EXPLORATIONS

Chapter 10

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 $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$ (the difference of the two population proportions) and the variance $\sigma_{p_1 - p_2}{}^2 = \sigma_{p_1}{}^2 + \sigma_{p_2}{}^2 = \dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}$ (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* $\mathbf{p_1 = p_2}$ (*or* $\mathbf{p_1 - p_2 = 0}$)*, 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.*

© 2001 Texas Instruments Incorporated

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.

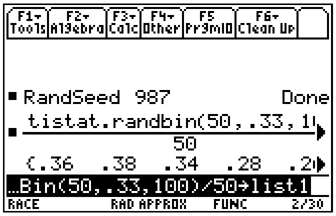
1. From the Home screen, set **RandSeed 987**.
2. Calculate **tistat.randbin(50,.33,100)/50→list1**.
3. Press [ENTER] to display screen 1.
4. Use [CATALOG] to calculate **mean(list1)**

   mean = .329 ≈ p_1 = .33.
5. Use [CATALOG] to calculate **stdDev(list1)**

   stdDev = $s_x = 0.063 \approx$

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

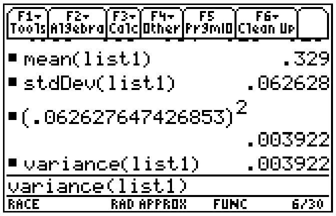
(1)



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.)

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

(2)



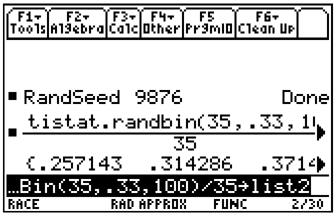
Compare with Topic 19, screen 7.

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 ≈ p_2 = .33.
5. Use [CATALOG] to calculate **stdDev(list2)**

(3)



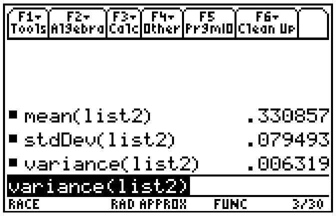
$$\text{stdDev} = .079493 \approx \sigma_{\hat{p}2} = \sqrt{\frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{(.33)(.67)}{35}} = .07948$$

© 2001 TEXAS INSTRUMENTS INCORPORATED

6. Use [CATALOG] to calculate **variance(list2)**

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

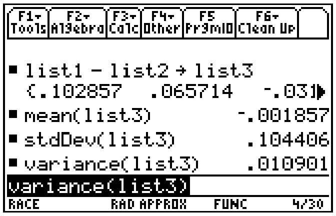
   (See screen 4.)



(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.



(5)

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

   $$\text{mean} = .001857 \approx \mu_{\hat{p}_1 - \hat{p}_2} = .00 = p_1 - p_2 = .33 - .33$$

   $$\text{stdDev} = .104406 \approx \sigma_{\hat{p}_1 - \hat{p}_2} = .104 = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{(.33)(.67)}{50} + \frac{(.33)(.67)}{35}}$$

   $$\text{variance} = .010901 \approx \sigma_{\hat{p}_1 - \hat{p}_2}^{\,2} = .0107 = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = \frac{(.33)(.67)}{50} + \frac{(.33)(.67)}{35}$$

   (See screen 5).

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

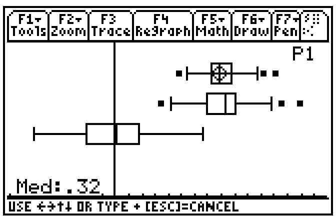
$$\sigma_{\hat{p}_1 - \hat{p}_2} = .10 \neq .07 + .08 = .15 = \sigma_{\hat{p}_1} + \sigma_{\hat{p}_2}$$

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

$$\sigma_{\hat{p}_1 - \hat{p}_2}^{\,2} = .01 = .004 + .006 = \sigma_{\hat{p}_1}^{\,2} + {}_{\hat{p}_2}^{\,2}$$

9. 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).

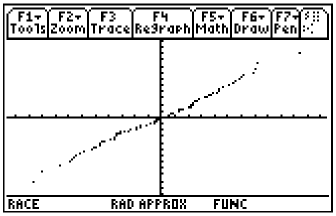


(6)

11. 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.

(7)



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 $\sigma_{\bar{x}_1-\bar{x}_2}{}^2=\sigma^2{}_{\bar{x}_1}+\sigma^2{}_{\bar{x}_2}=\dfrac{\sigma_1{}^2}{n_1}+\dfrac{\sigma_2{}^2}{n_2}$, and the

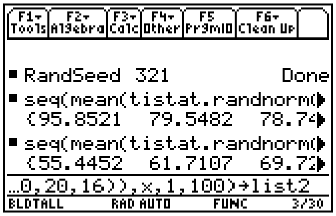
Standard Deviation $\sigma_{\bar{x}_1-\bar{x}_2}=\sqrt{\dfrac{\sigma_1{}^2}{n_1}+\dfrac{\sigma_2{}^2}{n_2}}$.

Simulate 100 samples of size 9 from a normal population with μ = **85** and σ = **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**.
2. Calculate **seq(mean(tistat.randnorm(85,15,9)),x,1,100)→list1** (second line of screen 8).

(8)

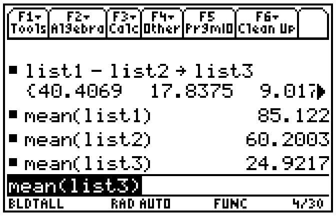


Simulate 100 samples of size 16 from a normal population with μ = **60** and σ = **20** and store in **list2**.

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

© 2001 TEXAS INSTRUMENTS INCORPORATED

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



(9)

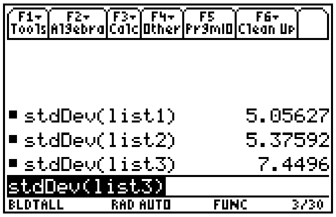
3. Find the mean, standard deviation, and variance of **list1**, **list2**, and **list3** (screens 9, 10, and 11).

   Note that:

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

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

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



(10)

Note: **stdDev(list3)** *does not equal* $\frac{\sigma_1}{\sqrt{n_1}} + \frac{\sigma_2}{\sqrt{n_2}}$ = 5+ 5=10.

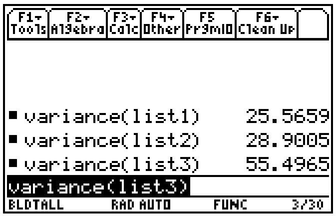
stdDev(list1) = 5.05627 ≈ 5 = $\frac{\sigma}{\sqrt{n_1}} = \frac{15}{\sqrt{9}} = \sigma_{\bar{x}_1}$

stdDev(list2) = 5.37592 ≈ 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 ≈ 25 = $\frac{\sigma_1{}^2}{n_1} = \frac{15^2}{9} = \sigma_{\bar{x}_1}{}^2$

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



(11)

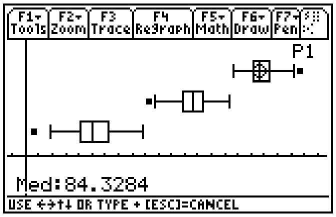
variance(list3) = 55.4965 ≈ 50 = $\sigma_{\bar{x}_1}{}^2 + \sigma_{\bar{x}_2}{}^2 = \frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2} = \sigma_{\bar{x}_1 - \bar{x}_2}{}^2$

4. 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.

© 2001 TEXAS INSTRUMENTS INCORPORATED

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.



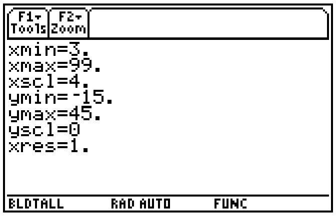
(12)

7. 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:

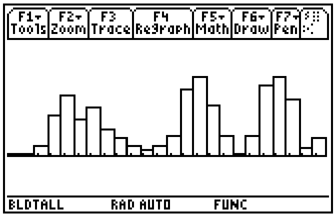
   - **xmin = 3**
   - **xmax = 99**
   - **xscl = 4**
   - **ymin = -15**
   - **ymax = 45**
   - **yscl = 0**
   - **xres = 1**

   (See screen 13.)



(13)

9. Press ◆ [GRAPH] (screen 14).



(14)

© 2001 TEXAS INSTRUMENTS INCORPORATED