Math 1700: Elementary Statistics

2^{nd} Week Summary (09/11/23)

• Measures of Position

Quartiles: $Q_1, Q_2(\tilde{x}), Q_3$

Percentile: P_k

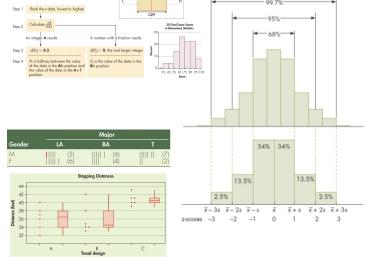
Five number summary: (L, Q_1, Q_2, Q_3, H)

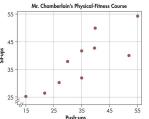
Interquartile range: $IQR = Q_3 - Q_1$

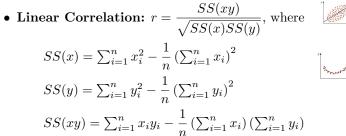
- Box-and-whiskers display
- Standard score, or z-score

$$z_i = \frac{x_i - \bar{x}}{s}$$

- Empirical Rule (68 95 99.7 Rule)
- Comparing the measures of center and spread
- Bivariate Data:
- Qualitative vs Qualitative Contingency table
- Qualitative vs Quantitative Side-by-side Box Plot
- Quantitative vs Quantitative Scatter diagram







- Properties of the Correlation r:

Takes values between -1 and 1

r=1 or r=-1 implies that the points lie on a straight line

r=0 implies that there is no **linear** association

r < 0 implies that there is a negative **linear** association & r > 0 implies positive **linear** association r is strongly affected by a few outliers

• Linear Regression: $\hat{y} = b_0 + b_1 x$

where, \hat{y} represents the predicted value of y that corresponds to a particular value of x

• The Least Squares Criterion: Finding b_0 and b_1 such that $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ is as small as possible

The slope: $b_1 = \frac{SS(xy)}{SS(x)}$, represents the predicted change in y per unit increase in x

The y-intercept: $b_0 = \bar{y} - b_1 \bar{x}$, is the value of y where the line of best fit intersects the y-axis

• Causation and Lurking Variable:

Simpson's Paradox