Error & One-Sample Tests

PSY 3307

Jonathan A. Pedroza, PhD

Cal Poly Pomona

2022-02-24

Inflated Error Rates

- **familywise error rate** or experimentwise error rate is the change in probability of making a type I error

$$familywise\ error = 1 - .95^n$$

$$0.40 = 1 - 0.95^{10}$$

- 40% chance of making a type I error (false positive)

Inflated Error Rates

- **Bonferroni Correction** is a correction to make the alpha accurate for the number of tests you are making
  - alpha is your pre-determined alpha/probability, k is the number of tests

$$P_{crit} = \frac{\alpha}{k}$$

$$P_{crit} = \frac{.05}{4}$$

$$P_{crit} = .0125$$

Statistical Power

- opposite issue from Type II error is known as statistical **power**
  - chance of not finding a true significant finding (false negative)
- false negative is beta
  - .2 of a probability is the norm for beta or power of .8
  - 80% chance that you detect an effect that is truly there

$$1 - \beta$$

Power

- we can think about increasing power with the following information
  - alpha, sample size, size of the effect desired
- calculate the power of the test (beta)
  - we have alpha and sample size
- calculate sample size to achieve a given level of power
  - we have alpha and beta

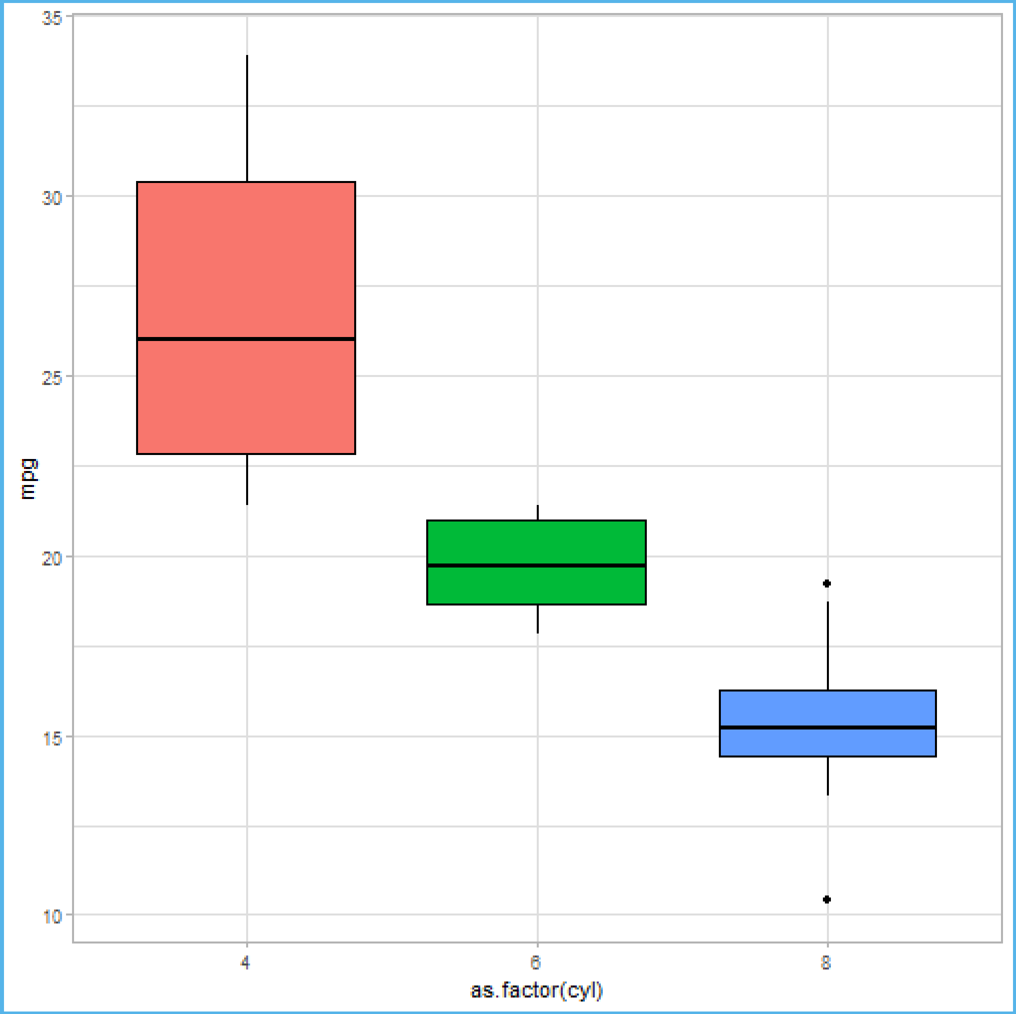
Power

- Power can be determined a priori (before analyses) or post hoc (after analyses)
- Ways to improve power
    - increase sample size
    - potentially use a one-tailed test over a two-tailed test
    - use parametric tests over nonparametric tests

Confidence Intervals & Statistical Significance

- if your confidence intervals barely touch on the ends then you'll have a significance level approximately at .01
- if there is a gap between your confidence intervals then you have a significance level less than .01
- slight overlap between your confidence intervals and you'll have a significant difference less than .05
- if confidence intervals overlap a lot then you have no difference between groups

Example



Sample Size & Statistical Significance

- even with means and sd at the same level for groups/conditions
  - with a large enough sample size, you'll see a difference
- large samples, small differences can be significant (evidence of significant effect)
  - small samples, large differences can be non-significant (no evidence of significant effect)

Reporting Significance Tests

- 95% confidence intervals are in brackets []
    - The age of participants (M = 23.5, SD = 1.34, 95% CI [21.1, 24.6])
- The age of students in PSY 3307-01 and PSY 3307-03 are significant different, $p < .05$ (using tables)
    - The age of students in PSY 3307-01 and PSY 3307-03 are significant different, $p = .034$

Effect Sizes

- there are many problems with NHST, but many of the findings from NHST don't tell us how strong of an effect there is
  - most statistics tell us we passed the critical value/threshold
- a standardized measure of an effect/strength of a relationship is known as an **effect size**
  - standardized = compare to other effects from other analyses or other studies
- many types of effect sizes, but let's focus on the most useful for this class
  - Cohen's d

Cohen's d/Hedges' g

- when samples are small (N < 20) --> Hedge's g
  - when greater than 20, both are roughly the same

$$g = \frac{\overline{X}_1 - \overline{X}_2}{SD_{pooled}} * \frac{N-3}{N-2.25} * \sqrt{\frac{N-2}{N}}$$

$$\hat{d} = \frac{\overline{X}_1 - \overline{X}_2}{S}$$

- helpful primarily for use in independent samples t-tests
  - where we are comparing two groups
- when groups have equal SD
  - use the SD

Cohen's d/Hedges' g

- when groups have unequal SD
    - use the SD for the control group/baseline OR
    - pool the SD of the two groups (if they are independent of one another)

$$S_p = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}}$$

- we'll cover this a lot more when conducting statistics that are not standardized

Logic of Hypothesis Testing

- At least for the current statistical testing we will be conducting (z-test)
- We are interested whether our DV score in our sample is representative of the population (HO) or if our sample's DV score is significantly different from the population DV score (H1)
- This means we will be using:
    - probability
    - region of rejection
    - criterion
    - critical value

Example

- We are interested in knowing if CPP students eat more or less fast food than the average CSU student H0: H1:
- The average number of a times a week that a CSU student eats fast food is 3.7, sd is 1.2
- Sampling from CPP, we find out the average CPP student eats fast food 2.4 times a week
- We would then conduct a z-test to see if our sample is statistically significantly different from CSU students
- Before doing any analyses, we would create our criterion for what counts as a significant finding

The z-Test Assumptions

1. Randomly selected a sample
2. DV is somewhat normally distributed in population
    - and is ratio or interval scale
3. Know population mean of raw scores under another condition of the IV
4. Know population standard deviation

Sampling Distribution of Two-tailed Test

1. Create sampling distribution of means from population raw scores
   - This will be what our H0 states
2. Identify what the population mean is for H0
3. Select an alpha
   - **alpha** greek letter for criterion probability (e.g., .05)
4. Identify regions of rejection
   - One-tailed or two-tailed test
5. Determine critical value
   - two-tailed is $z = 1.96$

Compute the z-score

$$z_{obt} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

- Need the standard error of the mean before getting the z-score
- z is now obtained from the data, that is why it is z obt in the formula above
- Choosing some random values
  - mu is 3.7
  - sigma is 1.2
  - N is 100
- We would calculate xbar from our sample, let's just call it 2.4

```r
# standard error calculation
1.2/sqrt(100)
```

```
## [1] 0.12
```

```r
# se is .12

(2.4 - 3.7)/.12
```

```
## [1] -10.83333
```

```r
# z-score is -10.83
```

Comparing Obtained z-value to Critical Calue

- We know that the critical value is +-1.96
- Since our value is outside of the critical value in the region of rejection
  - we can reject that CPP students eat fast food as much as the rest of the CSU system students
- We have rejected the null hypothesis

Interpretation of STATISTICALLY Significant & Nonsignificant Results

- If you reject the null, you have statistically significant findings
  - H1 is supported, there is a relationship/there are differences
- If you retain the null, you don't have statistically significant findings
  - H0 supported, there is no relationship/there are no differences

What Does a Statistically Significant z-test Finding Represent

- Significant MEANS NOTHING
- Statistically significant indicates you reject the null and your sample is different from the population (for z-tests)
- We can't prove that H0 is false
- We don't know the exact population mean represented by our sample
    - potential **sampling error**, your sample may not be representative of the population

What Does a Nonsignificant z-test Finding Represent

- our sample is representative of the population
    - CPP students eat fast food as much as CSU students
- Don't say insignificant
    - best way of stating this is: "there was no evidence supporting that CPP students are statistically different from CSU students"
- We didn't find evidence of a statistically significant difference/relationship
- We simply have failed to reject the null hypothesis

z-Test Summary

- Create hypothesis/es
- Set up sampling distribution, select alpha, location region of rejection, determine critical value
- compute xbar, standard error and z-score obtained from population mean and standard deviation
- compare obtained z-score to critical value
    - statistically significant = reject H0
    - nonsignificnat = retain H0
    - don't draw conclusions

The One-Tailed Test

- predict scores in a direction
- interested in whether DV scores are higher or lower, not both
- the alpha still needs to be determined prior to analyses

Example

- H0:
- H1: CSUDH students will eat more fast food than CSU students
- CSUDH students eat 4.1 times a week
- CSU students still eat 3.7 times a week, sigma (population standard deviation = 1.2)

```r
# standard error calculation
1.2/sqrt(100)
```

```
## [1] 0.12
```

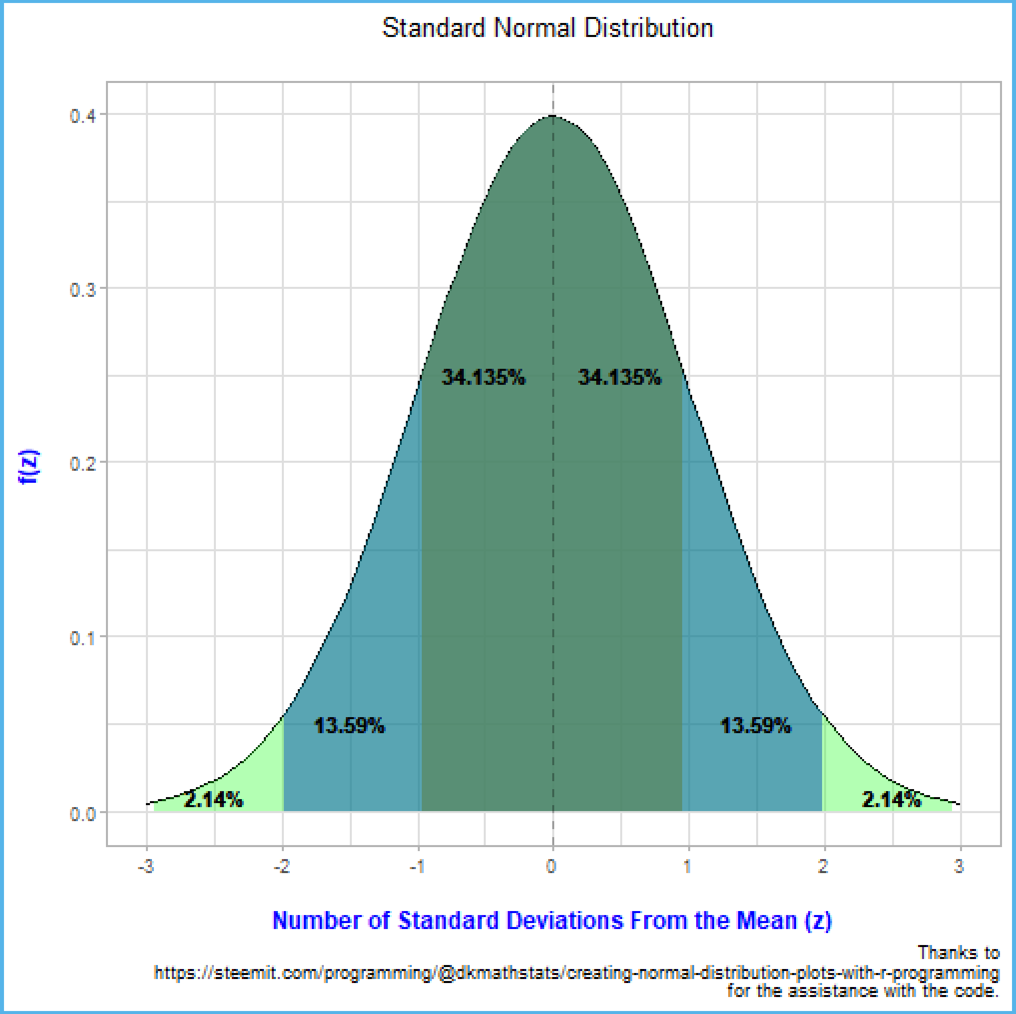
```r
# se is .12

(4.1 - 3.7)/.12
```

```
## [1] 3.333333
```

```r
# z-score is 3.33
```

Using The z-Table



Using The z-Table

- since the z-distribution is based on proportions, you can place the corresponding z-score to the proportion of the distribution that is being covered.

z-Scores

- 1.8
- -.4
- 1.96
- -2
- 0

What is a One-Sample t-test

- It's pretty similar to a z-test
  - t-test used more often in behavioral research
- z-test requires we know population standard deviation
  - often not possible in behavioral research
- uses unbiased estimators (N - 1 formulas)
- computes something like the z-score for our sample mean
  - t-score

One-Sample t-test

- parametric test for when the population standard deviation is unknown
- still compares the sample mean to the population mean

Steps to One-Sample t-test

1. Statistical Hypotheses
   - what is the population mean and is your sample mean different from that population mean
   - H0: sample mean equals the population mean
   - H1: sample mean is different from the population mean
2. Select an alpha
3. Check assumptions
   - Outcome needs to be continuous (interval or ratio scale)
   - Population score forms a normal distribution
   - variability of raw score population is estimated from the sample

Steps to One-Sample t-test

- All we need to know is the t critical value and if the t obtained value is within the regions of rejection

Steps to a z-test/One-Sample t-test

- get population variance/standard deviation (z-test)
- get estimated variance/standard deviation (t-test)
- get the standard error (SE) of the mean (z-test)
- get the **estimated** SE (t-test)
- calculate the score by subtracting the population mean from the sample mean and dividing by the SE
    - either obtained z or t value

Changes between the z-test and t-test

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1}}$$

Changes between the z-test and t-test

$$\sigma_{\overline{X}} = \frac{S}{\sqrt{N}}$$

$$z_{obt} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

$$S_{\overline{X}} = \frac{S}{\sqrt{N}}$$

$$t_{obt} = \frac{\overline{X} - \mu}{S_{\overline{X}}}$$

Example

```
##  [1] 4.770670 5.271329 6.899812 5.598090 4.219022 6.828034 6.046497 2.92124
##  [9] 4.993870 5.948126
```

1. Calculate the Variance and Standard Deviation

2. Calculate the SE

3. Compute t

```r
# population mean is 10

(4.770670 + 5.271329 + 6.899812 + 5.598090 + 4.219022 +
   6.828034 + 6.046497 + 2.921249 + 4.993870 + 5.948126)/10
```

```
## [1] 5.34967
```

```r
# 5.35 mean

(4.770670-5.35)^2 + (5.271329-5.35)^2 + (6.899812-5.35)^2 + (5.598090-5.
  (6.828034-5.35)^2 + (6.046497-5.35)^2 + (2.921249-5.35)^2 + (4.993870-
```

```
## [1] 13.1375
```

```r
# 13.14 sum of squares

10 - 1
```

```
## [1] 9
```

```r
# df is 9
```

```r
13.14/9
```

```
## [1] 1.46
```

```r
# 1.46 variance

sqrt(1.46)
```

```
## [1] 1.208305
```

```r
# standard deviation is 1.21
```

```r
1.21/sqrt(10)
```

```
## [1] 0.3826356
```

```r
# se is .38

# compute t
(5.35 - 10)/.38
```

```
## [1] -12.23684
```

```r
# t value of -12.24
```

t-distribution & Degrees of Freedom (df)

- we will now be working with the t-distribution
  - this also means we'll be working with a t-table
- **t-distribution** is the sampling distribution of all values of t when samples of a particular size (differing N size) are selected from the raw score population in the null hypothesis

t-distribution & Degrees of Freedom (df)

- higher values on the t-distribution are to the right of the population mean, lower values to the left of the population mean
- t-tests also have regions of rejection
- doesn't always represent a perfectly normal distribution
  - dependent on N value
  - larger the sample the more normal the distribution looks
- the different shapes are important because our regions of rejection will look different dependent on the sample size

t-distribution & Degrees of Freedom (df)

- the distribution changes based on the sample size, which then means that the 5% of the regions of rejection and critical value change
- remember to be conservative about estimating variance and SD, we have been using `N - 1`
- the name of that is the **degrees of freedom** or df
  - number of scores in a sample that reflect the variability in the population
  - determines shape of sampling distribution when estimating standard deviation for the population

t-distribution & Degrees of Freedom (df)

- since the df is the sample size - 1, the larger the df, the closer to resembling a normal distribution our data becomes
    - df of 120+ is the same as a z-distribution

Using the t-table

- the t-table is different from the z-table
- has df, \alpha = .05 and \alpha = .01
    - this is dependent on our sample size - 1, and what our alpha is `a priori`

t-table

- we need to figure out our `t critical value`
- we need our sample size, and a decision on what alpha we want to use (.05 or .01)
- since not all df are listed, if your df is between two values, a statistically significant finding is a t-value larger than the larger df and smaller than the smaller df

Examples

- sample size = 200
  - alpha = .05
- sample size = 90
  - alpha = .05
- sample size = 37
  - alpha = .01

t-test Interpretation

- If a statistically significant finding is found
  - your sample is significantly different from the population in whatever the outcome was

One-tailed test

- if you know if your sample will do better or worse than the population, you'd use a one-tailed test
- Example: you know that your sample will get higher grades than the population