COGS2020

WEEK 9: ASSUMPTIONS & ONE-WAY ANOVAS

Quick experimental design review

In experiments we generally want to see if one thing causes a change in another thing

- **Independent variable (IV)**: the variable we manipulate to see if it causes an effect
- **Dependent variable (DV)**: the outcome variable we measure/observe to see if there is change depending on the independent variable

- *Example: Does coffee improve memory performance?*
  - *IV is coffee: Does **coffee vs no coffee** cause an effect in memory?*
  - *DV is memory: Memory can be measured via a memory test, we want to see whether coffee vs no coffee makes a change in the memory test scores*

Independent variable levels pt. 1

- So far we have looked a lot at single sample tests in our examples:
  - Normal test
  - Z test
  - One sample t test
- These tests compare your singular sample (more specifically sample mean) to a known or hypothesised population mean
  - E.g. Does this year’s autumn temperature (our sample) differ significantly from previous years (known population mean)
  - E.g. Are MQ uni students (our sample) significantly taller than 170cm (hypothesised population mean)?
- What if we want to compare 2 or more samples to one another (and not just 1 sample to a known number)?

Independent variable levels pt. 2

- 2+ sample tests compare the means of 2 or more groups to see if they are significantly different from one another
- Example: Are MQ uni students (our sample) significantly taller than UNSW students (our other sample)?
  - Note there is no known population value that we are comparing against, we are comparing 2 unknown things

Independent samples t test – practice

I am curious whether there are height differences are different universities.

- I have recorded the height of 100 students who study exclusively at MQU, and 100 students that study exclusively at UNSW
- In my MQU sample, the average height was 170cm with a standard deviation of 5cm
- In my UNSW sample, the average height was 166cm with a standard deviation of 5cm
- Research Question: Are MQU students significantly taller than UNSW students?

Work through each of the steps from the previous slide, and conduct an appropriate statistical test. Highly recommend doing all the steps by hand on paper or on your devices. To calculate the exact p value – you can use R afterwards. Compare your manual results with t.test results()

How do we create a hypothetical null model?

1. **Define the null and alterative** (what is considered no effect in the population?). (H0 and H1 in terms of the population parameter θ)
2. **What does the null look like in the population** (what would the population look like **if null is true** – define the key moments)

   X ~ *N*(parameters if null was true)

3. **How are we estimating the population mean (θ-hat)?** (eg. sample mean)
4. What is the **sampling distribution of this estimate** (given null is true)? **Building the sampling distribution null model**..

   $\bar{x}$ ~ *distribution*(sampling dist parameters is null was true)

5. Do we have enough information about the population to create this null model distribution? If not, what alternative distribution could we use?
6. What is the **sample/test statistic (θ-hat$_{obs}$)** based on the distribution you use to model the null?
7. Set up your rejection zone(s). What is your alpha? What are/is your critical value(s)?
8. **Where does sample/test statistic lie when put into your null model (is it in the rejection zone or not)**? Is it likely to occur or unlikely?
   - If in rejection zone – it is unlikely we got our test statistic given this model, therefore we reject this null model
   - If not in rejection zone – it is likely we got our test statistic given this null model, therefore we do NOT reject this null model

2 sample t test on uni student heights, pt. 1

1. **Define the null and alterative**

X will represent the RV that generates MQU student heights

Y will represent the RV that generates UNSW student heights

H0: $\mu_X = \mu_Y$ OR H0: $\mu_X - \mu_Y = 0$

H1: $\mu_X > \mu_Y$ H1: $\mu_X - \mu_Y > 0$

2. What does the null look like in the population (what would the population look like **if null is true** – define the key moments). Draw out the distribution and label the key moments.

$X \sim N(\mu_X = \mu_Y , \sigma = ?)$

$Y \sim N(\mu_Y = \mu_X , \sigma = ?)$

If H0 true X and Y distributions should be centred at the same spot

$X - Y = W$

$W \sim N(\mu_w, \sigma)$

3. **How are we estimating the population mean (θ-hat)?**

$\hat{\mu}_X = \bar{x} = 170$

$\hat{\mu}_Y = \bar{y} = 166$

$\hat{\mu}_W = \hat{\mu}_X - \hat{\mu}_Y$

$\hat{\mu}_w = 170 - 166$

$= 4$

2 sample t test on uni student heights, pt. 2

4. What is the **sampling distribution of this estimate** (given null is true)? **Building the sampling distribution null model.** Draw out the distribution and label the key moments.

$\bar{X} - \bar{Y} = W$

$\bar{W} \sim N(\mu = 0, \sigma = ?)$

5. Do we have enough information about the population to create this null model distribution? If not, what alternative distribution could we use?

No, we do not know what the population standard deviation is. We will have to use a t distribution instead.

$t \sim t(df)$

$df = n1 + n2 - 2$ *equal variance*

$= 100 + 100 - 2 = 198$

Yes – we have both the key parameters to make our null model. We also have the information we need to convert our data into z scores.

6. What is the **sample/test statistic (θ-hat$_{obs}$)** based on the distribution you use to model the null?

$$t_{obs} = \frac{\bar{w} - \mu_{\bar{w}}}{s_{\bar{w}}} = \frac{(\bar{x} - \bar{y}) - (\mu_{X_{H_0}} - \mu_{Y_{H_0}})}{\sqrt{s^2_{pooled}\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

standard error of w

$$= \frac{(\bar{x} - \bar{y})}{s_{pooled}\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

$$s_{X-Y} = s_{pooled}$$

$$s^2_{pooled} = \frac{(n_X - 1)s^2_X + (n_Y - 1)s^2_Y}{n_X + n_Y - 2}$$

2 sample t test on uni student heights, pt. 3

5. Set up your rejection zone(s). What is your alpha? What are/is your critical value(s)?
6. **Where does sample/test statistic lie when put into your null model (is it in the rejection zone or not)**? Is it likely to occur or unlikely?
   - If in rejection zone – it is unlikely we got our test statistic given this model, therefore we reject this null model
   - If not in rejection zone – it is likely we got our test statistic given this null model, therefore we do NOT reject this null model

Do this by hand (e.g. use the pt() function in R), and then compare to your t.test function results

**Note if you do by hand and use rnorm() to generate your data, you will need to rescale the data first to ensure that your means and standard deviations are as exactly given in the question. Solutions in the R worksheet.*

T test assumptions

- **Independence** – observations are independent of each other
- **Normality** – data is normally distributed, can eyeball and use statistical tests
- **Scale of Measurement** – dependent variable is measured on a continuous scale

- **Equal variance** – in a 2 sample t test, ideally the variance should be equal, if not equal, the Welch’s t test is used instead

Shapiro-Wilk Test

- The **Shapiro-Wilk test** is a statistical test used to assess whether a dataset is normally distributed.
- **Hypotheses:**
  - **H_0**: The data are normally distributed
  - **H_1**: The data are not normally distributed
    → If **$p < .05$**, reject H_0 → data likely **not** normal

Shapiro-Wilk Test: Sample Size

Small samples (e.g., n < 30):

- The test has low power — meaning it's less likely to detect non-normality, even if the data are not normal.
- You might get a high p-value even if the distribution is *not quite* normal.
- Visual inspection (e.g., Q-Q plot, histogram) is especially important to supplement the test.

Large samples (e.g., n > 100):

- The test becomes very sensitive, often detecting minor and inconsequential deviations from normality.
- You might get a significant result (p < .05) even when the data are "normal enough" for practical purposes.
- In these cases, a significant result doesn't necessarily mean your analysis is invalid.

Best practice:

- Always combine the Shapiro-Wilk test with graphical methods:
    - Histogram
    - Q-Q plot (qqnorm(), qqline())

One-Way ANOVA

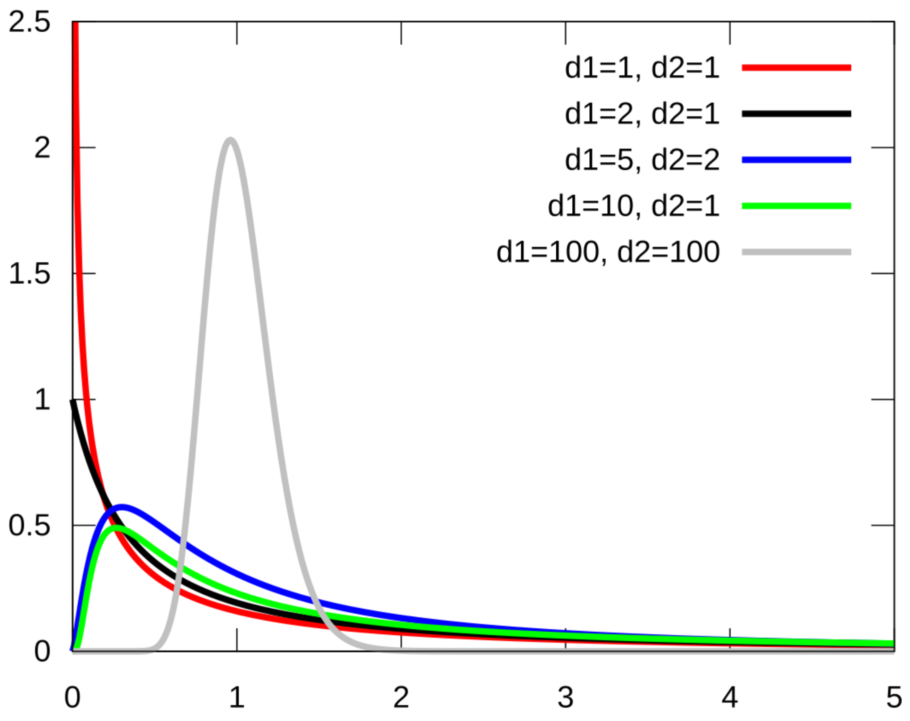
- A One-Way ANOVA tests whether three or more group means are significantly different from each other.
- ANOVA compares two sources of variation:
  - Within group variation: Reflects variation of individual scores around their group mean.
  - Between group variation: Reflects variation of group means around the grand mean.
- F distribution:
  - F-ratio = Between-group variance / Within-group variance
  - If the F-value is large, it suggests that group means differ more than expected by chance.

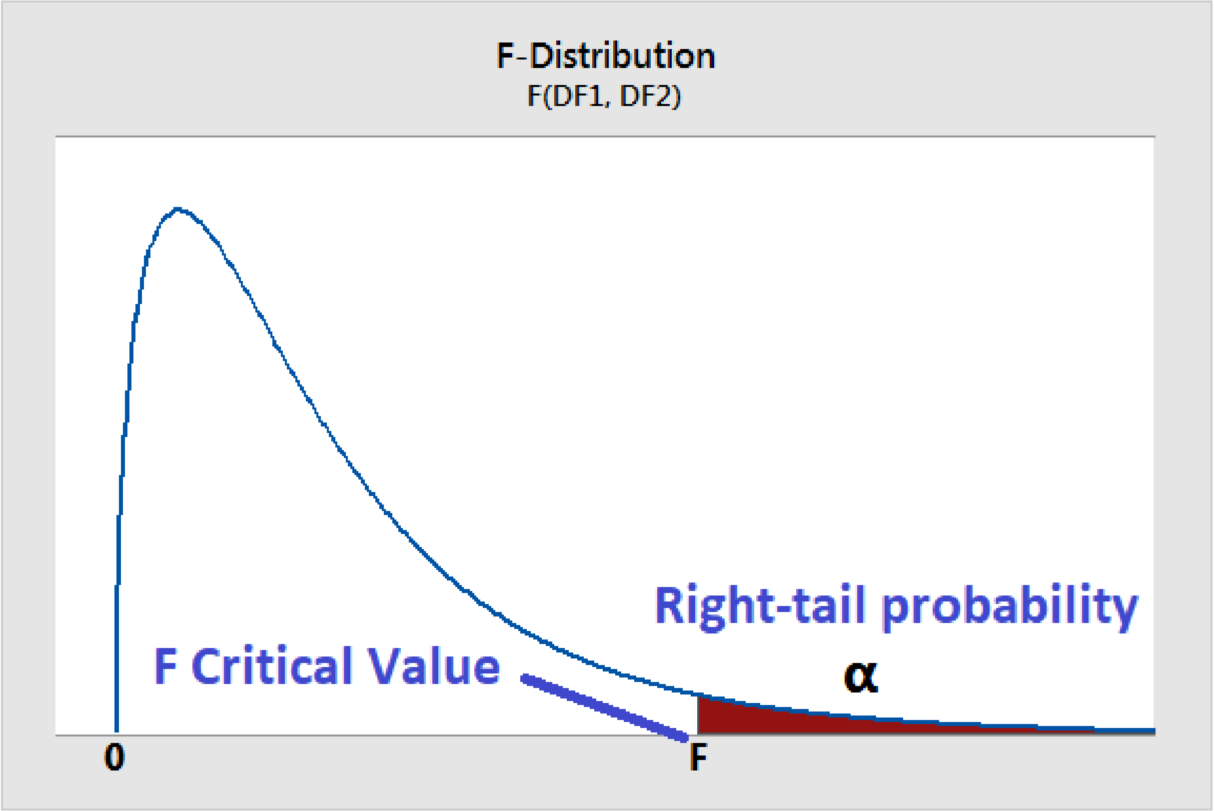
F distribution

- The distribution of f ratios

$$F \sim F(df_{between}, df_{within})$$

- F distribution is one tailed





Assumptions of one-way ANOVA

- **Independence:** Observations within and between groups are independent.
- **Normality:** Residual scores are normally distributed.
    - **Residuals** are the differences between each individual score and the group mean.
    - This assumption means that the distribution of the raw data within each group is normal, centered at that group's mean.
- **Homogeneity of variance:** Population variances are equal across groups.
- **Scale of measurement:** The dependent variable is continuous (i.e., interval or ratio scale). The independent variable is categorical.

ANOVA & Checking Normality of Residuals

Why check for normality?

- ANOVA assumes residuals (difference between observed values and group means) are normally distributional within each group.
- Ensures validity of F test
- ANOVA is robust to small violations, especially with:
    - Large samples
    - Equal group sizes (i.e., balanced designs)
- But severe non-normality can distort results.
- Test: histograms, Shapiro-wilk test on the residuals.

ANOVA & Homogeneity of Variance

Why check for equal variances?

- ANOVA assumes: All groups have equal variance (i.e., spread of scores is similar).
- This is called homogeneity of variance.
- Check: residual plots, Levene’s test