A person on ResearchGate asked the following question:

How can I correlate ordinal variables (attitude Likert scale) with continuous ratio data (years of experience)?

Currently, I am working on my dissertation which explores learning organisation characteristics at HEIs. One of the predictor demographic variables is the indication of the years of experience. Respondents were asked to fill in the gap the number of years. Should I categorise the responses instead? as for example:

1. from 1 to 4 years
2. from 4 to 10
3. and so on?

or is there a better choice/analysis I could apply?

My answer may also be of interest to others, so I post it here as well.

Normal practice is to treat likert scales as continuous variable even though they are not. As long as there are >=5 options, the bias from discreteness is not large.

I simulated the situation for you. I generated two variables with continuous random data from two normal distributions with a correlation of .50, N=1000. Then I created likert scales of varying levels from the second variable. Then I correlated all these variables with each other.

Correlations of continuous variable 1 with:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuous2</td>
<td>0.5</td>
</tr>
<tr>
<td>likert10</td>
<td>0.482</td>
</tr>
<tr>
<td>likert7</td>
<td>0.472</td>
</tr>
<tr>
<td>likert5</td>
<td>0.469</td>
</tr>
<tr>
<td>likert4</td>
<td>0.432</td>
</tr>
<tr>
<td>likert3</td>
<td>0.442</td>
</tr>
<tr>
<td>likert2</td>
<td>0.395</td>
</tr>
</tbody>
</table>

So you see, introducing discreteness biases correlations towards zero, but not by much as long as likert is >=5 level. You can correct for the bias by multiplying by the correction factor if desired:

Correction factor:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correction Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuous2</td>
<td>1</td>
</tr>
<tr>
<td>likert10</td>
<td>1.037</td>
</tr>
</tbody>
</table>
Psychologically, if your data does not make sense as an interval scale, i.e. if the difference between options 1-2 is not the same as between options 3-4, then you should use Spearman's correlation instead of Pearson's. However, it will rarely make much of a difference.

Here's the R code.

```r
#load library
library(MASS)

#simulate dataset of 2 variables with correlation of .50, N=1000
simul.data = mvrnorm(1000, mu = c(0,0), Sigma = matrix(c(1,0.50,0.50,1), ncol = 2), empirical = TRUE)
simul.data = as.data.frame(simul.data);colnames(simul.data) = c("continuous1","continuous2")

#divide into bins of equal length
simul.data["likert10"] = as.numeric(cut(unlist(simul.data[2]),breaks=10))
simul.data["likert7"] = as.numeric(cut(unlist(simul.data[2]),breaks=7))
simul.data["likert5"] = as.numeric(cut(unlist(simul.data[2]),breaks=5))
simul.data["likert4"] = as.numeric(cut(unlist(simul.data[2]),breaks=4))
simul.data["likert3"] = as.numeric(cut(unlist(simul.data[2]),breaks=3))
simul.data["likert2"] = as.numeric(cut(unlist(simul.data[2]),breaks=2))

#correlations
round(cor(simul.data),3)
```