Chapter Five Lecture Notes

Spring 2023

Distribution of Sample Mean:

- Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size *n* from *Y* which has the distribution $N(\mu, \sigma^2)$.

  THEN the distribution of $\bar{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$, the sample mean, is $N(\mu, \sigma^2 / n)$.
- Central Limit Theorem (CLT): When $Y_1, Y_2, \ldots, Y_n$ is a random sample of size *n* from *Y* which has expected value μ and variance $\sigma^2 < \infty$, then the distribution of $\bar{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$ is asymptotically $N(\mu, \sigma^2 / n)$. For a random sample of size *n*, it is always true that $E(\bar{Y}_n) = \mu$ and $\operatorname{var}(\bar{Y}_n) = \frac{\sigma^2}{n}$. The CLT allows probability calculations that increase in accuracy as the sample size increases.

Testing a Statistical Hypothesis (Variance Known)

- Null hypothesis: $H_0 : E(Y) = \mu_0$
- Alternative Hypothesis: $H_1 : E(Y) \neq \mu_0$
- Level of significance α
- Type I error: reject a null hypothesis that is true.
- The probability of a Type I error is α. Formally, $\Pr_0\{\text{Reject } H_0\} = \alpha$.
- Test statistic. Null distribution is the distribution of the test statistic under the null hypothesis.
- Type II error: accept a null hypothesis that is false.
- Typically, α is set to a small number (0.05 or 0.01), and *n* is chosen so that $\beta = \Pr_1\{\text{Accept } H_0\}$, where β is dependent on a setting of the alternative hypothesis, is small. Alternative distribution: distribution of the test statistic under the setting of the alternative hypothesis.

Example of Statistical Hypothesis (Variance Known)

- A research team took a sample of 8 observations from the random variable *Y*, which had a normal distribution $N(\mu, \sigma^2 = 625)$. They observed $\bar{y}_8 = 43.2$, where $\bar{y}_8$ is the average of the eight sampled observations. Test the null hypothesis that $H_0: E(Y) = 50$ against the alternative $H_1: E(Y) \neq 50$ at the 0.10, 0.05, and 0.01 levels of significance.

- The test statistic is $\bar{Y}_8$.
- The null distribution is $N(50, 625/8 = 78.125 = 8.84^2)$
- Put $\bar{Y}_8$ in standard score form: $Z = \dfrac{\bar{Y}_8 - \mu_0}{\sigma/\sqrt{8}}$.
- If $|Z| \geq 1.645$, reject $H_0: E(Y) = \mu_0$ at $\alpha = 0.10$. If $|Z| \geq 1.960$, reject $H_0: E(Y) = \mu_0$ at $\alpha = 0.05$. If $|Z| \geq 2.576$, reject $H_0: E(Y) = \mu_0$ at $\alpha = 0.01$.
- Calculate the standard score form of Z for the data given in the problem: $z = \dfrac{43.2 - 50}{8.84} = -0.769$
- Since $|z| < 1.645$, accept $H_0: E(Y) = 50$ at $\alpha = 0.10$. Of course, one should accept $H_0: E(Y) = 50$ at $\alpha = 0.05$ and $\alpha = 0.01$ as well.

Testing a Statistical Hypothesis (Variance Unknown).

- We cannot put $\bar{Y}_n$ in standard score form: $Z = \dfrac{\bar{Y}_n - \mu_0}{\sigma/\sqrt{n}}$ because we do not know σ.
- Instead, we use an estimate of σ^2, $\hat{\sigma}^2 = S^2 = \dfrac{\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2}{n-1}$, which has $n-1$ degrees of freedom.
- We put $\bar{Y}_n$ in studentized standard score form: $T_{n-1} = \dfrac{\bar{Y}_n - \mu_0}{\hat{\sigma}/\sqrt{n}}$.
- Student showed that the percentiles from the standard normal (here 1.645, 1.960, and 2.576) had to be stretched. The amount of stretching is determined by the number of degrees of freedom in $\hat{\sigma}^2 = S^2 = \dfrac{\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2}{n-1}$; here, $n-1$ degrees of freedom. These values are tabulated and will be given to you in your examinations.
- Example problem: Chapter 5 Study Guide, Problem 4:

  A research team took a sample of 8 observations from the random variable Y, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_8 = 43.2$, where $\bar{y}_8$ is the average of the eight sampled observations and $s^2 = 517.5$ is the observed value of the unbiased estimate of σ^2, based on the sample values. Test the null hypothesis that $H_0: E(Y) = 50$ against the alternative $H_1: E(Y) \neq 50$ at the 0.10, 0.05, and 0.01 levels of significance.

  - The degrees of freedom is $n - 1 = 8 - 1 = 7$.

- The studentized test statistic is $t_7 = \frac{43.2 - 50}{\sqrt{(517.5/8}} = \frac{-6.8}{\sqrt{64.7}} = \frac{-6.8}{8.04} = -0.845$
- Find the student t stretches for 1.645, 1.960, 2.576. They are 1.895, 2.365, 3.499.
- Make your decision. Here, it is to accept at the 0.10 level of significance (and of course at 0.05 and 0.01 as well) since $|-0.845| < 1.895$.

Confidence Interval, Variance Known

- The formal test of a null hypothesis addresses whether a single value (here a value for the expected value of the sampled random variable) is consistent with the data.
- Most researchers prefer a statement of what the data does show.
- The confidence interval is such a statement.
- The 99% confidence interval for *E(Y)* is $\bar{y}_n \pm 2.576 \frac{\sigma}{\sqrt{n}}$.
- In our first example is: $43.2 \pm 2.576(8.84) = 43.2 \pm 22.8$

Confidence Interval, Variance Unknown

- As always, we use the data S^2 to estimate σ^2 and stretch our normal percentile (here 2.576) using the degrees of freedom of S^2. The stretch of 2.576 is 3.499.
- The 99% confidence interval for *E(Y)* is $\bar{y}_n \pm t_{n-1,2.576} \frac{\hat{\sigma}}{\sqrt{n}}$.
- In the example problem, the 99% confidence interval for *E(Y)* is $\bar{y}_8 \pm t_{7,2.576} \frac{\hat{\sigma}}{\sqrt{n}} = 43.2 \pm 3.499 \sqrt{\frac{517.5}{8}} = 43.2 \pm 3.499(8.04) = 43.2 \pm 28.1$.