Economics 250 Formula Sheet

Descriptive Statistics

	Population	**Sample**
Mean	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
CV	$\frac{\sigma}{\mu} \times 100$	$\frac{s}{\overline{x}} \times 100$
Covariance	$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$	$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$
Correlation	$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	$r_{xy} = \frac{s_{xy}}{s_x s_y}$

Grouped Data

With K classes, with midpoints m_i and counts c_i, the sample mean is $\overline{x} = \frac{1}{n}\sum_{i=1}^{K} c_i m_i$, and the sample variance is $s^2 = \frac{1}{n-1}\sum_{i=1}^{K} c_i(m_i - \overline{x})^2$, where $n = \sum_{i=1}^{K} c_i$.

68–95–99.7 Rule

For a normal distribution 68% of the observations are in $\mu \pm 1\sigma$, 95% are in $\mu \pm 2\sigma$, and almost all (99.7%) are in $\mu \pm 3\sigma$.

Normal Distribution

For $-\infty < x < \infty$

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

with mean μ_x and standard deviation σ_x, then

$$x \sim N(\mu_x, \sigma_x)$$
$$z = \frac{x - \mu_x}{\sigma_x} \sim N(0,1)$$

Warning: Some people record the normal distribution as $N(\mu_x, \sigma_x^2)$ *i.e.* the second number in brackets is the variance rather than the standard deviation.

Uniform Distribution

For $a \leq x \leq b$

$$f(x) = \frac{1}{b-a} \qquad E(x) = \frac{a+b}{2} \qquad Var(x) = \frac{(b-a)^2}{12}$$

Random Variables

Let x be a discrete random variable, then:

$$\mu_x = E(x) = \sum_x xP(x)$$

$$\sigma_x^2 = \sum_x (x-\mu)^2 P(x)$$

The covariance of x and y is:

$$\text{cov}(x,y) = \sigma_{xy} = E(x-\mu_x)(y-\mu_y) = \sum_x \sum_y (x-\mu_x)(y-\mu_y)P(x,y)$$

The correlation between x and y is:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

For a continuous rv replace the sums by integrals.

Functions of Random Variables

If $y = a + bx$ then:

$$E(y) = \mu_y = a + b\mu_x,$$

and

$$\sigma_y^2 = b^2 \sigma_x^2.$$

If $w = cx + dy$ then:

$$E(w) = \mu_w = c\mu_x + d\mu_y,$$

and

$$\sigma_w^2 = c^2\sigma_x^2 + d^2\sigma_y^2 + 2cd\sigma_{xy}.$$

Sampling Distribution of the Sample Mean

For large samples,

$$\overline{x} \sim N\big(\mu, \frac{\sigma}{\sqrt{n}}\big)$$

Probability Theory

$$P(\overline{A}) = 1 - P(A) \quad \text{(complement rule)}$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{(addition rule)}$$
$$P(A \cap B) = P(A|B)P(B) \quad \text{(multiplication rule)}$$

A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

Marginal probabilities add entries in a joint probability table. If B_i are mutually exclusive and exhaustive events then:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

Bayes's Rule:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Binomial Distribution

For a success probability of p and sample size n:

$$P(x \text{ successes}) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},$$

and the discrete random variable x has mean

$$\mu_x = np$$

and variance

$$\sigma_x^2 = np(1-p)$$

Sampling Distribution of the Sample Proportion

For large samples,

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Inference for Means

1. For large samples, a $100(1-\alpha)\%$ CI for μ is:

(a)

$$\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

when the population variance is known to be σ^2;

(b)

$$\overline{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

when the population variance is unknown but estimated by s^2.

2. Differences in Means: Dependent Samples

$$d_i = x_{1i} - y_{2i},$$

then

$$\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

and

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \overline{d})^2}{n-1}}$$

can be used with $t_{n-1,\alpha/2}$ to form a confidence interval: $\overline{d} \pm t_{n-1,\alpha/2} s_d/\sqrt{n}$.

3. Differences in Means: Independent Samples

(variances unknown and not assumed equal):

$$s_{\overline{x}_1 - \overline{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

can be used with $t_{\nu,\alpha/2}$ and

$$\nu = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)},$$

or often degrees of freedom ν approximated by the smaller of $n_1 - 1$ and $n_2 - 1$.

Inference for Proportions

1. For large samples, a $100(1-\alpha)\%$ CI for p is:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

(Or replace $\hat{p} = x/n$ by $\tilde{p} = (x+2)/(n+4)$ when $\alpha = 1\%$, 5%, or 10%.)

2. Differences in Proportions

$$\hat{p}_1 - \hat{p}_2$$

has standard deviation:

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

which can be used with $z_{\alpha/2}$ to form a confidence interval. (Or add 1 success and 1 failure to each sample when $\alpha = 1\%$, 5%, or 10%.)

3. Testing the Hypothesis of Equal Proportions

For this test, use the pooled estimate of the common value of p_1 and p_2:

$$\hat{p}_{pool} = \frac{X_1 + X_2}{n_1 + n_2}$$

to form

$$SE_{Dp} = \sqrt{\hat{p}_{pool}(1-\hat{p}_{pool})(\frac{1}{n_1} + \frac{1}{n_2})}$$