MATH9102

HYPOTHESIS TESTING – MAKING A DECISION

Sources used in creation of this lecture:
Statistics and Data Analysis, Peck, Olsen and Devore; Discovering Statistics Using R, Field, Miles and Field; Understanding Basic Statistics, Brase and Brase;SPSS Survival Manual, Julie Pallant

Inspecting Ratio/Interval Data

1. Generate plots
   - Generate a histogram with a normal curve showing
   - Generate a Q-Q plot
2. Generate summary statistics
   - Central tendency and dispersion plus measures of skewness and kurtosis and assessing the percentage of standardised scores falling within acceptable limits.
3. Review your statistics and plots to see how far away from normal your data is
4. Report the correct statistics based on your assessment of whether your data can be considered to follow the normal distribution

When you perform a hypothesis test you make a decision:

Decision:

- **reject** H0

OR

- ***fail* to reject H0**

Each could possibly be a wrong decision;

- Therefore, there are two types of errors:
  - Type I
  - Type II

Type I error

The error of ***rejecting H0*** **when** ***H0 actually holds***

The probability of making a Type I error is denoted by **α**.

- **α** *is called th****e significance level*** *of the test*

Key Slide

Type II error

The error of ***failing to reject H0*** **when *H0 is does not hold***

The probability of making a Type II error is denoted by **β**

Key Slide

This is the lower-case Greek letter “beta”.

Power of a statistical test

The power of a statistical test is the probability that it correctly rejects the null hypothesis when the alternative hypothesis holds.

In other words, it’s the test's ability to detect a true effect.

The power of a test is calculated as:

Power=1−β

Key Slide

P-value, α – statistical significance

Key Slide

A probability measure of evidence about $H_{0.}$

$H_{0\,:}$ the assumption that null hypothesis holds.

- There is no effect or no difference.
- It represents the default state or a baseline assumption about the population

Statistical significance in simple terms:

- Given our presumption that the null hypothesis holds
  - α is the probability that the results could have been obtained purely because of chance alone.

P-value, α – statistical significance

Key Slide

The probability (under presumption that H_0 holds) the test statistic equals observed value or value even more extreme predicted by H_a

The ***P*-value** allows us to answer the question:

- Do our sample results allow us to reject H_0 in favour of H_a?
- If that probability (p-value) is small, it suggests the observed result cannot be easily explained by chance.

Key Slide

Statistical Significance

Working with random samples can never have 100% certainty that findings we derive from the sample will reflect real differences in the population as a whole.

Convention is that (for your field of study) there is an accepted level of probability such that it is considered so small that the finding from your sample is unlikely to have occurred by chance or sampling error.

- Normally, that line is drawn at $p=0.05$ or $p=0.01$.
  - In other words, when a statistical test tells us that the finding has less than a 5% or 1% chance of occurring due to sampling error then we tend to conclude that we can be sufficiently confident that the finding is therefore likely to reflect a 'real' characteristic of the population as a whole.
- When this occurs, you can say that your finding is **statistically significant**.

Hypothesis Testing

You start with the assumption (the 'null hypothesis' H_0) that there are no differences or relationships in the population as a whole.

You then state an alternative hypothesis (H_A) that there is a difference or a relationship.

You select a sample and find a difference/relationship in it.

You can then use a variety of tests for statistical significance to work out the probability of the difference/relationship you have found in your sample simply occurring by chance.

Hypothesis Testing

Using the standard level accepted by your domain (e.g. $p \leq 0.05$ or $p \leq 0.01$)

If the probability less than this value then you reject the null hypothesis and thus accept the alternative hypothesis and you can state that your findings are 'statistically significant'.

If the probability is greater this value then you conclude that there is no evidence to reject the null hypothesis and your findings are not 'statistically significant'.

- N.B. This is different from concluding that you have evidence to accept the null hypothesis.
- In these cases, your findings are said to be 'not significant'.

Hypothesis Testing

Caveat

- If we get a p-value of 0.051 should we accept the null hypothesis?
- Should we reject the null hypothesis if we get a p-value of 0.049?
- Need to allow for some flexibility in interpretation

Accepting and Rejecting Hypotheses

A non-statistically significant test result **does not** mean that the null hypothesis is true

- It means you do not have sufficient statistically significant evidence to reject the null in favour of the alternate

Key Slide

Accepting and Rejecting Hypotheses

A significant result **does not** mean that the null hypothesis is false

- It means you have sufficient statistically significant evidence to reject the null in favour of the alternate

Key Slide

What do I do if my data is far away from normal?

Transformations to Improve Normality

Many statistical methods require that the numeric variables you are working with have an approximately normal distribution.

Reality is that this is often not the case.

One of the most common departures from normality is skewness.

There are many different types of transformation available.

Assessing Normality Heuristic approach

According to some researchers, sometimes violations of normality are not problematic for running parametric tests.

Our Heuristic Approach is:

- Inspect Histogram and Boxplot
- Inspect Normal Quantile Plot
  (also called Normal Probability Plot)
- Generate and Inspect Recognized heuristics
  - Standardized skewness and kurtosis
  - Percentage of standardized scores falling within ranges
    (see previous slides and previous lectures for bounds)

Tools for Assessing Normality Goodness of Fit Tests

These will detect whether your dataset deviates from normality

You need to make a decision about whether your data can be treated as approximately normal or not

- Shapiro-Wilk Test (small dataset)
- Kolmogorov-Smirnov Test (larger dataset)
  - *****These are unreliable if your dataset is large**
- Anderson-Darling Test:extension of the Kolmogorov-Smirnov test and gives more weight to the tails of the distribution
- D'Agostino-Pearson Test:combines skewness and kurtosis to assess normality. It's more suitable for larger datasets, typically above 50 observations, and provides a good balance of sensitivity and specificity.

Transforming Data

When a variable is not normally distributed (a distributional requirement for many different analyses), we can create a transformed variable and test it for normality.

Transformation:

- Perform a mathematical operation on each of the scores in a set of data, and thereby converting the data into a new set of scores which are then employed to analyze the results of an experiment.
- If the transformed variable is normally distributed, we can substitute it in our analysis.

Transforming Data

Transformations are obtained by computing a new variable.

- R, either transform and save variable to dataset or apply transformation within whatever statistical formula

There may be data values which are not mathematically permissible.

- For example, the log of zero is not defined mathematically, division by zero is not permitted, and the square root of a negative number results in an "imaginary" value.
- You need to decide what to do about this.

Transformations

Logarithmic transforms

- “Pull in" distribution toward a more normal curve.
- Can be used for right (positive) skew, positive kurtosis, unequal variance, lack of linearity (log10).
- Reduces skewness by compressing high values more than low values.

Square root transformation

- Will bring larger scores closer to the centre
- Can be used for positive skew, positive kurtosis, unequal variance, lack of linearity.
- Especially for count data or when values are close to zero.
- Reduces the impact of extreme values while still preserving the data's original scale.

Transformations

Reciprocal transformation (1/x)

- Reduces impact of large scores
- Can be used for positive skew, positive kurtosis, unequal variance, lack of linearity
- Often used when there are large outliers.

Combination

- **Reverse score** transformation and then apply one of the above transformations
- For negatively skewed/kurtotic data

Transformations

Box-cox transformation

- Flexible family of power transformations that can accommodate various types of skewness by adjusting the power parameter λ lambda.
- λ =1: No transformation is applied; the data remains unchanged.
- λ=0: The transformation becomes a natural log, ln(y)ln(y), which is effective for reducing right skewness in positively skewed data.
- 0<λ<1: A power transformation that can moderately reduce right skewness while still compressing larger values
- λ<0: Negative values of λλ apply reciprocal transformations that can be effective for highly skewed data with large outliers
- λ>1: A stronger power transformation that might be useful for left-skewed data

Transformations

Two quick videos

Short and to the point:

https://www.youtube.com/watch?v=5571wc0iWCI&ab_channel=Udacity

More involved:

https://www.youtube.com/watch?v=kV9_qFR3ORg&ab_channel=PatrickManapat

Transforming Data

If transformation does not bring data to a normal distribution, the investigators might well choose a **nonparametric** procedure that does not make any assumptions about the shape of the distribution.

Good Information Source

Andy Field's The Beast of Bias

https://www.discoveringstatistics.com/repository/exploringdata.pdf