

Data characteristics.

Summary of what we have covered in class and what you will need to know for the test.

For supplementary reading look up the terms in the index of the book and go to the pages indicated.

Range, quantiles or percentiles.

Median; interquartile range; box-and whiskers plot; outliers; relative frequencies; cumulative frequency; histogram, ogive.

Moments.

Mean

$$\bar{X} = \frac{\sum X_i}{n} = \frac{\sum X_i f_i}{\sum f_i} = \sum X_i r_i,$$

(need to know).

Theorem. Average deviation = 0 (should be able to prove).

Mean squared deviation, MS (mean of squares).

Theorem. MS_a is minimized at the mean: if $a \neq \bar{Y}$, $MS_a > MS$. (know the result, understand the proof).

Standard deviation can be measured by $\sqrt{m_2}$.

Moments for data. average deviation: $m_1 = \frac{\sum(X_i - \bar{X})}{n}$; then $m_2 = \frac{\sum(X_i - \bar{X})^2}{n}$; $m_3 = \frac{\sum(X_i - \bar{X})^3}{n}$; $m_4 = \frac{\sum(X_i - \bar{X})^4}{n}$
need to know.

Standardized moments.

Properties. If $Y_i = bX_i$; then $\bar{Y} = \bar{X} \cdot b$; any moment

$$m_k(Y) = b^k m_k(X).$$

(need to know). Standardized moment

$$\alpha_k(X) = \frac{m_k(X)}{(m_2(X))^{\frac{k}{2}}}.$$

is **invariant to shifts and changes in scale** (need to know).

Standardized variables; coefficients of skewness and kurtosis.

Bivariate data.

Covariance:

$\text{Cov}(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n}$ (need to know).

Correlation coefficient.

$$r = \text{corr}(X, Y) = \frac{\sum_{i=1}^n \frac{(X_i - \bar{X})}{\sqrt{m_2(X)}} \frac{(Y_i - \bar{Y})}{\sqrt{m_2(Y)}}}{n} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n\sqrt{m_2(X)m_2(Y)}}$$

(need to know).

Theorem. The correlation coefficient takes values between -1 and 1 . (know result, understand the proof at least for $n=3$)

Relation between correlation coefficient and least squares regression (OLS). If $a + bX$ is the OLS line that minimizes the sum of squared errors:

$$\Sigma (Y_i - a - bX_i)^2 \rightarrow \min .$$

then

$$r(X, Y) = b \frac{\sqrt{m_2(X)}}{\sqrt{m_2(Y)}}.$$

(formula will be given on the test); you need to know the meaning of it all. Properties of OLS: $\Sigma e_i = 0$; the OLS line passes through the point of means: $(\bar{X}, \bar{Y})$ - need to know.

You should be able to compute and interpret all these characteristics of the data.