Econ 310 Formulas

1	Ch4	population mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	
2	Ch4	sample mean	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	
3	Ch4	population variance	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	
4	Ch 4	sample variance	$s^2 = \frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{n-1}$	
5	Ch4	shortcut for sample variance	$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]$	
6	Ch4	population standard deviation	$\sigma = \sqrt{\sigma^2}$	
7	Ch4	sample standard deviation	$s = \sqrt{s^2}$	
8	Ch4	Chebysheff's theorem	$1 - \frac{1}{k^2}$	→ No more than $\frac{1}{k^2}$ will be more than k St. Dev. away from the mean
9	Ch4	location of percentile	$L_P = (n+1)\frac{P}{100}$	
10	Ch4	sample covariance	$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$	
11	Ch4	shortcut f/ sample covariance	$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$	
12	Ch4	sample coefficient of correlation	$r = \frac{s_{xy}}{s_x s_y}$	$-1 \leq r \leq +1$
13	Ch4	population coefficient of correlation	$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$-1 \leq \rho \leq +1$
14	Ch 6	conditional prob	$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$	→ "A given B"
15	Ch6	independent events	$\left\{ \begin{array}{l} P(A\mid B) = P(A) \\ \text{or} \\ P(B\mid A) = P(B) \end{array} \right\}$	Definition of Independence
16	Ch6	multiplication rule f/ Independent events	$P(A \text{ and } B) = P(A)P(B)$	
17	Ch6	addition rule	$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$	
18	Ch6	Multiplication rule	$P(A \text{ and } B) = P(A \mid B) \times P(B)$	

Possian Dist. Properties: $E[x] = Var(x) = \mu$

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

Ch 6

also multiplication rule (joint prob. of 2 events if dependent events)

$$P(A \text{ and } B) = P(B \mid A) \times P(A)$$ → Alternate Multiplication Rule 19

Bayes Theorem!!

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$ 20

Ch 7

population mean + expected value

$$E(X) = \mu = \sum_{\text{all } x} xP(x)$$ → Expected Value = populu lation mean 21

Population Variance

$$V(X) = \sigma^2 = E[(X-\mu)^2] = \sum_{\text{all } x} (x-\mu)^2 P(x) = E(X^2) - E(X)^2$$ 23

Standard Deviation → $\sigma = \sqrt{\sigma^2}$ 24

Laws of Expected Value

$$\boxed{E(c) = c}$$ 25

$$\boxed{E(X + c) = E(X) + c}$$ 26

$$\boxed{E(cX) = cE(X)}$$ 27

Laws of Variance

$$\boxed{V(c) = 0}$$ 28

$$\boxed{V(X + c) = V(X)}$$ 29

$$\boxed{V(cX) = c^2V(X)}$$ 30

Marginal Probability

$$P(x) = \sum_y P(x,y)$$ 31

$$P(y) = \sum_x P(x,y)$$ 32

Covariance

$$COV(X,Y) = \sigma_{xy} = \sum_x \sum_y (x-\mu_x)(y-\mu_y)P(x,y) = E[(x-\mu_x)(y-\mu_y)]$$ 33

$$COV(X,Y) = \sigma_{xy} = \sum_x \sum_y xyP(x,y) - \mu_x\mu_y = E(XY) - E(X)E(Y)$$ 34

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$ – Correlation Coefficient 35

Expected Value of X+Y

$$E(X + Y) = \sum_x \sum_y (x+y)P(x,y) = E(X) + E(Y)$$ 36

Variance of X+Y

$$V(X + Y) = \sum_x \sum_y (x+y-\mu_{x+y})^2 P(x,y) = V(X) + V(Y) + 2COV(X,Y)$$ 37

Expected Value of Y given X

$$E(Y|X) = \mu_{y|x} = \sum_y yP(y|x)$$ 38

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$ } Binomial Probability Distribution 39

SD = stan. Dev = standard Deviation

Ch9 Central Limit Theorem (CLT)

3

The sampling dist. of the mean of a random sample drawn from any population is approx. normal for a sufficiently large sample ($n \geq 30$)

Ch7 Binomial Distribution
$\mu = np$ → Population mean
$\sigma^2 = np(1-p)$ → Variance
$\sigma = \sqrt{np(1-p)}$ → Stan. Dev

Poisson Distribution
$P(x) = \frac{e^{-\mu}\mu^x}{x!}$
$E(X) = V(X) = \mu$ Expected V = Var = μ

Ch8 - Continuous Probability Distributions & PDF
Uniform Probability Distribution (UPD) $f(x) = \frac{1}{b-a}$ → Area of a Continuous Normal Distribution

Ch8 Expected Value (of UPD) $E(X) = \frac{(a+b)}{2}$ EV of a Continuous Normal Dist.

Ch8 Variance (of UPD) $V(X) = \frac{(b-a)^2}{12}$ Variance of a Continuous Uniform Distribution

— Normal Dist —

Ch8 Probability Density Function of a Normal Random Variable $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$E(X) = \mu$
$V(X) = \sigma^2$ rules

$P(Z > Z_{Area}) = Area$

Ch 8 Z-Score / Standard Score for a Normal Distribution $\boxed{Z = \frac{x-\mu}{\sigma}}$ 51 → Z ≡ how many σ from the μ

Ch 9 $\boxed{\text{Sampling Dist.: Variance (Sample Variance → if population variance is known)} \longrightarrow \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}}$ 52 → Sample mean has the same expected Variance as each individual element of that sample (but divided by the sample size)

★ Ch 9 Standard Error of the Mean (Sample SD → from Population SD) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ 53 → its the stan. dev. of the sampling dist. of the statistic

Ch 9 Sampling Dist. of the Sample Mean $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 54 → if population is normal... if its not normal then if $n \geq 30$ then: $\bar{X} \overset{\cdot}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ (sample mean is approx. normal)

Ch 9 Standard Error of the Mean (SD of sample: $\frac{\sigma}{\sqrt{n}}$ • correlation factor) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \boxed{\sqrt{\frac{N-n}{N-1}}}$

population size (N)

finite population correlation factor (use when sample w/o replacement)

★1 Ch9 Sampling Distribution of a Proportion aka Sample Proportion $\boxed{\hat{p} = \frac{X}{n}}$
X — # of successes (binomial random variable)
n — sample size

★2 Ch9 Sampling Dist. of a Sample Proportion → $\hat{p} \overset{\cdot}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$

★3 Ch9 Standard Error of the proportion (aka stan. dev. of $\hat{p}$) → $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$

★ Normal Approx. to Binomial: $X \sim N(np, np(1-p))$
$\sigma^2 = np(1-p)$
$\sigma = \sqrt{np(1-p)}$
$\mu = np$

Ch9 Z-Score for a Sample Proportion $Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$ → standardizes sample proportions to a standard normal distribution

Ch9 Sampling Dist; Diff. of 2 Means from 2 independent samples $\mu_{\bar{x}-\bar{y}} = \mu_x - \mu_y$

independent random samples be drawn from each of the 2 normal pop.: $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ where $\bar{X} - \bar{Y}$ are normally distributed

$p = E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$

$\frac{p(1-p)}{n} = V(\hat{p})$

"difference between 2 means"

4

$P(\mu - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

Ch 9 Standard Error of the Δ between 2 Independent Sample Means $\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}$ 61

Standardize w/ $Z = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}$

Ch 9 Distribution of the Δ between sample means of 2 independent samples $\bar{X}-\bar{Y}\sim N\left(\mu_x-\mu_y, \frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}\right)$ 62

(just like $z = \frac{x-\mu}{\sigma}$) just now accounting f/ diff between 2 means

Ch 10 Confidence Interval (w/ Pop. SD is known) $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- consistent
- unbaised
- relatively efficient

Ch 10 ✱ Sample Size (n) given Bound on Error of Estimation (B) $n = \left(\frac{z_{\alpha/2}\sigma}{B}\right)^2$

bound on the error of estimation $B = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

Interval of Estimation $P(\bar{x}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{x}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

Ch 10/11 Standardized Test Statistic $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ error of estimation

- α = significance level
- $1-\alpha$ = confidence level

Ch 11 t-dist test-statistic → $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

$\bar{X} \mathrel{\dot\sim} N(\mu, \frac{\sigma^2}{n})$

Ch 11 t-statistic Confidence interval $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$

$Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$P(\chi^2_{1-\frac{\alpha}{2}} < \chi^2 < \chi^2_{\frac{\alpha}{2}}) = 1-\alpha$

Ch 11/12 Representation of Confidence Interval $N\left[\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$

$Z \mathrel{\dot\sim} N(0,1)$

Ch 12 Chi-Squared Dist. w/ n-1 DOF $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}$

$P\left(\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} > \sigma^2 > \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right)$ ✱ $P(-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$\bar{X}-\mu$ - error of estimation

Ch 12 Lower Confidence Limit of σ^2 f/ Chi-Squared dist $\frac{(n-1)s^2}{\chi^2_{\alpha/2}}$

✱ the Δ between $\bar{X}$ and μ lies between $\pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ w/ a probability of $1-\alpha$

Ch 12 Upper Confidence Limit of σ^2 f/ Chi-Squared dist $\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$

✱ Confidence intervals must come from a simple random sample preformed on the population

Confidence Interval for a Proportion from a Sample Proportion $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$

equal variance test. stat.

Population Variances Unknown Assumed to be equal $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\sim t_{n_1+n_2-2}$ 73

- Type I error = reject null when true
- Type II error = fail to reject the false null

Pooled Variance Estimate ↳ f/ $\bar{x}_1-\bar{x}_2$, Variances unknown, assumed to be equal $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ 74
→ aka: weighted average of two sample variances

Relations to Hypothesis tests f/Δ of 2 means

Confidence Interval f/ Pooled V.E. ↳ f/ Population Var unknown, assumed equal across 2 populations μ_1 & μ_2 $\bar{x}_1-\bar{x}_2 \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ 75

- P(type I error) = significance level (α)
- P(type II error) = (β)

test-statistic for $\mu_1-\mu_2$ when $\sigma_1^2 \neq \sigma_1^2$ (variances assumed unequal) $t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ 76

aka: unequal variance test-stat

calculate s_p^2 for each respective sample such that s_1^2, s_2^2 but $s_1^2 \neq s_2^2$ otherwise s_p^2

DOF - degrees of freedom

⑤

aka Satterwaite Formula

Ch 13 Dof (degrees of freedom) f/ test statistic f/ $\mu_1 - \mu_2$ when $\sigma_1^2 \neq \sigma_2^2$

★ always round up

$$v = (s_1^2/n_1 + s_2^2/n_2)^2/\left(\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}\right)$$ 77

Ch 13 Confidence Interval Estimator f/ 2 populations, Var unknown, Assumed Unequal $(\sigma_1^2 \neq \sigma_2^2)$

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$ 78

Ch 13 Ratio of Sample Variances

$$F = \frac{s_1^2}{s_2^2}$$ 79

Ch 13 F-Dist. Symmetry

$$F_{1-A, v_1, v_2} = \frac{1}{F_{A, v_2, v_1}}$$ 80

Ch 13 t-test of Matched Pairs Data

↳ Dof: $n_D - 1$ | CI: $\bar{x}_D \pm t_{\frac{\alpha}{2}} \frac{s_D}{\sqrt{n_D}}$

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n_D}} \sim t_{n_D - 1}$$ 81

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$ 82

$$\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$ 83

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$ 84

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ 85

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$ 86

Assume we are interested in the difference between a proportion from population 1 (p_1) and a proportion from population 2 (p_2)

- null: H_0: $p_1 - p_2 = D$
  - ★ D = hypothesized diff. between population proportions
  - ★ need $D \sim N(\mu, \sigma^2)$ (normally distributed)

We need:
- $n_1 \hat{p}_1 \geq 5$
- $n_1(1-\hat{p}_1) \geq 5$
- $n_2 \hat{p}_2 \geq 5$, $n_2(1-\hat{p}_2) \geq 5$

Ch 16 Simple linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$ ← error variable 87

$$\hat{y} = b_0 + b_1 x$$ 88

Ch 16 Least Squares line coefficients

$$b_1 = \frac{s_{xy}}{s_x^2}$$ 89

$$b_0 = \bar{y} - b_1 \bar{x}$$ 90

y = dependent variable
x = independent variable
β_0 = y-intercept
β_1 = Slope of reg. line
ε = error variable

$$SSE \equiv \sum_{i=1}^{n} \underbrace{(y_i - \hat{y}_i)^2}_{\text{sum of squared residuals}}$$ 91

$$s_\varepsilon^2 = \frac{SSE}{n-2}$$ 92

St

Ch 16 Standard error of the estimate

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$$ 93

b_0 = intercept estimator
b_1 = slope estimator
x = independent variable
$\hat{y}$ = predicted value (or fitted value) of the dependent variable

testing f/ a linear relationship

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$ 94

reg. coe.

Ch 16 test-stat f/ regression coefficient

$$t = \frac{b_1 - \beta_1}{s_{b_1}} \sim t_{n-2}$$ 95

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$ 96

Ch16 Coefficient of Determination (the proportion of the variation in the dependent variable that is explained by the variation in the independent variable)

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = 1 - \frac{SSE}{\sum (y_i - \bar{y})^2} = \frac{\text{explained variation}}{\text{Variation in } y}$$

* strength of a linear relationship

1) Find hyp.
2) Find rejection region
3) Find the test stat
4) Check if the test stat is in the rejection region
5) Reject / FTR null hyp.?