Statistics Rough Notes

Summary

- The following table summarises common critical values.

Level of Significance α	Critical Value c_α for Two-Sided Test	Critical Value c_α for One-Sided Test
1%	$2\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 2.576$	$\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 2.326$
5%	$2\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.960$	$\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.645$
10%	$2\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.645$	$\phi(-c_\alpha) = \alpha \Rightarrow c_\alpha = 1.282$

- The following table summarises standard error and test statistic computations for common hypothesis tests.

Parameter of Interest	Estimator	Standard Error Computation	Test Statistic Computation
Mean	$\overline{X}_n$	$se(\overline{X}_n) = \sqrt{\frac{s_{n,X}^2}{n}}$ (The intuition for this is that the sample variance is a consistent estimate of the population variance, the variance of the sample mean is the population variance divided by n.)	$t_n(\mu_0) = \frac{\overline{X}_n - \mu_0}{se(\overline{X}_n)}$
Difference of Means in One Sample	$\overline{D}_n := \hat{\mathbb{E}}(X - Y)$ $= \overline{X}_n - \overline{Y}_n$	$se(\overline{D}_n) = \sqrt{\frac{s_{n,D}^2}{n}}$ (This is exactly analogous to the above. The difference is treated as a variable.)	$t_n(\delta_0) = \frac{\overline{D}_n - \delta_0}{se(\overline{D}_n)}$ (This is again exactly analogous to the above)
Difference of Means in Two Independent Samples	$\hat{\delta} = \overline{X}_{1,n_1} - \overline{X}_{2,n_2}$	$se(\hat{\delta}) = \sqrt{\frac{s_{1,n_1}^2}{n_1} + \frac{s_{2,n_2}^2}{n_2}}$ (This is an estimator of $sd(\hat{\delta}) = \sqrt{var(\hat{\delta})}$ $= \sqrt{var(\overline{X}_{1,n_1}) + var(\overline{X}_{2,n_2})}$, where the second equality follows by the assumption of independent samples.)	$t_n(\delta_0) = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})}$
Proportion	$\hat{p}_n$	$se(\hat{p}_n) = \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$ (The numerator within the square root is a consistent estimator of the population variance, then the fraction within the square root is a consistent estimator of the variance of the sample mean.)	$t_n(p_0) = \frac{\hat{p}_n - p_0}{se(\hat{p}_n)}$
Difference of Proportions in One Sample	$\hat{d}_n := \hat{p}_n - \hat{q}_n$ (This is exactly analogous to the case of difference of means in one sample.)	$se(\hat{d}_n) = \sqrt{\frac{s_{n,d}^2}{n}}$	$t_n(\delta_0) = \frac{\hat{d}_n - \delta_0}{se(\hat{d}_n)}$
Difference of Proportions in Two Independent Samples	$\hat{\delta} := \hat{p}_{1,n_1} - \hat{p}_{2,n_2}$	$se(\hat{\delta}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}}$ where $\hat{\pi} = \frac{n_1}{n_1+n_2}\hat{p}_{1,n_1} + \frac{n_2}{n_1+n_2}\hat{p}_{2,n_2}$ (Under the null, the two proportions are equal, and the weighted average of the two proportions is a consistent estimate of the common proportion.)	$t_n(\delta_0) = \frac{\hat{\delta} - \delta_0}{se(\hat{\delta})}$

Estimation

Estimators

- Definition (Sample Mean)
  - The sample mean of the random variable X in a given random sample $\{(X_i, Y_i)\}_{i=1}^n$ is $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$, alternatively denoted as $\hat{\mu}_{n,X}$.

- The sample mean is an estimator of the population mean $\mathbb{E}X$, alternatively denoted as μ_X, where X is some arbitrary X_i.
- Definition (Sample Variance)
  - The sample variance of the random variable X (an arbitrary X_i) in a given sample $\{(X_i, Y_i)\}_{i=1}^n$ is $v\hat{a}r(X) := \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. This is alternatively denoted as $s_{n,X}^2$.
  - The sample variance is computed by applying Bessel's correction (multiplication by $\frac{n}{n-1}$) to the naive analogue estimator of population variance $\hat{\sigma}_{n,X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ (which is the sample analogue of $var(X) = \mathbb{E}(X - \mathbb{E}X)^2$)
- Definition (Sample Covariance)
  - The sample covariance of the random variables X, Y in a given sample $\{(X_i, Y_i)\}_{i=1}^n$ is $c\hat{o}v(X, Y) := \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$. This is alternatively denoted as $s_{n,XY}$.
  - The sample covariance is computed by applying Bessel's correction (multiplication by $\frac{n}{n-1}$) to the naive estimator of population covariance $\hat{\sigma}_{n,XY} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$ (which is the sample analogue of $cov(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$).
- Common Results
  - The sample mean, variance, and covariance share the properties of the population counterparts.
- Definition (Sample Standard Deviation)
  - The sample standard deviation of the random variable X in a given sample $\{(X_i, Y_i)\}_{i=1}^n$ is $\hat{sd}(X) = \sqrt{v\hat{a}r(X)}$.
  - The sample standard deviation relates to the sample variance in the way the population standard deviation relates to the population variance.
- Definition (Sample Correlation)
  - The sample correlation of the random variables X, Y in a given sample $\{(X_i, Y_i)\}_{i=1}^n$ is $c\hat{o}rr(X, Y) = \frac{c\hat{o}v(X,Y)}{\sqrt{v\hat{a}r(X)v\hat{a}r(Y)}}$.
  - The sample correlation relates to the sample covariance and sample variance in the way the population standard deviation relates to the population covariance and population variance.

Sampling Distribution of Estimators

- Definition (Mean Squared Error)
  - The mean squared error of an estimator $\hat{\theta}_n$ of population parameter θ is $mse(\hat{\theta}_n) := \mathbb{E}(\hat{\theta}_n - \theta)^2$. Note that $\hat{\theta}_n$ is itself a random variable because it is a function of random variables (that constitute the sample).
  - This is a measure of the precision of the estimator.
- Definition (Bias)
  - The bias of an estimator $\hat{\theta}_n$ of population parameter θ is $bias(\hat{\theta}_n) := \mathbb{E}\hat{\theta}_n - \theta$.
- Result (Decomposition of Mean Squared Error)
  - The mean squared error of an estimator $\hat{\theta}_n$ of population parameter θ can be decomposed into a function of its bias and its variance according to $mse(\hat{\theta}_n) = [bias(\hat{\theta}_n)]^2 + var(\hat{\theta}_n)$.
  - Then, the precision, as measured by mean squared error, of an unbiased estimator is entirely a function of the variance of that estimator. The efficient unbiased estimator is the unbiased estimator with the smallest variance.
- Definition (Standard Error)
  - The standard deviation of an estimator $\hat{\theta}_n$ of population parameter θ is $sd(\hat{\theta}_n) = \sqrt{var(\hat{\theta}_n)}$.
  - The standard error of an estimator $\hat{\theta}_n$ of population parameter θ is $se(\hat{\theta}_n)$, and is obtained by substituting each occurrence of a population parameter in the standard deviation with its estimate. The standard error of an estimator is an estimate of the standard deviation of that estimator.

Summary

- The sample mean of random variable X (an arbitrary X_i) in a given sample $\{X_i\}_{i=1}^n$ is $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. This is an estimator of the population mean $\mathbb{E}X$, and it can be verified that this is an unbiased estimator (by proving that its expectation is equal to $\mathbb{E}X$).
- The sample variance of X is $v\hat{a}r(X) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, this is alternatively denoted as $s_{n,X}^2$. This is an estimator of the population variance $var(X)$, which is alternatively denoted as σ_X^2. It can be verified that the sample variance is an unbiased estimator of the population variance.
  - In general, s denotes an estimator of σ that includes some efficiency-improving correction, such as Bessel's correction or a degrees of freedom correction (in linear regression).
- The analogue estimator of the population variance is $\hat{\sigma}_{n,X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)$, it can be verified that this is also an unbiased estimator.

- Note that ˆis used to distinguish estimators from population parameters (estimands), although in the case of sample variance s, the ˆis dropped because there is no population parameter denoted by s, hence no ambiguity.
- The (population) standard deviation of X is $sd(X) = \sqrt{var(X)}$, alternatively denoted σ_X, this is a measure of the dispersion of X.
- The sample standard deviation of X is $\hat{sd}(X) = \sqrt{\hat{var}(X)}$, alternatively denoted $s_{n,X}$, this is a an estimator of the population standard deviation.
- The variance of the sample mean is $var(\overline{X}_n) = \frac{\sigma_X^2}{n}$. The standard deviation of the sample mean $\overline{X}_n$ is $sd(\overline{X}_n)$, this is a measure of the dispersion of the sample mean $\overline{X}_n$ due to sampling variation. It can be verified that $sd(\overline{X}_n) = \frac{\sigma_X}{\sqrt{n}}$.
  - The variance of the sample mean converges to zero as n becomes large (and the expectation of the sample mean is the population mean) so the sample mean is a consistent estimator of the population mean.
- The standard error of the sample mean $\overline{X}_n$ is $se(\overline{X}_n)$, this is an estimate of $sd(\overline{X}_n)$ obtained by substituting each occurrence of a population parameter in $sd(\overline{X}_n)$ with its estimate. Then, $se(\overline{X}_n) = \frac{s_{n,X}}{\sqrt{n}}$.

Law of Large Numbers

- Definition (Convergence in Probability)
  - A sequence of random variables $\{\xi_i\}_{i=1}^n$ converges in probability to a constant c iff for all small $\epsilon > 0$, as $n \to \infty$, $\mathbb{P}\{|\xi_n - c| \le \epsilon\} \to 1$, (where ξ_n denotes the nth random variable in the given sequence), denoted as $\xi_n \to^p c$. In other words, as n tends to ∞, the probability that ξ_n lies within the range $[c - \epsilon, c + \epsilon]$ tends to 1, for all $\epsilon > 0$, however small.
- Definition (Consistency)
  - A sequence of estimators $\{\hat{\theta}_i\}_{i=1}^n = \{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n\}$ is consistent for the population parameter (estimand) θ iff $\hat{\theta}_n \to^p \theta$.
  - It can be verified that sample mean and sample variance are consistent estimators of the population counterparts.
- Theorem (Law of Large Numbers)
  - If the sequence of random variables $\{X_i\}_{i=1}^n$ is independently and identically distributed, and (for all i) $\mathbb{E}X_i = \mu$, where μ is some constant, then the sample mean $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to μ (i.e. $\overline{X}_n \to^p \mu$).
  - This follows from $\overline{X}_n$'s being an unbiased estimator of μ and that $var(\overline{X}_n)$ tends to 0 as n tends to ∞.

Central Limit Theorem

- Definition (Convergence in Distribution)
  - A sequence of random variables $\{\xi_i\}_{i=1}^n$ converges in distribution to some continuous distribution with cumulative distribution function F iff for all $x \in \mathbb{R}$, as $n \to \infty$, $F_{\xi_n}(x) := P\{\xi_n \le x\} \to F(x)$, denoted as $\xi_n \to^d \xi$, where ξ has cumulative distribution function F. In other words, as n tends to ∞, the cumulative distribution function F_{ξ_n} of ξ_n evaluated at every x tends to F evaluated at the same point.
- Definition (Central Limit Theorem)
  - If the sequence of random variables $\{X_i\}_{i=1}^n$ is independently and identically distributed, and (for all i) $\mathbb{E}X_i = \mu$ and $var(X_i) = \sigma^2$, then the standardised sample mean $Z_n := \frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ converges in distribution to $N[0,1]$.
    - It is not technically accurate to write $\overline{X}_n \sim^a N(\mu, \sqrt{\frac{\sigma^2}{n}})$.
  - Then, Z_n is approximately distributed as $N[0,1]$, denoted as $Z_n \sim^a N[0,1]$, and the cumulative distribution function of Z_n is approximated by the standard cumulative distribution function $\mathbb{P}\{Z_n \le z\} \approx \mathbb{P}\{N[0,1] \le z\} = \phi(z)$.

Inference

Hypothesis Tests

- A hypothesis test consists of (1) a statement of the null hypothesis and alternative hypothesis, (2) a computation of the test statistic, (3) a statement of the distribution of the test statistic under the null hypothesis, (4) a statement of the decision rule, (5) a statement of the level of significance and the corresponding critical value, and (6) a statement of the outcome of the hypothesis test.
- A hypothesis test, for the purposes of this course, supposes that observations in the given sample are independently and identically distributed, and that the central limit theorem holds even where the unknown standard deviation of the estimator is replaced by its estimate, the standard error.

Null Hypothesis and Alternative Hypothesis

- The null hypothesis of a hypothesis test is $H_0 : \theta = \theta_0$, where θ is the population parameter of interest and θ_0 is the hypothesised value of θ.
- The alternative hypothesis of a hypothesis test is either $H_1 : \theta \neq \theta_0$, $H_1 : \theta > \theta_0$, or $H_1 : \theta < \theta_0$. The former is the only appropriate alternative hypothesis if there is no a priori (prior to observing the data) reason to think that θ is no greater than θ_0 or that θ is no less than θ_0. Such reason is potentially supplied by economic theory. That a deviation from the null in one direction is in fact observed is not an admissible reason for favouring a one-sided test over a two-sided test.
- $H_1 : \theta > \theta_0$ is appropriate if there is a priori reason to think that θ is no less than θ_0, and $H_1 : \theta < \theta_0$ is appropriate if there is a priori reason to think that θ is no greater than θ_0. In each of these cases, the hypothesis test is a one-sided test. Otherwise, the hypothesis test is a two-sided test.

Test Statistic

- The t-statistic is $t_n(\theta_0) := \frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)}$. In words, the t-statistic is equal to the distance of the estimate of the parameter of interest from the hypothesised value of the parameter of interest divided by the standard error of the estimator.

Distribution of the Test Statistic Under the Null

- Under the null (given a sufficiently large independently and identically distributed random sample, by the central limit theorem), $t_n(\theta_0) \to^d N[0,1]$. Regardless of whether the null or the alternative hypothesis is true, $\frac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \to^d N[0,1]$, under the null, the left hand side is equal to the t-statistic.

Decision Rule

- For a two-sided test, the appropriate decision rule is:
  - reject H_0 if $|t_n(\theta_0)| > c_\alpha$.
- The intuition for this decision rule is that if the null hypothesis is true, then estimates that are distant from the hypothesised value (and hence yield large t-statistics) of the parameter are unlikely, so such estimates constitute evidence against the null hypothesis.
- For a one-sided test, with alternative hypothesis $H_1 : \theta > \theta_0$, the appropriate decision rule is:
  - reject H_0 if $t_n(\theta_0) > c'_\alpha$.
- The intuition for this is that low estimates (which yield low t-statistics) do not constitute evidence against the null hypothesis.
- For a one-sided test, with alternative hypothesis $H_1 : \theta < \theta_0$, the appropriate decision rule is:
  - reject H_0 if $t_n(\theta_0) < -c'_\alpha$.

Critical Value

- For a two-sided t-test, given level of significance α, the corresponding critical value c_α is such that the probability of rejecting a true null (the type I error rate) is equal to α. Then,
  $\alpha = \mathbb{P}\{\text{reject } H_0|H_0\} = \mathbb{P}\{|t_n(\theta_0)| > c_\alpha\} \approx \mathbb{P}\{|N[0,1]| > c_\alpha\} = 2\phi(-c_\alpha)$.
- The type II error rate is the probability of failing to reject a false null. The power of a statistical test is the probability of rejecting a false null, and is equal to 1 minus the type II error rate.
- The type I error rate is the probability of rejecting a true null. The type I error rate is also referred to as the size or significance of a statistical test.
  - The lower the level of significance α, informally, the more "evidence" is required to reject the null, hence the lower the type I error rate and the higher the type II error rate.
  - An increase in sample size increases the power of a test.
- For a one-sided t-test, given level of significance α, the corresponding critical value c'_α is such that the probability of rejecting a true null (the type I error rate) is equal to α. Then,
  $\alpha = \mathbb{P}\{\text{reject } H_0|H_0\} = \mathbb{P}\{t_n(\theta_0) > c'_\alpha\} \approx \mathbb{P}\{N[0,1] > c'_\alpha\} = \phi(-c'_\alpha)$.

p-values

- The p-value associated with some hypothesis test is the probability of observing a test statistic at least as unfavourable to the null as the test statistic actually observed. This is equivalent to the lowest level of significance at which the null is rejected.
- When computing the p-value, explain the relevant hypothesis test up to the decision rule.

- For a two-sided test, $p := \mathbb{P}_{H_0}\{|t_n(\theta_0)| > |t|\} \approx \mathbb{P}\{|N[0,1]| > |t|\} = 2\phi(-|t|)$, where t denotes the t-statistic actually observed.
- For a one-sided test with alternative hypothesis $H_1 : \theta > \theta_0$, $p := \mathbb{P}\{t_n(\theta_0) > t\} \approx \mathbb{P}\{N[0,1] > t\} = \phi(-t)$.
- For a one-sided test with alternative hypothesis $H_1 : \theta < \theta_0$, $p := \mathbb{P}\{t_n(\theta_0) < t\} \approx \mathbb{P}\{N[0,1] < t\} = \phi(t)$.

Confidence Intervals

- The $(1 - \alpha) \times 100\%$ confidence interval for a parameter θ with estimator $\hat{\theta}_n$ is $\mathcal{C}_n := \{\theta_0 \in \mathbb{R} : H_0 : \theta = \theta_0 \text{ is accepted}\}$ under the two-sided hypothesis test with level of significance α given some estimate $\hat{\theta}_n$.
- $\mathcal{C}_n = \{\theta_0 \in \mathbb{R} : |\frac{\hat{\theta}_n - \theta_0}{se(\hat{\theta}_n)}| \leq c_\alpha\} = \{\theta_0 \in \mathbb{R} : \hat{\theta}_n - c_\alpha se(\hat{\theta}_n) \leq \theta_0 \leq \hat{\theta}_n + c_\alpha se(\hat{\theta}_n)\} = [\hat{\theta}_n - c_\alpha se(\hat{\theta}_n), \hat{\theta}_n + c_\alpha se(\hat{\theta}_n)]$.
- In approximately $(1 - \alpha) \times 100\%$ of samples, the confidence interval will contain the true value of θ. $\mathcal{C}_n$ is random while θ is not, so it is inaccurate to say that θ lies in $\mathcal{C}_n$ with probability $1 - \alpha$.
- Offer both interpretations of a confidence interval: the confidence interval contains the true value with some probability, and the probability of a type I error when rejecting a null outside the interval is $1/5/10\%$.
  - When interpreting a confidence interval write "the probability that the **random interval** $\mathcal{C}$ contains the true value is . . . ".