TPS REPORT 05

Dispersion , Scaling, and Simple Linear Correlations in Excel

1. Download the TPS 05 Excel file from *www.terevaka.net/nau/ant568/tps01.html*

2. In this report, you'll be analyzing a variety of different hypothetical datasets. You'll begin with the Depression sheet data, which tracks the short-term benefits of an experimental drug. Our ultimate goal will be to determine which of the eight tests is most closely correlated with random variation in the experimental drug.

3. Although we might be interested in the distributions of data in Columns **A** through **I**, creating a histogram for each Test would be a time-consuming process for the sake of visual analysis. Instead, we can calculate skewness and kurtosis measures to get an idea of what our distributions look like. Remember:

$$Skewness = \frac{3(\bar{X} - M)}{\sigma} \qquad \text{and} \qquad Kurtosis = \frac{\sum\left(\frac{X-\bar{X}}{\sigma}\right)^4}{n} - 3$$

where:

- $\bar{X}$ is the mean
- M is the median
- σ is the standard deviation
- X is the individual score
- n is the sample size

Write a formula in Cell **A33** to calculate the skewness of the sample of doses in the experiment between Cell **A2** and Cell **A31**. (Hint: since we're dealing with a sample, remember to use the unbiased estimate for standard deviation.)

- Write your formula in your TPS Report.

Now Copy your formula and Paste it to Cells **B33** through **I33**.

- In your TPS Report, indicate which column of data is the least skewed (i.e., most symmetrical) of all.

4. Based on your measures of skewness, which Test do you think will have the strongest correlation coefficient with the Dose data?

- Write your answer and explain your reasoning in your TPS Report.

5. In Cell **B34**, write a formula to calculate the standard deviation for Test 1 scores. Copy your formula and Paste it to Cells **C34** through **I34**.

- In your TPS Report, indicate which column of data shows test scores most tightly bunched together (i.e., which Test shows the lowest average distance between data points and the mean)?

Add a label to Row **34** in Cell **A34**,

6. In Cell **B37**, write a single formula that can be Copied and Pasted to all Cells between **C37** and **I37** to calculate the Pearson product-moment correlation coefficient between each column of test scores and Doses in Column **A**. (Hint: make sure that when you Copy and Paste your formula from **B37** to other columns that you are always comparing the test data to the Dose data in Column **A**.)

   - Write the formula from **B37** in your TPS Report.

   Copy your formula from **B37** and Paste to Cells **C37** through **I37**.

   Add a label for Row **37** in Cell **A37**.

7. Suppose your job is to decide which of the tests provides the most valid reflection of behavioral changes associated with differences in dosage of the experimental drug.

   - In your TPS Report, explain which test you'd choose and explain your justification.

8. Now you'll focus on the Test-Retest sheet of the same Excel file. Suppose that the test scores in Column **B** have been selected as the most valid reflection of behavioral changes associated with differences in dosage of the experimental drug. Before we get too confident in that particular test, you'll need to determine if this very expensive test is reliable over multiple applications.

   - In your TPS Report, explain what level of test-retest reliability this test provides. Make sure to cite at least one specific statistic to support your explanation.

9. Now turn your attention to the Obama sheet in the same Excel file. Columns **A** through **F** represent different polls taken repeatedly in Los Angeles, aimed to get an idea whether the population that was sampled supports Obama's position(s) on minority-rights issues that have emerged over two terms of presidency. Each cell in Column **A** represents the mean value of poll scores when Poll 1 was administered to 100 randomly-selected residents of L.A., each cell in Column **B** represents the mean value of poll scores when Poll 2 was administered to 100 randomly-selected residents of L.A., etc.

   In Columns **A**, **B**, **C**, **E**, and **F**, higher numbers represent stronger support for Obama's actions, but clearly each poll calculated support on a different scale.

In order to convert Poll 4 to a numeric scale, select all of Column **D** and use Ctrl+H (or Command+H for Mac) to replace values based on the following scale:

Strongly support = 5
Support = 4
Neutral = 3
Disagree = 2
Strongly disagree = 1

10. Our next goal is to convert numeric values on various scales from Columns **A** through **F** to ***z***-scores so that all measurements are comparable on a single scale centered around a zero value. In general, we use the formula:

$$z = \frac{(X - \bar{X})}{\sigma}$$

where:

- $\bar{X}$ is the mean
- σ is the standard deviation
- X is the individual score

In Cell **I2**, write a formula that can be Copied and Pasted to Cells **I3** through **I51** that will calculate ***z***-scores for scores in Cells **A2** through **A51**.

- Write your formula in your TPS Report.

Copy your formula from Cell **I2** and Paste to Cells **I3** through **I51**.

In Columns **J** through **N**, calculate ***z***-scores for all data in Columns **B** through **F**, respectively.

11. In Cells **A53** and **B53**, calculate the median values of raw mean scores for Poll 1 and Poll 2.

- In your TPS Report, provide a detailed explanation as to which poll's median value, after 50 repeated implementations, indicates a stronger level of support for Obama's actions.

12. The ***z***-table at the end of this document offers a useful way to convert ***z***-scores to probability values when we are dealing with data that we suspect approximates a normal distribution. Data won't always be distributed normally. However, when we're calculating the mean of many ($n>30$) random variables (like 100 random poll-participants), and we repeat this random sampling process within the same

population (like Los Angeles), the mean value of our many means should approximate the shape of a normal distribution (remember the Central Limit Theorem?)

As an example of how to use the **z**-table, Cell **J34** should contain a **z**-score of approximately 1.26. If you look up the corresponding area on the **z**-table (look at the cell on the table where the **z** value on the left-most column 1.2 intersects with the **z** value on the top row of 0.06; i.e., 1.2+0.06=1.26), you'll see that the area under a normal curve between the mean and a **z**-score of 1.26 is 0.3962. Or in other words, approximately 40% of the total area under the normal curve is contained in the space between the mean and the **z**-score of 1.26.

This also tells you, from a predictive standpoint, that if our poll data truly approximates a normal distribution, that we would only expect to see **z**-scores higher than 1.26 (or equivalently raw mean scores higher than 9 in Poll 2) approximately 10% of the time. That's because under the normal curve, 50% of the area will be to the left of the mean, and we already figured that 40% of the area under the curve will be between the mean and the **z**-score of 1.26. That leaves only 10% of the area in the tail to the right of our **z**-score.

Using the **z**-table, calculate the predicted percentage of raw mean scores that will fall between 0.76 and 0.82 in Poll 6.

- In your TPS Report, write your answer and explain how you reached that answer.

13. Finally, turn your attention to the Emissions sheet in the same Excel file. Suppose that the scores represent emission scores for two prototypes and 100 randomly-selected vehicles as they performed in two different areas (highway and city).

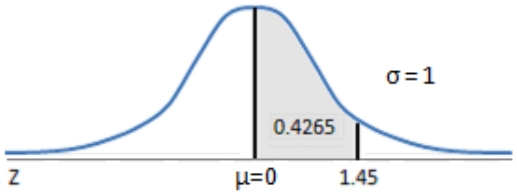
    Specifically, you are asked to determine whether the new Mitsubishi Volanz or the new Ford Ninny performed better when it came to emissions scores.

    - In your TPS Report, write your answer and explain how you reached that answer. (Hint: to assess the Volanz and Ninny on the same scale for Highway, City, and Average driving, calculate **z**-scores).

{ mean, median, stdev.s, correl, pearson }

Areas Under the One-Tailed Standard Normal Curve

This table provides the area between the mean and some Z score.
For example, when Z score = 1.45 the area = 0.4265.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000