Formula Sheet Mathematical Statistics

Probability Theory

$E(X+Y)=E(X)+E(Y)$ $\quad E(X-Y)=E(X)-E(Y)$ $\quad E(aX+b)=aE(X)+b$

$var(X)=E(X^2)-(EX)^2$ $\quad var(aX+b)=a^2var(X)$

If X and Y are independent: $\quad var(X+Y)=var(X)+var(Y)$, $\quad var(X-Y)=var(X)+var(Y)$

$var(T)=E(var(T|V))+var(E(T|V))$

Distribution	**Probability/Density function**	**Range**	$\boldsymbol{E(X)}$	$\boldsymbol{var(X)}$
Binomial (n,p)	$\binom{n}{x}p^x(1-p)^{n-x}$	$0,1,2,\ldots,n$	np	$np(1-p)$
Poisson (μ)	$e^{-\mu}\mu^x/x!$	$0,1,2,\ldots$	μ	μ
Uniform on (a,b)	$1/(b-a)$	$a<x<b$	$(a+b)/2$	$(b-a)^2/12$
Exponential (λ)	$\lambda\exp(-\lambda x)$	$x\geq 0$	$1/\lambda$	$1/\lambda^2$
Gamma (α,β)	$x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)/(\Gamma(\alpha)\beta^\alpha)$	$x>0$	$\alpha\times\beta$	$\alpha\times\beta^2$
Chi-square (χ^2_f)	is the Gamma distribution with $\alpha=f/2$ and $\beta=2$			

Testing procedure in 8 steps

1. Give a probability model of the observed values (the statistical assumptions).
2. State the null hypothesis and the alternative hypothesis, using parameters in the model.
3. Give the proper test statistic.
4. State the distribution of the test statistic if H_0 is true.
5. Compute (give) the observed value of the test statistic.
6. State the test and **a.** Determine the rejection region or **b.** Compute the p-value.
7. State your statistical conclusion: reject or fail to reject H_0 at the given significance level.
8. Draw the conclusion in words.

Bounds for Confidence Intervals:

* $\hat{p}\pm c\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
* $\overline{X}\pm c\dfrac{S}{\sqrt{n}}$ and $\left(\dfrac{(n-1)S^2}{c_2},\dfrac{(n-1)S^2}{c_1}\right)$
* $\overline{X}-\overline{Y}\pm c\sqrt{S^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}$, with $S^2=\dfrac{n_1-1}{n_1+n_2-2}S_X^2+\dfrac{n_2-1}{n_1+n_2-2}S_Y^2$ or: $\overline{X}-\overline{Y}\pm c\sqrt{\dfrac{S_X^2}{n_1}+\dfrac{S_Y^2}{n_2}}$
* $\hat{p}_1-\hat{p}_2\pm c\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
* (regression) $\hat{\beta}_i\pm c\times se(\hat{\beta}_i)$ and $\hat{\beta}_0+\hat{\beta}_1x_0\pm cS\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}$, with $\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{x}$,

  $S_{xy}=\sum_i(x_i-\overline{x})(y_i-\overline{y})$, $\hat{\beta}_1=\frac{S_{xY}}{S_{xx}}$, $se(\hat{\beta}_1)=\frac{S}{\sqrt{S_{xx}}}$ and $S^2=\frac{1}{n-k-1}\sum_{i=1}^n\left(Y_i-\hat{Y}_i\right)^2$.

Prediction intervals: $\overline{X} \pm c\sqrt{S^2\left(1+\frac{1}{n}\right)}$

(regression) $\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm cS\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}$

Test statistics

* X (number of successes for a binomial situation)
* $T = \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$ and S^2
* $T = \dfrac{(\overline{X}-\overline{Y})-\Delta_0}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, with $S^2 = \dfrac{n_1-1}{n_1+n_2-2}S_X^2 + \dfrac{n_2-1}{n_1+n_2-2}S_Y^2$ or: $Z = \dfrac{(\overline{X}-\overline{Y})-\Delta_0}{\sqrt{\frac{S_X^2}{n_1}+\frac{S_Y^2}{n_2}}}$
* $F = \dfrac{S_X^2}{S_Y^2}$
* $Z = \dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, with $\hat{p} = \dfrac{X_1+X_2}{n_1+n_2}$
* (regression) $T = \hat{\beta}_i / se(\hat{\beta}_i)$ and $F = \dfrac{SS_{Regr}/k}{SS_{Error}/(n-k-1)}$

Adjusted coefficient of determination: $R_{adj}^2 = 1 - \frac{n-1}{n-k-1} \times \frac{SS_{Error}}{SS_{Total}}$

Analysis of categorical variables

* 1 row and k columns: $\chi^2 = \displaystyle\sum_{i=1}^{k}\frac{(N_i - E_0N_i)^2}{E_0N_i}$ $(df = k-1)$
* $r \times c$-cross table: $\chi^2 = \displaystyle\sum_{j=1}^{c}\sum_{i=1}^{r}\frac{\left(N_{ij}-\hat{E}_0N_{ij}\right)^2}{\hat{E}_0N_{ij}}$, with $\hat{E}_0N_{ij} = \dfrac{\text{row total} \times \text{column total}}{n}$

  and $df = (r-1)(c-1)$.

Non-parametric tests

* Sign test: $X \sim B\left(n, \frac{1}{2}\right)$ under H_0
* Wilcoxon`s Rank sum test: $W = \displaystyle\sum_{i=1}^{n_1} R(X_i)$,

  under H_0 with: $E(W) = \dfrac{1}{2}n_1(N+1)$ and $var(W) = \dfrac{1}{12}n_1n_2(N+1)$

Test on the normal distribution

* Shapiro – Wilk`s test statistic: $W = \dfrac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$