STA258: Statistics with Applied Probability
Formula Sheet for Final Exam

Probability Distributions

Distribution	Distribution function	Support	Mean	Variance	MGF
Bernoulli	$f(x) = p^x(1-p)^{1-x}$	$x = 0, 1$	p	$p(1-p)$	$1 - p + pe^t$
Binomial	$f(x) = \binom{n}{x} p^x(1-p)^{n-x}$	$x = 0, 1, \ldots, n$	np	$np(1-p)$	$(1 - p + pe^t)^n$
Poisson	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$x = 0, 1, 2, \ldots$	λ	λ	$e^{\lambda(e^t - 1)}$
Normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	$-\infty \le x \le +\infty$	μ	σ^2	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
Gamma	$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$x > 0$	$\alpha\beta$	$\alpha\beta^2$	$(1 - \beta t)^{-\alpha},\ t < \frac{1}{\beta}$
Chi-Square	$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$	$x > 0$	k	$2k$	$(1 - 2t)^{-k/2},\ t < \frac{1}{2}$

Transformations

Let X be a normally distributed random variable with a mean μ and variance σ^2.

$$Z = \frac{x - \mu}{\sigma}$$

Let $\bar{x}$ represent the sample mean of n independent observations from a distribution with finite mean μ and finite variance σ^2. For normal populations or sufficiently large n:

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

Let $\bar{x}$ and s^2 represent the sample mean and sample variance of n independent observations from a distribution with finite mean μ and some unknown finite variance. For normal populations or sufficiently large n:

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Continuity Correction for the Normal Approximation

Binomial Probability	Continuity Correction	Normal Approximation
$P(X = x)$	$P(x - 0.5 \le X \le x + 0.5)$	$P\left(\frac{x - 0.5 - \mu}{\sigma} \le Z \le \frac{x + 0.5 - \mu}{\sigma}\right)$
$P(X \le x)$	$P(X \le x + 0.5)$	$P\left(Z \le \frac{x + 0.5 - \mu}{\sigma}\right)$
$P(X < x)$	$P(X \le x - 0.5)$	$P\left(Z \le \frac{x - 0.5 - \mu}{\sigma}\right)$
$P(X \ge x)$	$P(X \ge x - 0.5)$	$P\left(Z \ge \frac{x - 0.5 - \mu}{\sigma}\right)$
$P(X > x)$	$P(X \ge x + 0.5)$	$P\left(Z \ge \frac{x + 0.5 - \mu}{\sigma}\right)$
$P(a \le X \le b)$	$P(a - 0.5 \le X \le b + 0.5)$	$P\left(\frac{a - 0.5 - \mu}{\sigma} \le Z \le \frac{b + 0.5 - \mu}{\sigma}\right)$

Some Relationships Between Random Variables

Let $Z_1, Z_2, \ldots, Z_n \overset{\text{iid}}{\sim} N(0,1)$. Then

$$V = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Let $\bar{x}$ and S^2 represent the sample mean and variance of $X_1, X_2, \ldots, X_n$ respectively. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi_{n-1}^2$$

Let $Z \sim N(0,1)$ and let $W \sim \chi_n^2$. Then

$$T = \frac{Z}{\sqrt{\frac{W}{n}}} \sim t_n$$

Let $X_1, X_2, \ldots, X_{n_x} \overset{\text{iid}}{\sim} N(\mu_x, \sigma^2)$ with sample variance S_x^2, and let $Y_1, Y_2, \ldots, Y_{n_y} \overset{\text{iid}}{\sim} N(\mu_y, \sigma^2)$ with sample variance S_y^2. with the $X_i's$ and the $Y_j's$ all being independent. Then

$$F = \frac{\left(\frac{n_x - 1}{\sigma^2}\right) S_x^2/(n_x - 1)}{\left(\frac{n_y - 1}{\sigma^2}\right) S_y^2/(n_y - 1)} \sim F_{n_x - 1,\, n_y - 1}$$

Inference Procedures for the Population Mean μ

One-Sample

Inference Procedure for μ when σ is known

Confidence interval for μ:

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

To test $H_0 : \mu = \mu_0$, test statistic is:

$$Z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Inference Procedure for μ when σ is <u>not</u> known

Confidence interval for μ:

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

To test $H_0 : \mu = \mu_0$, test statistic is:

$$t^* = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Two-Sample

Inference Procedure for Difference between μ_1 and μ_2 when σ_1 and σ_2 are known

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x_1} - \bar{x_2}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

To test $H_0 : \mu_1 - \mu_2 = \mu_D$, test statistic is:

$$Z^* = \frac{(\bar{x_1} - \bar{x_2}) - \mu_D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Inference Procedure for Difference between μ_1 and μ_2 when σ_1 and σ_2 are <u>not</u> known

<u>*Case 1: Unequal Variances*</u>

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x_1} - \bar{x_2}) \pm t_{(\alpha/2,\ d)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where d is the smaller of $n_1 - 1$ and $n_2 - 1$.

To test $H_0 : \mu_1 - \mu_2 = \mu_D$, test statistic is:

$$t^* = \frac{(\bar{x_1} - \bar{x_2}) - \mu_D}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Case 2: Equal Variances

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(\alpha/2,\, n_1+n_2-2)} s_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

To test $H_0 : \mu_1 - \mu_2 = \mu_D$, test statistic is:

$$t^* = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_D}{s_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Pooled Sample Variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Inference Procedures for the Population Proportion p

One-Sample

Confidence interval for p:

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

To test $H_0 : p = p_0$, test statistic is:

$$Z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Two-Sample

Confidence interval for $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

To test $H_0 : p_1 - p_2 = p_D$, test statistic is:

$$Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - p_D}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

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

ANOVA

For an analysis of k groups:

Source	**DF**	**SS**	**MS**	**F**
Treatment	$k - 1$	$SSTrt$	$MSTrt = \frac{SSTrt}{k - 1}$	$F = \frac{MSTrt}{MSE}$
Error	$n - k$	SSE	$MSE = \frac{SSE}{n - k}$	—
Total	$n - 1$	$SSTotal$	—	—

Simple Linear Regression

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2\right) - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \left(\sum_{i=1}^{n} y_i^2\right) - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \left(\sum_{i=1}^{n} x_i y_i\right) - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \bar{y} = \hat{\beta}_1 \bar{x} + \hat{\beta}_0$$

Sum Square of Regression (SSR) $= \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2$

Sum Square Error (SSE) $= \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = S_{yy} - \hat{\beta}_1 S_{xy}$

Sum Square Total $SSTotal$ $= \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$

Coefficient of Correlation $r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

Coefficient of Determination $r^2 = \dfrac{SSR}{SSTotal} = 1 - \dfrac{SSE}{SS_{yy}}$

Residuals $e_i = y_i - \hat{y}_i$

Estimate of Variance of Random Error $s^2 = \dfrac{\sum_{i=1}^{n} e_i^2}{n-2} = \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \dfrac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}$

Inference Procedures on the Slope β_1

Confidence interval for β_1:

$$\hat{\beta}_1 \pm t_{(\alpha/2, n-2)} \frac{s}{\sqrt{S_{xx}}}$$

To test $H_0 : \beta_1 = 0$, test statistic is:

$$t^* = \frac{\hat{\beta}_1}{s/\sqrt{S_{xx}}}$$