Hypothesis Testing Recap

- Hypothesis testing helps an organization:
  1. Determine whether making a change to a process input (x) significantly changes the output (y) of the process.
  2. Statistically determine if there are differences between two or more process outputs.

1

Steps for Conducting a Hypothesis Test

1. Define the problem or issue to be studied.
2. Define the objective.
3. State the null hypothesis, identified as Ho.
   - Note: the null hypothesis is a statement of no difference between the before and after states.
4. The goal of the test is to either reject, or fail to reject Ho.

2

Decisions based on Data

- Every decision that is made has two possible outcomes with respect to the true nature of an event.

1. Correctly identifying the event, or
2. Incorrectly identifying the event

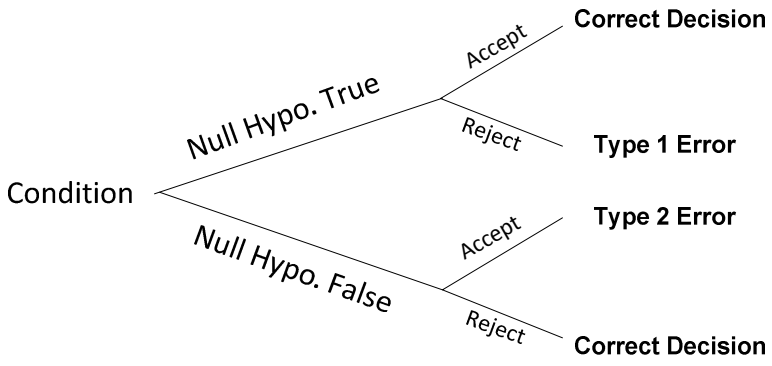
3

Decisions based on Data

	Condition	
Decision	Null Hypothesis True	Null Hypothesis False
Accept Null	Correct	Type 2 Error
Reject Null	Type 1 Error	Correct

4

Decisions based on Data



5

Salk Vaccine Experiment

- Jonas Salk, began his research on polio in 1947 at University of Pittsburgh Medical School.
- Polio (or infantile paralysis) is an infectious disease caused by the poliovirus which results in muscle weakness, especially in the legs.

6

Salk Vaccine Experiment

- Summertime in America in the 1940's and 1950's was a time of concern and worry for many parents as this was the season when children by the thousands became infected with the poliovirus, from which not everyone recovered.

7

Salk Vaccine Experiment

- The 1952 American Epidemic was the worst outbreak in the nations history.
- Of the nearly 58,000 cases reported that year, over 3000 people died and over 20,000 were left with mild to disabling paralysis.

8

Salk Vaccine Experiment

- An experiment concerning the use of the Salk vaccine in the control of polio was conducted in the United States in 1954.
- Children in grades in 1, 2, and 3 were chosen as the representative population, as it was among children of these ages that the disease was most prevalent.

9

Salk Vaccine Experiment

- Ho: “There is no difference in the percentage of 1st, 2nd, and 3rd grade children in the United States contracting polio within a year after being inoculated with the Salk vaccine.”
- Ha:”There is a difference . . .with the Salk vaccine.”

10

Salk Vaccine Experiment

- A completely randomized experiment was conducted involving about 1 million randomly selected children from grades 1 to 3 in areas of the United States where polio had been quite prevalent in the past.
- Of these 1 million, more than 400,000 parents consented to have their children participate in the study.

11

Salk Vaccine Experiment

- A mathematical model could be written as

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

- Where $Y_{ij} = 0$ or 1:

  0 - if no polio is diagnosed,

  1 - if polio is found.

12

Salk Vaccine Experiment

- μ is a constant or general avg. of the 0's and 1's for the whole population.
- τ_j is the treatment effect: $j = 1$ if treated, $j = 2$ if not treated.
- ε_{ij} is a random error associated with a child receiving treatment j.

13

Salk Vaccine Experiment

- To see the hypothesis to be tested, note that

$$p_j = \Sigma_{i=1}^{n_j} Y_{ij} / n_j$$

  is the proportion of children contracting polio who received treatment j.
- n_j is the number who received treatment j.

14

Salk Vaccine Experiment

- The statistical hypothesis to be tested is
  H_o: $p'_1 = p'_2$
  with the alternate
  H_a: $p'_1 < p'_2$
- The sign < indicates that we are interested in showing that the true proportion contracting polio will be less for treatment 1 (vaccine) than for treatment 2 (placebo).

15

Salk Vaccine Experiment Results

Treatment	Sample Size	Number Contracting Polio	Proportion Contracting Polio
Salk Vaccine	**n_1 = 200,745**	**56**	**p_1 = 28 x 10E-5**
Placebo	**n_2 = 201,229**	**142**	**p_2 = 71 x 10E-5**
Totals	**401,974**	**198**	**p = 49 x 10E-5**

16

Salk Vaccine Test Statistic

- The test statistic for testing the hypothesis stated is

  $z = (p_1 - p_2)/ ((pq)(1/n_1 + 1/n_2))^{0.5}$
- Where p is the proportion contracting polio in the total sample of $n_1 + n_2$ children.
  - Note: q = (1-p)

17

Salk Vaccine Experiment

- If one now assumes a risk of 0.001 *(taking one chance in 1000 of rejecting Ho when it is true)* the rejection region for z is given by $z < -3.09$ *(from the normal distribution table).*
- Substituting the values from the data slide yields the following:

18

Salk Vaccine Experiment

z = (28 x 10E-5) – (71x 10E-5) /
((49 x 10E-5x(1- 49 x 10E-5) x
(1/200,745 + 1/ 201,229))^0.5

z = -6.14

- Which yields a very, very strong indication that there is significant evidence to reject reject Ho, and conclude that the vaccine was indeed effective.

19

Salk Vaccine Experiment Conclusion

- In fact, the probability of getting a z as low as or lower than –6.14 when there is really no difference in the two groups is less than one chance in a billion.
- P(z=-6.14) = 4.13E-10 = 0.413 Ppb.
- These amazing results, or course, have been strongly substantiated by the nearly complete eradication of polio in the United States.

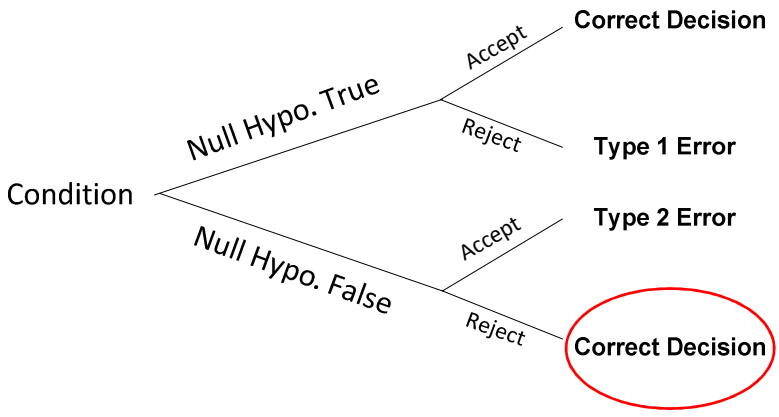
20

Salk Vaccine Experiment Conclusion

- The Salk polio vaccine was approved for widespread use in 1955, and Dr. Salk became world-famous overnight.
- In the following years the incidence of polio in America fell from 18 cases per 100,000 people to less than 2 per 100,000.

21

Decisions based on Data



22

The Space Shuttle Challenger

- On January 28, 1986, the space shuttle Challenger took off on the 25th flight in NASA's space shuttle program.
- Less than two minutes into the flight, the spacecraft exploded, killing all on board.
- A Presidential Commission headed by then Secretary of State William Rogers was appointed to determine the cause of the accident.

23

The Space Shuttle Challenger

- The space shuttle uses two booster rockets to help lift it into orbit.
- Each booster rocket consists of several pieces whose joints are sealed with rubber O-rings, which are designed to prevent the release of hot gases produced during combustion.

24

The Space Shuttle Challenger

- Each booster contains three primary O-rings (totaling six for the craft).
- In the 23 previous flights for which there were data (the hardware for flight #4 was lost at sea), the O-rings were examined for damage.

25

The Space Shuttle Challenger

- The forecasted temperature on launching day of the Challenger was 31deg F.
- The coldest previous launch temperature was 53 deg F.
  - Note: At 53 deg F. analysis of the O-rings revealed 3 Erosion Incidents and 2 Blow-by Incidents.

26

The Space Shuttle Challenger

- The sensitivity of the O-rings to temperature was well known.
- A warm O-ring will quickly recover its shape after a compression is removed, but cold one will not.
- The inability of the O-ring to recover its shape will lead to joints not being sealed and can result in a gas leak.
- It is the combustion of this leaking gas that resulted in the fiery explosion of the Challenger

27

The Space Shuttle Challenger

- There was good deal of discussion among the engineers in the hours preceding the launch.
- Should the launch go on as planned or not?
  - It is important to note that there were no statisticians involved in the discussion.
- A simplified version of one of the arguments made is as follows:

28

The Space Shuttle Challenger

- If we look at the table, there is no apparent relationship between temperature and the probability of damage;
  - *Higher damage occurred at both lower and higher temperatures.*
- Thus, that it was cold on the day of the flight doesn't imply that the flight should have been scrubbed.

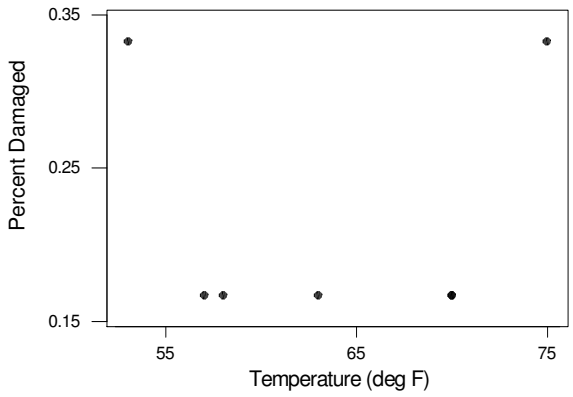
29

O-Ring Damage Table

Ambient Temp (deg F)	No. of O-Rings damaged	Percent defective
53	2	.333
57	1	.167
58	1	.167
63	1	.167
70	1	.167
70	1	.167
75	2	.333

30

Scatter Plot of Percent Damaged vs. Temperature



The Space Shuttle Challenger

- Unfortunately, this analysis was incorrect.
- 16 flights where there was no O-ring damage, were completely ignored and
- the information from those flights was deemed negligible.

32

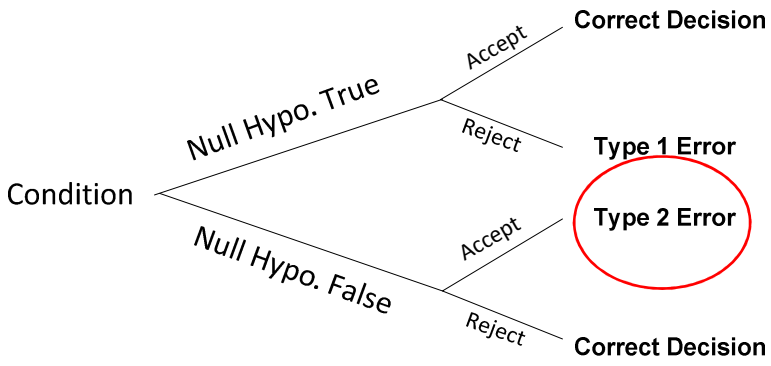
The Space Shuttle Challenger

- Statistically speaking:

  NASA engineers accepted the null hypothesis that the rest of the Space Shuttle launches (16 flights) contained no information regarding the effect that temperature had on O-Ring performance.

33

Decisions based on Data

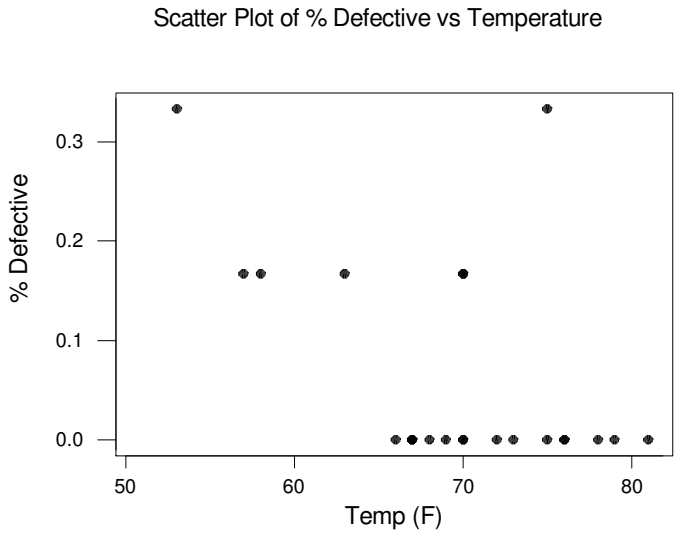


34

The Space Shuttle Challenger

- If flights with high temperature never had O-ring damage, for example, that would clearly tell us a lot about the relationship between temperature and O-ring damage!
- In fact, here is a scatter plot of the O-ring damage vs. temperature for all of the 23 flights for which information was available.

35



36

The Space Shuttle Challenger

- Except for the one observation in the upper right of the previous slide, there is a clear inverse relationship between the probability of O-ring damage and the ambient temperature.
- Lower temperature is associated with higher probability of damage.
  - The outlier is flight 61-A, Oct. 30, 1985.

37

The Space Shuttle Challenger

- A scatter plot of all data would certainly have raised some alarms about the advisability of launching the shuttle.
- Unfortunately, such a plot was never constructed.
- The basic flaw in the analysis of the thermal distress carried out before the launching was the failure to include flights in which were was no O-ring damage.

38

The Space Shuttle Challenger

- It is the conclusion to be drawn, as Sherlock Holmes said, from the "dog that did not bark."
- The Rogers Commission concluded that "a careful analysis of the flight history of O-ring performance would have revealed the correlation of O-ring damage in low temperature."

39

The Space Shuttle Challenger

- Dr. Edward R. Tufte, Yale University, complied a data matrix showing the complete history of temperature and O-ring condition for all previous launches.
- Entries are ordered by the possible cause, temperature, from coolest to warmest launch.

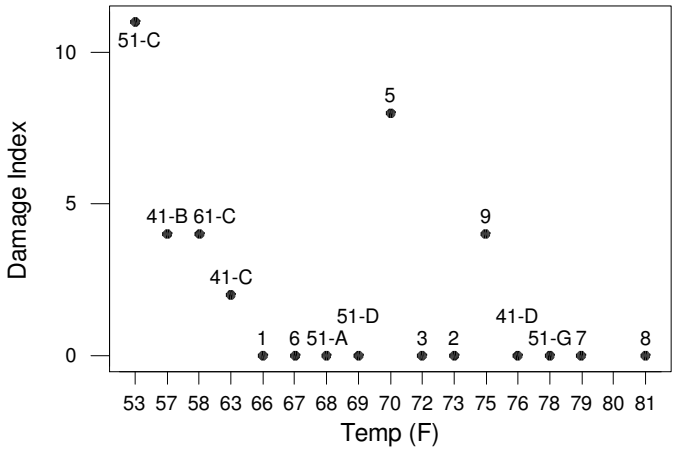
40

Challenger Data Table

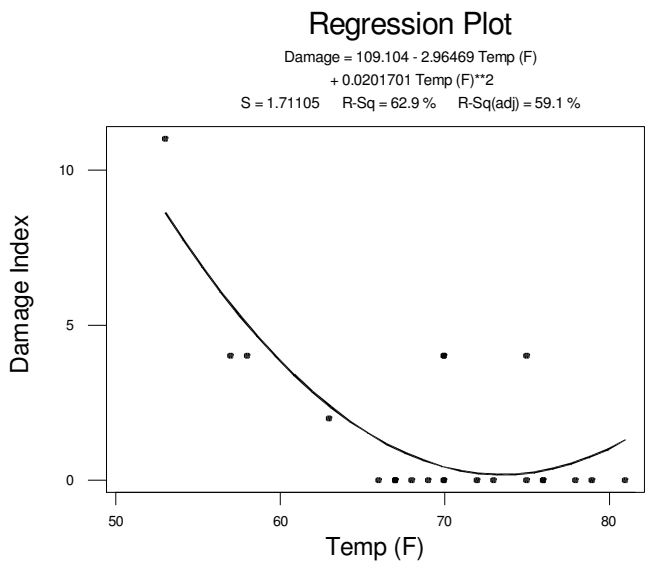
Flight	Date	Temp (F)	Erosion Incidents	Blow-by Incidents	Damage Index
51-C	24-Jan-1985	***53***	3	***2***	11
41-B	3-Feb-1984	57	***1***	0	4
61-C	12-Jan-1986	58	1	0	4
41-C	6-Apr-1984	63	***1***	0	2
1	12-Apr-1984	66	0	0	0
6	4-Apr-1983	67	0	0	0
51-A	8-Nov-1984	67	0	0	0
51-D	12-Apr-1985	67	0	0	0
5	11-Nov-1982	68	0	0	0
3	22-Mar-1982	69	0	0	0
2	12-Nov-1981	70	***1***	0	4
9	28-Nov-1983	70	0	0	0
41-D	30-Aug-1984	70	***1***	0	4
51-G	17-Jun-1985	70	0	0	0
7	18-Jun-1983	72	0	0	0
8	30-Aug-1983	73	0	0	0
51-B	29-Apr-1985	75	0	0	0
61-A	30-Oct-1985	***75***	0	***2***	4
51-I	27-Aug-1985	76	0	0	0
61-B	26-Nov-1985	76	0	0	0
41-G	5-Oct-1984	78	0	0	0
51-J	3-Oct-1985	79	0	0	0
4	27-Jun-1982	80	*	*	*
51-F	29-Jul-1985	81	0	0	0

41

O-Ring Damage Index by Shuttle



42



43

Challenger Conclusions

- In order to grasp how (statistically) significant launching the Challenger at 29 deg F, it is necessary to calculate a normal score for the distance between the average launch temperature and the launch temperature on January 28, 1986.
- Calculating a normal statistic for any variable is to divide an observed difference by an estimate of error.

44

Challenger Conclusions

- The average launch temperature for the previous 24 launches was 70 deg F.
- The corresponding standard deviation was 7.2232.
- The launch temperature on January 28, 1986 was 29 deg F.
- How large is the distance between 29 deg F and 70 deg F, in normalized units?

45

Challenger Conclusions

- Using 70 deg F as the mean (μ) temperature, and 7.2232 as the standard deviation (σ), we can calculate a statistic based upon the Z distribution.

  (29 deg F – 70 deg F)/ 7.2232 = -5.676.
- Assuming once again, a risk of 1/1000, the rejection region for z is given by $z < -3.09$ *(from the normal distribution table).*

46

Challenger Conclusions

- As in the Salk Vaccine experiment, the probability of getting a ***Z statistic*** as low as or lower than –5.676 when there is really no difference in the two groups is approximately 7 chances in a billion.
- P(z=-5.676) = 6.89 E-09 = 6.89 Ppb.

47

Challenger Conclusions

- There were 13 charts prepared for making the decision to launch on January 28, 1986.
- As analytical graphics, the displays failed to reveal a risk that was in fact present.
- *There was no cause and effect establishing O-ring performance as a function of temperature.*

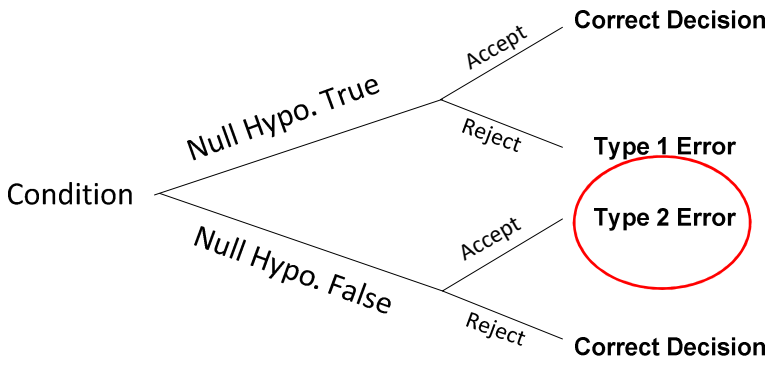
48

Challenger Conclusions

- As presentation graphics, the displays failed to persuade government officials that a cold-weather launch might be dangerous.
  - In fact the probability of there being no O-ring damage was approximately 7 Ppb.

49

Decisions based on Data



50

Challenger Conclusions

- There are right ways and wrong ways to show data; there are displays that reveal the truth and displays that do not.
- If the matter is an important one, then getting the displays of evidence right or wrong can possibly have momentous consequences.

51

Recommendations for Quality Practitioners

- Develop a proficiency in the use of simple and robust techniques for exploratory studies;
- Establish as a primary objective results that are

1. ***interpretable,***
2. ***understandable,***
3. ***and easily communicated to others.***

52