Statistical tests for quantitative variables

Objectives:

1. To be Able to understand the factors to apply for the choice of statistical tests in analyzing the data.
2. To be Able to apply appropriately Z-test, student's t-test and Karl Pearson's Correlation Coefficient.
3. To be Able to interpret the findings of the analysis using these tests.

Color Index:

■	Boys' Slides
■	Girls' Slides
■	Doctors Notes
■	Golden Notes
■	Important
■	Extra

MED439
KING SAUD UNIVERSITY

Research Team

RESEARCH TEAM
438

Choosing the appropriate statistical test

We have 2 types of data :
- A) Qualitative
- B) Quantitative

In this lecture we will focus of the statistical tests to observe the statistical significance of these quantitative data.

We have 2 classes of *Quantitative* data:

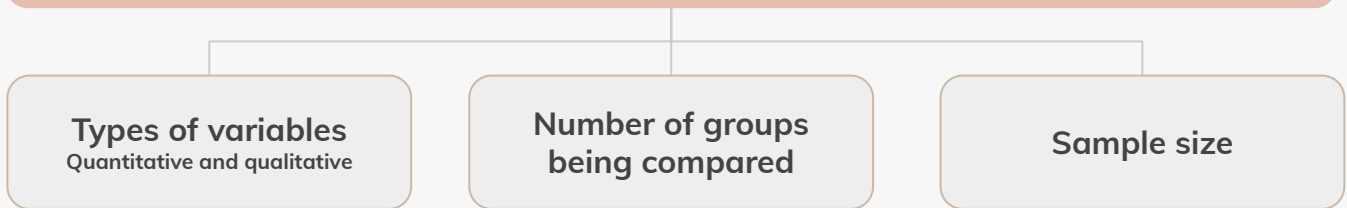1- **Discrete** e.g. whole numbers that we can <u>count</u> them (ex: members of family)

2- **Continuous** e.g. numbers with decimals, you need a device to <u>measure</u> them (ex: BMI, Hemoglobin, etc.)

Choosing the appropriate Statistical test
Based on the three aspects of the data

| **Types of variables** Quantitative and qualitative | **Number of groups being compared** | **Sample size** |

Statistical tests

Test	Z-test (Normal distribution test)	Student's t-test
Study variable Can be multiple, age, gender etc	Qualitative	Qualitative
Outcome variable In any study we have primary and secondary outcomes	Quantitative **Can also be used for Qualitative**	Quantitative
Comparison	• Sample mean with population mean • Two sample means • In z-test we have only these 2 scenarios.	• Sample mean with population mean • Two means (Independent samples) (Two groups) • Paired samples. one group before and after (2 observations for the <u>same</u> group and it's only found in t test not found in z test) • In t-test we have these 3 scenarios, • When the sample size is small you can't use Z-test, you only use t-test.
Sample size	Larger in each group (>30) & standard deviation is known Can only be applicable in large sample size it can't be applicable in small sample size.	Each group <30 (can be used even for large sample size) An advantage in this test is that it can also be applicable in larger sample size.

Example (Comparing sample mean with population mean):

1 The education department at a university has been accused of "grade inflation" (grade increasing) in medical students with higher GPAs than students in general

2 GPAs of all medical students should be compared with the GPAs of all other students

- There are 1000s of medical students, far too many to interview.
- How can this be investigated without interviewing all medical students ?

Comparing sample mean with population mean

Example Cont.

What we know	• The average GPA for **all** students is μ = 2.70. So This value is a **parameter**. (μ always represent population parameter) • To the right is the statistical information for a random sample of medical students?

$\bar{X}$ Sample mean	3.00
S Standard Deviation	0.70
N Sample size	117

Questions to ask	1. Is there a difference between the parameter (2.70) and the statistic (3.00) <u>P</u>arameter related to the <u>p</u>opulation and <u>s</u>tatistics is related to the <u>s</u>ample. 2. Could the observed difference have been caused by **random chance**? 3. Is the difference real (significant)?

Answers	1. The sample mean (3.00) is the same as the population mean (2.70).
	     - The difference is trivial and caused by random chance. (null hypothesis)
	 2. The difference is real (significant). We need P-value to rule out that it is not due to chance.
	     - Medical students are different from all other students. (alternative hypothesis)

Steps

> You don't need to worry about the mathematical equation and calculations, you just need to understand.

Step 1: Make assumption and meet test requirements

- **Random sampling**
  - Hypotheses testing assumes samples were selected using random sampling.
  - In this case, the sample of 117 cases was randomly selected from all medical students.
- **Level of Measurement of GPA is Ratio scale** (Quantitative variable)
  - So the mean is an appropriate statistic.
- **Sampling Distribution is normal in shape** (It's normal in shape because it is >30)
  - This is a "large" sample (n≥100).

Step 2: State the null hypothesis

$H_0 : \mu = 2.7$ (in other words, $H_0 : \bar{X} = \mu$)
- You can also state H_0 : No difference between the sample mean and the population parameter.
- (In other words, the sample mean of 3.0 really the same as the population mean of 2.7 – the difference is not real but is due to chance).
- The sample of 117 comes from a population that has a GPA of 2.7.
- The difference between 2.7 and 3.0 is trivial and caused by random chance.

$H_1 : \mu \neq 2.7$ (or, $H_0 : \bar{X} \neq \mu$) we use two tail hypothesis here because we said +/- in our hypothesis.
- Or H_1 : There is a difference between the sample mean and the population parameter.
- The sample of 117 comes a population that does not have a GPA of 2.7. In reality, it comes from a different population.
- The difference between 2.7 and 3.0 reflects an actual difference between medical students and other students.
- Note that we are testing whether the sample comes from is from a different population or is the same as the general student population.

Statistical tests

Steps Cont.

Sampling Distribution = Z
- Alpha (α) = 0.05 (False positive or level of significance)
- α is the indicator of "rare" events.
- Any difference with a probability <u>less than α</u> is rare and will cause us to <u>reject</u> the H_0 If it's <u>greater than α</u> we <u>won't reject</u> the H_0 .

Critical Region begins at Z = ± 1.96 (we use 1.96 to compare when alpha is 0.05)
- This is the critical **Z** score associated with **α** = 0.05, two-tailed test.
- If the obtained **Z** score falls in the Critical Region, or "the region of rejection," then we would reject the H_0.

Step 4:

- **Use Formula to Compute the Test Statistic (Z for large samples (≥100))**

  The denominator represents the standard error

  $$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{N}}$$

- **When the Population σ is not known, use the following formula:**

  $$Z = \frac{\overline{X} - \mu}{s/\sqrt{n-1}}$$

- **Test the Hypotheses :**
  Substituting the values into the formula, we calculate a Z score of **4.62**.
  Large Z value because the standard deviation is small because of close GPA values.

  $$Z = \frac{3.0 - 2.7}{.7/\sqrt{117-1}} = 4.62$$

- **Two-tailed Hypothesis Test:**

- When **α** = 0.05, then 0.025 of the area is distributed on either side of the curve in area **(C)** (Critical region or rejection region).
- The 0.95 in the **middle section** represents **no significant difference** between the population and the sample mean.
- The cut-off between the middle section and +/- 0.025 is represented by a **Z-value of +/- 1.96.**

  Accept H_0 (not significant)

  Z= -1.96   Z = +1.96

  c   c

  Reject H_0 (significant, not by chance)
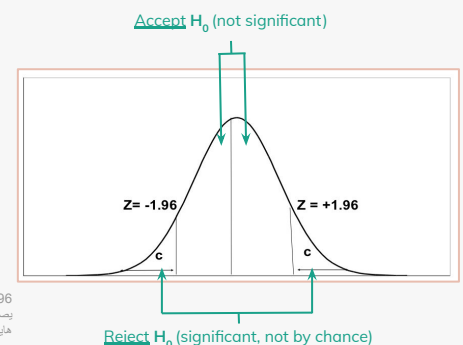
1.96 موجب من اليمين و 1.96 سالب من اليسار هذي حدود منطقة قبول النول هايبوثيسس ، يعني لو كانت القيمة (النتيجة الي تطلع لنا لما نطبق معادلة اختبار الـz على بحثنا) بين هذي الفترة و في هذي المنطقة يصير نقبل النول هايبوثيسس و نقول والله مافي فرق و نقدر نقول الفرق الي طلع كان صدفة و البحث كلام فاضي ، لكن لو حسبنا و طلع القيمة أكبر من +1.96 و لا أصغر من -1.96 نقول وقتها لا و الله هذي منطقة (رفض) النول هايبوثيسس يعني فيه الفرق و ما كان الفرق صدفة كان فرق حقيقي.

Step 5: Make a decision and interpret results

- The obtained **Z** score fell in the Critical Region, **so we reject the H_0.**
  - If the H_0 were true, a sample outcome of 3.00 would be unlikely.
  - Therefore, the H_0 is false and must be rejected.

- Medical students have a GPA that is significantly different from the non medical students (**Z** = 4.62, **P** < 0.05).

Statistical tests

Steps Cont.

Summary:

- The GPA of medical students is significantly different from the GPA of non-medical students.
- In hypothesis testing, we try to identify statistically significant differences that did not occur by random chance.
- In this example, the difference between the parameter 2.70 and the statistic 3.00 was large and **unlikely ($P < 0.05$) to have occurred by random chance.**

Example (Weight loss for diet vs Exercise):

Did dieters lose more fat than the exercisers?	
Diet only	**Exercise only**
Sample mean = 5.9 kg Sample standard deviation = 4.1 kg Sample size = **n** = 42 Standard error = $\textbf{SEM}_1$ = SD/Sample size = $4.1/\sqrt{42} = 0.633$	Sample mean = 4.1 kg Sample standard deviation = 3.7 kg Sample size = **n** = 47 Standard error = $\textbf{SEM}_2 = 3.7/\sqrt{47} = 0.540$ Measure of variability = $\sqrt{[(0.633)^2 + (0.540)^2]} = 0.83$

Steps

Step 1: Determine the null and alternative hypotheses.
- **Null hypothesis**: No difference in average fat lost in population for two methods. Population mean difference is zero.
- **Alternative hypothesis**: There is a difference in average fat lost in population for two methods. Population mean difference is not zero.

Step 2: Sampling distribution: Normal distribution (z-test)

Step 3: Assumptions of test statistic (sample size > 30 in each group)

Step 4: Collect and summarize data into a test statistic. (The outcome variables is quantitative variable)
- The sample mean difference = 5.9 – 4.1 = 1.8 kg and the standard error of the difference is 0.83. So the test statistic: **z** = (1.8 – 0) /0.83 = 2.17 ➔ (Then we find the corresponding number from Z table then subtract it from one) = 1- 0.9850 = 0.015

Step 5: Determine the p-value.
- **Recall the alternative hypothesis was two-sided.**
- p-value = 2 **x** [proportion of bell-shaped curve above 2.17]
- Z-test table => proportion is about 2 **x** 0.015 = 0.03. (We multiplied 0.015 by 2 because we have 2 tails)

Step 6: Make a decision:
- The **p-value of 0.03 is less than or equal to 0.05,** so …
- If really no difference between dieting and exercise as fat loss methods, would see such an extreme result only 3% of the time, or 3 times out of 100.
- Prefer to believe truth does not lie with null hypothesis. We conclude that there is a **statistically significant difference between average fat loss for the two methods.**

TIPS
- Anything greater than 0.05 (5%) we say that it's common, and anything less than 0.05 we say that it is rare. When it is rare it is statistically significant
-When the test statistic is large you will get small p-value and vise-versa

Student's t-test

Test for single mean:	Test for difference in means:	Test for Paired observation:
Whether the sample mean is equal to the predefined population mean?	Whether the CD4 level of patients taking treatment A is equal to CD4 level of patients taking treatment B? (two groups)	Whether the treatment conferred any significant benefit ?

Steps of test for single mean:

1- Questioned to be answered:
Is the Mean SBP (systolic BP) of the sample of 20 patients (small sample size) is 120?
N = 20, $\bar{x}$ = 135, sd = 5, μ = 120

2- Null Hypothesis:
The mean SBP of 20 patients is 120. That is, The sample mean is equal to normal value (population mean).

3-Test statistics: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ $- - -$ **t$_{(n-1)}$ df**

> (df= degrees of freedom) ,
> We subtracted n from 1 because we have 1 sample

4-Comparison with theoretical value
- ➢ if table **t$_{(n-1)}$** < calculated **t$_{(n-1)}$** reject H$_o$
- ➢ if table **t$_{(n-1)}$** > calculated **t$_{(n-1)}$** accept H$_o$

5- Inference

- **Test statistics**
  N = 20, $\bar{x}$ = 135, sd = 5, μ = 120

$t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ =(135-120)/1.12 =13.39

t$_{\alpha}$ = **t$_{0.05, 19}$** = 2.093

Accept H$_o$ if t < 2.093 , Reject H$_o$ if t >= 2.093

Inference :
We reject H$_o$ and conclude that the data is providing enough evidence, that the sample mean BP is significantly higher than the normal value.

Student's t-test

- **Example :**
  - Sample of size 25 (less than 30) was selected from healthy population, their mean SBP = 125 mmHg with SD of 10 mmHg.
    Another sample of size 17 was selected from the population of diabetics, their mean SBP was 132 mmHg, with SD of 12 mm Hg .
  - Test whether there is a significant difference in mean SBP of diabetics and healthy individual at 1% level of significance (Alpha here = 0.1 (1%) not 0.05 (5%) . Generally we use 0.05)

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S^2p}{n_1} + \frac{S^2p}{n_2}}}$$

$\overline{X}_1$ = mean of the first group
$\overline{X}_2$ = mean of the second group
S^2P = pooled variance

$$S^2P = \frac{(n_1 - 1)S^2_1 + (n_2 - 1)S^2_2}{n_1 + n_2 - 2}$$

- Critical t from table is detected
  - At degree of freedom= **n1+ n2-2** (two groups)
  - Level of significance 1% or 5%

State H_o $H_o : \mu_1 = \mu_2$
State H_1 $H_1 : \mu_1 \neq \mu_2$
Choose **α**   **α = 0.01**

$n_1 = 25, \overline{X}_1 = 125, S_1 = 12$
$n_2 = 17, \overline{X}_2 = 132, S_2 = 11$

$$S^2P = \frac{(25-1)10^2 + (17-1)12^2}{25+17-2} = 117.6$$

$$t = \frac{125-132}{\sqrt{\frac{117.6}{25} + \frac{117.6}{17}}} = -2.503$$

Critical t at **df** = 1% (column) **& 40** (row), Df=(25+17)-2=40 level of significance = 2.704
- **Decision:** Since the computed **t** (2.503) is smaller than critical **t** (2.704) so there is no significant difference between mean **SBP** of healthy and diabetic samples at 1 %.

If the alpha here= 0.05 critical t= 2.021 which is smaller than computed t, so we accept Ho (null hypothesis)

- **Uses:**
  To compare the means of two paired samples. Example, mean **SBP** before and after intake of drug.

- **Example:**
  The following data represents the reading of **SBP** before and after administration of certain drug. Test whether the drug has an effect on **SBP** at 1% level of significance. (Alpha here = 0.1 (1%) not 0.05 (5%) . Generally we use 0.05)

$$t = \frac{\overline{d}}{\frac{Sd}{\sqrt{n}}}$$

$$\overline{d} = \frac{\sum di}{n} = \text{mean difference}$$

$$Sd = \sqrt{\frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1}}$$

di = difference (after-before)
Sd = standard deviation of difference
n = sample size (we only have 1 sample but it's before and after)
Critical **t** from table at **df** = **n**-1 (one group)

Student's t-test

- **Answer:**

Serial No.	BP Before	BP After	di After-Before	di²
1	200	180	-20	400
2	160	165	5	25
3	190	175	-15	225
4	185	185	0	0
5	210	170	-40	1600
6	175	160	-15	225
Total			-85 Negative because most SBP has decreased	2475
			$\sum$di	$\sum$di2

$$\bar{d} = \frac{\sum di}{n} = \frac{-85}{6} = -14.17$$

$$Sd = \sqrt{\frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1}}$$

$$Sd = \sqrt{\frac{2475 - \frac{(-85)^2}{6}}{5}} = 15.942$$

$$t = \frac{\bar{d}}{\frac{Sd}{\sqrt{n}}}$$

$$\text{Computed } t = \frac{-14.17}{\frac{15.942}{\sqrt{6}}} = -2.17$$

Critical **t** at **df** = 6 -1 = 5
and 1% level of significance = 4.032

- **Decision:** Since t is < critical **t** so there is no significant difference between mean **SBP** before and after administration of drug at 1% Level.

Here even if we changed the alpha to 0.05 we still accept Ho (null hypothesis) and that's because the standard deviation is high with small sample size.

"Z and t" are the measures of: How difficult is it to believe the null hypothesis?

- **High z & t values:**
  Difficult to believe the null hypothesis - accept that there is a real difference.
- **Low z & t values:**
  Easy to believe the null hypothesis - have not proved any difference.

Karl pearson correlation coefficient

The 3rd concept/technique to deal with quantitative data. It sees if there was any relationship between 2 quantitative variables

Working with two variables (parameter)

All of these variables are quantitative

→ As age ↑, BP ↑

→ As height ↑, weight ↑

→ As age ↑, Cholesterol ↑

→ As duration of HIV ↑, CD4, CD8 ↓

- A number called the correlation measures both the <u>direction</u> and <u>strength</u> of the linear relationship between two related sets of quantitative variables.

- **Types of correlation:**

	Positive	Negative
Definition	Variables move in the <u>same</u> direction	Variables move in <u>opposite</u> direction
Examples	• Height and Weight • Age and BP	• Duration of HIV/AIDS and CD4 CD8 • Price and Demand • Sales and advertisement expenditure

What is meant by **"two related sets"** is that we must measure the data for 1 person.

e.g. Measuring the correlation of someone's Age of and his BP, we can't measure the correlation of someone's age with another one's BP

- **Measurement of correlation:**
  1. Scatter Diagram Assessing the relationship between 2 quantitative variables using graphical presentation. Once you get the diagram, you calculate the karl coefficient to see the pattern.
  2. Karl Pearson's Coefficient of Correlation

Scatter Diaphragm

- **Using the axes**
  - X-axis horizontally
  - Y-axis vertically
  - Both axes meet: origin of graph: 0/0
  - Both axes can have different units of measurement.
  - Numbers on graph are (x,y)

Answer 3 questions when seeing these graphs: What is the direction? Positive or negative? And what is the degree of strength?

Correlation = 0.67 (1)

Correlation = 0.993 (2)
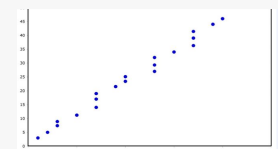
Correlation = - 0.975 (3)

Correlation = 0.003 (4)

1. The direction is positive because 0.67 is positive number and it is going from down to up. The degree of strength is moderate. It is <u>positive moderate correlation.</u>
2. The direction is positive because 0.993 is positive and it is going from down to up. Also, the number is closer to 1. The degree of strength is strong. It is perfect <u>positive</u> linear relationship or <u>strong correlation</u>.
3. The direction is negative because -0.975 is negative number and it is going from up to down. The degree of strength is strong. It is <u>negative</u> perfect or <u>strong correlation</u>.
4. It's very close to zero, very scattered and <u>weak correlation</u> or there is <u>no any correlation</u>.

Karl pearson correlation coefficient

The Pearson r:

We need:

- Sum of the Xs ΣX
- Sum of the Ys ΣY
- Sum of the Xs squared $(\Sigma X)^2$
- Sum of the Ys squared $(\Sigma Y)^2$
- Sum of the squared Xs ΣX^2
- Sum of the squared Ys ΣY^2
- Sum of Xs times the Ys ΣXY
- Number of Subjects   **(N)**

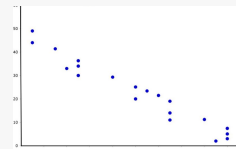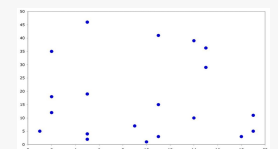$$r = \frac{\Sigma XY - \dfrac{(\Sigma X)(\Sigma Y)}{N}}{\sqrt{\left[\Sigma X^2 - \dfrac{(\Sigma X)^2}{N}\right]\left[\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{N}\right]}}$$

Example:

A sample of 6 children was selected, data about their age in years and weight in kilograms was recorded as shown in the following table. Find the correlation between age and weight.

Serial No.	Age (years) (X)	Weight (Kg) (Y)	XY	X²	Y²
1	7	12	84	40	144
2	6	8	48	36	64
3	8	12	96	64	144
4	5	10	50	25	100
5	6	11	66	36	121
6	9	13	117	81	169
Total	41	66	461	291	742

$$r = \frac{461 - \dfrac{41 \times 66}{6}}{\sqrt{\left[291 - \dfrac{(41)^2}{6}\right]\left[742 - \dfrac{(66)^2}{6}\right]}}$$

- r = 0.759
  Strong direct correlation
  (Positive correlation)

Example

The relationship between anxiety and test scores.

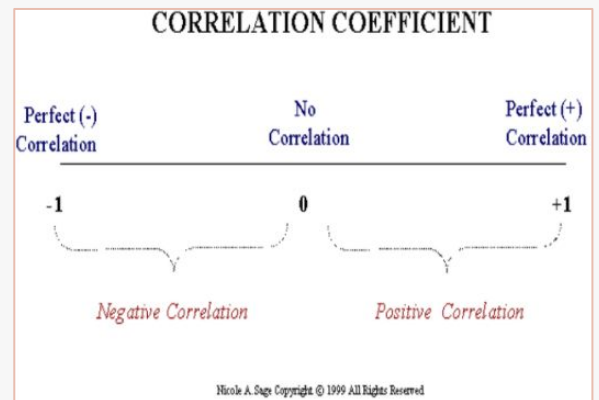| Serial No.	Anxiety (X)	Test score (Y)	X²	Y²	XY
1	10	2	100	4	20
2	8	3	64	9	24
3	2	9	4	81	18
4	1	7	1	49	7
5	5	6	25	36	30
6	6	5	36	25	30
Total	32	32	230	204	129

$$r = \frac{(6)(129) - (32)(32)}{\sqrt{(6(230) - 32^2)(6(204) - 32^2)}} = \frac{774 - 1024}{\sqrt{(356)(200)}} = -.94$$

- r = -0.94
  Indirect strong correlation
  (Strong negative correlation)

Correlation coefficient

- A correlation coefficient (**r**) provides a quantitative way to express the degree of <u>linear</u> relationship between two variables.
- Range: **r** is always between -1 and +1
- Sign of correlation indicates <u>direction</u>:
  - High with high and low with low -> positive
  - High with low and low with high -> negative
  - No consistent pattern -> near zero
- Magnitude (absolute value) indicates <u>strength</u> (-0.9 is just as strong as 0.9)
  - 0.10 to 0.40 weak
  - 0.40 to 0.80 moderate
  - 0.80 to 0.99 high
  - 1.00 perfect

CORRELATION COEFFICIENT

| Perfect (-) Correlation | No Correlation | Perfect (+) Correlation |

-1   0   +1

Negative Correlation   *Positive Correlation*

Nicole A. Sage Copyright © 1999 All Rights Reserved

About " r "

- **r** is not dependent on the units in the problem. Type of variables doesn't affect r value.
- **r** ignores the distinction between explanatory (study variables) and response variables (outcome variables).
- **r** is not designed to measure the strength of relationships that are not approximately straight line.
- **r** can be strongly influenced by outliers.

Limitations

1. Correlation coefficient is appropriate measure of relation <u>only when relationship is linear.</u>
2. Correlation coefficient is appropriate measure of relation when equal ranges of scores in the sample and in the population.
3. Correlation <u>doesn't imply causality</u>:
   A. Using U.S. cities a cases, there is a strong positive correlation between the number of churches and the incidence of violent crime.

   Does this mean churches cause violent crime, or violent crime causes more churches to be built?

   More likely, both related to population of city (3rd variable ,lurking or <u>confounding variable</u>)

   B. Ice-cream sales are strongly correlated with crime rates. Therefore, ice-cream causes crime.
   C. Smoking cause cancer, thus only smoking will cause cancer? No there's always other causes
- Without proper interpretation, causation **should not** be assumed, or even implied.

In conclusion ,

- Z-test will be used for both categorical (qualitative) and quantitative outcome variables.
- Student's t-test will be used for only quantitative outcome variables.
- Correlation will be used to quantify the linear relationship between two quantitative variables.

Tables
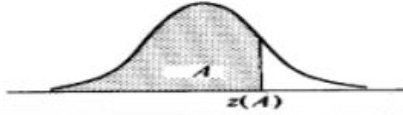
| z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

One-Tailed Test   Two-Tailed Test
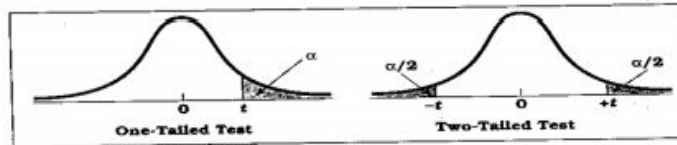
Table D.6   **Percentage Points of the t Distribution** (Source: The entries in this table were computed by the author.)

	Level of Significance for One-Tailed Test								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test								
df	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	63.662
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Summary

- Choosing the appropriate statistical test; based on the three aspects of data: Type of variables, number of groups being compared and sample size

Test	Z-test	Student's t-test
Study variable	Qualitative	Qualitative
Outcome variable	Quantitative	Quantitative
Comparison	Sample mean with population mean & two sample means	Sample mean with population mean, two means Independent samples & paired samples.
Sample size	Large in each group (>30) & standard deviation is known	Each group (<30) can be used even for large sample size
Steps	**Comparing sample mean with population mean** 1. Make assumptions and meet test requirements. 2. State the null hypothesis. 3. State the alternative hypothesis. 4. Select sampling distribution and establish the critical region. 5. Make a decision and interpret the result. **Comparison of two sample means** 1. Determine the null and alternative hypothesis. 2. Sampling distribution: normal distribution (z-test) 3. Assumption of test statistic (sample size > 30 in each group) 4. Collect and summarize data into a test statistic. 5. Determine the p-value. 6. Make a decision.	**Steps for test for single mean** 1. Questions to be answered. 2. Null hypothesis. 3. Test hypothesis. 4. Comparison with theoretical value. 5. Inference. **z-value & t-value** Measures: How difficult is it to believe the null hypothesis - **High z & t values:** Difficult to believe the null hypothesis - accept that there is a real difference. - **Low z & t values:** Easy to believe the null hypothesis - have not proved any difference.

Correlation coefficient provides a quantitative way to express the degree of linear relationship between two variables.
1. Range: r is always between -1 and 1
2. Sign of correlation indicates direction: high with high and low with low (positive), high with low and low with high (negative), No consistent pattern (near zero).
3. Magnitude (absolute value) indicates strength (-0.9 is just as strong as 0.9)

Limitations:
1. Correlation coefficient is appropriate measure of relation only when relationship is linear.
2. Correlation coefficient is appropriate measure of relation when equal ranges of scores in the sample and the population.
3. Correlation doesn't imply causality.

Questions

(1) What are the range of correlation coefficient?

    A) Between -1 to 0   C) Between -1 to 1

    B) Between 0 to 1   D) None of the above

(2) Which of the following is easy to believe the null hypothesis

    A) High z & t values   C) High z & low t values

    B) Low z & t values   D) Low z & High t values

(3) Which of the following has a real difference?

    A) High z & t values   C) High z & low t values

    B) Low z & t values   D) Low z & High t values

(4) Which of the following is used for <u>only</u> quantitative outcome variables?

    A) Z-test   C) Scatter diaphragm

    B) Student's t-test   D) Correlation Coefficient

(5) Which of the following is used for <u>both</u> categorical (qualitative) and quantitative outcome variables?

    A) Z-test   C) Scatter diaphragm

    B) Student's t-test   D) Correlation Coefficient

(6) Which correlation sign of the following indicates a negative direction?

    A) High with high and low with low   C) High with low and low with high

    B) No consistent pattern   D) Near zero

Answers:
1: C | 2: B | 3: A | 4: B | 5: A | 6: C

Questions

(1) Paired sample is measured by?

A) Z-test

B) Ratio

C) Student t-test

D) Correlation coefficient

(2) Which test should be used in this example:
The average test score for an entire school is 75 with a standard deviation of 10. What is the probability that a random sample of 5 students scored above 80?

A) Z- test

B) Student t-test

(3) Which test should be used in this example:
The average test score for an entire school is 75. The standard deviation of a random sample of 9 students is 10. What is the probability the average test score for the sample is above 80?

A) Z- test

B) Student t-test

(4) A neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each to neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug is 1.2 seconds. The mean of the 100 injected rats' response time is 1.05 seconds with a sample standard deviation of 0.5 seconds. Do you think that the drug has an effect on the response time?

A) Yes

B) No

For further explanation!

▷ Question 2+3

▷ Question 4

Answers:
1: C | 2: A | 3: B | 4: A

Thank you for checking our work!

Leaders:

Shuaa Khdary   Sarah AlQuwayz

Abdulrhman Alsuhaibany

Member:

Mona Alomiriny

Note Taker:

Mona Alomiriny   Hamad Almousa

Contact us:
Research4390@gmail.com

RESEARCH TEAM
438
Keep it simple, Keep it focused

Research Team