a Uniform distribution defined over $[0, 1]$ (density and histogram below)

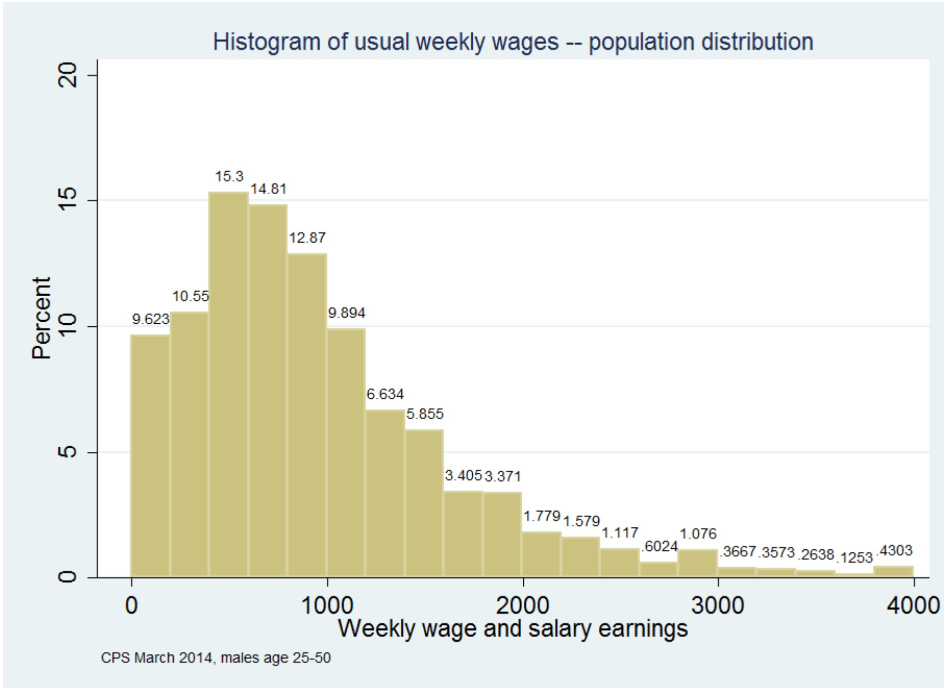


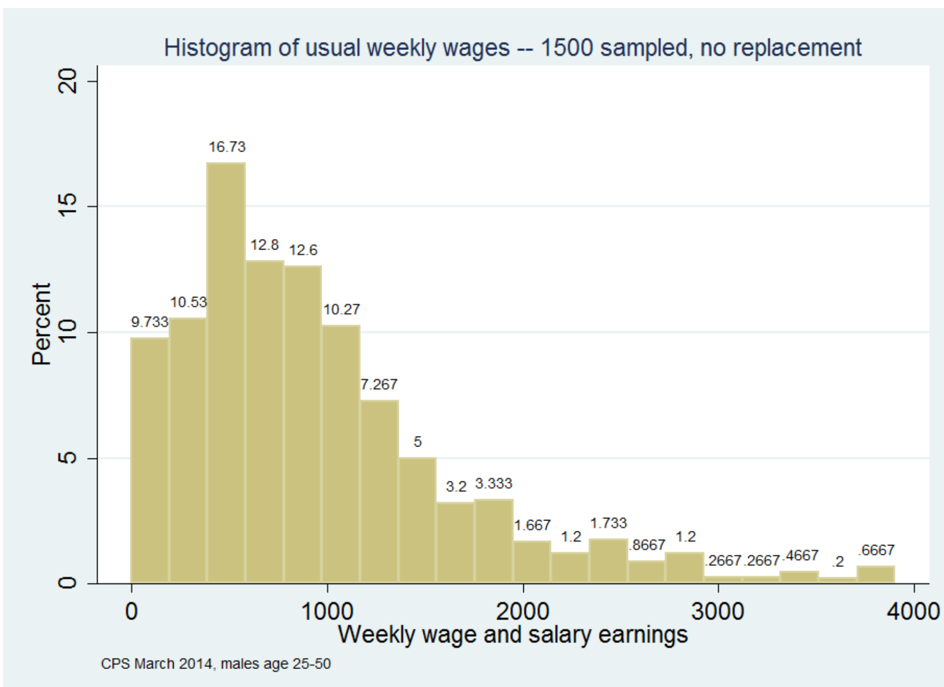
Q. What's a uniform distribution a good model for?
Q. Why isn't the histogram pictured here perfectly flat?

- A uniform distribution is parametric - what are its parameters?

- Other oft-seen parametric distributions: Dummy variables for things switched on or off (also called a Bernoulli distribution); Binomial and Poisson for count data; exponential and log-Normal for continuous non-negative r.v.s; multinomial for categorical data

- We sometimes explore statistical properties by sampling real-world distributions. Our m.o. here is to treat a large sample as a notional "population," and then draw smaller samples from this. The larger sample defines a realistic *empirical distribution*.
  - Consider the Current Population Survey (CPS), a large sample of US households (roughly 60,000 interviewed every month). With CPS samples playing the "population," I've drawn a much smaller sample of 1500.
  - Here's population and sample data on the usual weekly wage, which is computed by dividing annual earnings in a calendar year by annual weeks worked:

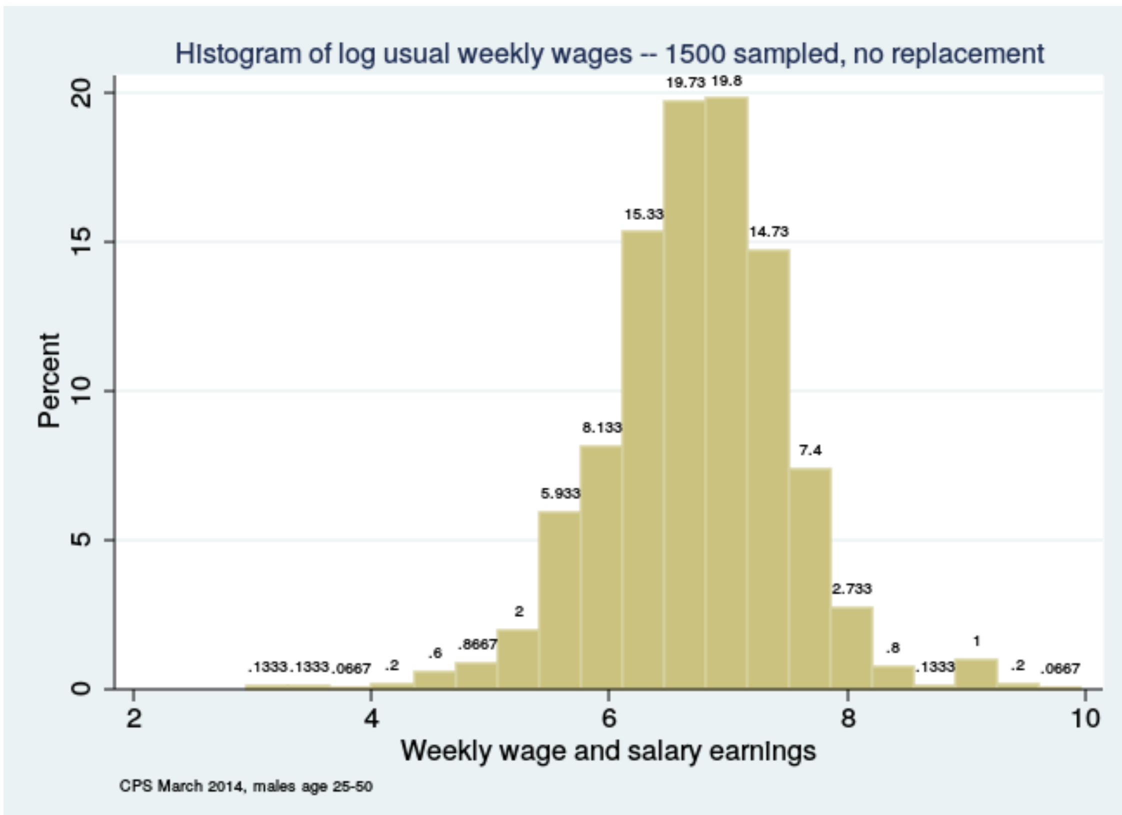




A few stats:

	N	Mean	s.d	Min	Max
UWW (“population”)	53,455	939.81	690.05	0	3996.6
UWW (random sample)	1500	943.96	714.68	0	3904.4

- The CPS is used to track and study the US economy. The U.S. Bureau of Labor Statistics uses the CPS to compute the unemployment rate. Each month's CPS data set is a sample, but we can sample from the CPS as if it's a population (for teaching purposes, or to get a smaller and more manageable data set, or as part of a *sampling experiment*, in which we study sampling distributions of sample statistics and econometric estimators).
- We distinguish random sampling from the construction of an *extract*. Often, we're interested in a particular population group, say nonwhite female college graduates. To study this population, we'll take as many sample observations in this category as we can get. A particular subsample of interest constitutes an extract.
- BTW, wages love to be logged:



- Q. Why are logs useful for this kind of data? (We'll log lots of time with logs later.)

2 Sampling Distributions

2.1 The Logic of Statistical Inference

1. Draw a sample from the population of interest
2. Use the this sample to derive conclusions about the population from which the sample was drawn. These conclusions come in two forms:
   (a) Evaluate (test) hypotheses about population parameters (are men and women equally likely to major in economics?)
   (b) Measure (estimate) population parameters as best we can, while quantifying the statistical uncertainty inherent in these estimates (what's the gender gap in econ major rates?)
3. Execute (a) and (b) by quantifying *sampling variance*. Sampling variance summarizes the randomness in *sampling distributions*, that is, distributions of sample statistics obtained in repeated samples.

- Statistical inference differs from causal inference
- Statistically significant differences need not reflect causal effects (though sometimes they do)

2.1.1 Expectation and Sampling Variance of the Sample Mean

Draw a random sample $\{X_i; i = 1, ..., n\}$. *Random sampling* means the observations are independent. For example, if we're sampling Course 14 majors and recording their gender (coded as a Bernoulli or dummy random variable), the fact that the first observation is female doesn't change the probability that the second is female.

Write the sample mean as:

$$\bar{X} = \frac{1}{n}\sum_i X_i.$$

Though we sample only once, the statistical behavior of $\bar{X}$ in repeated samples – it's *sampling distribution* – is easy to describe.

The expectation and sampling variance of the sample mean are:

$$E[\bar{X}] = \mu_X = E[X_i] \tag{1}$$

$$V(\bar{X}) = \frac{\sigma_X^2}{n} = \frac{V(X_i)}{n} \tag{2}$$

Prove it!

- Equation (1) says that the sample mean is an *unbiased estimator* of the population mean. Equation (2) says that variance of a sample mean declines with sample size at rate $\frac{1}{n}$.
- The ratio $\frac{\sigma_X}{\sqrt{n}}$ is called the *standard error of the sample mean*
- The *standard error* of $\bar{X}$ is distinct from the *standard deviation* of the underlying X_i, though they're clearly related
  - SE $(\frac{\sigma_X^2}{n})$ measures the statistical *precision* of $\bar{X}$, and declines with n. SD (σ_X) measures the dispersion of X_i and is a feature of the distribution of X_i.
  - Q. Which of the standard dev and the standard error decline as a function of sample size?

2.1.2 The Law of Large Numbers

The LLN says that the probability that the sample mean is close to the corresponding population mean approaches one as the sample size increases. How close? As close as you like! The odds that the sample mean is very close to the population mean are high in large samples. How large? Hard to say, but we often assume we have enough data for the LLN to be relevant.

Write $\bar{X}_n$ for the sample mean in a random sample of size n. Sample size is particularly important in this context, so we keep the n subscript. The LLN says:

$$\lim_{n\to\infty} P\left\{\left|\bar{X}_n - \mu_X\right| \geq \epsilon\right\} = 0,$$

for any number, ϵ, no matter how small. Equivalently, we write:

$$\underset{n\to\infty}{plim}\, \bar{X}_n = \mu_X,$$

where *plim* denotes a probability limit. Casino owners and insurance companies rely on the LLN. Why is the LLN reliable? Because it's a theorem. Pset 1 asks you to prove *Chebyshev's inequality*, which generates the LLN as a consequence.

2.1.3 Sampling Distributions: Two Ways

We know the expectation of $\bar{X}$ and its variance. We've also seen how precision grows as sample size increases. Still, we strive for more: we want to know the sampling distribution of $\bar{X}$ (and, later, other statistics) not just the mean and standard error.

Sampling distributions are done in two frameworks:

1. Normal distribution theory. When the data are Normally distributed:
   - the sample mean is unbiased and Normally distributed
   - the (appropriately-scaled) sample variance has a chi-square distribution
   - the ratio of the centered sample mean to the estimated standard error of the mean has a t distribution
   - appropriately scaled ratios of sample variances from the same population have an F distribution
2. Asymptotic distribution theory. When the data are distributed according to almost any distribution, then, in large samples:
   - sample moments are likely to be close to the corresponding population moments (the LLN)
   - the ratio of the centered sample mean to the estimated standard error of the mean is *approximately* distributed standard Normal
   - appropriately scaled ratios of sample variances have a chi-square distribution

- Asymptotic distribution theory relies on the *central limit theorem* (CLT) as well as the LLN.

3 Normal Distribution Theory

The Normal distribution is a symmetric, bell-shaped distribution that offers a good a model for data that are, well, kinda normal. In practice, our data are often abnormal, but this matters less than you might think. Since many theoretical results are easily derived when the data are Normally distributed, Normal distribution theory provides a valuable reference point for the more general *asymptotic* (large-sample) framework that doesn't rely on having Normally distributed data.

3.1 The Normal Distribution

- When X is Normal, we write $X \sim N(\mu_X, \sigma_X^2)$

$$Z = (X - \mu_X)/\sigma_X \sim N(0,1) \qquad \text{Standard Normal r.v.}$$

$$P(Z \leq z) = \Phi(z) =_{-\infty} \int^z \phi(t)dt \qquad \text{Std Normal } cdf$$

$$\text{where } \phi(t) = (1/\sqrt{2\pi})exp\{-t^2/2\} \qquad \text{Std Normal } density$$

$\Phi(z)$ is tabulated in books and by computers

3.2 Standardizing the Sample Mean

- Because linear combinations of Normal random variables are themselves Normally distributed, the transformed variable,

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}}, \tag{3}$$

has a *standard Normal distribution*. We write $Z \sim N(0,1)$

- In transforming any random variable by subtracting its mean and dividing by its standard deviation, we are said to *standardize* it

3.3 t-statistics

- Sample means standardized according to (3) are standard Normal. Alas, σ_X is unknown and must be estimated. Let Let $s_X^2 = \frac{\sum_i (X_i - \bar{X})^2}{n-1}$. (Note that here, s_X^2 is the unbiased sample variance estimator). We have

$$\begin{aligned} T_n &= [\bar{X} - \mu_X]/[\sigma_X/\sqrt{n}] \div (\sum_i (X_i - \bar{X})^2/\sigma_X^2(n-1))^{1/2} \\ &= \frac{\bar{X} - \mu_X}{s_X/\sqrt{n}} \\ &\sim t(n-1) \end{aligned}$$

because T_n has a t distribution, it's called a t-statistic. This result uses the fact that with Normal data, Z is standard Normal, $W = \sum_i (X_i - \bar{X})^2/\sigma_X^2 \sim \chi^2(n-1)$, independent of Z, and so $Z/[W/n-1]^{\frac{1}{2}} \sim t(n-1)$

- We refer to any econometric estimate divided by its standard error as a "t-statistic," regardless of whether the underlying data are Normal

4 An Asymptopian Alternative: The CLT on Toast

4.1 Sampling under Normality

Suppose your data (X_i) are Normally distributed, so that the statistic $T_n \sim t(n-1)$ in a sample of size n. Then,

$$\lim_{n \to \infty} P(T_n \leq c) = \Phi(c),$$

where $\Phi(c)$ is the standard Normal cdf evaluated at c. With large samples from a Normal distribution, standard Normal tables can be used for t-tests (entries in a t-table for $t(\infty)$ coincide with those for standard Normal).

4.2 Non-Normal Data

The LLN applies to all sample moments, whether or not the underlying data are Normal. The LLN is therefore reassuring (its the law!), though not enough for statistical inference. The *Central Limit Theorem* (CLT) is our ace in the hole.

- Suppose that $\bar{X}_n$ and s_X are the sample mean and standard deviation from an i.i.d. sample of n observations on random variable X_i – not Normally distributed – with mean μ_X. It remains true that

$$\lim_{n\to\infty} P(T_n \leq c) = \Phi(c),$$

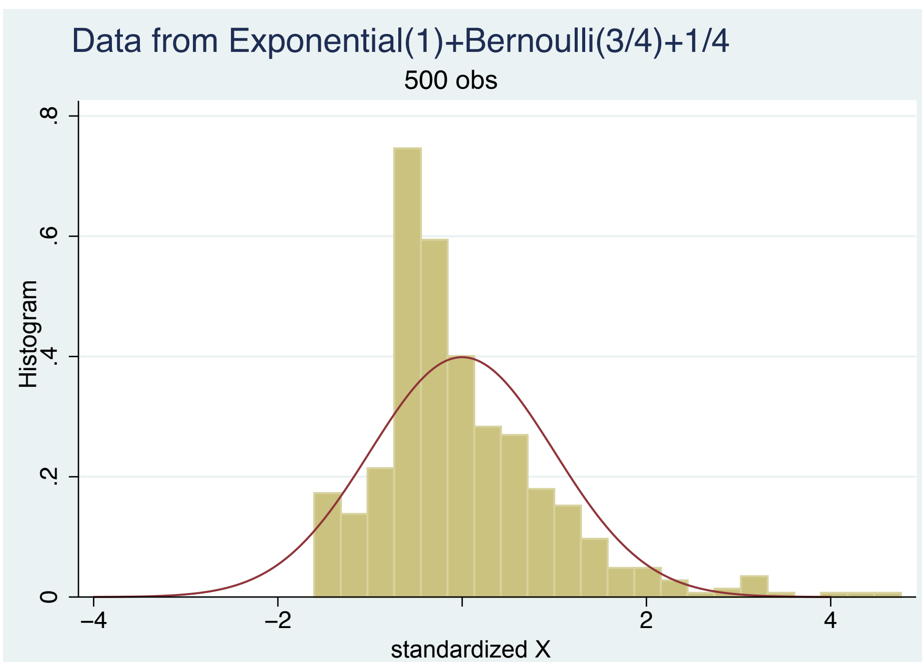
so the limiting (asymptotic) distribution of a t-statistic is still standard Normal.

- This remarkable fact, called the *Central Limit Theorem,* implies that appropriately standardized sample moments (and functions thereof, TBD) are approximately Normally distributed.
- Recap:
  - $E(\bar{X}_n) = \mu_X$, regardless of sample size or the distribution of the underlying data
  - The variance of $\bar{X}_n$ is $\frac{\sigma_X^2}{n}$, regardless of sample size or the distribution of the underlying data
  - The LLN promises that, in large samples, it's very likely that $\bar{X}_n$ is close to μ_X
  - The CLT says that if the sample is large enough, the distribution of T_n should be close to standard Normal, no matter the underlying data distribution
- With few exceptions, the CLT applies reliably to sample moments and functions of sample moments, including differences in means, regression coefficients, and the other econometric estimators we'll meet later in term. As a rule, the CLT says:

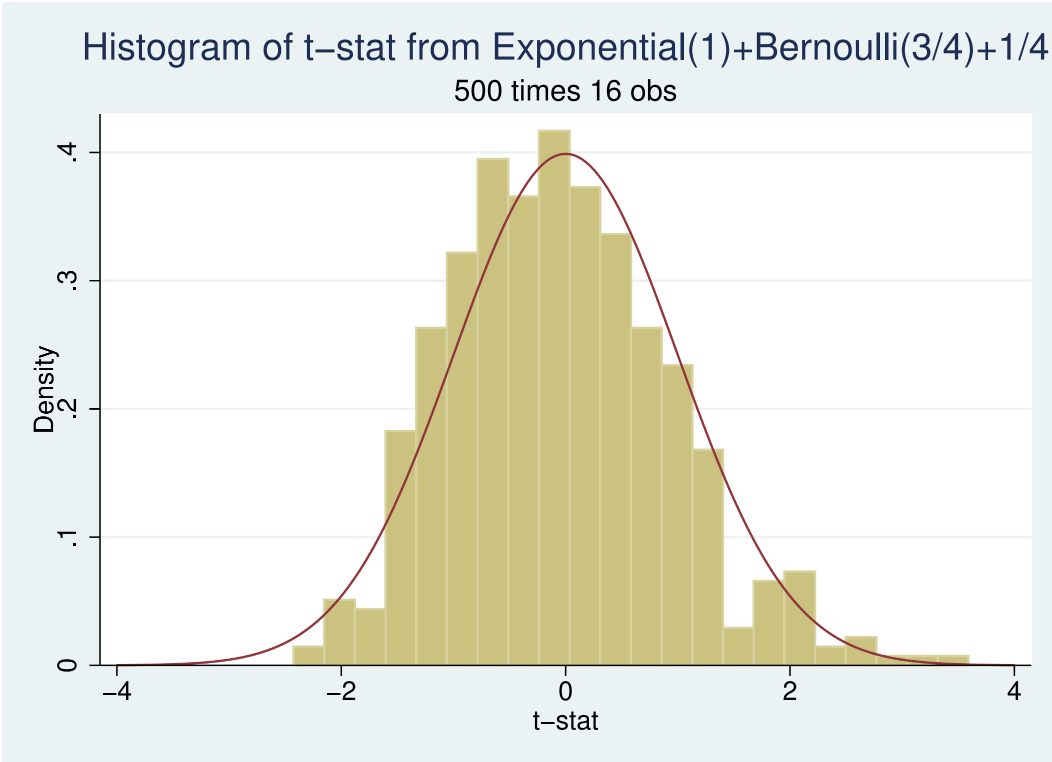
$$\frac{estimate - plim(estimate)}{SE(estimate)} \sim_a N(0,1),$$

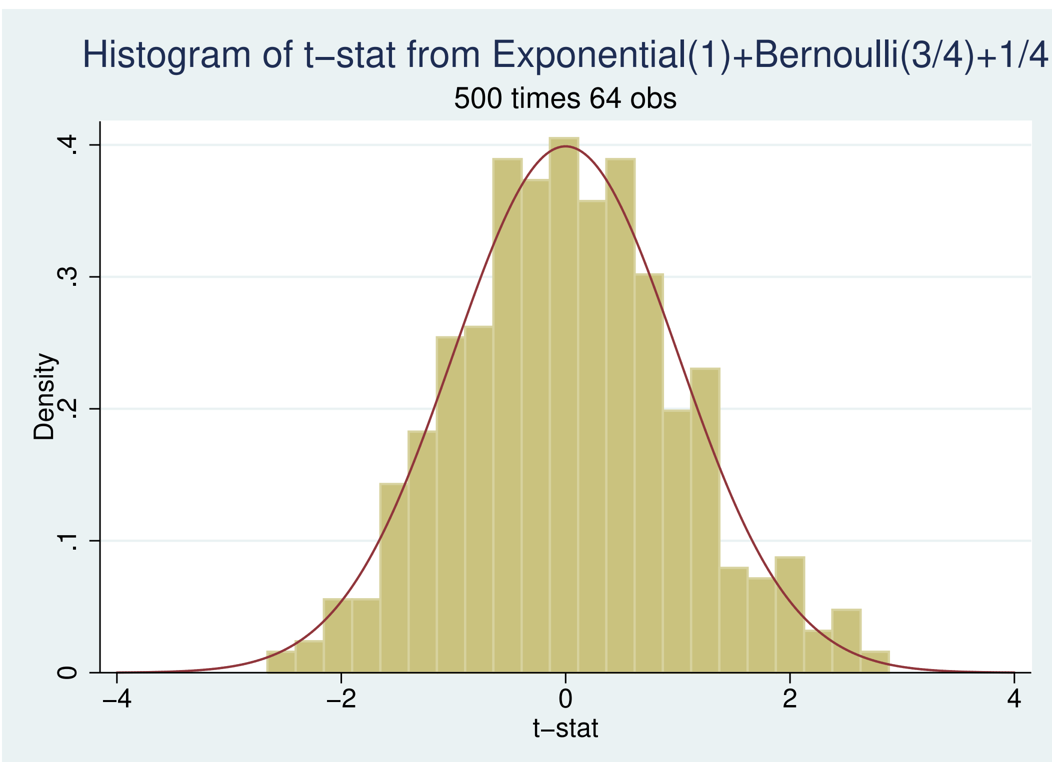
where $\sim_a$ means "approximately" or "asymptotically" distributed.

  - Masters of 'metrics mostly live and work in asymptopia
- Monte Carlo evidence for the CLT



– Sampling from an ab-Normal distribution





5 Testing with t

- We may want to test:

$$\begin{aligned} H_0: &\quad \mu_X = \mu_0 \\ H_1: &\quad \mu \neq \mu_0 \end{aligned}$$

- If $\mu_X = \mu_0$, then

$$T_n = \frac{\bar{X} - \mu_0}{s_X/\sqrt{n}} \sim t(n-1)$$

  If T_n is surprisingly large, given what we expect from a $t(n-1)$ distribution, we reject H_0. Of course, big t-stats just happen sometimes: we might wrongly reject.

- *Type I error*: reject a true null. The probability of Type I error is also called the *significance level or size* of a test. Sometimes this is denoted by α.

- *Type II error*: accept a false null. The probability of a Type II error is sometimes denoted β. *Statistical power* is $1 - \beta$. Low-powered tests rarely reject–even when the null is false (low power is the bane of an empiricist's existence)

Remember the two types of error by thinking about a criminal courtroom. In US jurisprudence, the null hypothesis is innocence: you're "innocent until proven guilty." Prosecutorial evidence weighs against innocence, perhaps causing the jury to reject the null and convict. But sometimes the evidence is misleading and juries wrongly convict. This is a Type I error. On the other hand, in the face of evidence of guilt that's too weak to banish reasonable doubt, juries fail to convict, even when the defendant is guilty. Letting a guilty man go free is a Type II error.

We face a trade-off between the the two types of errors, that is between significance and power. It's standard practice among empiricists to choose test size (often 5%) and hope for high power given this choice. Other testing approaches adjust the probability of the two types of error as a function of sample size.

In practice, we don't often follow orthodox testing theory, but rather use elements of the hypothesis testing machinery informally as suits our empirical agenda. For example, the t-ratio for a null hypothesis of zero divides the estimate of interest by its standard error. The asymptotic (Normal) critical value for a 5% test is 1.96.

- Estimates that are more than double their standard errors are often described as statistically significantly different from zero, meaning the fact that such an estimate differs from zero is unlikely to be due solely to chance.

- We don't fuss over whether the estimate is precisely double its standard error: estimates a little less than twice their standard error are said to be marginally significant, while those that are substantially more than twice their standard error support a decisive rejection of the null hypothesis.

5.1 Comparing Means

Everything interesting is relative. We're into comparisons in a big way, comparing treatment and control groups in RCTs, for instance. We'd like to know whether samples from two populations weigh against the null hypothesis that treatment and control populations are the same, in which case, we say that there's a nonzero *treatment effect*. But, some comparisons are merely descriptive. We might, for instance, compare men and women to see if they differ in some dimension of interest, such as the propensity to major in Economics.

- Given

  Sample 1: X_{1i} for $i = 1, ..., n_1$ $\quad$ $\bar{X}_1$ is the sample mean $\quad$ μ_1 is the population mean

	s_1^2 is the sample variance	σ_1^2 is the population variance
Sample 2: X_{2i} for $i = 1, ..., n_2$	$\bar{X}_2$ is the sample mean	μ_2 is the population mean
	s_2^2 is the sample variance	σ_2^2 is the population variance

- We test

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 \\ H_1 &: \quad \mu_1 \neq \mu_2 \end{aligned}$$

  Assuming the data are Normal, we have:

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right])$$

  Be sure you can derive the formula for the sampling variance of a difference in sample means:

$$\left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right], \tag{4}$$

  The square root of this is the corresponding *standard error of a difference in sample means.*

- Under the null hypothesis, H_0:

$$(\bar{X}_1 - \bar{X}_2)/([\sigma_1^2/n_1] + [\sigma_2^2/n_2])^{1/2} \sim N(0, 1)$$

  Assuming also that $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

$$Z = (\bar{X}_1 - \bar{X}_2)/[\sigma(\frac{1}{n_1} + \frac{1}{n_2})^{1/2}] \sim N(0, 1)$$

  and

$$T_n = (\bar{X}_1 - \bar{X}_2)/[s(1/n_1 + 1/n_2)^{1/2}] \sim t(n_1 + n_2 - 2)$$

  where $s^2 = [\sum_i (X_{1i} - \bar{X}_1)^2 + \sum_j (X_{2j} - \bar{X}_2)^2]/(n_1 + n_2 - 2)$ is the pooled variance estimate.[1]

- With separate variance estimates, s_1^2 and s_2^2, the t-stat is:

$$T_n = (\bar{X}_1 - \bar{X}_2)/([s_1^2/n_1] + [s_2^2/n_2])^{\frac{1}{2}} \tag{5}$$

  Even with Normal data, when pop variances differ, the distribution of T_n is only approximately $t(n_1 + n_2 - 2)$. But we still call it a $t-statistic$.

  - T_n can be compared with critical values for the $t(n_1 + n_2 - 2)$ distribution or a standard Normal. In large samples, the distinction between standard Normal and t-distributions matters little.
  - From a large-sample, asymptopian point of view, we expect T_n to be standard Normal under the null hypothesis regardless of whether the underlying data are Normal.

- We noted above that the standard error of a difference in means is:

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{6}$$

  - In practice, variances are unknown. We work, therefore, with the corresponding *estimated standard error*:

$$([s_1^2/n_1] + [s_2^2/n_2])^{\frac{1}{2}}.$$

    giving the t-stat formula in (5)
  - In practice, estimated SEs are the only ones we ever see, so we often forget to distinguish estimated from theoretical standard errors, referring to the quantity above simply as the "standard error."

[1]This comes from $s^2 = \frac{1}{n_1+n_2-2}[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]$, where s_j^2 uses an $n_j - 1$ denominator.

5.2 Inference for Treatment Effects in Randomized Trials

- 'Metrics masters Angrist, Lang, and Oreopoulos (2009) were disturbed by rumors of inattentive, low-performing college students. They decided to see whether students who won't boost study effort for love (of learning), might do it for money
- ALO randomly allocated 1,656 full-time freshmen at a Toronto-area college to four groups, three treated and a control:
  - The SSP (250 students): treated with peer advising and the opportunity to attend facilitated study groups
  - The SFP (250 students): treated with the opportunity to win merit scholarships based on grades
  - The SFSP (150): treated with extra services and the opportunity to win merit scholarships
  - The controls consisted of everyone in the original 1,656 not randomly selected for a treatment
- Did cash incentives and/or enhanced services boost college achievement? For whom and by how much?
- Here's the award scheme:

Appendix

Student Fellowship Program Award Schedule

	Award amount		
High school GPA quartile	\$1,000	\$2,500	\$5,000
0–25th percentile	2.3 (C+)	2.7 (B−)	3.0 (B)
25th–50th percentile	2.7 (B−)	3.0 (B)	3.3 (B+)
50th–75th percentile	3.0 (B)	3.3 (B+)	3.7 (A−)

Notes: Eligibillity was determined by the student's best four courses. Half of SFP/SFSP participants were offered the opportunity to qualify for a \$2,500 award.

• Descriptive statistics and check for balance (compare means between treatment and control groups)

TABLE 1—DESCRIPTIVE STATISTICS

	Control mean (1)	Contrasts by treatment status: SSP v. control (2)	SFP v. control (3)	SFSP v. control (4)	F-stat (all=control) (5)	Obs. (6)
Administrative variables						
Courses enrolled as of fall 2005	4.745	−0.053	0.015	−0.158	0.702	1,656
	{1.370}	[0.095]	[0.095]	[0.118]	(0.551)	
No show	0.054	0.002	−0.030	0.020	1.852	1,656
		[0.016]	[0.016]*	[0.019]	(0.136)	
Completed survey	0.898	−0.018	−0.010	−0.051	1.228	1,656
		[0.022]	[0.022]	[0.028]*	(0.298)	
Student background variables						
Female	0.574	−0.006	0.029	−0.005	0.272	1,571
		[0.036]	[0.035]	[0.045]	(0.845)	
High school GPA	78.657	0.170	0.238	−0.018	0.276	1,571
	{4.220}	[0.308]	[0.304]	[0.384]	(0.843)	
Age	18.291	−0.054	−0.033	0.026	0.752	1,571
	{0.616}	[0.045]	[0.044]	[0.056]	(0.521)	
Mother tongue is English	0.700	0.017	0.009	0.049	0.495	1,571
		[0.033]	[0.033]	[0.041]	(0.686)	
Survey response variables						
Lives at home	0.811	−0.040	0.009	−0.004	0.685	1,431
		[0.030]	[0.030]	[0.038]	(0.561)	
At first choice school	0.243	0.024	0.060	0.047	1.362	1,430
		[0.034]	[0.033]*	[0.042]	(0.253)	
Plans to work while in school	0.777	0.031	−0.066	0.037	2.541	1,431
		[0.032]	[0.031]**	[0.040]	(0.055)	
Mother a high school graduate	0.868	0.015	−0.021	−0.045	1.040	1,431
		[0.026]	[0.026]	[0.033]	(0.374)	
Mother a college graduate	0.358	0.053	−0.020	−0.052	1.487	1,431
		[0.037]	[0.036]	[0.046]	(0.216)	
Father a high school graduate	0.839	0.025	0.008	−0.017	0.416	1,431
		[0.028]	[0.027]	[0.035]	(0.741)	
Father a college graduate	0.451	0.021	−0.001	−0.024	0.216	1,431
		[0.038]	[0.037]	[0.048]	(0.885)	
Rarely puts off studying for tests	0.208	0.031	0.031	0.107	2.534	1,431
		[0.032]	[0.031]	[0.040]***	(0.055)	
Never puts off studying for tests	0.056	−0.019	−0.016	−0.032	1.206	1,431
		[0.016]	[0.016]	[0.021]	(0.306)	
Wants more than a BA	0.556	0.052	−0.029	0.073	(1.752)	1,431
		[0.038]	[0.037]	[0.048]	(0.155)	
Intends to finish in 4 years	0.821	−0.008	−0.006	−0.063	(0.942)	1,431
		[0.030]	[0.029]	[0.037]*	(0.419)	

Notes: Standard deviations are shown in braces in column 1. Standard errors are reported in brackets in columns 2–4. p-values for F-tests are reported in parentheses in column 5. The last column shows the number of nonmissing observations.

- Impact!

TABLE 5—TREATMENT EFFECTS ON FIRST YEAR OUTCOMES IN THE SAMPLE WITH FALL GRADES

	SFP by type			Any SFP		
	All (1)	Men (2)	Women (3)	All (4)	Men (5)	Women (6)
Panel A. Fall grade						
Control mean	64.225	65.935	62.958	64.225	65.935	62.958
	(11.902)	(11.340)	(12.160)	(11.902)	(11.340)	(12.160)
SSP	0.349	−0.027	0.737	0.344	−0.014	0.738
	[0.917]	[1.334]	[1.275]	[0.917]	[1.332]	[1.274]
SFP	1.824	0.331	2.602			
	[0.847]**	[1.233]	[1.176]**			
SFSP	2.702	−0.573	4.205			
	[1.124]**	[2.010]	[1.325]***			
SFP (any)				2.125	0.016	3.141
				[0.731]***	[1.164]	[0.972]***
Observations	1,255	526	729	1,255	526	729
Panel B. First year GPA						
Control mean	1.805	1.908	1.728	1.797	1.885	1.731
	(0.902)	(0.908)	(0.891)	(0.904)	(0.910)	(0.894)
SSP	0.073	0.011	0.116	0.071	0.008	0.116
	[0.066]	[0.107]	[0.082]	[0.066]	[0.107]	[0.082]
SFP	0.010	−0.110	0.086			
	[0.064]	[0.103]	[0.084]			
SFSP	0.210	0.084	0.267			
	[0.092]**	[0.162]	[0.117]**			
SFP (any)				0.079	−0.042	0.147
				[0.056]	[0.095]	[0.073]**
Observations	1,255	526	729	1,255	526	729

Notes: The table reports regression estimates of treatment effects on full grades and first-year GPA computed using the full set of controls. Robust standard errors are reported in brackets. The sample is limited to students registered for at least two courses as of November 1 with data on the relevant set of controls and at least one fall grade. The last three columns report estimates from a model that combines the SFP and SFSP treatment groups into "SFP (any)."
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

- Well, maybe.
  - See Angrist, Oreopoulos, and Tyler, "When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards," *J. Human Resources* 49 (Summer 2014) for an update
  - As at the movies, research sequels often disappoint

- Who noticed the experiment?

Table 3—Program Sign-up and Use of Services

	Signed up for STAR		Received SSP services		Met with/emailed an advisor		Attended FSGs	
	Basic controls (1)	All controls (2)	Basic controls (3)	All controls (4)	Basic controls (5)	All controls (6)	Basic controls (7)	All controls (8)
Panel A. All								
Offered SSP	0.519	0.549	0.238	0.255	0.204	0.217	0.106	0.118
	[0.032]***	[0.034]***	[0.028]***	[0.029]***	[0.026]***	[0.028]***	[0.020]***	[0.021]***
Offered SFP	0.863	0.867						
	[0.022]***	[0.022]***						
Offered SSP	0.762	0.792	0.412	0.431	0.383	0.397	0.131	0.139
and SFP	[0.036]***	[0.036]***	[0.041]***	[0.044]***	[0.041]***	[0.043]***	[0.029]***	[0.031]***
Observations	1,571	1,431	1,571	1,431	1,571	1,431	1,571	1,431
Panel B. Men								
Offered SSP	0.447	0.464	0.194	0.206	0.145	0.149	0.096	0.107
	[0.049]***	[0.052]***	[0.039]***	[0.042]***	[0.035]***	[0.038]***	[0.029]***	[0.032]***
Offered SFP	0.792	0.806						
	[0.040]***	[0.040]***						
Offered SSP	0.705	0.708	0.298	0.291	0.282	0.270	0.115	0.112
and SFP	[0.058]***	[0.065]***	[0.058]***	[0.063]***	[0.057]***	[0.061]***	[0.042]***	[0.046]**
Observations	665	594	665	594	665	594	665	594
Panel C. Women								
Offered SSP	0.571	0.605	0.273	0.287	0.251	0.264	0.113	0.124
	[0.043]***	[0.044]***	[0.038]***	[0.040]***	[0.037]***	[0.040]***	[0.027]***	[0.029]***
Offered SFP	0.912	0.908						
	[0.024]***	[0.026]***						
Offered SSP	0.800	0.835	0.506	0.532	0.466	0.489	0.146	0.155
and SFP	[0.046]***	[0.043]***	[0.056]***	[0.058]***	[0.056]***	[0.058]***	[0.040]***	[0.042]***
Observations	906	837	906	837	906	837	906	837

Notes: The table reports regression estimates of treatment effects on the dependent variables indicated in column headings. Robust standard errors are reported in brackets. The sample is limited to students registered for at least two courses as of November 1 with data on the relevant set of controls. "Basic controls" include sex, mother tongue, high school grade quartile, and number of credits enrolled. "All controls" includes basic controls plus responses to survey questions on procrastination and parents education.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

- Why does this matter?