Error & One-Sample Tests

PSY 3307

Jonathan A. Pedroza, PhD

Cal Poly Pomona

2022-02-22

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
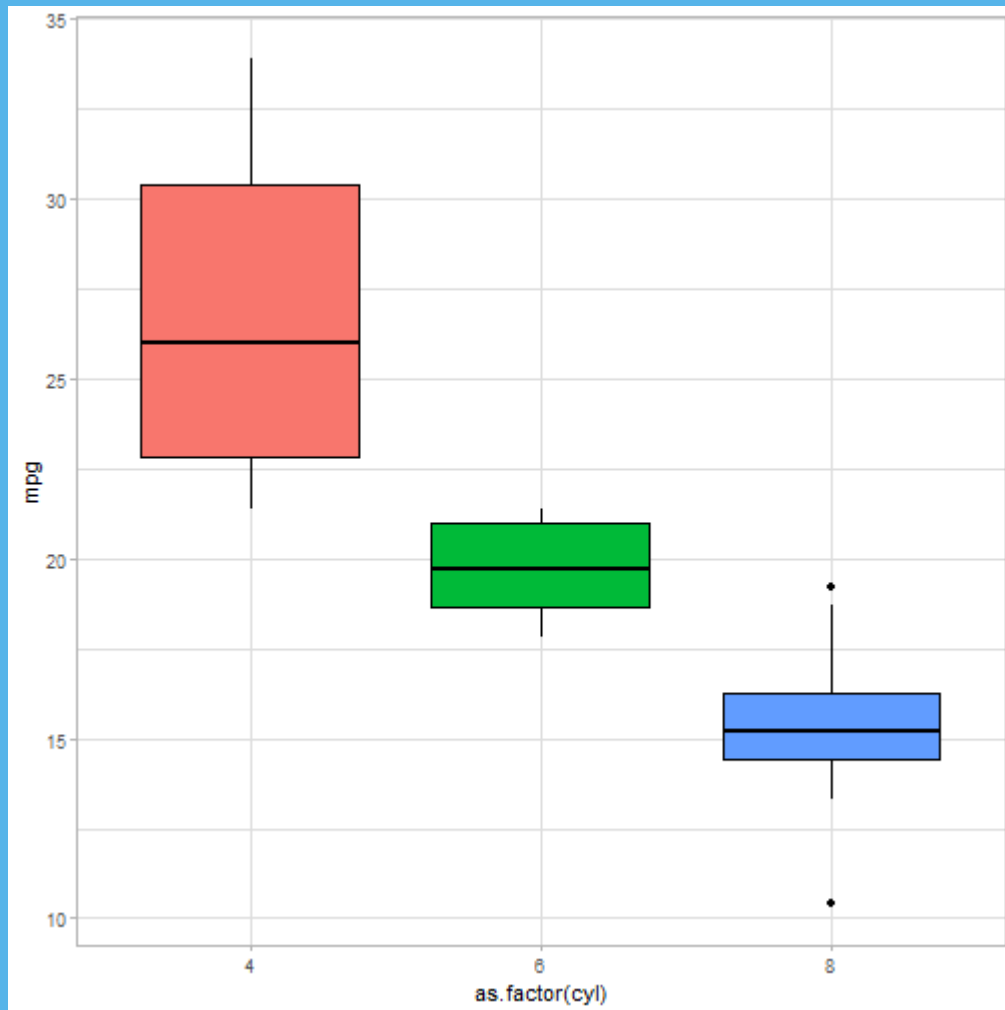
Sample Size & Statistical Significance

- 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

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

```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
```