Practical Applications of Statistical Analysis

M. Mokhtari, Ph.D.

Review
Probability Assumptions

- In order to do inference or make *generalizations to the entire population from sample* findings we use statistics (sample mean and variance) as an approximation to the population parameters (population mean and variance)
- Therefore, we have to account for sampling error (probability of being incorrect)
- Distribution of probability of each outcome is the probability distribution

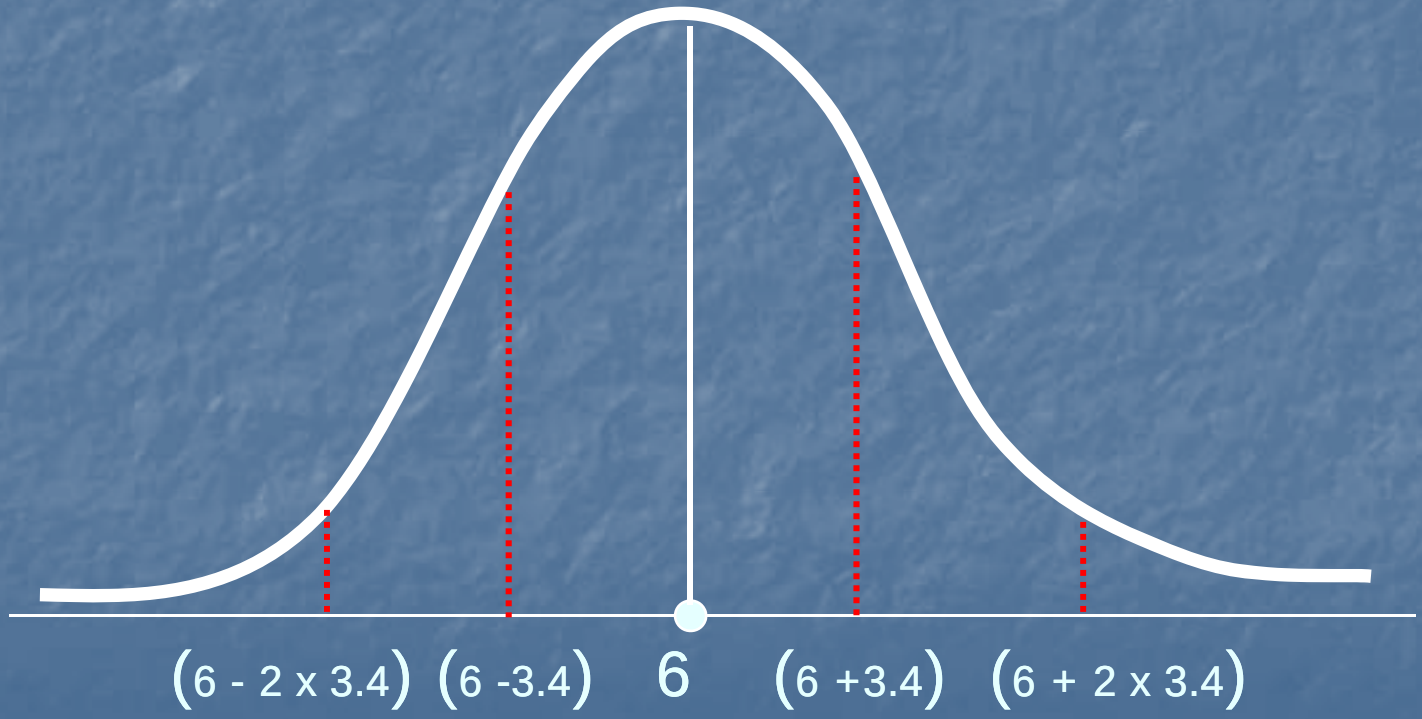
How to Deal With Uncertainty? Probability

- A probability is a number indicating the chance that a specific event will happen. The highest probability **1**, means that the even will certainly happen. The lowest probability is **0**, means the event will not occur.
- Event: a possible outcome of an experiment

Probability Density (Normal) Curve for a variable (for example, Total Revenue)

Mean = $\mu \approx \bar{x} = 6$ and Variance = $\sigma^2 \approx S^2 = 12$

$\sigma \approx$ Standard Deviation = S = $\sqrt{12} = 3.4$



Areas Under Normal Curve

If population Mean is =μ and population Standard Deviation is =σ, then:

- About **68%** of the area under the normal curve is within plus one and minus 1 standard deviation from the mean, **$\mu \pm 1\sigma$**. For previous curve: (6 ± 3.4)
- About **95%** of the area under the normal curve is within plus and minus 2 the mean, **$\mu \pm 2\sigma$**. For previous curve: $(6 \pm 2 \times 3.4)$
- Practically all **99%** of the area under the normal curve is within 3 standard deviation from the mean, **$\mu \pm 3\sigma$**

Review Sampling Distribution

- The sampling distribution is a distribution of sample means by their frequency of occurrence. If we were to make lots and lots of samples, each with (for example) 40 people randomly selected, and computed for each sample a mean value, we would get set of lots of means. If we were to order means by their frequency of occurrence, we would get a distribution. This distribution is known as sampling distribution.

Review Test Statistic

- The sampling distribution is described using test statistics such **z** and **t**, and also **F**-coefficient. We say that sample mean approximately has **z**-distribution or **t** -distribution
- Recall: when do we use **t** statistic?

$$t = \frac{\bar{Y} - \mu}{s / \sqrt{n}}$$

Confidence Interval Estimation

- Let's say we want to make a statement about the next year's Exports. What do we need to do to estimate next year's Exports with the **95** percent confidence?
- Let's say that Exports is normally distributed variable (that is, its distribution is bell-shape-and-symmetrical)
- Then we just need to use sample mean and sample standard deviation along with sample size and t value to get the 95 percent Confidence Interval.

Confidence Interval: Example

- Let's say the estimate of Exports is **73** Billion drams. The standard deviation is **7.4**. Assuming a **t** statistic equal to **2** Then the **95 %** confidence interval for the next year's Exports is:

$$73 \pm (2 \times 7.4) \frac{1}{\sqrt{4}}$$

- If we were to repeat the process with different samples many times, the Means that we get from each sample predictions will be captured in this interval in **95 %** of the time.

Modeling Sales: [Ordinary Least Squares (OLS)]

- Estimate of real relationship can be expressed in the following equation :

$$Sales = \alpha + \hat{\beta} \times Exports + \varepsilon$$

- Where, ε is what we cannot explain (=error term).

Parameter Estimation

$$\hat{\beta}_1 = \frac{\text{Covariance(sales, exports)}}{\text{Variation coefficient (exports)}}$$

$$\hat{\beta}_1 = \frac{5{,}641{,}795{,}918{,}367{,}350}{58{,}835{,}121{,}107{,}266{,}400} = 0.07$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_0 = 98{,}237{,}481 - 0.07 \times 127{,}235{,}294 = 89{,}331{,}010$$

Empirical Equation for Quarterly Data

Sales= 89,331,010+0.09*Exports

- The slope of this equation tells us the following: **For 100 million drams increase in Exports we should expect Sales to go up by 9 million drams!**

Summary and Conclusion

- In this session we reviewed major concepts of statistics
- We applied our knowledge to build a simple model of sales and exports using only 4 things:
- Mean
- Variance
- Covariance
- Distributional Assumptions