

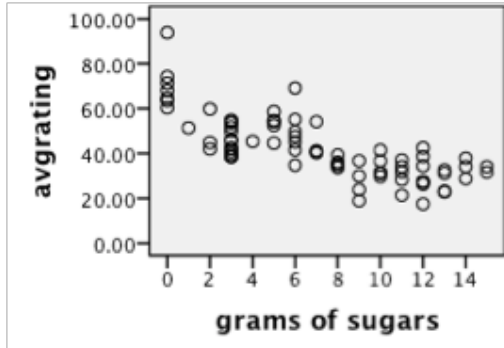
Correlation & Prediction Worksheet

Part 1: Answers - Correlation

Use the Cereals File

What's the strength of the relationship between sugar content and rating?

Checked linearity - OK.



Obtained PPMC: $-.76$

| Correlations | | | |
|-----------------|---------------------|-----------------|--------------|
| | | grams of sugars | avgrating |
| grams of sugars | Pearson Correlation | 1 | $-.764^{**}$ |
| | Sig. (2-tailed) | | 0.000 |
| | N | 76 | 76 |
| avgrating | Pearson Correlation | $-.764^{**}$ | 1 |
| | Sig. (2-tailed) | 0.000 | |
| | N | 76 | 77 |

Interpretation: strong negative correlation between $(-.76)$ grams of sugar and average cereal rating. As sugar content goes up, average rating goes down.

Convert average rating to ordinal data with 3 categories of similar size.

Used recode process with break points of:

- Low: up to 34.76
- Mid: 34.77 - 45.83
- High: > 45.83

That gave me 3 groups of similar size:

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---------|-----------------|-----------|---------|---------------|--------------------|
| Valid | low rating | 25 | 32.5 | 32.9 | 32.9 |
| | moderate rating | 26 | 33.8 | 34.2 | 67.1 |
| | high rating | 25 | 32.5 | 32.9 | 100.0 |
| | Total | 76 | 98.7 | 100.0 | |
| Missing | System | 1 | 1.3 | | |
| Total | | 77 | 100.0 | | |

What's the relationship between shelf and the ordinal version of rating?

Used Kendall's Tau - two ordinal variables.

Interpretation: No relationship between location on shelf and rating as an ordinal variable. Tau near zero.

| | | | display shelf - counting from floor | rating group |
|-----------------|-------------------------------------|-------------------------|-------------------------------------|--------------|
| Kendall's tau_b | display shelf - counting from floor | Correlation Coefficient | 1.000 | 0.066 |
| | | Sig. (2-tailed) | . | 0.524 |
| | | N | 76 | 75 |
| | rating group | Correlation Coefficient | 0.066 | 1.000 |
| | | Sig. (2-tailed) | 0.524 | . |
| | | N | 75 | 76 |

Note: could have used Cramer's V since both can be considered categorical. Correlation is then .33 which is still a low correlation but certainly greater than Tau=.07. Decision of which to report should be made on which level of measurement the data best fits. In this case, categorical.

| | | rating group | | | |
|-------------------------------------|--------------|--------------|-----------------|-------------|-------|
| | | low rating | moderate rating | high rating | Total |
| display shelf - counting from floor | bottom shelf | 5 | 6 | 9 | 20 |
| | middle shelf | 14 | 3 | 4 | 21 |
| | top shelf | 6 | 16 | 12 | 34 |
| | Total | 25 | 25 | 25 | 75 |

| | | Value | Approx. Sig. |
|--------------------|------------------|-------|--------------|
| Nominal by Nominal | Phi | 0.467 | 0.003 |
| | Cramer's V | 0.330 | 0.003 |
| | N of Valid Cases | 75 | |

Part 2: Answers - Correlation

Use the hoola data file

Is there a relationship between gender and transfer status?

Obtained Phi

| | | Transfer Student? | | Total |
|--------|--------|-------------------|-------|--------|
| | | yes | no | |
| gender | male | 51.2% | 48.8% | 100.0% |
| | female | 50.8% | 49.2% | 100.0% |
| | Total | 51.1% | 48.9% | 100.0% |

| Symmetric Measures | | | |
|--------------------|------------|-------|--------------|
| | | Value | Approx. Sig. |
| Nominal by Nominal | Phi | 0.004 | 0.963 |
| | Cramer's V | 0.004 | 0.963 |
| N of Valid Cases | 139 | | |

Interpretation: no relationship since $\phi = .004$. This means there are similar proportions of men and women who are transfer students.

Part 3: Answers - Correlation

Use the Calcium file

Note: value labels were given to the wrong variable. So, data file needed correction - in variable view, gave value labels to the treatment variable.

What's the relationship between treatment group and systolic blood pressure after the 12 week treatment? Point biserial correlation (-.21) is low.

Did Point Biserial by hand after getting info needed from PASW. Will need from PASW: Standard deviations, proportions, Means

| systolic blood pressure after treatment | | | | |
|---|--------|----|----------------|--------------|
| Treatment | Mean | N | Std. Deviation | % of Total N |
| placebo | 113.91 | 11 | 11.327 | 52.4% |
| calcium | 109.90 | 10 | 7.795 | 47.6% |
| Total | 112.00 | 21 | 9.783 | 100.0% |

$$r_{pb} = \frac{109.90 - 113.91}{9.561} \sqrt{.476(.524)} = -.21$$

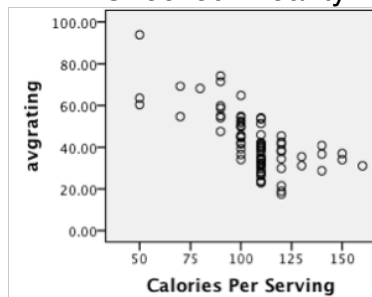
Interpretation based on comparison of means/medians - those in the calcium group had lower systolic blood pressure (SBP) than the placebo group after the calcium treatment. This group started out, before treatment, with a slightly higher SBP before the treatment than the placebo group.

Part 4: Answers - Prediction

Use the Cereals file

Obtain the prediction equation you would use to predict average rating from calories per serving.

Checked linearity - fine



Obtain PPMC, means, and standard deviations

| | | Calories Per Serving | avgrating |
|----------------------|---------------------|----------------------|-----------|
| Calories Per Serving | Pearson Correlation | 1 | -.682** |
| | Sig. (2-tailed) | | 0.000 |

| | | | |
|-----------|---------------------|---------|----|
| | N | 77 | 77 |
| avgrating | Pearson Correlation | -.682** | 1 |
| | Sig. (2-tailed) | 0.000 | |
| | N | 77 | 77 |

| | | Calories Per Serving | avgrating |
|----------------|---------|----------------------|-----------|
| N | Valid | 77 | 77 |
| | Missing | 0 | 0 |
| Mean | 106.88 | 42.5342 | |
| Std. Deviation | 19.484 | 13.95396 | |

Write out formula

$$\hat{y} = -.682\left(\frac{13.95396}{19.484}\right)(x - 106.88) + 42.5342$$

What would the SEE be for predicting rating from calorie content?

$$SEE = 13.95396\sqrt{1 - (-.682^2)} = 10.21$$

Interpretation: fairly large amount of error in predicting average rating from calories per serving.

Predict what the rating would be for a new cereal with 95 calories per serving

$$\hat{y} = -.682\left(\frac{13.95396}{19.484}\right)(95 - 106.88) + 42.5342 = 48.34$$