Didem Bulut Aykurt

RES500 – Fundamentals of Quantities Analysis

Colorado State University-Global Campus

Dr. Mohammad Sumadi

June 26,2022

Chapter 7.25)

a) The sampling distribution of the difference between mean help that you might not know the population parameter or might not even be easy to find. So, the way that SDDBM estimates a population parameter is by taking a sample. And this case size n_m=30 and $n_f = 48$ that can calculate statistics from a sample data used to estimate a population parameter, and this is a random sample. But we know this sample's statistical result will not necessarily be the same as the population parameter. So the mean of the sampling distribution for the difference in sample means will be between the man's and woman's mean of the sample distribution so that man's mean minus the woman's mean.

$\overline{Xm}$ ~ N (μ_m +2, 20)

$\overline{Xw}$ ~ N (μ_w ,20)

Thus, the mean of the sampling distribution of the difference between means is

$\sum_i^{20}(\overline{Xm} - \overline{Xw})$ = 2 that is mean all of the different scenarios of the sample mean in each of the combinations gave a total of difference two as men's Anger-out score minus woman's Anger-out score total. These sample test results are statistically significate because the two samples mean not overlap.

b) We created an Anger-out score bar chart that shows crazy distribution as not usual. That is the reason for taking a sample to develop normal distribution. Barely can take sample data from normal distribution also. As the sample size is essential, the mean does not change when the sample size change, but the standard deviation gets smaller or bigger. The small sample size is not a perfect normal distribution but increases the sample size; a few things are happening charts are becoming more normal deviation going to have tighter. Let's talk about how-to affect sample size to the variance, which means that the standard error increases sample size and decreases the standard error. And the other is sample size increases the standard error of the difference between two sample means decreases, so variable increases SED increases. The reason is that as we increase the sample size, we increase the confidence with which we can place a result related to the difference between two sample means. Additionally, SE is critical in determining whether the population has a proper relationship between males and females. Let's look at the problem to see SEDM;

The standard error for men or variance = σ^2 m= σ^2/nm =20

The standard error for women or variance = σ^2 w= σ^2/nw = 20

SED = $\sqrt{\left(\frac{Variance\ of\ 1st\ sample}{sample\ size\ 1}\right) + \left(\frac{Variance\ of\ 2nd\ sample}{sample\ size\ 2}\right)}$

= $\sqrt{\left(\frac{20}{30}\right) + \left(\frac{20}{48}\right)}$ = 1.040833 that means men and women average amount differ by is 1.04 as SE tells how far from two sample's data distribution.

c) Probability shows the parameters in a model on the sample data are trustworthy. Probability can range from 0 to 1 as O means the model or sample data do not occur, and 1 notifies a particular event. Probability can calculate with a mean and standard deviation that also remembers module 4 about Z-score calculation. Let's figure out the Anger-out possible.

P[$\overline{Xm}$- $\overline{Xw}$ ≤ 16.5666666666667 - 15.7708333333333]

P[$\overline{Xm}$ - $\overline{Xw}$ ≤ 0.796]

$$[(\frac{(Xm)^{-} - (Xw)^{-} - (\mu m - \mu w))}{\sqrt{(\frac{\sigma^2 m}{nm}) + (\frac{\sigma^2 w}{nw})}} \leq \frac{0.796 - (\overline{Xm} - \overline{Xw})}{\sqrt{(\frac{\sigma^2 m}{nm}) + (\frac{\sigma^2 w}{nw})}}$$

P[Z ≤ $\frac{0.796-2}{1.041}$]

P[Z ≤ -1.921537846] = 0.02733197 that means the probability of 0.2% that the percentage of the likelihood, lower probability for the model or case hypothesis as the area to the left of the -1.922. Thus, we could reject the null hypothesis Ho at 0.0273 < 0.05 at the most significate level 0.05.

Chapter 8.26) Descriptive statistics for Anger-In by athletes and non-athletes with excel

Athletes	
Mean	16.68
Standard Error	0.736478
Median	17
Mode	17
Standard Deviation	3.682391
Sample Variance	13.56
Kurtosis	-0.43732
Skewness	0.227878
Range	14
Minimum	10
Maximum	24
Sum	417
Count	25
Confidence Level(95.0%)	1.520016

non-Athletes	
Mean	19.4717
Standard Error	0.671093
Median	20
Mode	22
Standard Deviation	4.88563
Sample Variance	23.86938
Kurtosis	-0.23179
Skewness	0.187655
Range	21
Minimum	10
Maximum	31
Sum	1032
Count	53
Confidence Level(95.0%)	1.346646

To test H_0: $\overline{Xa} - \overline{Xnona}$ than H_1 : $\overline{Xa} \neq \overline{Xnona}$

Let's calculate statistic

SED = $\sqrt{\left(\frac{Variance\ \ of\ 1st\ sample}{sam\,ple\ \ size}\right) + \left(\frac{Variance\ \ of\ 2nd\ sample}{sam\,ple\ \ size}\right)}$

= $\sqrt{\left(\frac{3.682391^\wedge 2}{25}\right) + \left(\frac{4.88563^\wedge 2}{53}\right)}$ = 0.99637623

$$Z_{test} = \frac{\overline{Xa} - \overline{Xnon-a}}{\sqrt{\left(\frac{Variance\ of\ 1st\ sample}{sample\ size\ 1}\right) + \left(\frac{Variance\ of\ 2nd\ sample}{sample\ size\ 2}\right)}} = \frac{16.68 - 19.4717}{1.5387760356070266790187230192 8}$$

The value of the test statistic Z_{test} ~ -2.80222 that tells negative Z-score means is 2.80222 standard deviations below the mean. According to the Z table, 0.2% of values fall below this value.

The Tabulated $Z_{0.025}$ = 1.96

Thus, $|Z_{test}|$ = 2.80222 > $Z_{0.025}$ = 1.96 that mean Z test data point of standard deviation more than tabulated z-score. For test where lower z scores indicated better performance. Also result reject H_0 at 5% level of significance than can tell there is significantly diff between the mean Anger-In for athletes and non-athletes

$\overline{Xm} - \overline{Xw}$ = $16.68 - 19.4717$ = - 2.7916981

So the 95% confidence interval is,

Confidence Interval = (Mean(difference) ± Z0.025* SE(difference))

Confidence Interval for a Difference in Means

Confidence interval = $(x_1 – x_2)$ +/- t*√$((s_p^2/n_1) + (s_p^2/n_2))$

I used to excel to calculate C.I.

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\overline{Xm} - \overline{Xw}$	-2.7916981	0.99637623	60.993605	-4.7840797	-0.79931646

Z-score means the number of standard deviations away from the population mean that a specific data point is a particular data point. It helps to compare which hypothesis is greater than the other if the Z score is more than the mean, far from the mean. Both lower and upper (-4.78, -0.80) bounds are not so far from each other. H_0 will follow or observe a standard normal distribution. Therefore the 95% confidence interval for the difference between the Anger-In score of athletes and non-athletes was found to be (-4.7840797 to -0.7993164). Since the interval only consists of negative values, it can be concluded that athletes do not express their anger as much as the non-athletes.

Chapter 9.27) Descriptive statistics for Anger-Out by male and female with excel.

Male		*Female*	
Mean	16.56667	Mean	15.77083
Standard Error	0.796087	Standard Error	0.597898
Median	16	Median	15
Mode	15	Mode	18
Standard Deviation	4.360349	Standard Deviation	4.142358
Sample Variance	19.01264	Sample Variance	17.15913
Kurtosis	0.333605	Kurtosis	-0.27554
Skewness	0.670366	Skewness	0.537253
Range	18	Range	17
Minimum	9	Minimum	9
Maximum	27	Maximum	26
Sum	497	Sum	757
Count	30	Count	48
Confidence Level(95.0%)	1.628181	Confidence Level(95.0%)	1.202815

We have the means of two groups. A t-test is often used in hypothesis testing that helps to find males and females with a high score behavior to improve an angry mood. We could figure out the mean and average of each of that samples. Above the table shows the average in field male is higher than the average female, but that's only part of the picture. The means only tells us so much because we could have different distributions, and depending on that distribution or the variance within that sample, there could be a statistically significant difference between the two or not, and that's where the t value comes in.

Two sample T hypothesis test;

$\overline{Xm}$: Mean of the Anger-out where gender 1

$\overline{Xw}$: Mean of the Anger-out where gender 2

$\overline{Xm} - \overline{Xw}$: Difference between the two will tell how much signal (point) there is how much difference there.

SED = $\sqrt{\left(\frac{Variance\ of\ 1st\ sample}{sample\ size\ 1}\right) + (\frac{Variance\ of\ 2nd\ sample}{sample\ size\ 2}})$ that tells our data from the mean square that calls variance and so if increase the variance that's going to lower t value. It's like giving no more noise. Other increase the number of samples that will actually increase the signal up to a point and so again the difference between the mean is going to give more signal higher T value. Let's formulate T-value

Module 6: Hypothesis

T-value = $\frac{|\overline{Xa} - \overline{Xnon-a}|}{\sqrt{(\frac{Variance\ of\ 1st\ sample}{sample\ size\ 1}) + (\frac{Variance\ of\ 2nd\ sample}{sample\ size\ 2}}}$

I used to excel to calculate t-value.

To test H_0: $\overline{Xa} - \overline{Xnona}$ than H_1 : $\overline{Xa} \neq \overline{Xnona}$ (without pooled variance)

Hypothesis test results;

Difference	**Sample Diff.**	**Std. Err.**	**DF**	**T-Stat**	**P-value**
$\overline{Xm} - \overline{Xw}$	0.79583333	0.9956087	59.301096	0.79934349	0.4273

An independent samples t-test was conducted to analyze whether the demonstration of anger using verbal or physical medium differed for men and women; at p<.05, no significant difference between the anger-out scores of the two genders was observed, t(59.30)= .799, p= .427. Hence, it was concluded that there is not enough evidence to claim a difference in how much males and females use aggressive behavior to improve an angry mood.

Module 6: Hypothesis

Reference;

What is Sampling Distribution of the Difference Between Two Means? definition and meaning - Business Jargons

Using Confidence Intervals to Compare Means - Statistics By Jim

Standard Error Definition (investopedia.com)

Standard Error in Statistics - Understanding the concept, formula and how to calculate - Machine Learning Plus