

Sampling Distribution of Differences Topic 21 covers the distribution of a difference between two independent sample proportions or two independent sample means.

Topic 21—Sampling Distribution of a Difference Between Two Independent Sample Proportions or Two Independent Sample Means (Simulations)

Sampling Distribution of a Difference Between Two Independent Sample Proportions

This simulation will reinforce the mathematical reality that

the mean of the distribution of $p_1 - p_2$ is $p_1 - p_2$ (the difference of the two population proportions) and the

variance
$$\sigma_{p_1-p_2}^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2 = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

(the sum of the two population variances) and thus

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Note: In later topics it will be hypothesized that $\mathbf{p}_1 = \mathbf{p}_2$ (or $\mathbf{p}_1 - \mathbf{p}_2 = \mathbf{0}$), so you will look at that case here. The difference need not be zero, however, and in the next section you will observe such an example. *Example*: $n_1 = 50$, $p_1 = 0.33$, and $n_2 = 35$, $p_2 = 0.33$. One hundred samples from population one were generated in Topic 19, screens 5, 6, and 7 and saved in **list1**.

Change to folder **RACE**, since you will be using some data from Topic 19.

- 1. From the Home screen, set RandSeed 987.
- 2. Calculate tistat.randbin(50,.33,100)/50>list1.
- 3. Press ENTER to display screen 1.
- 4. Use CATALOG to calculate mean(list1)

5. Use CATALOG to calculate stdDev(list1)

0.000

stablev =
$$S_x = 0.003 \approx$$

 $\sigma_{\hat{p}_1} = \sqrt{\frac{p_1(1-p_1)}{n_1}} = \sqrt{\frac{(.33)(.67)}{50}} = .066$.

- 6. Press [2nd] [ANS] ^ 2 and then press [ENTER] to display the third line in screen 2.
- 7. Use CATALOG to calculate variance(list1). (See screen 2.)

variance =
$$s_x^2 \approx .004 \approx \sigma_{\hat{p}_1}^2 = \frac{p_1(1-p_1)}{n_1} = .0044$$
.

Compare with Topic 19, screen 7.

Simulate 100 samples from population two and store in **list2**, similar to Topic 19 and shown in screens 3 and 4.

- 1. Set RandSeed 9876.
- 2. Calculate tistat.randbin(35,.33,100)/35>list2.
- 3. Press ENTER to display screen 3.
- 4. Use CATALOG to calculate mean(list2)

mean = .330857
$$\approx$$
 p₂ = .33

5. Use CATALOG to calculate stdDev(list2)

stdDev = .079493
$$\approx \sigma_{\hat{p}_2} = \sqrt{\frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{(.33)(.67)}{35}} = .07948$$

	F1+ Tools A1:	F2+ 9ebra	F3+) F4+ a1c0the	r F5 r Pr3ml	F6+ OC1ean	
	Rand	Seed	987			Done
	_ tis	tat.	randb	in(50	9,.33	5,11
				50	~~	
	٤.30	· ·	ა8 .	J4	.28	-20
1)	Bin(50,.	33, 10	90)/5	O⇒lis	301
• /	RACE	F	KAD APPR	0X FI	INC	2/30

(

	F1+ F2+ F3+ F4+ F ToolsAlgebraCalcOtherPrs	5 F6+ ImIOC1ean UP
	■mean(list1)	.329
	■stdDev(list1)	.062628
	•(.06262764742685	53) ²
		.003922
	■variance(list1)	.003922
	variance(list1)	
(2)	RACE RAD APPROX	FUNC 6/30



6. Use CATALOG to calculate variance(list2)

variance =.006319
$$\approx \sigma_{\hat{p}_2}^2 = \frac{p_2(1-p_2)}{n_2} = \frac{(.33)(.67)}{35} = .006317$$

(See screen 4.)

7. Store the difference using **list1 – list2>list3** and find the mean, standard deviation, and variance of this distribution of differences (screen 5).

Note that list3[1] = list1[1] – list2[1] = .36 - .257143 = .102857 from screens 1, 3, and 5.

8. Calculate the mean, stdDev, and variance of list3.

mean = .001857
$$\approx \mu_{\hat{p}_{1-}\hat{p}_{2}} = .00 = p_{1} - p_{2} = .33 - .33$$

stdDev = .104406 $\approx \sigma_{\hat{p}_{1}-\hat{p}_{2}} = .104 = \sqrt{\frac{p_{1}(1-p_{1})}{n_{1}} + \frac{p_{2}(1-p_{2})}{n_{2}}} = \sqrt{\frac{(.33)(.67)}{50} + \frac{(.33)(.67)}{35}}$
variance = .010901 $\approx \sigma_{\hat{p}_{1}-\hat{p}_{2}}^{2} = .0107 = \frac{p_{1}(1-p_{1})}{n_{1}} + \frac{p_{2}(1-p_{2})}{n_{2}} = \frac{(.33)(.67)}{50} + \frac{(.33)(.67)}{35}$
(See screen 5)

It is important to note that it is *not* true that $\sigma_{\hat{p}_1 - \hat{p}_2} = \sigma_{\hat{p}_1} + \sigma_{\hat{p}_2}$.

 $\sigma_{\hat{p}_1 - \hat{p}_2} = .10 \neq .07 + .08 = .15 = \sigma_{\hat{p}_1} + \sigma_{\hat{p}_2}$

On the other hand it *is* true that $\sigma_{\hat{p}_1 - \hat{p}_2}^2 = \sigma_{\hat{p}_1}^2 + \hat{p}_2^2$. $\sigma_{\hat{p}_1 - \hat{p}_2}^2 = .01 = .004 + .006 = \sigma_{\hat{p}_1}^2 + \hat{p}_2^2$

- From the Stats/List Editor, set up and define Plot 1, Plot 2, and Plot 3 as modified boxplots with Mark: Square, using list1, list2, and list3, respectively.
- 10. From the Plot Setup screen, press F5 **ZoomData** and F3 **Trace**, which graphically shows the distribution of \hat{p}_1 's at the top, \hat{p}_2 's in the middle with the same mean as \hat{p}_1 's but with greater spread, and then the distribution of $\hat{p}_1 - \hat{p}_2$'s at the bottom centered at zero with still greater spread (screen 6).



F17 Tools	F2+ A19ebra	F3+	F4+ Ither	F5 Pr9ml		67 IN UP	\Box
me	an(li	st2))		.3	308	357
■st	dDev(list	t2)		.0)794	493
∎ va	rianc	e(1)	isti	2)	.0	0063	319
var:	iance	li≤	st2)				
RACE		RAD A	PPRO)	(FI	UNC	3	730

	F1+ F2+ F3+ F4+ F ToolsAl9ebraCalcOtherPrs	75 FI 9ml0C1ea	4 U A
	∎list1-list2→1	ist3	
	{.102857 .0657	'14 ·	•.031)
	∎mean(list3)	0	01857
	∎stdDev(list3)	. 1	04406
	variance(list3)	.0	10901
(5)	variance(list3)		
(S)	RACE RAD APPROX	FUNC	4/30

- From the Stats/List Editor, press F2 Plots and select
 3:Plotsoff to turn off the plots.
- 12. Press F2 Plots and 2:Norm Prob Plot to define Plot 4 using list3 with Mark: Dot.
- 13. Press ENTER to return to the Stats/List Editor.
- 14. Select F2 Plots, 1:Plot Setup and F5 Zoom Data to get a probability plot on the distribution of the differences in **list3** (screen 7).

Observe that the distribution of the differences also is normally distributed because the sample size is large enough.

Sampling Distribution of a Difference Between Two Independent Sample Means

The simulation will reinforce the fact that $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$ and the variance $\sigma_{\bar{x}_1-\bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1}{n_1}^2 + \frac{\sigma_2}{n_2}^2$, and the Standard Deviation $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Simulate 100 samples of size 9 from a normal population with $\mu = 85$ and $\sigma = 15$ and store in list1 (similar to Topic 20, screen 17, but in that case it was a uniform distribution).

For this section, change to folder **BLDTALL**.

- 1. From the Home screen, set **RandSeed 321**.
- Calculate seq(mean(tistat.randnorm(85,15,9)),x,1,100)→list1 (second line of screen 8).

Simulate 100 samples of size 16 from a normal population
with $\mu = 60$ and $\sigma = 20$ and store in list2.

1. Calculate

seq(mean(tistat.randnorm(60,20,16)),x,1,100)→**list2** (third line of screen 8).



	F1+ F2+ F3+ F4+ F5 Too1sA13ebraCa1c0therPr3m10C1+	F6+ 2an Up
	■RandSeed 321	Done
	seq(mean(tistat.rand (95.8521 79.5482)	inorm() 78.74
	seq(mean(tistat.rano (55.4452 61.7107	inorm() 69.72)
(8)	0,20,16)),×,1,100)→) BLDTALL RAD AUTO FUNC	list2 3/30

Store the differences using list1 – list2→list3 (top of screen 9).

3. Find the mean, standard deviation, and variance of **list1**, **list2**, and **list3** (screens 9, 10, and 11).

mean(list1) = 85.122 $\approx 85 = \mu_1 = \mu_{\bar{x}_1}$

mean(list2) = $60.2003 \approx 60 = \mu_2 = \mu_{\bar{x}_2}$

Note that:

	F1+ F2+ F3+ F4+ F5 ToolsA19ebraCa1cOtherPr9miDClean Up						
	■ list1 - 1 (40.4069	ist2 → 1 17.83	ist3 75 9.	.0176			
	∎mean(lis	t1)	85	5.122			
	∎mean(lis	t2)	60.	2003			
	■ mean(list	t3)	24.	9217			
(0)	mean(list3	3)					
(9)	BLDTALL RE	AD AUTO	FUNC	4/30			

	F1+ F2+ ToolsAl9ebra	F3+ F4+ CalcOther	F5 Pr9mi0C1	F6+ ean Up
			-	
	■ stdDev(list1)		6.05627
	■ stdDev(list2)	5	5.37592
	stdDev(list3)		7.4496
(10)	stdDev(1	ist3)		
(10)	BLDTALL	RAD AUTO	FUNC	3/30

Note: stdDev(list3) does not

$$equal \frac{\sigma_1}{\sqrt{n_1}} + \frac{\sigma_2}{\sqrt{n_2}} = 5 + 5 = 10.$$

o 15

mean(list3) = 24.9217 $\approx 25 = \mu_1 - \mu_2 = \mu_{\bar{x}_1 - \bar{x}_2}$

stdDev(list1) = 5.05627
$$\approx 5 = \frac{6}{\sqrt{n_1}} = \frac{15}{\sqrt{9}} = \sigma_{\bar{x}_1}$$

stdDev(list2) = 5.37592
$$\approx 5 = \frac{\sigma_2}{\sqrt{n_2}} = \frac{20}{\sqrt{16}} = \sigma_{\bar{x}_2}$$

stdDev(list3) = 7.4496 =
$$\sqrt{55.4965} \approx \sqrt{50} = \sqrt{\sigma_{\bar{x}_1 - \bar{x}_2}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

variance(list1) = 25.5659
$$\approx 25 = \frac{{\sigma_1}^2}{n_1} = \frac{15^2}{9} = {\sigma_{\bar{x}_1}}^2$$

variance(list2) = 28.9005
$$\approx 25 = \frac{\sigma_2^2}{n_2} = \frac{20^2}{16} = \sigma_{\bar{x}_2}^2$$

	F1+) F2+ F3+ F4+ F Too1sA19ebraCa1cOtherPr3	'S F6+ 9miOClean UP
	■variance(list1) ■variance(list2)	25.5659 28.9005
	■variance(list3)	55.4965
(11)	variance(list3) BLOTALL BAD AUTO	FUNC 3/30

variance(list3) = 55.4965
$$\approx 50 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma_{\bar{x}_1 - \bar{x}_2}^2$$

- From the Stats/List Editor, set up and define Plot 1, Plot 2, and Plot 3 as modified boxplots with Mark: Square, using list1, list2, and list3.
- 5. Deselect all other plots.

- 6. From the Plot Setup screen, press **F5 ZoomData** to display screen 12 with **list1** at the top, **list2** in the middle, and the distribution of the difference with the greater spread at the bottom.
- For another look at the results, set up and define Plot 1, Plot 2, and Plot 3 as histograms with x: list1, list2, and list3, and Hist. Bucket Width: 4.
- 8. Set up the window using [WINDOW] with the following entries:
 - xmin = 3
 - xmax = 99
 - xscl = 4
 - ymin = -15
 - ymax = 45
 - yscl = 0
 - xres = 1

(See screen 13.)

9. Press • [GRAPH] (screen 14).

ĺ	F1+ F2+ F3 F4 F5+ F6+ F7+ ⁵ oolsZoomTraceRe9raphMathDrawPen+:
	P1 ⊢₽-•
	•
	•
(12)	Med: 84.3284 se +>t+ DR TYPE + LESCI=CANCEL

	(F1+) F2 Tools[Zool xmin=3 xscl=4 ymin=- ymax=4 yscl=6 xres=1) 15. 5.		
(13)	BLDTALL	RAD AUTO	FUNC	

	F17 T0015	F2+ Zoom	F3 Trace	F4 Re9rapt	F5+ Math	F6+ Draw	F7+50 Pen
					_		_
		Л	-	Γ	11	Γ	Ih
		٦H	h	ᆔ			
						!!	
(14)	BLDTA	LL	RAD	AUTO	FUI	NC	