



## Statistics Review

- A statistic is a *function* of a *sample* of data
- An *estimator* is a statistic
- Population parameter  $\rightarrow$  unknown
- Estimator  $\rightarrow$  used to estimate an unknown population parameter
- The sample,  $y$ , will be considered random
- Since  $y$  is random, estimators using  $y$  will be random

sample (like the die rolls in assign 1)

Since estimators are random, they have a probability function, given a special name: sampling distribution.

We will obtain properties of the sampling distribution to see if the estimator is "good" or not.

1

### 3.1 Random Sampling from the Population

contains truth

- Typically, we want to know something about a population
- The population is considered to be very large (infinite), and contains some unknown "truth"
- We likely won't observe the whole population, but a sample from the pop.
- We'll use the sample,  $y$ , to estimate that something

2

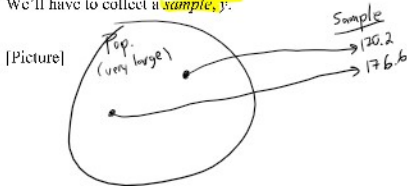
Example: suppose we want to know the mean height of a U of M student

Let  $y$  = height of a ~~single~~ student

- Population: all ~~single~~ students
- Population parameter of interest:  $\mu_y$

We can't afford to observe the whole pop.

We'll have to collect a sample,  $y$ .



3

We want the sample to reflect the population.

Question: How should the sample be selected from the population?  
randomly

In particular we want the sample to be i.i.d.

- Identically  $\rightarrow$  come from pop. of U of M students (no min: = U students)
- Independently  $\rightarrow$  no link/connection (entire basketball team)
- Distributed

4

So, the sample  $y$  is random!!

- Could have gotten a different  $y$
- Parallel universe

Table 3.1: Entire population of heights (in cm). The true (unobservable) population mean and variance are  $\mu_y = 176.8$  and  $\sigma_y^2 = 39.7$ .

177.3	170.2	187.2	178.3	170.3	179.4	181.2	180.0	173.9
178.7	171.7	160.5	183.9	175.7	175.9	182.6	181.7	180.2
181.5	176.5	162.1	180.3	175.6	174.9	165.7	172.7	178.9
175.3	178.7	175.6	166.4	173.1	173.2	175.6	183.7	181.3
174.2	180.9	179.9	171.2	171.0	178.6	181.4	175.2	182.2
171.7	178.4	168.1	186.0	189.9	173.1	168.7	180.0	175.1
175.7	180.8	176.2	170.8	177.3	163.4	186.3	177.1	191.2
171.0	180.3	169.5	167.2	178.0	172.9	176.0	176.5	171.9
175.1	184.2	165.3	180.2	178.3	183.4	178.9	178.6	177.9
184.5	184.1	180.9	187.1	179.9	167.1	172.0	167.4	172.7
171.6	186.6	182.4	185.5	174.8	178.8	192.8	179.3	172.0

5

How could i.i.d. be violated in the heights example?

Example: mean income of Canadians. How could i.i.d. be violated?

How should we estimate the mean height?

### 3.2 Estimators and Sampling Distributions

An estimator uses the sample  $y$  to "guess" something about the pop.

We collect our sample,  $y = \{173.9, 171.7, 182.6, 181.5, 162.1, 174.9, 165.7, 182.2, 171.7, 168.1, 180.9, 175.7, 163.4, 186.3, 169.5, 171.9, 173.9, 173.9, 172.7, 172.0\}$ . How should we use this sample to estimate the mean height?

Sample mean / sample average / average

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

6

#### 3.2.1 Sample mean

A popular choice for estimating a population mean is by using a sample mean (or sample average or just average)



(3.1)

$$\frac{\sum_{i=1}^n y_i}{n}$$

• From heights example,  $y = \{174.9\}$ ,  $\mu_y = 176.8$

• There are many ways to estimate  $\mu_y$ . Examples?

• Why is (3.1) so popular? It's the best

• How good is  $\bar{y}$  at estimating  $\mu_y$  in general?

• To answer these questions: idea of a sampling distribution

median / mode / geometric average / harmonic avg.  
 $\frac{\min(y) + \max(y)}{2}$

7

Recall that the sample,  $y$ , is random. Each element of  $y$  was selected randomly from the population. We could have selected a different sample of size  $n = 20$ . For example, in a parallel universe, we could have gotten  $y = \{175.9, 175.3, 182.2, 178.6, 175.2, 180.3, 178.3, 183.7, 176.0, 167.4, 178.7, 178.7, 185.0, 175.6, 180.0, 168.7, 178.6, 173.1, 173.2, 187.1\}$ , where the  $\bar{y}$  in  $y'$  denotes that we are in the parallel universe. In this parallel universe, we got  $\bar{y}' = 177.6$ . But in every universe, the population (table 3.1), is the same.  $\bar{y} = 176.8$

• Randomly sample from the population  $\rightarrow$  get  $y$

◦  $y$  is random

• Use  $y$  to calculate  $\bar{y}$

◦  $\bar{y}$  is random

◦ could have gotten a different sample  $\rightarrow$  could have gotten a different  $\bar{y}$

◦ population is always the same ( $\mu_y$ )

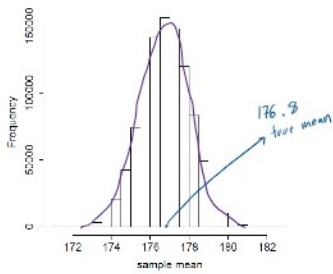
8

### 3.2.2 Sampling distribution of the sample mean

- $\bar{y}$  is **random variable** (it's an estimator, all estimators are random)
- **random variables** usually have **probability functions**
- $\bar{y}$  has a **sampling distribution** (probability function for an estimator)
- **sampling distribution** - imagine all possible values for  $\bar{y}$  that you could get - plot a histogram
- Using a computer, I drew **1 mil. different random samples of  $n=20$**  from table 3.1. Calculate  $\bar{y}$  each time. Plot histogram:

9

Figure 3.1: Histogram for 1 million  $\bar{y}$ s



10

Normal dist<sup>n</sup>

Which probability function is right for  $\bar{y}$ ? Why?

- Look at figure 3.1
- Notice the **summation operator** in equation 3.1
- Answer: Normal Reason: CLT (adding in sample mean)

$\bar{y}$  is random. We'll derive its:

- mean
- variance

Use these to determine if it's a "good" estimator via three statistical properties:

- **Bias**
- **Efficiency**
- **Consistency**

11

### 3.2.3 Bias

An estimator is unbiased if its expected value is equal to the population parameter it's estimating.

That is,  $\bar{y}$  is unbiased if  $E[\bar{y}] = \mu_y$ .

Unbiased if it gives "the right answer on average".

Biased if it gives the wrong answer on average.

Rules of the mean  
 (i)  $E[cY] = c E[Y]$   
 (ii)  $E[X+Y] = E[X] + E[Y]$

$$\begin{aligned}
 E[\bar{y}] &= E\left[\frac{1}{n} \sum y_i\right] \\
 &= \frac{1}{n} E\left[\sum y_i\right] = \frac{1}{n} E[y_1 + y_2 + \dots + y_n] \\
 &= \frac{1}{n} \{E[y_1] + E[y_2] + \dots + E[y_n]\} \\
 &= \frac{1}{n} \{\mu_y + \mu_y + \dots + \mu_y\} \\
 &= \frac{1}{n} n \mu_y = \mu_y \quad \bar{y} \text{ is unbiased}
 \end{aligned}$$

i.i.d.  
↳ "identical"

12

$$\begin{aligned}
 E[\bar{y}] &= E\left[\frac{1}{n} \sum_{i=1}^n y_i\right] \\
 &= \frac{1}{n} E\left[\sum_{i=1}^n y_i\right] \\
 &= \frac{1}{n} E[y_1 + y_2 + \dots + y_n] \quad (3.2) \\
 &= \frac{1}{n} (E[y_1] + E[y_2] + \dots + E[y_n]) \\
 &= \frac{1}{n} (\mu_y + \mu_y + \dots + \mu_y) \\
 &= \frac{n\mu_y}{n} = \mu_y
 \end{aligned}$$

13

### 3.2.4 Efficiency $\rightarrow$ accuracy/spread of estimator

An estimator is **efficient** if it has the **smallest variance** among all other potential **estimators** (for us, potential = linear, unbiased)

Need to get the variance of  $\bar{y}$ .

Rules of variance

- (i)  $\text{var}(cY) = c^2 \text{var}(Y)$
- (ii)  $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X,Y)$

If 2 variables are independent  $\Rightarrow \text{cov} = 0$

$$\begin{aligned}
 \text{var}(\bar{y}) &= \text{var}\left(\frac{1}{n} \sum y_i\right) \\
 &= \frac{1}{n^2} \text{var}\left(\sum y_i\right) = \frac{1}{n^2} \text{var}(y_1 + y_2 + \dots + y_n) \\
 &= \frac{1}{n^2} \{ \text{var}(y_1) + \text{var}(y_2) + \dots + \text{var}(y_n) \} \\
 &= \frac{1}{n^2} \{ \sigma_y^2 + \sigma_y^2 + \dots + \sigma_y^2 \} \quad \rightarrow \text{i.i.d. } \hookrightarrow \text{"independent"} \\
 &= \frac{1}{n^2} n \sigma_y^2 = \frac{\sigma_y^2}{n}
 \end{aligned}$$

14

$$\begin{aligned}
 \text{Var}[\bar{y}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n y_i\right] \\
 &= \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n y_i\right] \\
 &= \frac{1}{n^2} \text{Var}[y_1 + y_2 + \dots + y_n] \quad (3.3) \\
 &= \frac{1}{n^2} (\text{Var}[y_1] + \text{Var}[y_2] + \dots + \text{Var}[y_n]) \\
 &= \frac{1}{n^2} (\sigma_y^2 + \sigma_y^2 + \dots + \sigma_y^2) \\
 &= \frac{n\sigma_y^2}{n^2} = \frac{\sigma_y^2}{n}
 \end{aligned}$$

any other estimator for  $\mu$

$\text{var}(\bar{y}) < \text{var}(\hat{\mu}_1)$

$\bar{y}$  is BLUE (best linear unbiased estimator)

- Gauss-Markov theorem proves this is minimum variance
- We'll also need this to **prove consistency**, and for **hyp. testing**

15

### 3.2.5 Consistency

Suppose we had a lot of information, ( $n \rightarrow \infty$ )

What value should we get for our estimator? **right answer, every time**

How would state this mathematically?

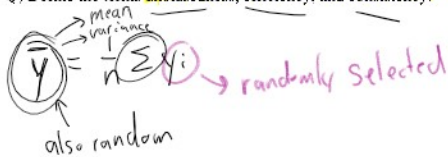
$$\lim_{n \rightarrow \infty} \text{var}(\bar{y}) \rightarrow 0 \quad \checkmark \quad \lim_{n \rightarrow \infty} \text{bias}(\bar{y}) \rightarrow 0 \quad \checkmark$$

Q) Prove that the **sample mean** is a **consistent estimator** for the population mean.

$$\text{var}(\bar{y}) = \frac{\sigma^2}{n} \quad \lim_{n \rightarrow \infty} \frac{\sigma^2}{n} \rightarrow 0$$

"variance goes away"

Q) Define the terms **unbiasedness**, **efficiency**, and **consistency**.



$$\begin{aligned}
 E[\bar{y}] &= \mu_y \quad \text{unbiased} \\
 \text{var}[\bar{y}] &= \frac{\sigma_y^2}{n} \rightarrow \text{efficient} \\
 &\quad \rightarrow \text{consistent}
 \end{aligned}$$

16

3.3 Hypothesis tests (known  $\sigma_y^2$ )

very unrealistic

alternative  $H_1: \mu_y \neq \mu_{y,0}$  (2-sided alternative)  $\rightarrow$  almost always 2-sided in Econ

null  $H_0: \mu_y = \mu_{y,0}$   $\rightarrow$  market

(3.4)

3.3 Hypothesis tests (known  $\sigma_y^2$ )  
 null  $H_0: \mu_y = \mu_{y,0}$   
 alternative  $H_A: \mu_y \neq \mu_{y,0}$  (2-sided alternative) (3.4)  
 almost always 2-sided in Econ

- Estimate  $\mu_y$  (using  $\bar{y}$  for example)
- See if  $\bar{y}$  appears "close" to  $\mu_{y,0}$ 
  - Remember,  $\bar{y}$  is random! (and Normal)
- If it's close  $\rightarrow$  fail to reject
- If it's far  $\rightarrow$  reject

17

Example:

- Hypothesize that mean height of a U of M student is 173cm

$H_0: \mu_y = 173$  (3.5)  
 $H_A: \mu_y \neq 173$

$(174.1 - 173) = 1.1 \text{ cm}$

- Collect a sample:  $y = \{173.9, 171.7, \dots, 172.0\}$

- Calculate  $\bar{y} = 174.1$

- Suppose (very unrealistically that we know that)  $\sigma_y^2 = 39.7$

- What now?

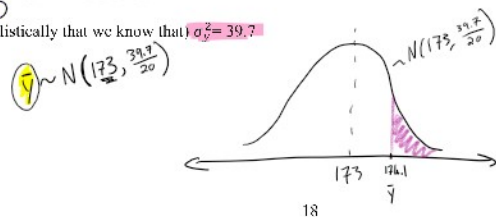
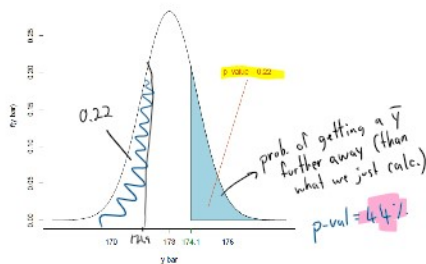


Figure 3.2: Normal distribution with  $\mu = 173$  and  $\sigma^2 = 39.7/20$ . Shaded area is the probability that the normal variable is greater than 174.1.



19

The p-value for the above test is 0.44. How to interpret this?  
 44% chance of getting a  $\bar{y}$  that is more adverse to  $H_0$ .  $\rightarrow$  fail to reject

3.3.1 Significance of a test

$\hookrightarrow$  pre-determined p-value that decides reject/fail to reject  
 $\alpha = 10\%, 5\%, 1\% \rightarrow$  if  $p\text{-val} < \alpha \Rightarrow$  reject

3.3.2 Type I error

$P_{\text{I}}(\text{reject } H_0 \mid H_0 \text{ is true}) = \alpha$

3.3.3 Type II error (and power)

(fail to reject  $H_0$  |  $H_0$  is false)

power =  $1 - \text{type II} = P_{\text{I}}(\text{reject } H_0 \mid H_0 \text{ is false}) = ?$

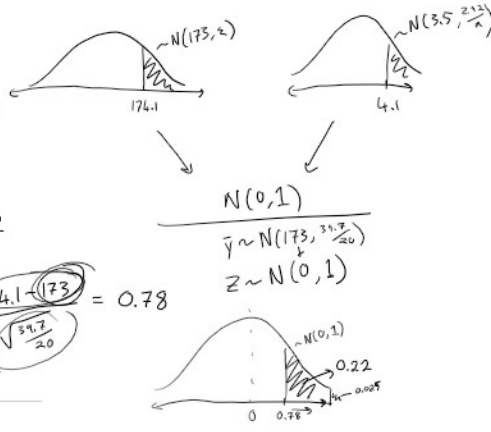
depends on how false  $H_0$

$H_0: \mu_y = 20$   
 in reality 20.01 vs. 1,000,000

20

### 3.3.4 Test statistics

- Just a more convenient way of getting the p-value for the test
- Each hypothesis test would present us with a new normal curve that we would have to draw, and calculate a new area (see fig. 3.2)
- Instead: **standardize**
- This gives us **one curve for all testing problems** (the standard normal curve)
- Calculate a bunch of areas under the curve, and tabulate them
- Not an issue with modern computers, but this is still the way we do things
- How to get a z test statistic?
- Do a z test for our heights example.



$$Z = \frac{\text{estimate} - \text{hypothesis}}{\sqrt{\text{var}(\text{estimator})}} = \frac{\bar{y} - \mu_{H_0}}{\sqrt{\sigma^2/n}} = \frac{174.1 - 173}{\sqrt{2/n}} = 0.78$$

$\beta\text{-val} = 0.22 \times 2 = 0.44 > 0.05$   
 $\hookrightarrow$  fail to reject

Table 2.3 Area under the standard normal curve to the right of z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.4960	0.4961	0.4963	0.4964	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970
0.1	0.4970	0.4971	0.4972	0.4973	0.4974	0.4975	0.4976	0.4977	0.4978	0.4979
0.2	0.4979	0.4980	0.4981	0.4982	0.4983	0.4984	0.4985	0.4986	0.4987	0.4988
0.3	0.4988	0.4989	0.4990	0.4991	0.4992	0.4993	0.4994	0.4995	0.4996	0.4997
0.4	0.4997	0.4998	0.4999	0.5000	0.5001	0.5002	0.5003	0.5004	0.5005	0.5006
0.5	0.5006	0.5007	0.5008	0.5009	0.5010	0.5011	0.5012	0.5013	0.5014	0.5015
0.6	0.5015	0.5016	0.5017	0.5018	0.5019	0.5020	0.5021	0.5022	0.5023	0.5024
0.7	0.5024	0.5025	0.5026	0.5027	0.5028	0.5029	0.5030	0.5031	0.5032	0.5033
0.8	0.5033	0.5034	0.5035	0.5036	0.5037	0.5038	0.5039	0.5040	0.5041	0.5042
0.9	0.5042	0.5043	0.5044	0.5045	0.5046	0.5047	0.5048	0.5049	0.5050	0.5051
1.0	0.5051	0.5052	0.5053	0.5054	0.5055	0.5056	0.5057	0.5058	0.5059	0.5060
1.1	0.5060	0.5061	0.5062	0.5063	0.5064	0.5065	0.5066	0.5067	0.5068	0.5069
1.2	0.5069	0.5070	0.5071	0.5072	0.5073	0.5074	0.5075	0.5076	0.5077	0.5078
1.3	0.5078	0.5079	0.5080	0.5081	0.5082	0.5083	0.5084	0.5085	0.5086	0.5087
1.4	0.5087	0.5088	0.5089	0.5090	0.5091	0.5092	0.5093	0.5094	0.5095	0.5096
1.5	0.5096	0.5097	0.5098	0.5099	0.5100	0.5101	0.5102	0.5103	0.5104	0.5105
1.6	0.5105	0.5106	0.5107	0.5108	0.5109	0.5110	0.5111	0.5112	0.5113	0.5114
1.7	0.5114	0.5115	0.5116	0.5117	0.5118	0.5119	0.5120	0.5121	0.5122	0.5123
1.8	0.5123	0.5124	0.5125	0.5126	0.5127	0.5128	0.5129	0.5130	0.5131	0.5132
1.9	0.5132	0.5133	0.5134	0.5135	0.5136	0.5137	0.5138	0.5139	0.5140	0.5141
2.0	0.5141	0.5142	0.5143	0.5144	0.5145	0.5146	0.5147	0.5148	0.5149	0.5150
2.1	0.5150	0.5151	0.5152	0.5153	0.5154	0.5155	0.5156	0.5157	0.5158	0.5159
2.2	0.5159	0.5160	0.5161	0.5162	0.5163	0.5164	0.5165	0.5166	0.5167	0.5168
2.3	0.5168	0.5169	0.5170	0.5171	0.5172	0.5173	0.5174	0.5175	0.5176	0.5177
2.4	0.5177	0.5178	0.5179	0.5180	0.5181	0.5182	0.5183	0.5184	0.5185	0.5186
2.5	0.5186	0.5187	0.5188	0.5189	0.5190	0.5191	0.5192	0.5193	0.5194	0.5195
2.6	0.5195	0.5196	0.5197	0.5198	0.5199	0.5200	0.5201	0.5202	0.5203	0.5204
2.7	0.5204	0.5205	0.5206	0.5207	0.5208	0.5209	0.5210	0.5211	0.5212	0.5213
2.8	0.5213	0.5214	0.5215	0.5216	0.5217	0.5218	0.5219	0.5220	0.5221	0.5222
2.9	0.5222	0.5223	0.5224	0.5225	0.5226	0.5227	0.5228	0.5229	0.5230	0.5231
3.0	0.5231	0.5232	0.5233	0.5234	0.5235	0.5236	0.5237	0.5238	0.5239	0.5240
3.1	0.5240	0.5241	0.5242	0.5243	0.5244	0.5245	0.5246	0.5247	0.5248	0.5249
3.2	0.5249	0.5250	0.5251	0.5252	0.5253	0.5254	0.5255	0.5256	0.5257	0.5258
3.3	0.5258	0.5259	0.5260	0.5261	0.5262	0.5263	0.5264	0.5265	0.5266	0.5267
3.4	0.5267	0.5268	0.5269	0.5270	0.5271	0.5272	0.5273	0.5274	0.5275	0.5276
3.5	0.5276	0.5277	0.5278	0.5279	0.5280	0.5281	0.5282	0.5283	0.5284	0.5285
3.6	0.5285	0.5286	0.5287	0.5288	0.5289	0.5290	0.5291	0.5292	0.5293	0.5294
3.7	0.5294	0.5295	0.5296	0.5297	0.5298	0.5299	0.5300	0.5301	0.5302	0.5303
3.8	0.5303	0.5304	0.5305	0.5306	0.5307	0.5308	0.5309	0.5310	0.5311	0.5312
3.9	0.5312	0.5313	0.5314	0.5315	0.5316	0.5317	0.5318	0.5319	0.5320	0.5321
4.0	0.5321	0.5322	0.5323	0.5324	0.5325	0.5326	0.5327	0.5328	0.5329	0.5330
4.1	0.5330	0.5331	0.5332	0.5333	0.5334	0.5335	0.5336	0.5337	0.5338	0.5339
4.2	0.5339	0.5340	0.5341	0.5342	0.5343	0.5344	0.5345	0.5346	0.5347	0.5348
4.3	0.5348	0.5349	0.5350	0.5351	0.5352	0.5353	0.5354	0.5355	0.5356	0.5357
4.4	0.5357	0.5358	0.5359	0.5360	0.5361	0.5362	0.5363	0.5364	0.5365	0.5366
4.5	0.5366	0.5367	0.5368	0.5369	0.5370	0.5371	0.5372	0.5373	0.5374	0.5375
4.6	0.5375	0.5376	0.5377	0.5378	0.5379	0.5380	0.5381	0.5382	0.5383	0.5384
4.7	0.5384	0.5385	0.5386	0.5387	0.5388	0.5389	0.5390	0.5391	0.5392	0.5393
4.8	0.5393	0.5394	0.5395	0.5396	0.5397	0.5398	0.5399	0.5400	0.5401	0.5402
4.9	0.5402	0.5403	0.5404	0.5405	0.5406	0.5407	0.5408	0.5409	0.5410	0.5411
5.0	0.5411	0.5412	0.5413	0.5414	0.5415	0.5416	0.5417	0.5418	0.5419	0.5420
5.1	0.5420	0.5421	0.5422	0.5423	0.5424	0.5425	0.5426	0.5427	0.5428	0.5429
5.2	0.5429	0.5430	0.5431	0.5432	0.5433	0.5434	0.5435	0.5436	0.5437	0.5438
5.3	0.5438	0.5439	0.5440	0.5441	0.5442	0.5443	0.5444	0.5445	0.5446	0.5447
5.4	0.5447	0.5448	0.5449	0.5450	0.5451	0.5452	0.5453	0.5454	0.5455	0.5456
5.5	0.5456	0.5457	0.5458	0.5459	0.5460	0.5461	0.5462	0.5463	0.5464	0.5465
5.6	0.5465	0.5466	0.5467	0.5468	0.5469	0.5470	0.5471	0.5472	0.5473	0.5474
5.7	0.5474	0.5475	0.5476	0.5477	0.5478	0.5479	0.5480	0.5481	0.5482	0.5483
5.8	0.5483	0.5484	0.5485	0.5486	0.5487	0.5488	0.5489	0.5490	0.5491	0.5492
5.9	0.5492	0.5493	0.5494	0.5495	0.5496	0.5497	0.5498	0.5499	0.5500	0.5501
6.0	0.5501	0.5502	0.5503	0.5504	0.5505	0.5506	0.5507	0.5508	0.5509	0.5510
6.1	0.5510	0.5511	0.5512	0.5513	0.5514	0.5515	0.5516	0.5517	0.5518	0.5519
6.2	0.5519	0.5520	0.5521	0.5522	0.5523	0.5524	0.5525	0.5526	0.5527	0.5528
6.3	0.5528	0.5529	0.5530	0.5531	0.5532	0.5533	0.5534	0.5535	0.5536	0.5537
6.4	0.5537	0.5538	0.5539	0.5540	0.5541	0.5542	0.5543	0.5544	0.5545	0.5546
6.5	0.5546	0.5547	0.5548	0.5549	0.5550	0.5551	0.5552	0.5553	0.5554	0.5555
6.6	0.5555	0.5556	0.5557	0.5558	0.5559	0.5560	0.5561	0.5562	0.5563	0.5564
6.7	0.5564	0.5565	0.5566	0.5567	0.5568	0.5569	0.5570	0.5571	0.5572	0.5573
6.8	0.5573	0.5574	0.5575	0.5576	0.5577	0.5578	0.5579	0.5580	0.5581	0.5582
6.9	0.5582	0.5583	0.5584	0.5585	0.5586	0.5587	0.5588	0.5589	0.5590	0.5591
7.0	0.5591	0.5592	0.5593	0.5594	0.5595	0.5596	0.5597	0.5598	0.5599	0.5600
7.1	0.5600	0.5601	0.5602	0.5603	0.5604	0.5605	0.5606	0.5607	0.5608	0.5609
7.2	0.5609	0.5610	0.5611	0.5612	0.5613	0.5614	0.5615	0.5616	0.5617	0.5618
7.3	0.5618	0.5619	0.5620	0.5621	0.5622	0.5623	0.5624	0.5625	0.5626	0.5627
7.4	0.5627	0.5628	0.5629	0.5630	0.5631	0.5632	0.5633	0.5634	0.5635	0.5636
7.5	0.5636	0.5637	0.5638	0.5639	0.5640	0.5641	0.5642	0.5643	0.5644	0.5645
7.6	0.5645	0.5646	0.5647	0.5648	0.5649	0.5650	0.5651	0.5652	0.5653	0.5654
7.7	0.5654	0.5655	0.5656	0.5657	0.5658	0.5659	0.5660	0.5661	0.5662	0.5663
7.8	0.5663	0.5664	0.5665	0.5666	0.5667	0.5668	0.5669	0.5670	0.5671	0.5672
7.9	0.5672	0.5673	0.5674	0.5675	0.5676	0.5677	0.5678	0.5679	0.5680	0.5681
8.0	0.5681	0.5682	0.5683	0.5684	0.5685	0.5686	0.5687	0.5688	0.5689	0.5690
8.1	0.5690	0.5691	0.5692	0.5693	0.5694	0.5695	0.5696	0.5697	0.5698	0.5699
8.2	0.5699	0.5700	0.5701	0.5702	0.5703	0.5704	0.5705	0.5706	0.5707	0.5708
8.3	0.5708	0.5709	0.5710	0.5711	0.5712	0.5713	0.5714	0.5715	0.5716	0.5717
8.4	0.5717	0.5718	0.5719	0.5720	0.5721	0.5722	0.5723	0.5724	0.5725	0.5726
8.5	0.5726	0.5727	0.5728	0.5729	0.5730	0.5731	0.5732	0.5733	0.5734	0.5735
8.6	0.5735	0.5736	0.5737	0.5738	0.5739	0.5740	0.5741	0.5742	0.5743	0.5744
8.7	0.5744	0.5745	0.5746	0.5747	0.5748	0.5749	0.5750	0.5751	0.5752	0.5753
8.8	0.5753	0.5754	0.5755	0.5756	0.5757	0.5758	0.5759	0.5760	0.5761	0.5762
8.9	0.5762	0.5763	0.5764	0.5765	0.5766	0.5767	0.5768	0.5769	0.5770	0.5771
9.0	0.5771	0.5772	0.5773	0.5774	0.5775	0.5776	0.5777	0.5778	0.5779	0.5780
9.1	0.5780	0.5781	0.5782	0.5783	0.5784	0.5785	0.5786	0.5787	0.5788	0.5789
9.2	0.5789	0.5790	0.5791	0.5792	0.5793	0.5794	0.5795	0.5796	0.5797	0.5798
9.3	0.5798	0.5799	0.5800	0.5801	0.5802	0.5803	0.5804	0.5805	0.5806	0.5807
9.4	0.5807	0.5808	0.5809	0.5810	0.5811					

3.4.1 Estimating  $\sigma_y^2$

$\text{var}[\bar{y}] = E[(\bar{y} - E(\bar{y}))^2]$  estimate  $(\frac{1}{n} \sum y_i)$

- A "natural" estimator:

$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$

(3.15)

$E[\frac{1}{n} \sum (y_i - \bar{y})^2] = \frac{n-1}{n} \sigma^2$  **BIASED**

$\rightarrow E[\frac{n}{n-1} \hat{\sigma}^2] = E[\frac{1}{n-1} \sum (y_i - \bar{y})^2] = \frac{n}{n-1} E[\hat{\sigma}^2] = \frac{n}{n-1} \frac{n-1}{n} \sigma^2$  **UNBIASED**

- Is this a good estimator? Why or why not?
- A better estimator:

$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$  (3.17)

- Degrees-of-freedom correction

$y = \{1, 3, ?\}$   
 $\bar{y} = 3$   
 calc.  $\bar{y} \rightarrow$  lose 1 piece info.

$\frac{1}{n-1} \sum (y_i - \bar{y})^2 = S^2$

So:

Estimated variance of  $\bar{y}$   $\frac{\sigma_y^2}{n} \rightarrow \frac{1}{n-1} E(y_i - \bar{y})^2$

We can implement hypothesis testing by replacing the unknown  $\sigma_y^2$  with its estimator  $s_y^2$ . The z test statistic now becomes:

$\frac{\bar{y} - \mu_{y,0}}{\sqrt{s_y^2/n}} = t$  using  $s_y^2$  instead of  $\sigma^2$  makes z become t

only random thing (Normal)  $\rightarrow z$  also Normal  
 $Z = \frac{\bar{y} - \mu_{y,0}}{\sqrt{\sigma_y^2/n}}$

Note: for large n, the t test is equivalent to the z test  $t \sim t_{n-1}$

