t-test Results Comparison

Context

For this task, we will use data from Schroeder and Epley (2015), which we have also used in the t-test data skills chapter. The aim of the study was to investigate whether delivering a short speech to a potential employer would be more effective at landing you a job than writing the speech down and the employer reading it themselves. Thirty-nine professional recruiters were randomly assigned to receive a job application speech as either a transcript for them to read, or an audio recording of the applicant reading the speech.

The recruiters then rated the applicants on perceived intellect, their impression of the applicant, and whether they would recommend hiring the candidate. All ratings were originally on a Likert scale ranging from 0 (low intellect, impression etc.) to 10 (high impression, recommendation etc.), with the final value representing the mean across several items.

This means we have a between-subjects design with one independent variable containing two levels: an audio group and a transcript group. It is an experiment as recruiters were randomly allocated to one group or another. There was one dependent variable for the mean intellect rating of a candidate interviewing for a position.

The research question was: Would there be a difference in the perceived intellect of candidates depending on if a recruiter reads the transcript of a job interview compared to if a recruiter listens to an audio recording of a job interview?

The hypothesis was: Recruiters who heard an audio recording of a job interview would rate a candidate as higher in perceived intellect compared to recruiters who read a transcript of a job interview.

**Version 1**

We hypothesised that the recruiters in the audio group would provide a higher candidate intellect rating than recruiters in the transcript group. The mean candidate intellect rating was 5.63 (*SD* = 1.91) in the audio group and the mean rating in the transcript group was 3.65 (*SD* = 1.61). Figure 1 provides a violin-boxplot showing the difference between the two groups.

Figure 1

*A violin-boxplot showing higher candidate intellect ratings in the audio group compared to the transcript group.*



When checking outliers, there were four participants in the audio group with ratings outside the boxplot limits, meaning they were more than ± 1.5 times the interquartile range. However, given the 0-10 bounded scale, we did not consider them as outliers and retained them in the analysis. Data met the assumption of normality, and we used Welch’s t-test which Delacre et al. (2017) recommend using as default as it does not assume homogeneity of variance.

A two-tailed Welch’s t-test showed recruiters who listened to an audio recording rated the candidate's intellect as significantly higher than recruiters who read a transcript, *t* (33.43) = 3.48, *p* = .001. Our results support our hypothesis that recruiters who heard an audio recording of a job interview would rate a candidate as higher in perceived intellect compared to recruiters who read a transcript of a job interview.

**Version 2**

We hypothesised that the recruiters in the audio group would provide a higher candidate intellect rating than recruiters in the transcript group. The mean candidate intellect rating was 5.63 (*SD* = 1.91) in the audio group and the mean rating in the transcript group was 3.65 (*SD* = 1.61). Figure 1 provides a bar plot showing the difference between the two groups.

Figure 1

*A bar plot showing higher candidate intellect ratings in the audio group compared to the transcript group.*



When checking outliers, there were four participants in the audio group with ratings outside the boxplot limits (see Appendix B1 for boxplots), meaning they were more than ± 1.5 times the interquartile range. However, given the 0-10 bounded scale, we did not consider them as outliers and retained them in the analysis. Data met the assumption of normality (see Appendix B2 for qq plots), and we used Welch’s t-test which Delacre et al. (2017) recommend using as default as it does not assume homogeneity of variance.

We found recruiters who listened to an audio recording rated the candidate's intellect as 1.99 units higher (95% CI = [0.83, 3.15]) than recruiters who read a transcript, where a two-tailed Welch's t-test was statistically significant, *t* (33.43) = 3.48, *p* = .001, Cohen's d = 1.12, 95% CI = [0.43, 1.81].

Our results support our hypothesis that recruiters who heard an audio recording of a job interview would rate a candidate as higher in perceived intellect compared to recruiters who read a transcript of a job interview.

**Version 3**

We hypothesised that the recruiters in the audio group would provide a higher candidate intellect rating than recruiters in the transcript group. The mean candidate intellect rating was 5.63 (*SD* = 1.91) in the audio group and the mean rating in the transcript group was 3.65 (*SD* = 1.61). Table 1 provides a table of descriptive statistics for the two groups.

Table 1

*Descriptive statistics for the mean intellect rating between recruiters in the audio and transcript conditions.*

|  |  |  |
| --- | --- | --- |
|  | Mean Intellect Rating | SD Intellect Rating |
| Transcript Group | 3.65 | 1.61 |
| Audio Group | 5.63 | 1.91 |

When checking outliers, there were four participants in the audio group with ratings outside the boxplot limits (see Appendix B1 for boxplots), meaning they were more than ± 1.5 times the interquartile range. However, given the 0-10 bounded scale, we did not consider them as outliers and retained them in the analysis. Data met the assumption of normality (see Appendix B2 for qq plots), and we used Welch’s t-test which Delacre et al. (2017) recommend using as default as it does not assume homogeneity of variance.

We found recruiters who listened to an audio recording rated the candidate's intellect as 1.99 units higher (95% CI = [0.83, 3.15]) than recruiters who read a transcript, where a two-tailed Welch's t-test was statistically significant, *t* (33.43) = 3.48, *p* = .001, Cohen's d = 1.12, 95% CI = [0.43, 1.81].

Our results support our hypothesis that recruiters who heard an audio recording of a job interview would rate a candidate as higher in perceived intellect compared to recruiters who read a transcript of a job interview.

**Version 4**

We hypothesised that the recruiters in the audio group would provide a higher candidate intellect rating than recruiters in the transcript group. The mean candidate intellect rating was 5.63 (*SD* = 1.91) in the audio group and the mean rating in the transcript group was 3.65 (*SD* = 1.61). Figure 1 provides a violin-boxplot showing the difference between the two groups.

Figure 1

*A violin-boxplot showing higher candidate intellect ratings in the audio group compared to the transcript group.*



When checking outliers, there were four participants in the audio group with ratings outside the boxplot limits, meaning they were more than ± 1.5 times the interquartile range. However, given the 0-10 bounded scale, we did not consider them as outliers and retained them in the analysis. Data met the assumption of normality (see Appendix B1 for qq plots), and we used Welch’s t-test which Delacre et al. (2017) recommend using as default as it does not assume homogeneity of variance.

We found recruiters who listened to an audio recording rated the candidate's intellect as 1.99 units higher (95% CI = [0.83, 3.15]) than recruiters who read a transcript, where a two-tailed Welch's t-test was statistically significant, *t* (33.43) = 3.48, *p* = .001, Cohen's d = 1.12, 95% CI = [0.43, 1.81].

Our results support our hypothesis that recruiters who heard an audio recording of a job interview would rate a candidate as higher in perceived intellect compared to recruiters who read a transcript of a job interview.

**Version 5**

The mean candidate intellect rating was 5.634921 (*SD* = 1.911343) in the audio group and the mean rating in the transcript group was 3.648148 (*SD* = 1.608674). Figure 1 provides a violin-boxplot showing the difference between the two groups.

Figure 1



The data met all parametric assumptions. We found recruiters who listened to an audio recording rated the candidate's intellect as significantly higher than recruiters who read a transcript, *t* (33.434) = 3.4787, *p* = .001421, Cohen's d = 1.124693, 95% CI = [0.4293182, 1.805776].